

Research Paper

Smad2 Cooperating with TGIF2 Contributes to EMT and Cancer Stem Cells Properties in Pancreatic Cancer via Co-Targeting SOX2

Fuqiang Zu^{1#}, ChuanPing Chen^{2#}, Qilong Geng^{3#}, Haoyu Li³, Boyuan Chan³, Guopei Luo⁴, Mengcheng Wu⁵, Matthias Ilmer⁶, Bernhard W Renz⁶, Lutterodt Bentum-Ennin⁷, Hao Gu⁷✉, Weiwei Sheng⁵✉

1. Department of General Surgery, the Second Affiliated Hospital of Anhui Medical University, Hefei, 230601, China.
2. Department of Pharmacy, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui 230022, China.
3. Department of Clinical Medicine, Anhui Medical University, Hefei, 230022, China.
4. Department of Pancreatic Surgery, Fudan University Shanghai Cancer Center, Shanghai, 200032, China.
5. Department of General Surgery, the First Affiliated Hospital of Anhui Medical University, Hefei, 230022, China.
6. German Cancer Consortium (DKTK), Partner Site Munich and German Cancer Research Center (DKFZ), Heidelberg, Germany.
7. Department of Immunology, School of Basic Medical Sciences, Anhui Medical University, Hefei 230032, China.

The authors contribute equally to this work.

✉ Corresponding authors: guhao@ahmu.edu.cn (H.G.) and cmu_wwsheng@sina.com (W. S.).

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). See <https://ivyspring.com/terms> for full terms and conditions.

Received: 2024.08.15; Accepted: 2024.11.18; Published: 2025.01.01

Abstract

The underlying mechanisms between cancer stem cells (CSC) and epithelial-mesenchymal transition (EMT) in pancreatic cancer (PC) remain unclear. In this study, we identified TGIF2 as a target gene of CSC using sncRNA and machine learning. TGIF2 is closely related to the expression of SOX2, EGFR, and E-cadherin, indicating poor prognosis. Mechanistically, TGIF2 promoted the EMT phenotype and CSC properties following the activation of SOX2, Slug, CD44, and ERGF/MAPK signaling, which were rescued by SOX2 silencing. TGIF2 silencing contributes to the opposite phenotype via SOX2. Notably, Smad2 cooperates with TGIF2 to co-regulate the SOX2 promoter, which in turn promotes EMT and CSC signaling by transactivating Slug and EGFR, respectively. The transactivation of EGFR/MAPK signaling by SOX2 promotes TGIF2 nuclear translocation, forming a positive feedback loop *in vitro*. Moreover, the interaction of TGIF2 and SOX2 with EGFR inhibitors promoted subcutaneous tumors and liver metastasis *in vivo*. Thus, the TGIF2/SOX2 axis contributes to CSC, EMT, and chemoresistance, providing a promising target for PC therapy.

Keywords: TGIF2, SOX2, EMT, cancer stem cells, pancreatic cancer

Introduction

Pancreatic cancer (PC) is a lethal tumor with approximately 11% surviving over 5 years [1, 2] owing to drug resistance, local progression, and distant metastasis. Cancer stem cells (CSC) and epithelial-mesenchymal transition (EMT) play significant roles in tumor spread and distant metastasis [3]. CSC act as a self-renewing pool of tumorigenic cells, which can lead to drug resistance, recurrence, metastasis, and aggressiveness of cancer [4, 5]. The active tumorigenesis of CSC also paves the way for EMT, which has been demonstrated in PC chemoresistance and metastasis [6-8]. For example,

zinc finger E-box binding homeobox 1 (ZEB1) positively regulates CD44 to regulate CSC properties [9]. Our previous studies identified several critical oncogenes as potential EMT promoters in CSC [10-13]. It is essential to elaborate on the molecular interconnection between CSC and EMT and their regulation.

Previous Single-cell RNA (scRNA) analyses have revealed the composition and interactions of CSC in the tumor microenvironment. Highly expressed genes associated with CSC have been identified that interact with EMT, oxidative stress, proteasomes, and

immunotherapy [14-16]. However, comprehensive crosstalk between CSC signatures and the EMT landscape remains limited [17, 18]. Here, we investigated the cellular composition and transcriptome profiles of CSC using sncRNAs and machine learning [19]. A novel machine learning framework that incorporates 10 algorithms was established, composed of six genes, and indicated excellent prognosis prediction. Among these, TGIF2 has been identified as the most significant factor at the transcriptional intersection of CSC and EMT. Therefore, we explored its role in maintaining CSC pluripotency and promoting EMT in PC.

TGF- β -induced factor homeobox (TGIF) is a family of transcription factors, containing multiple members, such as TGIF1, TGIF2, TGIF2LX, and TGIF2LY [20, 21]. As a DNA-binding transcription factor, TGIF2 is important for regulating many crucial developmental processes, including cell proliferation and differentiation, and is involved in the development of several cancers, including melanoma, osteosarcoma, oral squamous cell carcinoma, glioblastomas, and lung cancer [22-25]. However, its potential role in PC has not been studied. In this study, the TGIF2/SOX2 transcriptional axis promoted PC development by facilitating EMT and maintaining CSC pluripotency. This study identified a novel transcriptional regulatory network for revealing the molecular mechanisms underlying the malignant biology of PC.

Materials and methods

Patients and samples

All three surgical resection samples diagnosed with PC were collected at the First Affiliated Hospital of Anhui Medical University. Briefly, the samples were chopped on ice, dissociated in a collagenase digestion solution, and oscillated at 50 rpm. The cells were filtered using a strainer and the survival rate was measured. For Single-cell RNA sequencing (scRNA), the single-cell suspension was adjusted to 1×10^5 cells/mL and sequenced by $10 \times$ Genomics on an Illumina NovaSeq instrument with 150bp paired-end reads. Raw sequencing data were converted into FASTQ files using Illumina bcl2fastq, and aligned to the human genome reference sequence (GRCH38). The SeekSoul® Tools was used to sample demultiplexing, barcode processing, and gene counting to generate a digital gene-cell matrix. The gene expression matrix was processed and analyzed using the Seurat package (version 5.0.3) [26].

ScRNA and bulk analysis

Seurat was used for dimensionality reduction,

clustering, and visualization [27]. Filtered gene expression was examined at 300-10000, and the mitochondrial percentage was set at lower than 20. Data were normalized and identified for variable genes, clustered with a resolution of 0.5, and visualized using t-SNE. Cluster markers were identified using FindAllMarkers and were re-annotated for further analysis.

The transcriptional data and clinical information were downloaded from four public databases: E-MTAB-6134, TCGA, GSE21501 and GSE62452 [28, 29]. Further, 335 stem genes were identified using the StemChecker [30]. Briefly, the PC stemness scores of PC was estimated using single-sample gene set enrichment analysis (ssGSEA) via GSVA packages (V 1.50.1) [31] and categorized into CSC_high and CSC_low groups. For machine learning, we validated the CSC index of PC prognosis in the E-MTAB-6134 database (training cohort) and TCGA database (validation cohort) and selected the best model of the machine learning algorithms [32]. The gene coefficients of the model were calculated by multivariate Cox regression analysis, and the final CSC-index = $\sum (\text{Coefi} \times \text{Exp})$ was obtained. We compared the OS of the groups via Kaplan-Meier analysis, evaluated the predictive performance through receiver operating characteristic (ROC) curve analysis, and validated independent prognostic factors by univariate and multivariate Cox regression analysis using survival packages (V 3.5-8).

Functional analysis was performed using clusterProfiler (V 4.10.1) and org. Hs. eg. db packages (V 3.18.0) using GSEA analysis [33]. Differences in the enriched HALLMARK pathways were compared between patients with high and low CSC-index.

Target genes of TGIF2 were predicted using the TFDB 3.0 database [34]. Common genes were identified and the score was calculated using TGIF2. ATAC-seq data were downloaded from the GSE213394 database [35]. The chromatin openness of these genes was also determined. Moreover, we downloaded EMT genes from the MSigDB database and evaluated the EMT signatures of PC and Gastric cancer (GC) from previous studies [36-38].

Tissue specimens and PC cell lines

Here, 88 cases of paraffin-embedded PC specimens and 56 adjacent pancreatic specimens were collected for the IHC assays. Twenty patients with PC and their adjacent tissues were randomly selected and used for PCR assays. All samples were obtained from patients who underwent radical surgery without neoadjuvant therapy, and the diagnosis was confirmed by a pathologist. All patients provided written informed consent and the study was

approved by the Ethics Committee of the Second Affiliated Hospital of Anhui Medical University (2023558). Cells were cultured as described previously [10-13]. Five cell lines (AsPC-1, BxPC-3, PANC-1, Miapaca-2, and HEK-293) were purchased from the National Cell Culture in Shanghai using recommended growth media containing 10% fetal calf serum (Hyclone, Logan, UT, USA) [10].

Immunohistochemistry assays

Immunohistochemistry was performed as previously described [10-13]. The following primary antibodies were used and incubated overnight: anti-TGIF2 (Proteintech, Chicago, IL, USA, Cat #11522-1-AP, dilution 1:100), SOX2 (Proteintech, Cat #66411-1-Ig, dilution 1:100), EGFR (Proteintech, Cat #66455-1-Ig, 1:500 dilution), and anti-E-cad (Abcam, Cambridge, UK, Cat #ab40772, dilution 1:200).

Western blot analysis

Western blot (WB) was performed as previously described [10-13]. Nucleoplasmic proteins were extracted from the cells using a special lysis buffer (BB-36021, BestBio) following the manufacturer's instructions. The antibodies used were TGIF2 (Proteintech, dilution at 1:500), SOX2 (Proteintech, dilution at 1:1000), E-cad (Abcam, dilution at 1:1000), N-cadherin (N-cad, Proteintech, Cat # 22018-1-AP, dilution at 1:1000), Vimentin (Proteintech, Cat # 10366-1-AP, dilution at 1:2000), Snail1 (Proteintech, Cat #13099-1-AP, dilution at 1:500), Slug (Proteintech, Cat #12129-1-AP dilution at 1:500), CD133 (Proteintech, Cat #18470-1-AP, dilution at 1:1000), CD44 (Proteintech, Cat #15675-1-AP, dilution at 1:1000), EGFR (Proteintech, dilution at 1:1000), p-EGFR (Abcam, Cat # ab97613, dilution at 1:500), p-ERK (Cell Signaling Technology, Beverly, USA, Cat #137F5, dilution at 1:1000), Smad2 (Cell Signaling Technology, Cat #5339, dilution at 1:1000), and GAPDH (Proteintech, Cat #60004-1-Ig, 1:3000).

Co-Immunoprecipitation (Co-IP)

Based on our previous study [10-13], the lysate of BxPC-3 was extracted using RIPA lysis buffer. TGIF2 antibodies mixed with magnetic beads were incubated together on a rotator overnight in 4°C. The final immunocomplex was stripped by boiling in WB loading buffer. The input and IgG panels were used in parallel as positive and negative controls, respectively.

qRT-PCR

qRT-PCR was performed as described previously [10-12]. mRNA levels in tissues and cell lines were estimated using a LightCycler kit for the

fast qRT-PCR system. The primers used are listed in Table S1. The quality of PCR products was monitored using post-PCR melt-curve analysis. The expression level of these target genes was calculated by the $-\Delta\Delta C_t$ method.

siRNA and lentivirus vector-mediated TGIF2/SOX2 overexpression

The siRNA and shRNA sequences for TGIF2 and SOX2 were summarized in Table S2, which were synthesized by GenePharma Co., Ltd. (Shanghai, China). shRNA-mediated TGIF2 and SOX2 silencing and lentivirus vector (GV492 and CV186)-mediated TGIF2, SOX2, and Smad2 overexpression (TGIF2-OE, SOX2-OE, and Smad2-OE, respectively) were purchased from GeneChem (Shanghai, China). siRNAs and plasmids were mixed with Oligofectamine 3000 (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions.

EMT model construction

EMT model construction has been described in our previous studies [10-13]. Briefly, cells were induced by TGF β and calculated the percentage of residual epithelial cells to the whole cell area. We validated EMT phenotypes by western blotting and detected cell mobility in EMT-stimulated cells using invasion and migration assays.

MTT and migration assays

Based on our previous studies [10-13], MTT assay was used to detect cell proliferation and gemcitabine resistance. Briefly, transfected cells were incubated for growth with or without Erlotinib treatment (5 μ M for 2 h twice), different concentrations of Gemcitabine, respectively. Then, we treated the cells with MTT and DMSO successively and measured the absorbance at 570 nm using an ELISA microtiter plate reader (Bio-Rad 680, California, USA). For the migration assays, transfected cells were implanted into the upper chamber with Matrigel or serum-free medium. Medium supplemented with 10% FBS was added to the bottom as a stimulus. The migrated cells were fixed, co-stained with crystal violet, and counted in five random fields [10].

CSC culture

According to previous studies [39], transfected PANC-1 and BxPC-3 cell lines were cultured in serum-free DMEM-F12 medium containing 0.4% BSA (Sigma), N-2 Plus media (Gibco), B-27 (Gibco), FGF (10ng/ml) and EGF (20ng/ml, for Kras wildtype BxPC-3 cells only) (Preprotech, Rocky Hill, NJ, USA) at a density of 15000 cells/ml in low-attachment dishes (Corning, NY, USA). Round aggregates containing six or more cells were considered as

'spheres' for quantification purposes. The secondary spheres formed following a 1-2-week incubation were counted and photographed.

Organoids culture from pancreatic acinar cells

Wild-type C57BL/6 mice and Pdxcre; LSL-KrasG12D (KC) mice were a gift from Timothy Wang's laboratory at Columbia University and were kept in the animal experimental department of Anhui Medical University under SPF conditions. The pancreatic tissue was rinsed in cold PBS and chopped into pieces, mixed with collagenase and stirred at 37°C for 20 min. The cells were resuspended in PBS, incubated in Pancreatic Medium, mixed with Matrigel, and seeded in a 24 well plate for 3–5 days. Once the organoids were formed, the whole organoids released from the 3D Matrigel were transfected with TGIF2 and SOX2 siRNAs to achieve the highest transfection efficiency, according to protocols described in a previous study [40]. After transfection, the organoids were cultured for another seven days, counted, and photographed.

Immunofluorescence (IF) staining

As detailed in our previous studies [10-13], transfected cells were implanted with or without erlotinib treatment, fixed in paraformaldehyde, permeabilized with Triton, incubated with BSA, and stained with primary antibodies against TGIF2 and SOX2 overnight. The plates were then incubated with secondary antibodies of different origins (rabbit FITC for TGIF2 and mouse TRITC for SOX2). Hoechst33258 was used for nuclear visualization. IF was performed in triplicates.

Chromatin immunoprecipitation (ChIP) assay

The chip assay was performed according to manufacturer's instructions (9003; Cell Signaling Technology). Briefly, cultured cells were lysed in protease inhibitor buffer and sonicated to extract approximately 150–800bp chromatin fragments. Following dilution with an IP dilution buffer, the lysates were incubated at 4°C overnight with TGIF2, Smad2 or SOX2 antibodies. The antibody-bound chromatin complex was precipitated using protein A/G magnetic beads and salmon sperm DNA for 4 h at room temperature. Finally, DNA was isolated from the immunoprecipitated chromatin using a DNA Elution Buffer. The corresponding PCR-amplified primer pairs flanking the consensus-binding sites in the SOX2, Slug, and EGFR promoters are shown in Table S1. All the PCR was carried out for 35 cycles with the primers annealed at 58°C, and the PCR products were resolved on a 2% agarose gel in TBE

buffer for the final gel imaging (Bio-Rad).

Dual-luciferase reporter assay

SOX2-OE transfected HEK-293 cells were seeded in 24-well plates and co-transfected with a pGL3-Basic-SOX2 promoter plasmid (SOX2-WT), the corresponding mutant plasmid (SOX2-Mutant, SOX2-MT) for TGIF2, and the corresponding mutant plasmid (SOX2-Mutant2, SOX2-MT2) for Smad2. For the dual-luciferase assay, wild-type (wt) and mutant (mut) EGFR promoter plasmids were constructed. HEK293 cells in SOX2-OE and scramble groups were seeded and co-transfected with 0.5–1 µg of EGFR-WT and EGFR-Mut promoter plasmid and empty vector using Lipofectamine 3000. After incubation for 48 h, the luciferase activity was measured using the Dual-Luciferase® Assay Kit (Cat. #E2920, Promega). The results were repeated at least three times.

In vivo xenograft model

All animal experiments were approved by the Animal Care Committee of Anhui Medical University (LISC20231510). Twelve nude mice were randomly divided into four groups: scramble, TGIF2-OE, TGIF2-OE/sh-SOX2, and TGIF2-OE plus erlotinib, and injected into the axillae using transfected PANC-1 cells. According to a previous study [12], the TGIF2-OE plus erlotinib group was orally administered 100 mg/kg/day erlotinib for 5 days per week, whereas the other groups were orally administered 1% DMSO as a control. All mice were sacrificed after 3 weeks. Tumor volumes were calculated using the following formula: length × width × height × 0.5 in mm³. Moreover, BxPC-3 cells were injected into the spleens of 15 nude mice to construct a distant liver metastasis model (n=5 in each group), which was assessed by the number of liver metastases and liver weight/body weight ratio [41, 42]. All the mice were euthanized 4 weeks later. Tissue samples were extracted and used for hematoxylin, eosin (HE) and IHC staining.

Statistical analysis

All statistical analyses were performed using the SPSS software (version 21.0; R Software 4.3.1). Significant differences were analyzed using Welch's t-test for two-group comparisons, one-way ANOVA for multiple group comparisons, two-way ANOVA with Tukey's test or Sidak's test for interactions, log-rank test for survival data, Pearson's correlation analysis, chi-square test, and Kruskal-Wallis test for nonparametric data analysis [43]. Statistical significance was set at P <0.05 was considered significant.

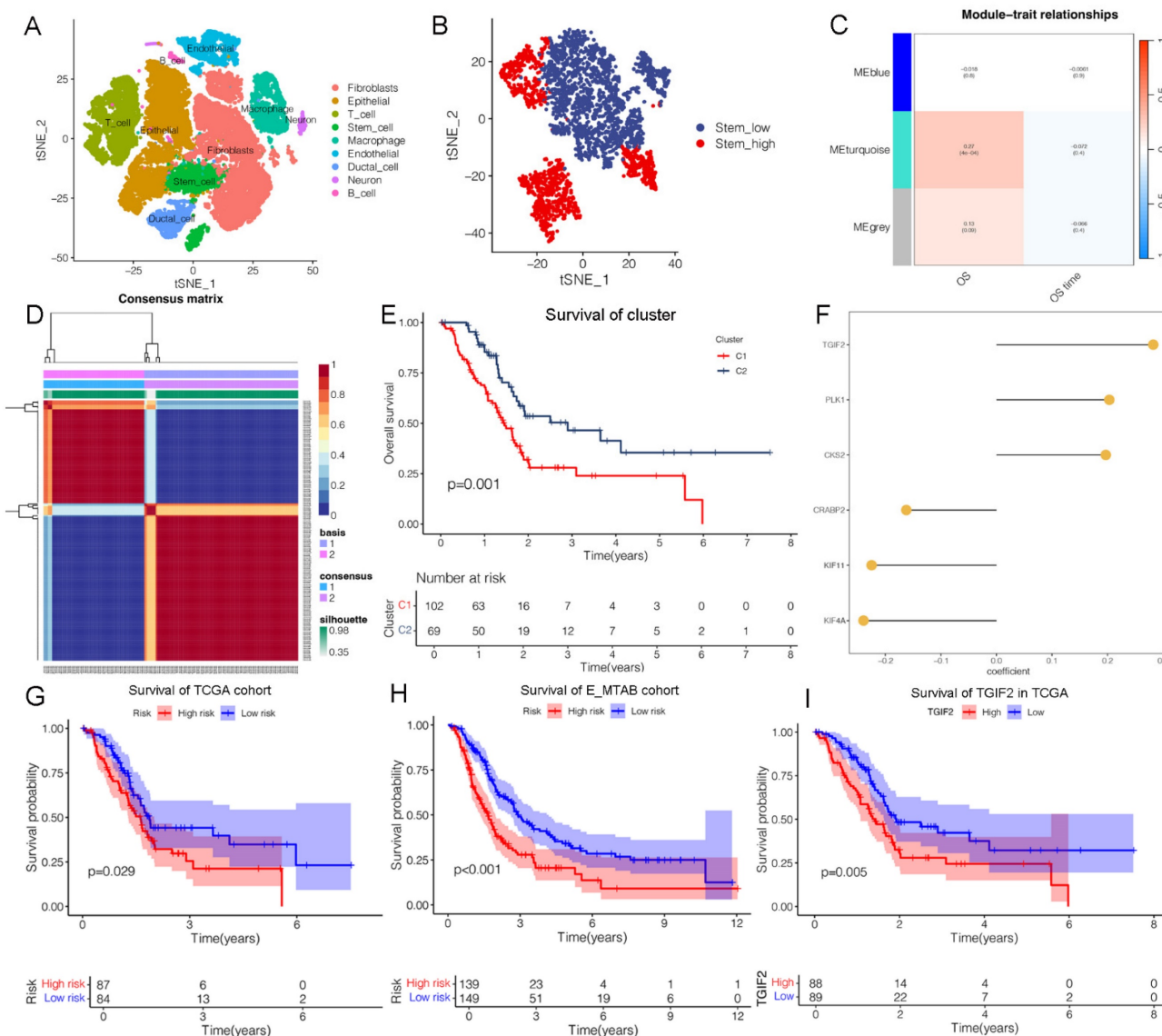


Figure 1. Identification of TGIF2 as a CSC gene. **A.** Nine clusters of tSNE distribution in scRNA analysis. **B.** Cell_type of tSNE plotter in stem subcluster. **C.** Correlation map of 3 modules with OS time and status. **D.** the Optimal grouping of risk scores via NFM algorithm. **E.** Survival difference between stem subgroups. **F.** The qualified CSC genes in risk model. **G and H.** The survival difference between the high and low-risk in TCGA cohort and E-MTAB cohort. **I.** The overall survival of TGIF2 in TCGA database by Kaplan Meier database.

Results

Identification of TGIF2 as a targeted gene of CSC and EMT

Based on marker annotation, we preliminarily identified nine cell subtypes, including fibroblasts, epithelial cells, T cells, stem cells, macrophages, endothelial cells, ductal cells, neurons, and B cells (Figure 1A, Figure S1A). Stem cells were subset and re-clustered into stem_high and stem_low groups using ssGSEA (Figure 1B). In total, 335 stem genes were identified and selected for further analysis (Figure S1E). Three modules were identified in the stemness signature using the WGCNA network (Figure 1C). The turquoise module showed the strongest positive correlation with OS (ME = 0.27, P =

4e-4), from which the genes were validated for further analyses (Figure S1E). We then calculated the risk scores of the genes and divided them into two categories (Figure 1D, Figure S1C-D), from which 574 subtype genes were identified (Figure S1E). Patients with subtype C2 had a significantly longer OS than those with subtype C1 (Figure 1E), which was well distinguished in the PCA plot. Thirty CSC target genes were identified (Figure S1E). The heatmap shows significant differences in the expression levels of the subtypes (Figure S1H).

We developed a machine learning model and calculated the C-index of 45 predictive models (Figure S2A). The two algorithms with the highest C-indices were the EMTAB and TCGA databases. Notably, six genes were identified in the CSC model (Figure 1F),

including TGIF2, PLK1, CKS2, CRABP2, KIF11, and KIF4A, and were validated using the E-MTAB and TCGA databases (Figure S2B-G). Our findings revealed a valuable AUC for predicting the 1-, 3-, and 5-year survival rates of patients in both the E-MTAB and TCGA databases (Figure S2B-C). The CSC model showed significance only in the E-MTAB cohort (Figure S2D) compared to the TCGA cohort (Figure S2E). Additionally, a prognostic nomogram indicated good prediction of 1-, 3-, and 5-year overall survival (Figure S2F-G). Calibration curves confirmed that the nomogram outperformed the other predictors. A significant difference in survival was observed between the high- and low-risk groups (Figure 1G-H).

There exist enriched pathways of E-MTAB cohort (Figure S3A), including TGF- β , mTORC1 and Kras signaling, which are closely related to MAPK activating. An interaction between CSC and EMT was predicted, which is consistent with our hypothesis. Additionally, we validated that the CSC model was enriched in KRAS and EMT signaling in TCGA cohort (Figure S3B-C). A significant difference in survival was observed between the high- and low-risk groups based on KRAS signaling and EMT (Figure S3D-G). Correlation analysis revealed that TGIF2 expression was positively associated with Snail2 (Slug) and EGFR expression in PC tissues (Figure S3H). Therefore, we identified TGIF2 as having the highest expression, indicating poor survival (Figure 1I), and selected it for further analysis.

TGIF2 is identified as a transcript factor previously [24, 25]. Based on the TFDB database, eight stem cell genes were identified as target genes (Figure S4A and Table S4), including MYC, SOX2, TP53, LMNB1, YY1, OTX2, CBX3, and HDAC1, but not OCT4 or NANOG. Next, we evaluated the correlation between TGIF2 and these genes (Figure S4D-I). SOX2 expression was strongly correlated with TGIF2 expression ($r=0.3$, $P<0.05$; Figure S4E). Moreover, SMAD3 binds to TGIF2 to increase SOX levels by forming a transcriptional complex [44], which further indicates a specific interaction between TGIF2 and SOX2.

Consistency exists between PC and GC in their EMT response. We identified 11 genes and compared their EMT signatures between PC and GC. The EMT signatures of five genes were both changed (RUNX2, SNAI2, ZEB1, TEAD1, and NUA1), while the others were not, which indicated consistency between PC and GC in the EMT response (Figure S4B & C).

Overexpression of TGIF2 and SOX2 indicated advanced clinical stage and dismal prognosis

SOX2, EGFR, and E-cadherin are significant regulators of CSCs and EMT [45, 46]. In this study,

both TGIF2 and SOX2 were localized in the cytoplasm and nucleus (Figure 2A), whereas EGFR showed membrane and cytoplasmic expression (Figure 2B). According to a previous study [12], an abnormal expression pattern can be classified as absent, cytoplasmic, or heterogeneous (Figure 2B). As shown in Figure 2A, the expression levels of TGIF2 and SOX2 were much higher in PC tissues than in healthy tissues (51.1% vs. 23.2%, $P<0.01$; and 39.8% vs. 3.6%, respectively). $P<0.001$). Correlation analysis showed that TGIF2 was positively associated with SOX2 and EGFR expression but negatively associated with E-cadherin expression (Table 1). High expression of TGIF2, SOX2, EGFR, and E-cadherin was observed in #21 PC specimen, #37 (Figure 2B-C).

Table 1. The relationship among TGIF2, SOX2, EGFR and E-cad expression in 88 cases of clinical PC samples

Parameters	TGIF2		<i>r</i>	<i>P</i>
	Low	High		
SOX2	Low	33	0.330	0.002
	High	10		
EGFR	Negative	36	0.263	0.014
	Positive	7		
E-cad	Abnormal	21	-0.227	0.033
	Normal	22		

Similarly, high levels of TGIF2 and SOX2 were also observed in randomly selected 20 cases of PC tissue samples (Table S3) compared with those in the adjacent pancreas, which was further verified according to the GEPIA database (Figure 2D-E and Figure S1F-G). The relationships between TGIF2, SOX2, and clinical traits are shown in Table 2-3. Briefly, TGIF2 overexpression was closely associated with T stage ($P=0.011$), lymph node metastasis ($P=0.032$), and TNM stage ($P=0.028$), whereas SOX2 overexpression was associated with tumor size ($P=0.029$) and T stage ($P=0.026$). High TGIF2 expression was associated with poor OS ($P=0.002$) (Figure 2F), whereas SOX2 expression was not significantly associated (Figure 2G). The combined expression of TGIF2 and SOX2 contributed to worse prognosis (Figure 2H).

Interaction of TGIF2/SOX2 promotes EMT of PC

WB and qRT-PCR showed that TGIF2 and SOX2 were more highly expressed in Miapaca-2 and BxPC-3 cells than in other cells (Figure 3A-B, Figure S5A). PANC-1 and BxPC-3 cells were used for TGIF2 overexpression (TGIF2-OE) and silencing (si TGIF2) separately. TGIF2-OE led to the activation of Vimentin and Slug and the inactivation of E-cadherin and SOX2 expression. SOX2 silencing partially downregulates N-Cadherin, Vimentin and Slug

expression and upregulates E-cadherin expression. Moreover, the loss of E-cadherin expression caused by TGIF2-OE was rescued by silencing SOX2 (Figure 3C, Figure S5B). Conversely, EMT signaling blunted by

silencing TGIF2 was rescued by SOX2 overexpression (SOX2-OE) (Figure 3D, Figure S5C), suggesting an interaction effect of TGIF2/SOX2 on EMT signaling.

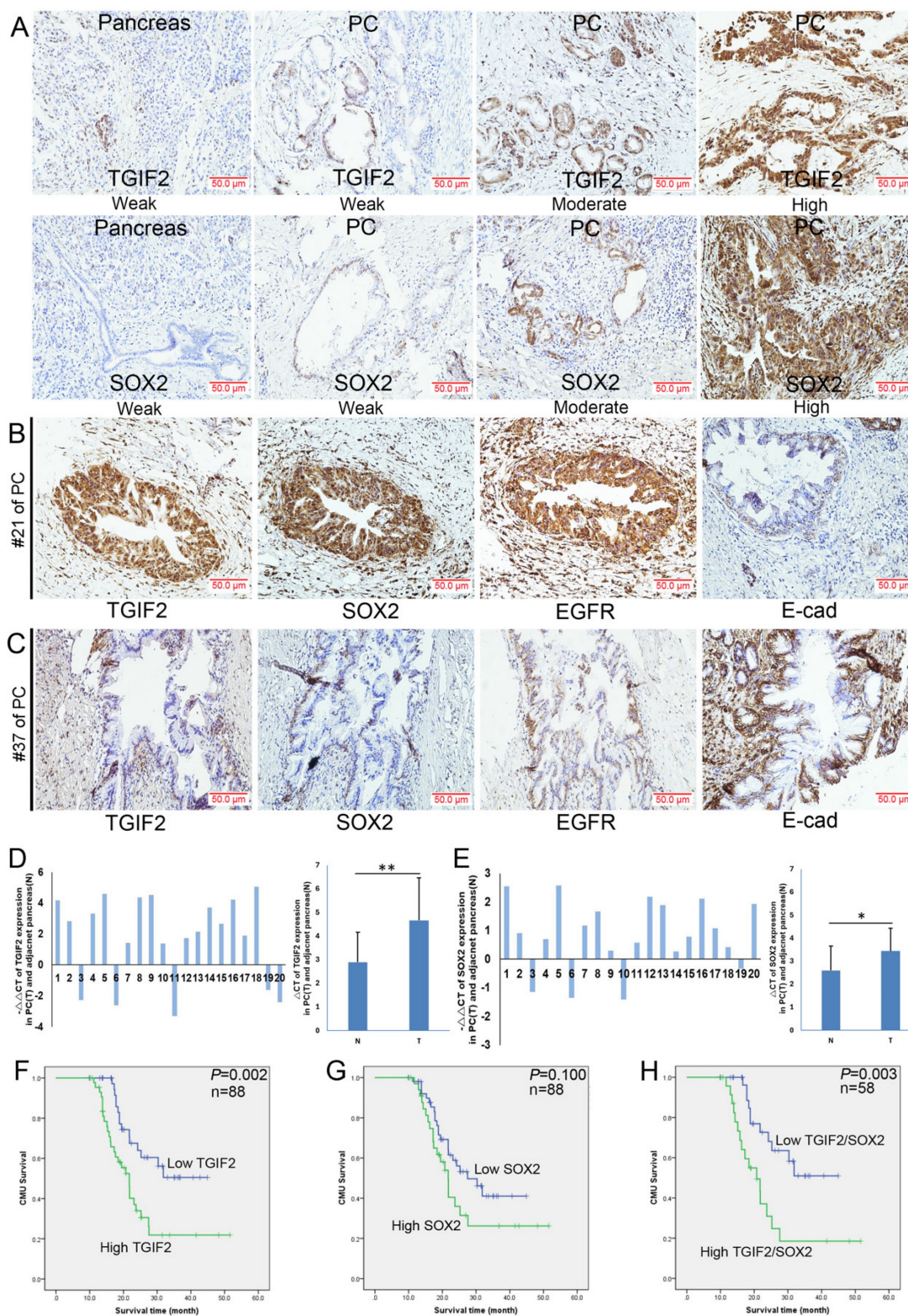


Figure 2. The expression of TGIF2 and SOX2 in human PC and adjacent pancreas. **A.** TGIF2 and SOX2 expression in PC and paired pancreas specimens by IHC. **B and C.** TGIF2, SOX2, EGFR and E-cad expression in two PC specimens (#21 and #37). **D.** The mRNA level of TGIF2 in 20 cases of human PC and adjacent pancreas by qRT-PCR (T: PC; N: paired pancreas) and SOX2. **E.** The mRNA level of SOX2 in our cohort. **F and G.** High and low expression of TGIF2 and SOX2 against prognosis. **H.** Combination of TGIF2 and SOX2 against prognosis.

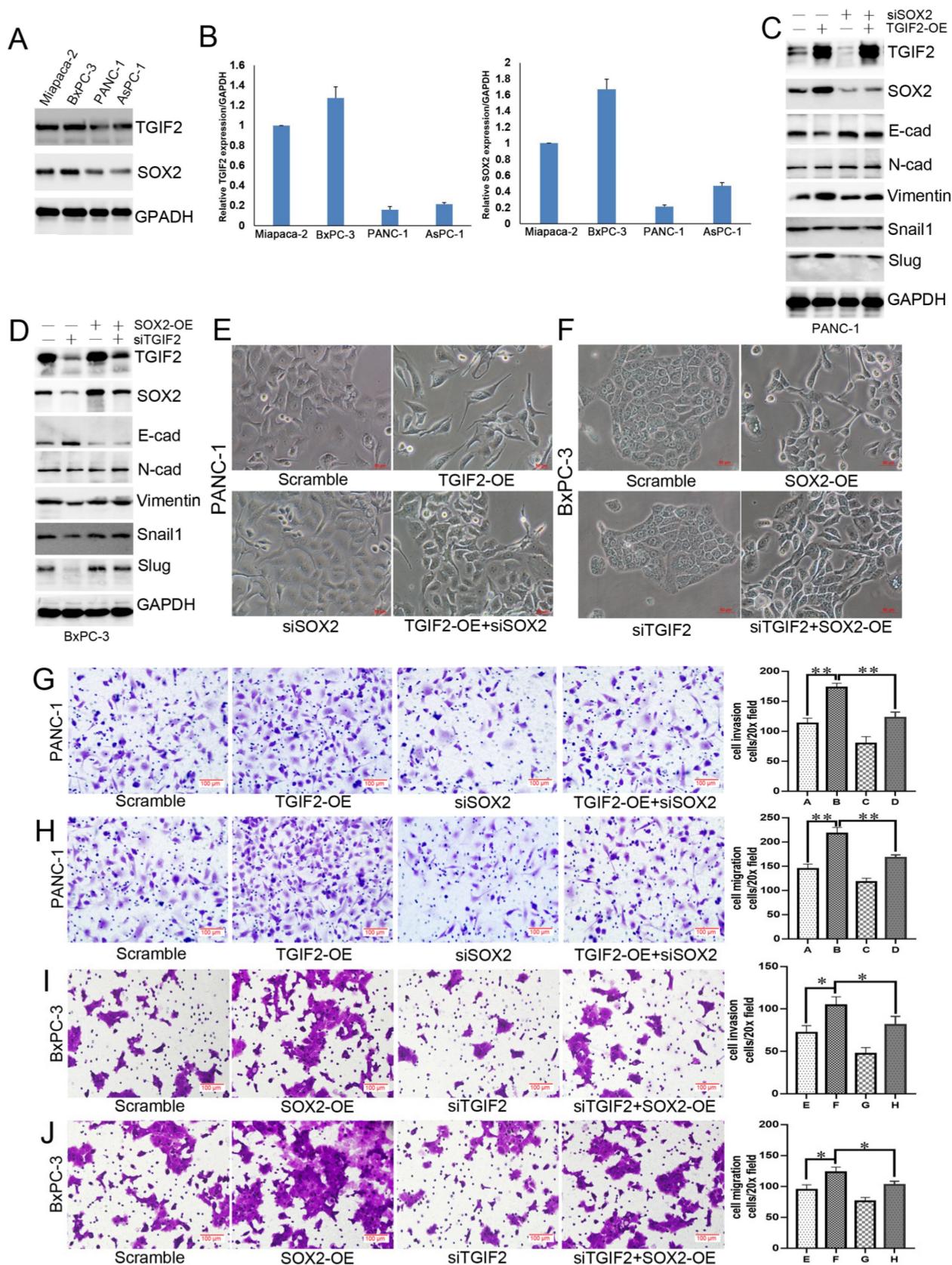


Figure 3. Coordination of TGIF2/SOX2 promotes EMT of PC *in vitro*. **A and B.** TGIF2 and SOX2 protein (A) and mRNA (B) level in 4 PC cell lines. **C.** The protein expression of TGIF2, SOX2 and EMT signaling in Scramble, TGIF2-OE, siSOX2 and TGIF2-OE/siSOX2 groups in PANC-1 cells. **D.** The protein expression of TGIF2, SOX2 and EMT signaling in Scramble, siTGIF2, SOX2-OE, and siTGIF2/SOX2-OE groups in BxPC-3 cells. **E.** The EMT phenotype in Scramble, TGIF2-OE, siSOX2 and TGIF2-OE/siSOX2 groups in PANC-1 cells. **F.** The EMT phenotype in Scramble, SOX2-OE, siTGIF2 and siTGIF2/SOX2-OE groups in BxPC-3 cells. **G and H.** Cell invasion (G) and migration (H) in Scramble, TGIF2-OE, siSOX2 and TGIF2-OE/siSOX2 groups in PANC-1 cells. **I and J.** Cell invasion (I) and migration (J) in Scramble, SOX2-OE and siTGIF2/SOX2-OE groups Scramble, SOX2-OE, siTGIF2 and siTGIF2/SOX2-OE in BxPC-3 cells. Bars indicate \pm S.E.*, $P < 0.05$; **, $P < 0.01$ compared with the control.

Table 2. The clinicopathological significance of TGIF2 and SOX2 expression in 88 cases of clinical PC samples

Parameters	No. of patients	TGIF2		P	SOX2		P
		Low	High		Low	High	
Cases	88	43	45		53	35	
Age (years)							
≤65	51	24	27	0.691	31	20	0.900
>65	37	19	18		22	15	
Gender							
Male	57	25	32	0.203	32	25	0.288
Female	31	18	13		21	10	
Tumor size (cm)							
<3	35	21	14	0.127	26	9	0.029
≥3	53	22	31		27	26	
Tumor location							
Head	62	33	29	0.206	35	27	0.264
Body-tail	26	10	16		18	8	
Differentiation							
Well	35	20	15	0.207	23	12	0.393
Moderate to Poor	53	23	30		30	23	
T stage ^a							
T1+T2	43	27	16	0.011	31	12	0.026
T3	45	16	29		22	23	
LN metastasis ^b							
N0 (negative)	60	34	26	0.032	38	22	0.383
N1 (positive)	28	9	19		15	13	
UICC stage ^a							
I+IIA stage	62	35	27	0.028	40	22	0.204
IIB+III stage	26	8	18		13	13	
Vascular permeation							
Absent	50	28	22	0.124	31	19	0.697
Present	38	15	23		22	16	

a. According to 8th TNM stage of AJCC. b. Lymph node

Therefore, it is essential to induce cell transfer to maintain EMT morphology. TGIF2-OE drove EMT by losing epithelial characteristics and presenting a spindle-shaped morphology (Figure 3E). Although SOX2 silencing did not affect the cell morphology, it reversed the TGIF2-OE EMT phenotype and decreased the spindle-shaped phenotype. Conversely, TGIF2 silencing inhibited the SOX2-OE phenotype in BxPC-3 cells (Figure 3F). Moreover, TGIF2 overexpression promotes the invasion and migration of PANC-1 cells. Although SOX2 silencing partially inhibited cell mobility, this inhibition was reversed in the TGIF2-OE + siSOX2 group (Figure 3G-H). In contrast, TGIF2 silencing inhibits the invasion and migration of BxPC-3 cells. However, SOX2-OE rescued the invasive and migratory capabilities of BxPC-3 cells (Figure 3I-J). These findings demonstrate the coordination of TGIF2/SOX2 in EMT.

TGIF2/SOX2 cooperatively promote CSCs and drug resistance

SOX2 plays a significant role in regulating tumor initiation and stem cell function. Thus, we

investigated the role of the TGIF2/SOX2 axis in the self-renewal of CSCs using a sphere formation assay (Figure 4A-B, Figure S5D-E). The protein expression of TGIF2, SOX2, CD133, CD44, p-EGFR, and p-ERK was upregulated during the transition from monolayer cells to spheres, indicating a vital role for the TGIF2/SOX2 axis and EGFR/MAPK signaling in the self-renewal capacity of CSCs. Although SOX2 silencing partially decreased the number of spheres compared to the scramble group, TGIF2-OE with SOX2 silencing rescued the decrease in the number of spheres caused by SOX2 silencing (Figure 4C). WB showed that TGIF2-OE promoted the expression of SOX2, CD133, EGFR, p-EGFR, and p-ERK, whereas SOX2 silencing partially downregulated the expression of these proteins. When TGIF2 overexpression and SOX2 silencing were performed concurrently, TGIF2-OE rescued the downregulated expression of SOX2, CD133, EGFR, p-EGFR, and p-ERK (Figure 4D and S5F). Conversely, TGIF2 silencing inhibited the self-renewal capacity of CSCs in BxPC-3 cells, which was significantly reversed by SOX2 overexpression (Figure 4E-F and S5G).

The initiation event of PC is caused by pancreatic intraepithelial neoplasia (PanIN). We found that TGIF2 and SOX2 silencing suppressed the number of spheroid pancreatic acinar cells in wild-type (treated with EGF to activate EGFR/ERK signaling) and KRAS-mutant (constant activation of EGFR/ERK signaling) mice. This trend was stronger in the TGIF2/SOX2 double-silenced group (Figure 4G-H). TGIF2, SOX2, CD133, and p-ERK were highly expressed in PanIN KC mice (6 months) (Figure 4I). All results indicated that the TGIF2/SOX2 axis not only promotes the self-renewal capacity of CSCs in PC by activating EGFR/MAPK signaling, but also participates in the early transformation events that occur in precancerous lesions.

With Gemcitabine treatment, the proliferation rate of TGIF2-OE PANC-1 cells was significantly higher than that in the scramble group. And SOX2 silencing partially inhibited cell proliferation and blunted the proliferation in the TGIF2-OE group when SOX2 silencing was done concurrently (Figure 4J-K). Erlotinib, an inhibitor of EGF receptor (EGFR), also significantly inhibited TGIF2-OE induced Gemcitabine resistance, indicating a vital role of EGFR/MAPK signaling in TGIF2/SOX2 axis-mediated drug resistance. Conversely, SOX2-OE reversed TGIF2 inhibited Gemcitabine sensitivity in BxPC-3 cells (Figure 4J-K). Taken together, the coordination of TGIF2/SOX2 promotes CSCs and drug resistance in PC.

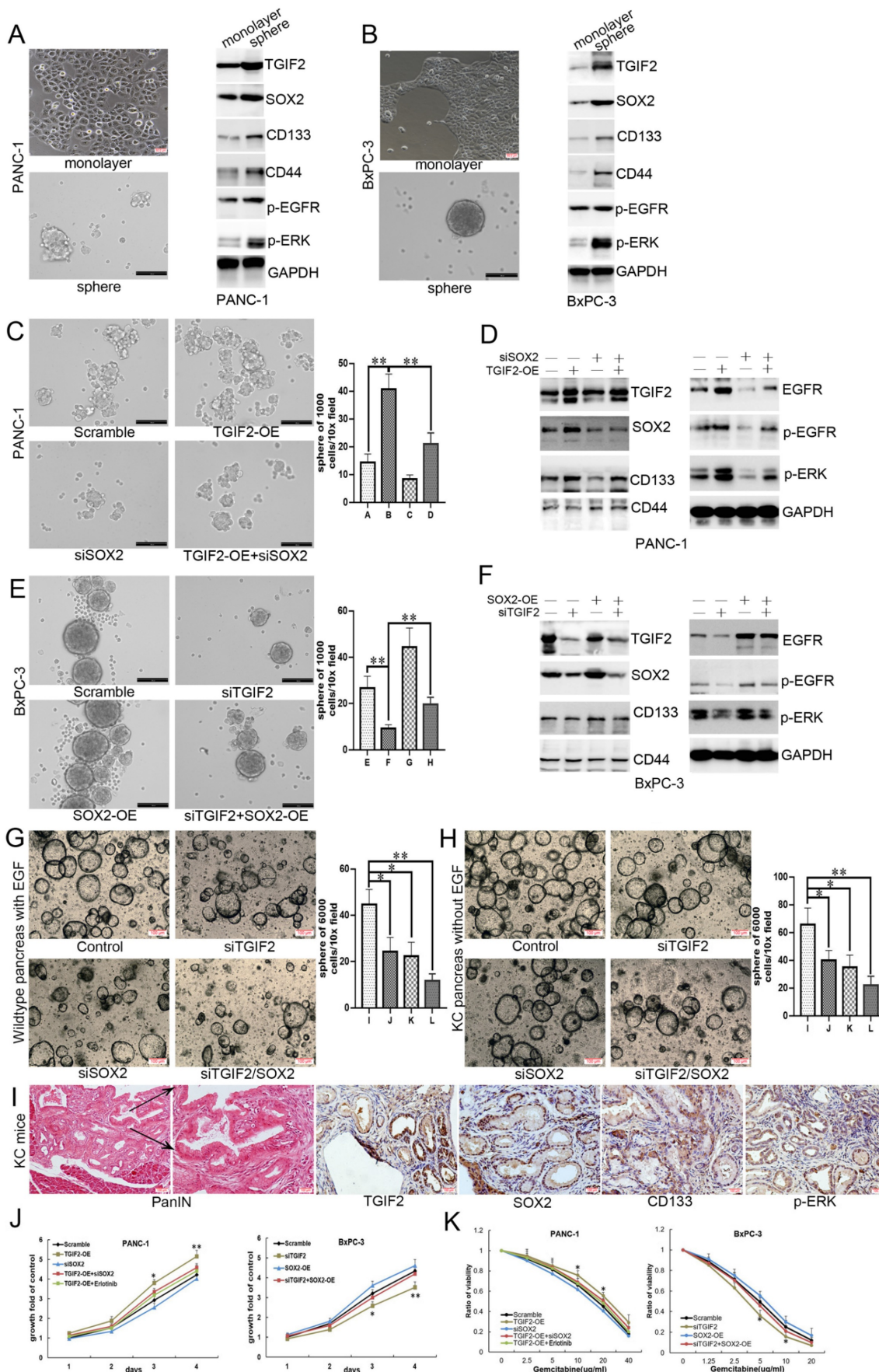


Figure 4. Coordination of TGIF2/SOX2 promotes CSCs and drug resistance of PC *in vitro*. **A and B.** Sphere formation in PANC-1 cells (A) and BxPC-3 (B) cells following the activation of TGIF2, SOX2, CD133, CD44 and EGFR/MAPK signaling. **C and D.** The sphere number (C) and the protein expression of TGIF2, SOX2, CD133, CD44 and

EGFR/MAPK signaling (D) in Scramble, TGIF2-OE, siSOX2 and TGIF2-OE/siSOX2 groups in PANC-1 cells. **E and F** The sphere number (E) and the protein expression of TGIF2, SOX2, CD133, CD44 and EGFR/MAPK signaling (F) in Scramble, siTGIF2, SOX2-OE and siTGIF2/SOX2-OE groups in BxPC-3 cells. **(G-H)** the spheroids number of PC in wildtype (treated with EGF to activate EGFR/ERK signaling) and Kras-mutant (constant activation of EGFR/ERK signaling) mice. **(I)** The overexpression of TGIF2, SOX2, CD133 and p-ERK were in pancreatic intraepithelial neoplasia (PanIN) of KC mice. **J.** Cell growth in Scramble, TGIF2-OE, siSOX2, TGIF2-OE/siSOX2 and TGIF2-OE/Erlotinib groups of PANC-1 cells and in Scramble, siTGIF2, SOX2-OE and siTGIF2/SOX2-OE of BxPC-3 cells by MTT. **K.** Under various concentration of Gemcitabine treatment, Cell proliferation rate in Scramble, TGIF2-OE, siSOX2, TGIF2-OE/siSOX2 and TGIF2-OE/Erlotinib groups of PANC-1 cells and in Scramble, siTGIF2, SOX2-OE and siTGIF2/SOX2-OE of BxPC-3 cells by MTT. Bars indicate \pm S.E.*, P <0.05; **, P <0.01 compared with the control.

TGIF2 with Smad2 activating SOX2 via Slug and EGFR expression

We first showed that TGIF2-OE or silence also upregulated or downregulated the mRNA level of SOX2, respectively (Figure 5A). The binding site of TGIF2 on SOX2 was obtained from the JASPAR database (Figure 5B). We designed a specific primer to amplify the SOX2 promoter and the corresponding wild-type SOX2 (SOX2-WT) and mutant SOX2 (SOX2-Mut) plasmids (Figure 5C). Upon immunoprecipitation with an anti-TGIF2 antibody, the DNA fragment containing site A was amplified to a significantly higher level. The amplification was more significant in TGIF2 overexpression group (Figure 5D). Furthermore, the luciferase activity of the SOX2-WT promoter was significantly stimulated by the co-transfection of TGIF2-OE in HEK-293 cells (Figure 5E). However, when the binding site of TGIF2 on SOX2 was mutated, the activating effect was markedly attenuated (Figure 5E). Thus, TGIF2 transactivates SOX2 in PC cells.

The protein interaction network of TGIF2 showed strong interaction TGIF2 with Smad2 and Smad3 (Figure 5F). Co-IP further verified that TGIF2 and Smad2 co-immunoprecipitated in anti-TGIF2 cells (Figure 5G and 5H). Smad2 silencing attenuated SOX2 expression, indicating that Smad2 may be a potential TF for SOX2 (Figure 5H-I, Figure S5I). We predicted Smad2, but not Smad3, as the potential TF of SOX2 using the hTFtarget and Jaspasr databases (Figure 5J) and designed the corresponding binding sites in SOX2-WT and SOX2-Mut (Figure 5K). After pull-down with anti-Smad2, the DNA fragment containing site A, but not site B, was amplified at a significantly higher level from the chromatin of BxPC-3 cells (Figure 5L). Additionally, luciferase activity of the SOX2-WT promoter was significantly stimulated by co-transfection with Smad2-OE in HEK-293 cells (Figure 5M). This activating effect was markedly attenuated (Figure 5M) when the binding site of Smad2 to SOX2 was mutated. Moreover, the upregulation of SOX2 induced by TGIF2-OE was significantly reversed by Smad2 silencing (Figure 5N and 5J), further indicating that Smad2 influences TGIF2 induced SOX2 expression.

SOX2-OE or silence upregulated or downregulated the mRNA levels of EGFR and Slug, respectively, in PANC-1 and BxPC-3 cells (Figure 6A). The slug-binding site is shown in Figure 6B, and a

specific primer for the slug promoter was obtained (Figure 6C). Upon anti-SOX2 treatment, the DNA fragment containing site A was amplified at a significantly higher level from the chromatin of PANC-1 and BxPC-3 cells (Figure 6D-E). The predicted binding sites in the EGFR promoter (sites A and B) are shown in Figure 6F. The amplification of site A was significantly higher than that of site B in the chromatin of cells (Figure 6G), revealing that TGIF2 is a transcriptional activator of SOX2, which in turn promotes EMT and CSCs by transactivating Slug and EGFR expression. Moreover, the luciferase activity was augmented in the EGFR-WT promoter plasmid rather than in EGFR-Mut, which further verified the specific binding site of SOX2 to the EGFR promoter (Figure 6H).

Notably, WB revealed that SOX2 overexpression promoted TGIF2 nuclear translocation in PANC-1 cells, which was reversed by erlotinib treatment (Figure 6I). It further showed that TGIF2 and SOX2 colocalized in the cytoplasm and nucleus of PANC-1 cells. SOX2 overexpression promoted TGIF2 nuclear translocation, which was reversed by erlotinib treatment (Figure 6J). Thus, transactivation of SOX2 by TGIF2 promotes EGFR/MAPK signaling, which in turn promotes TGIF2 nuclear translocation, forming a positive feedback loop *in vitro*. Moreover, the binding sites of TGIF2/Smad2 with SOX2 were detected in ESC and DE at three time points: ESC, DE-24h, and DE-48h. The openness and promoters of the three genes were activated and increased with time, indicating the overlap of ATAC-seq peaks among genes (Figure S4).

Coordination of TGIF2/SOX2 promotes tumor and liver metastasis

As shown in Figure 7A-B, the subcutaneous tumors in the TGIF2-OE group were much larger than those in the scramble group in a time-dependent manner. However, SOX2 silencing or erlotinib treatment significantly reversed TGIF2 overexpression-promoted increase in tumor size (TGIF2-OE vs. TGIF2-OE/shSOX2; TGIF2-OE vs TGIF2-OE/Erlotinib) (Figure 7A-B). Hematoxylin and eosin (HE) staining confirmed the tumor pathology (Figure 7C). IHC verified that the expression of TGIF2, SOX2, EGFR, and Vimentin was increased and that of E-cadherin was decreased in the TGIF2-OE group compared with the scramble group and reversed in

the TGIF2-OE combined with shSOX2 group (Figure 7D-E).

The number of liver metastases and the corresponding liver body ratio in TGIF2 silencing group (shTGIF2 group) were much lower than those in the scrambled control group (Figure 8A-B). However, SOX2 overexpression (SOX2-OE) significantly reversed TGIF2 silencing, which inhibited the decrease in liver metastasis and the corresponding liver body ratio (shTGIF2 group vs.

shTGIF2/SOX2-OE group) (Figure 8A-C). Liver body IHC further verified that TGIF2, SOX2, EGFR, and Vimentin were decreased, but E-cadherin was increased in the shTGIF2 group compared with the scramble group, which was significantly reversed in the shTGIF2/SOX2-OE group (Figure 8D-E). Taken together, the coordination of TGIF2/SOX2 promotes subcutaneous tumor size and liver metastasis by activating EMT and EGFR/MAPK signaling (Figure 9).

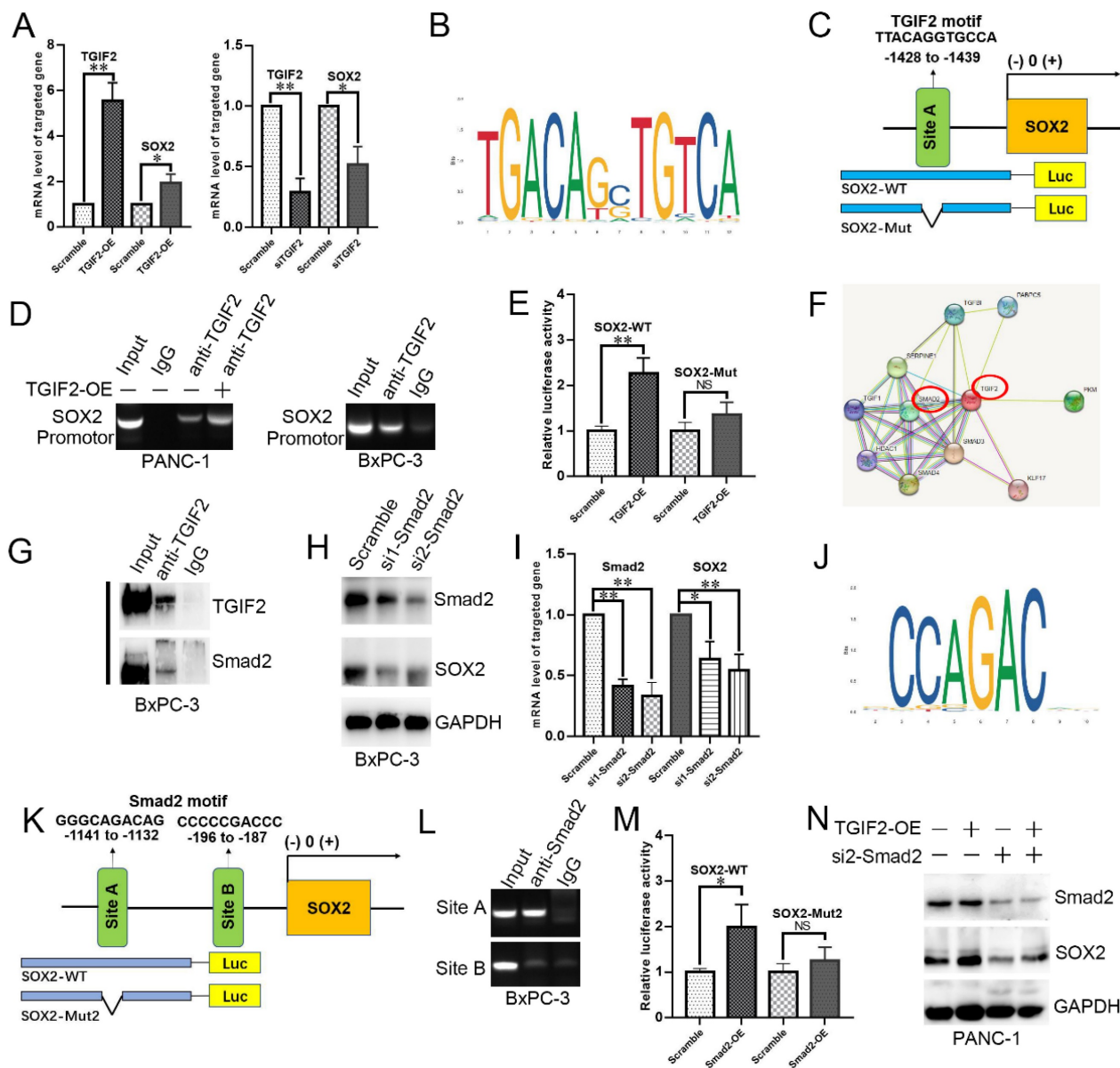


Figure 5. TGIF2 is a transactivation factor of SOX2 and interacts with Smad2 to co-regulate SOX2. **A.** The mRNA level of TGIF2 and SOX2 in TGIF2 overexpressing PANC-1 and TGIF2 silencing BxPC-3 cells, respectively. **B.** The binding site of TGIF2 in its zinc fingers domain was obtained from the JASPAR database. **C.** The predicted potential binding site of TGIF2 to SOX2 promoter, as well as wild-type/mutant SOX2 promoter plasmids (SOX2-WT or SOX2-Mut) designed accordingly. **D.** Chip assays in PANC-1 and BxPC-3 cells. **E.** Luciferase assay in 293 T cells co-transfected with SOX2-WT or SOX2-Mut promoter plasmid and TGIF2 overexpression plasmid. **F.** The protein interaction network of TGIF2 via String database. **G.** Co-IP was performed in PANC-1 and BxPC-3 cells. **H and I.** The protein and mRNA level of Smad2 and SOX2 in Smad2 silencing BxPC-3 cells. **J.** The binding site of Smad2 in its zinc fingers domain was obtained from the JASPAR database. **K.** The predicted potential binding site of Smad2 to SOX2 promoter, as well as wild-type/mutant SOX2 promoter plasmids (SOX2-WT or SOX2-Mut2) designed accordingly. **L.** Chip assays in BxPC-3 cells. **M.** Luciferase assay in 293 T cells co-transfected with SOX2-WT or SOX2-Mut2 promoter plasmid and Smad2 overexpression plasmid (or empty vector scramble). **N.** The protein level of Smad2 and SOX2 in Scramble, TGIF2-OE, si2-Smad2 and TGIF2-OE/si2-Smad2 groups in PANC-1 cells. Bars indicate \pm S.E. *, P < 0.05; **, P < 0.01 in contrast with the control.

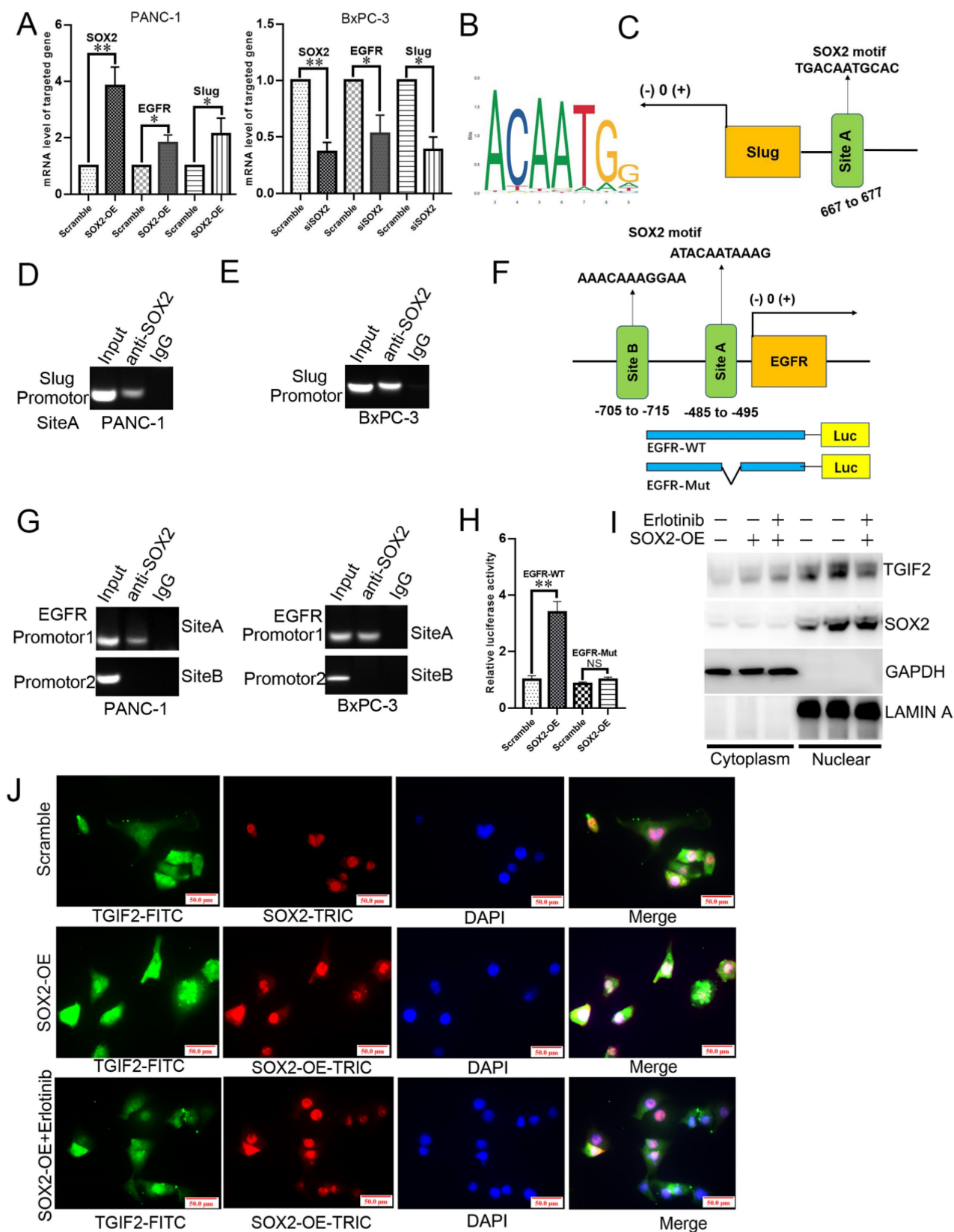


Figure 6. SOX2 is a transcription factor of EGFR and Slug. **A.** The mRNA level of SOX2, EGFR and Slug in SOX2 overexpressing PANC-1 and SOX2 silencing BxPC-3 cells, respectively. **B.** The DNA binding site of SOX2 was obtained from the JASPAR database. **C.** The predicted potential binding site of SOX2 to Slug promoter. **D and E.** Chip assays in PANC-1 (D) and BxPC-3 (E) cells. **F.** The predicted potential binding site of SOX2 to EGFR promoter, as well as wild-type/mutant EGFR promoter plasmids designed accordingly. **G.** Chip assays in PANC-1 and BxPC-3 cells. **H.** Luciferase assay in 293 T cells co-transfected with EGFR-WT (or EGFR-Mut) promoter plasmid and SOX2 overexpression plasmid (or empty vector corresponding to scramble group). **I.** The roles of SOX2-OE and Erlotinib treatment to TGIF2 nuclear translocation. **J.** IF staining of TGIF2 and SOX2 in Scramble, SOX2-OE and SOX2-OE plus Erlotinib groups of PANC-1. Bars indicate \pm S.E.*; $P < 0.05$; **, $P < 0.01$ in contrast with the control.

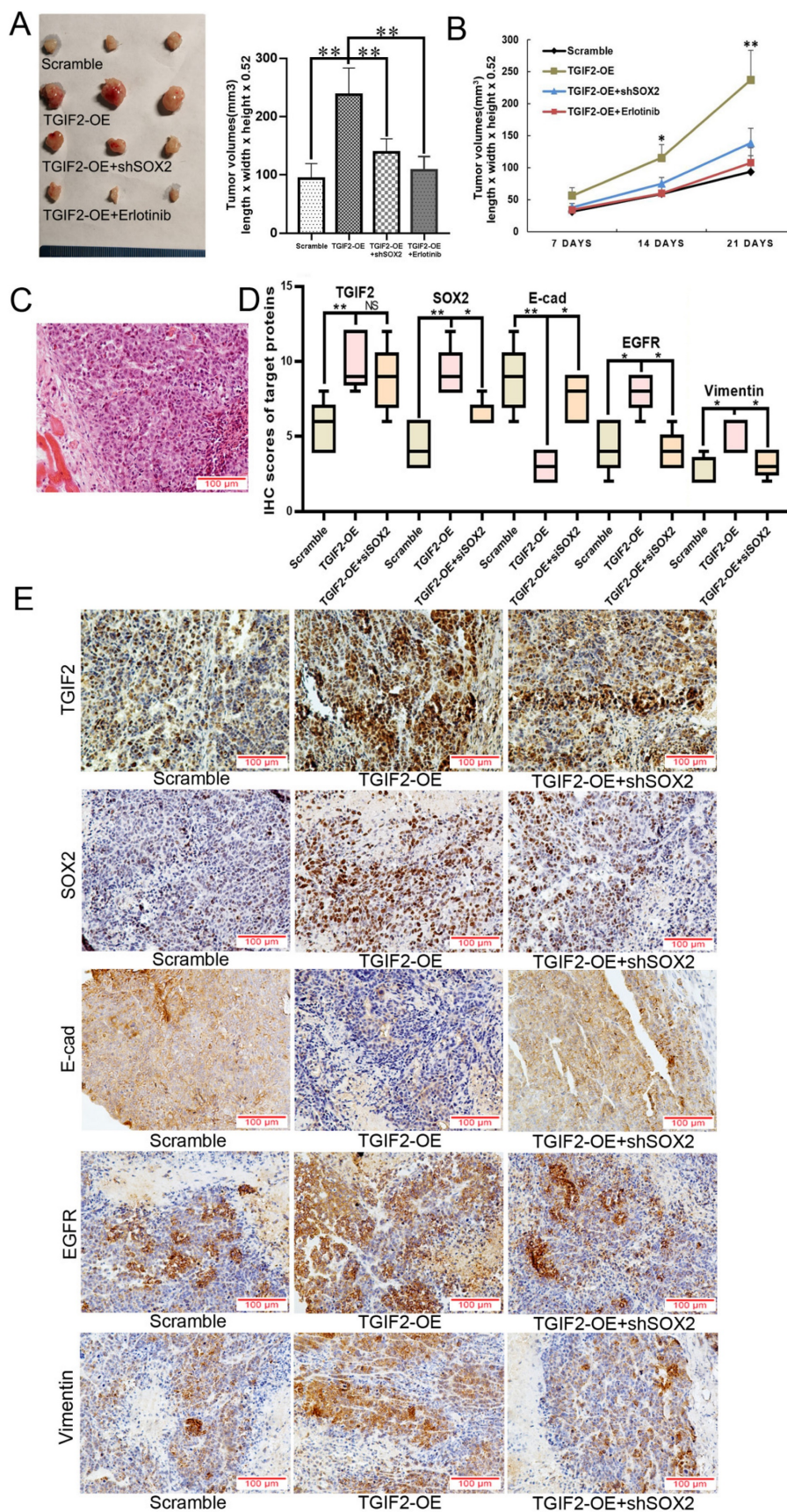


Figure 7. Coordination of TGIF2/SOX2 promotes subcutaneous tumor size *in vivo*. **A.** Tumor volumes in Scramble, TGIF2-OE, TGIF2-OE/shSOX2, and TGIF2-OE plus Erlotinib groups implanted with PANC-1 cells. **B.** Tumor growth curve in above group. **C.** HE staining of harvested tumor. **D.** The statistical data of IHC assays. **E.** The different expression of TGIF2, SOX2, E-cad, EGFR and Vimentin in Scramble, TGIF2-OE, and TGIF2-OE/shSOX2 groups by IHC. Bars indicate \pm S.E. * $P < 0.05$; ** $P < 0.01$ compared with control.

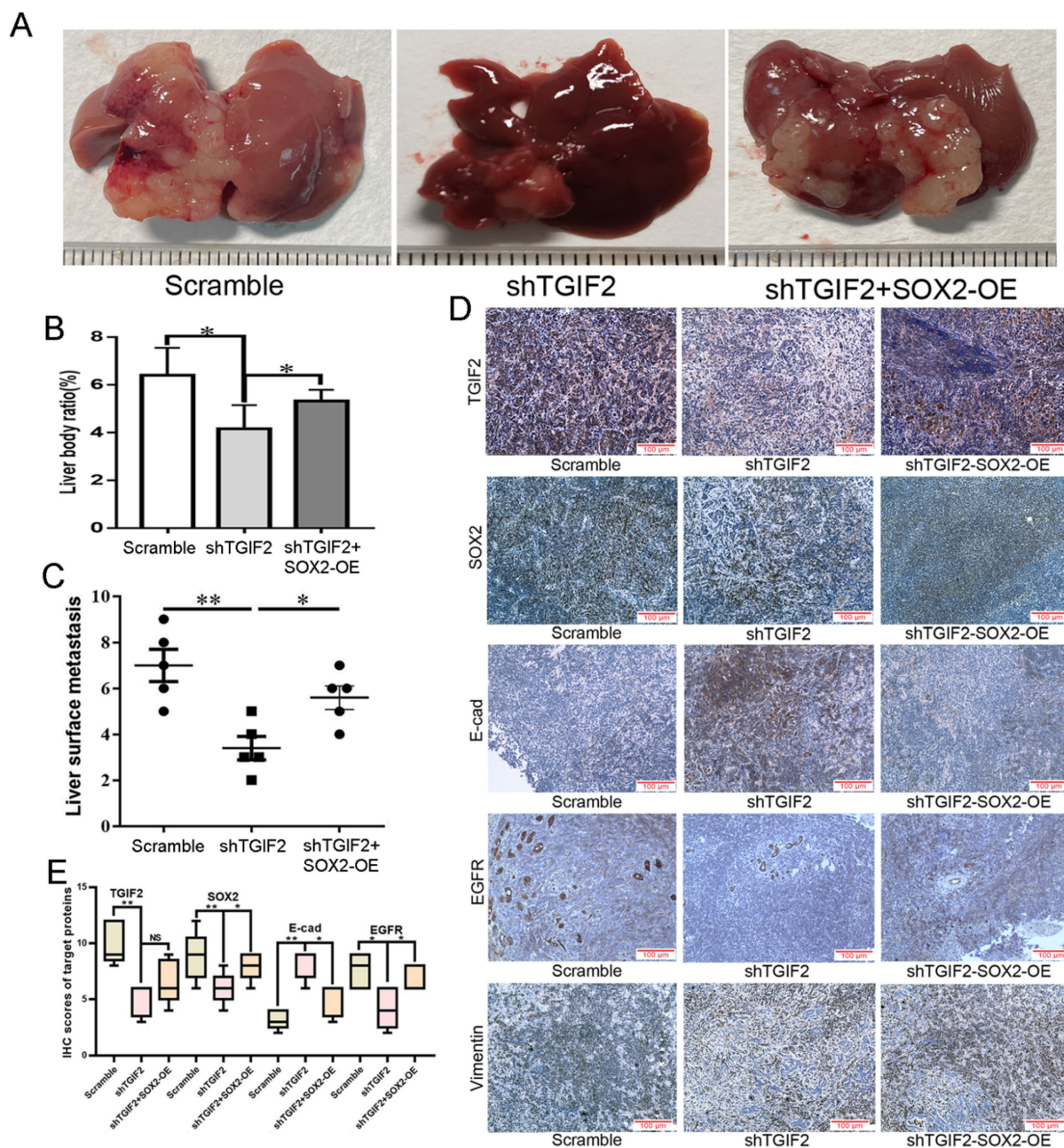


Figure 8. Coordination of TGIF2/SOX2 promotes liver metastasis *in vivo*. **A.** Liver metastasis in Scramble, shTGIF2 and shTGIF2/SOX2-OE groups. **B-C.** The Liver body ratio and the number of liver metastases in above groups. **D.** The different expression of TGIF2, SOX2, E-cad, EGFR and Vimentin in Scramble, shTGIF2 and shTGIF2/SOX2-OE groups by IHC. **E.** The statistical data of IHC assays. Bars indicate \pm S.E. *, $P < 0.05$; **, $P < 0.01$ compared with control.

Discussion

However, the transcriptional regulatory network of CSC and EMT in PC progression is poorly understood. In the present study, TGIF2 and Smad2 were identified as novel SOX2 transcription factors. Transactivation of SOX2 through cooperation between TGIF2 and Smad2 transcriptionally activates Snail2 and EGFR, which in turn promotes EMT and CSC. Meanwhile, SOX2 activates EGFR-ERK/MAPK signaling, promoting TGIF2 nuclear translocation and forming a positive feedback loop, which has not been

reported, to our knowledge.

In this study, we identified TGIF2 as a key player in the CSC signature using scRNA analysis and machine learning. TGIF2 overexpression indicates advanced clinical stage and poor prognosis. We then explored eight targets of TGIF2, and SOX2 showed strong correlations. The clinical relevance of SOX2 in PC was consistent with that of TGIF2, which also prevailed of TGIF2 in lung cancer and cervical cancer [47, 48] and SOX2 in cervical cancer [49], colorectal cancer [50], and urethral carcinoma [51]. TGIF2 was positively correlated with SOX2 and EGFR, but

negatively associated with E-cadherin expression, and the cooperation of TGIF2 and SOX2 contributed to a worse prognosis for patients with PC. These findings

indicate a close interaction between the TGIF2/SOX2 axis in CSC and EMT of PC, which prompted us to further explore the mechanism.

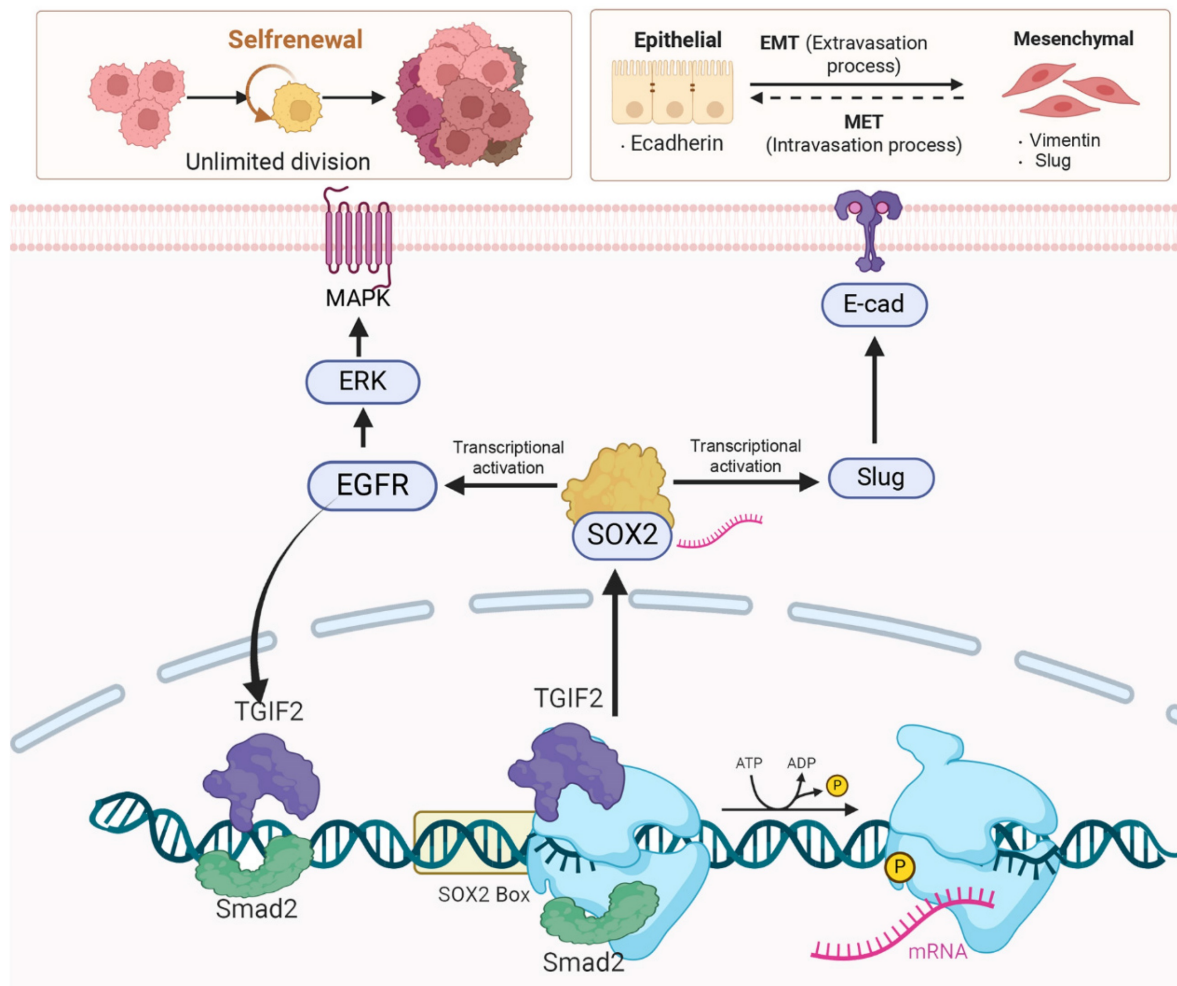


Figure 9. Smad2 cooperating with TGIF2 contributes to CSC and EMT via co-targeting SOX2. TGIF2 activates SOX2 promoter via interacting with Smad2, which stimulates EMT and EGFR/MAPK signaling by transactivating Slug and EGFR, and promoting EMT and CSCs function. Moreover, the stimulation of EGFR/MAPK signaling by SOX2 promotes TGIF2 nuclear translocation, forming a positive feedback loop.

Table 3. Univariate and Multivariate analysis in survival time

Parameters	Median survival (months)	Univariate analysis <i>P</i> (log rank)	Multivariate analysis hazard ratio (95% CI)	<i>P</i>
Differentiation				
(Well/Moderate to Poor)	31.8 vs 21.8	0.065	-	-
T stage				
(T1+T2/T3)	30.3 vs 21.1	0.072	-	-
Lymph nodes metastasis (N0/N1)	31.3 vs 19.4	0.002	1.39(0.61-3.11)	0.430
Vascular permeation (absent/present)	31.8 vs 18.8	0.004	1.62 (0.73-3.60)	0.233
UICC stage	30.3 vs 17.8	0.002	2.39(1.24-4.56)	0.008
(I+IIA/IIIB+III)				
TGIF2				
(Low/High)	34.1 vs 21.7	0.002	2.16(1.10-4.24)	0.024
SOX2				
(Low/High)	27.4 vs 21.7	0.100	-	-
E-cad				
(Normal/ Abnormal)	27.5 vs 23.1	0.240	-	-
EGFR				
(Negative/Positive)	27.4 vs 22.1	0.123	-	-

Moreover, TGIF2-OE induced an EMT-like phenotype following an increase in SOX2, Vimentin and Slug expression and a decrease in E-cadherin expression, which was reversed by SOX2 silencing. In lung cancer, TGIF2 phosphorylation is a therapeutic target that drives EMT and metastasis [48]. However, TGIF2 was reported to repress EMT of oral squamous cell carcinoma [25]. The dual role of TGIF2 in EMT may be due to differences in cancer types and microenvironments, which helps us understand the general or cancer-type-specific EMT response. Specifically, genomic alterations in EMT have been systematically studied in gastric cancer, which is related to the fact that the stomach and pancreas are both endoderm-derived tissues that share some TFs. We have previously investigated the molecular mechanisms mediating the initiation and development of EMT in PC [10-13, 52]. Various key regulators (CALR, MSI2-Numb, ZNF263/ZNF31, and GINS2) play significant roles in promoting EMT following genomic alterations under different tumor microenvironment conditions. Most of these genes also participate in EMT in GC [53-56]. Similarly, common EMT gene signatures found in GC also exist in PC, including ZEB1, RUNX and AP-1 [36-38]. For example, Zeb1 induces PC plasticity as an EMT activator [13], RUNX acts as a master transcription factor in PC [57], and AP-1 regulates the activation of Akt signaling during PC tumorigenesis [58].

Dormant cells stimulate EMT phenotypes to promote the metastasis in CSC-like phenotypes [51], which is key in cancer drug resistance acquisition and malignant plasticity [59]. Our findings showed that TGIF2-OE promoted the self-renewal capacity of CSCs and drug resistance following the upregulation of SOX2, CD133, CD44, and EGFR/ERK signaling, which was also reversed by SOX2 silencing. PanIN is one of the earliest events involved in exocrine PC development [48, 60]. Oncogenic KrasG12D inhibits healing progression by blocking redifferentiation and promoting PanIN formation [61]. Here, we found that the TGIF2/SOX2 axis induces PanIN by increasing the number of acinar-derived spheroids in KRAS-mutant mice. Thus, the TGIF2/SOX2 axis not only promotes the self-renewal capacity of CSCs but also participates in the early events of PC involving the initiation and progression of PanIN.

Mechanistically, TGIF2 immunoprecipitated with Smad2 and induced the transcription of SOX2. Smad2, but not Smad3, acts as a parallel transcriptional activator of SOX2 by binding to its promoter. This difference between Smad2 and Smad3 in DNA binding is likely due to a sequence insert in the N domain of Smad2, immediately before the DNA binding β hairpin, which might be a block of DNA

recognition [62]. TGIF2 interacted with Smad2/3 in a TGF- β -inducible manner, resulting in the recruitment of TGIF2 to Smad-responsive DNA elements [63]. Though TGIF2 interacts with TGF- β -activated Smads and represses TGF beta-responsive transcription, it appears to be a context-dependent transcriptional activator or repressor consistent with other TALE homeoproteins [23]. For example, TGIF2 promotes lung cancer stemness by transactivating OCT4 [48]. It can also bind to the CDH1 promoter and activate CDH1 expression in epithelial cells of colon cancer cells [64]. The mechanism of transcriptional activation by Smad is most likely defined by the combined requirement of interactions with co-transcription activators and promoter DNA sequences. In the current study, Smad2 cooperated with TGIF2 to promote the transcriptional activity of SOX2. Smad2 is key for maintaining the pluripotent stem cell state [65]. For example, Smad2/3 is previously known to directly binds and regulates the expression of NANOG to sustain human ESC self-renewal [66]. ChIPseq in human ESCs also revealed the binding of SMAD2/3 to OCT4, TERT, MYC, and DPPA4 [67]. Here, Smad2 was first identified as a transcriptional activator of SOX2 via its interaction with TGIF2, which, to the best of our knowledge, has not yet been clarified.

Transactivation of SOX2 has also been identified as a novel TF activator of Slug and EGFR via binding to their promoters. The transcription factor Slug represses E-cadherin expression and induces EMT in several cancers [68], while EGFR/MAPK signaling promotes CSC function in colorectal, breast, and squamous cell carcinoma [69-71]. Thus, the TGIF2/SOX2 axis promotes EMT and CSCs properties via activation of Slug and EGFR signaling. Interestingly, the transactivation of EGFR/MAPK signaling by SOX2 promotes TGIF2 nuclear translocation, forming a positive feedback loop. EGFR-RAS-ERK signaling phosphorylates TGIF2 and increases its stability [48]. TGIF2 also stimulates EGFR/MAPK signaling by activating SOX2, which, in turn, increases TGIF2 stability by promoting its nuclear translocation.

Taken together, TGIF2 and smad2 targeting SOX2 promoted EMT and CSCs properties via the activation of Slug and EGFR signaling, respectively. The transactivation of EGFR/MAPK signaling by SOX2 promotes TGIF2 nuclear translocation, forming a positive feedback loop (Figure 9). The TGIF2/SOX2 transcriptional axis contributes to EMT, cancer stem cell properties, and chemoresistance in PC and is a promising target for PC therapy.

Abbreviations

EMT: Epithelial-mesenchymal transition; CSC: cancer stem cell; scRNA: Single-cell RNA; WGCNA: weighted gene co-expression network; t-SNE: t-Distributed Stochastic Neighbor Embedding; TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; Gastric cancer: GC; Pancreatic cancer: PC; TGIF2: TGF- β -induced factor homeobox 2; SOX2: sex determining region Y box 2; Smad2: SMAD family member 2; ERK: Extracellular regulated protein kinases; EGFR: epidermal growth factor receptor; MAPK: mitogen-activated protein kinase; E-cad: E-cadherin; N-cad: N-cadherin.

Supplementary Material

Supplementary figures and tables.
<https://www.ijbs.com/v21p0524s1.pdf>

Acknowledgements

The authors thank the Herbert Irving Comprehensive Cancer Center of Columbia University and the Center of Scientific Research of Anhui Medical University and for valuable help in our experiments.

Funding

This study was funded by the Research Fund of Anhui Institute of Translational Medicine (2023zhyx-C36), Clinical Science Fund Project of Anhui Medical University (2023xkj140), Education University Research Program of Anhui province (2024AH050827), the project of Anhui province key laboratory of tumor immune microenvironment and immunotherapy (No. 20240004), the Entrepreneurship Project for College Students of Anhui Medical University (202210366008, 202410366024, X202410366048), the Graduate Research and Practice Innovation Project of Anhui Medical University (YJS20230176), the Undergraduate Training Program for Innovation and Entrepreneurship (S202410366032), the Talent Introduction Fund Project of the Second Affiliated Hospital of Anhui Medical University (JFN20230114) and the National Natural Incubation Program of the Second Affiliated Hospital of Anhui Medical University (2023GQFY02).

Ethics approval and consent to participate

All patients received written informed consent with approval by the academic committee at the First Affiliated Hospital of Anhui Medical University (2023558). The animal study was approved by the ethics committee of Anhui Medical University (LISC20231510).

Consent for publication

All authors have read and approved of its submission to this journal.

Availability of data and materials

The scRNA-seq data can be requested from the authors. The RNA-seq transcript data and clinical information of PC patients are downloaded from: E-MTAB-6134 (<https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-6134>), TCGA (<https://portal.gdc.cancer.gov/projects/TCGA-PAA-D>), GSE21501 and GSE62452.

Author contributions

Conception: MI, BR, HG, WS. Methodology: FZ, CC, QG, HL, MW. Acquisition of data: GL, BJ, HG, WS. Analysis of data: FZ, CC. Writing and reviewing the manuscript: GL, MI, BR, HG, WS. Administrative, technical, or material support: GL, MI, BR. Study supervision: LB-E, HG and WS. All authors read and approved the final manuscript.

Competing Interests

The authors have declared that no competing interest exists.

References

1. Miller KD, Nogueira L, Devasia T, Mariotto AB, Yabroff KR, Jemal A, et al. Cancer treatment and survivorship statistics, 2022. *CA Cancer J Clin.* 2022; 72: 409-36.
2. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024; 74: 229-63.
3. Pan G, Liu Y, Shang L, Zhou F, Yang S. EMT-associated microRNAs and their roles in cancer stemness and drug resistance. *Cancer Commun (Lond).* 2021; 41: 199-217.
4. Shibue T, Weinberg RA. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. *Nat Rev Clin Oncol.* 2017; 14: 611-29.
5. Tanabe S, Quader S, Cabral H, Ono R. Interplay of EMT and CSC in Cancer and the Potential Therapeutic Strategies. *Front Pharmacol.* 2020; 11: 904.
6. Hermann PC, Huber SL, Herrler T, Aicher A, Ellwart JW, Guba M, et al. Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. *Cell Stem Cell.* 2007; 1: 313-23.
7. Zhou P, Li B, Liu F, Zhang M, Wang Q, Liu Y, et al. The epithelial to mesenchymal transition (EMT) and cancer stem cells: implication for treatment resistance in pancreatic cancer. *Mol Cancer.* 2017; 16: 52.
8. Palamaris K, Felekouras E, Sakellariou S. Epithelial to Mesenchymal Transition: Key Regulator of Pancreatic Ductal Adenocarcinoma Progression and Chemoresistance. *Cancers (Basel).* 2021; 13: 5532.
9. Preca BT, Bajdak K, Mock K, Sundararajan V, Pfanstiel J, Maurer J, et al. A self-enforcing CD44s/ZEB1 feedback loop maintains EMT and stemness properties in cancer cells. *Int J Cancer.* 2015; 137: 2566-77.
10. Sheng W, Tang J, Cao R, Shi X, Ma Y, Dong M. Numb-PRRL promotes TGF- β 1- and EGF-induced epithelial-to-mesenchymal transition in pancreatic cancer. *Cell Death Dis.* 2022; 13: 173.
11. Sheng W, Wang G, Tang J, Shi X, Cao R, Sun J, et al. Calreticulin promotes EMT in pancreatic cancer via mediating Ca(2+) dependent acute and chronic endoplasmic reticulum stress. *J Exp Clin Cancer Res.* 2020; 39: 209.
12. Huang L, Chen S, Fan H, Ji D, Chen C, Sheng W. GINS2 promotes EMT in pancreatic cancer via specifically stimulating ERK/MAPK signaling. *Cancer Gene Ther.* 2021; 28: 839-49.
13. Sheng W, Shi X, Lin Y, Tang J, Jia C, Cao R, et al. Musashi2 promotes EGF-induced EMT in pancreatic cancer via ZEB1-ERK/MAPK signaling. *J Exp Clin Cancer Res.* 2020; 39: 16.
14. Peng J, Sun BF, Chen CY, Zhou JY, Chen YS, Chen H, et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* 2019; 29: 725-38.

15. Moncada R, Barkley D, Wagner F, Chiodin M, Devlin JC, Baron M, et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat Biotechnol.* 2020; 38: 333-42.
16. Bärthel S, Falcomatà C, Rad R, Theis FJ, Saur D. Single-cell profiling to explore pancreatic cancer heterogeneity, plasticity and response to therapy. *Nat Cancer.* 2023; 4: 454-67.
17. Liu Z, Liu L, Weng S, Guo C, Dang Q, Xu H, et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat Commun.* 2022; 13: 816.
18. Wang L, Liu Z, Liang R, Wang W, Zhu R, Li J, et al. Comprehensive machine-learning survival framework develops a consensus model in large-scale multicenter cohorts for pancreatic cancer. *Elife.* 2022; 11: e80150.
19. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008; 9: 559.
20. Goossens S, Vandamme N, Van Vlierbergh P, Bex G. EMT transcription factors in cancer development re-evaluated: Beyond EMT and MET. *Biochim Biophys Acta Rev Cancer.* 2017; 1868: 584-91.
21. Yang L, Shi P, Zhao G, Xu J, Peng W, Zhang J, et al. Targeting cancer stem cell pathways for cancer therapy. *Signal Transduct Target Ther.* 2020; 5: 8.
22. Agrawal P, Fontanals-Cirera B, Sokolova E, Jacob S, Vaiana CA, Argibay D, et al. A Systems Biology Approach Identifies FUT8 as a Driver of Melanoma Metastasis. *Cancer Cell.* 2017; 31: 804-19.e7.
23. Melhuish TA, Gallo CM, Wotton D. TGIF2 interacts with histone deacetylase 1 and represses transcription. *J Biol Chem.* 2001; 276: 32109-14.
24. Vincure OS, Sharma V, Tabasum S, Ghosh S, Singh RP, Sarkar C, et al. Polycomb complex mediated epigenetic reprogramming alters TGF- β signaling via a novel EZH2/miR-490/TGIF2 axis thereby inducing migration and EMT potential in glioblastomas. *Int J Cancer.* 2019; 145: 1254-69.
25. Tanaka F, Yoshimoto S, Okamura K, Ikebe T, Hashimoto S. Nuclear PKM2 promotes the progression of oral squamous cell carcinoma by inducing EMT and post-translationally repressing TGIF2. *Oncotarget.* 2018; 9: 33745-61.
26. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol.* 2018; 36: 411-20.
27. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, et al. Comprehensive Integration of Single-Cell Data. *Cell.* 2019; 177: 1888-902.e21.
28. Sarkans U, Gostev M, Athar A, Behrang E, Melnichuk O, Ali A, et al. The BioStudies database-one stop shop for all data supporting a life sciences study. *Nucleic Acids Res.* 2018; 46: D1266-d70.
29. Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol.* 2020; 38: 675-8.
30. Pinto JP, Kalathur RK, Oliveira DV, Barata T, Machado RS, Machado S, et al. StemChecker: a web-based tool to discover and explore stemness signatures in gene sets. *Nucleic Acids Res.* 2015; 43: W72-7.
31. Jin Y, Wang Z, He D, Zhu Y, Chen X, Cao K. Identification of novel subtypes based on ssGSEA in immune-related prognostic signature for tongue squamous cell carcinoma. *Cancer Med.* 2021; 10: 8693-707.
32. Luo J, Luo F, Li Q, Liu Q, Wang J. An immunogenic cell death-related lncRNA signature correlates with prognosis and tumor immune microenvironment in bladder cancer. *Sci Rep.* 2024; 14: 13106.
33. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics.* 2012; 16: 284-7.
34. Hu H, Miao YR, Jia LH, Yu QY, Zhang Q, Guo AY. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* 2019; 47: D33-d8.
35. Luo R, Yan J, Oh JW, Xi W, Shigaki D, Wong W, et al. Dynamic network-guided CRISPRi screen identifies CTCF-loop-constrained nonlinear enhancer gene regulatory activity during cell state transitions. *Nat Genet.* 2023; 55: 1336-46.
36. Razavi-Mohseni M, Huang W, Guo YA, Shigaki D, Ho SWT, Tan P, et al. Machine learning identifies activation of RUNX/AP-1 as drivers of mesenchymal and fibrotic regulatory programs in gastric cancer. *Genome Res.* 2024; 34: 680-95.
37. Ho SWT, Sheng T, Xing M, Ooi WF, Xu C, Sundar R, et al. Regulatory enhancer profiling of mesenchymal-type gastric cancer reveals subtype-specific epigenomic landscapes and targetable vulnerabilities. *Gut.* 2023; 72: 226-41.
38. Sheng T, Ho SWT, Ooi WF, Xu C, Xing M, Padmanabhan N, et al. Integrative epigenomic and high-throughput functional enhancer profiling reveals determinants of enhancer heterogeneity in gastric cancer. *Genome Med.* 2021; 13: 158.
39. Herrerros-Villanueva M, Zhang JS, Koenig A, Abel EV, Smyrk TC, Bamlet WR, et al. SOX2 promotes dedifferentiation and imparts stem cell-like features to pancreatic cancer cells. *Oncogenesis.* 2013; 2: e61.
40. Takeuchi K, Tabe S, Takahashi K, Aoshima K, Matsuo M, Ueno Y, et al. Incorporation of human iPSC-derived stromal cells creates a pancreatic cancer organoid with heterogeneous cancer-associated fibroblasts. *Cell Rep.* 2023; 42: 113420.
41. Xie Z, Gao Y, Ho C, Li L, Jin C, Wang X, et al. Exosome-delivered CD44v6/CIQP complex drives pancreatic cancer liver metastasis by promoting fibrotic liver microenvironment. *Gut.* 2022; 71: 568-79.
42. Astuti Y, Raymant M, Quaranta V, Clarke K, Abudula M, Smith O, et al. Efferocytosis reprograms the tumor microenvironment to promote pancreatic cancer liver metastasis. *Nat Cancer.* 2024; 5: 774-90.
43. Thomas SK, Wattenberg MM, Choi-Bose S, Uhlik M, Harrison B, Coho H, et al. Kupffer cells prevent pancreatic ductal adenocarcinoma metastasis to the liver in mice. *Nat Commun.* 2023; 14: 6330.
44. Kim EJ, Kim JY, Kim SO, Hong N, Choi SH, Park MG, et al. The oncogenic JAG1 intracellular domain is a transcriptional cofactor that acts in concert with DDX17/SMAD3/TGIF2. *Cell Rep.* 2022; 41: 111626.
45. Sommariva M, Gagliano N. E-Cadherin in Pancreatic Ductal Adenocarcinoma: A Multifaceted Actor during EMT. *Cells.* 2020; 9: 1040.
46. Zhu Y, Huang S, Chen S, Chen J, Wang Z, Wang Y, et al. SOX2 promotes chemoresistance, cancer stem cells properties, and epithelial-mesenchymal transition by β -catenin and Beclin1/autophagy signaling in colorectal cancer. *Cell Death Dis.* 2021; 12: 449.
47. Jiang J, Wu RH, Zhou HL, Li ZM, Kou D, Deng Z, et al. TGIF2 promotes cervical cancer metastasis by negatively regulating FCMR. *Eur Rev Med Pharmacol Sci.* 2020; 24: 5953-62.
48. Du R, Shen W, Liu Y, Gao W, Zhou W, Li J, et al. TGIF2 promotes the progression of lung adenocarcinoma by bridging EGFR/RAS/ERK signaling to cancer cell stemness. *Signal Transduct Target Ther.* 2019; 4: 60.
49. Yuan D, Wang J, Yan M, Xu Y. SOX2 as a prognostic marker and a potential molecular target in cervical cancer: A meta-analysis. *Int J Biol Markers.* 2021; 36: 45-53.
50. Bao Z, Zhan Y, He S, Li Y, Guan B, He Q, et al. Increased Expression Of SOX2 Predicts A Poor Prognosis And Promotes Malignant Phenotypes In Upper Tract Urothelial Carcinoma. *Cancer Manag Res.* 2019; 11: 9095-106.
51. Weidenfeld K, Schiff-Zuck S, Abu-Tayeh H, Kang K, Kessler O, Weissmann M, et al. Dormant tumor cells expressing LOXL2 acquire a stem-like phenotype mediating their transition to proliferative growth. *Oncotarget.* 2016; 7: 71362-77.
52. Zhan J, Chen C, Geng Q, Li H, Wu M, Chan B, et al. ZNF263 cooperates with ZNF31 to promote the drug resistance and EMT of pancreatic cancer through transactivating RNF126. *J Cell Physiol.* 2024; 239: e31259.
53. Wang L, Chen J, Zuo Q, Wu C, Yu T, Zheng P, et al. Calreticulin enhances gastric cancer metastasis by dimethylating H3K9 in the E-cadherin promoter region mediating by G9a. *Oncogenesis.* 2022; 11: 29.
54. Sun J, Sheng W, Ma Y, Dong M. Potential Role of Musashi-2 RNA-Binding Protein in Cancer EMT. *Onco Targets Ther.* 2021; 14: 1969-80.
55. Feng H, Zeng J, Gao L, Zhou Z, Wang L. GINS Complex Subunit 2 Facilitates Gastric Adenocarcinoma Proliferation and Indicates Poor Prognosis. *Tohoku J Exp Med.* 2021; 255: 111-21.
56. Xu G, Li K, Zhang N, Zhu B, Feng G. Screening Driving Transcription Factors in the Processing of Gastric Cancer. *Gastroenterol Res Pract.* 2016; 2016: 8431480.
57. Ervin EH, French R, Chang CH, Pauklin S. Inside the stemness engine: Mechanistic links between deregulated transcription factors and stemness in cancer. *Semin Cancer Biol.* 2022; 87: 48-83.
58. Shin S, Asano T, Yao Y, Zhang R, Claret FX, Korc M, et al. Activator protein-1 has an essential role in pancreatic cancer cells and is regulated by a novel Akt-mediated mechanism. *Mol Cancer Res.* 2009; 7: 745-54.
59. Loret N, Denys H, Tummers P, Bex G. The Role of Epithelial-to-Mesenchymal Plasticity in Ovarian Cancer Progression and Therapy Resistance. *Cancers (Basel).* 2019; 11: 838.
60. Storz P. Acinar cell plasticity and development of pancreatic ductal adenocarcinoma. *Nat Rev Gastroenterol Hepatol.* 2017; 14: 296-304.
61. Marstrand-Daucé L, Lorenzo D, Chassac A, Nicole P, Couvelard A, Haumaitre C. Acinar-to-Ductal Metaplasia (ADM): On the Road to Pancreatic Intraepithelial Neoplasia (PanIN) and Pancreatic Cancer. *Int J Mol Sci.* 2023; 24: 9946.
62. Shi Y, Wang YF, Jayaraman L, Yang H, Massagué J, Pavletich NP. Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling. *Cell.* 1998; 94: 585-94.
63. Massagué J, Wotton D. Transcriptional control by the TGF-beta/Smad signaling system. *Embo j.* 2000; 19: 1745-54.
64. Yang Y, Ren P, Liu X, Sun X, Zhang C, Du X, et al. PPP1R26 drives hepatocellular carcinoma progression by controlling glycolysis and epithelial-mesenchymal transition. *J Exp Clin Cancer Res.* 2022; 41: 101.
65. Hamabe A, Konno M, Tanuma N, Shima H, Tsunekuni K, Kawamoto K, et al. Role of pyruvate kinase M2 in transcriptional regulation leading to epithelial-mesenchymal transition. *Proc Natl Acad Sci U S A.* 2014; 111: 15526-31.
66. Xu RH, Sampsel-Barron TL, Gu F, Root S, Peck RM, Pan G, et al. NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs. *Cell Stem Cell.* 2008; 3: 196-206.
67. Brown S, Teo A, Pauklin S, Hannan N, Cho CH, Lim B, et al. Activin/Nodal signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors. *Stem Cells.* 2011; 29: 1176-85.
68. Yang S, Liu Y, Li MY, Ng CSH, Yang SL, Wang S, et al. FDX1 promotes tumor growth and metastasis by activating Wnt/ β -catenin signaling pathway and EMT in non-small cell lung cancer. *Mol Cancer.* 2017; 16: 124.
69. Pan T, Xu J, Zhu Y. Self-renewal molecular mechanisms of colorectal cancer stem cells. *Int J Mol Med.* 2017; 39: 9-20.
70. Gallo M, Carotenuto M, Frezzetti D, Camerlingo R, Roma C, Bergantino F, et al. The EGFR Signaling Modulates in Mesenchymal Stem Cells the Expression

of miRNAs Involved in the Interaction with Breast Cancer Cells. *Cancers* (Basel). 2022; 14: 1851.

71. Schinke H, Shi E, Lin Z, Quadt T, Kranz G, Zhou J, et al. A transcriptomic map of EGFR-induced epithelial-to-mesenchymal transition identifies prognostic and therapeutic targets for head and neck cancer. *Mol Cancer*. 2022; 21: 178.