



# Bioinformatics analysis of SOXF family genes reveals potential regulatory mechanism and diagnostic value in cancers

Dan Wu<sup>1</sup>, Chuan Jiang<sup>1</sup>, Jing-Jing Zheng<sup>1</sup>, De-Sheng Luo<sup>1</sup>, Ji Ma<sup>1</sup>, Hai-Feng Que<sup>1</sup>, Chao Li<sup>2</sup>, Chao Ma<sup>2</sup>, Hui-Yong Wang<sup>2</sup>, Wei Wang<sup>2</sup>, Hong-Tao Xu<sup>1</sup>

<sup>1</sup>Department of Gastrointestinal Surgery, Li Shui Municipal Central Hospital, Lishui, China; <sup>2</sup>Smartquerier Gene Technology (Shanghai) Co., Ltd., Shanghai, China

**Contributions:** (I) Conception and design: D Wu, C Jiang; (II) Administrative support: HF Que, W Wang; (III) Provision of study materials or patients: J Ma, C Ma; (IV) Collection and assembly of data: JJ Zheng, C Li; (V) Data analysis and interpretation: DS Luo, HY Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

**Correspondence to:** Hong-Tao Xu. Department of Gastrointestinal Surgery, Li Shui Municipal Central Hospital, 289 Kuocang Road, Baiyun Street, Lishui 323000, China. Email: xht0071@sina.com.

**Background:** SOXF family genes (*SOX7*, *SOX17*, *SOX18*) have been reported to be involved in tumorigenesis and development in previous articles, separately. But data sources, analysis contents and criteria are not the same. Here, we focused on SOXF genes to analyze the regulatory mechanisms and diagnostic value at the same standards.

**Methods:** This study analyzed functions, expressions, methylations, and mutations of SOXF genes through public databases including Metascape, Gene Expression Profiling Interactive Analysis (GEPIA), cBioPortal, Tumor Immune Estimation Resource (TIMER), and Kaplan-Meier Plotter. TIMER applies a deconvolution method to infer the abundance of tumor-infiltrating immune cells (TIICs) from gene expression profiles. Metascape combines several biological functions and over 40 independent knowledge bases within one integrated portal. GEPIA analyzes RNA sequencing expression data from the The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) projects. The cBioPortal visualizes and analyzes genetic data from cancer studies.

**Results:** This study found that SOXF genes had low expressions in multiple types of cancer, such as lung cancer and breast cancer (ANOVA differential methods,  $|\log_2FC|$  cutoff: 1, q value cutoff: 0.01). The lung adenocarcinoma (LUAD) patients with high expression of *SOX7* [HR =0.72 (0.61–0.85), logrank P=8.1e-05] and *SOX17* [HR =0.54 (0.45–0.64), logrank P=1.7e-12] had a higher overall survival (OS) rate. Expression of *SOX7* was significantly related to the copy number variation (CNV) (P=3.02e-8) and promoter methylation level (P=5.33e-14), while *SOX17* was only related to the promoter methylation level (P=3.32e-12). The expression of SOXF genes was positively correlated with CD4<sup>+</sup> T cell infiltration (*SOX7*: P=8.32e-07, *SOX17*: P=4.93e-06, *SOX18*: P=1.61e-11). The AUC for cg07660671 site of *SOX7*, cg15377283 site of *SOX17*, and cg24199599 site of *SOX18* in distinguishing between normal and tumor in LUAD, intestinal cancer, and breast cancer reached 0.9. SOXF genes were mainly involved in transcriptional regulation, and the Wnt signaling pathway and low expression of SOXF genes in tumor tissue had a strong negative correlation with tumor hypoxia (correlation: -0.35, P<0.001).

**Conclusions:** This study implied that the expression of *SOX7* and *SOX17* are potential prognosis markers for patients with Lung cancer and the SOXF genes methylation is potential biomarkers for pan-cancer screening. The *SOX7* and *SOX17* might modulate the Wnt signaling pathway and the expression of SOXF family genes was significantly negatively correlated with tumor hypoxia.

**Keywords:** SOXF family genes; methylation; immune infiltration; prognosis; diagnosis

Submitted May 07, 2022. Accepted for publication Jun 20, 2022.

doi: 10.21037/atm-22-2749

**View this article at:** <https://dx.doi.org/10.21037/atm-22-2749>

## Introduction

Cancer is one of the biggest public health challenges today. According to the World Health Organization (WHO), 9.6 million people died of cancer in 2018. The most common cancers include lung cancer, breast cancer, colorectal cancer, prostate cancer, skin cancer (non-melanoma), and stomach cancer (1). Tumors occur due to genetic and epigenetic changes causing more and more healthy cells to transform into cancer cells. These cancer cells are characterized by uncontrolled proliferation, high survival rate, unlimited growth, replication potential, and strong angiogenesis, invasion, and metastasis ability (2). Therefore, studies on cancer are very necessary, which can help determine the cause of the disease and explore new treatment methods to prevent, diagnose, and treat different types of cancer (3).

During the development of cancer, various key factors are involved, all of which contribute to tumorigenesis. The Hallmarks of Cancer may include genome instability & mutation, resisting cell death, sustaining proliferative signaling, non-mutational epigenetic reprogramming, avoiding immune destruction, etc. Many of the factors are identified as proto-oncogenes, and play an important role in tumorigenesis and development. For example, platelet-derived growth factor (4), insulin-like growth factor axis (5), forkhead/wing helix box transcription factor (Fox) family (6), signal transduction pathways such as Wnt (7) and Notch (8) pathways, and viruses such as human papilloma virus, Epstein-Barr virus, and the hepatitis B and C viruses.

SOX family genes are a series of important transcription factors involved in tumorigenesis and development. They contain a variety of transcriptional regulators, which can mediate DNA binding through the highly-conserved high mobility group (HMG) domain. Some of these SOX family transcription factors tightly control cell differentiation in cancer, and some are involved in progression and metastasis (9). The SOX gene family includes SOXA (*SRY*), SOXB1 (*SOX1*, *SOX2*, *SOX3*), SOXB2 (*SOX14*, *SOX21*), SOXC (*SOX4*, *SOX11*, *SOX12*), SOXD (*SOX5*, *SOX6*, *SOX13*), SOXE (*SOX8*, *SOX9*, *SOX10*), SOXF (*SOX7*, *SOX17*, *SOX18*), SOXG (*SOX15*), and SOXH (*SOX30*).

The sex determining region Y gene (*SRY*) is a member of the SRY-like-box (SOX) family of DNA binding proteins and contains a central HMG region (10). In the process of sexual development, *SRY* is a determinant of the testis (11), which is also the physiological evidence that it plays a key role in gender differences during embryonic development (12).

In addition, it has also been reported in the literature that the up-regulated expression of *SRY* is associated with poor prognosis of liver cancer, and there is no gender difference in this correlation (13). The *SOX1*, *SOX2*, and *SOX3* are the three members of SOXB1 subclass transcription factors, and they have similar sequences, expression patterns, and overexpression phenotypes (14). The data show that high levels of SRY box transcription factor 1 (*SOX1*) are associated with lower overall survival (OS) rates in some patients, and suggest that *SOX1* is a potential target for the glioma stem cell (GSC) population in glioblastoma (15). The radiation-activated PI3K/AKT pathway promotes the induction of tumor stem cell-like cells by up-regulating *SOX2* in colorectal cancer (16). As the two members of the SOXB2 subclass transcription factors, *SOX14* can promote the proliferation and invasion ability of cervical cancer cells by activating the Wnt/ $\beta$ -catenin pathway (17). The methylation of *SOX21* gene promoter has great potential to be an epigenetic biomarker for early detection of colorectal cancer (18). Together, *SOX4*, *SOX11*, and *SOX12* constitute the C group of SRY-related HMG box proteins. Both *SOX4* and *SOX11* regulate cell differentiation, proliferation, and survival in a number of basic processes, and they may function in a redundant manner to control more developmental, physiological, and pathological processes than currently known (19). The SOXD family genes (*SOX5*, *SOX6*, *SOX13*, and *SOX23*) are involved in the transcriptional regulation of developmental processes, including embryonic development, nerve growth and cartilage formation. The SOXD gene family was also identified as an important transcriptional regulator associated with a variety of cancers (20). Since the discovery of the SOX factor, members of the SOXF family (*SOX7*, *SOX17*, and *SOX18*) have been identified to play a wide range of roles, especially in cardiovascular development. Recently, SOXF factor was discovered to be a key factor in determining cell fate and regulating cancer cells (21). The *SOX15* gene can regulate the proliferation and migration of endometrial cancer cells, and the up-regulation of *SOX5* may be valuable for the treatment of endometrial cancer (22). As the main switch of desmosome genes, *SOX30* inhibits the proliferation, migration, and invasion of lung adenocarcinoma (LUAD) cells by activating the transcription of desmosome genes (23).

In this study, we focused on the investigation of SOXF family genes' regulatory mechanisms and diagnostic value using the same data sources, analysis contents and criteria. Using public databases and public data, the relationships

between gene expression, gene mutation, and gene methylation in LUAD were studied, which have not been reported in other studies. In addition, we also evaluated the potential of SOXF family genes as diagnostic markers in multiple cancers, and revealed their potential mechanisms affecting prognosis. We present the following article in accordance with the STARD reporting checklist (available at <https://atm.amegroups.com/article/view/10.21037/atm-22-2749/rc>).

## Methods

We comprehensively analyzed the biological functions of the SOXF family using a variety of tools. Firstly, functional analysis of SOX gene family was conducted, and expression analysis was done according to sub-group classification. Then, we put our focus on SOXF family. The clinical value of SOXF family expression was analyzed in LUAD. Subsequently, the factors affecting SOXF family expression were analyzed from the visions of mutation, immune microenvironment, methylation and regulation. Meanwhile, the biological functions of SOXF family were conducted, that include Gene Oncology, Pathway, single-cell level functional, etc.

### Data collection

In this study, RNAseq data in HTSeq-fragments per kilobase of exon per million reads (HTSeq-FPKM) format, clinical information, and prognostic data of The Cancer Genome Atlas (TCGA) LUAD were downloaded in Xena (<https://xenabrowser.net/>) for subsequent analysis. The RNAseq data in FPKM format was converted to transcripts per million (TPM) and processed by  $\log_2$  for analysis. In addition, the processed Beta value data of SOXF group genes corresponding to lung cancer, intestinal cancer, and breast cancer were also downloaded in this database. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

### Gene function enrichment analysis

The analysis tool Metascape (24) (<http://metascape.org/gp/index.html#/main/step1>) was used to analyze the function of the 20 member genes of the SOX gene family.

The Search Tool for the Retrieval of Interacting Genes/Proteins (25) (STRING; version 11.0, [http://www.string-](http://www.string-db.org/)

[db.org/](http://www.string-db.org/)) database was used to predict genes related to SOXF family genes. At the same time, the online analysis tool GeneMANIA (26) (<http://genemania.org/>) was used to predict gene information related to SOXF Group genes. The intersection of the results was drawn from the two databases, the intersection gene and the three genes of the SOXF family were put in R (version 3.6.3; The R Foundation for Statistical Computing, Vienna, Austria), and the clusterProfiler package (version 2.14.3) was used to perform the enrichment analysis of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway.

### Gene expression analysis

The analysis tool Gene Expression Profiling Interactive Analysis 2 (27) (GEPIA2; <http://gepia2.cancer-pku.cn/#>) analysis was used to analyze the expression of 20 gene members of the SOX gene family in multiple cancer types. The analysis function of this tool “multiple genes comparison” was used to categorize according to the family. The expression of SOX genes in pan-cancer was displayed. In addition, the profile tool under Expression DIY was selected for difference analysis, and the analysis of variance (ANOVA) method was chosen for comparison. The difference selection criterion was  $|\log_2(\text{fold change})| > 1$ ,  $q$  value  $< 0.01$ .

The analysis tool GEPIA (28) (<http://gepia.cancer-pku.cn/>) was also used in the correlation analysis between SOXF group expression and LUAD tumor stage. Later, the “Correlation function” of this tool was used to analyze the relationship between the SOXF Group gene and the expression of immune cell marker genes.

### Prognostic analysis

The analysis tool Kaplan-Meier Plotter (29) (<https://kmplot.com/analysis/index.php?P=service>) was used to analyze the prognostic value of SOXF Group expression in LUAD.

### Receiver operating characteristic curve drawing

The receiver operating characteristic (ROC) curve in this study was analyzed in R version 3.6.3. After using the R package pROC (version 1.17.0.1) for analysis, the ggplot2 (version 3.3.3) package was used to visualize the results.

### *Single factor Cox analysis*

In this study, the single factor Cox analysis method was used to explore the effects of T stage, N stage, pathological stage, gender, age, and smoking history on the OS rate of patients with LUAD. The R version used for the analysis was 3.6.3, and the R package was survival (version 3.2-10).

### *Gene mutation analysis*

The online analysis tool cBioPortal (30) (<https://www.cbioportal.org/>) was used to analyze the gene mutation frequency and main mutation types of SOXF Group genes in LUAD, and to explore the relationship between SOXF group gene copy number variation (CNV) and gene expression. At the same time, the Cancer Cell Line Encyclopedia (CCLE) database (31) (<https://depmap.org/portal/>) was used to verify the relationship between SOXF group gene CNV and gene expression.

### *Immune infiltration analysis*

The online analysis tool Tumor IMMune Estimation Resource (TIMER; <https://cistrome.shinyapps.io/timer/>) (32,33) was used to analyze the correlation between the SOXF group gene copy number and the level of immune cell infiltration.

### *Methylation analysis*

The online analysis tool University of ALabama at Birmingham CANcer data analysis Portal (UALCAN) (34) (<http://ualcan.path.uab.edu/>) was used in selecting TCGA data to analyze the methylation of SOX7, SOX17, and SOX18 gene promoter regions of normal and LUAD samples.

The online analysis tool cBioPortal was used to select LUAD (TCGA, Firehose Legacy) and Lung Adenocarcinoma (TCGA, Nature 2014), and analyze the relationship between the methylation of SOXF Group gene promoter region and gene expression in the Plots module.

We used the online analysis tool DiseaseMeth V2.0 (35) (<http://bio-bigdata.hrbmu.edu.cn/diseasemeth/analyze.html>). In LUAD, we selected ARRAY and NGS as the platform. We then selected *t*-test; significant P value: 0.01; absolute methylation difference >0.2 as the comparison method, to verify the methylation status of SOXF Group and RASSF1 genes.

### *Single cell function analysis*

The online analysis tool CancerSEA (36) (<http://biocc.hrbmu.edu.cn/CancerSEA/home.jsp>) was used to analyze the functions of SOXF group genes in single-cell sequencing data.

### *Statistical analysis*

We performed all the statistical analysis using the R package (V3.6.3) and the default statistical analysis algorithm of the online analysis tool (Metascape, STRING, GEPIA, TIMER, etc.).

## **Results**

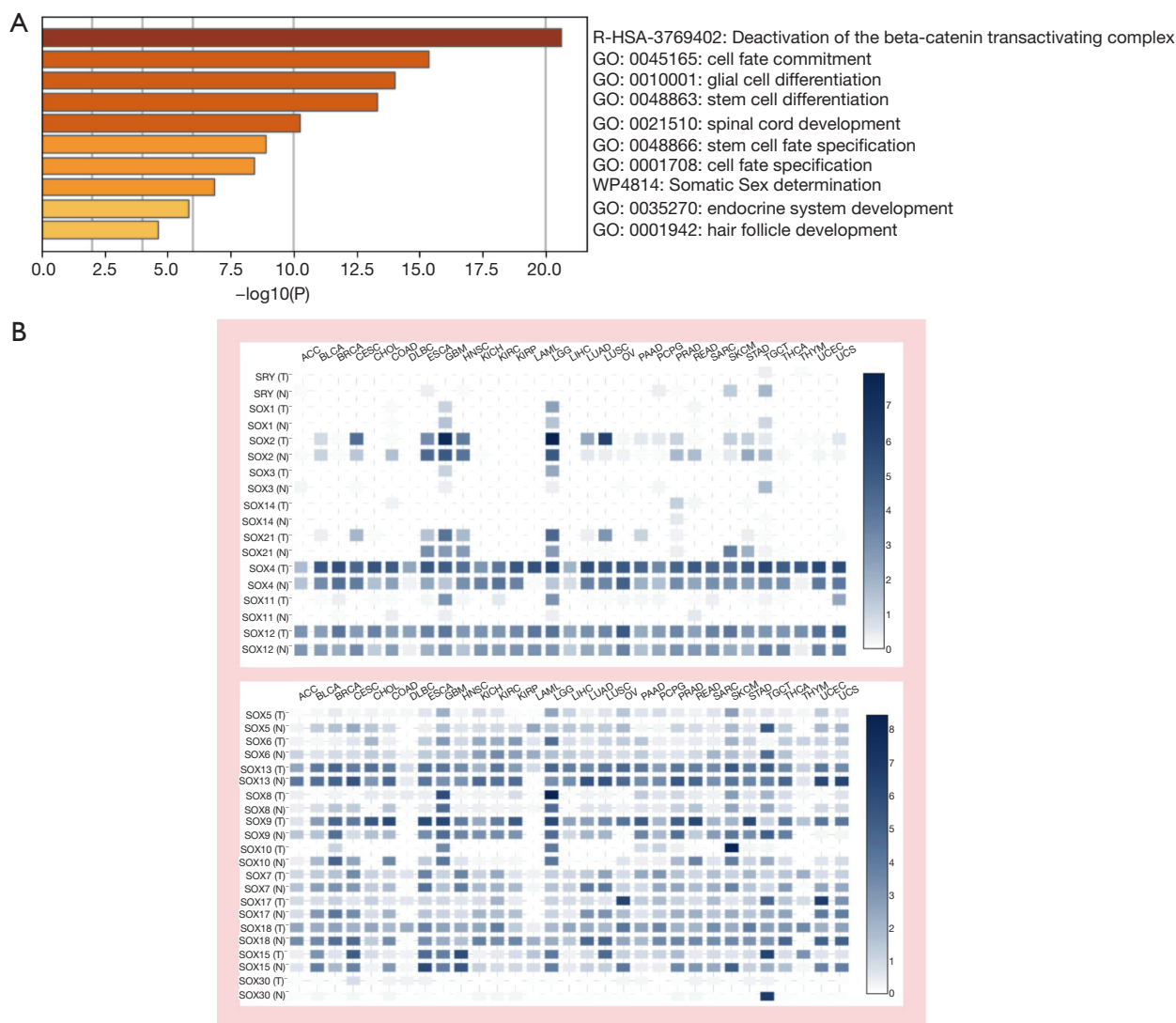
### *SOX gene family function and expression analysis*

In order to explore the functions of the SOX gene family, we conducted an enrichment analysis of the 20 genes in the SOX gene family. The results showed that genes of the family were involved in a variety of life processes including deactivation of the beta-catenin transactivating complex, cell fate commitment, and stem cell fate specification (*Figure 1A*). Subsequently, we continued to explore the expression of this gene family in different cancer types. The results showed that most of the genes in this family were abnormally expressed in a variety of cancer tissues, such as SOX4, SOX7, SOX9, SOX12, SOX13, SOX17, SOX18, and so on (*Figure 1B*). Later, we selected the SOXF (*SOX7*, *SOX17*, *SOX18*) family genes we were more interested in for further study.

### *Expression analysis of SOXF family genes*

First, we further explored the expression details of SOXF family genes (*SOX7*, *SOX17*, *SOX18*) in pan-cancer. The results showed that the expression of SOXF family genes in six cancer tissues, including lung squamous cell carcinoma (LUSC), LUAD, and breast cancer (BRCA), was significantly lower than that in normal samples (*Figure 2A-2C*). Since lung cancer is a relatively prevalent cancer type in China, LUAD was taken as an example for further exploration and research. *Figure 3A-3C* shows the expression of SOX7, SOX17, and SOX18 in LUAD, and it can be seen that their expressions in tumor tissues are significantly lower than that in normal tissues. However, there is no significant difference in their expression in different stages of LUAD (*Figure 3D-3F*).



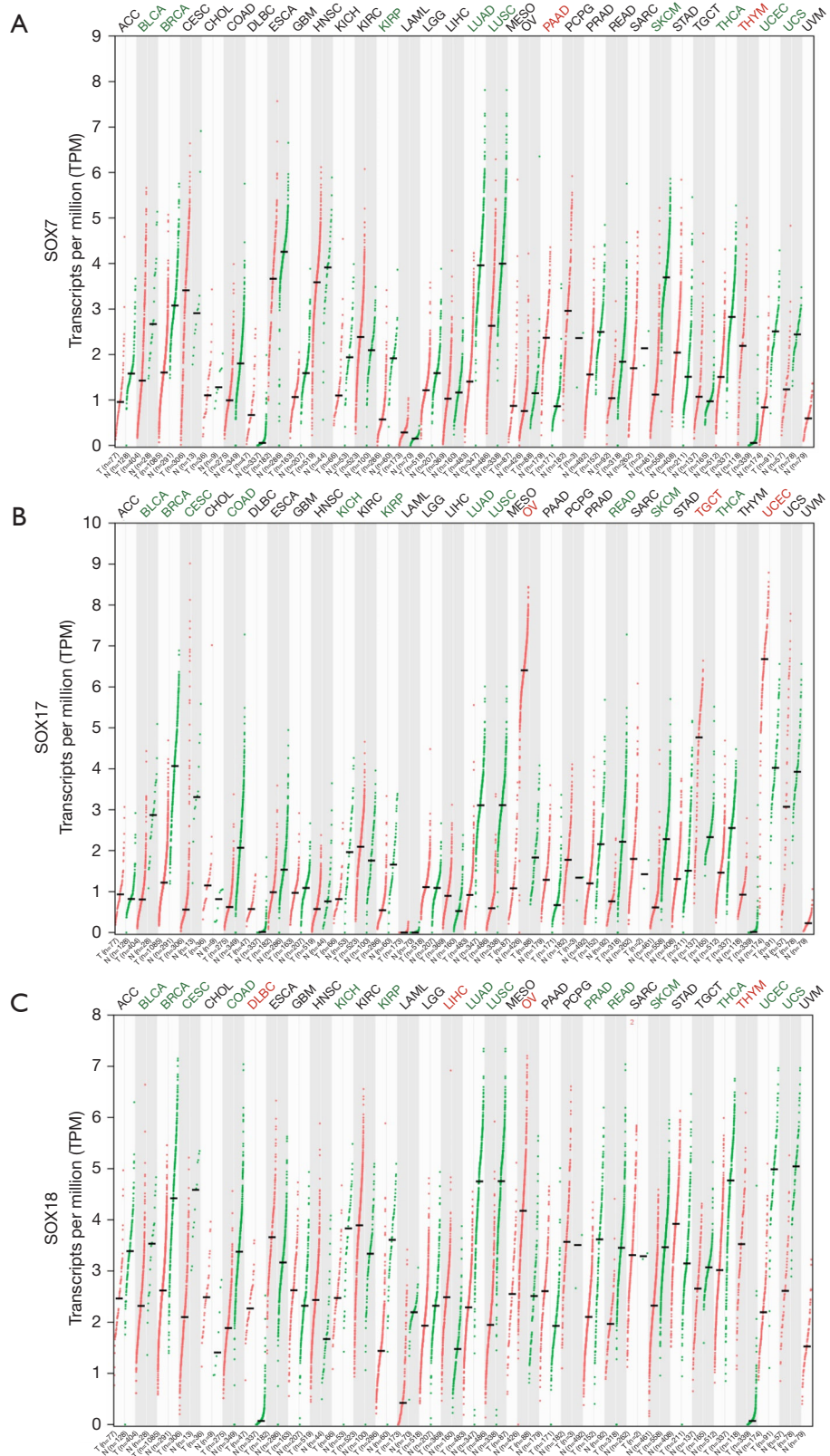


**Figure 1** SOX family gene function and expression analysis. (A) SOX family gene function enrichment analysis; (B) SOX family gene expression analysis in pan-cancer.

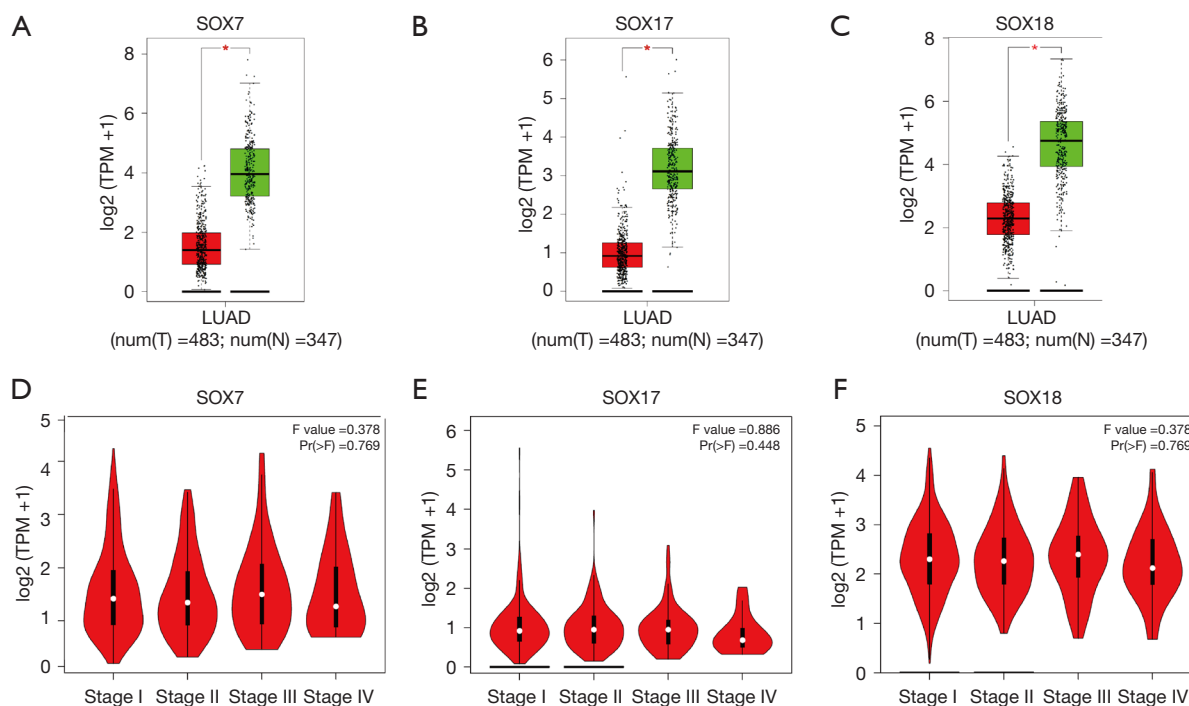
### Prognostic value evaluation of SOXF family genes

In order to explore the clinical significance of SOXF family genes, we studied the OS of patients with different SOXF family gene expressions. The results showed that patients in the SOX7 and SOX17 high expression group had a higher OS rate (Figure 4A,4B). However, there was no significant difference in OS rate for patients in the SOX18 high expression group and SOX18 low expression group (Figure 4C). Subsequently, we drew the ROC curve to evaluate the prognostic predictive ability of SOXF family genes in different stages. The results showed that the

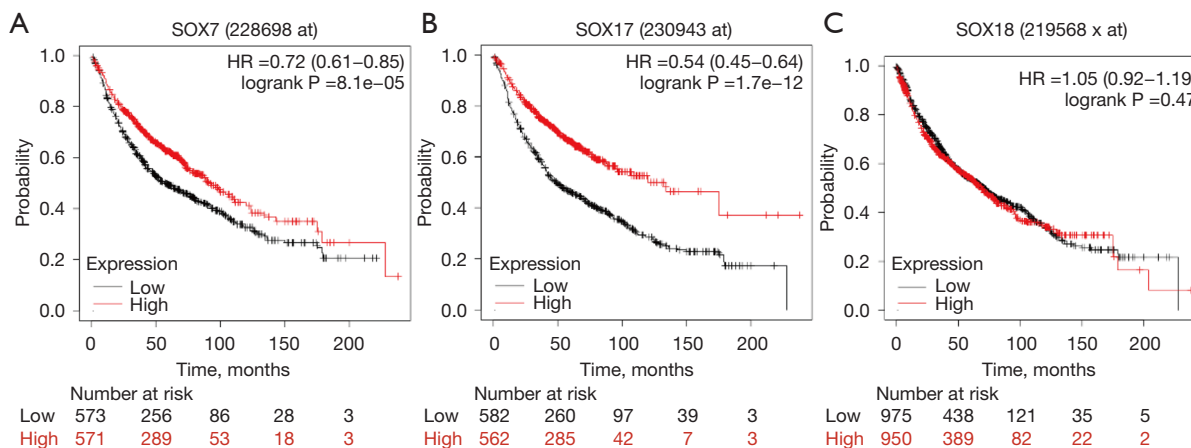
SOXF family genes showed good accuracy in the normal population and patients with LUAD (Figure 5A), and that in the case of further distinguishing cancer stages, SOXF family genes still had high accuracy for the predicting the endpoints of normal people and patients with LUAD (Figure 5B-5E). In addition, we also used single factor Cox analysis to explore the impact of different clinical factors on the OS of patients. The results showed that for LUAD patients, the later the pathological stage, the worse the OS state (Figure 5F), and that there is no relationship between smoking history, age, and gender of the patients and the OS rate of LUAD patients.



**Figure 2** The expression of SOXF family genes SOX7 (A), SOX17 (B), and SOX18 (C) in pan-cancer.



**Figure 3** The expression of SOXF family genes in lung adenocarcinoma. The expression of SOX7 (A), SOX17 (B), and SOX18 (C) in LUAD. The expression of SOX7 (D), SOX17 (E), and SOX18 (F) in different LUAD stages. \*, the gene expression is significant different between tumor and normal samples. Pr(F), is the P value of the F-statistic. LUAD, lung adenocarcinoma. num(T), number of tumor samples. num(N), number of normal samples.

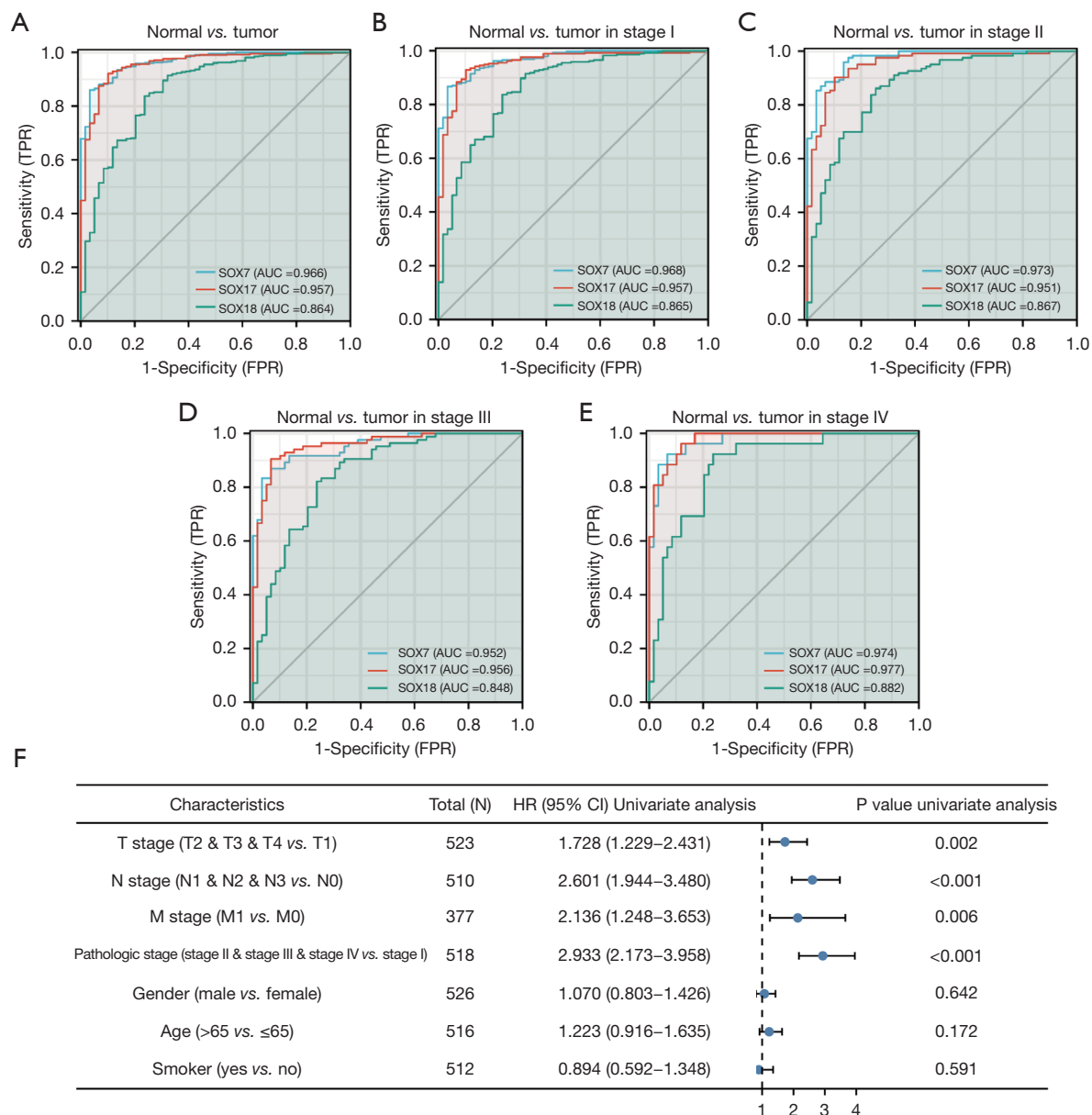


**Figure 4** Prognostic analysis. The overall survival analysis of different SOX7 (A), SOX17 (B), and SOX18 (C) expression populations. HR, hazard ratio. Logrank P, P value of the logrank test.

**Mutations of SOXF family genes**

Gene mutations are of great significance to the tumorigenesis and development of cancer. Therefore, we explored the mutations of SOXF family genes in LUAD. In various

published data (TCGA, Memorial Sloan Kettering Cancer Center (MSKCC), etc.), the overall mutation frequency of SOXF family genes is 7% (SOX7), 3% (SOX17), and 5% (SOX18) respectively (Figure 6A). The SOX7 gene has more



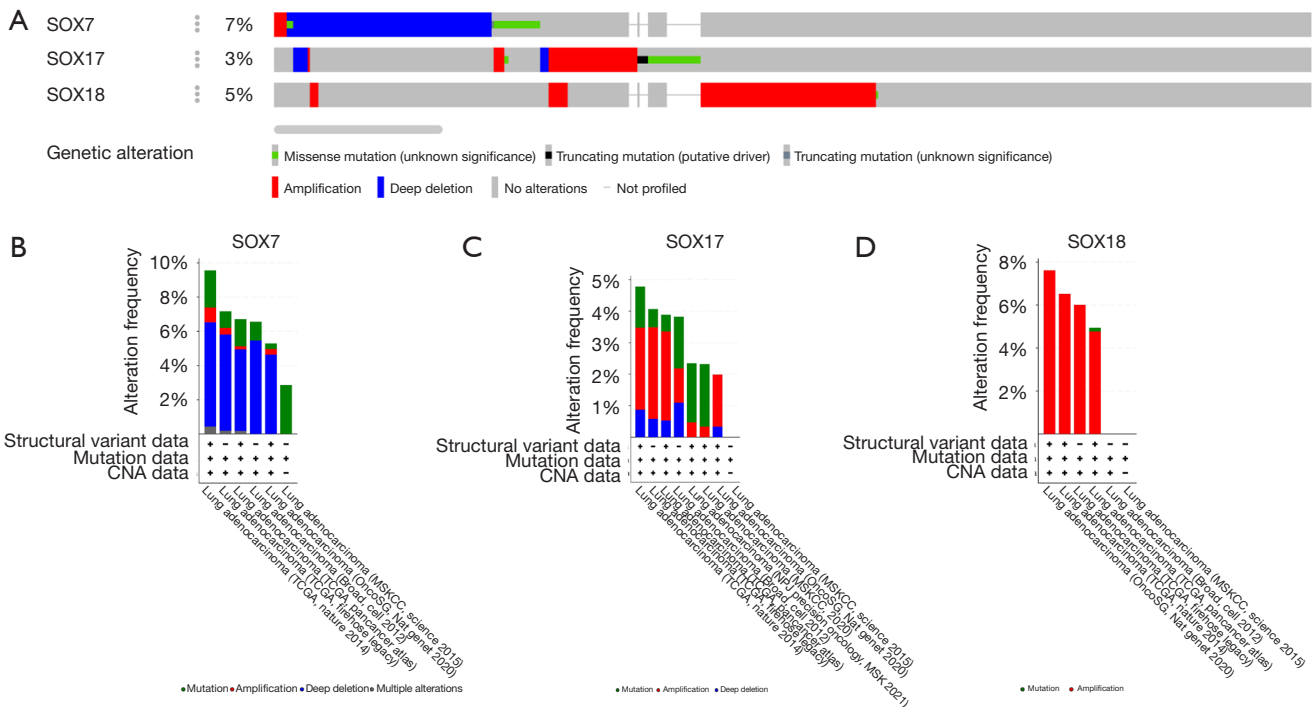
**Figure 5** ROC analysis and single factor Cox analysis. The accuracy of SOX7, SOX17, and SOX18 in predicting clinical endpoint of normal and cancer population (A), patients with different cancer stages (B-E). (F) Single factor Cox analysis. ROC, receiver operating characteristic; TPR, true positive rate; FPR, false positive rate; HR, hazard ratio; 95% CI, 95% confidence interval; T/N/M stage, cancer staging system.

deep deletion mutations (Figure 6B); SOX17 has more evenly distributed mutations, and relatively more occurrence of amplification (Figure 6C); and SOX18 has more amplification mutations (Figure 6D). In the same type of mutation data, the frequency of gene mutations is quite different.

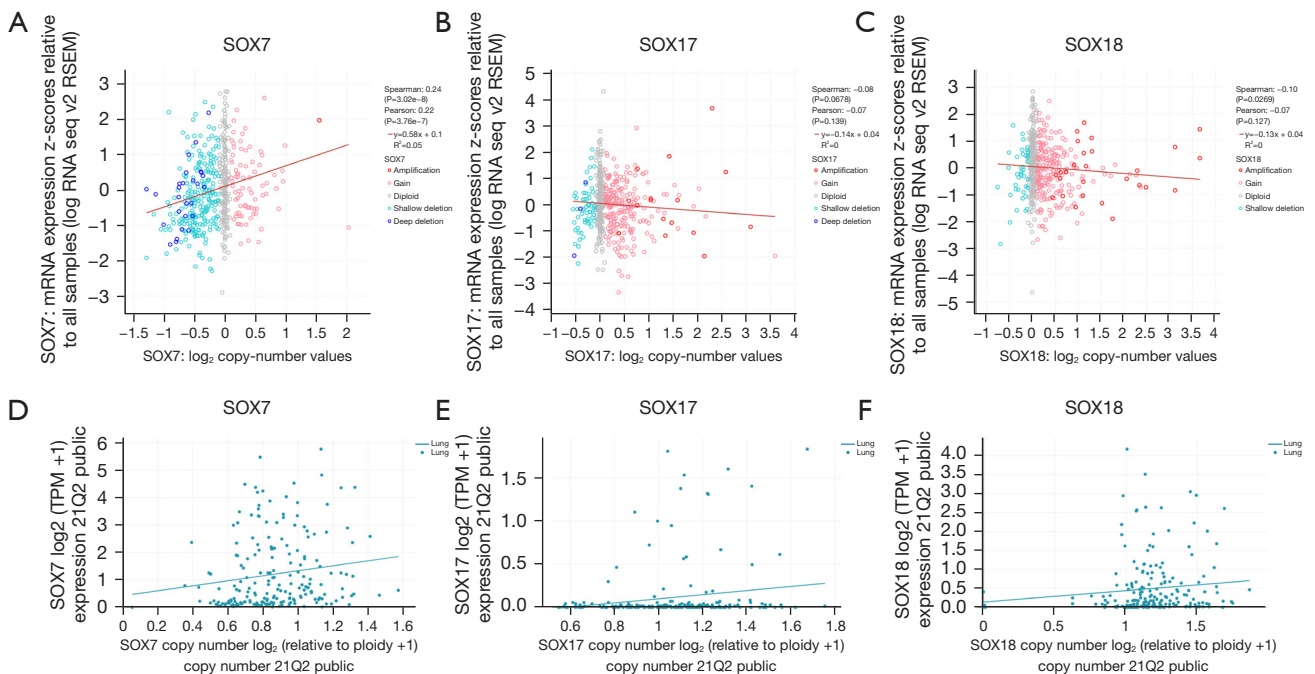
In order to explore the correlation between CNV and SOXF family gene expression, we performed correlation

analysis between the two. The results showed that the expression of SOX7 was significantly affected by the copy number (correlation coefficient was about 0.2,  $P < 0.05$ ), that deep deletion may be the main reason for the low expression of this gene (Figure 7A), and that the expression of SOX17 and SOX18 had little correlation with copy number, which may be caused by other factors (Figure 7B, 7C). The above





**Figure 6** Gene mutation analysis. (A) The overall mutation frequency of SOXF family genes. SOX7 (B), SOX17 (C), and SOX18 (D) specific mutation type analysis. +, this kind of mutation result exists in that study. -, this kind of mutation result doesn't exist in that study.



**Figure 7** Correlation analysis between SOXF group copy number variation and mRNA expression. cBioPortal database, analyzing the correlation between SOX7 (A), SOX17 (B), SOX18 (C) CNV and mRNA expression. CCLE database, analyzing the correlation between SOX7 (D), SOX17 (E), and SOX18 (F) CNV and mRNA expression. CNV, copy number variation; mRNA, messenger RNA; CCLE, cancer cell line encyclopedia.

conclusions were verified in another database (CCLE; *Figure 7D-7F*).

#### ***The relationship between SOXF group gene copy number, gene expression, and immune cell infiltration level***

The above results confirm that there is a certain relationship between gene copy number and expression, and gene expression can affect the infiltration of immune cells. Therefore, we further explored the relationship between copy number and immune cell infiltration level. The results showed that in LUAD, the B cell and CD8<sup>+</sup> T cell immune infiltration levels were significantly reduced in the SOX7 high amplification population (*Figure 8A*); the B cell, CD4<sup>+</sup> T cell, macrophage cell, neutrophils, and dendritic cell (DC) infiltration levels were reduced in the SOX17 high amplification population (*Figure 8B*); only the DC infiltration level was significantly reduced in the SOX18 high amplification population (*Figure 8C*); and that SOXF group gene expression was positively correlated with CD4<sup>+</sup> T cell infiltration (*Figure 8D-8F*); macrophage cell infiltration was positively correlated with the expression of SOX7 and SOX17 (*Figure 8D,8E*); and neutrophils and DC infiltration had a more significant relationship with SOX7 expression (*Figure 8D*). This is roughly the same as the relationship between CNV and immune cell infiltration.

In order to further explain the influence of SOXF group gene expression on immune cells, we explored the relationship between SOXF group gene and some immune cell markers. The results showed that SOX7 expression was highly positively correlated with *CD4* expression, while SOX17 and SOX18 showed a weak correlation. However, their *NOS2* expression was highly positively correlated. This is consistent with the results of the previous immune infiltration investigation (*Table 1*).

#### ***SOXF group gene family methylation analysis and biomarker exploration***

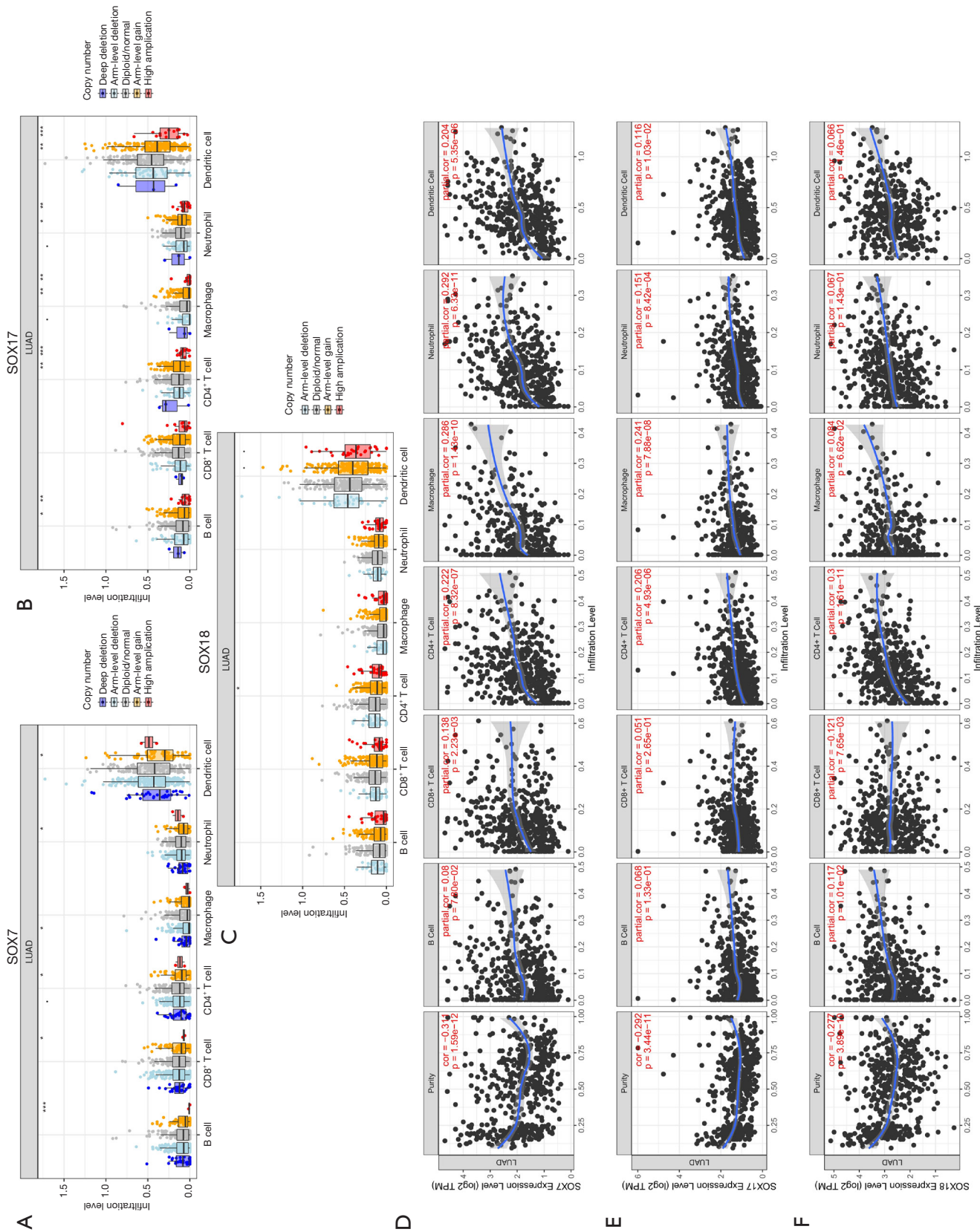
Gene methylation will also affect gene expression (37), so we further explored the methylation of the SOXF group gene family. The results showed that in people with of pathological stages (*Figure 9A-9C*), different age groups (*Figure 9D-9F*), different genders (*Figure 9G-9I*), and different lymph node metastases (*Figure 9J-9L*), the promoter methylation level of SOXF family genes in normal samples was very low, while the promoter methylation level in tumor samples was significantly increased. However, the promoter methylation

level of SOX18 in normal samples was relatively high, higher than that of SOX7 and SOX17. Subsequently, correlation analysis was performed in the different LUAD cohorts of the two TCGAs to explore the relationship between methylation and gene expression in promoter region. The results showed that SOX7 (*Figure 10A,10B*) and SOX17 (*Figure 10C,10D*) methylation significantly negatively regulated gene expression. For SOX18, there was no such result (*Figure 10E,10F*).

The methylation in the *RASSF1* gene promoter region can be used as an effective biomarker for the early diagnosis of lung cancer (38). Therefore, we compared the methylation status of the SOXF group with the methylation status of *RASSF1* to evaluate whether SOXF family genes can be used as potential biomarkers. First, we observed the methylation status of *RASSF1* in different pathological stages. The methylation status of this gene was low in the normal group and the methylation level in the tumor tissues was high (*Figure 11*). In comparison of tumor and normal samples, the methylations of four gene promoter regions were all significantly different, with a P value of <0.01. Among them, the difference between the average methylation degree of SOX17 gene tumor sample and the average methylation degree of normal sample was 0.194, which was the most significant of the four genes, and the degree of difference was better than that of *RASSF1* (*Table 2*). In order to further evaluate the feasibility of SOXF group gene methylation for early diagnosis of cancer, its diagnostic capability was explored in LUAD, intestinal cancer, and breast cancer. It was found that the AUC for the cg07660671 site of SOX7, the cg15377283 site of SOX17, and the cg24199599 site of SOX18 in distinguishing between normal and tumor tissue in LUAD, intestinal cancer, and breast cancer reached 0.9 (*Table 3*).

#### ***SOXF group related genes and function analysis***

In order to further analyze the functions of SOXF family genes, we used two online databases (*Figure 12A,12B*) to predict the proteins related to SOXF family genes and constructed a protein interaction network. After taking intersection of the results of the two databases, six related genes were obtained (*Figure 12C*), namely: *CTNNB1*, *LEF1*, *POU5F1*, *TCF7*, *TCF7L1*, and *TCF7L2*. Then, three SOXF family genes and the above six related genes were analyzed by KEGG and GO. The results showed that these nine genes mainly participated in processes such as transcriptional regulation, Wnt signaling pathway, and so



**Figure 8** Correlation analysis of immune cell infiltration. Correlation analysis between SOX7 (A), SOX17 (B), and SOX18 (C) copy number and immune infiltration. Correlation analysis of SOX7 (D), SOX17 (E), and SOX18 (F) expression and immune infiltration. LUAD, the lung adenocarcinoma. \*, cutoff 0.1; \*\*, cutoff 0.01; \*\*\*, cutoff 0.001. cor, correlation.

**Table 1** Correlation analysis between SOXF group and immune cells biomarkers in LUAD by using the GEPIA database

| Immune cell             | Biomarker | SOX7    |          | SOX17   |          | SOX18   |          |
|-------------------------|-----------|---------|----------|---------|----------|---------|----------|
|                         |           | R value | P value  | R value | P value  | R value | P value  |
| B cell                  | CD19      | 0.21    | 2.90E-06 | 0.2     | 9.90E-06 | 0.18    | 5.30E-05 |
|                         | CD79A     | 0.18    | 5.30E-05 | 0.19    | 3.50E-05 | 0.17    | 2.60E-04 |
| CD8 <sup>+</sup> T cell | CD8A      | 0.15    | 0.0011   | 0.085   | 0.061    | 0.064   | 0.16     |
|                         | CD8B      | 0.069   | 0.13     | 0.049   | 0.28     | 0.032   | 0.49     |
| CD4 <sup>+</sup> T cell | CD4       | 0.45    | 2.90E-25 | 0.37    | 2.50E-17 | 0.3     | 2.30E-11 |
| M1 macrophage           | NOS2      | 0.39    | 1.40E-18 | 0.43    | 1.70E-23 | 0.4     | 1.20E-19 |
|                         | IRF5      | 0.2     | 1.10E-05 | 0.13    | 3.10E-03 | 0.18    | 7.40E-05 |
|                         | PTGS2     | 0.21    | 1.90E-06 | 0.2     | 8.00E-06 | 0.089   | 5.20E-02 |
| M2 macrophage           | CD163     | 0.33    | 4.10E-14 | 0.31    | 4.50E-12 | 0.2     | 1.20E-05 |
|                         | VSIG4     | 0.27    | 2.70E-09 | 0.22    | 1.00E-06 | 0.12    | 7.80E-03 |
|                         | MS4A4A    | 0.34    | 1.10E-14 | 0.28    | 2.80E-10 | 0.14    | 2.10E-03 |
| Neutrophil              | CEACAM8   | 0.19    | 2.80E-05 | 0.19    | 1.70E-05 | 0.073   | 1.10E-01 |
|                         | ITGAM     | 0.34    | 1.80E-14 | 0.27    | 1.00E-09 | 0.23    | 2.40E-07 |
|                         | CCR7      | 0.35    | 1.30E-15 | 0.29    | 1.50E-10 | 0.21    | 2.30E-06 |
| Dendritic cell          | HLA-DPB1  | 0.28    | 4.40E-10 | 0.21    | 4.00E-06 | 0.17    | 1.30E-04 |
|                         | HLA-DQB1  | 0.11    | 1.30E-02 | 0.12    | 7.10E-03 | 0.17    | 1.40E-04 |
|                         | HLA-DRA   | 0.21    | 4.40E-06 | 0.13    | 3.50E-03 | 0.067   | 1.40E-01 |
|                         | HLA-DPA1  | 0.23    | 2.90E-07 | 0.15    | 9.30E-04 | 0.096   | 3.60E-02 |
|                         | CD1C      | 0.29    | 1.50E-10 | 0.24    | 5.60E-08 | 0.14    | 2.10E-03 |
|                         | NRP1      | 0.22    | 1.70E-06 | 0.19    | 1.80E-05 | 0.09    | 4.90E-02 |
|                         | ITGAX     | 0.34    | 1.00E-14 | 0.26    | 5.40E-09 | 0.29    | 4.30E-11 |

LUAD, lung adenocarcinoma; GEPIA, Gene Expression Profiling Interactive Analysis.

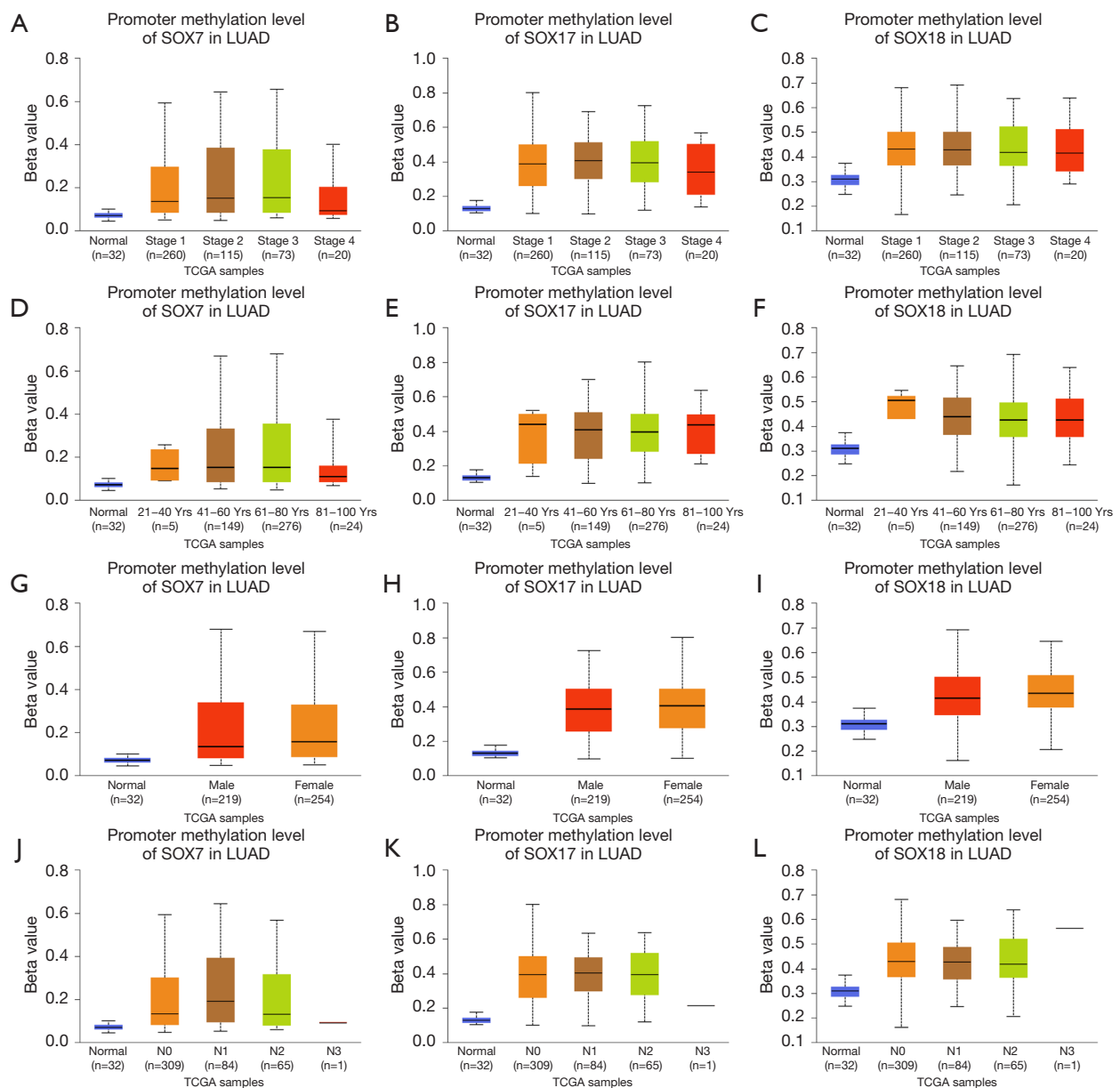
on (Figure 12D). Subsequently, a more in-depth analysis of SOXF group functions was performed through the CancerSEA single-cell sequencing database. From the results, it can be concluded that the expression of SOXF Group was significantly negatively correlated with tumor hypoxia (Figure 13). In the hypoxia environment, in order to maintain growth and proliferation, tumor cells will change their metabolism and cell behaviors, which may be related to extracellular matrix remodeling and increased migration and metastasis behavior (39).

## Discussion

This article first explored the functions of SOX family genes through enrichment analysis and found that the SOX

gene family was mainly involved in signal transduction, cell fate determination, stem cell differentiation, and other pathways and functions, which is consistent with previous reports on breast cancer (40) and ovarian cancer (41). Then the expression of SOX family genes in pan-cancer was observed. The results showed that most genes of this family were abnormally expressed in a variety of cancer tissues, such as SOX4 (42), SOX7 (43), SOX9 (44), SOX12 (45), SOX13 (46), SOX17 (47), SOX18 (48,49), and so on. Subsequently, we conducted follow-up studies taking SOXF family genes as the main research object.

The SOXF group includes three genes, SOX7, SOX17, and SOX18. The expression levels of these three genes in a variety of cancers, including lung cancer (49,50), intestinal cancer (51), breast cancer (50), and so on, were significantly

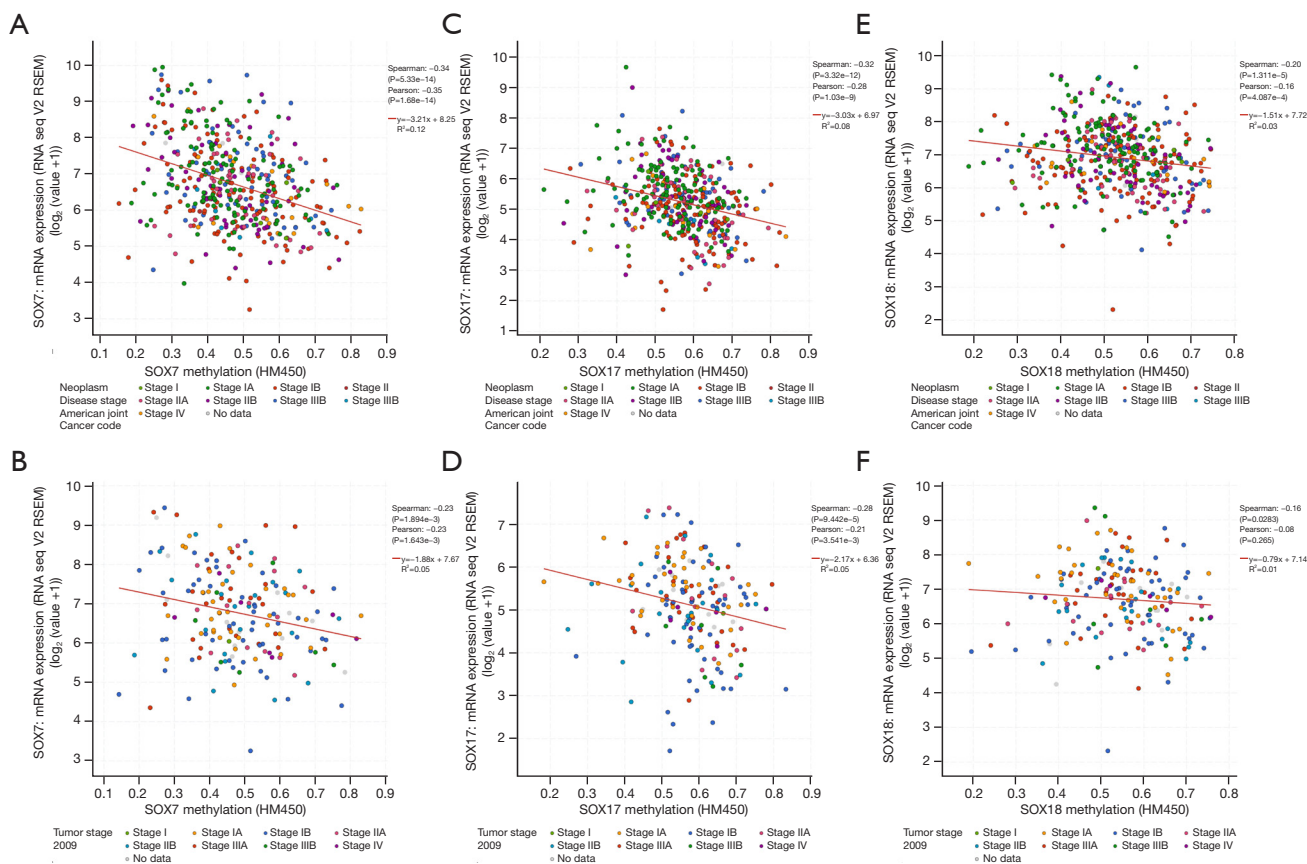


**Figure 9** SOXF Group promoter methylation assessment. The methylation status of SOX7 (A), SOX17 (B), and SOX18 (C) in different pathological stages. The methylation status of SOX7 (D), SOX17 (E), and SOX18 (F) in different age groups. The methylation status of SOX7 (G), SOX17 (H), and SOX18 (I) in different genders. The methylation status of SOX7 (J), SOX17 (K), and SOX18 (L) in people with different degrees of lymph node metastasis. Beta Value, represents the ratio between the methylated array intensity and total array intensity, falls between 0 (lower levels of methylation) and 1 (higher levels of methylation).

lower than those in normal samples. All of them were solid tumors, which might result from the function of SOX7, SOX17, and SOX18. It was reported that SOX7 (52), SOX17 (53), and SOX18 (54) are related to angiogenesis. Angiogenesis is considered the key and early step of

tumorigenesis. Furthermore, it was also revealed as a marker of solid tumor and a key promoter of tumor recurrence (55). Hence, these genes showed abnormal expression in most of solid tumors to promote the cancer development. Then, taking LUAD as an example, we studied the relationship





**Figure 10** The correlation analysis SOXF group methylation and gene expression. The correlation analysis of SOX7 (A), SOX17 (C), and SOX18 (E) promoter region methylation and gene expression (TCGA Firehose data set). The correlation analysis of SOX7 (B), SOX17 (D), and SOX18 (F) promoter region methylation and gene expression (TCGA Nature data set). TCGA, The Cancer Genome Atlas.

between the expression, mutation, and methylation of SOXF family genes and their prognosis. The results showed that the expression of SOXF family genes decreased in LUAD, but there was no significant difference in their expression in different LUAD pathological stages. Subsequently, according to the expression level of SOXF family genes, the patients were divided into a high expression group and low expression group. It was found that the patients in the SOX7 and SOX17 high expression groups had better OS rate, although there was no significant difference in the survival rate of the patients in SOX18 high expression group and low expression group. This may be related to their own gene functions. According to report, SOX7 is a tumor suppressor. Highly expressed SOX7 can inhibit the cell cycle, promote cell apoptosis, and thereby inhibit cancer (56). There is also similar report on SOX17 in esophageal squamous cell carcinoma. Highly expressed SOX17 can reduce cell migration ability and slow down

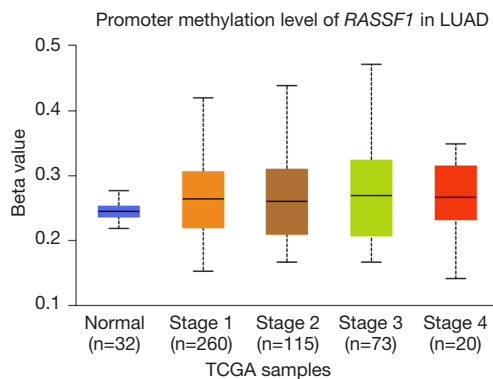
tumor growth (57). This also explains why SOX7 and SOX17 high expression group patients had a higher OS rate. In addition, we also analyzed the predictive ability of SOXF family genes for the clinical endpoints of cancer patients at different stages, and the results showed that all three genes had good predictive capabilities.

In order to further explore the reasons for the differential expression of SOXF family genes, we analyzed the mutations of three genes. The results showed that SOX7 had relatively more deep deletion mutations; SOX17 had more evenly distributed mutations; and SOX18 has more amplification mutations. According to report, the CNV is linearly correlated to gene expression (58), which may be one of the reasons for the abnormal expression of SOXF family genes in cancer tissues. Subsequently, we also performed a correlation analysis on the relationship between the copy number and the expression of SOXF family genes. The results showed that only SOX7 had a copy number

significantly correlated with its expression, suggesting that the abnormal expression of SOX7 may be mainly caused by CNV. Since gene expression can help predict immune cell infiltration, which is also one of the main reasons that affect the OS of patients, we also explored the relationship between SOXF family copy number and gene expression and immune cell infiltration. The B cell and CD8<sup>+</sup> T cell immune infiltration levels were significantly reduced in the SOX7 high amplification population. The B cell, CD4<sup>+</sup> T cell, macrophage cell, neutrophils, and DC infiltration levels were reduced in the SOX17 high amplification population. Only the DC infiltration level was significantly reduced in the SOX18 high amplification population. Although the gene expression of SOX7 was related to its CNV, the CNV of SOX17 and SOX18 also had a certain relationship with immune cell infiltration, suggesting that SOX17 and SOX18 may be more inclined to control downstream genes to achieve regulatory goals during the regulation process. The SOXF group gene expression was positively correlated with CD4<sup>+</sup> T cell infiltration, and macrophage cell infiltration

was positively correlated with the expression of SOX7. The LUAD patients enriched with CD4<sup>+</sup> T and macrophage cells had a better prognosis (59). The DC have been shown to perform antigen presentation, which in turn triggers the anti-tumor immune response mediated by T cells (60). This also indicates that perhaps affecting immune infiltration is one of the ways that SOXF family genes affect the prognosis of patients. The SOXF family genes are positively correlated with the expression of immune cell marker genes *CD4* and *NOS2*, which again confirms the above inference. Further, the activation of SOX7 could modulate interleukin (IL)-33 to recruit tumor-associated macrophages leading to metastasis, and SOX17 deletion also resulted in decreased inflammatory DC in the lungs (61).

As only SOX7 had a significant correlation between CNV and its expression, we continued to explore factors that may affect the expression of SOX17 and SOX18 from other perspectives. It has been reported that the methylation of the promoter region can affect the expression of genes (37). Therefore, we explored the methylation of SOXF family genes. The results showed that in people of different pathological stages, age groups, genders, and lymph node metastases, the promoter methylation level of SOXF family genes in normal samples was very low, and promoter methylation in tumor samples was significantly increased. The methylation of SOX7 and SOX17 significantly negatively regulated gene expression. This indicates that the low expression of SOX7 in LUAD is not only due to CNV, but is also affected by the level of methylation. The expression of SOX17 is more affected by the level of methylation, but the factors affecting the changes in SOX18 expression still need to be explored. It has been reported that the methylation of the promoter region of the *RASSF1* gene can be used as an effective biomarker for the early diagnosis of lung cancer (37). To explore whether SOXF family genes have similar capabilities, we compared SOXF



**Figure 11** The methylation status of *RASSF1* in different pathological stages. LUAD, lung adenocarcinoma; TCGA, The Cancer Genome Atlas.

**Table 2** Verification of methylation results

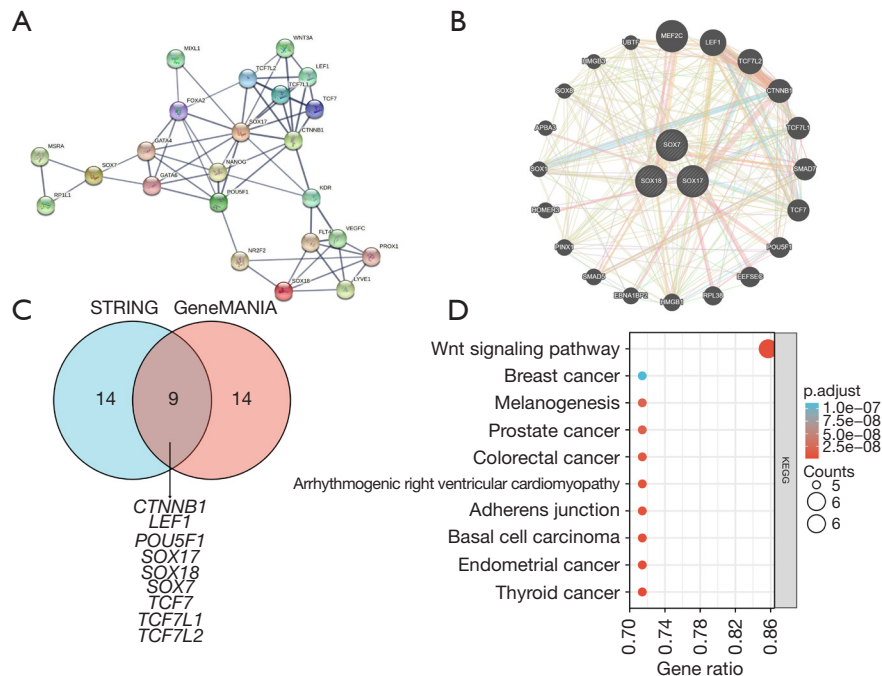
| Tumor | Genomic region          | Transcript | Gene          | P value   | Mean (MethylTumor)-Mean (MethylNormal) |
|-------|-------------------------|------------|---------------|-----------|--|
| LUAD  | chr3:50377867-50380367  | NM_007182  | <i>RASSF1</i> | 0.000e+00 | 0.122                                  |
| LUAD  | chr8:55368494-55370994  | NM_022454  | <i>SOX17</i>  | 0.000e+00 | 0.194                                  |
| LUAD  | chr20:62680479-62682979 | NM_018419  | <i>SOX18</i>  | 1.624e-12 | 0.099                                  |
| LUAD  | chr8:10587584-10590084  | NM_031439  | <i>SOX7</i>   | 1.624e-12 | 0.114                                  |

LUAD, lung adenocarcinoma; Mean (MethylTumor), mean of methylation level in tumor samples; Mean (MethylNormal), mean of methylation level in normal samples.

**Table 3** SOXF Group pan-cancer methylation marker analysis

| Cancer type | Gene  | Probe      | AUC (best) | Sensitivity | Specificity |
|-------------|-------|------------|------------|-------------|-------------|
| LUAD        | SOX7  | cg07660671 | 0.924      | 0.841       | 1           |
| LUAD        | SOX17 | cg15377283 | 0.978      | 0.931       | 1           |
| LUAD        | SOX18 | cg24199599 | 0.915      | 0.865       | 1           |
| COAD        | SOX7  | cg07660671 | 0.923      | 0.847       | 1           |
| COAD        | SOX17 | cg26059468 | 0.982      | 0.922       | 1           |
| COAD        | SOX18 | cg23509896 | 0.973      | 0.942       | 1           |
| BRCA        | SOX7  | cg22008625 | 0.920      | 0.779       | 0.952       |
| BRCA        | SOX17 | cg00123055 | 0.956      | 0.854       | 0.976       |
| BRCA        | SOX18 | cg24199599 | 0.885      | 0.748       | 0.940       |

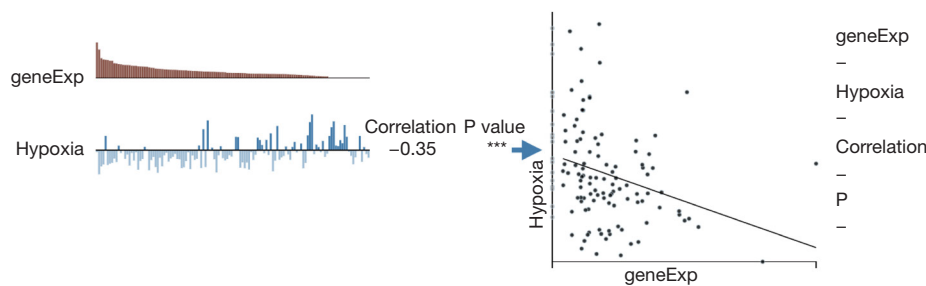
LUAD, lung adenocarcinoma; COAD, colon adenocarcinoma; BRCA, breast invasive carcinoma; AUC, area under the curve.



**Figure 12** Functional analysis. (A) STRING analyzing the protein interaction network of SOXF family genes. (B) GeneMANIA predicting the SOXF family gene-related gene network. (C) Venn diagram, searching for the intersection genes predicted by STRING and GeneMANIA. (D) Pathway enrichment analysis of SOXF family genes and their function-related genes. STRING, Search Tool for the Retrieval of Interacting Genes/Proteins; GeneMANIA, tool for predicting gene sets function.

family genes with *RASSF1*. In comparison of tumor and normal samples, the methylations of four gene promoter regions were all significantly different. Among them, the difference between the average methylation degree of SOX17 gene tumor sample and the average methylation

degree of normal sample was 0.194, and the degree of difference was better than that of *RASSF1*. Through further analysis, it was found that the AUC for the cg07660671 site of SOX7, the cg15377283 site of SOX17, and the cg24199599 site of SOX18 in distinguishing between



**Figure 13** SOXF group single cell function analysis. \*\*\*, the gene expression of SOXF group is significant related with hypoxia. geneExp, gene expression.

normal and tumor in LUAD, intestinal cancer, and breast cancer reached 0.9, indicating that SOXF family genes have the potential to become the biomarkers for methylation in lung cancer diagnosis.

Finally, we analyzed the function of SOXF family genes and found six related genes, *CTNNB1*, *LEF1*, *POU5F1*, *TCF7*, *TCF7L1*, and *TCF7L2*. After enrichment analysis, it was found that these genes mainly participated in transcriptional regulation, the Wnt signaling pathway, and a variety of cancer signaling pathways. These pathways are of great significance to the tumorigenesis and development of cancer (7,62), further confirming the important value of these genes for the exploration of cancer. The SOX7 gene disrupts  $\beta$ -catenin (*CTNNB1*)/*BCL9* interaction to suppress Wnt signaling and also inhibits *CTNNB1*/TCF-mediated transcription (63). The SOX17 gene decreased expression of  $\beta$ -catenin and proteins in the Wnt signaling pathway (64), whereas  $\beta$ -catenin was demonstrated to be inversely correlated with CD8<sup>+</sup> T cell infiltration (65). One might assume that inhibition of Wnt/ $\beta$ -catenin signaling can improve CD8<sup>+</sup> T cell infiltration and initiation, so it may produce a more favorable immune checkpoint inhibition scheme (66). Conversely, the induction of Wnt signaling plays an important role in maintaining the dryness of memory CD8<sup>+</sup> T cells by blocking T cell differentiation (67). The Wnt/ $\beta$ -catenin pathway has also been associated with modulation of innate immunity, such as DC (68). Hence, SOX7 and SOX17 might modulate  $\beta$ -catenin and TCF to regulate the Wnt signaling pathway, resulting in the change of immune cell infiltration. In most solid tumors, due to insufficient oxygen supply from abnormal blood vessels, the needs of rapid proliferation required by cancer cells cannot be met, and hypoxia occurs to a certain extent. The oxygenation level in the same tumor varies greatly between different regions and will change over time. Tumor hypoxia

is an important obstacle to effective tumor treatment. In radiotherapy, the main mechanism is producing reactive oxygen species; therefore, tumor hypoxia is radiation resistant (69) and contributes to tumor metastasis (39). In our study, we found that the low expression of SOXF group genes in tumor tissue had a strong negative correlation with tumor hypoxia. This also provides more theoretical support for the promotion effect of SOXF group gene low expression in tumors.

## Conclusions

In this article, we integrated bioinformatics tools and public databases to explore the mechanism and diagnostic value of the SOXF gene family. First, we found that the SOXF family genes are lowly expressed in multiple cancer types, but the mechanisms that affect the expression of each gene are very different. For the SOX7 gene, deep deletion, promoter methylation, and immune infiltration were found to affect its expression in tumor samples together. For the SOX17 gene, methylation and immune infiltration are the main factors that affected its expression. For the SOX18 gene, amplification and methylation play important role in its expression. These methods provide new ideas for study of gene expression mechanism in cancer.

Second, through systematic analysis of gene expression and methylation in multiple cancer types, SOXF genes expression affect the prognosis of patients and the promoter methylation are potential biomarkers for pan-cancer screening. Diagnostic methods and reagents can be developed based on these results.

## Acknowledgments

*Funding:* None.

## Footnote

**Reporting Checklist:** The authors have completed the STARD reporting checklist. Available at <https://atm.amegroups.com/article/view/10.21037/atm-22-2749/rc>

**Conflicts of Interest:** All authors have completed the ICMJE uniform disclosure form (available at <https://atm.amegroups.com/article/view/10.21037/atm-22-2749/coif>). CL, CM, HYW, and WW are from Smartquerier Gene Technology (Shanghai) Co., Ltd. The other authors have no conflicts of interest to declare.

**Ethical Statement:** The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

**Open Access Statement:** This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

- Koni M, Pinnarò V, Brizzi MF. The Wnt Signalling Pathway: A Tailored Target in Cancer. *Int J Mol Sci* 2020;21:7697.
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144:646-74.
- Grimm D, Bauer J, Wise P, et al. The role of SOX family members in solid tumours and metastasis. *Semin Cancer Biol* 2020;67:122-53.
- Jitariu AA, Raica M, Cîmpean AM, et al. The role of PDGF-B/PDGFR-BETA axis in the normal development and carcinogenesis of the breast. *Crit Rev Oncol Hematol* 2018;131:46-52.
- Mutgan AC, Besikcioglu HE, Wang S, et al. Insulin/IGF-driven cancer cell-stroma crosstalk as a novel therapeutic target in pancreatic cancer. *Mol Cancer* 2018;17:66.
- Chen C, Aihemaiti M, Zhang X, et al. FOXD4 induces tumor progression in colorectal cancer by regulation of the SNAI3/CDH1 axis. *Cancer Biol Ther* 2018;19:1065-71.
- Taciak B, Pruszyńska I, Kiraga L, et al. Wnt signaling pathway in development and cancer. *J Physiol Pharmacol* 2018;69:10.26402/jpp.2018.2.07.
- Gao J, Long B, Wang Z. Role of Notch signaling pathway in pancreatic cancer. *Am J Cancer Res* 2017;7:173-86.
- Abadi AJ, Zarrabi A, Hashemi F, et al. The role of SOX family transcription factors in gastric cancer. *Int J Biol Macromol* 2021;180:608-24.
- Phillips NB, Jancso-Radek A, Ittah V, et al. SRY and human sex determination: the basic tail of the HMG box functions as a kinetic clamp to augment DNA bending. *J Mol Biol* 2006;358:172-92.
- Jordan BK, Vilain E. Sry and the genetics of sex determination. *Adv Exp Med Biol* 2002;511:1-13; discussion 13-4.
- Sekido R, Lovell-Badge R. Sex determination and SRY: down to a wink and a nudge? *Trends Genet* 2009;25:19-29.
- Xue TC, Zhang L, Ren ZG, et al. Sex-determination gene SRY potentially associates with poor prognosis but not sex bias in hepatocellular carcinoma. *Dig Dis Sci* 2015;60:427-35.
- Archer TC, Jin J, Casey ES. Interaction of SOX1, SOX2, SOX3 and Oct4 during primary neurogenesis. *Dev Biol* 2011;350:429-40.
- Garcia I, Aldaregia J, Marjanovic Vicentic J, et al. Oncogenic activity of SOX1 in glioblastoma. *Sci Rep* 2017;7:46575.
- Restoration and compensation of functional disorders in patients with injuries to the spine and spinal cord using exercise therapy. *Vopr Kurortol Fizioter Lech Fiz Kult* 1987;(6):48-50.
- Li F, Wang T, Tang S. SOX14 promotes proliferation and invasion of cervical cancer cells through Wnt/ $\beta$ -catenin pathway. *Int J Clin Exp Pathol* 2015;8:1698-704.
- Moradi K, Babaei E, Rezvani N, et al. High potential of SOX21 gene promoter methylation as an epigenetic biomarker for early detection of colorectal cancer. *Indian J Cancer* 2020;57:166-71.
- Seok J, Gil M, Dayem AA, et al. Multi-Omics Analysis of SOX4, SOX11, and SOX12 Expression and the Associated Pathways in Human Cancers. *J Pers Med* 2021;11:823.
- Liang Z, Xu J, Gu C. Novel role of the SRY-related high-mobility-group box D gene in cancer. *Semin Cancer Biol* 2020;67:83-90.
- Higashijima Y, Kanki Y. Molecular mechanistic insights:



- The emerging role of SOXF transcription factors in tumorigenesis and development. *Semin Cancer Biol* 2020;67:39-48.
22. Rui X, Xu Y, Jiang X, et al. SOX15 regulates proliferation and migration of endometrial cancer cells. *Biosci Rep* 2017;37:BSR20171045.
  23. Hao X, Han F, Ma B, et al. SOX30 is a key regulator of desmosomal gene suppressing tumor growth and metastasis in lung adenocarcinoma. *J Exp Clin Cancer Res* 2018;37:111.
  24. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019;10:1523.
  25. Szklarczyk D, Gable AL, Nastou KC, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* 2021;49:D605-12.
  26. Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 2010;38:W214-20.
  27. Tang Z, Kang B, Li C, et al. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* 2019;47:W556-60.
  28. Tang Z, Li C, Kang B, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* 2017;45:W98-W102.
  29. Györfy B, Surowiak P, Budczies J, et al. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One* 2013;8:e82241.
  30. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013;6:pl1.
  31. Ghandi M, Huang FW, Jané-Valbuena J, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* 2019;569:503-8.
  32. Yi L, Wu G, Guo L, et al. Comprehensive Analysis of the PD-L1 and Immune Infiltrates of m6A RNA Methylation Regulators in Head and Neck Squamous Cell Carcinoma. *Mol Ther Nucleic Acids* 2020;21:299-314.
  33. Lou W, Wang W, Chen J, et al. ncRNAs-mediated high expression of SEMA3F correlates with poor prognosis and tumor immune infiltration of hepatocellular carcinoma. *Mol Ther Nucleic Acids* 2021;24:845-55.
  34. Chandrashekar DS, Bashel B, Balasubramanya SAH, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* 2017;19:649-58.
  35. Xiong Y, Wei Y, Gu Y, et al. DiseaseMeth version 2.0: a major expansion and update of the human disease methylation database. *Nucleic Acids Res* 2017;45:D888-95.
  36. Yuan H, Yan M, Zhang G, et al. CancerSEA: a cancer single-cell state atlas. *Nucleic Acids Res* 2019;47:D900-D908.
  37. Lee K, Moon S, Park MJ, et al. Integrated Analysis of Tissue-Specific Promoter Methylation and Gene Expression Profile in Complex Diseases. *Int J Mol Sci* 2020;21:5056.
  38. Zhang T, Li Y, Zhang H, et al. The Role of RASSF1 Methylation in Lung Carcinoma. *Adv Exp Med Biol* 2020;1255:99-108.
  39. Majidpoor J, Mortezaee K. Steps in metastasis: an updated review. *Med Oncol* 2021;38:3.
  40. Liu Y, Guo W. SOX factors as cell-state regulators in the mammary gland and breast cancer. *Semin Cell Dev Biol* 2021;114:126-33.
  41. Ashrafizadeh M, Taeb S, Hushmandi K, et al. Cancer and SOX proteins: New insight into their role in ovarian cancer progression/inhibition. *Pharmacol Res* 2020;161:105159.
  42. Liu J, Qiu J, Zhang Z, et al. SOX4 maintains the stemness of cancer cells via transcriptionally enhancing HDAC1 revealed by comparative proteomics study. *Cell Biosci* 2021;11:23.
  43. Qin CX, Yang XQ, Zhan ZY. Connection between SOX7 Expression and Breast Cancer Prognosis. *Med Sci Monit* 2020;26:e921510.
  44. Zu G, Gao J, Zhou T. The Clinicopathological and Prognostic Significance of SOX9 Expression in Gastric Cancer: Meta-Analysis and TCGA Analysis. *Front Oncol* 2021;11:668946.
  45. Xu J, Zhang J, Li L, et al. SOX12 expression is associated with progression and poor prognosis in human breast cancer. *Am J Transl Res* 2020;12:8162-74.
  46. Feng M, Fang F, Fang T, et al. SOX13 promotes hepatocellular carcinoma metastasis by transcriptionally activating Twist1. *Lab Invest* 2020;100:1400-10.
  47. Xiong X, Xu W, Gong J, et al. miR-937-5p targets SOX17 to modulate breast cancer cell cycle and cell proliferation through the Wnt signaling pathway. *Cell Signal* 2021;77:109818.
  48. Ornat M, Kobierzycki C, Grzegorzolka J, et al. SOX18 Expression in Non-melanoma Skin Cancer. *Anticancer Res* 2016;36:2379-83.

49. Jethon A, Pula B, Olbromski M, et al. Prognostic significance of SOX18 expression in non-small cell lung cancer. *Int J Oncol* 2015;46:123-32.
50. Sun QY, Ding LW, Johnson K, et al. SOX7 regulates MAPK/ERK-BIM mediated apoptosis in cancer cells. *Oncogene* 2019;38:6196-210.
51. Zhang W, Glöckner SC, Guo M, et al. Epigenetic inactivation of the canonical Wnt antagonist SRY-box containing gene 17 in colorectal cancer. *Cancer Res* 2008;68:2764-72.
52. Hu C, Bai X, Liu C, et al. Long noncoding RNA XIST participates hypoxia-induced angiogenesis in human brain microvascular endothelial cells through regulating miR-485/SOX7 axis. *Am J Transl Res* 2019;11:6487-97.
53. Lee SH, Lee S, Yang H, et al. Notch pathway targets proangiogenic regulator SOX17 to restrict angiogenesis. *Circ Res* 2014;115:215-26.
54. Young N, Hahn CN, Poh A, et al. Effect of disrupted SOX18 transcription factor function on tumor growth, vascularization, and endothelial development. *J Natl Cancer Inst* 2006;98:1060-7.
55. Majidpoor J, Mortezaee K. Angiogenesis as a hallmark of solid tumors - clinical perspectives. *Cell Oncol (Dordr)* 2021;44:715-37.
56. Hayano T, Garg M, Yin D, et al. SOX7 is down-regulated in lung cancer. *J Exp Clin Cancer Res* 2013;32:17.
57. Kuo IY, Wu CC, Chang JM, et al. Low SOX17 expression is a prognostic factor and drives transcriptional dysregulation and esophageal cancer progression. *Int J Cancer* 2014;135:563-73.
58. Shao X, Lv N, Liao J, et al. Copy number variation is highly correlated with differential gene expression: a pan-cancer study. *BMC Med Genet* 2019;20:175.
59. Zhao J, Cheng M, Gai J, et al. SPOCK2 Serves as a Potential Prognostic Marker and Correlates With Immune Infiltration in Lung Adenocarcinoma. *Front Genet* 2020;11:588499.
60. Wang K, Chen R, Feng Z, et al. Identification of differentially expressed genes in non-small cell lung cancer. *Aging (Albany NY)* 2019;11:11170-85.
61. Ha EH, Choi JP, Kwon HS, et al. Endothelial SOX17 promotes allergic airway inflammation. *J Allergy Clin Immunol* 2019;144:561-573.e6.
62. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. *Oncogene* 2017;36:1461-73.
63. Fan R, He H, Yao W, et al. SOX7 Suppresses Wnt Signaling by Disrupting  $\beta$ -Catenin/BCL9 Interaction. *DNA Cell Biol* 2018;37:126-32.
64. Zhou W, Wang K, Wang J, et al. SOX17 Inhibits Tumor Metastasis Via Wnt Signaling In Endometrial Cancer. *Onco Targets Ther* 2019;12:8275-86.
65. Pai SG, Carneiro BA, Mota JM, et al. Wnt/beta-catenin pathway: modulating anticancer immune response. *J Hematol Oncol* 2017;10:101.
66. Chen DS, Mellman I. Oncology meets immunology: the cancer-immunity cycle. *Immunity* 2013;39:1-10.
67. Gattinoni L, Zhong XS, Palmer DC, et al. Wnt signaling arrests effector T cell differentiation and generates CD8+ memory stem cells. *Nat Med* 2009;15:808-13.
68. Swafford D, Manicassamy S. Wnt signaling in dendritic cells: its role in regulation of immunity and tolerance. *Discov Med* 2015;19:303-10.
69. Li Y, Zhao L, Li XF. Hypoxia and the Tumor Microenvironment. *Technol Cancer Res Treat* 2021;20:15330338211036304.

(English Language Editor: J. Jones)

**Cite this article as:** Wu D, Jiang C, Zheng JJ, Luo DS, Ma J, Que HF, Li C, Ma C, Wang HY, Wang W, Xu HT. Bioinformatics analysis of SOXF family genes reveals potential regulatory mechanism and diagnostic value in cancers. *Ann Transl Med* 2022;10(12):701. doi: 10.21037/atm-22-2749