



SMAD4 mutation correlates with poor prognosis in non-small cell lung cancer

Yue Wang^{1,2} · Qianqian Xue^{1,2} · Qiang Zheng^{1,2} · Yan Jin^{1,2} · Xuxia Shen^{1,2} · Mu Yang³ · Xiaoyan Zhou^{1,2} · Yuan Li^{1,2}

Received: 31 July 2020 / Revised: 7 November 2020 / Accepted: 11 November 2020 / Published online: 10 December 2020
© The Author(s), under exclusive licence to United States and Canadian Academy of Pathology 2020

Abstract

SMAD4 is an intracellular signaling mediator of the TGF- β pathway. Its mutation was commonly observed in gastrointestinal cancers, such as pancreatic cancer. The loss of *SMAD4* on immunohistochemical staining is often used to suggest a pancreaticobiliary differentiation in evaluating a metastatic adenocarcinoma with unknown origin. However, the function and molecular mechanism of *SMAD4* in non-small cell lung cancer (NSCLC) development are largely unknown. Thus, we studied the correlation between *SMAD4* mutations and clinico-molecular features in the patients with NSCLC. We reported the frequencies and prognostic values of *SMAD4* mutations in a Chinese NSCLC cohort using next-generation sequencing. The NSCLC cases from several public databases, including The Cancer Genome Atlas and others, were also used in this study to elucidate *SMAD4*-related molecular partners and mechanisms. Integrated bioinformatics analyses were conducted, such as analysis of Gene Ontology enrichment analysis, gene set enrichment analysis (GSEA), and survival analysis. Immunohistochemistry showed that the tissues harboring *SMAD4* mutations tended to show *SMAD4* deficiency or loss, while *SMAD4* expression was significantly reduced at all stages of NSCLC cases. We found that reduced *SMAD4* expression was more frequent in the patients with poor disease-free survival and resistance to platinum-based chemotherapy. *SMAD4* mutation was an independent risk factor for the survival of NSCLC patients. The expression of *SMAD4* was associated with that of *SMAD2*. The GSEA showed that *SMAD4* might promote NSCLC progression by regulating proliferation, adhesion, and immune response. In conclusion, these data suggest that *SMAD4* mutation or loss as well as reduced expression can be used to identify the NSCLC patients with poor survival and resistance to platinum-based chemotherapy. *SMAD4* may be a predictive marker or therapeutic target in NSCLC. The source code and user's guide are freely available at Github: <https://github.com/wangyue77-ab/sm4d4>.

These authors contributed equally: Yue Wang, Qianqian Xue

Supplementary information The online version of this article (<https://doi.org/10.1038/s41374-020-00517-x>) contains supplementary material, which is available to authorized users.

- ✉ Yuan Li
whliyuan@hotmail.com
- ✉ Mu Yang
yangmu1021@hotmail.com

- ¹ Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai, China
- ² Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China
- ³ Department of Pathology, Shanghai First People's Hospital, Shanghai Jiaotong University School of Medicine, Shanghai 200032, China

Introduction

Lung cancer is the most commonly diagnosed cancer and the leading cause of cancer-related deaths worldwide [1]. Non-small cell lung cancer (NSCLC) is comprised mainly of lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), which accounts for 80–85% of all lung cancer patients [2, 3]. In the past decade, remarkable progress has been made in the treatment of NSCLC in patients whose tumors harbor targetable somatic mutations, and *EGFR*, *ALK*, *PD-L1*, or *CTLA-4* inhibitors have been proven to be more effective than conventional chemotherapy [4–7]. Despite the advances in screening and treatment of NSCLC, only a small portion of NSCLC patients benefit from these approaches, and the 5-year survival rate is <20% [2]. The potential markers for prediction of outcomes and responses to specific therapy in NSCLC are limited.

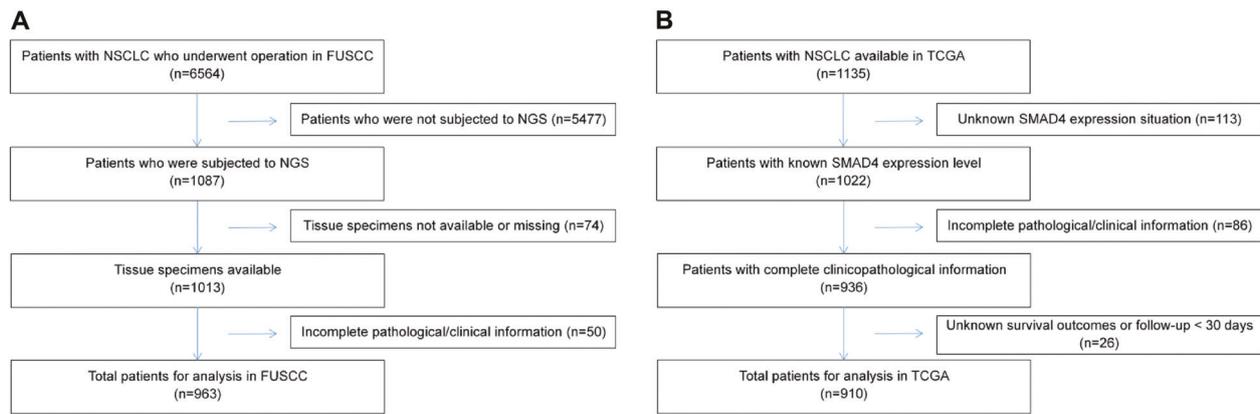


Fig. 1 Flow chart of cases selection. Cases from FUSCC (a) and TCGA (b).

The TGF- β /Smad signaling pathway is one of the most commonly altered cellular pathways in human cancers and plays a dual role in tumorigenesis [8]. TGF- β serves as a tumor suppressor by inhibiting proliferation and accelerating apoptosis during initiation and early progression of the tumor, but as the disease progresses, it promotes tumor formation by facilitating migration, invasion, angiogenesis, and evasion of the immune system [9, 10]. The loss of sensitivity to TGF- β is frequently observed in human cancers, and inactivation of Smad family members is a prominent mechanism in disruption of the TGF- β pathway [11].

SMAD4 (also called deleted in pancreatic carcinoma 4) was initially identified as a tumor suppressor gene at a homozygous deleted region on human chromosome 18q21.1 in pancreatic ductal adenocarcinoma by Hahn et al. [12] and encodes the only one co-Smad. It has been confirmed that the loss of *SMAD4* predicts a worse prognosis and the development of metastatic disease in patients with pancreatic cancer [13, 14]. In combination with other biomarkers, the loss of *SMAD4* on immunohistochemical staining is often used to suggest a pancreaticobiliary primary site in the evaluation of metastatic adenocarcinoma with a unknown primary site [15]. Recently, several studies have shown that mutations and deletions of *SMAD4* are detected in colon, ovarian, appendiceal, esophageal, and lung cancers [16–22], in addition to pancreatic cancer. Although mutation and loss of *SMAD4* have been described in NSCLC, where they are associated with lymph node metastases, increased angiogenesis and more aggressive cellular behavior in vitro [22, 23], detailed documentation of the prognostic role of *SMAD4* and its association with other molecular parameters in a large patient cohort are still lacking. In this study, we detected 24 *SMAD4*-mutated cases in a 963-case Chinese NSCLC cohort using next-generation sequencing (NGS). The correlation between *SMAD4* mutation and the clinico-molecular features of the patients was further evaluated. We found that the mutation of *SMAD4*, as well as loss or reduction of its expression,

was linked to the progression of NSCLC and patient survival.

Materials and methods

Patient characteristics

A total of 6564 patients diagnosed with NSCLC who underwent surgical resection were identified at Fudan University Shanghai Cancer Center (FUSCC) between June 2017 and July 2019. Of them, NGS was performed on 1087 patients. The patients with tumor samples not available or missing ($n = 74$) and the patients with incomplete follow-up information ($n = 50$) were excluded. Finally, 963 patients from FUSCC who provided fully informed consents were used for the analysis (Fig. 1A). For all patients, the following data were extracted from the electric medical record: gender, age, pathological grade, and TNM stage. Tumor stage was determined based on the American Joint Committee on Cancer tumor staging system (eighth edition, 2017). The median follow-up of the patients was 23 months (IQR, 10–30 months), and the last date of follow-up was June 1, 2020. In addition, the expression of 1135 NSCLC based on RNA-seq data from The Cancer Genome Atlas (TCGA) were available, while those cases with unknown *SMAD4* expression ($n = 113$), incomplete pathological or clinical information ($n = 86$), and unknown survival outcomes or <30 days of follow-up time ($n = 26$) were removed, and 910 cases from TCGA were enrolled in this study for analysis (Fig. 1B).

Next-generation sequencing

Genomic DNA was extracted from the FFPE tissue samples using the QIAamp DNA Mini Kit, and targeted deep sequencing of mutational hotspots was conducted using a capture-based targeted sequencing panel, which included all

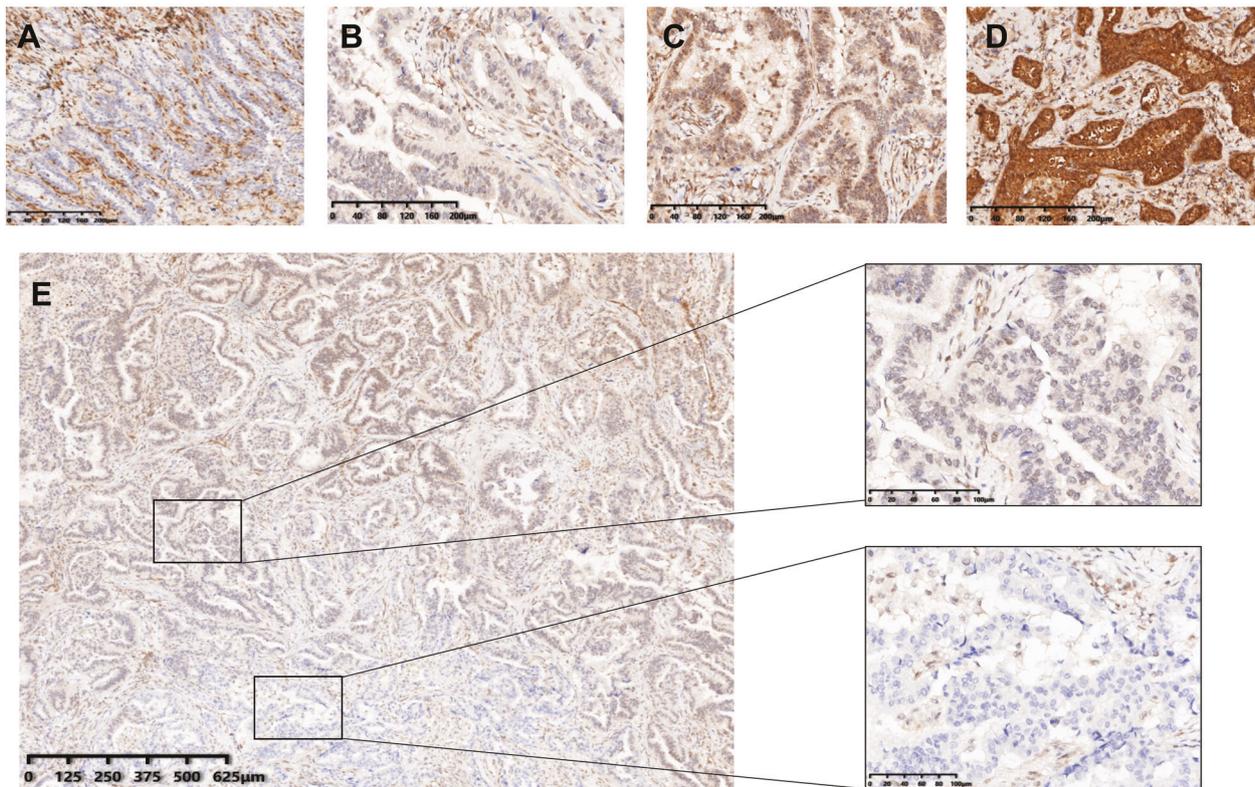


Fig. 2 Immunohistochemical analysis of Smad4 in NSCLC specimens. A different score (0–3) was given to each sample based on the intensity of staining signal in both nucleus and cytoplasm. A final *H*-score (0–300) was obtained by multiplying the intensity and reactivity

extension. Graphics indicate different staining: “0” in (a), “1+” in (b), “2+” in (c), and “3+” in (d). **e** The sample image comprised mainly of negative or low Smad4 expression, and *H*-score was about 85 that was classified as low Smad4 group.

exons of 68 genes and selected introns in *ALK*, *RET*, and *ROS1* for the detection of translocation events. The library DNAs were prepared by amplifying the targeted regions using multiplex polymerase chain reaction, followed by adapter DNA ligation. Multiplexed sequencing was performed using the Illumina HiSeq 2500 platform. Mutations and variants, including indels, substitutions, rearrangements, and amplifications, were identified using Illumina suite software.

Immunohistochemistry

For histopathological analysis, tissue samples were fixed for 48 h in 4% paraformaldehyde, decalcified with 20% EDTA, and embedded in paraffin. Next, 4- μ m sections were cut and incubated in sodium citrate (pH 6.0) at 100 °C for 15 min. After endogenous peroxidase activity elimination by 3% H₂O₂, the sections were blocked with 1% PBS. Primary antibody specific for SMAD4 (1:100; ab40759, Abcam) was applied at 4 °C overnight. After the samples were washed with PBS, horseradish peroxidase-conjugated anti-rabbit/mouse secondary antibodies (Gene Technology, Shanghai, China) were applied and incubated for 1 h at room temperature. Normal epithelium adjacent to the tumor served as the internal positive control. The immunohistochemical

staining was scored in a semiquantitative approach incorporating both the intensity and distribution of specific staining as follows. The intensity of specific staining was characterized as not present (0), weak (1+), distinct (2+), and very strong (3+) based on the average expression of both nucleus and cytoplasm. The percentage of positive tumor cells per slide (0 to 100%) was multiplied by the dominant intensity pattern of staining (Fig. 2). The *H*-score (0–300) was calculated according to the following formula:

$$[1 \times (\% \text{ cells } 1+) + 2 \times (\% \text{ cells } 2+) + 3 \times (\% \text{ cells } 3+)].$$

An *H*-score of ≥ 75 was the threshold value for a “positive” immunohistochemical assay [24]. Smad4 loss was considered if *H*-score < 75 in order to minimize confounding effects introduced by varying degrees of Smad4 down-regulation. The median score of positive expression was utilized to stratify patients into low Smad4 (75–187) and high Smad4 (188–300) groups.

Gene Ontology (GO) enrichment analysis

The original RNA-seq read counts of NSCLC cases were acquired via the TCGA website (<https://portal.gdc.cancer.gov>)

using the *gdc-client* tool. Normalization of expression matrix data was executed by the R command *normalizeBetweenArrays*. The cases were divided into two groups depending upon the median mRNA expression level of *SMAD4*. The Genomic Analysis of Differentially Expressed Genes (DEGs) analysis was performed by the *limma* R package. The two cutoff values for the DEGs were an expression fold change >2 and a *p* value <0.05. The Database for Annotation, Visualization, and Integrated Discovery (<http://www.david.niaid.nih.gov>) integrates GO, KEGG, UniProt, and DrugBank, among other authoritative data sources, to provide detailed annotation of genes from cell components, molecular functions, and biological processes. Point plots were generated by R software.

Gene set enrichment analysis (GSEA)

GSEA is a computing method for exploring the statistical significance and concordant differences of defined gene sets or pathways between two biological states. In the present study, GSEA was used to deeply analyze biological information, enlightening our understanding of relevant biological events. Files of genomic expression data and contrast information were inputted, and the analysis was carried out using R software.

Statistical analysis

Frequencies were compared using Fisher's exact and Pearson's χ^2 tests, as appropriate. Overall survival (OS) was measured from the date of surgical resection to the date of death or last follow-up. Disease-free survival (DFS) was defined from the date of surgery to cancer recurrence, death, or last follow-up whichever was the earliest. Progression-free survival (PFS) was defined from the date of surgery to progression as determined by the treating physician based on radiologic, biochemical, and/or clinical criteria. Survival curves were generated using Kaplan–Meier methods and compared using the log-rank test. Univariate and multivariate analyses were performed to identify prognostic factors. Analyses were performed using R software (version 3.6.3), GraphPad Prism (version 7.0), and IBM SPSS software (version 24.0). *p* values were considered significant if <0.05 in all analyses.

Results

Patient characteristics and clinicopathological variables

Of the 963 NSCLC cases from FUSCC included in the study, 24 (2.5%) had *SMAD4* mutations, including 23

adenocarcinomas and 1 adenosquamous carcinoma. The patients were divided into two groups: mutated type (*n* = 24) and wild-type (*n* = 939) *SMAD4* according to the NGS results. The demographic and baseline clinicopathologic data, including gender, diagnostic age, pathological grade, histological type, TNM stage, and driver mutation status, are shown in Table 1. Most patients in this study had an adenocarcinoma (91.5%), and the vast majority of the patients were at an early stage (85.8%). The patients with *SMAD4* gene mutations were more likely to be at T stage III or IV (*p* = 0.015, Table 1) and show poorer tumor differentiation (*p* = 0.044, Table 1) than the patients without mutations. In contrast, *SMAD4* mutation status was not significantly associated with gender, age, N stage, M stage, histology type, or mutations in dominant cancer driver genes, such as *EGFR*, *ALK*, and *KRAS*.

Genetic variations of *SMAD4* in NSCLC

The *SMAD4* alterations in pan cancers were analyzed using cBioPortal. The genetic variations of *SMAD4* showed incidence rate of 5% in NSCLC, including mutation and copy number deletion mainly (Fig. 3A). One major limitation of NGS approaches was the ability to detect long copy number variations, which was probably one reason of different alterations between the FUSCC cohort and the public cases. To further assess the genetic variations of *SMAD4*, we examined *SMAD4* alterations of 819 NSCLC cases from five studies (University of Turin [25]; MSKCC [26]; TRACERx [27, 28]; MSK [29]; MSK [30]). The results showed that the alteration frequency of *SMAD4* in NSCLC was 4% (Fig. 3B), and mutation was the most common type, comprising more than half of all genetic variations (Fig. 3C). Therefore, mutation might be one of the main mechanisms by which *SMAD4* is dysfunctional in NSCLC. Moreover, the *SMAD4* mutation sites on the peptide sequence in lollipop plots were portrayed. Four types of mutations were confirmed (Fig. 3E): missense mutations (66.7%), deletions (20.8%), nonsense mutations (8.3%), and frameshift mutations (4.2%). We observed the mutational hotspots were D351 in cases with NSCLC from cBioPortal (Fig. 3D) and R361 in the Chinese cohort from FUSCC (Fig. 3E). R361 is located on the *SMAD4* homotrimer interaction interface, and this hotspot normally stabilizes homo- or heterotrimer oligomerization. Mutation at R361 in *SMAD4* correlates with metastasis and decreased survival in colon cancer [31]. Those mutations could have widespread effects, because *SMAD4* is a binding partner for all Smad dependent transcriptional regulation. The *SMAD* family peptide chains share conserved sequences in amino and carboxyl-terminal regions with certain structural similarities, known as Mad homology (MH) domains 1 and 2, respectively [13]. In spite of different hotspots, the missense

Table 1 Clinicopathologic characteristics in patients with mutated and wild-type SMAD4 from FUSCC.

Characters	Total	SMAD4 Mutated type	SMAD4 Wild type	<i>p</i>
<i>n</i>	963	24 (2.5%)	939 (97.5%)	
Gender				0.299
Male	413 (42.9%)	13 (54.2%)	400 (42.6%)	
Female	550 (57.1%)	11 (45.8%)	539 (57.4%)	
Age, year				0.421
<60	439 (45.6%)	9 (37.5%)	430 (45.8%)	
≥60	524 (54.4%)	15 (62.5%)	509 (54.2%)	
Pathological grade				0.044
Well/moderate	821 (85.3%)	17 (70.8%)	804 (85.6%)	
Poor	142 (14.7%)	7 (29.2%)	135 (14.4%)	
Histology subtype				0.280
AD	881 (91.5%)	23 (95.8%)	858 (91.4%)	
SCC	60 (6.2%)	0	60 (6.4%)	
ASC	11 (1.1%)	1 (4.2%)	10 (1.1%)	
Others	11 (1.1%)	0	11 (1.2%)	
T stage				0.015
1, 2	923 (95.8%)	20 (83.3%)	903 (96.2%)	
3, 4	40 (4.2%)	4 (16.7%)	36 (3.8%)	
N stage				0.792
0	781 (81.1%)	19 (79.2%)	762 (81.2%)	
1, 2	182 (18.9%)	5 (20.8%)	177 (18.8%)	
M stage				1.000
0	960 (99.7%)	24 (100%)	936 (99.7%)	
1	3 (0.3%)	0	3 (0.3%)	
TNM stage				0.960
1, 2	826 (85.8%)	20 (83.3%)	806 (85.8%)	
3, 4	137 (14.2%)	4 (16.7%)	133 (14.2%)	
EGFR mutation				0.512
Mutated	643 (66.8%)	18 (75.0%)	625 (66.6%)	
Wild type	320 (33.2%)	6 (25.0%)	314 (33.4%)	
ALK mutation				0.189
Mutated	63 (6.5%)	0	63 (6.7%)	
Wild type	900 (93.5%)	24 (100%)	876 (93.3%)	
KRAS mutation				0.479
Mutated	82 (8.5%)	3 (12.5%)	79 (8.4%)	
Wild type	881 (91.5%)	21 (87.5%)	860 (91.6%)	

AD adenocarcinoma, SCC squamous cell carcinoma, ASC adenosquamous carcinoma.

mutations were most common, and mutations were mostly located in the MH2 domain in two data sets.

Patterns of SMAD4 expression in NSCLC

To determine the correlation between the mutation and expression of SMAD4 in NSCLC, we randomly selected 45 wild-type SMAD4 NSCLC cases with sufficient tissues to investigate the SMAD4 expression and distribution by tissue microarray. Immunohistochemistry was performed on 45

wild-type SMAD4 and 24 mutated SMAD4 samples. The prevalence of SMAD4 expression in the patients with SMAD4 mutations was significantly different from that in the matched group, and the results are shown in Table 2. We found that wild-type SMAD4 samples tended to show high IHC expression, and the loss of SMAD4 was almost not seen. Mutated cases showed loss of SMAD4 expression levels with loss areas ranging from 10 to 100%, thus most cases showed low SMAD4 expression. This result demonstrated that the patients harboring SMAD4 mutations tended

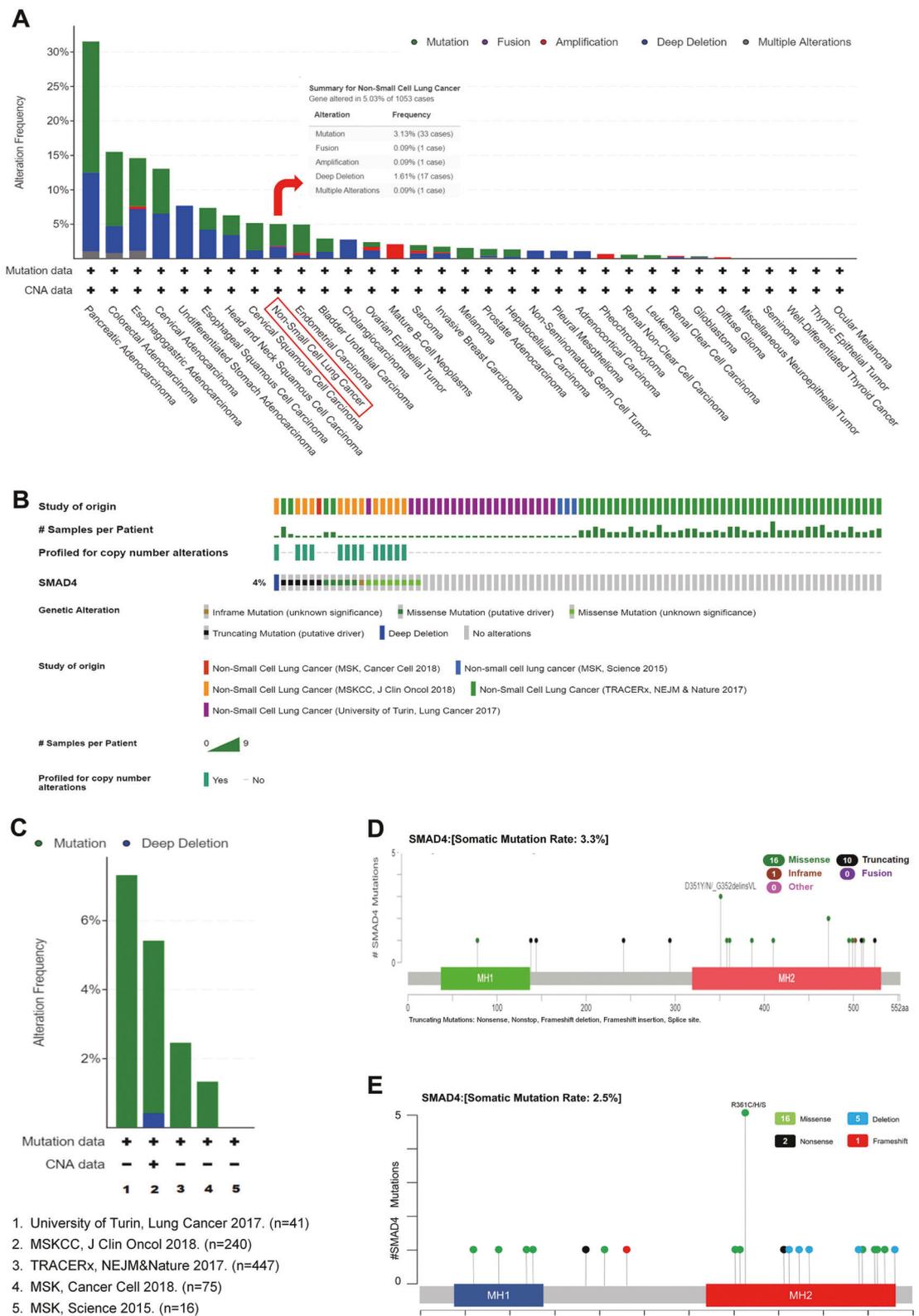


Fig. 3 Analyses of genetic variations in SMAD4 for NSCLC. **a** Alteration frequencies of SMAD4 across cancer studies. **b** OncoPrint overview of the genetic variations of SMAD4. **c** Alteration frequencies

of SMAD4 in different studies. **d** Distribution of SMAD4 mutations from cBioPortal. **e** Distribution of SMAD4 mutations from FUSCC.

Table 2 Comparison between SMAD4 gene and SMAD4 IHC statuses.

	SMAD4 IHC status			Total	<i>p</i>
	Loss	Low	High		
SMAD4 gene status					
Wild type, <i>n</i> (%)	0	10 (22)	35 (78)	45	<0.001
Mutated, <i>n</i> (%)	2 (8)	19 (79)	3 (13)	24	
Total	2	29	38	69	

to show deficient *SMAD4* expression. Further studies confirmed that the staining intensity was weakly affected by synonymous mutations; tissues with missense mutations, deletions, and especially frameshift mutations were found to have apparent *SMAD4* deficiency. These data indicate that individual discrepancies and heterogeneous expression exist. Given the presence of different types of mutations and possibly contaminating *SMAD4*-positive mesenchymal cells, the specific governing mechanisms remain to be elucidated.

To further explore the expression of *SMAD4* in NSCLC, we performed data mining in NSCLC cohort in TCGA database using cBioPortal. We found that the mRNA expression levels of wild-type *SMAD4* were much higher than mutated (Fig. 4A), while the correlations between mRNA and protein expression levels were feeble (Fig. 4B). These correlations seem to provide some evidence for *SMAD4* deficiency with significant heterogeneity in IHC. Moreover, the top 50 genes positively correlated with *SMAD4* expression in LUAD and LUSC were identified (Fig. 4C) using LinkedOmics. The heatmap showed that *SMAD2*, another member of the *SMAD* family, and *SMAD4* had high correlation coefficients, which was verified in cBioPortal and ENCORI (Fig. 4D, E). Phosphorylated *SMAD2* and *SMAD3* form a heteromeric complex with *SMAD4* and are transported to the nucleus, where it binds with other DNA-binding transcription factors and consequently regulates the transcription of TGF- β target genes [8, 13]. Abundant evidence suggests that depletion of *SMAD2* results in enhanced cell invasion, metastasis [10, 14], and angiogenesis [32–34] in skin squamous carcinoma, NSCLC, and breast cancer. There seems to be a synergy between the *SMAD2* and *SMAD4* in inhibiting tumors.

SMAD4 is significantly reduced in all stages of NSCLC

To assess the relationship between *SMAD4* expression and lung cancer, we first analyzed the mRNA expression profile of *SMAD4* in 910 NSCLC cases from TCGA. We found that the *SMAD4* mRNA levels were significantly lower in

the NSCLC patients than in the healthy individuals ($p < 0.001$; Fig. 5A) and lower in LUAD than in LUSC ($p < 0.001$; Fig. 5B). No significant differences were observed in age ($p = 0.330$; Fig. 5C) or gender ($p = 0.211$; Fig. 5D). Moreover, *SMAD4* expression was reduced in all stages of NSCLC compared with that in normal tissues (Fig. 5E). We found that *SMAD4* was expressed at lower levels in the LUAD and LUSC tissues than in the adjacent normal tissues using GEPIA (Fig. 5F). These results suggest that there may remain a link between *SMAD4* expression and the occurrence and development of NSCLC.

Mutation or reduced expression of SMAD4 predicts poor survival and chemotherapy resistance

The low expression of *SMAD4* in NSCLC compared with adjacent normal tissues indicates its role as a tumor suppressor. To assess the relationship between *SMAD4* expression and clinical outcomes of NSCLC patients, we investigated OS of LUAD and LUSC using the online tool ENCORI. However, no significant association of *SMAD4* expression with OS was found (Fig. 6A). To further research its prognostic value, we stratified 910 NSCLC patients from TCGA into two groups using optimal cutoff values determined by the “surv_cutpoint” function of the “survminer” R package: high expression (high, $n = 555$) and low expression (low, $n = 355$) of *SMAD4*. The results showed no significant association of *SMAD4* expression with OS (HR = 0.95, $p = 0.24$; Fig. 6B), but a significantly longer DFS (HR = 0.84, $p = 0.027$; Fig. 6B) was detected when *SMAD4* expression was high. Among the 332 patients received drug treatments, the overwhelming majority of which were platinum compounds ($n = 322$, 97.0%), such as cisplatin and carboplatin. As a consequence, we split the patients into two groups according to the preceding method to explore the relationship between *SMAD4* expression and the effects of platinum-based chemotherapy, including high expression (high, $n = 200$) and low expression (low, $n = 122$) of *SMAD4*. We found that low *SMAD4* expression was associated with poor OS (HR = 0.83, $p = 0.038$; Fig. 6C) as well as DFS (HR = 0.85, $p = 0.048$; Fig. 6C). OS of reduced *SMAD4* in NSCLC identified patients with poor DFS and resistance to platinum-based chemotherapy. The PFS curves of altered ($n = 28$) and wild-type ($n = 791$) *SMAD4* cases were analyzed using cBioPortal. No significant association was found ($p = 0.410$; Fig. 6D). To further analyze the impact of *SMAD4* mutations on survival of NSCLC patients, we downloaded the information of five studies as mentioned above, including follow-up and gene mutations. We found that *SMAD4*-mutated cases ($n = 14$) were more likely to progress than wild-type *SMAD4* cases ($n = 147$) in patients with driver (*EGFR*, *ALK*, *KRAS*, and *MET*) mutations ($p = 0.042$; Fig. 6E). We also

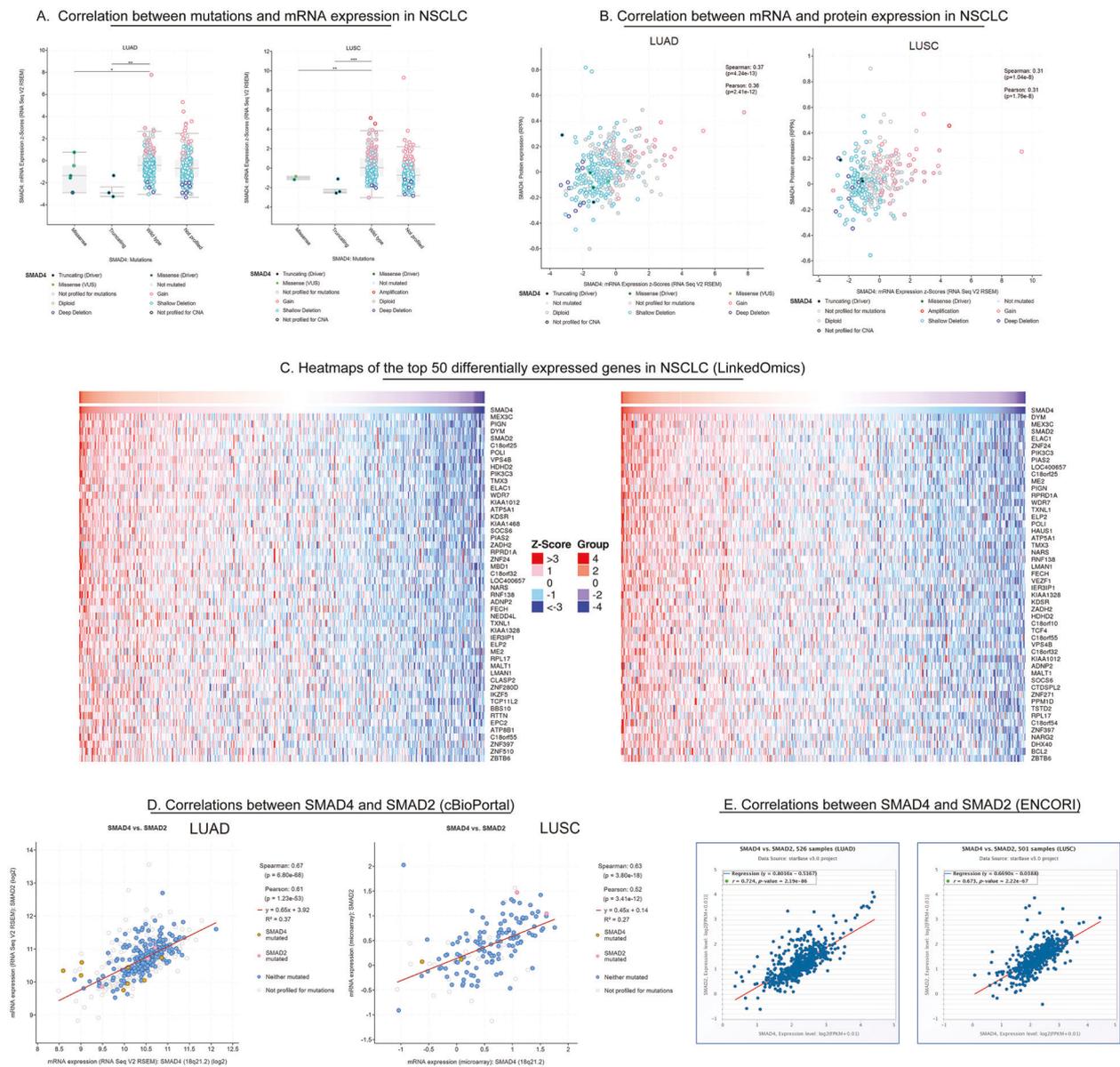


Fig. 4 Patterns of SMAD4 expression in NSCLC. Correlations between mutations and mRNA expression levels (**a**), mRNA and protein expression levels (**b**). **c** The top 50 genes positively correlated

compared the survivals between LUAD and LUSC patients in Fig. S1.

In 24 *SMAD4*-mutated cases from FUSCC, 18 (75%) harboring *EGFR* mutations were selected (Table 1). To further investigate the effect of *SMAD4* and *EGFR* mutations on survival outcomes, we randomly matched 72 cases with *EGFR* mutations and *SMAD4* wild type according to TNM stage. Univariable and Cox proportional hazards models for OS and PFS of these 90 NSCLC patients were created (Table 3). *SMAD4* was an independent prognostic factor for OS in NSCLC ($p = 0.038$, Table 3). TNM stage ($p < 0.001$, Table 3), smoking status ($p = 0.025$, Table 3), and *SMAD4* status ($p = 0.002$, Table 3) were independent

to *SMAD4* expression. **d**, **e** Correlations between *SMAD4* and *SMAD2*. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

prognostic factors for PFS of NSCLC patients according to the univariate and multivariate regression models. These analyses demonstrated that *SMAD4* mutation was strongly associated with poor prognosis. Patients with coexisting mutations in the *EGFR* and *SMAD4* genes had a significantly worse OS and PFS than those with single *EGFR* mutations.

Gene Ontology analysis and GSEA

In order to explore the potential functions and molecular mechanisms of *SMAD4* in NSCLC progression, we conducted GO enrichment analysis and GSEA on NSCLC

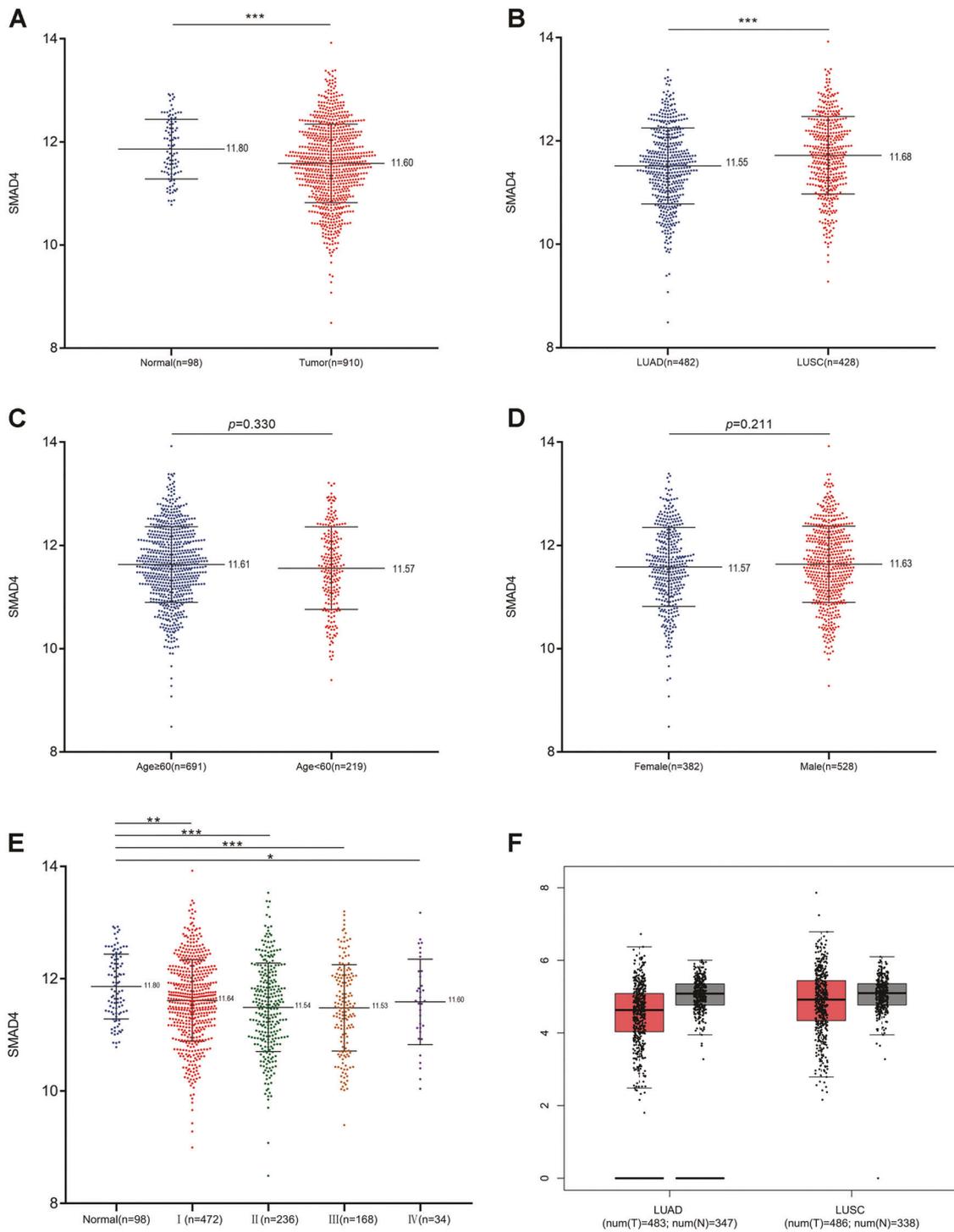


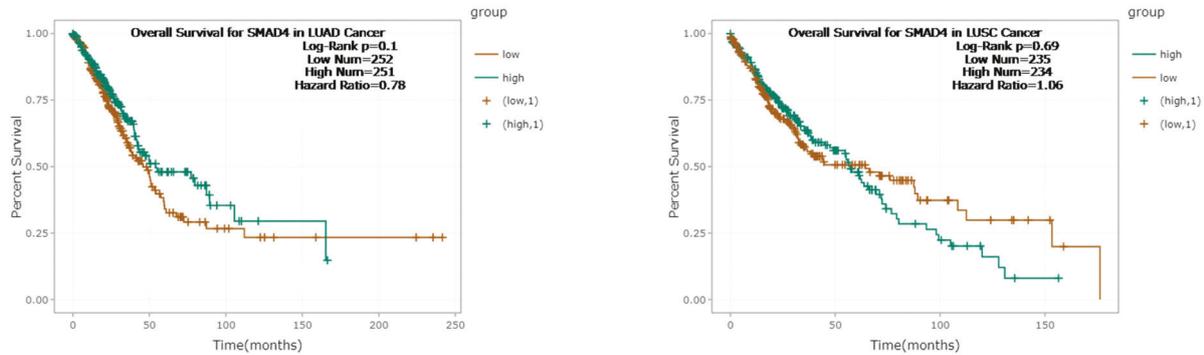
Fig. 5 SMAD4 was significantly downregulated in NSCLC cases from TCGA. The expression of SMAD4 at the mRNA level in 910 NSCLC from TCGA. Box plots show the association of SMAD4 expression with normal/tumor tissues (a), histological types (b), age (c),

gender (d), and TNM stage (e). f GEPIA was used to detect the expression of SMAD4 in TCGA NSCLC data set. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

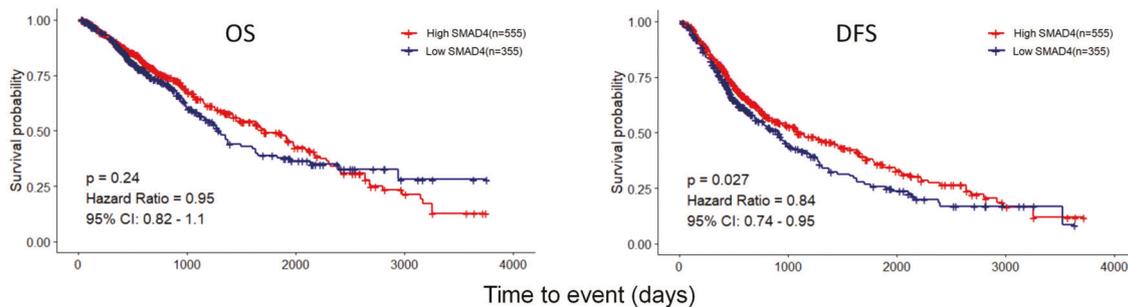
cases from TCGA database. The GO terms, in which the DEGs were significantly enriched, included integral components of nucleoplasm, cytoplasm, regulation of apoptotic process and gene expression, immune response, etc.

(Fig. 7A). GSEA indicated enrichments of gene sets including cell cycle, focal adhesion, adherens junction, and pyrimidine metabolism in the samples with low SMAD4 expression (Fig. 7B); KEGG jak stat signaling pathway,

A. OS for SMAD4 in NSCLC (ENCORI)



B. OS and DFS among NSCLC cases from TCGA



C. OS and DFS among chemotherapy-treated NSCLC cases from TCGA

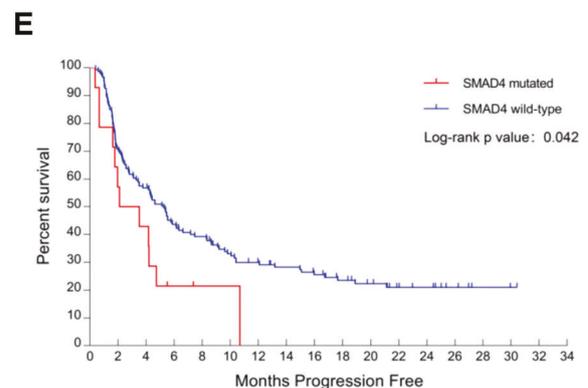
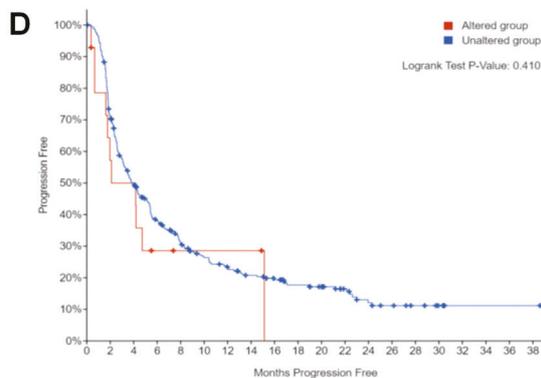
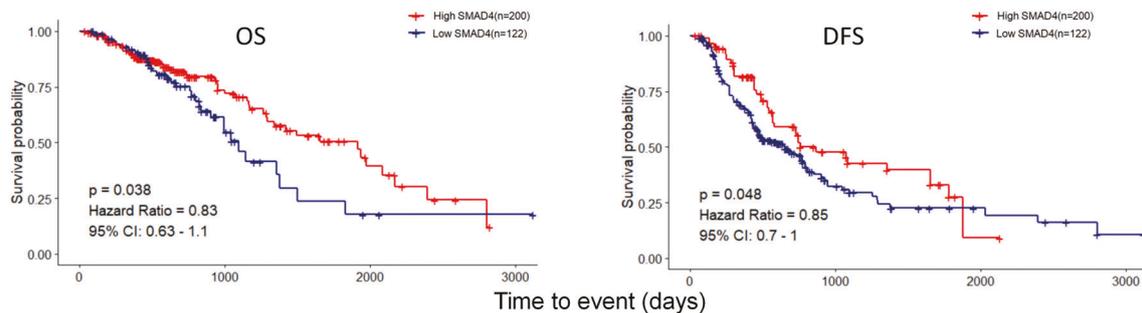


Fig. 6 Kaplan–Meier survival analysis according to SMAD4 mutation or expression. a The overall survival curves analyzed using ENCORI. **b** The overall and disease-free survival curves of NSCLC cases from TCGA. **c** The overall and disease-free survival curves of

chemotherapy-treated patients from TCGA. **d** The progression-free survival curves of altered and unaltered SMAD4 cases analyzed using cBioPortal. **e** The progression-free survival curves of mutated and wild-type SMAD4 cases both with driver mutation from cBioPortal.

Table 3 Univariable and multivariate analyses for OS and PFS in NSCLC patients from FUSCC.

Characters	OS			PFS		
	Univariable		Multivariable	Univariable		Multivariable
	HR (95% CI)	p	HR (95% CI)	HR (95% CI)	p	HR (95% CI)
Gender (male vs. female)	2.033 (0.211–9.576)	0.539	NI	1.291 (0.403–4.136)	0.667	NI
Age (≥60 vs. <60)	1.261 (0.177–8.968)	0.817	NI	0.819 (0.260–2.577)	0.732	NI
Smoking status (current/former vs. never)	0.608 (0.063–5.852)	0.667	NI	3.397 (1.084–10.649)	0.036	2.940 (1.149–7.526)
Stage (III, IV vs. I, II)	1.261 (0.177–8.968)	0.817	NI	8.174 (3.150–21.211)	<0.001	7.933 (3.103–20.284)
SMAD4 (mutation vs. wild)	1.453 (1.397–5.317)	0.042	1.924 (1.547–6.775)	4.467 (1.707–11.690)	0.002	4.593 (1.771–11.912)

Characteristics with a *p* value of <0.05 in the univariate analysis were entered into the multivariate model in a stepwise fashion. OS overall survival, PFS progression-free survival, HR hazard ratio, CI confidence interval, NI not included due to the *p* value > 0.05.

KEGG toll-like receptor signaling pathway, KEGG T-cell receptor signaling pathway, KEGG B-cell receptor signaling pathway in the samples with high SMAD4 expression (Fig. 7C).

Discussion

SMAD family members are a series of downstream effectors of the TGF-β signaling pathway that play a critical role in tumor progression. Smad2 and Smad3 proteins are phosphorylated by activated TGF-β receptors and then combine with the common Smad4 to form a complex, and the resulting trimer translocates into the nucleus, where it functions as a gene expression regulator [8]. Smad4 protein has been demonstrated as an important factor in IHC to suggest pancreaticobiliary primary site [15] and an independent prognostic factor for both DFS and OS in advanced colorectal cancer [35]. Varying SMAD4 mutations have been detected in different types of cancers by large-scale sequencing. Compared with 35% in pancreatic cancer and 12% in colon cancer cases [36–40], a significantly lower frequency of SMAD4 mutations have been found in other types of cancers. In previous cohort studies, point mutations of SMAD4 were identified in 0.21%, 2.24%, 2.46%, and 8.86% of kidney, lung, esophageal, and biliary tract cancers, respectively [40–42]. In this study, we found that the mutation of SMAD4 in the Chinese NSCLC cohort was 2.5%, which was consistent with the previous reported 2.24% in general, and the mutations were mostly located in the MH1 and MH2 domains. Although point mutations are rare, protein deletions are common in cancers due to the presence of various other influencing factors, such as alternative splicing. MH2 has been reported to play a key role in transcriptional activation and homo- and hetero-oligomer formation, while MH1 functions as a negative regulatory domain [13, 43]. These two domains are thought to always affect each other because MH1 disturbs the MH2 interaction with phosphorylated R-Smads, and the connection between MH1 and cDNAs is prevented by MH2. The linker region that maintains the interaction between MH1 and MH2 is often deleted during alternative splicing. Wan et al. [44] found several alternatively spliced variants that are associated with exon deletion in the linker region, showing significant effects on EMT and regulating E-cad and VIM protein expression.

In our study, the expression level of SMAD4 gradually decreased as the disease progressed, except it increased in stage IV. We speculated this was due to the small sample size because there were a very small number of samples of stage IV. Another possible reason is that SMAD4 is involved in the dual function of the TGF-β signaling pathway. TGF-β serves as a tumor suppressor by inhibiting

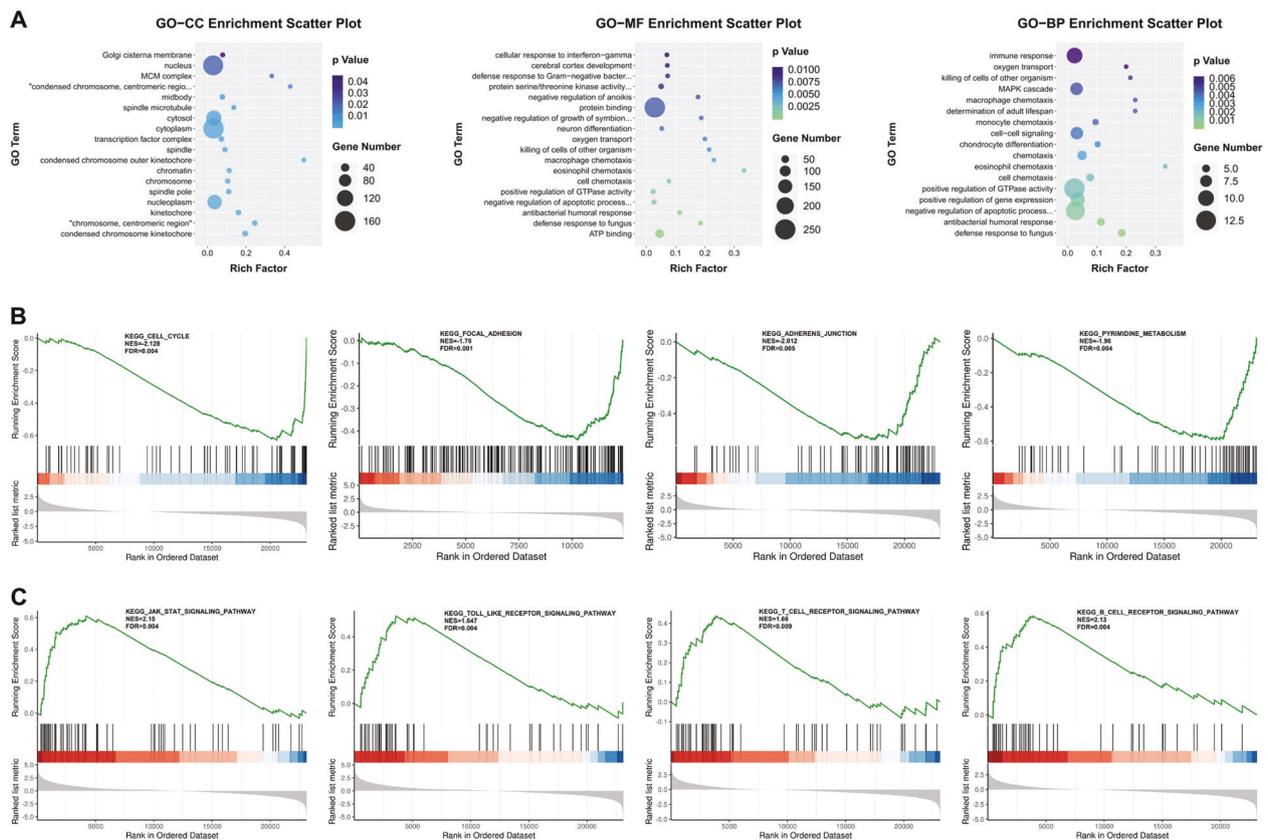


Fig. 7 GO and GSEA of *SMAD4*-related enrichment gene sets in NSCLC cases from TCGA. **a** GO analysis of DEGs based on *SMAD4* expression. **b** Enrichment gene sets in the samples with low

SMAD4 expression by GSEA. **c** Enrichment gene sets in the samples with high *SMAD4* expression by GSEA.

proliferation and accelerating apoptosis during initiation and early progression of the tumor, whereas it could promote tumor formation by facilitating migration, invasion, and angiogenesis in the late stage [9, 10], the specific molecular mechanism of which has not been fully elucidated. It is possible that inactivation of *SMAD4* is one of the involved mechanisms. In addition, *TGFBR2* serves as an initial regulator of the TGF- β signaling pathway, and its loss has been implicated as a mechanism for lung cancer formation, especially in LUSC formation. It has been reported that dysregulated *TGFBR2/ERK-Smad4/SOX2* signaling drives LUSC formation [45].

Although *SMAD4* mutation is not deemed as a tumor initiating event in most cases, loss or mutation of *SMAD4* is thought to enhance tumor progression, for example, after *KRAS* mutational activation in pancreatic adenocarcinoma [42]. Researchers found that low expression of *SMAD4* could promote A549 cell growth [46, 47], which indicated that the deletion of *SMAD4* was associated with apoptosis by affecting the Bcl-2/Bax balance in NSCLC, thereby contributing to carcinogenesis. Kawaguchi et al. [48] showed that double or triple gene mutations in *RAS*, *TP53*, and *SMAD4* are associated with a worse survival and earlier

recurrence than the cases with mutations in only one or none of these genes after resection of colorectal liver metastases. Several previous studies have shown that *SMAD4* depletion in a head and neck squamous cell carcinoma cell line induces cetuximab resistance and results in worse survival in an orthotopic mouse model in vivo [49]. Wasserman et al. [47] also suggested that *SMAD4* loss correlates with worse clinical outcome and resistance to chemotherapy. In our study, we first analyzed the prognostic role of *SMAD4* mutation and its association with other molecular parameters in a large Chinese NSCLC cohort. *SMAD4* mutation or loss, as well as reduced expression, was associated with poor DFS and platinum-based chemotherapy resistance. Coexisting mutations in *EGFR* and *SMAD4* genes conferred significantly worse OS and PFS than single *EGFR* mutations. However, Ziemke et al. [50] found that NSCLC with reduced *SMAD4* expression is more sensitive to treatment regimens containing DNA topoisomerase inhibitors than those without. To detect more potential mechanisms of *SMAD4* in NSCLC progression, we conducted GO analysis and GSEA in NSCLC cases from TCGA database and found that reduced *SMAD4* expression might promote NSCLC progression by

promoting proliferation and adhesion. We also found that the increased *SMAD4* expression (vs. reduced expression) was associated with immune response. Indeed, *SMAD4* participated in the expression of the anti-inflammatory cytokine, IL-10, in TH17 cells [51, 52]. A TGF β -SKI-Smad4 pathway that critically and specifically directs resident CD103⁺ CD8⁺ T-cell generation for protective immunity against primary and secondary viral infection was revealed recently [53]. The correlations and mechanisms of *SMAD4* and immune microenvironment warrant further investigation.

This study has several limitations. First, although our cohort of patients with NSCLC was large, the number of *SMAD4*-mutated cases was rather moderate. Second, the patients diagnosed with NSCLC undergoing surgical resection for the past 2 years were included in this study, and the follow-up time was limited. A long-term study is needed to confirm our findings. Third, our study was limited by the inherent issues of retrospective studies, including potential selection bias. This should be considered when generalizing our findings to other populations.

In conclusion, we found that *SMAD4* alteration was associated with poor survival and resistance to platinum-based chemotherapy, suggesting that *SMAD4* alteration might be a predictive marker or therapeutic target in these NSCLC. Further prospective studies are needed to evaluate these findings, and the clinical significance of *SMAD4* alteration in NSCLC.

Acknowledgements This study was supported by National Nature Science Foundation of China [Grant Numbers 81972171 and 81472173].

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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

References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68:394–424.
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020;70:7–30.
- Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016;66:115–32.
- Brigatinib outperforms crizotinib as first-line therapy. *Cancer Discov.* 2020;10:OF5.
- Suzuki S, Ishida T, Yoshikawa K, Ueda R. Current status of immunotherapy. *Jpn J Clin Oncol.* 2016;46:191–203.
- Borghaei H, Paz-ares L, Horn L, Zhou C, Wang J, Mok T, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med.* 2015;373:1627–39.
- Herbst RS, Baas P, Kim D-W, Felip E, Pérez-Gracia JL, Han JY, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet.* 2016;387:1540–50.
- Pickup M, Novitskiy S, Moses HL. The roles of TGF β in the tumour microenvironment. *Nat Rev Cancer.* 2013;13:788–99.
- Nkolice A, Kojic S, Knezevic S, Krivokapic Z, Ristanovic M, Radojkovic D. Structural and functional analysis of *SMAD4* gene promoter in malignant pancreatic and colorectal tissues: detection of two novel polymorphic nucleotide repeats. *Cancer Epidemiol.* 2011;35:265–71.
- Fleming NI, Jorissen RN, Mouradov D, Christie M, Sakthianandeswaren A, Palmieri M, et al. *SMAD2*, *SMAD3* and *SMAD4* mutations in colorectal cancer. *Cancer Res.* 2013;73:725–35.
- Morén A, Itoh S, Moustakas A, Dijke P, Heldin CH. Functional consequences of tumorigenic missense mutations in the amino-terminal domain of Smad4. *Oncogene.* 2000;19:4396–404.
- Hahn SA, Schutte M, Hoque AT, Moskaluk CA, da Costa LT, Rozenblum E, et al. *DPC4*, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science.* 1996;271:350–3.
- Drabsch Y, ten Dijke P. TGF- β signalling and its role in cancer progression and metastasis. *Cancer Metastasis Rev.* 2012;31:553–68.
- Hoot KE, Lighthall J, Han G, Lu SL, Li A, Ju W, et al. Keratinocyte-specific Smad2 ablation results in increased epithelial-mesenchymal transition during skin cancer formation and progression. *J Clin Invest.* 2008;118:2722–32.
- Wilentz RE, Su GH, Dai JL, Sparks AB, Argani P, Sohn TA, et al. Immunohistochemical labeling for *dpc4* mirrors genetic status in pancreatic adenocarcinomas: a new marker of *DPC4* inactivation. *Am J Pathol.* 2000;156:37–43.
- Ji H, Isacson C, Seidman JD, Kurman RJ, Ronnett BM. Cytokeratins 7 and 20, *Dpc4*, and *MUC5AC* in the distinction of metastatic mucinous carcinomas in the ovary from primary ovarian mucinous tumors: *Dpc4* assists in identifying metastatic pancreatic carcinomas. *Int J Gynecol Pathol.* 2002;21:391–400.
- Blackford A, Serrano OK, Wolfgang CL, Parmigiani G, Jones S, Zhang X, et al. *SMAD4* gene mutations are associated with poor prognosis in pancreatic cancer. *Clin Cancer Res.* 2009;15:4674–9.
- Maitra A, Molberg K, Albores-Saavedra J, Lindberg G. Loss of *Dpc4* expression in colonic adenocarcinomas correlates with the presence of metastatic disease. *Am J Pathol.* 2000;157:1105–11.
- Tascilar M, Skinner HG, Rosty C, Sohn T, Wilentz RE, Offerhaus GJ, et al. The *SMAD4* protein and prognosis of pancreatic ductal adenocarcinoma. *Clin Cancer Res.* 2001;7:4115–21.
- Davison JM, Hartman DA, Singhi AD, Choudry HA, Ahrendt SA, Zureikat AH, et al. Loss of *SMAD4* protein expression is associated with high tumor grade and poor prognosis in disseminated appendiceal mucinous neoplasms. *Am J Surg Pathol.* 2014;38:583–92.
- Singhi AD, Foxwell TJ, Nason K, Cressman KL, McGrath KM, Sun W, et al. Smad4 loss in esophageal adenocarcinoma is associated with an increased propensity for disease recurrence and poor survival. *Am J Surg Pathol.* 2015;39:487–95.
- Ke Z, Zhang X, Ma L, Wang L. Expression of *DPC4/Smad4* in non-small-cell lung carcinoma and its relationship with angiogenesis. *Neoplasma.* 2008;55:323–9.
- Goldstraw P, Ball D, Jett JR, Le Chevalier T, Lim E, Nicholson AG, et al. Non-small-cell lung cancer. *Lancet.* 2011;378:1727–40.
- Budwit-Novotny DA, McCarty KS, Cox EB, Soper JT, Mutch DG, Creasman WT, et al. Immunohistochemical analyses of estrogen receptor in endometrial adenocarcinoma using a monoclonal antibody. *Cancer Res.* 1986;46:5419–25.

25. Vavalà T, Monica V, Lo Iacono M, Mele T, Busso S, Righi L, et al. Precision medicine in age-specific non-small-cell-lung-cancer patients: integrating biomolecular results into clinical practice—a new approach to improve personalized translational research. *Lung Cancer*. 2017;107:84–90.
26. Rizvi H, Sanchez-Vega F, La K, Chatila W, Jonsson P, Halpenny D, et al. Molecular determinants of response to anti-programmed cell death (PD)-1 and anti-programmed death-ligand 1 (PD-L1) blockade in patients with non-small-cell lung cancer profiled with targeted next-generation sequencing. *J Clin Oncol*. 2018;36:633–41.
27. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the evolution of non-small-cell lung cancer. *N Engl J Med*. 2017;376:2109–21.
28. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*. 2017;545:446–51.
29. Hellmann MD, Nathanson T, Rizvi H, Creelan BC, Sanchez-Vega F, Ahuja A, et al. Genomic features of response to combination immunotherapy in patients with advanced non-small-cell lung cancer. *Cancer Cell*. 2018;33:843–52.
30. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. 2015;348:124–8.
31. Mehrvarz Sarshekeh A, Advani S, Overman MJ, Manyam G, Kee BK, Fogelman DR, et al. Association of SMAD4 mutation with patient demographics, tumor characteristics, and clinical outcomes in colorectal cancer. *PLoS ONE*. 2017;12:e0173345.
32. Ying Z, Tian H, Li Y, Lian R, Li W, Wu S, et al. CCT6A suppresses SMAD2 and promotes prometastatic TGF- β signaling. *J Clin Invest*. 2017;127:1725–40.
33. Hoot KE, Oka M, Han G, Bottinger E, Zhang Q, Wang XJ. HGF upregulation contributes to angiogenesis in mice with keratinocyte-specific Smad2 deletion. *J Clin Invest*. 2010;120:3606–16.
34. Petersen M, Pardali E, van der Horst G, Cheung H, van den Hoogen C, van der Pluijm G, et al. Smad2 and Smad3 have opposing roles in breast cancer bone metastasis by differentially affecting tumor angiogenesis. *Oncogene*. 2010;29:1351–61.
35. Oyanagi H, Shimada Y, Nagahashi M, Ichikawa H, Tajima Y, Abe K, et al. SMAD4 alteration associates with invasive-front pathological markers and poor prognosis in colorectal cancer. *Histopathology*. 2019;74:873–82.
36. Liao X, Hao Y, Zhang X, Ward S, Houldsworth J, Polydorides AD, et al. Clinicopathological characterization of SMAD4-mutated intestinal adenocarcinomas: a case-control study. *PLoS ONE*. 2019;14:e0212142.
37. Rozenblum E, Schutte M, Goggins M, Hahn SA, Panzer S, Zahurak M, et al. Tumor-suppressive pathways in pancreatic carcinoma. *Cancer Res*. 1997;57:1731–4.
38. Yatsuoka T, Sunamura M, Furukawa T, Fukushige S, Yokoyama T, Inoue H, et al. Association of poor prognosis with loss of 12q, 17p, and 18q, and concordant loss of 6q/17p and 12q/18q in human pancreatic ductal adenocarcinoma. *Am J Gastroenterol*. 2000;95:2080–5.
39. Chang YC, Chang JG, Liu TC, Lin CY, Yang SF, Ho CM, et al. Mutation analysis of 13 driver genes of colorectal cancer-related pathways in Taiwanese patients. *World J Gastroenterol*. 2016;22:2314–25.
40. Martin D, Abba MC, Molinolo AA, Vitale-Cross L, Wang Z, Zaida M, et al. The head and neck cancer cell oncogenome: a platform for the development of precision molecular therapies. *Oncotarget*. 2014;5:8906–23.
41. India Project Team of the International Cancer Genome Consortium. Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. *Nat Commun*. 2013;4:2873.
42. Boone BA, Sabbaghian S, Zenati M, Marsh JW, Moser AJ, Zureikat AH, et al. Loss of SMAD4 staining in pre-operative cell blocks is associated with distant metastases following pancreaticoduodenectomy with venous resection for pancreatic cancer. *J Surg Oncol*. 2014;110:171–5.
43. Yanagisawa K, Uchida K, Nagatake M, Masuda A, Sugiyama M, Saito T, et al. Heterogeneities in the biological and biochemical functions of Smad2 and Smad4 mutants naturally occurring in human lung cancers. *Oncogene*. 2000;19:2305–11.
44. Wan R, Xu X, Ma L, Chen Y, Tang L, Feng J. Novel alternatively spliced variants of Smad4 expressed in TGF- β -induced EMT regulating proliferation and migration of A549 cells. *Onco Targets Ther*. 2020;13:2203–13.
45. Wang Y, Tan X, Tang Y, Zhang C, Xu J, Zhou J, et al. Dysregulated Tgfr2/ERK-Smad4/SOX2 signaling promotes lung squamous cell carcinoma formation. *Cancer Res*. 2019;79:4466–79.
46. Weis-Banke SE, Lerdrup M, Kleine-Kohlbrecher D, Mohammad F, Sidoli S, Jensen ON, et al. Mutant FOXL2^{C134W} highjacks SMAD4 and SMAD2/3 to drive adult granulosa cell tumors. *Cancer Res*. 2020;80:3466–79.
47. Wasserman I, Lee LH, Ogino S, Marco MR, Wu C, Chen X, et al. SMAD4 loss in colorectal cancer patients correlates with recurrence, loss of immune infiltrate, and chemoresistance. *Clin Cancer Res*. 2019;25:1948–56.
48. Kawaguchi Y, Kopetz S, Newhook TE, De Bellis M, Chun YS, Tzeng CD, et al. Mutation status of RAS, TP53, and SMAD4 is superior to mutation status of RAS alone for predicting prognosis after resection of colorectal liver metastases. *Clin Cancer Res*. 2019;25:5843–51.
49. Ozawa H, Ranaweera RS, Izumchenko E, Makarev E, Zhavoronkov A, Fertig EJ, et al. SMAD4 loss is associated with cetuximab resistance and induction of MAPK/JNK activation in head and neck cancer cells. *Clin Cancer Res*. 2017;23:5162–75.
50. Ziemke M, Patil T, Nolan K, Tippimanchai D, Malkoski SP. Reduced Smad4 expression and DNA topoisomerase inhibitor chemosensitivity in non-small cell lung cancer. *Lung Cancer*. 2017;109:28–35.
51. Zhang S, Takaku M, Zou L, Gu AD, Chou WC, Zhang G, et al. Reversing SKI-SMAD4-mediated suppression is essential for TH17 cell differentiation. *Nature*. 2017;551:105–9.
52. Xu H, Agaloti T, Zhao J, Steglich B, Wahib R, Vesely MCA, et al. The induction and function of the anti-inflammatory fate of TH17 cells. *Nat Commun*. 2020;11:3334.
53. Wu B, Zhang G, Guo Z, Wang G, Xu X, Li JL, et al. The SKI proto-oncogene restrains the resident CD103+CD8+ T cell response in viral clearance. *Cell Mol Immunol*. 2020;10:1038.