109:120-129

7 Martin J, Ginsberg RJ, Venkatraman ES, et al. Long-term results of combined-modality therapy in resectable non-small-cell lung cancer. J Clin Oncol 2002; 20:1989-1995

8 Yoshino I, Yohena T, Kitajima M, et al. Survival of non-small cell lung cancer patients with postoperative recurrence at distant organs. Ann Thorac Cardiovasc Surg 2001; 7:204-209

9 Maeda R, Yoshida J, Hishida T, et al. Late recurrence of non-small cell lung cancer more than 5 years after complete resection: incidence and clinical implications in patient follow-up. Chest 2010; 138:145-150

10 Martini N, Rusch VW, Bains MS, et al. Factors influencing ten-year survival in resected stages I to IIIa non-small cell lung cancer. J Thorac Cardiovasc Surg 1999; 117:32-36; discussion 37-38

11 Okada M, Nishio W, Sakamoto T, et al. Long-term survival and prognostic factors of five-year survivors with complete resection of non-small cell lung carcinoma. J Thorac Cardiovasc Surg 2003; 126:558-562

12 Endo C, Sakurada A, Notsuda H, et al. Results of long-term follow-up of patients with completely resected non-small cell lung cancer. Ann Thorac Surg

2012; 93:1061-1068

13 Sugimura H, Nichols FC, Yang P, et al. Survival after recurrent nonsmall-cell lung cancer after complete pulmonary resection. Ann Thorac Surg 2007; 83:409-417; discussioin 417-408

14 Williams BA, Sugimura H, Endo C, et al. Predicting postrecurrence survival among completely resected nonsmall-cell lung cancer patients. Ann Thorac Surg 2006; 81:1021-1027

15 Maemondo M, Inoue A, Kobayashi K, et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. N Engl J Med 2010; 362:2380-2388

16 Mitsudomi T, Morita S, Yatabe Y, et al. Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. Lancet Oncol 2010; 11:121-128

17 Mok TS, Wu YL, Thongprasert S, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 2009; 361:947-957

18 Paz-Ares L, de Marinis F, Dediu M, et al. Maintenance therapy with pemetrexed plus best supportive care versus placebo plus best supportive care after induction

therapy with pemetrexed plus cisplatin for advanced non-squamous non-small-cell lung cancer (PARAMOUNT): a double-blind, phase 3, randomised controlled trial. Lancet Oncol 2012; 13:247-255

19 Reck M, von Pawel J, Zatloukal P, et al. Overall survival with cisplatin-gemcitabine and bevacizumab or placebo as first-line therapy for nonsquamous non-small-cell lung cancer: results from a randomised phase III trial (AVAiL). Ann Oncol 2010; 21:1804-1809

20 Sandler A, Gray R, Perry MC, et al. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. N Engl J Med 2006; 355:2542-2550

21 Scagliotti GV, Parikh P, von Pawel J, et al. Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. J Clin Oncol 2008; 26:3543-3551

22 Travis WD, Brambilla E, Muller-Hermelink HK et al. World Health Organization Classification of Tumors: Pathology and Genetics of Tumors of the Lung, Pleura, Thymus and Heart. Lyon, France: IARC Press; 2004.

23 International Union Against Cancer. TNM Classification of Malignant Tumours.

7 th ed. Oxford, UK: Wiley-Blackwell; 2009.

24 Jones DR, Daniel TM, Denlinger CE, et al. Stage IB nonsmall cell lung cancers: are they all the same? Ann Thorac Surg 2006; 81:1958-1962; discussion 1962

25 Brechot JM, Chevret S, Charpentier MC, et al. Blood vessel and lymphatic vessel invasion in resected nonsmall cell lung carcinoma. Correlation with TNM stage and disease free and overall survival. Cancer 1996; 78:2111-2118

26 Ichinose Y, Yano T, Asoh H, et al. Prognostic factors obtained by a pathologic examination in completely resected non-small-cell lung cancer. An analysis in each pathologic stage. J Thorac Cardiovasc Surg 1995; 110:601-605

27 Kobayashi N, Toyooka S, Soh J, et al. Risk factors for recurrence and unfavorable prognosis in patients with stage I non-small cell lung cancer and a tumor diameter of 20 mm or less. J Thorac Oncol 2007; 2:808-812

28 Maeda R, Yoshida J, Ishii G, et al. Long-term survival and risk factors for recurrence in stage I non-small cell lung cancer patients with tumors up to 3 cm in maximum dimension. Chest 2010; 138:357-362

29 Maeda R, Yoshida J, Ishii G, et al. Prognostic impact of intratumoral vascular invasion in non-small cell lung cancer patients. Thorax 2010; 65:1092-1098

30 Maeda R, Yoshida J, Ishii G, et al. Poor prognostic factors in patients with stage

IB non-small cell lung cancer according to the seventh edition TNM classification. Chest 2011; 139:855-861

31 Miyoshi K, Moriyama S, Kunitomo T, et al. Prognostic impact of intratumoral vessel invasion in completely resected pathologic stage I non-small cell lung cancer. J Thorac Cardiovasc Surg 2009; 137:429-434

32 Ruffini E, Asioli S, Filosso PL, et al. Significance of the presence of microscopic vascular invasion after complete resection of Stage I-II pT1-T2N0 non-small cell lung cancer and its relation with T-Size categories: did the 2009 7th edition of the TNM staging system miss something? J Thorac Oncol 2011; 6:319-326

33 Shimada Y, Ishii G, Hishida T, et al. Extratumoral vascular invasion is a significant prognostic indicator and a predicting factor of distant metastasis in non-small cell lung cancer. J Thorac Oncol 2010; 5:970-975

34 Shimizu K, Yoshida J, Nagai K, et al. Visceral pleural invasion is an invasive and aggressive indicator of non-small cell lung cancer. J Thorac Cardiovasc Surg 2005; 130:160-165

35 Shimizu K, Yoshida J, Nagai K, et al. Visceral pleural invasion classification in non-small cell lung cancer: a proposal on the basis of outcome assessment. J

Thorac Cardiovasc Surg 2004; 127:1574-1578

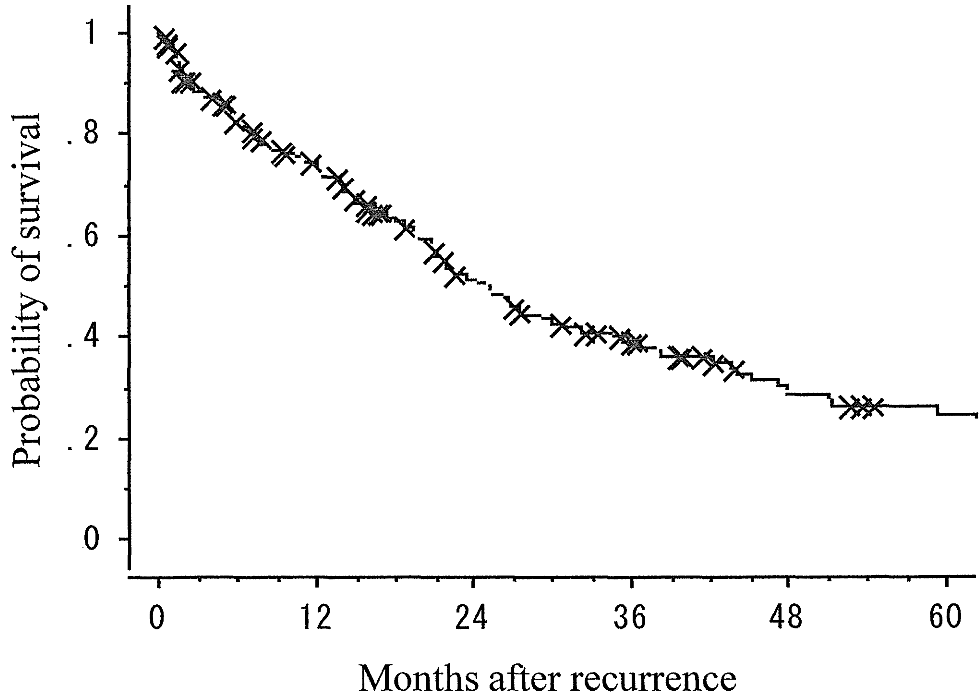
36 Tsuchiya T, Akamine S, Muraoka M, et al. Stage IA non-small cell lung cancer: vessel invasion is a poor prognostic factor and a new target of adjuvant chemotherapy. Lung Cancer 2007; 56:341-348

37 Garfield D, Normanno N, Cadranel J. Prognostic factor for non-small cell lung cancer with bone metastases at the time of diagnosis. Lung Cancer 2012; 78:168

38 Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 2004; 350:2129-2139

39 Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004; 304:1497-1500

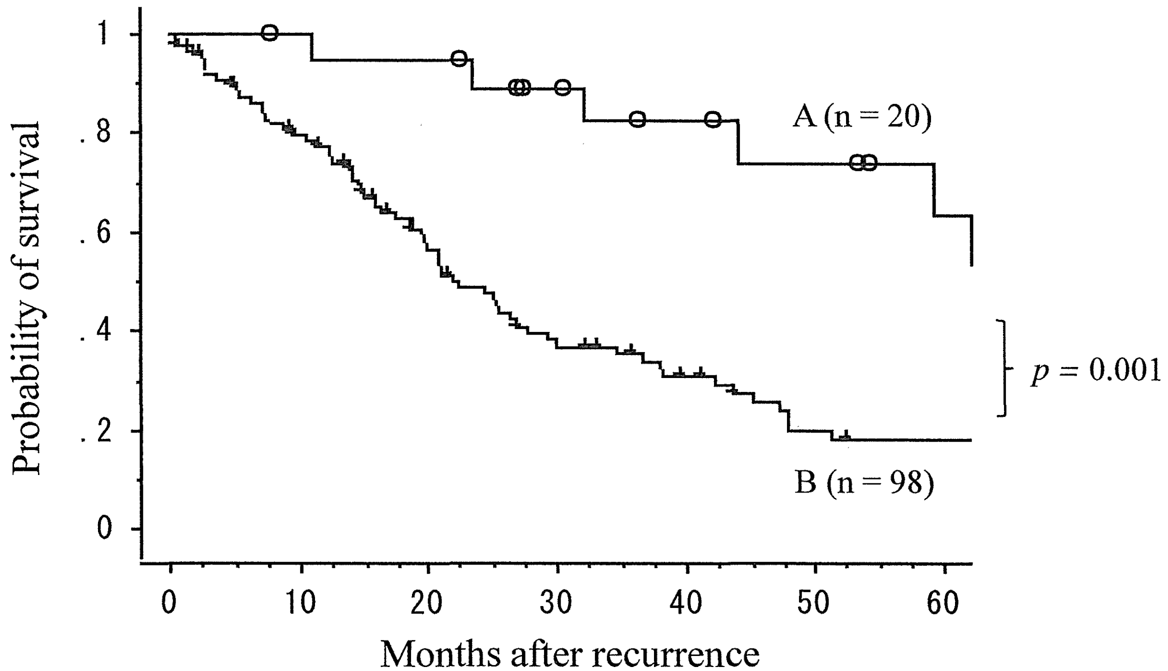
Figure 1



Patients at risk of death (n = 170)

170	109	65	42	25	19

Figure 2



Patients at risk of death (n = 118)

A	20	18	16	12	9	6
B	98	67	37	24	11	9

A: The patients lacking all 5 unfavorable factors (not receiving EGFR-TKIs and chemotherapy, presence of liver or bone metastasis, non-adenocarcinoma)
B: The patients with one of the above mentioned risk factors

Table 1

Patient characteristics, and univariate and multivariate analyses of recurrence

Factors	Univariate analysis			Multivariate analysis		
	Number	5-y RFP (%)	*p*-value	HR	95% CI	*p*-value
Age (years; median 65)						
< 65	439	84.1				
≥ 65	480	80.4	0.129			
Gender						
Male	542	78.0				
Female	377	87.8	< 0.001			
Smoking status						
Never smoker	347	85.2	0.134			
Ever smoker	572	80.2				
T category						
T1	512	84.7				
T2	407	78.9	0.100			
Tumor size						
0-30 mm	663	84.0				
> 30 mm	256	81.5	0.112			
Pathological vascular invasion						
Absent	481	91.0		1		
Present	421	72.1	< 0.001	2.306	1.621-3.280	< 0.001
Pleural invasion						
Absent	719	84.9		1		
Present	191	71.8	< 0.001	1.489	1.048-2.115	0.026
Histology						
Adenocarcinoma	706	83.8				
Nonadenocarcinoma	213	76.3	0.039			
Differentiation						
Well or moderate	656	86.7		1		
Poor	216	67.7	< 0.001	1.842	1.328-2.555	< 0.001
Type of surgery						
Single lobectomy	873	81.9				

Bilobectomy or pneumonectomy	46	87.2	0.942

5y-RFP: 5 year recurrence-free proportion, HR: hazard ratio, CI: confidence interval

Table 2
Initial recurrence site and postrecurrence therapy

	Number
Overall	170
Type of recurrence	
Distant	113
Local	43
Both	14
Initial recurrence site	
Ipsilateral lung	23
Contralateral/Bilateral lung	43
Regional lymph nodes	37
Malignant effusion/dissemination	13
Stump	9
Brain	30
Bone	21
Liver	16
Adrenal gland	10
Others	14
Postrecurrence therapy	
Initial therapy	
Surgery	8 (lung 3, brain 3, adrenal gland 1, lymph nodes 1)
Surgery alone	6
Surgery + CT	3
CT	79
RT	10
CRT	21
None	41
Unknown	11
Second-line or the subsequent therapy	66
CT	58
EGFR-TKIs	27 (gefitinib 22/erlotinib 3/both 2)
EGFR mutation status/histology	positive 12 (Ad 11/Sq 1)

	wild	4 (Ad 3/LCC 1)
	unknown	11 (Ad 10/LCC 1)
Others		7

CT: chemotherapy, RT: radiation therapy, CRT: chemoradiotherapy, EGFR-TKIs: epidermal growth factor receptor-tyrosine kinase inhibitors, Ad: adenocarcinoma, Sq: squamous cell carcinoma, LCC: large cell carcinoma

Table 3
Postrecurrence survival analyses

Factors	Univariate analysis			Multivariate analysis		
	Number	Median PRS (months)	*p*-value	HR	95% CI	*p*-value
Age at recurrence (median: 66)						
< 66	76	18.9				
≥ 66	94	15.8	0.242			
Gender						
Male	118	15.5		1		
Female	52	25.6	< 0.001	0.487	0.297-0.801	0.005
Smoking status						
Never smoker	59	25.0				
Ever smoker	111	14.1	0.006			
T category						
T1	87	15.8				
T2	83	19.6	0.476			
Tumor size						
0-30 mm	132	16.9				
> 30 mm	38	20.9	0.632			
Pathological vascular invasion						
Absent	53	15.8				
Present	113	17.0	0.088			
Pleural invasion						
Absent	115	15.8				
Present	53	18.8	0.393			
Histology						
Adenocarcinoma	124	20.9				
Nonadenocarcinoma	46	12.4	< 0.001			
Differentiation						
Well or moderate	97	20.8		1		
Poor	65	14.1	0.002	1.810	1.194-2.743	0.005

Type of surgery						
Single lobectomy	162	17.3	0.152			
Bilobectomy or pneumonectomy	8	19.5				
Adjuvant therapy						
Without	134	15.9	0.547			
With	36	21.0				
Postrecurrence therapy						
Without	41	7.2		1		
With	118	21.4	0.021	0.542	0.344-0.853	0.008
Recurrence free interval						
≤ 24 months	82	16.2				
> 24months	88	18.4	0.021			
Type of recurrence						
Distant	127	15.8				
Local only	43	18.8	0.087			
Number of recurrent sites						
Single	132	16.8				
Multiple	38	18.6	0.305			

PRS: postrecurrence survival, HR: hazard ratio, CI: confidence interval

Table 4

Postrecurrence survival analyses in 118 patients who underwent postrecurrence therapy

Factors	Univariate analysis			Multivariate analysis		
	Number	Median PRS (months)	*p*-value	HR	95% CI	*p*-value
Age at recurrence						
< 66	63	22.4				
≥ 66	55	19.5	0.151			
Gender						
Male	79	20.0				
Female	39	27.2	0.002			
Smoking status						
Never smoker	43	27.6				
Ever smoker	75	17.6	0.035			
Histology						
Adenocarcinoma	84	24.4		1		
Nonadenocarcinoma	34	13.9	< 0.001	2.136	1.273-3.585	0.004
Differentiation						
Well or moderate	66	23.1				
Poor	46	18.8	0.019			
Lung metastasis						
Absent	68	19.8				
Present	49	21.4	0.053			
Brain metastasis						
Absent	96	19.6				
Present	21	22.6	0.584			
Bone metastasis						
Absent	100	21.9		1		
Present	17	15.8	0.001	3.288	1.783-6.062	< 0.001
Liver metastasis						
Absent	110	21.9		1		
Present	7	10.5	0.001	4.518	1.793-11.379	0.001
Chemotherapy						
Without	15	9.6		1		

With	103	22.7	0.009	0.478	0.236-0.975	0.040
Surgical resection						
Without	110	20.8				
With	8	33.7	0.209			
EGFR-TKI therapy						
Without	91	17.0		1		
With	27	41.4	0.002	0.460	0.245-0.862	0.015
Second line therapy						
Without	52	14.0				
With	66	27.2	0.004			
Recurrence free interval						
≤ 24 months	59	17.0				
> 24 months	59	22.4	0.394			
Type of recurrence						
Distant	85	20.8				
Local only	33	21.8	0.086			
Number of recurrent sites						
Single	89	21.0				
Multiple	29	20.8	0.049			

PRS: postrecurrence survival, HR: hazard ratio, CI: confidence interval, CT: chemotherapy, EGFR-TKI: epidermal growth factor receptor-tyrosine kinase inhibitor

Table 5

Postrecurrence survival of patients with stage I non-small cell lung cancer in previous series

Series	No. of patients	Incidence of recurrence (%)	PRS	Type of recurrence	Independent favorable factors of PRS
Martini (1995) [6]	598	159 (26.6)	NR	L/ D	NR
Al-Kattan (1997) [1]	123	36 (29.3)	NR	L/ D	NR
Nakagawa (2008) [4]	397	87 (21.9)	67.7% (1y) 34.4% (3y)	L/ D	Symptom at recurrence (-) Cervico-mediastinum meta. (-) Liver meta. (-) PRT (Surgery/ non-surgery)
Hung (2009) [2]	933	74 (7.9)	48.7% (1y) 17.6% (2y)	L	PRT (Surgery, CT and/or RT)
Hung (2010) [3]	933	166 (17.8)	30.2% (1y) 15.1% (2y)	D	Disease-free interval > 16 mo PRT
Our series (2012)	919	170 (18.5)	73.5% (1y) 51.4% (2y)	L/ D	PRT Female

PRS: postrecurrence survival, L: Local recurrence, D: Distant recurrence, NR: not reported,
PRT: postrecurrence therapy, CT: chemotherapy, RT: radiotherapy

Open

Oncogene (2012), 1–7
© 2012 Macmillan Publishers Limited All rights reserved 0950-9232/12
www.nature.com/onc

REVIEW

Genome-wide mechanisms of Smad binding

M Morikawa[1,2], D Koinuma[2], K Miyazono[1,2] and C-H Heldin[1]

A dual role of transforming growth factor β (TGF-β), to both suppress and promote tumor progression and metastasis, has been well established, but its molecular basis has remained elusive. In this review, we focus on Smad proteins, which are central mediators of the signal transduction of TGF-β family members. We describe current knowledge of cell-type-specific binding patterns of Smad proteins and mechanisms of transcriptional regulation, obtained from recent studies on genome-wide binding sites of Smad molecules. We also discuss potential application of the genome-wide analyses for cancer research, which will allow clarification of the complex mechanisms occurring during cancer progression, and the identification of potential biomarkers for future cancer diagnosis, prognosis and therapy.

Oncogene advance online publication, 21 May 2012; doi:10.1038/onc.2012.191

Keywords: ChIP-chip; ChIP-sequencing; TGF-β; BMP; Smad

INTRODUCTION

Members of the transforming growth factor β (TGF-β) family, which include three TGF-β isoforms, as well as activins, nodal and bone morphogenetic proteins (BMPs), regulate a variety of cellular processes including differentiation, proliferation, migration and cell death in cell-type-specific and context-dependent manners.[1–3] The biological effects of TGF-β family members are highly contextual, for example, their responses may differ in different tissues, local environments and stage of disease. Since TGF-β activates cytostatic and cell death processes that maintain homeostasis in mature tissues, it functions as a suppressor of epithelial cell tumorigenesis at early stages. Inactivation of the TGF-β signaling pathway through mutation and/or loss of heterozygosity of TGF-β receptors or Smad proteins has been found in certain types of cancer and is related to poor prognosis for the patients (reviewed in Levy and Hill[4]). However, TGF-β promotes tumor progression by enhancing migration, invasion and survival of tumor cells during the later stages of tumorigenesis, through stimulating extracellular matrix deposition and tissue fibrosis, perturbing immune and inflammatory function, stimulating angiogenesis and promoting epithelial–mesenchymal transition (reviewed in Yoshimura *et al.*[5], Roberts and Wakefield[6], Moustakas and Heldin[7] and Miyazono *et al.*[8]). Accumulating evidences also indicate critical roles of TGF-β/activin signaling in the maintenance of stem cell-like properties of certain cancer-initiating cells, such as glioma-initiating cells,[9,10] breast cancer-initiating cells,[11] pancreatic cancer-initiating cells,[12] and leukemia-initiating cells in chronic myeloid leukemia.[13] Intriguingly, small molecular inhibitors for type I receptors have therapeutic effects at least in animal models.[9,10,12,13] These observations suggest that targeting the TGF-β/activin signaling pathways could be an attractive therapy in certain advanced cancers, although it is possible that shutdown of these pathways in normal tissues will increase the risk for the development of other tumors. Thus, one of the major questions that remain to be addressed in this field is what defines the dual role of TGF-β in cancer biology.

Identification of the signaling components of TGF-β family members, including membrane receptor serine/threonine kinases and Smad transcription factors, has led to an understanding of the molecular mechanisms underlying this highly contextual process.[14,15] Genome-wide transcriptome analyses in various cell types have identified many target genes that are required for ligand-mediated cellular responses. Direct binding of Smad complexes was confirmed by *in vitro* binding assays, promoter assays and chromatin immunoprecipitation (ChIP) followed by polymerase chain reaction. Until recently, however, regulatory elements were mainly identified in the promoter regions of the target genes, especially 1–2 kb upstream of their transcription start sites.

ChIP with promoter array analysis (ChIP-chip) and ChIP followed by sequencing (ChIP-seq) have become powerful tools to analyze genome-wide mapping of protein-binding sites and epigenetic marks.[16,17] In this case, a DNA sample obtained after ChIP procedure is analyzed using promoter-tiling arrays, or massively parallel sequencing (Supplementary Figure 1), which provides a comprehensive chromatin-binding landscape of target transcription factors. Information obtained by these analyses has shed light on previously unrecognized mechanisms and sometimes challenged notions previously characterized in a specific situation. Recently, several groups have reported that Smad proteins tend to co-occupy target sites with cell-type-specific master transcription factors.[18–20] The results also indicate that co-occupied regions mainly overlap with enhancer elements, although previous studies have identified numerous Smad-responsive elements in the promoter regions of their target genes. In addition, recent ChIP-chip/ChIP-seq studies have identified a group of direct target genes, or target gene signatures, in specific cell types and cellular contexts. Intriguingly, Kennedy *et al.*[21] reported that the TGF-β/Smad4 target gene signature identified in ovarian cancer cell lines predicts patient survival.

In this review, we discuss current knowledge of cell-type-specific binding patterns of Smad proteins and mechanisms of transcriptional regulation obtained from recent ChIP-chip/

[1]Ludwig Institute for Cancer Research, Science for Life Laboratory, Uppsala University, Biomedical Center, Uppsala, Sweden and [2]Department of Molecular Pathology, Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo, Japan. Correspondence: Professor K Miyazono, Department of Molecular Pathology, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.
E-mail: miyazono@m.u-tokyo.ac.jp
Received 1 March 2012; revised 5 April 2012; accepted 13 April 2012

ChIP-seq studies (Supplementary Table 1). We also highlight applications of the genome-wide analyses for cancer research. These insights contribute to the unraveling of the complex mechanisms of TGF-β signaling in cancer biology.

OVERVIEW OF SIGNALING PATHWAYS OF TGF-β FAMILY MEMBERS

The TGF-β family consists of 33 members in mammals. Two types of serine/threonine kinase transmembrane receptors, that is, type II and type I receptors, are required for intracellular signal transduction by the TGF-β family members.[14] Five type II receptors and seven type I receptors are present in mammals.[22] Ligand binding assembles specific type II and type I receptors into heterotetramers. Then the type II receptor transphosphorylates and activates the type I receptor, which subsequently transduces the signal by phosphorylating the carboxyl terminus of receptor-regulated (R)-Smad. In most cell types, TGF-β and activin induce phosphorylation of Smad2 and Smad3 (activin/TGF-β-specific R-Smads, or AR-Smads) and BMPs induce phosphorylation of Smad1, Smad5 and Smad8 (BMP-specific R-Smads, or BR-Smads). Activated R-Smads form heterooligomeric complexes with common-partner (co)-Smad (Smad4). The complexes translocate into the nucleus where they regulate the expression of target genes, such as the genes for *Serpine1* (plasminogen activator inhibitor-1), inhibitory (I)-Smads (Smad6 and Smad7) and *Id1* (inhibitor of differentiation-1 or inhibitor of DNA binding-1) (Figure 1). Because of their relatively low DNA-binding affinity, Smad complexes interact with a wide variety of DNA-binding proteins and cooperatively regulate a synexpression group of target genes (Figure 2a).[2] So far, several transcription factors, such as AP-1,[23] ETS,[24,25] basic helix-loop-helix proteins,[26,27] C/EBPβ,[28] FoxH1[29,30] and FoxO[31] have been identified and validated as important cofactors of TGF-β/BMP signaling pathways. In addition, Smad complexes recruit coactivators, such as p300 and CREB-binding protein,[32,33] or corepressors, such as ATF-3.[34] For example, TGF-β represses transcription of the *Id1* gene in epithelial cells through formation of a complex with ATF-3, while TGF-β induces Id1 in cells which do not express ATF-3, such as glioma-initiating cell-like cells.[35] Since ATF-3 is induced by tumor necrosis factor-α, signaling crosstalk between TGF-β and tumor necrosis factor-α pathways determines the transcriptional regulation of Id1. Thus, crosstalk with other signaling pathways and interaction with other DNA-binding cofactors define the specific binding patterns of Smads; in addition, interaction with coactivators/corepressors modulates their transcriptional activity (Figure 1).

Smad proteins are targets of protein modifications, such as phosphorylation, ubiquitination and ADP-ribosylation. The cyclin-dependent kinases (CDKs) CDK8 and CDK9, which are downstream effectors of extracellular-signal-regulated kinase (ERK) MAP kinase, phosphorylate the linker region of Smads in the nucleus.[36–39] Glycogen synthase kinase-3β (GSK3β) also phosphorylates the linker region of Smads, which requires priming phosphorylation by ERK MAP kinase.[40] These phosphorylations mark the proteins for polyubiquitination and promote proteasome-mediated degradation of Smad complexes. Several WW domain proteins have been reported to recognize the phosphorylated linker regions and interact with R-Smads.[41] Smurf1 is a member of the E3 ubiquitin ligase family, which can target BR-Smads for degradation,[42] while NEDD4L (also known as NEDD4-2) is an E3 ubiquitin ligase for AR-Smads.[43,44] Consequently, endogenous ERK MAPK and GSK3β signaling pathways are able to antagonize Smad activity through proteasome-mediated degradation. Recently, deubiquitinating enzymes (DUBs) for Smad proteins have been identified.[45,46] Monoubiquitination of the lysine-519 (K519) residue of Smad4 prevents its association with R-Smads and negatively regulates TGF-β/BMP signaling pathway. USP9x (also known as FAM) has been identified as a DUB that reverts this modification.[45] R-Smads are monoubiquitinated in their DNA-binding domains, which attenuates their affinity for DNA. This monoubiquitination is opposed by another DUB, USP15.[46] Recently, Lonn *et al.*[47] found that Smad proteins are targets of ADP-ribosylation. Poly(ADP-ribose) polymerase-1 (PARP-1) interacts with and ADP-ribosylates Smad3 and Smad4 in the nucleus, and affects the binding affinity of Smad complexes in a context-dependent manner.[47,48] Thus, posttranslational modifications of Smad proteins affect their signal transduction capacities; some of these modifications are regulated by other signaling pathways (Figure 1).

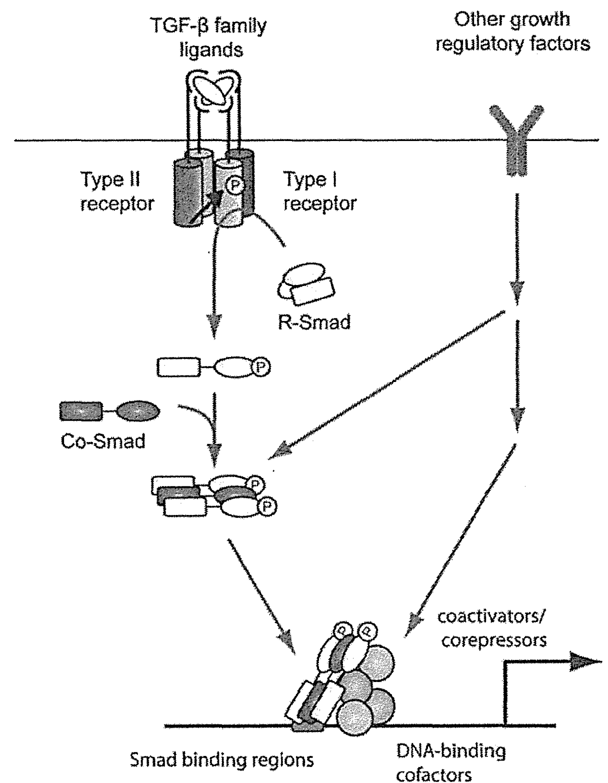


Figure 1. Signaling of TGF-β family members through Smad complexes. Smad proteins are central mediators of the signal transduction of TGF-β family members. Ligand binding assembles specific type II and type I receptors into heterotetramers. The type II receptor transphosphorylates (P) and activates the type I receptor, which subsequently activates receptor-regulated (R)-Smads. Activated R-Smads form heterooligomeric complexes with common-partner (co)-Smad. In the nucleus, Smad complexes interact with DNA-binding cofactors and cooperatively regulate a group of target genes. Crosstalk with other growth regulatory factors affects the specific binding patterns and transcriptional activity of Smads.

SMAD-BINDING MOTIFS

The R-Smads and Smad4 are composed of two evolutionally conserved domains named Mad Homology 1 and 2 (MH1 and MH2). The MH2 domain plays an important role for the formation of heterooligomeric Smad complexes and transcriptional activation, whereas the MH1 domain is responsible for sequence-specific DNA-binding activity. Using a polymerase chain reaction-based random-oligonucleotide selection process, an 8-bp palindromic DNA sequence, GTCTAGAC, was identified as a Smad3

© 2012 Macmillan Publishers Limited

and Smad4 binding motif.[49] In contrast to Smad3 and Smad4, Smad2 does not directly bind to DNA due to steric hindrance by an inserted sequence in the DNA-binding region.[50] The crystal structures of the MH1 domain of Smad1 and Smad3 have revealed that R-Smads recognize and directly bind to half of the palindrome, that is, GTCT or AGAC sequences, through an 11-amino-acid residue β-hairpin loop in the MH1 domain.[51–53] The amino-acid sequences of the loop are completely conserved among R-Smads and show a high level of similarity between R-Smads and Smad4. The half-site sequences are usually referred to as the CAGA box or Smad binding element (SBE). Recent ChIP-chip/ChIP-seq studies have confirmed that the SBE is enriched in the Smad2/3-binding regions.[18,24,26,54,55]

Although the MH1 domain of Smad1 has high affinity for SBE,[52,53] BR-Smads seem to prefer a GC-rich sequence, such as GCCGnCGC, which was originally identified in *Drosophila*.[56] In mammals, GC-rich sequences, such as GCCG and (T)GGCGCC, have been identified in the promoter regions of several BMP target genes. Using a *de novo* motif-finding method, we identified a Smad1/5-binding motif, which is consistent with the previously reported GC-rich sequences and thus named as GC-rich SBE (GC-SBE).[57] Importantly, both GC-SBE and SBE are enriched in the Smad1/5-binding sites identified in both endothelial cells (ECs) and pulmonary arterial smooth muscle cells (PASMCs).[57] Since binding motifs for R-Smads have been identified *in vitro* and *in vivo*, candidate Smad-binding sites can be predicted in the promoter regions of the target genes. However, these motifs are common throughout the genome, and the majority of them are not occupied by R-Smads when examined using ChIP-chip/ChIP-seq. Thus, additional mechanisms operate to determine the binding patterns of Smads.

FACTORS THAT DETERMINE THE BINDING PATTERNS OF SMADS

Recent studies have suggested that Smad complexes colocalize with master transcription factors that specify and maintain cell identities.[18–20] Chen *et al.*[20] pointed out that Smad1 colocalizes in the multiple transcription factor-binding loci with embryonic stem (ES) cell-specific transcription factors, such as Oct4 and Sox2 in mouse ES cells (mESCs). Mullen *et al.*[18] reported that binding regions of Smad3 also overlap with those of Oct4 in both human and mouse ES cells. Intriguingly, at least some of these co-occupied regions are still enriched after tandem ChIP-re-ChIP experiments, indicating that Oct4 and Smad3 bind to similar regions in mESCs simultaneously.[18] Moreover, Smad3 colocalizes with MyoD (encoded by *Myod1*) or PU.1, master transcription factors controlling muscle or hematopoietic differentiation, respectively, in specific cell types which express these genes; forced expression of MyoD in mESCs is sufficient to redirect Smad3 to muscle specific binding sites, where they colocalize.[18] In addition, Trompouki *et al.*[19] reported that induction of the myeloid lineage regulator C/EBPα shifted Smad1 to sites newly occupied by C/EBPα in the human erythroleukemia cell line K562. Overexpression of the erythroid regulator GATA1 restricts Smad1 binding to erythroid genes, while binding to genes expressed in other lineages is diminished.[19] These findings suggest that Smad complexes are passively recruited to cell-type-specific binding sites through the interaction with master transcription factors.

On the other hand, we recently found that HNF4α, one of the master regulators of hepatocyte differentiation and liver function, contributes to the hepatocyte-specific binding pattern of Smad2/3.[58] Interestingly, 32.5% of the Smad2/3-binding regions overlapped with those of HNF4α. This is against the simple model in which cell-type-specific master regulators recruit R-Smads to their binding sites and determine their function. In addition, through the analysis of the distances between the Oct4 peak and the peaks of Sox2 and Smad3 in mESCs, Mullen *et al.*[18] found that Oct4 sites are more closely associated with Sox2 sites than Smad3 sites, suggesting that Oct4 and Smad3 do not interact in a direct manner. They revealed that nucleosomes were relatively depleted at the sites co-occupied by cell-type-specific master transcription factors and Smad3, and hypothesized that master transcription factors increase the accessibility of SBEs and contribute to Smad3 binding. Intriguingly, MyoD binding has been reported to be associated with local histone acetylation.[59] PU.1 and C/EBPα binding has been reported to induce nucleosome remodeling, followed by monomethylation of H3K4.[60] John *et al.*[61] reported that cell-type-specific glucocorticoid receptor binding patterns are comprehensively predetermined by cell-specific differences in baseline chromatin accessibility patterns, with secondary contributions from local sequence features. Similarly, comparison of Smad1/5-binding patterns of ECs and PASMCs suggested that the endothelial-specific binding pattern of Smad1/5 is predetermined by baseline chromatin accessibility patterns.[57] Thus, these facts support the notion that Smad complexes determine their target sites together with other DNA-binding cofactors in two different ways: (1) cell-type- or lineage-specific transcription factors, or pioneer factors,[62] open up local chromatin structure to make SBE and GC-SBE accessible and (2) DNA-binding cofactors, induced and activated in context-dependent manner, strengthen the interaction between Smad and DNA (Figure 2b).

Intriguingly, it has been observed that different levels of activation of Smad signaling pathways cause different binding patterns of Smad complexes, possibly correlating to the amount of activated Smad complexes in the nucleus.[63] It has been well described that different concentrations of activin regulate the expression of distinct subsets of target genes.[64] Lee *et al.*[54] confirmed that phospho-Smad2 is dose-dependently able to bind to different subsets of target genes and regulate their transcription in mESCs. Comparing the ChIP-seq data of different BMP isoforms in ECs, we found that each binding site has different binding affinity for Smad complexes and that the strength of Smad1/5 signaling affects the number and distribution of Smad-binding sites over the genome.[57] Thus, these findings suggest that a distinct dose-dependency occurs in the regulation of different subsets of target genes, which may cause phenotypic change.

SMAD BINDING AND HISTONE MODIFICATION MARKERS

As discussed above, local chromatin structure or accessibility affects the binding patterns of Smads. Recent studies have emphasized the importance of enhancers for the precise regulation of expression of target genes.[18–20,54,57] On the other hand, several groups have found that most of the Smad-binding sites are located at promoters of known genes.[30,65,66] Kim *et al.*[30] reported that 50–60% of Smad2/3 binding occurs in exons and promoters in human ES cells (hESCs), while only 10–15% of Smad binding occurs in exons and promoters in derived endoderm. This finding suggests that the preference of binding pattern of Smads to either promoters or enhancers is modulated by the differentiation stages.

Smad proteins have also been shown to induce local chromatin remodeling and modification at their binding sites. Both Smad1/5 and Smad2/3 have been reported to physically interact with a histone demethylase, KDM6B (also known as JMJD3), to recruit it to the *NOG* (encoding noggin) and *NODAL* promoter regions, respectively, and to cause the loss of the repressive mark histone H3 lysine-27 trimethylation (H3K27me3) in mESCs.[67,68] Recently, Kim *et al.*[30] reported that Smad2/3 and KDM6B are simultaneously enriched in the *GSC* (encoding goosecoid) and *EOMES* (encoding eomesdermin) promoter of hESCs after activin treatment, followed by the loss of the H3K27me3 repressive mark (Figure 3a). Interestingly, Fei *et al.*[65] identified that KDM6B is one of

© 2012 Macmillan Publishers Limited