Open

npg

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.

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

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
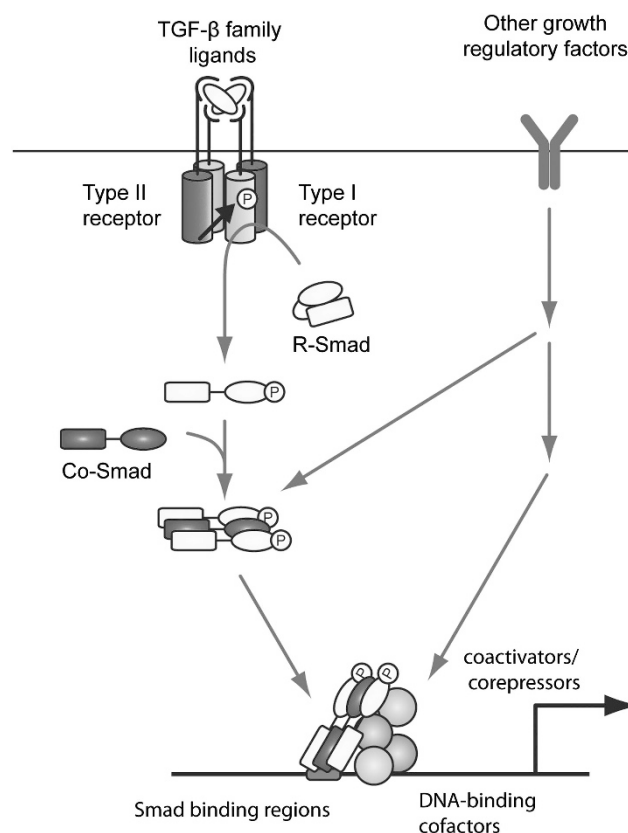

Figure 1. Signaling of TGF-β family members through Smad complexes. Smad proteins are central mediators of the signal transduction of TGF-β family members. Ligand binding assembles specific type II and type I receptors into heterotetramers. The type II receptor transphosphorylates (P) and activates the 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.

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).

SMAD-BINDING MOTIFS

The R-Smads and Smad4 are composed of two evolutionarily 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

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
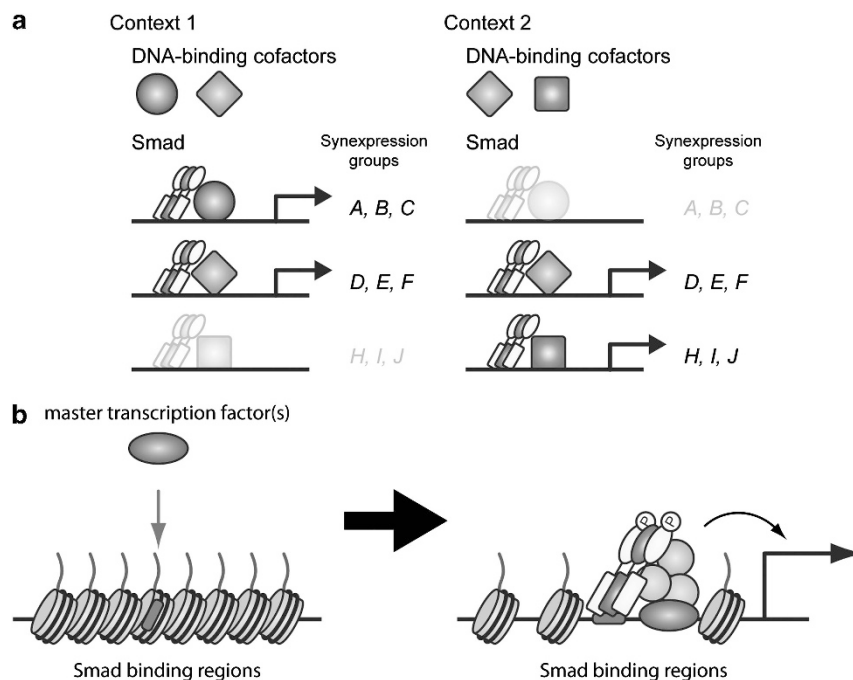
a

Context 1

DNA-binding cofactors

Smad   Synexpression
              groups

              A, B, C

              D, E, F

              H, I, J

Context 2

DNA-binding cofactors

Smad   Synexpression
              groups

              A, B, C

              D, E, F

              H, I, J

b master transcription factor(s)

Smad binding regions

Smad binding regions

Figure 2. Factors that determine the binding patterns of Smads. (**a**) A group of genes that are simultaneously regulated by a specific Smad–cofactor complex is known as a synexpression group. Distinct combinations of DNA-binding cofactors in different contexts determine the set of genes regulated by Smad complexes. (**b**) Cell-type- or lineage-specific master transcription factors (purple) open up local chromatin structure to make Smad-binding regions (red) accessible. The master transcription factors also physically interact with Smads and, in some cases, recruit them to their binding sites. DNA-binding cofactors, induced and activated in context-dependent manner, strengthen the interaction between Smad and DNA. Interaction with coactivators/corepressors also affects the regulation of their target genes. A full colour version of this figure is available at the *Oncogene* journal online.

the BMP4-modulated early neural differentiation regulators, suggesting that loss of repressive histone marks through the Smad–KDM6B pathway explains the transcriptional regulation especially at later time points.

In addition to sequence-specific DNA-binding transcription factors, histone code reader proteins, which are recruited and bound to specific histone modifications, are reported to help to determine the binding sites of Smad proteins. Massagué and colleagues have reported that tripartite motif 33 (TRIM33, also known as TIF1γ or Ectodermin), physically interacts with Smad2 and Smad3, to make a TRIM33–Smad2/3 complex without Smad4.[69] The TRIM33 contains an N-terminal RING finger/B-box/coiled coil (RBCC) or TRIM domain, and a plant homeodomain (PHD) zinc finger and a Bromo domain in the C-terminus. They reported that the PHD-Bromo cassette recognized histone H3 lysine-9 trimethylation (H3K9me3) and H3 acetylation especially at lysine residues 18 and 23 (H3K18ac and H3K23ac). During mESC differentiation, nodal signaling triggered TRIM33–Smad2/3 complex formation. The TRIM33–Smad2/3 complex recognizes and binds to H3K9me3-K18ac dual histone marks and displaces the chromatin-compacting factor heterochromatin protein 1γ (HP1γ) in the *GSC* and *MIXL1* promoters, resulting in the remodeling of the local chromatin structure (Figure 3b).[70] Agricola *et al.*[71] also found that TRIM33 recognizes and binds to H3K18ac/K23ac. On the other hand, TRIM33 has been reported to bind Smad4 and function as a RING-type ubiquitin ligase for Smad4.[72] Consistent with this model, Agricola *et al.*[71] reported that TRIM33 inhibits Smad4 function through ubiquitin-mediated degradation of Smad4, and that its E3 ubiquitin ligase activity is induced after binding to histones. The detailed mechanisms have not been settled, but TRIM33 recognizes a specific histone code and modulates TGF-β/BMP signaling. Since the relationship between Smad proteins and histone modification marks has not been fully

elucidated on a genome-wide scale, future analyses will address a possible mechanistic link between Smad proteins and epigenetic marks using ChIP-chip/ChIP-seq approach.

SMAD BINDING AND GENE REGULATION

Previous studies have indicated that binding of transcription factors detected by ChIP-chip/ChIP-seq experiments are not necessarily associated with transcriptional regulation of nearby genes (reviewed in Farnham[73]). It has frequently been observed that changing the level of a DNA-binding transcription factor alters the expression level of only 1–10% of its potential target genes. Most of the recent studies have confirmed that 1–20% of Smad-binding sites are associated with the regulation of expression of nearby genes. This discrepancy is in part due to the fact that mRNA levels do not only reflect transcriptional activities, since mRNA levels are also regulated by other biological processes, for example, degradation. Another explanation for the discrepancy is related to the definition of target genes. Although most studies assign binding sites to the nearest gene within 50 kb, this is not always the case. For example, Trompouki *et al.* revealed that several transcription factors, including Smad1, cooperatively regulate the expression of the hematopoietic gene *LMO2* through binding to the known enhancer region at 72 kb upstream of the transcription start site in K562 cells.[19,74] We also observed that Smad1/5 bound to a region 57 kb upstream of the transcription start site of *Smad6* in ECs, as well as the *LMO2* − 72 kb enhancer.[57] This region has been reported to be associated with Smad6 expression in the heart, vasculature and hematopoietic organs,[75] suggesting that the binding to this region, as well as the promoter region, plays an important role in these cell types. Recently, methods that characterize the chromatin architecture have been developed. Chromosome conformation capture (3C) assays make
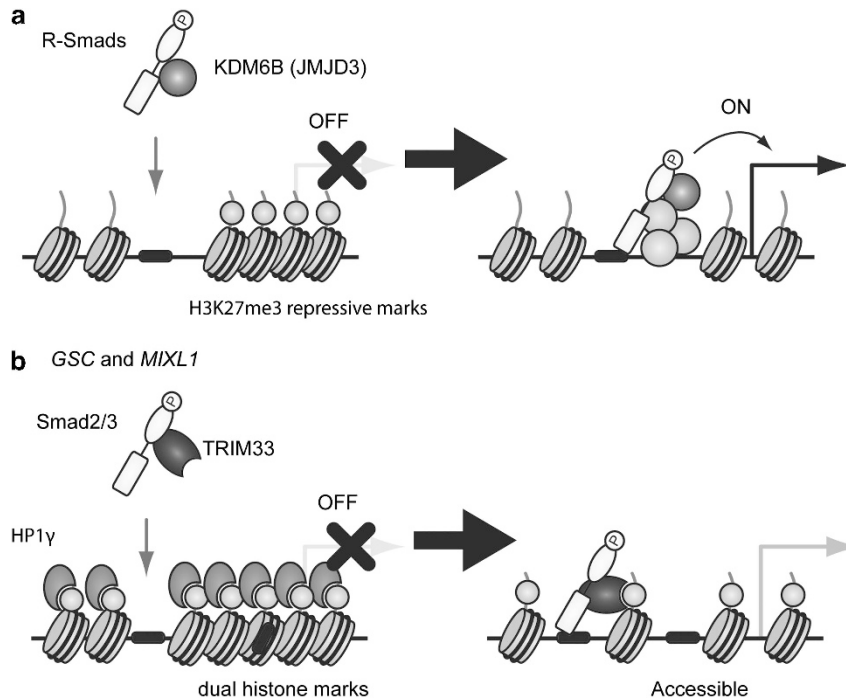
Figure 3. Smad proteins and histone modification marks. Smad proteins have been reported to induce local chromatin remodeling and modification at their binding sites. Several models are described in ES cells, where early developmental genes are poised and ready to be activated in response to extracellular signals, such as nodal. (**a**) R-Smads physically interact with a histone demethylase, KDM6B (also known as JMJD3), and recruit it to their target sites, followed by the loss of the H3K27me3 repressive mark (light green). (**b**) Xi et al.[70] reported that nodal signaling triggered TRIM33–Smad2/3 complex formation. The TRIM33–Smad2/3 complex recognizes and binds to H3K9me3-K18ac dual histone marks (light blue) and displaces the chromatin-compacting factor HP1γ (heterochromatin protein 1γ) in the GSC and MIXL1 promoters, resulting in the remodeling of the local chromatin structure to make Smad-binding region(s) (red) accessible. A full colour version of this figure is available at the Oncogene journal online.

it possible to study long-distance regulation of genes by enhancers through formation of chromatin loops (reviewed in Simonis et al.[76]). Application of these technologies will help to identify the functional relationship between Smad-binding sites and genes implicated in cancer progression.

It is also possible that for many sites, binding of Smads is not sufficient for transcriptional regulation, but additional stimuli are required to drive the expression of the target genes. For example, costimulation with tumor necrosis factor-α, which induces the transcriptional repressor ATF-3, affects the expression regulation of the Id1 gene and cellular response.[34,35] Sometimes, ligand stimulation itself induces these cofactors and makes a feed-forward circuit, like in myotube differentiation. The myogenic transcription factor MyoD directly regulates genes expressed during skeletal muscle differentiation together with other transcription factors such as MEF2[77] and Zfp238 (also known as RP58).[78] These transcription factors are also induced by MyoD, and MEF2 functions with MyoD in a positive feed-forward circuit,[77] while Zfp238 participates in a negative feed-forward circuit.[78] Comparison of MyoD-binding patterns of mouse C2C12 myoblasts and differentiated myotubes has revealed that most binding events in myoblasts are not directly associated with gene regulation. However, MyoD binding increases during myogenic differentiation at many of the regulatory regions associated with genes expressed in skeletal muscle. Intriguingly, the myotube-increased binding sites are enriched for MEF2-like motifs, while the myotube-decreased peaks are enriched for Zfp238-like motifs,[59] consistent with the fact that MEF2 positively and Zfp238 negatively cooperate with MyoD. It is possible that TGF-β stimulation induces certain transcription factors, which take part in feed-forward regulatory loops and cooperatively regulate gene expression especially at late time points.

IDENTIFICATION OF A TGF-β GENE SIGNATURE

The notion of 'gene signature' comes from the early work on cancer classification and prognosis prediction using genome-wide gene expression profiles obtained from microarray analyses of cancer patients.[79] Identification of a group of genes that reflect the activity of a common function, pathway or other property in a specific context, are sometimes more revealing compared with the analysis of single genes. Gene expression signatures obtained in experimental conditions has proved to subcategorize patients and predict their prognosis. Concerning TGF-β, Coulouarn et al.[80] reported that TGF-β-responsive genes at late time points, or a late TGF-β signature, which were identified in mouse primary hepatocytes, successfully discriminate distinct subgroups of hepatocellular carcinoma and possess a predictive value for hepatocellular carcinoma patients.

Combination of ChIP-chip/ChIP-seq and genome-wide transcriptome analyses provides an accurate prediction of target genes of Smad proteins. TGF-β family members regulate a variety of target genes both directly and indirectly, and modulate many biological processes. The chromatin-binding landscape of Smad proteins, obtained by ChIP-chip/ChIP-seq, will help to identify specific genes that are directly regulated by Smad proteins. It will also help to dissect a specific cellular program regulated by TGF-β family members, for example, the growth inhibitory and apoptosis programs of TGF-β. So far, many groups have identified groups of direct TGF-β target genes by using this strategy. Importantly, the TGF-β/Smad4 target gene signature identified in an ovarian cancer cell line predicts patient survival, based on in silico mining of publically available patient data bases.[21] Since TGF-β functions as a tumor suppressor in low-grade carcinoma cells, while it promotes metastasis in advanced carcinoma cells, a direct

comparison of the Smad-binding sites of these two stages of tumorigenesis, obtained from experimental models or from cancer patients, may reveal specific gene signatures of TGF-β correlating to its tumor suppressive and tumor-promoting roles, respectively. This may provide us more novel predictive indicators and biomarkers for TGF-β targeting treatments.

CONCLUSIONS AND PERSPECTIVES

The signaling pathways of TGF-β family members are key players in tumorigenesis and cancer progression. TGF-β can function both as a tumor-suppressing and a tumor-promoting factor during cancer progression. BMP signaling has been reported to play critical roles in oncogene-induced senescence, which is part of the tumorigenesis barrier and blocks cellular proliferation by inducing irreversible growth arrest.[66] Interestingly, BMP signaling induces differentiation of certain cancer-initiating cells, such as glioma-initiating cells,[81] while TGF-β/activin signaling maintains their stem cell-like properties.[9,10] Since Smad proteins are central mediators of the signal transduction, studies on global and genome-wide binding sites of Smad proteins may reveal important insights into their complex biological functions.

Identification of an appropriate antibody is the first and most important step for ChIP-chip and ChIP-seq analyses, because the quality of ChIP data depends crucially on the quality of the antibody used.[16] Since MH1 and MH2 domains are conserved among R-Smads, several specific antibodies for Smad proteins recognize their linker region. However, linker regions are targets of posttranslational modification and protein interactions, as discussed above. It is possible that such changes may attenuate the affinities of antibodies under specific conditions. Although ChIP-grade antibodies for Smad proteins have been established (Supplementary Table 2), careful interpretation of the results will be required.

In summary, genome-wide analysis of the binding sites of Smad proteins have led to important discoveries of their cell-type-specific and context-dependent functions. Application of genome-wide techniques to experimental models and human samples derived from cancer patients, will help to clarify their complex mechanisms during cancer progression, and may also provide potential prognostic biomarkers for future cancer therapy.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1 Massague J. TGFbeta in cancer. *Cell* 2008; **134**: 215–230.
2 Ikushima H, Miyazono K. TGFbeta signalling: a complex web in cancer progression. *Nat Rev Cancer* 2010; **10**: 415–424.
3 Feng XH, Derynck R. Specificity and versatility in TGF-beta signaling through Smads. *Annu Rev Cell Dev Biol* 2005; **21**: 659–693.
4 Levy L, Hill CS. Alterations in components of the TGF-beta superfamily signaling pathways in human cancer. *Cytokine Growth Factor Rev* 2006; **17**: 41–58.
5 Yoshimura A, Wakabayashi Y, Mori T. Cellular and molecular basis for the regulation of inflammation by TGF-beta. *J Biochem* 2010; **147**: 781–792.
6 Roberts AB, Wakefield LM. The two faces of transforming growth factor beta in carcinogenesis. *Proc Natl Acad Sci USA* 2003; **100**: 8621–8623.
7 Moustakas A, Heldin CH. Non-Smad TGF-beta signals. *J Cell Sci* 2005; **118**: 3573–3584.
8 Miyazono K, Ehata S, Koinuma D. Tumor-promoting functions of transforming growth factor-beta in progression of cancer. *Ups J Med Sci* 2012; **117**: 143–152.
9 Ikushima H, Todo T, Ino Y, Takahashi M, Miyazawa K, Miyazono K. Autocrine TGF-beta signaling maintains tumorigenicity of glioma-initiating cells through Sry-related HMG-box factors. *Cell Stem Cell* 2009; **5**: 504–514.
10 Penuelas S, Anido J, Prieto-Sanchez RM, Folch G, Barba I, Cuartas I et al. TGF-beta increases glioma-initiating cell self-renewal through the induction of LIF in human glioblastoma. *Cancer Cell* 2009; **15**: 315–327.
11 Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 2008; **133**: 704–715.
12 Lonardo E, Hermann PC, Mueller MT, Huber S, Balic A, Miranda-Lorenzo I et al. Nodal/Activin signaling drives self-renewal and tumorigenicity of pancreatic cancer stem cells and provides a target for combined drug therapy. *Cell Stem Cell* 2011; **9**: 433–446.
13 Naka K, Hoshii T, Muraguchi T, Tadokoro Y, Ooshio T, Kondo Y et al. TGF-beta-FOXO signalling maintains leukaemia-initiating cells in chronic myeloid leukaemia. *Nature* 2010; **463**: 676–680.
14 Heldin CH, Miyazono K, ten Dijke P. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature* 1997; **390**: 465–471.
15 Shi Y, Massague J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* 2003; **113**: 685–700.
16 Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 2009; **10**: 669–680.
17 Pepke S, Wold B, Mortazavi A. Computation for ChIP-seq and RNA-seq studies. *Nat Methods* 2009; **6**: S22–S32.
18 Mullen AC, Orlando DA, Newman JJ, Loven J, Kumar RM, Bilodeau S et al. Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell* 2011; **147**: 565–576.
19 Trompouki E, Bowman TV, Lawton LN, Fan ZP, Wu DC, DiBiase A et al. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell* 2011; **147**: 577–589.
20 Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 2008; **133**: 1106–1117.
21 Kennedy BA, Deatherage DE, Gu F, Tang B, Chan MW, Nephew KP et al. ChIP-seq defined genome-wide map of TGFbeta/SMAD4 targets: implications with clinical outcome of ovarian cancer. *PLoS One* 2011; **6**: e22606.
22 Miyazono K, Kamiya Y, Morikawa M. Bone morphogenetic protein receptors and signal transduction. *J Biochem* 2010; **147**: 35–51.
23 Zhang Y, Feng XH, Smad3 Derynck R. and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-beta-induced transcription. *Nature* 1998; **394**: 909–913.
24 Koinuma D, Tsutsumi S, Kamimura N, Taniguchi H, Miyazawa K, Sunamura M et al. Chromatin immunoprecipitation on microarray analysis of Smad2/3 binding sites reveals roles of ETS1 and TFAP2A in transforming growth factor beta signaling. *Mol Cell Biol* 2009; **29**: 172–186.
25 Koinuma D, Tsutsumi S, Kamimura N, Imamura T, Aburatani H, Miyazono K. Promoter-wide analysis of Smad4 binding sites in human epithelial cells. *Cancer Sci* 2009; **100**: 2133–2142.
26 Yoon SJ, Wills AE, Chuong E, Gupta R, HEB Baker JC. and E2A function as SMAD/FOXH1 cofactors. *Genes Dev* 2011; **25**: 1654–1661.
27 Ikushima H, Komuro A, Isogaya K, Shinozaki M, Hellman U, Miyazawa K et al. An Id-like molecule, HHM, is a synexpression group-restricted regulator of TGF-beta signalling. *EMBO J* 2008; **27**: 2955–2965.
28 Gomis RR, Alarcon C, Nadal C, Van Poznak C, Massague J. C/EBPbeta at the core of the TGFbeta cytostatic response and its evasion in metastatic breast cancer cells. *Cancer Cell* 2006; **10**: 203–214.
29 Silvestri C, Narimatsu M, von Both I, Liu Y, Tan NB, Izzi L et al. Genome-wide identification of Smad/Foxh1 targets reveals a role for Foxh1 in retinoic acid regulation and forebrain development. *Dev Cell* 2008; **14**: 411–423.
30 Kim SW, Yoon SJ, Chuong E, Oyolu C, Wills AE, Gupta R et al. Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev Biol* 2011; **357**: 492–504.
31 Gomis RR, Alarcon C, He W, Wang Q, Seoane J, Lash A et al. A FoxO-Smad synexpression group in human keratinocytes. *Proc Natl Acad Sci USA* 2006; **103**: 12747–12752.
32 Janknecht R, Wells NJ, Hunter T. TGF-beta-stimulated cooperation of smad proteins with the coactivators CBP/p300. *Genes Dev* 1998; **12**: 2114–2119.
33 Feng XH, Zhang Y, Wu RY, Derynck R. The tumor suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are coactivators for smad3 in TGF-beta-induced transcriptional activation. *Genes Dev* 1998; **12**: 2153–2163.

34 Kang Y, Chen CR, Massague J. A self-enabling TGFbeta response coupled to stress signaling: Smad engages stress response factor ATF3 for Id1 repression in epithelial cells. *Mol Cell* 2003; **11**: 915–926.

35 Anido J, Saez-Borderias A, Gonzalez-Junca A, Rodon L, Folch G, Carmona MA et al. TGF-beta receptor inhibitors target the CD44(high)/Id1(high) glioma-initiating cell population in human glioblastoma. *Cancer Cell* 2010; **18**: 655–668.

36 Alarcon C, Zaromytidou AI, Xi Q, Gao S, Yu J, Fujisawa S et al. Nuclear CDKs drive Smad transcriptional activation and turnover in BMP and TGF-beta pathways. *Cell* 2009; **139**: 757–769.

37 Sapkota G, Alarcon C, Spagnoli FM, Brivanlou AH, Massague J. Balancing BMP signaling through integrated inputs into the Smad1 linker. *Mol Cell* 2007; **25**: 441–454.

38 Kretzschmar M, Doody J, Massague J. Opposing BMP and EGF signalling pathways converge on the TGF-beta family mediator Smad1. *Nature* 1997; **389**: 618–622.

39 Aubin J, Davy A, Soriano P. *In vivo* convergence of BMP and MAPK signaling pathways: impact of differential Smad1 phosphorylation on development and homeostasis. *Genes Dev* 2004; **18**: 1482–1494.

40 Fuentealba LC, Eivers E, Ikeda A, Hurtado C, Kuroda H, Pera EM et al. Integrating patterning signals: Wnt/GSK3 regulates the duration of the BMP/Smad1 signal. *Cell* 2007; **131**: 980–993.

41 Aragon E, Goerner N, Zaromytidou AI, Xi Q, Escobedo A, Massague J et al. A Smad action turnover switch operated by WW domain readers of a phosphoserine code. *Genes Dev* 2011; **25**: 1275–1288.

42 Zhu H, Kavsak P, Abdollah S, Wrana JL, Thomsen GHA. SMAD ubiquitin ligase targets the BMP pathway and affects embryonic pattern formation. *Nature* 1999; **400**: 687–693.

43 Kuratomi G, Komuro A, Goto K, Shinozaki M, Miyazawa K, Miyazono K et al. NEDD4-2 (neural precursor cell expressed, developmentally down-regulated 4-2) negatively regulates TGF-beta (transforming growth factor-beta) signalling by inducing ubiquitin-mediated degradation of Smad2 and TGF-beta type I receptor. *Biochem J* 2005; **386**: 461–470.

44 Gao S, Alarcon C, Sapkota G, Rahman S, Chen PY, Goerner N et al. Ubiquitin ligase Nedd4L targets activated Smad2/3 to limit TGF-beta signaling. *Mol Cell* 2009; **36**: 457–468.

45 Dupont S, Mamidi A, Cordenonsi M, Montagner M, Zacchigna L, Adorno M et al. FAM/USP9x, a deubiquitinating enzyme essential for TGFbeta signaling, controls Smad4 monoubiquitination. *Cell* 2009; **136**: 123–135.

46 Inui M, Manfrin A, Mamidi A, Martello G, Morsut L, Soligo S et al. USP15 is a deubiquitylating enzyme for receptor-activated SMADs. *Nat Cell Biol* 2011; **13**: 1368–1375.

47 Lonn P, van der Heide LP, Dahl M, Hellman U, Heldin CH, Moustakas A. PARP-1 attenuates Smad-mediated transcription. *Mol Cell* 2010; **40**: 521–532.

48 Huang D, Wang Y, Wang L, Zhang F, Deng S, Wang R et al. Poly(ADP-ribose) polymerase 1 is indispensable for transforming growth factor-beta Induced Smad3 activation in vascular smooth muscle cell. *PLoS One* 2011; **6**: e27123.

49 Zawel L, Dai JL, Buckhaults P, Zhou S, Kinzler KW, Vogelstein B et al. Human Smad3 and Smad4 are sequence-specific transcription activators. *Mol Cell* 1998; **1**: 611–617.

50 Yagi K, Goto D, Hamamoto T, Takenoshita S, Kato M, Miyazono K. Alternatively spliced variant of Smad2 lacking exon 3. Comparison with wild-type Smad2 and Smad3. *J Biol Chem* 1999; **274**: 703–709.

51 Chai J, Wu JW, Yan N, Massague J, Pavletich NP, Shi Y. Features of a Smad3 MH1-DNA complex. Roles of water and zinc in DNA binding. *J Biol Chem* 2003; **278**: 20327–20331.

52 Shi Y, Wang YF, Jayaraman L, Yang H, Massague J, Pavletich NP. Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling. *Cell* 1998; **94**: 585–594.

53 BabuRajendran N, Palasingam P, Narasimhan K, Sun W, Prabhakar S, Jauch R et al. Structure of Smad1 MH1/DNA complex reveals distinctive rearrangements of BMP and TGF-beta effectors. *Nucleic Acids Res* 2010; **38**: 3477–3488.

54 Lee KL, Lim SK, Orlov YL, Yit le Y, Yang H, Ang LT et al. Graded Nodal/Activin signaling titrates conversion of quantitative phospho-Smad2 levels into qualitative embryonic stem cell fate decisions. *PLoS Genet* 2011; **7**: e1002130.

55 Zhang Y, Handley D, Kaplan T, Yu H, Bais AS, Richards T et al. High throughput determination of TGFbeta1/SMAD3 targets in A549 lung epithelial cells. *PLoS One* 2011; **6**: e20319.

56 Kim J, Johnson K, Chen HJ, Carroll S, Laughon A. Drosophila Mad binds to DNA and directly mediates activation of vestigial by Decapentaplegic. *Nature* 1997; **388**: 304–308.

57 Morikawa M, Koinuma D, Tsutsumi S, Vasilaki E, Kanki Y, Heldin CH et al. ChIP-seq reveals cell type-specific binding patterns of BMP-specific Smads and a novel binding motif. *Nucleic Acids Res* 2011; **39**: 8712–8727.

58 Mizutani A, Koinuma D, Tsutsumi S, Kamimura N, Morikawa M, Suzuki HI et al. Cell type-specific target selection by combinatorial binding of Smad2/3 proteins and hepatocyte nuclear factor 4alpha in HepG2 cells. *J Biol Chem* 2011; **286**: 29848–29860.

59 Cao Y, Yao Z, Sarkar D, Lawrence M, Sanchez GJ, Parker MH et al. Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev Cell* 2010; **18**: 662–674.

60 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010; **38**: 576–589.

61 John S, Sabo PJ, Thurman RE, Sung MH, Biddie SC, Johnson TA et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* 2011; **43**: 264–268.

62 Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* 2011; **25**: 2227–2241.

63 Harvey SA, Smith JC. Visualisation and quantification of morphogen gradient formation in the zebrafish. *PLoS Biol* 2009; **7**: e1000101.

64 Green JB, New HV, Smith JC. Responses of embryonic Xenopus cells to activin and FGF are separated by multiple dose thresholds and correspond to distinct axes of the mesoderm. *Cell* 1992; **71**: 731–739.

65 Fei T, Xia K, Li Z, Zhou B, Zhu S, Chen H et al. Genome-wide mapping of SMAD target genes reveals the role of BMP signaling in embryonic stem cell fate determination. *Genome Res* 2010; **20**: 36–44.

66 Kaneda A, Fujita T, Anai M, Yamamoto S, Nagae G, Morikawa M et al. Activation of Bmp2-Smad1 signal and its regulation by coordinated alteration of H3K27 trimethylation in Ras-induced senescence. *PLoS Genet* 2011; **7**: e1002359.

67 Akizu N, Estaras C, Guerrero L, Marti E, Martinez-Balbas MA. H3K27me3 regulates BMP activity in developing spinal cord. *Development* 2010; **137**: 2915–2925.

68 Dahle O, Kumar A, Kuehn MR. Nodal signaling recruits the histone demethylase Jmjd3 to counteract polycomb-mediated repression at target genes. *Sci Signal* 2010; **3**: ra48.

69 He W, Dorn DC, Erdjument-Bromage H, Tempst P, Moore MA, Massague J. Hematopoiesis controlled by distinct TIF1gamma and Smad4 branches of the TGFbeta pathway. *Cell* 2006; **125**: 929–941.

70 Xi Q, Wang Z, Zaromytidou AI, Zhang XH, Chow-Tsang LF, Liu JX et al. A poised chromatin platform for TGF-beta access to master regulators. *Cell* 2011; **147**: 1511–1524.

71 Agricola E, Randall RA, Gaarenstroom T, Dupont S, Hill CS. Recruitment of TIF1gamma to chromatin via its PHD finger-bromodomain activates its ubiquitin ligase and transcriptional repressor activities. *Mol Cell* 2011; **43**: 85–96.

72 Dupont S, Zacchigna L, Cordenonsi M, Soligo S, Adorno M, Rugge M et al. Germ-layer specification and control of cell growth by Ectodermin, a Smad4 ubiquitin ligase. *Cell* 2005; **121**: 87–99.

73 Farnham PJ. Insights from genomic profiling of transcription factors. *Nat Rev Genet* 2009; **10**: 605–616.

74 Landry JR, Bonadies N, Kinston S, Knezevic K, Wilson NK, Oram SH et al. Expression of the leukemia oncogene Lmo2 is controlled by an array of tissue-specific elements dispersed over 100 kb and bound by Tal1/Lmo2, Ets, and Gata factors. *Blood* 2009; **113**: 5783–5792.

75 Pimanda JE, Donaldson IJ, de Bruijn MF, Kinston S, Knezevic K, Huckle L et al. The SCL transcriptional network and BMP signaling pathway interact to regulate RUNX1 activity. *Proc Natl Acad Sci USA* 2007; **104**: 840–845.

76 Simonis M, Kooren J, de Laat W. An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* 2007; **4**: 895–901.

77 Penn BH, Bergstrom DA, Dilworth FJ, Bengal E, Tapscott SJ. A MyoD-generated feed-forward circuit temporally patterns gene expression during skeletal muscle differentiation. *Genes Dev* 2004; **18**: 2348–2353.

78 Yokoyama S, Ito Y, Ueno-Kudoh H, Shimizu H, Uchibe K, Albini S et al. A systems approach reveals that the myogenesis genome network is regulated by the transcriptional repressor RP58. *Dev Cell* 2009; **17**: 836–848.

79 van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002; **347**: 1999–2009.

80 Coulouarn C, Factor VM, Thorgeirsson SS. Transforming growth factor-beta gene expression signature in mouse hepatocytes predicts clinical outcome in human cancer. *Hepatology* 2008; **47**: 2059–2067.

81 Piccirillo SG, Reynolds BA, Zanetti N, Lamorte G, Binda E, Broggi G et al. Bone morphogenetic proteins inhibit the tumorigenic potential of human brain tumour-initiating cells. *Nature* 2006; **444**: 761–765.

Supplementary Information accompanies the paper on the Oncogene website (http://www.nature.com/onc)