as a negative control region (Fig. 1*A*, *lower panel*). We confirmed significant enrichment of Smad2/3 binding to the *SERPINE1*, *SERPINA1*/α-antitrypsin, and *APOA1* loci using ChIP-qPCR (Fig. 1*B*). We identified 3,636 significant Smad2/3 binding regions that had detection p values of less than 10^{-4} within the promoter regions of known genes.

Next, we compared the identified Smad2/3 binding regions in HepG2 cells to those in HaCaT cells (11). We also obtained 1,270 Smad2/3 binding regions in Hep3B hepatocellular carcinoma cells to determine their overlaps (Fig. 2*A*). We found that only 25.2% of the Smad2/3 binding regions in HaCaT cells (n = 682) were shared with those in HepG2 cells. In contrast, 58.3% of the binding regions in Hep3B cells overlapped with those identified in HepG2 cells, although the number of overlapping binding regions (n = 741) was similar to that between HaCaT and HepG2 cells. Many of the Smad2/3 binding regions were thus unique to each cell type. We determined the candidate target genes of Smad2/3 using the dataset of Smad2/3 binding regions that were either common to HepG2 and HaCaT, HepG2-specific, or HaCaT-specific (supplemental Tables S3-S5), and performed gene ontology analysis of each category by DAVID (27). We did not observe remarkable differences in the top five enriched annotation clusters between the common Smad2/3 binding regions and HepG2-specific Smad2/3 binding region. Conversely, enrichment of cell death and cytoskeleton-related annotations was found in HaCaT-specific binding regions (supplemental Table S6).

To identify specific motifs in the Smad2/3 binding regions in HepG2 cells, *de novo* motif prediction was performed using the CisGenome Gibbs motif sampler (supplemental Fig. S1*A*). We searched for known motifs that had similarity to the calculated motifs using the JASPAR data base (28). As shown in Fig. 2*C*, we found that one predicted motif (Fig. 2*B*) was strongly similar to the HNF4α-binding motif (Fig. 2*C*, 96.9% score). The frequency of the HNF4α motif in HepG2-specific Smad2/3 binding regions was 41.2%, although that in HaCaT-specific Smad2/3 binding regions and its matched random genomic regions was 23.7 and 17.6%, respectively (Fig. 2*D*). We also analyzed the sequences in Smad2/3 binding regions using the CEAS analysis tool as we did in our previous report (11, 12, 25), and the HNF4α-binding motif was identified as one of the top three enriched motifs in the binding sequences (supplemental Fig. S1*B*) (25). It should be noted that canonical Smad-binding element (M00974.SMAD, "CAGAC") was also identified as enriched motif through CEAS analysis and present in 40.6% of the HepG2-specific Smad2/3 binding regions (data not shown). These findings suggested that the HNF4α motif was enriched in HepG2-specific Smad2/3 binding regions and had roles for cell type specificity of Smad2/3 binding and TGF-β-induced transcription in HepG2 cells. HNF4α is one of the "master genes" of hepatocytes and is essential for hepatocyte-specific gene expressions and functions. Because HNF4α was not expressed in HaCaT cells (Fig. 2*E*), we decided to determine HNF4α binding regions *in vivo* in the presence of TGF-β.

HNF4α Binding to Its Binding Regions Is Not Extensively Altered by TGF-β—ChIP-chip and ChIP-seq studies of HNF4α binding regions have been reported using HepG2 cells (29–31). We retrieved data (30) from the data base and found that 20.7% of the Smad2/3 binding regions were common to HNF4α binding regions *in vivo*. However, no reports have yet determined the changes in the binding of HNF4α by extracellular stimulations, including that by TGF-β. We therefore acquired the HNF4α binding data using the newly available ChIP-seq technology to compare the Smad2/3 binding with the HNF4α binding under TGF-β stimulation. We identified 25,105 significant HNF4α binding regions in the absence of TGF-β and 23,368 regions in the presence of TGF-β, at an FDR of less than 0.1%. The *APOA4/APOC3/APOA1* gene cluster that is a known target of HNF4α is shown in Fig. 3*A*. We observed significant HNF4α binding to several of these regions in the absence of TGF-β, and the binding was not extensively changed following stimulation. We also found that there was significant Smad2/3 binding to the *APOA1* promoter, which was absent in HaCaT cells. Our data showed that Smad2/3 binding to the *APOC3* promoter was not significant.

We then examined the changes in HNF4α binding by TGF-β stimulation. More than 80% of the HNF4α binding regions overlapped between TGF-β-treated and untreated cells. However, there were also specific binding regions for both TGF-β-treated and untreated cells (Fig. 3*B*). In addition, we calculated the changes in the normalized read numbers within the HNF4α binding regions by TGF-β stimulation and found that some regions indeed had either decreased or increased sequence reads following TGF-β stimulation (data not shown). Percent input values of HNF4α binding to the *APOC3* and *APOA1* loci were also up-regulated to some extent (Fig. 3*C*), suggesting that there were some, if limited, roles of TGF-β for HNF4α binding. Using the HNF4α binding data with TGF-β stimulation, we determined the frequency of HNF4α binding in Smad2/3 binding regions *in vivo*. We found that 32.5% of the Smad2/3 binding regions in HepG2 cells were indeed common to HNF4α binding regions. In contrast, only 13.2% of Smad2/3 binding regions in HaCaT cells were common, and the frequency decreased to 7.7% when HaCaT-specific Smad2/3 binding regions were examined (Fig. 3*D*). These results suggested that HNF4α and Smad2/3 binding regions are located in close prox-

FIGURE 3. **Identification of HNF4α binding regions in the presence and absence of TGF-β stimulation.** *A*, graphical representation of HNF4α binding to the *APOA4/APOC3/APOA1* gene loci. Sequence read numbers of 100-bp sliding window were plotted for HNF4α ChIP-seq samples. Smad2/3 bindings as determined by ChIP-chip analysis were shown in the *upper two panels* as in Fig. 1*A*. *Black bars* represent significant binding regions (FDR, <0.1%). *B*, Venn diagrams showing overlap between TGF-β-treated and untreated HNF4α binding regions. HNF4α binding regions were determined for each sample (FDR, <0.1%). HNF4α binding regions that have overlapping regions within 500 bp from their positions of maximum read numbers were considered as shared binding regions. *C*, changes in the HNF4α binding to *APOC3* and *APOA1* loci were quantitatively determined by ChIP-qPCR analysis. *Error bars*, S.D. *D*, frequencies of *in vivo* HNF4α binding to the Smad2/3 binding regions. Percentages of HNF4α binding within 250 bp from the peak signal position of Smad2/3 binding regions were calculated for the indicated Smad2/3 binding groups. *E*, frequencies of canonical Smad-binding elements in HNF4α binding regions compared with Smad binding regions in HepG2 cells. A Smad-binding element, M00974.SMAD that was identified as an enriched motif in HepG2-specific Smad2/3 binding regions using CEAS (see text), was selected for calculation. CisGenome was used for mapping of the motif. Presence of the motif for each HNF4α binding region and Smad2/3 binding region was determined using PerlScript.

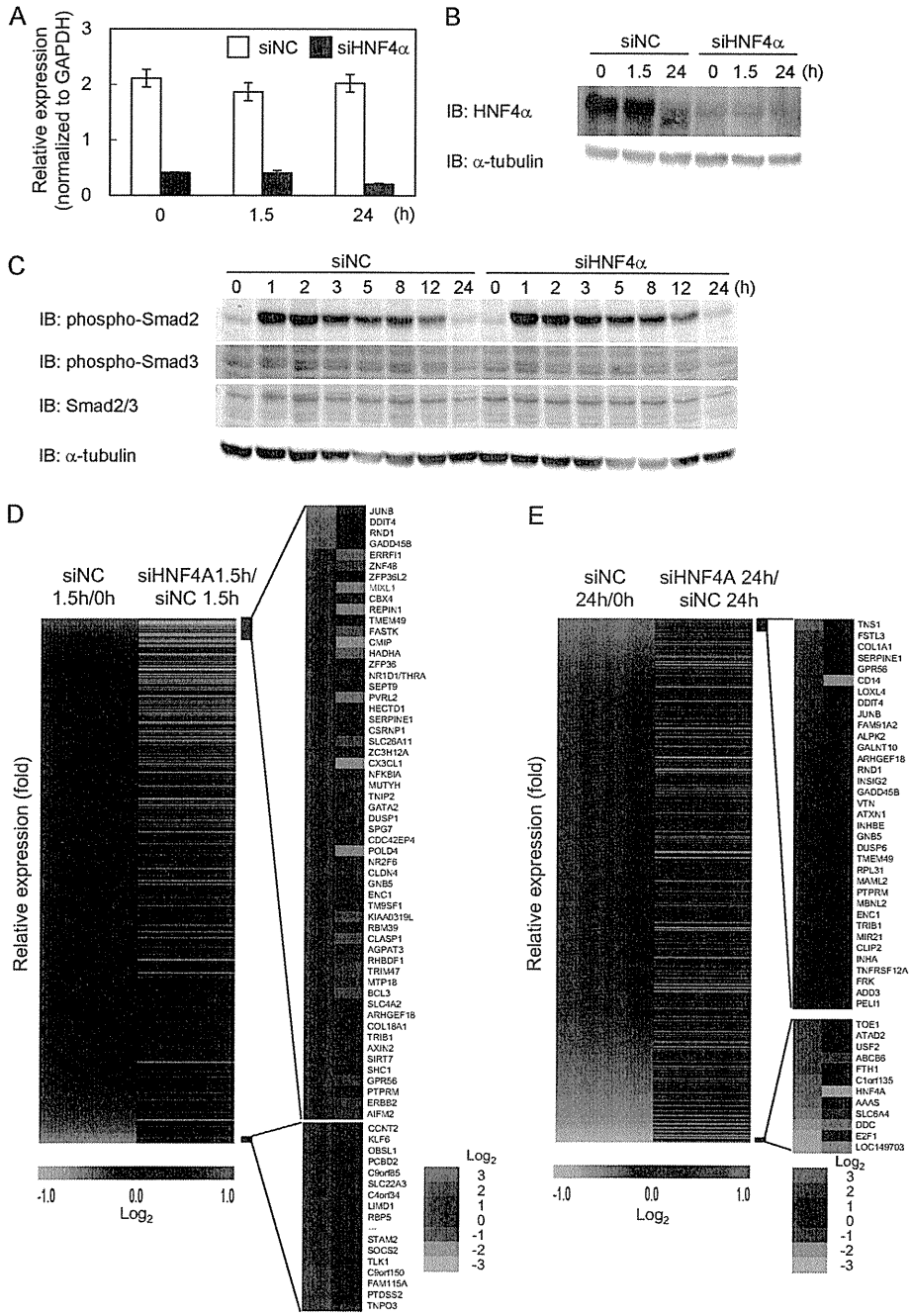

FIGURE 4. **Effect of knockdown of HNF4α on TGF-β-induced gene expression in HepG2 cells.** *A,* confirmation of HNF4α-knocked down samples for microarray analysis. HepG2 cells were transfected with HNF4α siRNA and treated with 120 pM TGF-β for the indicated times and harvested. Expression of HNF4α was determined by RT-qPCR and normalized by glyceraldehyde-3-phosphate dehydrogenase (GAPDH). *siNC,* negative control siRNA. *B,* down-regulation of HNF4α protein expression by siRNA. *C,* phosphorylation levels of Smad2/3 by using HNF4α siRNA. The *top two panels* show phosphorylation of Smad2 and Smad3. The *3rd panel* indicates the expression of total Smad2/3, and the *bottom panel* is a loading control. *IB,* immunoblot. *D,* heat map of the TGF-β-induced expression of target genes of Smad2/3 and HNF4α and effect of HNF4α siRNA. Target genes that have overlapping binding regions of Smad2/3 and HNF4α were sorted by their induction of probe signal values by TGF-β stimulation for 1.5 h and are represented by *color bars* in the *1st column,* using the TM4 microarray software (60). Relative expression of these genes in HNF4α siRNA samples to the control siRNA is shown in the *2nd column.* In addition, a list of genes whose expressions changed more than 1.5-fold is shown in the *right panel* with their expression changes. *E,* heat map of target genes of TGF-β with Smad2/3 binding regions common to HNF4α at 24 h after TGF-β stimulation are shown as in *D.* Genes whose expressions were changed more than 2-fold are shown in the *right panel.*

imity to each other in HepG2 cells, although we could not determine whether HNF4α- and Smad2/3-binding "elements" overlapped within the binding regions because of the limited resolution of ChIP-chip- and ChIP-seq-based assays. We then calculated the frequency of Smad-binding element CAGAC in HNF4α binding regions. 40.4% of the binding regions common to HNF4α and Smad2/3 had Smad-binding elements, compared with 24.3% in the total HNF4α binding regions (Fig. 3*E*).

TABLE 1

TGF-β-induced changes in gene expression in relation to Smad2/3 binding

Expression array data transfected with control siRNA and stimulated with TGF-β were compared with Smad2/3 ChIP-chip data. A total of 8,653 genes that had values of more than 100 at least at one time point for one of their probes (n = 13,720) was used for the analysis. Up-regulated or down-regulated genes were determined compared with 0-h values. The positions of peak signals of Smad binding regions (SBRs) relative to the nearby RefSeq genes were first determined, and regions within 5 kb upstream from the transcription start site and the first intron were selected. *a indicates number of genes analyzed by microarray. *b indicates number of genes which have Smad2/3 binding regions.

		All genes		Genes with SBRs		
		*a	%	*b	%	*b/*a (%)
Total		8653	100.0	1941	100.0	22.4
Increase						
>2-Fold	1.5 h	25	0.3	14	0.7	56.0
	24 h	223	2.6	59	3.0	26.5
>1.5-Fold	1.5 h	273	3.2	89	4.6	32.6
	24 h	837	9.7	250	12.9	29.9
Decrease						
>2-Fold	1.5 h	16	0.2	2	0.1	12.5
	24 h	174	2.0	25	1.3	14.4
>1.5-Fold	1.5 h	217	2.5	47	2.4	21.7
	24 h	877	10.1	160	8.2	18.2

Effect of HNF4α on the Expression of Smad2/3 Target Genes—To elucidate the effect of HNF4α on TGF-β-induced transcriptional regulation, we knocked down HNF4α by using siRNA (Fig. 4, *A* and *B*). The phosphorylation of Smad2 and Smad3 was not affected by the siRNA under the applied condition (Fig. 4*C*). We obtained expression microarray data and calculated the changes in the expression of genes with binding regions shared by Smad2/3 and HNF4α in the presence of TGF-β and siRNA. We first analyzed the data of cells transfected with control siRNA. Twenty four hours after TGF-β stimulation, 4.3 and 21.1% of the genes with Smad2/3 binding regions were regulated (either up- or down-regulated) more than 2- and 1.5-fold, respectively (Table 1). We observed that Smad2/3 binding regions were weakly enriched in the genes up-regulated by TGF-β at 1.5 h (supplemental Fig. S2). Many of the genes with Smad2/3 binding regions were not transcriptionally regulated by TGF-β, and these findings were essentially similar to those in our previous analysis in HaCaT cells (11). We then found that HNF4α siRNA inhibited the expression changes of common target genes of HNF4α and Smad2/3 by TGF-β 1.5 h after stimulation (Fig. 4*D*). This result underscored the general roles of HNF4α in hepatocyte-specific transcriptome regulation by TGF-β. In contrast, the effect of HNF4α silencing was not so obvious in the TGF-β-induced expression changes 24 h after stimulation, compared with the setting after 1.5 h of TGF-β stimulation, although TGF-β-induced expression changes of a subset of genes appeared to be rather enhanced by HNF4α knockdown (Fig. 4*E*). We focused on the changes at 1.5 h, when we obtained Smad2/3 and HNF4α binding data, and we listed target genes of TGF-β and the effect of HNF4α knockdown (Table 2). We identified *MIXL1* as both TGF-β- and HNF4α-regulated gene with no Smad2/3 binding regions in HaCaT cells.

HNF4α Provides a New Mechanism of TGF-β-induced MIXL1 Expression—As shown in Fig. 5*A*, significant binding of Smad2/3 and HNF4α to the *MIXL1* promoter was observed (Fig. 5*A*). We confirmed the binding of these transcription factors by ChIP-qPCR (Fig. 5, *B* and *C*) and changes in the expression of *MIXL1* by HNF4α siRNA by RT-qPCR (Fig. 5*D*). We then determined the sequence of the *MIXL1* promoter bound by Smad2/3 and HNF4α. We first found two possible HNF4α-binding motifs (Fig. 6*A*). Using a promoter reporter assay, we found

TABLE 2

TGF-β-induced genes with Smad2/3 and HNF4α binding at 1.5 h

Target genes of TGF-β in HepG2 cells that were induced more than 2-fold at 1.5 h and that have common binding regions for Smad2/3 and HNF4α were sorted by their expression changes in the presence or absence of HNF4α siRNA. Presence of Smad2/3 binding regions in HaCaT cells is also shown in the 2nd column.

Gene symbol	Smad2/3 binding in HaCaT cells	Relative expression (siNC/siHNF4α)	Induction by TGF-β
		-fold	*-fold*
CMIP	+	0.3	2.1
REPIN1	+	0.3	2.2
MIXL1	−	0.3	2.3
ERRFI1	+	0.4	2.7
FASTK	+	0.5	2.1
ZNF48	−	0.5	2.6
CBX4	−	0.6	2.2
TMEM49	−	0.8	2.1
ZFP36L2	−	0.9	2.6
JUNB	+	1.0	9.8
DDIT4	+	1.1	5.6
GADD45B	+	1.3	4.9
RND1	−	1.4	5.5

that the transcriptional activity of the reporter containing the Smad2/3-HNF4α binding regions was up-regulated by TGF-β, which was significantly repressed by mutations in the HNF4α-binding sequences (Fig. 6*B*). We next searched for canonical Smad-binding elements conserved between mouse and human. We identified three Smad-binding elements between the two HNF4α motifs, and one just upstream of the distal HNF4α motif, termed SBE1 to -4 (Fig. 6*C*). Of them, only a mutation in SBE2 lost TGF-β-induced transcription (Fig. 6*D*). These results suggested that both HNF4α-binding motifs and SBE2 are required for *MIXL1* reporter activity induced by TGF-β.

The transcriptional activity of the reporter was inhibited by HNF4α siRNA, which was observed even without TGF-β, suggesting that preceding binding of HNF4α to its binding motifs as observed in Fig. 5*A* was important both for the basal and TGF-β-induced transcriptional activation of *MIXL1* promoter (Fig. 7*A*). We also investigated the effect of forced HNF4α expression in HaCaT cells to determine whether HNF4α was able to activate the *MIXL1* transcriptional activity in these cells. As shown in Fig. 7*B*, HNF4α induced the transcriptional activity of the *MIXL1* promoter reporter in HaCaT cells. We then examined the effect of a mutant of HNF4α that cannot bind to DNA (HNF4α CR mutant) (32) and found that DNA binding

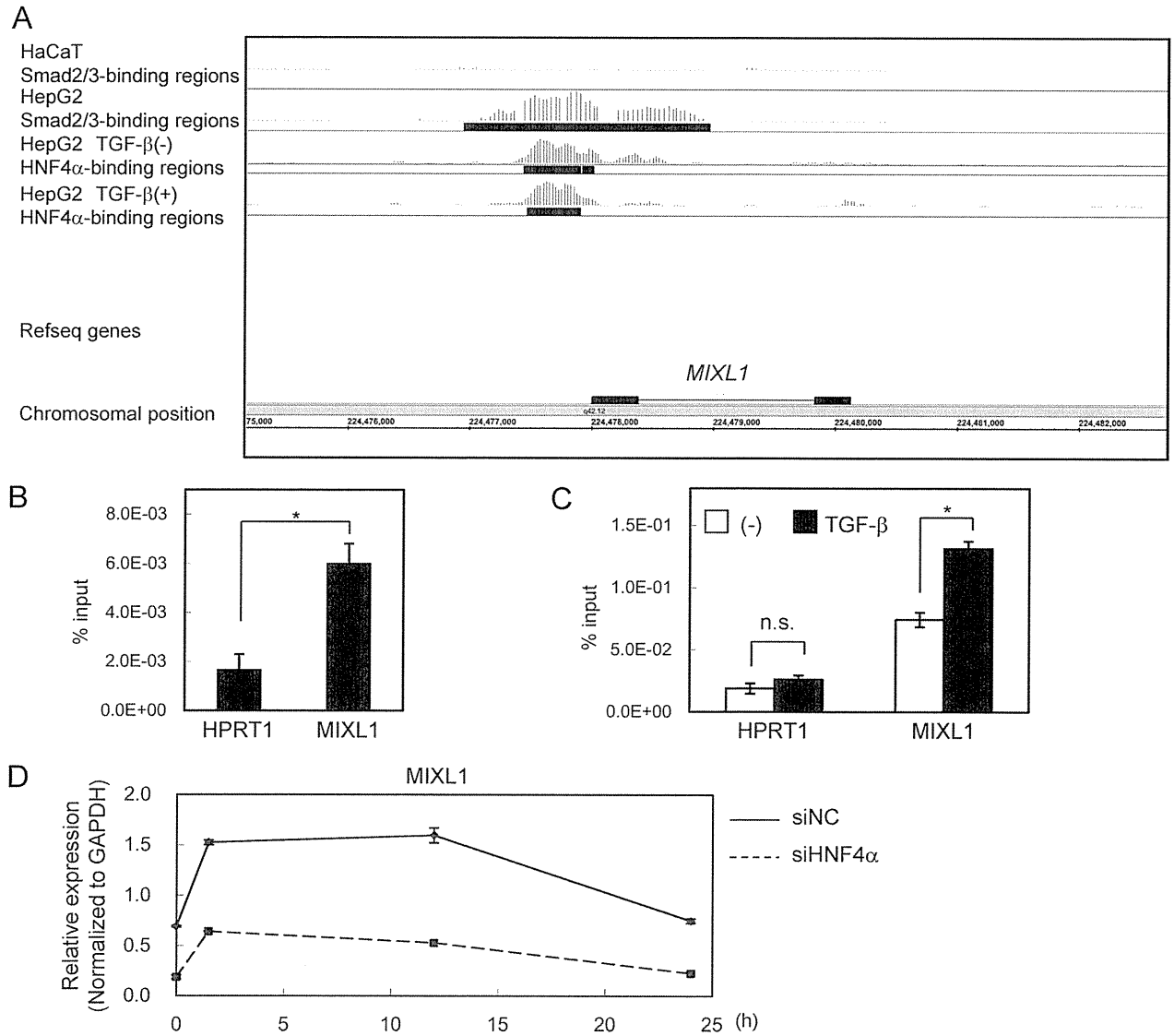

FIGURE 5. **Smad2/3 and HNF4α bindings in the *MIXL1* locus.** *A*, Smad2/3- and HNF4α-enriched regions in the *MIXL1* locus are shown as in Fig. 3*A*. *B*, HepG2 cells were treated with 120 pM TGF-β for 1.5 h, fixed in formaldehyde, and harvested. Smad2/3 binding to the *MIXL1* locus was verified by ChIP-qPCR. *HPRT1* served as a negative control. *C*, HepG2 cells were treated with or without 120 pM TGF-β for 1.5 h, and ChIP-qPCR analysis of the *MIXL1* locus using anti-HNF4α was performed as in *B*. *n.s.*, not significant. *D*, effects of knockdown of HNF4α on TGF-β-induced expression changes of *MIXL1*. HepG2 cells were transfected with control siRNA (*siNC*) or siHNF4α, treated with 3 ng/ml TGF-β for the indicated times, and harvested. HNF4α expression was quantified by RT-qPCR. *, $p < 0.05$; *error bars*, S.D.

activity was required for its TGF-β-induced transcriptional activation (Fig. 7*C*). Finally, the effect of HNF4α siRNA on Smad2/3 binding to the *MIXL1* promoter was determined. HNF4α siRNA inhibited the TGF-β-induced Smad2/3 binding to the *MIXL1* promoter, indicating that the recruitment of Smad2/3 was one of the mechanisms of transcriptional regulation by HNF4α under TGF-β stimulation (Fig. 7*D*).

Taken together, these findings propose that the preceding binding of HNF4α on *MIXL1* promoter enables the recruitment of Smad2/3 to this promoter after TGF-β stimulation and confers TGF-β-mediated HepG2-specific *MIXL1* induction.

DISCUSSION

Recent technological advances, including ChIP-chip and ChIP-seq, provide a functional platform for comprehensive understanding of transcriptional regulation. This study revealed that Smad2/3 binding regions specifically observed in HepG2 cells were enriched in HNF4α binding regions. HNF4α was also expressed in Hep3B cells, and HNF4α-binding motif was identified in Smad2/3 binding regions in Hep3B cells by CEAS analysis (data not shown), which suggests that the functional relation between Smad2/3 and HNF4α is commonly observed in hepatocyte-derived cells. Based on the findings on the HNF4α-Smad interaction (18), physical interaction between HNF4α and Smads is important, at least in part, for TGF-β-induced Smad2/3 binding and transcriptional activation in HepG2 cells. It is also possible that HNF4α has additional indirect interactive functions for TGF-β signaling. Many regulatory mechanisms control the expression of a proper set of genes in various cells and tissues. At the genome level, CpG methylation plays a cen-

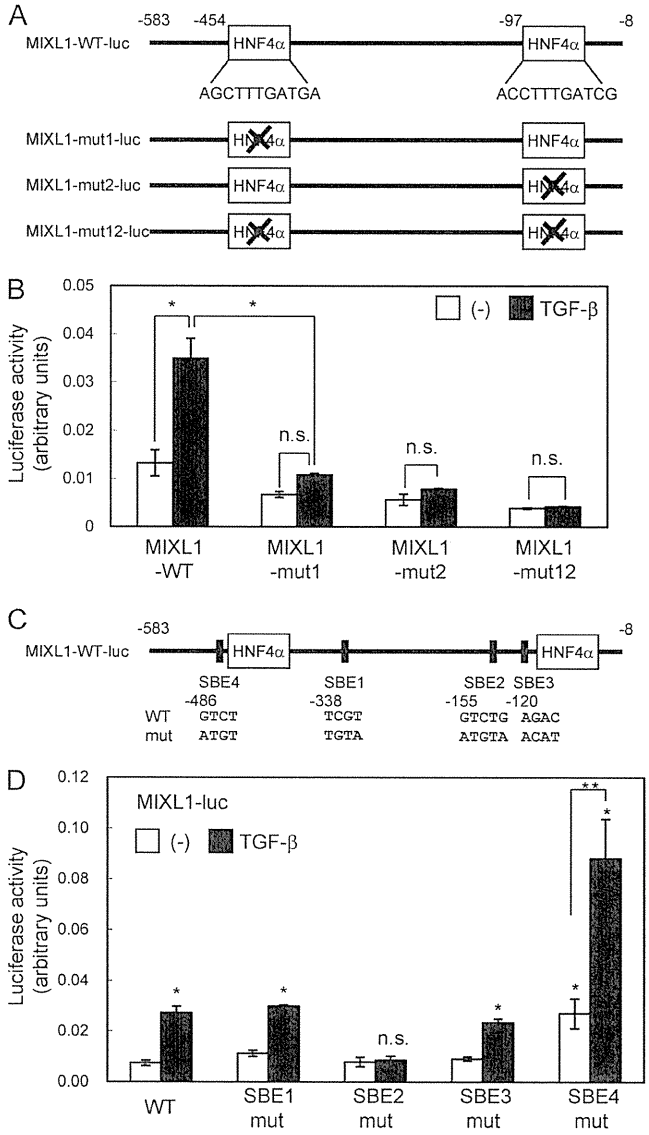

FIGURE 6. **Identification of regulatory elements important for TGF-β-induced transactivation of the *MIXL1* promoter.** *A*, schematic representation of HNF4α-binding motifs in the Smad2/3 and HNF4α binding region of the *MIXL1* promoter. Promoter reporters with mutations in their HNF4α-binding motifs are shown in the *lower panel*. *B*, activation of the *MIXL1* gene promoter by TGF-β and effects of mutations in putative HNF4α-binding motifs. HepG2 cells were transfected with the *MIXL1* promoter and its mutants and treated with TGF-β for 24 h. *C*, conserved Smad-binding elements (*SBEs*) of the *MIXL1* promoter. Four Smad-binding elements that were conserved between mouse and human (SBE 1–4) are shown with their relative positions from the transcription start site. Nucleotide sequences of Smad-binding elements and their mutations used in *D* are also shown. *WT*, wild-type; *mut*, mutant. *D*, effect of mutations in Smad-binding elements on TGF-β-induced transcriptional activity of *MIXL1* promoter. Cells were treated as in *B*, and luciferase activities were determined. *, $p < 0.05$ compared with WT without TGF-β; **, $p < 0.05$ compared with SBE4 mutant, without TGF-β; *n.s.*, not significant compared with WT and SBE2 mutant, without TGF-β; *error bars*, S.D.

tral role to avoid unintended expression of genes that are not suitable for the given tissue (33). Modification of the histone tail is also well known to lead to the formation of either euchromatin or heterochromatin. These modifications of the genome or histones allow transcription factors and cofactors to access the cell- and tissue-specific genomic loci to exert their actions. Modifications of the genome and histones are sometimes induced by *trans* factors during differentiation of the cells and tissues (34, 35). HNF4α physically interacts with the histone acetyltransferase complex and chromatin remodeling complex (29), and it is thus possible that HNF4α induces such epigenomic changes in the liver and indirectly provides Smad2/3 to access to hepatocyte-specific binding regions.

Identification of Smad binding regions downstream of the TGF-β/activin signaling by ChIP-chip analysis has been performed using several cell lines. Recently, Fei *et al.* (36) reported promoter analysis of Smad2 binding regions in mouse embryonic stem cells by ChIP-chip. We and Qin *et al.* (12, 37) analyzed Smad4 binding regions under TGF-β stimulation using HaCaT and ovarian surface epithelial cells, respectively. It has been reported that transcription factor binding regions in the same target gene loci differ among the five vertebrate species (38); it is thus difficult to compare the ChIP-chip or ChIP-seq data obtained from mouse and human. Differences in the ChIP efficiencies of the antibodies also make the comparison of the data difficult (12). Importantly, we used the same antibody and sample preparation procedures for HaCaT cells and HepG2 cells. Our present analysis thus revealed for the first time that Smad binding regions greatly differ among cell lines. Analysis of HaCaT-specific *trans* factors will facilitate our understanding of cell type-specific TGF-β-induced transcription in the future. However, comparison of the number of binding regions in different cell types is still difficult. We found a greater number of Smad2/3 binding regions in HepG2 cells than in HaCaT cells. Because the phosphorylation of Smad3 was weaker and the percent input value of the Smad2/3 ChIP sample was smaller in HepG2 than HaCaT cells, we cannot conclude that HepG2 cells have more Smad2/3 binding regions than HaCaT. It should also be noted that we cannot fully exclude that the antibody recognizes unknown genome-bound molecules in addition to Smad2/3.

Comparison of ChIP-chip and ChIP-seq data of the same transcription factor has been reported (39). In general, ChIP-seq is reported to be more sensitive and specific than ChIP-chip. Oligonucleotide-based array analysis has a potential risk of cross-hybridization and false discovery. Conversely, ChIP-seq also has difficulty in identifying GC-rich sequences (10, 39). We primarily focused on the comparison of our previously reported Smad2/3 binding regions to those of different cell types by the same platform. However, based on the known problems as described above, comparison of the Smad2/3-HNF4α binding regions will be more accurately performed by the ChIP-seq in the future.

Interaction of several transcription factors at the same enhancer positions has been recognized, and the complex is called "enhanceosome." Structure of such complex and their binding DNA motifs have been analyzed in the interferon-β promoter as reviewed by Panne (40). In enhanceosome, each transcription factor physically interacts with others to provide its adequate surface that can bind to the series of their corresponding DNA motifs. Several reports have identified HNF4α binding regions by ChIP-chip and ChIP-seq analyses (29–31, 38, 41–43). Many transcription factors, *e.g.* FOXA2, GABP, HNF1α, HNF4γ, HNF6, cohesin, and CDX2, were identified to

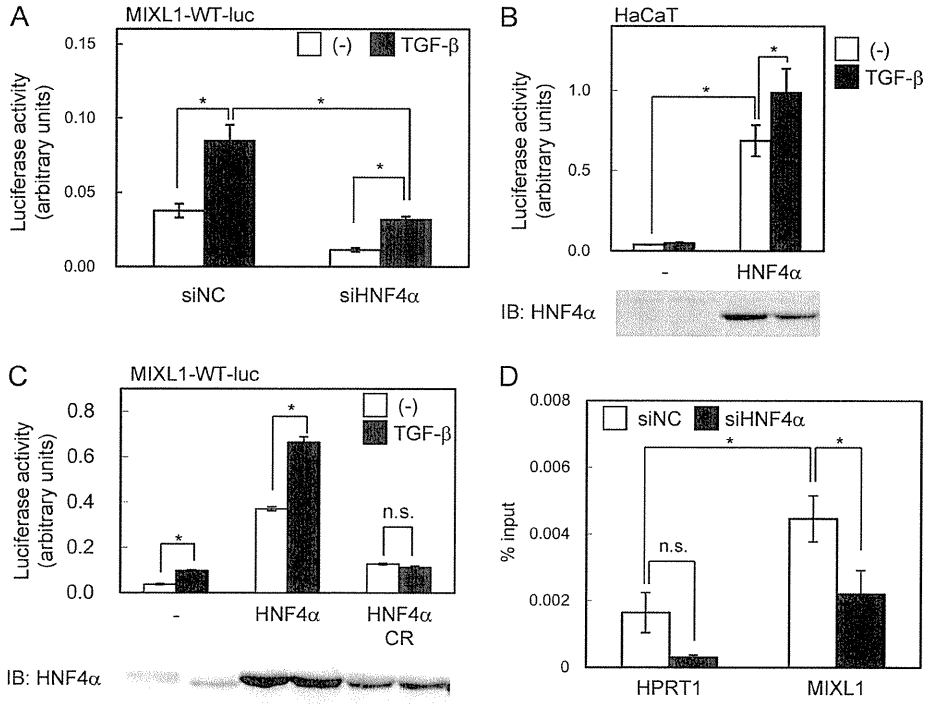

FIGURE 7. **Roles of HNF4α on Smad2/3 binding and transcriptional activity of *MIXL1* promoter.** *A*, effects of HNF4α knockdown on transactivation of the *MIXL1* promoter. HepG2 cells were transfected with siRNAs 1 day before transfection with the reporter constructs. *siNC*, negative control siRNA. *B*, effects of exogenous HNF4α on transactivation of the *MIXL1* promoter. HNF4α was exogenously expressed in HaCaT cells, and transcriptional activity of MIXL1 reporter was determined. The *lower panel* shows the protein expression of HNF4α. *IB*, immunoblot. *C*, HNF4α (variant 2, RefSeq ID: NM_000457) or its C115R (*CR*) mutant, which does not bind to DNA, was overexpressed in HepG2 cells. The *lower panel* shows the protein expression of HNF4α and its mutant. *D*, effect of HNF4α siRNA on Smad2/3 binding to the *MIXL1* locus. HepG2 cells were transfected with siRNAs 24 h before TGF-β stimulation. Cells were fixed 1.5 h after treatment, and ChIP-qPCR was performed as in Fig. 1*B*. *Error bars*, S.D.; *, $p < 0.05$; *n.s.*, not significant.

co-localize with HNF4α through these analyses. Other reports also revealed interaction of FOXO1 or retinoic acid receptor/retinoid X receptor with HNF4α on specific promoters (44, 45). These findings clearly revealed steady-state binding regions of HNF4α on the genome and suggested that transcription factors that co-localize or interact with HNF4α may form enhanceosome with HNF4α. Changes in the HNF4α binding regions were found during differentiation of an intestinal epithelial cell line CaCo2 (43); however, to our knowledge, the effect of single extracellular stimulation on genome-wide HNF4α binding regions has not yet been elucidated. Our present analysis provides the data of HNF4α binding regions following TGF-β stimulation, which were compared with the Smad2/3 binding regions in HaCaT cells that lack the expression of HNF4α. We have found that large proportions of HNF4α binding regions in HepG2 cells were unchanged by TGF-β stimulation. However, some changes in HNF4α binding regions were observed with regard to their positions and their strength, suggesting that TGF-β might regulate a subset of HNF4α binding regions. de Boussac *et al.* (46) reported that hepatocyte growth factor inhibited HNF4α binding to the *ABCC6* promoter, which together suggest the importance of changes in the HNF4α binding positions by external stimuli. We also found that the effect of HNF4α on the TGF-β-induced gene expression after 24 h of TGF-β stimulation was different from that after 1.5 h of TGF-β stimulation. Studies on the changes in the genome-wide HNF4α and Smad2/3 binding after TGF-β stimulation at several time points and ChIP-seq analysis of HNF4α with other interactive factors in relation to their binding DNA sequences will reveal new mechanisms of the regulation of HNF4α-induced transcription in the context of the enhanceosome.

MIXL1 is an ortholog of *Xenopus* Mix.1, a transcription factor rapidly induced by activin during the early stage of *Xenopus* development (47). There are six known homologs that have been identified in *Xenopus* to engage in the formation of mesoderm and endoderm (48, 49). However, only one ortholog of Mix.1 is known in human and mouse (50). MIXL1 is required for the development of the chordamesoderm, heart, and gut in mouse (51). Forced expression of MIXL1 in embryonic stem cells resulted in the differentiation of the cells to endoderm (52). TGF-β is reported to induce *Mix.2* promoter activity by formation of a Smad2/Smad4/FAST-1 (FoxH1) complex (53). In mouse, Smads and FAST-1 interact to up-regulate the transcriptional activity of the *MIXL1* promoter (54, 55). However, FAST-1 is not expressed in HepG2 cells (56). Our finding of TGF-β-induced *MIXL1* expression in HepG2 cells suggests a previously unrecognized regulatory mechanism of its expression by HNF4α in the absence of FAST-1. During development, HNF4α is expressed in the visceral endoderm during the gastrulation stage and plays a role in the differentiation of the embryonic mesoderm (57). MIXL1 is also expressed in the visceral endoderm and induces migration of the embryonic endoderm. HNF4α-null mice embryo showed impaired development of mature visceral endoderm, indicating that HNF4α acts upstream of MIXL1, at least in the visceral endoderm. Notably, both HNF4α and MIXL1 positively reg-

ulate the E-cadherin expression (52, 58), and the HNF4α expression was repressed in a model of progression of hepatocellular carcinoma (59). Functional analysis of MIXL1 in liver fibrosis and hepatocellular carcinoma in relation to TGF-β signaling might reveal the roles of MIXL1 in the adult liver in the future.

Acknowledgments—We thank Keiko Yuki, Hiroko Meguro, and Kaori Shiina for technical assistance.

REFERENCES

1. Ikushima, H., and Miyazono, K. (2010) *Nat. Rev. Cancer* **10,** 415–424
2. Heldin, C. H., Miyazono, K., and ten Dijke, P. (1997) *Nature* **390,** 465–471
3. Feng, X. H., and Derynck, R. (2005) *Annu. Rev. Cell Dev. Biol.* **21,** 659–693
4. Miyazono, K., Kamiya, Y., and Morikawa, M. (2010) *J. Biochem.* **147,** 35–51
5. Shi, Y., Wang, Y. F., Jayaraman, L., Yang, H., Massagué, J., and Pavletich, N. P. (1998) *Cell* **94,** 585–594
6. Zawel, L., Dai, J. L., Buckhaults, P., Zhou, S., Kinzler, K. W., Vogelstein, B., and Kern, S. E. (1998) *Mol. Cell* **1,** 611–617
7. Ikushima, H., and Miyazono, K. (2010) *Cancer Sci.* **101,** 306–312
8. Gomis, R. R., Alarcón, C., He, W., Wang, Q., Seoane, J., Lash, A., and Massagué, J. (2006) *Proc. Natl. Acad. Sci. U.S.A.* **103,** 12747–12752
9. Eeckhoute, J., Lupien, M., and Brown, M. (2009) *Methods Mol. Biol.* **556,** 155–164
10. Park, P. J. (2009) *Nat. Rev. Genet.* **10,** 669–680
11. Koinuma, D., Tsutsumi, S., Kamimura, N., Taniguchi, H., Miyazawa, K., Sunamura, M., Imamura, T., Miyazono, K., and Aburatani, H. (2009) *Mol. Cell. Biol.* **29,** 172–186
12. Koinuma, D., Tsutsumi, S., Kamimura, N., Imamura, T., Aburatani, H., and Miyazono, K. (2009) *Cancer Sci.* **100,** 2133–2142
13. Sladek, F. M., Zhong, W. M., Lai, E., and Darnell, J. E., Jr. (1990) *Genes Dev.* **4,** 2353–2365
14. Si-Tayeb, K., Lemaigre, F. P., and Duncan, S. A. (2010) *Dev. Cell* **18,** 175–189
15. Jiang, G., Nepomuceno, L., Hopkins, K., and Sladek, F. M. (1995) *Mol. Cell. Biol.* **15,** 5131–5143
16. Lucas Sd, S., López-Alcorocho, J. M., Bartolomé, J., and Carreño, V. (2004) *Biochem. Biophys. Res. Commun.* **321,** 688–694
17. Ishikawa, F., Nose, K., and Shibanuma, M. (2008) *Exp. Cell Res.* **314,** 2131–2140
18. Kardassis, D., Pardali, K., and Zannis, V. I. (2000) *J. Biol. Chem.* **275,** 41405–41414
19. Chou, W. C., Prokova, V., Shiraishi, K., Valcourt, U., Moustakas, A., Hadzopoulou-Cladaras, M., Zannis, V. I., and Kardassis, D. (2003) *Mol. Biol. Cell* **14,** 1279–1294
20. Mizutani, A., Saitoh, M., Imamura, T., Miyazawa, K., and Miyazono, K. (2010) *J. Biochem.* **148,** 733–741
21. Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K., and Peters, J. M. (2008) *Nature* **451,** 796–801
22. Kaneshiro, K., Tsutsumi, S., Tsuji, S., Shirahige, K., and Aburatani, H. (2007) *Genomics* **89,** 178–188
23. Johnson, W. E., Li, W., Meyer, C. A., Gottardo, R., Carroll, J. S., Brown, M., and Liu, X. S. (2006) *Proc. Natl. Acad. Sci. U.S.A.* **103,** 12457–12462
24. Ji, H., Jiang, H., Ma, W., Johnson, D. S., Myers, R. M., and Wong, W. H. (2008) *Nat. Biotechnol.* **26,** 1293–1300
25. Ji, X., Li, W., Song, J., Wei, L., and Liu, X. S. (2006) *Nucleic Acids Res.* **34,** W551–W554
26. Nagano, Y., Koinuma, D., Miyazawa, K., and Miyazono, K. (2010) *J. Biochem.* **147,** 545–554
27. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) *Nat. Protoc.* **4,** 44–57
28. Portales-Casamar, E., Thongjuea, S., Kwon, A. T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W. W., and Sandelin, A. (2010) *Nucleic Acids Res.* **38,** D105–D110
29. Daigo, K., Kawamura, T., Ohta, Y., Ohashi, R., Katayose, S., Tanaka, T., Aburatani, H., Naito, M., Kodama, T., Ihara, S., and Hamakubo, T. (2011) *J. Biol. Chem.* **286,** 674–686
30. Wallerman, O., Motallebipour, M., Enroth, S., Patra, K., Bysani, M. S., Komorowski, J., and Wadelius, C. (2009) *Nucleic Acids Res.* **37,** 7498–7508
31. Schmidt, D., Schwalie, P. C., Ross-Innes, C. S., Hurtado, A., Brown, G. D., Carroll, J. S., Flicek, P., and Odom, D. T. (2010) *Genome Res.* **20,** 578- 588
32. Taylor, D. G., Haubenwallner, S., and Leff, T. (1996) *Nucleic Acids Res.* **24,** 2930–2935
33. Bernstein, B. E., Meissner, A., and Lander, E. S. (2007) *Cell* **128,** 669–681
34. Kim, M. S., Kondo, T., Takada, I., Youn, M. Y., Yamamoto, Y., Takahashi, S., Matsumoto, T., Fujiyama, S., Shirode, Y., Yamaoka, I., Kitagawa, H., Takeyama, K., Shibuya, H., Ohtake, F., and Kato, S. (2009) *Nature* **461,** 1007–1012
35. Lan, F., Nottke, A. C., and Shi, Y. (2008) *Curr. Opin. Cell Biol.* **20,** 316–325
36. Fei, T., Zhu, S., Xia, K., Zhang, J., Li, Z., Han, J. D., and Chen, Y. G. (2010) *Cell Res.* **20,** 1306–1318
37. Qin, H., Chan, M. W., Liyanarachchi, S., Balch, C., Potter, D., Souriraj, I. J., Cheng, A. S., Agosto-Perez, F. J., Nikonova, E. V., Yan, P. S., Lin, H. J., Nephew, K. P., Saltz, J. H., Showe, L. C., Huang, T. H., and Davuluri, R. V. (2009) *BMC Syst. Biol.* **3,** 73
38. Schmidt, D., Wilson, M. D., Ballester, B., Schwalie, P. C., Brown, G. D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C. P., Mackay, S., Talianidis, I., Flicek, P., and Odom, D. T. (2010) *Science* **328,** 1036–1040
39. Ho, J. W., Bishop, E., Karchenko, P. V., Nègre, N., White, K. P., and Park, P. J. (2011) *BMC Genomics* **12,** 134
40. Panne, D. (2008) *Curr. Opin. Struct. Biol.* **18,** 236–242
41. Boyd, M., Bressendorff, S., Møller, J., Olsen, J., and Troelsen, J. T. (2009) *BMC Gastroenterol.* **9,** 68
42. Odom, D. T., Zizlsperger, N., Gordon, D. B., Bell, G. W., Rinaldi, N. J., Murray, H. L., Volkert, T. L., Schreiber, J., Rolfe, P. A., Gifford, D. K., Fraenkel, E., Bell, G. I., and Young, R. A. (2004) *Science* **303,** 1378–1381
43. Verzi, M. P., Shin, H., He, H. H., Sulahian, R., Meyer, C. A., Montgomery, R. K., Fleet, J. C., Brown, M., Liu, X. S., and Shivdasani, R. A. (2010) *Dev. Cell* **19,** 713–726
44. Ganjam, G. K., Dimova, E. Y., Unterman, T. G., and Kietzmann, T. (2009) *J. Biol. Chem.* **284,** 30783–30797
45. Mosialou, I., Zannis, V. I., and Kardassis, D. (2010) *J. Biol. Chem.* **285,** 30719–30730
46. de Boussac, H., Ratajewski, M., Sachrajda, I., Köblös, G., Tordai, A., Pulaski, L., Buday, L., Váradi, A., and Arányi, T. (2010) *J. Biol. Chem.* **285,** 22800–22808
47. Rosa, F. M. (1989) *Cell* **57,** 965–974
48. Ecochard, V., Cayrol, C., Rey, S., Foulquier, F., Caillol, D., Lemaire, P., and Duprat, A. M. (1998) *Development* **125,** 2577–2585
49. Vize, P. D. (1996) *Dev. Biol.* **177,** 226–231
50. Robb, L., Hartley, L., Begley, C. G., Brodnicki, T. C., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., and Elefanty, A. G. (2000) *Dev. Dyn.* **219,** 497–504
51. Hart, A. H., Hartley, L., Sourris, K., Stadler, E. S., Li, R., Stanley, E. G., Tam, P. P., Elefanty, A. G., and Robb, L. (2002) *Development* **129,** 3597–3608
52. Lim, S. M., Pereira, L., Wong, M. S., Hirst, C. E., Van Vranken, B. E., Pick, M., Trounson, A., Elefanty, A. G., and Stanley, E. G. (2009) *Stem Cells* **27,** 363–374
53. Watanabe, M., and Whitman, M. (1999) *Development* **126,** 5621–5634
54. Hart, A. H., Willson, T. A., Wong, M., Parker, K., and Robb, L. (2005) *Biochem. Biophys. Res. Commun.* **333,** 1361–1369
55. Izzi, L., Silvestri, C., von Both, I., Labbé, E., Zakin, L., Wrana, J. L., and Attisano, L. (2007) *EMBO J.* **26,** 3132–3143
56. Hayashi, H., Abdollah, S., Qiu, Y., Cai, J., Xu, Y. Y., Grinnell, B. W., Richa-

rdson, M. A., Topper, J. N., Gimbrone, M. A., Jr., Wrana, J. L., and Falb, D. (1997) *Cell* **89,** 1165–1173
57. Chen, W. S., Manova, K., Weinstein, D. C., Duncan, S. A., Plump, A. S., Prezioso, V. R., Bachvarova, R. F., and Darnell, J. E., Jr. (1994) *Genes Dev.* **8,** 2466–2477
58. Späth, G. F., and Weiss, M. C. (1998) *J. Cell Biol.* **140,** 935–946
59. Lazarevich, N. L., Cheremnova, O. A., Varga, E. V., Ovchinnikov, D. A., Kudrjavtseva, E. I., Morozova, O. V., Fleishman, D. I., Engelhardt, N. V., and Duncan, S. A. (2004) *Hepatology* **39,** 1038–1047
60. Saeed, A. I., Bhagabati, N. K., Braisted, J. C., Liang, W., Sharov, V., Howe, E. A., Li, J., Thiagarajan, M., White, J. A., and Quackenbush, J. (2006) *Methods Enzymol.* **411,** 134–193

Homozygously deleted gene DACH1 regulates tumor-initiating activity of glioma cells

Akira Watanabe[a,b,1], Hideki Ogiwara[a,c,1], Shogo Ehata[d], Akitake Mukasa[c], Shumpei Ishikawa[a,e], Daichi Maeda[e], Keisuke Ueki[f], Yasushi Ino[c], Tomoki Todo[c], Yasuhiro Yamada[b], Masashi Fukayama[e], Nobuhito Saito[c], Kohei Miyazono[d], and Hiroyuki Aburatani[a,g,2]

[a]Genome Science Division, Research Center for Advanced Science and Technology, University of Tokyo, Tokyo 153-8904, Japan; [b]Department of Reprogramming Science, Center for iPS Research and Application, Kyoto University, Kyoto 606-8507, Japan; Departments of [c]Neurosurgery, [d]Molecular Pathology, and [e]Human Pathology, Graduate School of Medicine, University of Tokyo, Tokyo 113-0033, Japan; [f]Department of Neurosurgery, Dokkyo University Medical School, Tochigi 321-0293, Japan; and [g]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Saitama 332-0012, Japan

Edited* by Charles R. Cantor, Sequenom, Inc., San Diego, CA, and approved June 1, 2011 (received for review June 22, 2009)

Loss or reduction in function of tumor suppressor genes contributes to tumorigenesis. Here, by allelic DNA copy number analysis using single-nucleotide polymorphism genotyping array and mass spectrometry, we report homozygous deletion in glioblastoma multiformes at chromosome 13q21, where *DACH1* gene is located. We found decreased cell proliferation of a series of glioma cell lines by forced expression of DACH1. We then generated U87TR-Da glioma cells, where DACH1 expression could be activated by exposure of the cells to doxycycline. Both ex vivo cellular proliferation and in vivo growth of s.c. transplanted tumors in mice are reduced in U87TR-Da cells with DACH1 expression (U87-DACH1-high), compared with DACH1-nonexpressing U87TR-Da cells (U87-DACH1-low). U87-DACH1-low cells form spheroids with CD133 and Nestin expression in serum-free medium but U87-DACH1-high cells do not. Compared with spheroid-forming U87-DACH1-low cells, adherent U87-DACH1-high cells display lower tumorigenicity, indicating DACH1 decreases the number of tumor-initiating cells. Gene expression analysis and chromatin immunoprecipitation assay reveal that fibroblast growth factor 2 (FGF2/bFGF) is transcriptionally repressed by DACH1, especially in cells cultured in serum-free medium. Exogenous bFGF rescues spheroid-forming activity and tumorigenicity of the U87-DACH1-high cells, suggesting that loss of DACH1 increases the number of tumor-initiating cells through transcriptional activation of bFGF. These results illustrate that DACH1 is a distinctive tumor suppressor, which does not only suppress growth of tumor cells but also regulates bFGF-mediated tumor-initiating activity of glioma cells.

neural differentiation | gliomagenesis

Glioblastoma multiformes (GBMs), the most frequent primary malignant brain tumor in adults, are aggressive and highly invasive tumors (1). Genetic alterations of GBMs, including aberration of DNA copy number such as gene amplifications, loss of heterozygosity (LOH), and homozygous deletions, leads to activation of oncogenes and inactivation of tumor suppressor genes (1–3). DNA copy number analysis by single-nucleotide polymorphism (SNP) genotyping enables the high-resolution analysis of allelic DNA copy number and has been used to obtain a genome-wide view of DNA copy number alterations in human cancers (4–10). Particularly, pairwise analysis of normal and tumor DNAs is crucial in detecting homozygous deletion in clinical specimens, because infiltrating nontumorous cells are significant in GBMs.

In this study we examined the allelic copy number of paired glioma and blood DNAs by SNP genotyping array analysis by using Genome Imbalanced Map (GIM) algorithm (5, 11), which could calculate the signal ratio of SNP genotyping array in an allelic manner. We identified a unique homozygous deletion at *DACH1* gene region on chromosome 13q21, and we demonstrated forced expression of DACH1 reduced proliferation of cultured glioma cells and in vivo tumor growth in ortholotopic xenograft model. We also found that DACH1 inhibited formation of tumor-initiating spheroids, presumably by directly repressing expression of fibroblast growth factor-2 (*FGF2*), suggesting *DACH1* is a unique tumor suppressor of glioblastoma, which not only suppresses tumor growth but also inhibits generation of tumor-initiating cells.

Results

DACH1 Gene on Chromosome 13q21 Is Homozygously Deleted in Glioblastoma. To identify genomic alterations involved in gliomagenesis, we performed DNA copy number analysis of eight GBMs by using SNP genotyping array (Fig. 1*A* and Fig. S1*A*), as well as analysis of a corresponding normal blood DNA for highlighting tumor-specific alterations. We observed high-level amplification at chromosome 7q21 (inferred total copy number >8) and copy number reduction within chromosome arm 4q, 10p, 13q, 16q, 17q, and 18q, and we detected homozygously deleted loci at chromosome 9q21 and 10q23, which spanned known tumor suppressor genes *CDKN2A* and *PTEN*, respectively. In addition, we found a unique homozygous deletion at chromosome 13q21 in GBM case 4 and LOH at the region in GBM case 1 and 3 (Fig. 1*A* and Fig. S1*A*). Although loss of chromosome 13q14.2 spanning *RB1* gene is frequently observed in human malignancies including GBMs (12, 13), the homozygous deletion of chromosome 13q21 has not been reported.

To examine allelic DNA copy numbers at the chromosome 13q21 region in additional GBM cases, we performed targeted genotyping analysis of 28 paired GBMs and blood cells by high-density mass spectrometric analysis using MassARRAY (14, 15). The chromosomal losses were found at least in 11 samples (GBM case 1, 3, 4, 5, 18, 19, 24, 25, 26, 27, and 28; 39.3%), and three of them (GBM case 4, 5, and 27; 10.7% of GBMs) displayed homozygous deletion (Fig. 1*B* and Fig. S1*B*). By combination of SNP genotyping array with MassARRAY analysis, homozygously deleted region at chromosome 13q21 of GBM case 4 was restricted to rs1999603 (probe S10)–rs1326684 (probe M8), which might be extended from rs9542598 (probe M2) to rs1421280 (probe S15). In two additional GBM cases, homozygous deletions found by MassARRAY at this locus were from rs10492537 (probe M6) to rs3818437 (probe M7), which might be extended

Author contributions: A.W. and H.A. designed research; A.W., H.O., S.E., A.M., D.M., Y.Y., and K.M. performed research; S.E., A.M., K.U., Y.I., T.T., M.F., and N.S. contributed new reagents/analytic tools; A.W. and S.I. analyzed data; and A.W., A.M., and H.A. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]A.W. and H.O. contributed equally to this work.

[2]To whom correspondence should be addressed: E-mail: haburata-tky@umin.ac.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.0906930108/-/DCSupplemental.

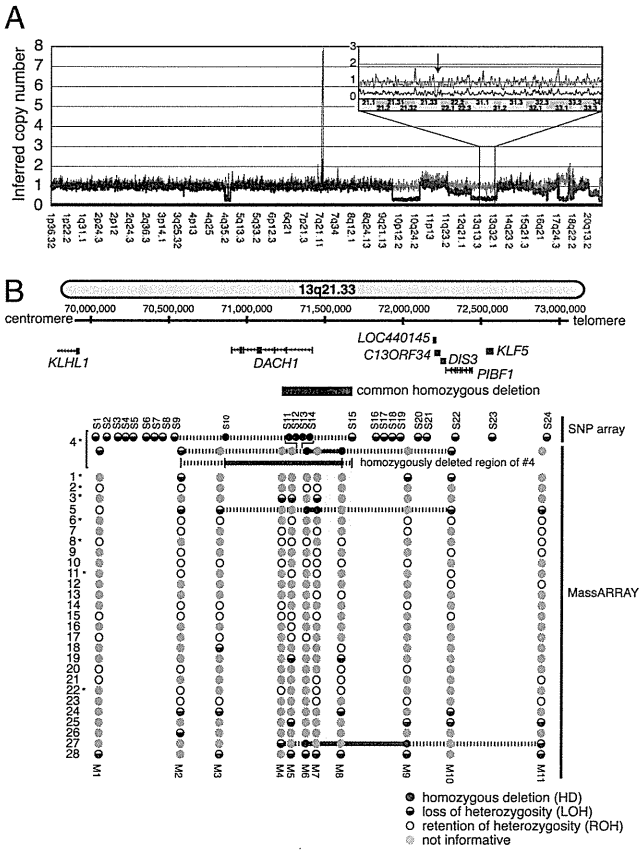

Fig. 1. DACH1 is homozygously deleted in GBMs. (*A*) An integrated view of DNA copy numbers and allelic alternations of GBM case 4. Scatter plot of inferred allelic copy numbers (red and blue) was estimated by GIM algorithm. An arrow indicated homozygously deleted loci. (*Right*) Magnified view of DNA copy number of GBM case 4 from chromosome 13q21 to 13q34. One allele showed DNA copy number reduction in whole region (blue), whereas the other allele (red) showed copy number reduction only at chromosome 13q21. (*B*) Summary of SNP genotyping in the 13q21 region. S1-24 and M1-11 are SNP ID available in SNP genotyping array and MassARRAY, respectively. Black line, the homozygously deleted region; dotted line, the possible extended region of homozygous deletions; *, also examined by SNP arrays shown in Fig. S1*A*.

to rs1931443 (probe M3)–rs7332388 (probe M10) in GBM case 5, and from rs10492537 (probe M6) to rs1981186 (probe M9), which might be extended to rs10507796 (probe M4)–rs1886452 (probe M11) in GBM case 27. The boundary of the common homozygously deleted region on the centromeric side was estimated at rs10507796 (probe M4), which was genotyped as LOH in case 27, whereas the telomeric boundary was estimated at rs1421280 (probe S15), which was genotyped as LOH in case 4 (Fig. 1*B*). Because the homozygously deleted region from rs10507796 to rs1421280 overlapped with *DACH1* gene and did not span the adjacent genes to *DACH1* such as *LOC440145* and *KLHL1*, we thought *DACH1* was a target of these deletions and could be a potential candidate as a tumor suppressor gene of GBMs. Quantitative PCR analysis demonstrated that DNA copy numbers of GBM samples with a homozygous deletion at *DACH1*-region were much lower than that of whole brain and normal fibroblast cells KMS-6, which has a normal karyotype as (46, XX) (Fig. S2*A*). In addition, we confirmed protein expression of DACH1 was reduced in glioma cells with a homozygous deletion at *DACH1* region, whereas it was detectable in vascular endothelial cells (Fig. S2*B*).

DACH1 Expression Reduces Growth of Glioma Cells. We screened an expression level of *DACH1* gene in a series of glioma cell lines. Forced expression of lentivirus-carrying *DACH1* reduced cell proliferation of SF188, U87MG, T98G, and LNG-308 glioma cell lines, where endogenous expression of *DACH1* was not detectable (Fig. S3 *A* and *B*). We then generated two U87MG-derived cell lines: U87TR-Da clone-16 and -18, where DACH1 expression could be induced by the addition of doxycycline (Fig. 2*A* and Fig. S4*A*). DACH1 decreased viability of the cells at 4 and 6 d after addition of doxycycline (Fig. 2*B*) and also abrogated anchorage-independent growth of the cells in soft agar (Fig. S3 *C* and *D*). We next examined the impact of DACH1 expression on in vivo growth of s.c.-injected tumors of U87TR-Da clone-16 and -18. Increased expression of DACH1 was observed at 8 d after replacement of doxycycline-free drinking water by doxycycline-supplemented water (Fig. 2 *C* and *D*). Growth of the U87TR-Da clone-16 and clone-18 tumors was significantly decreased by administration of doxycycline compared with U87TR-Da tumors supplemented with normal drinking water (Fig. 2 *E* and *F* and Fig. S4*B*). Tumor formation was reproduced by stereotactic intracerebral inoculation of DACH1-nonexpressing U87TR-Da cells, whereas tumor was not detectable by the injection of DACH1-expressing U87TR-Da cells (Fig. 2 *G* and *H* and Fig. S3*E*). These results demonstrated that *DACH1* decreased growth and proliferation of glioma cells both ex vivo and in vivo, supporting that *DACH1* is a tumor suppressor gene of GBMs.

DACH1 Inhibits Formation of Tumor-Initiating Spheroids of Glioma Cells. DACH1 is structurally related to c-Ski and SnoN, which act as transcriptional repressors of the transforming growth factor-β (TGF-β) signaling pathway through the interaction with Smad proteins (16). Previous studies showed that human DACH1 inhibited TGF-β signaling through repressing cyclin D1 (*CCND1*) expression and decreased proliferation of breast cancer cells (17, 18). We examined whether DACH1 expression affected TGF-β signaling in glioma cells. Induced DACH1 expression reduced the proliferation of the U-373MG cells, which are widely used as a cell model for analyzing TGF-β signaling (19), and repressed TGF-β3–stimulated $(CAGA)_9$- and p800-luciferase activity (Fig. S5 *A*–*C*). However, the expression of cyclin D1, which was proposed as a transcriptional target of DACH1 in breast cancer cell lines (18), was not affected by induced expression of DACH1 (Fig. S5*D*). Because DACH1 might repress expressions of the other target genes, but except for cyclin D1, we sought DACH1-regulated genes, which could affect growth of glioma cells. By global expression analysis of DACH1-high and -low cells, we found *FGF2* expression was repressed by DACH1 (Table S1, Fig. 3*A*, and Fig. S4*A*). Because FGF2, also named basic FGF (bFGF), is an essential factor for maintenance of self-renewal of glioma-initiating cells (20, 21), we thought DACH1 affected manintenance of self-renewal of glioma-initiating cells. We discovered DACH1, grown in serum-free neurobasal (NBE) medium (22), blocked spheroid formation of U87TR-Da cells, whereas cell morphology in serum-containing medium was not changed by DACH1 expression (Fig. 3*B*). Because glioma-initiating cells have been proposed to form spheroids (21), we thought that DACH1-nonexpressing U87TR-Da cells, which formed spheroids in NBE medium, showed high tumorigenicity compared with DACH1-expressing U87TR-Da cells in NBE medium. We confirmed that spheroid of DACH-nonexpressing cells cultured in NBE medium showed high expression of *CD133*, which has been reported as a marker of cells that are capable of tumor initiation (21), and a neural stem cell marker Nestin (Fig. 3 *C* and *D*). We then performed s.c. injection of U87TR-Da cells to examine the relationship between spheroid formation and tumorigenicity of the cells. We found tumor formation of DACH1-nonexpressing U87TR-Da cells (3 of 4 mice with 5×10^3 cells and all mice with 2×10^4 and 1×10^5 cells), whereas DACH1-

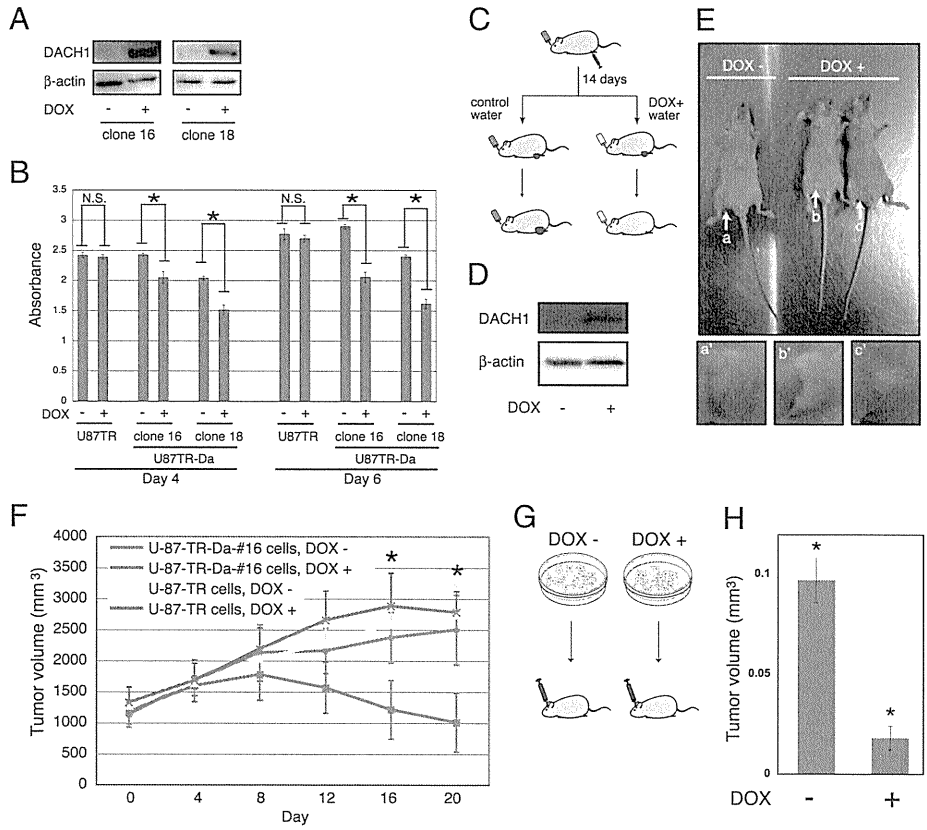

Fig. 2. DACH1 expression repressed growth of glioma cells. (*A*) Immunoblotting of U87TR-Da clone-16 and clone-18 cells with an anti-DACH1 antibody. (*B*) Cell proliferation of U87TR-Da cells, counted by WST-8 assay at 4 or 6 d after induction of *DACH1* by doxycycline. (*C*) Experimental model of tumor progression affected by DACH1 expression. Serially diluted U87TR-Da cells were injected s.c. into the backs of BALB/c nude mice, and the tumor formation was observed at 28 d after s.c. injection of the cells. (*D*) Expression of DACH1 in U87TR-Da tumor. DACH1 was detected by immunoblotting of xenografted U87TR-Da clone-16 tumors with an anti-DACH1 antibody (*Upper*). Tumor tissues were resected from mice drinking doxycycline-supplemented (left lane) or control water (right lane). β-actin was detected as a loading control (*Lower*). (*E*) Tumor formation of xenografted DACH1-nonexpressing or -expressing U87TR-Da clone-16 cells at 20 d after inoculation. (*F*) Growth of xenografted U87TR-Da clone-16 or U87TR tumors. DOX, doxycycline (1 μg/mL); points, mean (n = 6); bars, SEM of 6 amimals on per group; N.S., not statistically significant; $^*P < 0.05$ (two-way ANOVA), statistically significant compared with doxycycline minus control. (*G*) Orthotopic xenograft model for assessing the effect of DACH1 expression on tumor progression. U87TR-Da clone-16 cells were precultured in doxycycline-free (*Left*) or doxycycline-containing (*Right*) medium. (*H*) Size of tumors of the brain stereotactically implanted with DACH1-nonexpressing (DOX-) or -expressing (DOX+) cells at 5 wk after implantation. Bars, SEM of 4 amimals per group; N.S., not statistically significant; $^*P < 0.05$ (two-way ANOVA), statistically significant compared with doxycycline minus control.

expressing U87TR-Da cells did not form any tumors (Fig. 3*E*). Because DACH1-expressing U87TR-Da cells in serum-containing DMEM did not form tumors with so few cells (2×10^4 to 1×10^5 cells), tumor-initiating cells might be enriched in spheroids of DACH1-nonexpressing U87TR-Da cells grown in serum-free NBE medium.

FGF2 Rescues DACH1-Repressed Tumorigenicity. *FGF2* expression was highly induced under serum-free culture condition, however, DACH1 repressed *FGF2* expression at low levels (Fig. 4*A*). We verified that DACH1 reduced a reporter activity of luciferase cis-regulated by *FGF2* promoter (Fig. S6*A*) and directly bound to FGF2 promoter region in DACH1-expressing U87TR-Da cells cultured in both serum-containing DMEM and serum-free NBE medium by ChIP analysis with an anti-DACH1 antibody (Fig. S6*B*). Because overexpression of *FGF2* was frequently observed in high-grade gliomas and involved in malignant progression of gliomas (23, 24) and a previous study showed that bFGF enhanced tumor-initiating spheroid formation of glioma cells (20), we examined whether spheroid formation of DACH1-expressing U87TR-Da cells was enhanced by exposure of cells to bFGF. DACH1-expressing U87TR-Da cells did not form spheroid, but exogenous bFGF-induced spheroid formation of DACH1-expressing U87TR-Da cells, indicating that bFGF, which was repressed by DACH1, increased the number of spheroid-forming tumor-initiating cells (Fig. 4*B* and Fig. S4*C*). Morphology of primary tumor spheroid, which did not express *DACH1*, was not markedly different under culture conditions with or without bFGF. However, by lentiviral expression of DACH1, the spheroid formation was partially disrupted, and most spheroid-forming cells started to differentiate even under bFGF-supplemented condition (Fig. S4*D*). We confirmed the reduced expression of a glioma stem cell marker CD133 in primary tumor-derived cells with DACH1 expression under both bFGF2-supplemented and -unsupplemented culture conditions (Fig. S4*E*). To examine whether bFGF overexpression can confer an increased tumorigenicity, the intracerebral implantation of the cells ectopically overexpressing DACH1 and FGF2 was performed (Fig. 4*C*). Tumor formation in DACH1-expressing cells with ectopic expression of FGF2 was much higher than that in control DACH1-expressing cells, showing that rescue of FGF2 repression in DACH1-expressing cells increases intracerebral tumor formation (Fig. 4 *D* and *E* and Fig. S6*C*). These results suggested that DACH1 suppress tumor formation through transcriptional repression of *FGF2*.

Discussion

Loss of chromosome 13q has been reported to occur frequently in GBMs (12, 25, 26). Homozygous deletions, LOH, and mutations in *RB1* gene, which is located at chromosome 13q14.2, 23 Mb centromeric to *DACH1*, are found in human cancers including GBMs (13, 26). In this study, we found another locus with a homozygous deletion at 13q21 by SNP genotyping array and targeted genotyping analysis with a mass spectrometer. A previous study reported LOH of chromosome 13q was more common in secondary than in primary glioblastomas (3), and the mutations of *IDH1* and *IDH2* genes have suggested to associate with development of secondary glioblastomas (27–30). However, because significant relationship between loss of *DACH1* and mutations of *IDH1* and *IDH2* were not found, we think that the mechanisms of *DACH1* loss and gliomagenesis categorized into primary or secondary glioblastoma are different (Table S2).

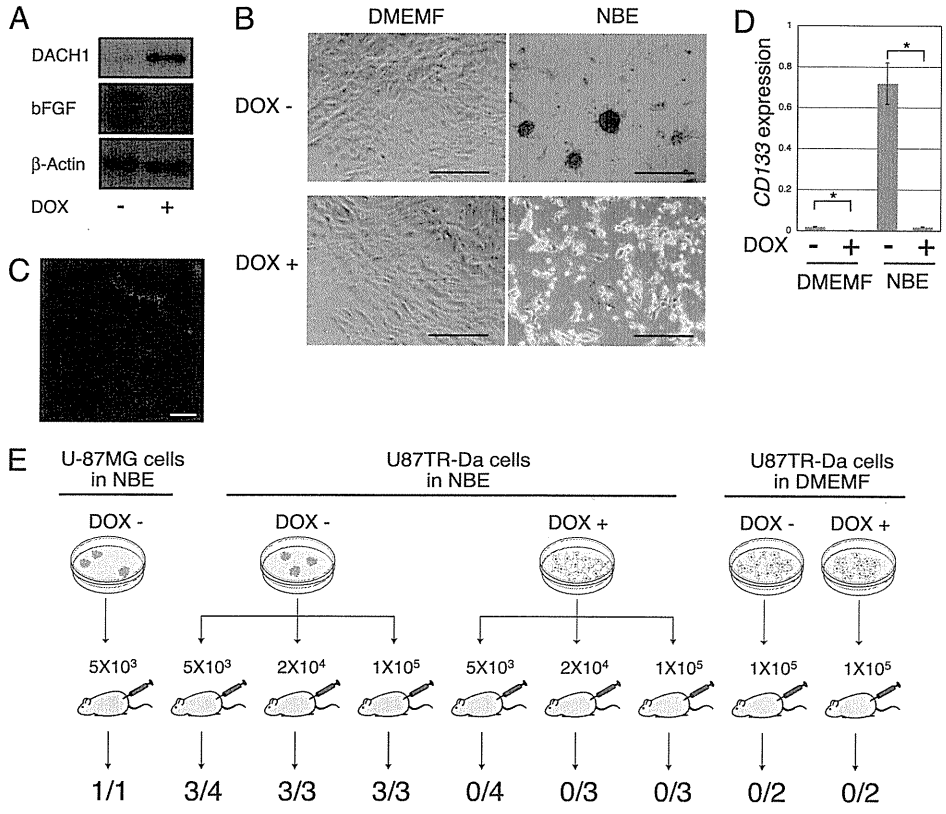

Fig. 3. Spheroid formation was inhibited by DACH1 expression. (*A*) Protein expression levels of DACH1 and bFGF in U87TR-Da clone-16 cells. (*B*) Images of cultured cells. U87TR-Da clone-16 cells were first cultured in FBS-containing DMEM in the absence (*Upper Left*) or presence (*Lower Left*) of doxycycline for 72 h, and then culture medium was replaced into serum-free NBE or serum containing DMEMF medium. Branched projection of adherent cells was observed in DACH1-nonexpressing cells (red arrows). (Scale bars: 500 μm.) (*C*) Nestin staining of spheroid (green). Nucleus was stained with DAPI (blue). (*D*) CD133 expression of DACH1-nonexpressing (DOX-) or -expressing (DOX+) cells under serum-containing (DMEMF) or serum-free (NBE) culture condition. Columns, mean ($n = 3$); bars, S.D.; *, **$P < 0.05$. (*E*) Tumor formation of DACH1-expressing or DACH1-nonexpressing U87TR-Da cells cultured in NBE medium. Tumor formation in 3 of 4 mice with 5×10^3 DACH1-nonexpressing U87TR-Da clone-16 cells and in all mice with 2×10^4 and 1×10^5 DACH1-nonexpressing U87TR-Da clone-16 cells was observed, whereas s.c. injection of DACH1-expressing U87TR-Da clone-16 cells did not form any tumors. DOX-, mice with U87TR-Da clone-16 cells cultured in doxycycline-free NBE medium; DOX+, mice with U87TR-Da cells cultured in doxycycline-supplemented medium.

DACH1-spanning region is indicated as a target of copy number variations (CNV), where healthy individuals also exhibit copy number changes. Although it is hard to detect tumor-specific small deletions by a general copy number analysis using an algorithm based on moving window, our GIM algorithm can normalize aberrant copy number change even in CNV region by using signal data of both tumor DNA and corresponding normal DNA (4, 11). We overcame CNV effect and could detect small homozygous deletions at DACH1 region by an improved algorithm for SNP genotyping array with the targeted genotyping analysis using a mass spectrometer.

To explore additional mechanisms for *DACH1* inactivation other than genomic loss, we first examined sequencing of *DACH1* exonic regions by using genomic DNA from 25 GBM tumors and 8 glioma cell lines and did not observe any somatic mutations, whereas GBM case 25 showed LOH and the remained allele with nonsynonymous polymorphism, which could be translated into amino acid-substituted DACH1, V464M (Fig. S2*C*). It should require further examination such as loss-of-function of the protein and frequency of mutational events with amino acid substitution of DACH1 protein in GBMs. We then examined methylation status of *DACH1* promoter region, because a CpG island existed in the vicinity of transcription start site (TSS) of *DACH1*. Because DNA hypermethylation at *DACH1* promoter region was frequently observed (Fig. S2*D*), hypermethylation of the *DACH1* promoter could be an additional mechanism for DACH1 inactivation in gliomagenesis.

Homozygous deletions observed in all three cases contained a deduced promoter region and the first exon with ATG start codon of *DACH1* gene, indicating the homozygous deletions cause functional disruption of DACH1. Forced expression of DACH1 decreases growth of glioma cells and inhibits formation of spheroids, which are proposed as a tumor-initiating cell population and chemoresistant population of glioma cells. Even though downstream signaling of bFGF is well described, only a few reports have provided the mechanism of transcriptional regulation of *FGF2* (31, 32). Here, we demonstrated that DACH1, a member of corepressor complexes, repressed *FGF2* expression. Campanelli et al. found the expression level of DACH1 in glial precursors was much higher than that in neural stem cells (NSCs) (33). Withdrawal of bFGF, which is known as a critical component of culture medium for human embryonic and neural stem cells (34), drives astrocytic differentiation (2, 34, 35). So we think that loss or mutation of DACH1 may interrupt astrocytic differentiation of glial-restricted progenitor cells through sustained expression of bFGF and, subsequently, trigger gliomagenesis by convertion of progenitor cells to tumor-initiating cells.

Although loss of *DACH1* was detected in some GBM specimens and glioma cell lines at genomic and/or transcriptional level, there were other glioma cell lines that exhibited detectable DACH1 expression, such as U251MG (Fig. S3*A*). When we silenced *DACH1* expression of U251MG by RNA interference (RNAi) (Fig. S7*A*), slightly increased cell proliferation was observed (Fig. S7*B*). *FGF2* and *CD133* expression was slightly upregulated when *DACH1* expression was silenced by RNAi (Fig. S7 *C* and *D*). We also observed increased spheroid formation in *DACH1*-silenced cells under serum-free culture condition (Fig. S7*E*). However, these effects of *DACH1* silencing in U251MG cells were relatively slight. Because U251MG exhibited cell-autonomous growth even with high DACH1 expression, there must be additional alterations in oncogenes/tumor suppressor genes other than *DACH1*, which might obscure the effect of *DACH1* silencing on tumor suppressing activity.

In the analysis with expression microarrays, we observed that DACH1 decreased expression of secreted factors including TGF-β2, leukemia inhibitory factor (LIF), and interleukin-6 (IL-6) (Table S1). Recent reports demonstrated that TGF-β enhanced the self-renewal property of glioma-initiating cells by activating

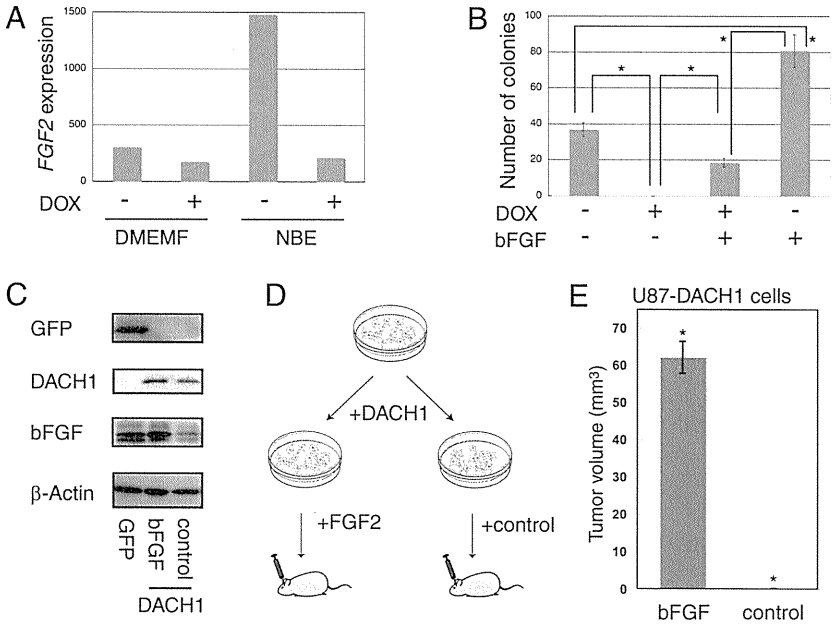

Fig. 4. Ectopic DACH1 expression decreases FGF2 expression, spheroid formation, and tumor growth. (*A*) Repressed *FGF2* expression by DACH1. DMEMF, serum-containing DMEM; NBE, serum-free Neurobasal medium. (*B*) Spheroid formation in U87TR-Da in the absence and presence of bFGF. Columns, mean (n = 3); bars, SEM of three experiments; $*P < 0.05$ (unpaired t-test), statistically significant compared with doxycycline minus control. (*C*) Protein expression of green fluorescent protein (GFP), DACH1, and bFGF. (*D*) Orthotopic xenograft model for assessing the effect of *FGF2* on tumor progression of *DACH1*-expressing cells. U87MG cells were lentivirally transduced with DACH1 and cultured in zeocin-containing medium. Then, DACH1-expressing cells were infected with lentivirus carrying FGF2 (*Left*) or control (*Right*) vector. (*E*) Size of tumors of the brain implanted with *FGF2*-expressing or control U87MG cells. Bars, SEM of 5 amimals on per group; N.S., not statistically significant; $*P < 0.05$ (two-way ANOVA), statistically significant compared with doxycycline minus control.

LIF-JAK/STAT pathway (36), and autocrine TGF-β signaling is involved in maintenance of tumorigenicity of glioma-initiating cells (19). Both LIF and IL-6 also are activators of JAK/STAT pathway. So we presumed that DACH1 may suppress tumorigenesis through repression of not only *FGF2* but also the above tumor-initiating factors. Potential tumor-initiating factors, such as bFGF, TGF-β, LIF, and IL-6, will also be good candidates drug targets for GBMs.

The previous integrated genomic analysis of glioblastoma identified genomic alterations in genes belonging not only to a variety of cellular process pathways, which were likewise altered in many types of cancers, but also to nervous system-specific cellular pathways (27). From our results, we think that *DACH1* may function as one of tumor suppressor genes regulating through nervous system-specific cellular pathways. Little is yet known about the origin of glioblastoma cells, but some reports suggested that NSCs in the subventricular zone (SVZ) of the adult brain might be a candidate origin of cancer stem cells of glioblastoma (37, 38). Although we could not detect DACH1 expression at adult SVZ, a high expression of DACH1 was observed at the ventricular wall of fetus (Fig. S2 *E* and *F*), implicating involvement of DACH1 in maturation of neural cells. Li and colleagues demonstrated murine Dach1 regulated retinogenesis and pituitary development through tissue-specific gene regulation by recruiting the corepressors (39, 40). Functional relevance of *DACH1* expression in development and neural differentiation of both invertebrates and vertebrates has been reported (39–46). DACH1-expressed embryonic and postnatal brain-derived cells displayed neural stem cell-like property (44), suggesting that DACH1 functions in neural differentiation. From these facts together, we think that DACH1 act as a guardian of differentiation in the glial lineage, and loss of DACH1 would result in dysregulation of normal differentiation and drive gliomagenesis.

Taking advantages of our allelic DNA copy number analysis for tumors and corresponding normal cells by using SNP genotyping array as well as MassARRAY, we provide evidence that DACH1 is homozygously deleted in GBMs. As far as we know, this is the first study showing loss of DACH1 gene function in tumor cells at the genomic level. Our observation supports recent studies indicating that DACH1 scarcely expressed in tumor-initiating cells and such low expression correlated with poor prognosis of breast cancers (17, 18, 47). We showed that DACH1 expression decreases proliferation of glioma cells and suppresses tumorigenicity through inhibition of bFGF-dependent spheroid formation. We here propose DACH1 is a unique tumor suppressor gene, which does not only suppresses tumor growth but also inhibits generation of tumor-initiating cells during neural differentiation. Understanding molecular basis of DACH1-mediated epigenetic regulation may provide mechanism of both neural differentiation and gliomagenesis.

Methods

Materials. All clinical samples were obtained with the informed consent of the patients after permission by the ethics committees of Tokyo University Hospital. Tumors were diagnosed according to World Health Organization classification (48).

SNP Genotyping Array. SNPs of peripheral blood cells or GBM samples were genotyped by 50K Xba SNP mapping arrays (Affymetrix) according to GeneChip Mapping 50K Assay Manual (4). Allelic and total DNA copy numbers were calculated by GIM algorithm (11).

MassARRAY Analysis. SNP genotyping of 28 GBM samples, including the 8 samples used in the initial screening by SNP mapping array analysis, was performed with MassARRAY Genotyping system (Sequenom). Paired DNA samples from blood or tumor were genotyped in duplicate. Quantitation of the peak area was performed by Sequenom's MassARRAY RT software. The threshold for LOH was defined as 40% reduction of one allele in tumor sample, as described (14).

Gene Expression Analysis. Total RNA was extracted with TRIzol (Life Technologies). Biotin-labeled cRNA was synthesized and hybridized to GeneChip U133 plus 2 (Affymetrix) according to the manufacturer's instruction. Gene expression data were normalized by using the MAS5 algorithm. Data were deposited in National Center for Biotechnology Information Gene Expression Omnibus (GEO) database (GEO accession no. GSE19678; www.ncbi.nlm.nih.gov/geo/).

Immunoblotting. An anti-DACH1 antibody was generated by immunization of GST-fused DACH1 protein and purified by affinity-purification with Immunogen-bound beads. Immunoblot was performed by using anti-DACH1 or anti–β-actin (AC-40; Sigma) antibody according to a previous report (49). Cell lysates at 3 d after the addition of doxycycline or RNAi treatment were used.

Spheroid Formation Assay. Expression of DACH1 was induced by exposure of U87TR-Da-clone16 or -clone18 cells to doxycycline for 72 h and then cells were grown in NBE medium (22, 50) composed of Neurobasal Medium (Life

Technologies), *N*-2, B-27 supplement (0.5× each; Life Technologies), human recombinant bFGF (50 ng/mL; PeproTech), and EGF (50 ng/mL; Life Technologies). At 10 d after medium change, the number of spheroid-forming colonies was counted. Paraformaldehyde-fixed spheroids were stained with an anti-Nestin monoclonal antibody (BD Biosciences) and Alexa488-conjugated anti-mouse IgG antibody (Life Technologies).

In Vivo Studies. Male 6- to 8-wk-old male athymic (nude) mice (nu/nu) of BALB/c were purchased from Charles River Laboratory Japan. For assessment of in vivo tumor growth, U87TR-Da cells were injected s.c. into the flank of mice. At the start point (day 0), drinking water was replaced with doxycycline-containing water for induction of DACH1 expression. For assessment of tumorigenicity, before injection, DACH1 expression was induced by the exposure of the U87TR-Da clone-16 cells with doxycycline for 10 d. The tumor formation was observed after s.c. injection of the serially diluted cells. For intracerebral stereotactic inoculation, cells in 6 μL of PBS were implanted into the right corpus striatum of the anesthetized mouse brain. Animal studies were performed according to institutional guidelines. See details of in vivo tumor growth and tumorigenicity assay in *SI Methods*.

Full methods and any associated references are available in *SI Methods*. Primers used were listed in Table S3.

ACKNOWLEDGMENTS. We thank Saori Kawanabe for technical assistance, Hiroko Meguro for GeneChip analysis, Kaoru Nakano for MassARRAY analysis, Dr. Junji Shibahara and Kotaro Onishi for immunohistochemistry, Drs. Aya Nonaka, Asuka Morizane, and Jun Takahashi for cell culture, and Drs. Vincent Stanton, Haruhiko Sugimura, Toru Niwa, Hiroaki Ikushima, Shogo Yamamoto, Takayuki Isagawa, and Genta Nagae for helpful comments. This work was supported by Grant-in-Aid for Young Scientists (B) 21790380 (to A.W.); and Grants-in-Aid for Scientific Research (S) 20221009 and Scientific Research on Priority Areas 17015008 from the Ministry of Education, Culture, Sports, Science and Technology-Japan, and Core Research for Evolutional Science and Technology from Japan Science and Technology (all to H.A.).

1. Furnari FB, et al. (2007) Malignant astrocytic glioma: Genetics, biology, and paths to treatment. *Genes Dev* 21:2683–2710.
2. Holland EC (2001) Gliomagenesis: Genetic alterations and mouse models. *Nat Rev Genet* 2:120–129.
3. Ohgaki H, Kleihues P (2007) Genetic pathways to primary and secondary glioblastoma. *Am J Pathol* 170:1445–1453.
4. Ishikawa S, et al. (2005) Allelic dosage analysis with genotyping microarrays. *Biochem Biophys Res Commun* 333:1309–1314.
5. Zhao X, et al. (2004) An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res* 64:3060–3071.
6. Nannya Y, et al. (2005) A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res* 65:6071–6079.
7. Midorikawa Y, et al. (2006) Molecular karyotyping of human hepatocellular carcinoma using single-nucleotide polymorphism arrays. *Oncogene* 25:5581–5590.
8. Gaasenbeek M, et al. (2006) Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex changes and multiple forms of chromosomal instability in colorectal cancers. *Cancer Res* 66: 3471–3479.
9. Midorikawa Y, et al. (2009) Allelic imbalances and homozygous deletion on 8p23.2 for stepwise progression of hepatocarcinogenesis. *Hepatology* 49:513–522.
10. Cancer Genome Atlas Research Network (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455:1061–1068.
11. Komura D, et al. (2006) Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome Res* 16:1575–1584.
12. Ueki K, et al. (1996) CDKN2/p16 or RB alterations occur in the majority of glioblastomas and are inversely correlated. *Cancer Res* 56:150–153.
13. Yin Y, Shen WH (2008) PTEN: A new guardian of the genome. *Oncogene* 27:5443–5453.
14. van Puijenbroek M, et al. (2005) Mass spectrometry-based loss of heterozygosity analysis of single-nucleotide polymorphism loci in paraffin embedded tumors using the MassEXTEND assay: Single-nucleotide polymorphism loss of heterozygosity analysis of the protein tyrosine phosphatase receptor type J in familial colorectal cancer. *J Mol Diagn* 7:623–630.
15. Tai AL, et al. (2006) High-throughput loss-of-heterozygosity study of chromosome 3p in lung cancer using single-nucleotide polymorphism markers. *Cancer Res* 66:4133–4138.
16. Miyazono K, Maeda S, Imamura T (2006) Smad transcriptional co-activators and co-repressors. *Smad Signal Transduction*, eds Dijke Pt, Heldin C-H (Springer, Dordrechet, The Netherlands), pp 277–293.
17. Wu K, et al. (2007) Cell fate determination factor DACH1 inhibits c-Jun-induced contact-independent growth. *Mol Biol Cell* 18:755–767.
18. Wu K, et al. (2006) DACH1 is a cell fate determination factor that inhibits cyclin D1 and breast tumor growth. *Mol Cell Biol* 26:7116–7129.
19. Ikushima H, et al. (2009) Autocrine TGF-beta signaling maintains tumorigenicity of glioma-initiating cells through Sry-related HMG-box factors. *Cell Stem Cell* 5:504–514.
20. Galli R, et al. (2004) Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Res* 64:7011–7021.
21. Singh SK, et al. (2003) Identification of a cancer stem cell in human brain tumors. *Cancer Res* 63:5821–5828.
22. Lee J, et al. (2006) Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell* 9:391–403.
23. Takahashi JA, et al. (1992) Correlation of basic fibroblast growth factor expression levels with the degree of malignancy and vascularity in human gliomas. *J Neurosurg* 76:792–798.
24. Morrison RS, et al. (1994) Basic fibroblast growth factor and fibroblast growth factor receptor I are implicated in the growth of human astrocytomas. *J Neurooncol* 18:207–216.
25. Wong KK, et al. (2006) Genome-wide allelic imbalance analysis of pediatric gliomas by single nucleotide polymorphic allele array. *Cancer Res* 66:11172–11178.
26. Fan X, et al. (2002) Genetic profile, PTEN mutation and therapeutic role of PTEN in glioblastomas. *Int J Oncol* 21:1141–1150.
27. Parsons DW, et al. (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science* 321:1807–1812.
28. Yan H, et al. (2009) IDH1 and IDH2 mutations in gliomas. *N Engl J Med* 360:765–773.
29. Watanabe T, Nobusawa S, Kleihues P, Ohgaki H (2009) IDH1 mutations are early events in the development of astrocytomas and oligodendrogliomas. *Am J Pathol* 174:1149–1153.
30. Zhao S, et al. (2009) Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1alpha. *Science* 324:261–265.
31. Biesiada E, Razandi M, Levin ER (1996) Egr-1 activates basic fibroblast growth factor transcription. Mechanistic implications for astrocyte proliferation. *J Biol Chem* 271: 18576–18581.
32. Xie TX, et al. (2006) Activation of stat3 in human melanoma promotes brain metastasis. *Cancer Res* 66:3188–3196.
33. Campanelli JT, et al. (2008) Expression profiling of human glial precursors. *BMC Dev Biol* 8:102.
34. Zhang SC, Wernig M, Duncan ID, Brüstle O, Thomson JA (2001) In vitro differentiation of transplantable neural precursors from human embryonic stem cells. *Nat Biotechnol* 19:1129–1133.
35. Noble M, Mayer-Pröschel M (1997) Growth factors, glia and gliomas. *J Neurooncol* 35: 193–209.
36. Peñuelas S, et al. (2009) TGF-beta increases glioma-initiating cell self-renewal through the induction of LIF in human glioblastoma. *Cancer Cell* 15:315–327.
37. Quiñones-Hinojosa A, Chaichana K (2007) The human subventricular zone: A source of new cells and a potential source of brain tumors. *Exp Neurol* 205:313–324.
38. Sanai N, Alvarez-Buylla A, Berger MS (2005) Neural stem cells and the origin of gliomas. *N Engl J Med* 353:811–822.
39. Li X, Perissi V, Liu F, Rose DW, Rosenfeld MG (2002) Tissue-specific regulation of retinal and pituitary precursor cell proliferation. *Science* 297:1180–1183.
40. Li X, et al. (2003) Eya protein phosphatase activity regulates Six1-Dach-Eya transcriptional effects in mammalian organogenesis. *Nature* 426:247–254.
41. Mardon G, Solomon NM, Rubin GM (1994) dachshund encodes a nuclear protein required for normal eye and leg development in Drosophila. *Development* 120: 3473–3486.
42. Shen W, Mardon G (1997) Ectopic eye development in Drosophila induced by directed dachshund expression. *Development* 124:45–52.
43. Lecuit T, Cohen SM (1997) Proximal-distal axis formation in the Drosophila leg. *Nature* 388:139–145.
44. Machon O, et al. (2002) Forebrain-specific promoter/enhancer D6 derived from the mouse Dach1 gene controls expression in neural stem cells. *Neuroscience* 112:951–966.
45. Heanue TA, et al. (2002) Dach1, a vertebrate homologue of Drosophila dachshund, is expressed in the developing eye and ear of both chick and mouse and is regulated independently of Pax and Eya genes. *Mech Dev* 111:75–87.
46. Jing Y, et al. (2011) In vitro differentiation of mouse embryonic stem cells into neurons of the dorsal forebrain. *Cell Mol Neurobiol* 31:715–721.
47. Wu K, et al. (2011) Cell fate determination factor Dachshund reprograms breast cancer stem cell function. *J Biol Chem* 286:2132–2142.
48. Louis DN, et al. (2007) The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* 114:97–109.
49. Watanabe A, et al. (2003) An opposing view on WWOX protein function as a tumor suppressor. *Cancer Res* 63:8629–8633.
50. Galli R, et al. (2002) Emx2 regulates the proliferation of stem cells of the adult mammalian central nervous system. *Development* 129:1633–1644.

8712–8727 Nucleic Acids Research, 2011, Vol. 39, No. 20
doi:10.1093/nar/gkr572

Published online 15 July 2011

ChIP-seq reveals cell type-specific binding patterns of BMP-specific Smads and a novel binding motif

Masato Morikawa[1], Daizo Koinuma[1], Shuichi Tsutsumi[2], Eleftheria Vasilaki[3], Yasuharu Kanki[4], Carl-Henrik Heldin[3], Hiroyuki Aburatani[2] and Kohei Miyazono[1,3,*]

[1]Department of Molecular Pathology, Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, [2]Genome Science Division, Research Center for Advanced Science and Technology, University of Tokyo, Meguro-ku, Tokyo 153-8904, Japan, [3]Ludwig Institute for Cancer Research, Box 595 Biomedical Center, SE-751 24 Uppsala, Sweden and [4]Laboratory for Systems Biology and Medicine, Research Center for Advanced Science and Technology, University of Tokyo, Meguro-ku, Tokyo 153-8904, Japan

Received May 13, 2011; Revised and Accepted June 24, 2011

ABSTRACT

Dysregulated bone morphogenetic protein (BMP) signaling in endothelial cells (ECs) and pulmonary arterial smooth muscle cells (PASMCs) are implicated in human genetic disorders. Here, we generated genome-wide maps of Smad1/5 binding sites in ECs and PASMCs. Smad1/5 preferentially bound to the region outside the promoter of known genes, and the binding was associated with target gene upregulation. Cell-selective Smad1/5 binding patterns appear to be determined mostly by cell-specific differences in baseline chromatin accessibility patterns. We identified, for the first time, a Smad1/5 binding motif in mammals, and termed GC-rich Smad binding element (GC-SBE). Several sequences in the identified GC-SBE motif had relatively weak affinity for Smad binding, and were enriched in cell type-specific Smad1/5 binding regions. We also found that both GC-SBE and the canonical SBE affect binding affinity for the Smad complex. Furthermore, we characterized EC-specific Smad1/5 target genes and found that several Notch signaling pathway-related genes were induced by BMP in ECs. Among them, a Notch ligand, JAG1 was regulated directly by Smad1/5, transactivating Notch signaling in the neighboring cells. These results provide insights into the molecular mechanism of BMP signaling and the pathogenesis of vascular lesions of certain genetic disorders, including hereditary hemorrhagic telangiectasia.

INTRODUCTION

Bone morphogenetic proteins (BMPs) are members of the transforming growth factor-β (TGF-β) family, which regulate a variety of cellular processes including differentiation, proliferation, migration and cell death in a cell type-specific and context-dependent manner (1). Perturbations of BMP signaling pathways have been implicated in a diverse set of developmental disorders, tumorigenesis and diseases including ectopic ossification and cardiovascular diseases. Mutations in *ENG*, *ACVRL1* or *SMAD4* genes have been shown to cause hereditary hemorrhagic telangiectasia (HHT) (2–4), which is a multisystemic vascular disorder characterized by epistaxis, telangiectases and arteriovenous malformation (AVM). The *ACVRL1* gene encodes an endothelial-specific type I receptor for TGF-β members, ALK-1, whose signals are transmitted through BMP-specific receptor-regulated Smads (BR-Smads; Smad1/5/8) (5). Recent work has indicated that haploinsufficiency of ALK-1 causes HHT (6). The *ENG* gene encodes Endoglin, which is an endothelial expressed co-receptor and modulates ALK-1 signaling (7). The *SMAD4* gene encodes a common mediator Smad (co-Smad), which makes a heterotrimeric complex with BR-Smads and regulate transcription of specific target genes (8). Therefore, dysregulated BMP signaling through ALK-1 in endothelial cells (ECs) is implicated in the pathogenesis of HHT. Interestingly, BMP signaling activated by BMP type I receptors, other than ALK-1 in ECs, are not able to compensate for the loss of function of ALK-1. On the other hand, aberrant BMP signaling through BMP type II receptor (encoded by *BMPR2*), especially in pulmonary arterial smooth muscle cells (PASMCs), are implicated in the pathogenesis of pulmonary arterial hypertension (PAH) (9,10). Therefore, readout of BMP signaling depends on the strength of BMP signaling, types I and II receptors and co-receptors, and cell types.

A binding sequence for BR-Smad was originally identified in *Drosophila*. Kim and colleagues (11) indicated that GCCGnCGC is a consensus binding sequence for Mad (*Drosophila* Smad1). In mammals, similar GC-rich

*To whom correspondence should be addressed. Tel: +81 3 5841 3356; Fax: +81 3 5841 3354; Email: miyazono@m.u-tokyo.ac.jp

sequences, e.g. GCCG or GGCGCC, have been evaluated in the promoter regions of well-known BMP target genes (12,13). Recombinant protein of the DNA binding domain of mouse Smad1 (Smad1 MH1) has been shown to bind to the GGCGCC sequence *in vitro* (14). This GGCGCC sequence is widely accepted as a binding sequence for BR-Smads, while a binding motif for BR-Smad has not been clearly defined in mammals.

Recent advances in microarray and sequencing technologies have made it possible to analyze global gene expression profiles and genome-wide maps of protein binding sites or epigenetic marks (15). Two groups have reported genome-wide analyses of BR-Smads in mouse ES cells (mESCs) using chromatin immunoprecipitation (ChIP) coupled with promoter array (ChIP-chip) and ChIP followed by sequencing (ChIP-seq) analyses (16,17). Through profiling of the global binding sites of 13 transcription factors and 2 transcription regulators in mESCs, Chen and colleagues (16) hypothesized that Smad1 makes an enhancer complex with Sox2-Oct4 (also known as Pou5F1), which defines ES-specific binding patterns of Smads. However, it has not been clarified whether a transcription factor complex, or an enhancer complex, determines the cell type-specific binding patterns of Smads in other cell types.

Here, we performed ChIP-seq to map Smad1/5 occupancy at high resolution in two different primary human cells treated with several BMP isoforms; human umbilical vein endothelial cells (HUVECs) with BMP-9 or BMP-6 and PASMCs with BMP-4. Smad1/5 preferentially bound to the region outside the promoter of known genes, and their binding was associated with upregulation of target genes. In HUVECs, Smad1/5 binding regions overlapped with reported enhancer regions. Comparison of HUVECs and PASMCs revealed that about 20% of the binding regions were overlapped. In contrast, most of the Smad1/5 binding sites in HUVECs treated with BMP-6 overlapped with those with BMP-9, especially in the regions with higher affinity for Smads. Cell-selective Smad1/5 binding patterns appear to be determined mostly by cell-specific differences in baseline chromatin accessibility patterns. In addition, a Smad1/5 binding motif was identified and termed a GC-rich Smad Binding Element (GC-SBE). Interestingly, GGAGCC sequence was enriched in the HUVEC- or PASMC-specific Smad1/5 binding regions compared with the GGCGCC sequence. We revealed that mutations of GC-SBE affected binding of Smad complex in a cell type-specific manner. Furthermore, we characterized EC-specific Smad1/5 target genes and found that several Notch signaling pathway-related genes were induced in ECs. Among them, a Notch ligand, JAG1 was regulated directly by Smad1/5, transactivating Notch signaling in the neighboring cells. These results provide insights into the molecular mechanism of BMP signaling and the pathogenesis of vascular lesions of HHT.

MATERIALS AND METHODS

Cell culture

HEK293T, HepG2 and HeLa cells were obtained from the American Type Culture Collection (ATCC). HUVECs and PASMCs were obtained from Lonza. HMEC-1, an immortalized human dermal microvascular EC line, was obtained from Dr T. Lawley (Emory University, Atlanta, GA, USA). 293T, HepG2 and HeLa cells were maintained in Dulbecco's modified Eagle's medium (Gibco), supplemented with 10% (v/v) fetal bovine serum (FBS) (HyClone) and 1% penicillin–streptomycin (Gibco). HUVECs and HMEC-1 were cultured in EGM-2 medium (Lonza). PASMCs were cultured in SmGM-2 (Lonza).

Reagents and antibodies

Recombinant human BMP-4, BMP-6 and BMP-9 were purchased from R&D Systems. TNF-α was from PeproTech. Cycloheximide (CHX) was purchased from Sigma-Aldrich.

The following antibodies were used: anti-Flag M2 (Sigma-Aldrich), anti-α-tubulin (AC-15; Sigma-Aldrich), anti-HDAC-1 (2E10; Upstate Millipore), anti-Smad1 (Bio Matrix Research, Chiba, Japan), which recognizes both Smad1 and Smad5 for ChIP procedure, anti-Smad1/5/8 (N-18; Santa Cruz Biotechnology) for western blotting, anti-phospho-Smad1/5 (Cell Signaling Technology), anti-phospho-Smad1/5/8 (Cell Signaling) and anti-JAG1 (H-114; Santa Cruz).

ChIP

Chromatin isolation, sonication and immunoprecipitation (IP) using anti-Smad1/5 antibody were performed essentially as described (18).

ChIP-sequencing (ChIP-seq) and data analysis

High-throughput sequencing of the ChIP fragments was performed using Illumina Genome Analyzer (Illumina) following the manufacturer's protocols. One flow cell lane was used for sequencing of each pooled sample. Unfiltered 36 bp sequence reads were aligned against the human reference genome (NCBI Build 36, hg18) using ELAND (Illumina). Peaks were called using CisGenome v1.2 (19) by two-sample analysis, where input genomic DNA was used as a negative control (Supplementary Table S1). Assigning a binding site to the nearest gene within 100 kb from a peak was performed using CisGenome.

A set of random genomic control regions for 3750 Smad1/5 binding regions was generated by randomly picking up the same number of 301 bp chromosome-matched sequences. In order to calculate the frequency of transcription factor binding site (TFBS)-positive sequences, MATCH score of position-specific scoring matrix (PSSM) for each transcription factor was computed. The highest MATCH score (HMS) was assigned to each sequence, and the number of sequences with HMS greater than or equal to a threshold was counted. For obtaining background data against those of 3750 Smad1/5 binding regions, the chromosome-matched sequences were generated randomly for 1000 times. The distribution of HMSs in 1000 sets of 3750 sequences was used as background control for each PSSM. The threshold was set to the mode of the distribution of HMSs. PSSMs

were obtained from JASPAR database (20). A set of non-overlapping matched genomic control sequences was generated by CisGenome. The frequency of TFBSs in these sequences (motif counts per set of sequences) was computed with the likelihood ratio ≥500 (default value of CisGenome) or 200 (for shorter motifs such as MEME4). Fifty sets of matched control sequences were used as background data. Mapping of TFBSs to the specific genomic regions were calculated by the CisGenome.

The sequences of Smad1/5 binding sites were input to the MEME (21) with several options: mod = oops, nmotifs = 5, minw = 6, maxw = 8, revcomp and other settings default. The logo plots were generated using the seqLogo package in R (http://bioconductor.org/packages/2.6/bioc/html/seqLogo.html). Enriched binding motifs were also obtained from the Cis-regulatory Element Annotation System (CEAS) website as described (http://ceas.cbi.pku.edu.cn/index.html) (22,23). Overrepresented gene ontology (GO) categories for genes associated with Smad1/5 binding regions were determined using the Database for Annotation, Visualization and Integrated Discovery (DAVID v6.7; http://david.abcc.ncifcrf.gov) (24).

ChIP and quantitative-PCR

The real-time PCR was conducted as described (23). Primer sequences are given in Supplementary Table S2 in the Supplementary Data. The amount of immunoprecipitated DNA was calculated relative to the input.

RNA isolation, quantitative real-time reverse transcription–PCR and conventional RT–PCR

Extraction of total RNA, qRT–PCR and conventional RT–PCR were performed as described (23). Primer sequences are given in Supplementary Table S2.

Gene expression profiling

HUVECs and PASMCs were serum starved overnight and treated with or without BMP-9 (1 ng/ml), BMP-6 or BMP-4 (50 ng/ml) treatment for 2 or 24 h. Gene expression profiling was performed with a GeneChip Human Genome U133 Plus 2.0 Array (Affymetrix) as described (18). The 8544 and 8067 genes, whose signal intensity exceeded 100 at any time point were considered to be expressed and functional in HUVECs and PASMCs, respectively. The heatmaps were produced using the heatmap.2 function from the gplots library in R (http://cran.r-project.org/web/packages/gplots/).

Histone modification data

Genome-wide histone modification map for monomethylation of histone H3 lysine 4 (H3K4me1), trimethylation of histone H3 lysine 4 (H3K4me3) and acetylation of histone H3 lysine 27 (H3K27ac) of HUVECs were produced and released from the ENCODE Project (25) and were downloaded from UCSC (http://hgdownload.cse.ucsc.edu/goldenPath/hg18/encodeDCC/wgEncodeBroadChipSeq/).

Plasmid construction

FLAG-tagged Smad constructs were previously described (18). Each fragment of Smad1/5 binding regions was amplified from human genomic DNA by PCR, cloned into a modified pGL4.10 reporter plasmid (Promega) driven by minimal adenoviral major late promoter (MLP) (12). A point mutation was introduced by site-directed mutagenesis using PCR with specific primers. A reporter construct with six multimerized CTGGAGCC sequence (pGL4-6xGC-SBE-Luc) was constructed as follows. A fragment with one copy of the CTGGAGCC sequence was cloned into the pcDNA3.1 vector (Invitrogen). The sequences of the oligonucleotides were 5′-AGATCTTCGAACAGCTC TGGAGCCAGATGGCCTGGATCC-3′ (sense) and 5′-G GATCCAGGCCATCTGGCTCCAGAGCTGTTCGAA GATCT-3′ (antisense). This fragment was multimerized in tandem, and the fragment containing six tandem copies was subcloned into the modified pGL4-MLP plasmid. Six multimerized dimeric CBF1/Suppressor of Hairless/Lag1 (CSL) binding sites with Epstein-Barr virus (EBV) TP1 promoter sequence of the pGA981-6 (26) was transferred into pGL4.10 reporter plasmid (pGL4-12xCSL-Luc) and used for Notch reporter assay. A plasmid encoding GST-hSmad1-MH1 was constructed by PCR amplification of the MH1 domain of human Smad1 (1–143 amino acid). The fragment was subcloned into pGEX-6P-1 vector (GE Healthcare, Chalfont St Giles, UK). All constructs were DNA sequence verified.

Protein production and purification

The bacterially expressed GST fusion proteins containing residues for human $Smad1_{1-143}$ (GST-hSmad1-MH1) were purified with Glutathione Sepharose 4B beads (GE Healthcare) followed by cleavage with PreScission Protease (GE Healthcare) at 4°C overnight according to the recommendations of the manufacturer. The concentration of the protein was measured by BCA Protein Assay Kit (Pierce).

Electrophoretic mobility shift assays

Electrophoretic mobility shift assays (EMSA) was conducted essentially as described previously (14) and detected with LightShift Chemiluminescent EMSA kit (Pierce). The sequences of the DNA oligos are provided in Supplementary Table S2.

Lentiviral infection and luciferase assays

Since transfection efficacy is very low in HUVECs and it is rather toxic, we adapted lentiviral expression system. pGL4 constructs were subcloned between EcoRI and XhoI sites of the lentiviral vector construct CS-CDF-CG-PRE. Recombinant lentiviral vectors were generated as reported previously (23).

Stably expressing cells were stimulated with indicated doses of BMP-9 or BMP-6, and then they were harvested and assayed for luciferase activity at 12 h after stimulation. Luciferase activities of the cell lysates were determined using the Dual-luciferase Reporter Assay System (Promega).

Transient transfection and dual-luciferase assays

Transient transfection was carried out using FuGENE 6 (Roche) for HEK293T cells, Lipofectamine 2000 (Invitrogen) for HepG2 cells and Lipofectamine LTX (Invitrogen) for HMEC-1 cells and PASMCs according to the recommendations of the manufacturer.

Cells were transiently transfected with 1 μg of the luciferase reporter constructs along with 0.05 μg of Renilla luciferase reporter vector pGL4.74[hRluc/TK] (Promega) as an internal control. For HMEC-1 and PASMCs, the medium was changed at 3 h after transfection. Cells were stimulated with 1 ng/ml BMP-9 (HMEC-1), 50 ng/ml BMP-6 (HepG2) or 30 ng/ml BMP-4 (PASMCs) at 24 h after transfection, and then they were harvested and assayed for luciferase activity at 16 h after stimulation.

RNA interference

Duplexes of small interfering RNA (siRNA) against human Smad4 (D-003902-05) were synthesized by Dharmacon (Thermo Fisher Scientific), and were transfected using Lipofectamine RNAiMAX (Invitrogen) according to the recommendations of the manufacturer. The final concentration of siRNA in the culture media was 10 nM. At 24 h after transfection, cells were serum starved overnight, treated with or without BMP-9 for 2 or 24 h and subjected to qRT–PCR.

Western blotting

Western blotting was performed essentially as described (18). Cytoplasmic and nuclear fractions were isolated using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Pierce) according to the recommendations of the manufacturer.

Immunofluorescence microscopy

HUVECs were treated with 1 ng/ml BMP-9 for 24 h, fixed in 10% formalin for 20 min and incubated overnight at 4°C with primary antibodies (JAG1, 1:100 dilution) diluted in Blocking One solution (Nacalai Tesque, Kyoto, Japan). The cells were washed with PBST (PBS with 0.1% Triton X-100), and then incubated with secondary antibodies (Alexa Fluor 488 goat anti-rabbit IgG, Invitrogen, 1:500 dilution) for 2 h and TOTO-3 (Invitrogen) for 10 min at room temperature. Images were obtained with a Zeiss LSM 510 Meta confocal microscope (Carl Zeiss).

Transactivation (coculture) Notch assay

One day prior to transfection, HeLa cells were seeded at a density of 5.0×10^4 cells per well in 12-well plate. Cells were transiently transfected with 1 μg of the pGL4-12xCSL-Luc reporter construct along with 0.05 μg of pGL4.74[hRluc/TK] (Promega) using Lipofectamine 2000 (Invitrogen). After 16 h transfection, medium was changed to 1:1 mixture of DMEM and EGM-2 and then 1.0×10^5 HUVECs were added. After adhesion of HUVECs (about 2 h later), cells were treated with or without 5 ng/ml BMP-9 for 24 h, and subjected to luciferase assay.

Statistical analysis

The difference between experimental groups of equal variance was analyzed using Student's *t*-test with $P < 0.05$ being considered significant. All experiments were performed at least three times independently and similar results were obtained.

RESULTS

Genome-wide identification and characterization of Smad1/5 binding sites in HUVECs and PASMCs

We performed ChIP-seq analyses using HUVECs stimulated with BMP-9 (1 ng/ml) or BMP-6 (50 ng/ml) and PASMCs treated with BMP-4 (50 ng/ml). Doses of the ligands for HUVECs were determined based on the phosphorylation status of BR-Smads and the physiological range of the circulating ligands. BMP-9 has been identified as a major circulating ligand for ALK-1 (27). Serum concentration of BMP-9 ranges from 1 to 12 ng/ml, which is enough for full activation of ALK-1 (Supplementary Figure S1A) (28,29). Thus, it is thought to play important roles in the control of vascular quiescence. BMP-6 transduces its signal mainly through the BMP type I receptor ALK-2 (encoded by *ACVR1*), which is also a receptor for BMP-9 (1). Notably, BMP-6 exists in FBS at concentrations of 2–10 ng/ml (29), and BMP-6 has been reported to activate ECs (30). However, 10 ng/ml was not enough to activate Smad1/5 in HUVECs (Supplementary Figure S1A). We selected a BMP-6 concentration of 50 ng/ml for our experiments, which gave an equivalent induction of *ID1* mRNA (Supplementary Figure S1B), and almost as high phospho-Smad1/5/8 level in the nuclear fraction as stimulation with 1 ng/ml BMP-9 (Supplementary Figure S1C). We also confirmed that 50 ng/ml BMP-4 was enough for full activation in PASMCs (Supplementary Figure S1D). Both HUVECs and PASMCs expressed Smad1 and Smad5 (Supplementary Figure S1E).

The anti-Smad1/5 antibody worked efficiently in IP under formalin-fixed condition (Supplementary Figure S1F and S1G). Human genomic DNA sequences that corresponded to known BMP responsive elements (BREs) in mouse *Id1* (12) and mouse *Hey1* (13) promoters were used as positive control regions. In HUVECs, a comparable enrichment of Smad1/5 was confirmed at the *ID1* promoter after BMP-6 or BMP-9 treatment, while weak Smad1/5 binding was observed at the *HEY1* promoter after BMP-6 stimulation (Supplementary Figure S1H). Since maximal Smad1/5 binding was observed at 1.5 h after stimulation with BMPs, we adopted this time of stimulation for ChIP-seq analyses, the same stimulation time that was used in similar studies of Smad2/3 (18) and Smad4 (31).

The ChIP DNA and the control input DNA were then submitted to high-throughput sequencing analyses. The enriched genomic regions were determined using CisGenome (19). Using a false discovery rate (FDR) cut off of 0.1, a total of 3750 Smad1/5 binding regions were identified in the ChIP-seq data of HUVECs treated with BMP-9, 880 in HUVECs treated with BMP-6 and 2745 in

PASMCs treated with BMP-4 (Figure 1A and B; Supplementary Figure S1I). To validate the results, BMP-9-dependent Smad1/5 enrichment was confirmed by ChIP–qPCR at 20 novel Smad1/5 binding regions of variable peak intensity (Figure 1C). The ChIP-seq peaks were annotated to a total of 2179, 563 and 1,609 genes, respectively (Supplementary Table S1). Approximately 30% of the binding sites were located in the introns of known genes and 20% in the promoter regions within 10 kb upstream of known transcription start sites (TSSs) (Figure 1D). These Smad1/5 binding regions were highly conserved among multiple vertebrate species (Supplementary Figure S1J).

Comparison of the three ChIP-seq data revealed that ~20% of Smad1/5 binding regions overlapped between HUVECs and PASMCs (Figure 2A), while most of the Smad1/5 binding sites in HUVECs after BMP-6 stimulation overlapped with those after BMP-9 stimulation, especially in the higher ranked peaks (Figure 2A and B). Common Smad1/5 binding sites shared with HUVECs treated with BMP-9 and those with BMP-6, including those at *ID1* and *ID3* loci, had comparable levels of Smad1/5 binding, suggesting that these sites had higher affinity for Smad1/5, while the BMP-9 specific sites (Figure 1A and B and Supplementary Figure S1I) had weaker affinity. In line with this hypothesis, increasing concentrations of BMP-6 dose dependently enhanced the Smad1/5 binding to the BMP-9 specific binding sites, e.g. at *HEY1* and *JAG1* loci (Figure 2C), whereas common Smad1/5 binding sites, e.g. at *ID1* and *ENG* loci, had enough enrichment when stimulated with only 20 ng/ml BMP-6 (Figure 2C). These dose response data also indicated that 50 ng/ml BMP-6 was not enough to achieve Smad1/5 binding to target sites with relatively lower affinity for Smad1/5 in ECs.

Smad1/5 bind to enhancer regions already accessible in specific cell types

To investigate the biological functions associated with Smad1/5 binding in HUVECs and PASMCs, the significance of functional annotation clustering of the GO of the genes related to Smad1/5 binding was assessed using DAVID (24). This analysis showed that the highest enriched GO category of biological function for HUVEC-specific genes with BMP-9 stimulation was related to blood vessel development, while that for PASMC-specific genes with BMP-4 stimulation was related to extracellular matrix production (Figure 3A). Thus, Smad1/5 bind to different sets of target sites in different cell types, which may be related to the cell type-specific function.

In order to identify the cell type-specific binding mechanism for Smad complex, we sought known binding motifs enriched in the Smad1/5 binding regions using the CEAS website (22). Interestingly, ETS, AP-1, AP-2 and SP-1 binding sites were enriched in the Smad1/5 binding regions in both HUVECs and PASMCs, while other motifs occurred only in a small proportion of sequences analyzed (Supplementary Table S3). We also conducted *de novo* motif prediction in order to find overrepresented motifs in the HUVEC- and PASMC-specific Smad1/5 binding regions using MEME (21). Obtained motifs were then compared with TRANSFAC (32) and JASPAR (20) database of known motifs, and ranked by their similarity using the TOMTOM program (33). The predicted motifs were similar to the ones identified by CEAS (Supplementary Figure S2). These results suggest that these transcription factors do not determine the cell type-specific BR-Smad binding pattern.

To evaluate the association between Smad1/5 binding and gene expression regulation, expression microarray analyses were performed at several time points (0, 2 and 24 h). We confirmed an equivalent induction of ID proteins after BMP-6 or BMP-9 stimulation in HUVECs (Supplementary Figure S3A). Combining the mapping data with gene expression profiles revealed that Smad1/5 binding regions were enriched in early upregulated genes (Figure 3B and C). Notably, in HUVECs treated with BMP-9, 108 genes were upregulated and 37 were downregulated more than 2-fold in early phase, and 70 of the 108 upregulated genes (64%) and 9 of the 37 downregulated genes (24%) were associated with Smad1/5 binding regions (Supplementary Figure S3B). We consider these 70 upregulated genes (corresponding to 170 binding sites) as putative direct target genes of ALK-1 in ECs (Supplementary Table S4). We also identified 19 putative direct target genes in PASMCs using the same criteria (Supplementary Table S4).

Smad1/5 binding regions in HUVECs were further characterized using differential histone modification marks, which were produced and released from the ENCODE Project (25). H3K4me3 is associated with promoters and H3K4me1 is preferentially associated with enhancers. H3K27ac is associated with active regulatory regions (34). As many as 3651 Smad1/5 binding peaks (97.4%) overlapped with H3K4me1 or H3K4me3 regions of HUVECs. Among them, 3201 Smad1/5 binding peaks (85.4%) overlapped with enhancer regions, characterized with both H3K4me1 and H3K27ac (Figure 3D and Supplementary Figure S3C) (34). In PASMCs, 86.5% (724/837) of common Smad1/5 binding peaks shared with HUVECs and PASMCs overlapped with enhancer regions of HUVECs characterized with both H3K4me1 and H3K27ac, while only 54.3% (1036/1908) of PASMC-specific peaks overlapped with endothelial enhancers (Figure 3D). Thus, these data also suggest that Smad1/5 preferentially bind to enhancer regions already accessible in specific cell types.

GC-SBE is a direct binding motif for Smad1/5

To elucidate a specific binding motif in Smad1/5 binding regions, a *de novo* motif prediction was performed using MEME (21). Since ChIP experiments may detect indirect Smad1/5-DNA binding through protein–protein interaction, we focused on the 170 Smad1/5 binding regions of BMP-9 target genes in HUVECs. Five overrepresented motifs were identified and designated as MEME1–5 (Figure 4A and Supplementary Figure S4A). These motifs were validated in three ways. First, the fold

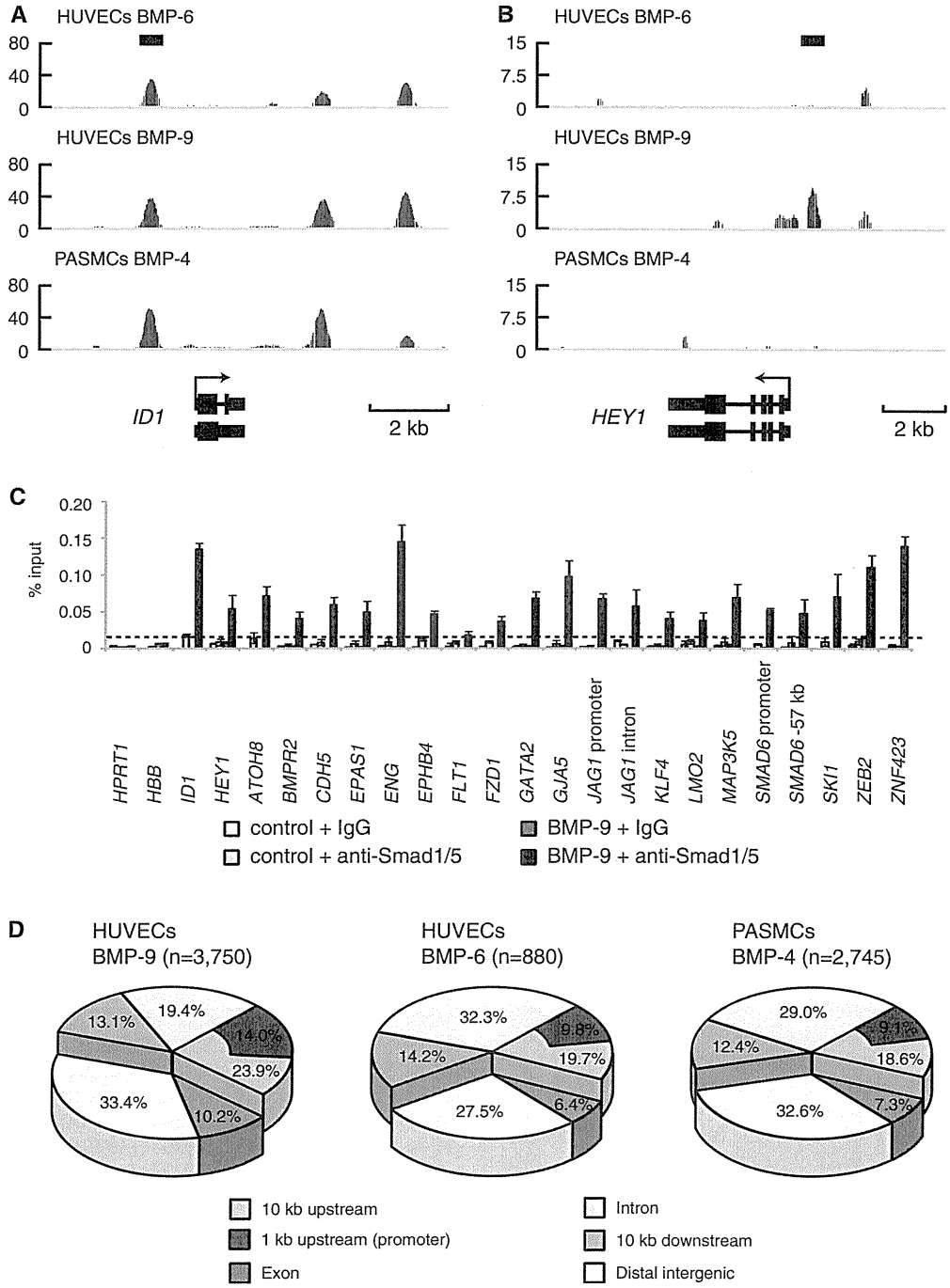

Figure 1. Genome-wide identification and characterization of Smad1/5 binding sites in HUVECs. (**A** and **B**) Genomic loci of *ID1* (A) and *HEY1* (B) are shown together with the results of ChIP-seq (in red). The direction of transcription is shown by the arrow beginning at the TSS. Horizontal black bars represent the positions of previously reported Smad1/5/8 binding regions. (**C**) HUVECs were stimulated with 1 ng/ml BMP-9 for 1.5 h and subjected to ChIP assays with anti-Smad1/5 antibody or control IgG. The ChIP samples were quantified by real-time PCR with locus specific primers and normalized to input DNA. The dashed line indicates 0.01% of input. The data are the mean of triplicate values ± SD. (**D**) Genome-wide location summary of Smad1/5 binding regions relative to known genes in human genome (hg18). Ten kilo base pairs upstream and downstream regions are defined as ≤10 kb upstream from the TSS or ≤10 kb downstream from the transcription end site (TES) of a gene, respectively. Distal intergenic refers to all locations outside the intragenic and the 10 kb flanking regions.