Supplementary Figure Legends

Supplementary Figure S1. Genome-wide identification and characterization of Smad1/5-binding sites in HUVECs

(A) Dose response study of BR-Smad phosphorylation after BMP-6 or BMP-9 stimulation in HUVECs. HUVECs were starved overnight and stimulated with the indicated concentration of BMP-6 or BMP-9 for 1 h. The cells were then subjected to immunoblot analysis to determine the phospho-Smad1/5 (pSmad1/5) protein level. α-tubulin was used as a loading control.

(B) Comparison of induction of ID1 after BMP-6 or BMP-9 stimulation. HUVECs were starved overnight and stimulated with BMP-9 (1 ng/ml) or BMP-6 (50 ng/ml) for indicated time periods. The cells were then subjected to quantitative RT-PCR analysis for ID1. Values were normalized to the amount of housekeeping GAPDH mRNA. The data are the mean of triplicate values ± standard deviation.

(C) Comparison of protein level of phospho-Smad1/5/8 after BMP-6 or BMP-9 stimulation. HUVECs were starved overnight and stimulated with BMP-9 (1 ng/ml) or BMP-6 (50 ng/ml) for 1 h. The cells were then fractionated into nuclear and cytosolic fractions and subjected to immunoblot analysis to determine the phospho-Smad1/5/8 (pSmad1/5/8) protein level. Separation of the fractions was verified by immunoblotting of histone deacetylase 1 (HDAC-1) (nuclear marker) and α-tubulin (cytosolic marker), which are also served as a loading control. TCL indicates total cell lysates.

(D) Dose response study of BR-Smads phosphorylation after BMP-4 stimulation in PASMCs. PASMCs were starved overnight and stimulated with the indicated concentration of BMP-4 for 1 h. The cells were then subjected to immunoblot analysis to
determine the phospho-Smad1/5/8 (pSmad1/5/8) protein level. α-tubulin was used as a loading control.

(E) Conventional RT-PCR analysis of BR-Smad genes. RNA samples from HUVECs, PASMCs, HepG2, and HMEC-1 cells were analyzed by conventional RT-PCR for expression of BR-Smad genes. Primers used for SMAD1, SMAD5, SMAD8, and SMAD4 specifically detect the endogenous transcripts. GAPDH was used as a loading control. H2O indicates no template control reactions, where the cDNA was substituted with nuclease-free water.

(F) Validation of anti-Smad1/5 antibody used for chromatin immunoprecipitation (ChIP) procedure. HEK293T cells were transfected with FLAG tagged-Smad constructs and subjected to immunoblot analysis with anti-Smad1/5 antibody or anti-FLAG (M2) antibody.

(G) Immunoprecipitation (IP) was performed with the same procedure to chromatin immunoprecipitation (ChIP) experiments. Briefly, 1.5x10^6 cells were starved overnight and stimulated with BMP-9 (1 ng/ml) for 1.5 h, and then cross-linked with 1% formaldehyde. The cross-linked cells were harvested by scraping, pelleted, resuspended in 1 ml of lysis buffer, and sonicated. Protein-DNA complex were immunoprecipitated for 8 h by anti-mouse IgG Dynabeads preincubated with 5 µg of anti-Smad1/5 antibody. The beads were then washed five times with RIPA buffer and once with TE buffer. Proteins were eluted using SDS sample buffer and subjected to immunoblot analysis with anti-Smad1/5/8 antibody.

(H) Time course study of Smad1/5 binding to ID1, HEY1 locus or HPRT1 1st intron. HUVECs were starved overnight and stimulated with BMP-9 (1 ng/ml) or BMP-6 (50
ng/ml) for indicated time periods and subjected to ChIP assays with anti-Smad1/5 antibody. The ChIP samples were quantified by real-time PCR with locus specific primers (see Table S4) and normalized to input DNA. The dashed line indicates 0.01% of input. The data are the mean of triplicate values ± standard deviation.

(I) Visualization of several Smad1/5 binding regions. Genomic loci are shown together with the results of ChIP-seq. Red peaks represent ChIP regions (top panel). The direction of transcription is shown by the arrow beginning at the transcription start site (TSS). Horizontal blue bars represent the positions of PCR primers used in this study.

(J) Conservation plots of Smad1/5 binding regions. The mean phastCons conservation score was calculated by CisGenome and was plotted against the distance from peak summits (within 1.5 kb from peak summit).

**Supplementary Figure S2. De novo motif search for overrepresented motifs in Smad1/5 binding regions in different cell types.**

(A-C) Results of de novo motif discovery performed on 815 common Smad1/5 binding sites shared with HUVECs and PASMCs (A), 2,935 HUVEC-specific binding sites (B), and 1,908 PASMC-specific binding sites (C). For each predicted motif, matched motifs in the database are indicated with TOMTOM $p$-value.

**Supplementary Figure S3. Genome-wide analysis of the relation between Smad1/5 binding and gene expression profiles.**

(A) Graphical representation of microarray expression data of ID proteins after BMP-6 or BMP-9 stimulation.
(B) Graphical presentation of the number of up- or down-regulated genes associated with Smad1/5 binding regions. The number of the genes up- or down-regulated more than 2 fold relative to time 0 are graphically displayed (red and blue, respectively). The genes associated with Smad1/5 binding regions are also represented (light red or light blue, respectively).

(C) A screenshot of chromosome 2 with DNA-binding profiles of Smad1/5 treated with BMP-9 and histone modification markers. RefSeq Genes track (Green) shows known protein-coding genes taken from the NCBI mRNA reference sequences collection (RefSeq). The red peaks represent Smad1/5 binding, and blue represent histone modification markers. The data for H3K4me1 and H3K27ac are derived from ENCODE projects. Each vertical black line above the peak represents the regions with signal intensity exceeded the threshold value. Areas where Smad1/5 binding regions are overlapped with the H3K4me1-positive or the H3K27ac-positive regions are also indicated as black vertical lines (‘overlap’ track).

**Supplementary Figure S4. De novo prediction of Smad1/5 binding motif.**

(A) The sequences of 170 Smad1/5 binding regions associated with putative direct target genes were analyzed for overrepresented motifs using MEME. The top 5 motifs selected by MEME software are named as MEME1 to MEME5 and displayed as sequence logos. The sequence logo of MEME2 is shown in Figure 4A.

(B) TFBS enrichment calculated against chromosome-matched control sequences. Enrichment of transcription factor binding site (TFBS) in Smad1/5 binding regions was calculated. For obtaining background data, chromosome-matched sequences were
generated randomly for 1,000 times. In order to calculate the frequency of 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. Data are given as boxplot. The circles represent outlier values. The black circles indicate the number of matched motifs observed in the Smad1/5 binding regions.

(C) MEME2 motif occurs in about 45% of all Smad1/5 binding regions in HUVECs treated with BMP-6 and PASMCs treated with BMP-4.

(D) Fraction of peaks with at least one MEME2 motif. ChIP-seq peaks of Smad1/5 in HUVECs treated with BMP-9 are ranked by peak height. Fraction of peaks with at least one MEME2 motif is calculated for every 100 peak and plotted. Peaks with motifs are enriched in the high ranked peaks.

(E) Distribution of MEME2 motif around the peak summits. (Upper) Schematic presentation of the analysis. The MEME2 motif was mapped in regions within 500 bp from the peak summits using CisGenome with the likelihood ratio more than 500 (default value). Data from all peaks were accumulated and presented as a line histogram. (Lower) The number of the MEME2 motif around the peak summits was counted and plotted in a 7-bp sliding window against the distance from the summits (within 500 bp from the summits) (blue). The motifs closest to the summits are located within 100 bp from the peak summit (1st motif; green). Five separate matched control regions were randomly chosen by CisGenome and used as a control. The number of the MEME2 motif in those regions was counted and the average was plotted (red).
(F) Distribution of identified motifs around the peak summits. The number of each identified motif is counted and plotted in a 7-bp sliding window against the distance from peak summits (within 500 bp from peak summits), as in Figure 4D.

**Supplementary Figure S5. Validation of GGAGCC sequence as a novel BMP responsive element.**

(A) Schematic presentation of the enhancer sequences and luciferase reporter constructs. Smad1/5 binding regions, which are located in BMPR2 intron 3 and the JAG1 promoter, contain GGAGCC. These regions were subcloned into pGL4 vector with adenoviral major late promoter (MLP). Reporter constructs named SBEmut contain mutation in SBE. (B) pGL4-JAG1 reporter constructs were introduced into HUVECs using lentiviral vector system in order to evaluate their enhancer activity. The 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. The data are the mean of triplicate values ± standard deviation. (C) PASMCs were transfected with indicated reporter constructs to evaluate their enhancer activity. The cells were stimulated with BMP-4 (30 ng/ml) at 24 h after transfection, and then they were harvested and assayed for luciferase activity at 16 h after stimulation. The data are the mean of triplicate values ± standard deviation. (D and E) HMEC-1 cells (D) and HepG2 cells (E) were transfected with indicated reporter constructs to evaluate their enhancer activity. The cells were stimulated with BMP-9 (1 ng/ml) for HMEC-1 or BMP-6 (50 ng/ml) for HepG2 at 24 h after transfection,
and then they were harvested and assayed for luciferase activity at 16 h after stimulation. The data are the mean of triplicate values ± standard deviation.

**Supplementary Figure S6. The Notch ligand JAG1 is a direct target gene of Smad1/5.**

(A) HUVECs were starved overnight and stimulated with 1 ng/ml BMP-9 for 2 h. To access whether activation of \( JAG1 \) by BMP-9 was independent of \textit{de novo} protein synthesis, cells were treated with cycloheximide (CHX, 10 \( \mu \)M) for 30 min before the treatment of BMP-9. Cells were then subjected to quantitative RT-PCR analysis for \( JAG1 \). Values were normalized to the amount of housekeeping \( GAPDH \) mRNA. The data are the mean of triplicate values ± standard deviation.

(B) HUVECs were transfected with siRNA against Smad4 or no template control (NTC). Next day, cells were starved overnight, stimulated with 1 ng/ml BMP-9 for 2 h and subjected to quantitative RT-PCR analysis for \( JAG1 \). Values were normalized to the amount of housekeeping \( GAPDH \) mRNA. The data are the mean of triplicate values ± standard deviation.
Supplementary Figure S1 (Continued)
### A. Common Smad1/5 binding sites shared with HUVECs and PASMCs (n=815)

| motif | p-value |
|-------|---------|
| motif1 | M00339 (c-Ets-1) 1.3E-06  
M00340 (c-Ets-2) 1.4E-06  
M00971 (Ets) 2.9E-06 |
| motif2 | M00005 (AP-4) 2.8E-04  
M00983 (MAF) 5.4E-04  
M00199 (AP-1) 5.4E-04 |
| motif3 | M00629 (Eve) 4.5E-05 |
| motif4 | M01207 (ETS2) 9.1E-04  
M00743 (c-Ets-1) 1.6E-03  
M00321 (Muscle) 2.5E-03 |
| motif5 | M001295 (ATF5) 3.3E-04  
M00649 (MAZ) 1.1E-03  
MA0079.2 (SP1) 1.1E-03 |

### B. HUVEC-specific Smad1/5 binding regions (n=2,935)

| motif | p-value |
|-------|---------|
| motif1 | M00340 (c-Ets-2) 3.0E-07  
M00339 (c-Ets-1) 3.5E-07  
M01207 (ETS2) 5.3E-07 |
| motif2 | M00804 (E2A) 1.2E-04  
M00929 (MyoD) 7.9E-04  
M00773 (MYB) 1.9E-03 |
| motif3 | M00701 (SMAD3) 2.5E-04  
M00974 (SMAD) 4.3E-04  
M00733 (SMAD4) 2.3E-03 |
| motif4 | MA0079.2 (SP1) 5.1E-05  
M00720 (CAC-binding) 5.5E-05 |
| motif5 | M00008 (Sp1) 7.3E-05  
M00808 (Pax) 2.8E-03  
M00238 (Barbie) 4.0E-03 |

### C. PASMC-specific Smad1/5 binding regions (n=1,908)

| motif | p-value |
|-------|---------|
| motif1 | M00706 (TFII-I) 4.1E-04  
M00658 (PU.1) 5.0E-04  
M00971 (Ets) 5.8E-04 |
| motif2 | M00629 (Eve) 5.2E-04  
M00704 (TEF-1) 8.8E-04 |
| motif3 | M00983 (MAF) 8.1E-05  
M00037 (NF-E2) 1.5E-04  
M00005 (AP-4) 2.1E-04 |
| motif4 | M01119 (KAISO) 2.0E-03  
M00238 (Barbie) 3.1E-03  
M01288 (NeuroD) 3.2E-03 |
| motif5 | M00733 (SMAD4) 1.2E-04  
M00687 (alpha-CP1) 2.1E-04  
M00701 (SMAD3) 4.4E-04 |
Supplementary Figure S3

A

ID1 (208937_s_at)

Signal intensity

0 2000 4000 6000 8000

0 8 16 24 (h)

-○- BMP-6 stimulation

-●- BMP-9 stimulation

ID2 (201565_s_at)

Signal intensity

0 1000 2000 3000 4000

0 8 16 24 (h)

ID3 (207826_s_at)

Signal intensity

0 1000 2000 3000 4000

0 8 16 24 (h)

B

Up-regulated

The number of the genes

400 200 0

Down-regulated

2 h 24 h

Genes expressed in HUVECs

n = 8,544

Up-regulated genes

| Number | Ratio |
|--------|-------|
| 2 h    | 108   | 1    |
| bound  | 70    | 0.64 |
| 24 h   | 318   | 1    |
| bound  | 116   | 0.36 |

Down-regulated genes

| Number | Ratio |
|--------|-------|
| 2 h    | 37    | 1    |
| bound  | 9     | 0.24 |
| 24 h   | 264   | 1    |
| bound  | 58    | 0.22 |

C

overlap

H3K4me1

overlap

H3K27ac

Smad1/5

Refseq (+)

Refseq (-)
Supplementary Figure S4

A

MEME1

MEME3

MEME4

MEME5

B

Random control regions (n=1,000)

The number of TFBS-positive sequences

motif折叠 enrichment

MEME1 1.8
MEME2 1.9
MEME3 1.6
MEME4 1.3
MEME5 1.6
Smad_1 1.8
Smad_2 1.4
Pou5f1 0.4
Sox2 0.5
TFAP4 2.0
GATA2 0.7

C

HUVECs BMP-6

394/880 (44.7%)

PASMCs BMP-4

1,167/2,745 (42.5%)

D

MEME2 positive ratio

n≥1

n≥2

n≥5

n≥10

Peak rank
Supplementary Figure S4 (Continued)

**E**

Data from all peaks were accumulated

- Motif nearest to a peak summit (1st motif)
- Within 100 bp from a peak summit

**HUVECs BMP-6**

- Counts vs Distance from peak summit (bp)
- Blue: Smad1/5 binding regions
- Green: 1st motif
- Red: Control

**PASMCs BMP-4**

- Counts vs Distance from peak summit (bp)
- Blue: Smad1/5 binding regions
- Green: 1st motif
- Red: Control

**F**

MEME1

- Counts vs Distance from peak summit (bp)

MEME3

- Counts vs Distance from peak summit (bp)

MEME4

- Counts vs Distance from peak summit (bp)

MEME5

- Counts vs Distance from peak summit (bp)
Supplementary Figure S5

A. **BMPR2 3rd intron**

- Human: cAGAC
- Mouse: cAGAC
- Mutants: cATCT

**JAG1 promoter**

- Human: CTGGCTCAAG
- Mouse: CTGACTCAAG
- Mutants: CTGGATCTAG

B. **JAG1 promoter**

| Relative luciferase activity (fold) |
|-----------------------------------|
| 0 1 2 3                         |

- GC: GGAGCC (WT)
- GGCGC
- GGTGCC
- GGGGC
- GC mut
- SBE mut
- 2nd intron

C. **Relative luciferase activity**

- Cell: HUVECs
- BMPR2
- JAG1
- BMPS-6: 50 ng/ml
- BMPS-6: 200 ng/ml
- BMPS-9: 1 ng/ml

D. **BMPR2 3rd intron**

- Cell: HMEC-1

E. **JAG1 promoter**

- Cell: HepG2

- Control
- BMP stimulation
