Supplemental Materials

Genome-wide Mapping of SMAD Target genes Reveals the Role of BMP Signaling in Embryonic Stem Cell Fate Determination

Teng Fei[1#], Kai Xia[2#], Zhongwei Li[1^], Bing Zhou[2^], Shanshan Zhu[2^], Hua Chen[1], Jianping Zhang[1], Zhang Chen[2], Huasheng Xiao[3], Jing-Dong J. Han[2]*, Ye-Guang Chen[1]*

Supplemental Methods

Cell culture and antibodies

Murine R1 ES cells were first plated on mitomycin-treated murine embryonic fibroblasts (MEFs) and grown under typical ES cell conditions (DMEM supplemented with 15% fetal calf serum, leukemia inhibitory factor (LIF), non-essential amino acids, GlutaMax-I (Invitrogen Gibco), beta-mercaptoethanol and penicillin/streptomycin. For ChIP assay, cells were cultured on gelatinized tissue culture plates for two passages to eliminate MEFs. KO-DMEM and Knockout serum replacement (Invitrogen Gibco) were used to substitute DMEM and fetal calf serum for feeder free culture. For ChIP experiments and immunoblotting, we used affinity-purified SMAD1/5 and SMAD4 antibodies that can recognize a specific band corresponding to endogenous SMAD proteins in R1 ES cells (Fig. S1A). Rabbit anti-SMAD1/5 antibody was generated with recombinant human SMAD1-linker (aa145-267)-His, and rabbit anti-SMAD4 antibody with recombinant human SMAD4-linker (aa144-316)-His. Anti-SMAD1/5 antibody can recognize both overexpressed SMAD1 and SMAD5 proteins but minimally SMAD8 (Fig. S1B). Phospho-SMAD1/5 antibody was purchased from Chemicon International (AB3848) and anti-nestin antibody from Chemicon International (MAB353). H3K27me3 antibody (ab6002) and histone H3 antibody (ab1791) were from Abcam. GFP (SC-8334), GST (SC-138), His (SC-803) antibody were from Santa Cruz Biotech.

Chromatin immunoprecipitation assay (ChIP)

R1 ES cells ($5x10^7$ – $1x10^8$ cells) were chemically cross-linked by addition of one tenth volume of freshly-prepared 11% formaldehyde solution to cell media for 15 min at room

temperature. Cells were rinsed twice with 1xPBS, harvested using a silicon scraper, frozen in liquid nitrogen and stored at –80°C prior to use. Cells were resuspended, lysed (1% SDS, 50 mM Tris pH 8.0, 5 mM EDTA, proteinase inhibitors), and sonicated to obtain DNA fragments of about 300-500bp length on average. Samples were then centrifuged at 14,000 rpm for 10 min. The supernatant was diluted (20 mM Tris pH 8.0, 2 mM EDTA, 1% Trion X-100, 150 mM NaCl, proteinase inhibitors) and pre-absorbed by 50 μl protein A beads (Zymed) and then incubated with 10 μg antibodies (anti-SMAD1/5 and anti-SMAD4) overnight at 4°C. The immunocomplex was collected with 100 μl protein A beads by 3 hour co-incubation and then washed sequentially with TSE I (0.1% SDS, 20 mM Tris pH 8.0, 2 mM EDTA, 1% Trion X-100, 150 mM NaCl, proteinase inhibitors), TSE II (0.1% SDS, 20 mM Tris pH 8.0, 2 mM EDTA, 1% Trion X-100, 500 mM NaCl, proteinase inhibitors), LiCl buffer (10 mM Tris pH 8.0, 1 mM EDTA, 0.25 mM LiCl, 0.1% NP-40, 1% deoxycholate sodium) and TE (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0). The bound immunocomplex was eluted with 400 μl fresh elution buffer (25 mM Tris pH 8.0, 10 mM EDTA, 0.5% SDS) by heating at 65°C with occasional vortex for 15 min and crosslinking was reversed by overnight incubation at 65°C. Whole cell extract (WCE) DNA (Input fraction reserved from the sonication step) was also treated for crosslinking reversal. Immunoprecipitated DNA and WCE DNA were then purified by treatment with RNaseA, proteinase K and multiple phenol:chloroform:isoamyl alcohol extraction.

RNA isolation, quantitative real-time PCR, and analysis of transcript levels

To determine transcript levels by RT-PCR, total RNA was isolated from ES cells, EB or neural precursor cells using Trizol reagent (Invitrogen). For transient BMP stimulation, cells were

treated by 20ng/ml BMP4 (R&D, 314-BP). In neural precursor differentiation experiments, BMP4 was used at a concentration of 10ng/ml. Reverse transcriptase (Toyobo) was employed for oligo dT primed first strand cDNA synthesis. qRT-PCR was carried out on Mx3000P detection system (Stratagene) using EvaGreen dye (Biotium). Primers are designed with assistance of Primer bank (Wang and Seed 2003) and listed in Supplemental Table S9. To comparatively quantify the amount of mRNA level, ΔΔCt method was used. The significance of expression was analyzed by Student *t* test.

Derivation of neural precursor cells from mouse ES cells

Serum-free monolayer neural differentiation was modified from a protocol (Ying et al. 2003). Briefly, undifferentiated R1 ES cells cultured in feeder-free conditions were dissociated and plated onto 0.1% gelatin-coated six-well tissue culture plastic at a density of $0.5–1.5\times10^4/cm^2$ in KO-SR ES medium without LIF. About 24 hours later, the medium was changed into N2B27 medium (1:1 mixture of DMEM/F12 supplemented with N2 and NeuralBasal medium supplemented with B27 and with beta-mercaptoethanol and Glutamax-I) and this time was recorded as day 0 of differentiation in N2B27 medium. Medium was changed every other day. At day 5, cells were collected for RNA isolation and immunofluoresence analysis.

Knockdown of *Smad1, Smad4, Dpysl2* and *Kdm6b* in R1 ES cells

The short hairpin nucleotides corresponding to the genes under investigation were cloned into pll3.7 RNAi plasmid by *Hpa* I and *Xho* I sites. Nucleotides targeting the firefly luciferase gene which had no homology with mouse transcriptome was served as control shRNA. The RNAi oligonucleotides used are as follows: Control RNAi: 5’-GCGACCAACGCCTTGATTG-3’; *Smad1* RNAi: 5’-GCCTCTGGAATGCTGTGAG-3’;

Smad4 RNAi: 5'GGAGTGCAGTTGGAATGTA-3'; *Dpysl2* RNAi 1#: 5'-GATGGGTTGATCAAGCAAA-3' (Vincent et al. 2005); *Dpysl2* RNAi 2#: 5'-ACTCCTTCCTCGTGTACAT-3' (Mimura et al. 2006); *Kdm6b* RNAi 1#: 5'-GCCCAGTCTGTGAAACCCAAG-3'(De Santa et al. 2007), *Kdm6b* RNAi 2#: 5'-GCACTCGATGCCTCATTCATA-3' (De Santa et al. 2007). To generate knockdown cells, R1 cells were transfected separately with the shRNA constructs by Lipofectamine 2000 and 1 μg/ml puromycin was added 48 hours after transfection. About 10 days later, stably transfected cells were expanded and then stored for further analysis. To determine knockdown effectiveness, RNA was isolated and qRT-PCR was employed to determine the mRNA expression of *Smad1*, *Dpysl2* and *Kdm6b*. SMAD4 expression was examined by immunoblotting. SMAD1/5 antibody recognizes both SMAD1 and SMAD5 and was not used for examination of *Smad1* knockdown effectiveness.

Construction of *Dpysl2* reporter plasmids

To construct *Dpysl2* promoter luciferase reporters, the *Dpysl2* (NC_000080) promoter covering -968 bp to +619 bp or -968 bp to +212 bp was amplified by PCR with the primers containing *Xho* I and *Hind* III sites and subcloned into pGL3-Basic vector (Promega, Madison, WI). The genomic location number was coordinated by transcriptional start site. The upstream primer: 5'-GCGCTCGAGCTTAGTGACTAGAGGTTCTGTTTC-3'; and downstream primers: 5'-GCCAAGCTTCGGAGAACAATATCAGAGCGTCGA-3' for 619-Luc and 5'- GGCAAGCTTAGTAATCTCAAGAGGACAGCTTTA-3' for 212-Luc.

Ectopic expression of *Dpysl2* during early neural differentiation

Mouse *Dpysl2* cDNA (NM_009955) was cloned by PCR from mouse ES cells cDNA library

and inserted into pEGFP-C3 plasmid using *Hind* III and *Pst* I sites. The upstream primer is 5'-CGGAAGCTTATGTCTTATCAGGGGAAGAAAAATATTCC-3' and downstream primer 5'-ATACTGCAGTTAGCCCAGGCTGGTGATGTTGGCA-3'. For neural differentiation experiment, pEGFP-Dpysl2 or pEGFP-C3 empty vector were first transfected into R1 ES cells using Lipofectamine 2000. At 24 hour post-transfection, the cells were replated for neural differentiation. BMP4 (10ng/ml) were added at neural differentiation day 1. Cells were harvested for further analysis at day 3 when neural precursor markers emerged in normal differentiation process.

Luciferase assay

HEK293T and R1 ES cells seeded in 24-well plate were transfected by calcium phosphate and Lipofectamine 2000, respectively, with the indicated plasmids. The internal control pRenilla-TK vector (10ng) was used to normalize luciferase activity. Each experiment was performed in triplicate and the data represent the mean ± SEM of three independent experiments after normalized to Renilla activity.

DNA oligonucleotide pull-down assay

5'-Biotinylated DNA oligonucleotides were synthesized by Sangon (Shanghai, China). The oligos sequences are as follows: Dpy WT, 5'-GGTTGAAGATGCAATAACAGTTATAAGGCACGCAATTTGCAAGGCATTGAAATAGGTGC-3'; Dpy Mut, 5'-GGTTGAAGATGCAATAACAGTTATAAGGTACATAATTTGTAAGATATTGAAATAGGTGC-3'; ARE, 5'-TATCTGCTGCCCTAAAATGTGTATTCCATGGAAATGTCTGCCCTTCTCTCC-3';

SMAD1-M1, 5'- GAGTGTGGAAGGGTGTGGTTGAGTGTGGAAGGGTGTGG-3';

SMAD1-M2, 5'- GAGGCGGGAAGAGGTGGGTTGAGGCGGGAAGAGGTGGG-3';

SMAD1-M3, 5'-GGCGGGGCAAGGCGGGGTTTGGCGGGGCAAGGCGGGGT-3';

SMAD1-M4, 5'- CCGCCCCAAACCGCCCCTTTCCGCCCCAAACCGCCCCT-3';

SMAD4-M1, 5'- TGGGAAGAAATGGGCAGATTTGGGGAGAAATGGGAAGA-3';

SMAD4-M2, 5'- CTTCACTGAACTTCCCTGTTCTTCTCTGAACTTCACTG-3';

SMAD4-M3, 5'- TCTGCACTAATCTGCCCTTTTCTGCTCTAATCTGCACT-3';

SMAD4-M4, 5'- CCTTCCCAAACCTTCTCATTCCTTCCCAAACCTTCTCA-3'.

Biotinylated oligos were first annealed with non-biotinylated complementary strand and then incubated together with *Escherichia coli* purified SMAD proteins in 500 μl HKMG buffer (10 mM Hepes, pH 7.9, 100 mM KCl, 5 mM $MgCl_2$, 10% glycerol, 1 mM dithiothreitol and 1.5% Nonidet P40) plus 10 μg salmon sperm DNA to reduce background. After 2 hours binding reaction, streptavidin beads (Pierce) were added in for 1 hour and then washed for at least three times using HKMG buffer with increasing KCl concentration up to 400 mM for the most stringent wash condition. The precipitated protein were separated by boiling the beads in 2XSDS loading buffer and then analyzed by immunoblotting.

Promoter array design and data normalizations

For chip analysis, purified DNA was blunted, ligated to a universal linker and amplified by PCR. DNA was labeled with CGH labeling kits (Invitrogen) (immuno-enriched DNA was labeled with Cy5 fluorophore, WCE DNA was labeled with Cy3 fluorophore). Labeled and purified DNA was combined and hybridized to DNA chips (Agilent Mouse Promoter

two-color Array, NCBI 35, UCSC mm7) according to manufacturer's instruction. The raw data were processed by the R package limma (Smyth and Speed 2003). Briefly, the signals were corrected for background with the 'minimum' method (Ritchie et al. 2007). Then Cy5 and Cy3 signal values were separately subjected to inter-array median normalization. For each slide, the median signal over the common probes was computed separately for each channel. For each channel, the average of this median is computed over the two separate slides. Finally, a signal was normalized by dividing it by the ratio of the median of its replicate's common probe over the average computed above. Finally the log-ratio of IP over WCE was normalized within arrays using 'loess' method (Yang et al. 2001) and the normalized log-ratios were then used for SMAD target selection.

Identification of SMAD bound regions

For each sample, we calculated the standard deviation (SD) and average of the log-ratio (avgLogRatio) of all the probes. The ratio cutoff was defined as m times of SD larger than the avgLogRatio (ratio cutoff = avgLogRatio +m*SD). A probe is defined as the significant probe if the ratio of the probe is larger than ratio cutoff. We then defined the binding site as the region covered by at least n continuous significant probes, where each pair of neighboring probes should be apart from each other by at most 500 bp. We run this two times under different cutoffs (run 1, m=2, n=2, run 2, m=1.6, n=3) and took the union of the two sets as SMAD target genes.

Confirmation of ChIP-chip targets by ChIP-seq (Illumina deep sequencing)

ChIP-seq was carried out according to Illumina standard protocols. Using the SOAP program (Li et al. 2008), 16.3, 11.4 and 11.1 million unique reads were mapped to the mouse genome

for SMAD1/5 ChIP-seq, SMAD4 ChIP-seq and ChIP input DNA sequencing, respectively. Binding peaks were identified using the SICER program (Zang et al. 2009) by comparing unique reads from a ChIP experiment to those from ChIP input DNA ($P < 10^{-3}$). The peaks were then mapped to genes if they appear at the gene promoter regions (-8 kb to +3 kb of TSS).

Site-specific PCR for binding validation

To validate binding events identified by ChIP-chip, site-specific PCR analysis was performed. A subset of primers corresponding to the predicted binding sites were designed to amplify a 200~300bp region around the peak probes. PCR was carried out on ligation-mediated PCR amplified samples. 10 ng of immunoprecipitated (IP) DNA and a range amount (10 ng, 20 ng, 50 ng) of whole cell extract (Input) DNA were used in PCR reactions. PCR products were visualized on agarose gel stained with ethidium bromide. "Unbound" region was defined as the PCR product yield of 10ng IP DNA less than that of 10ng Input DNA; "Bound" defined as the PCR product yield of 10ng IP DNA more than that of 20ng Input DNA; and "Marginally bound" as the PCR product yield of 10ng IP DNA fallen between the ones of 10ng and 20ng Input DNA.

Monte Carlo simulation for calculating empirical *P*-value of the overlap counts between two groups of genes

In order to evaluate if the target genes of SMAD1/5 and SMAD4 have significant overlap, we did 100 times of Monte Carlo simulation to calculate the empirical *P*-value. For each simulation, we selected two random groups of genes of same number as the real gene sets, and counted the number of overlapping genes between the two random groups. We run such

simulations 100 times and defined *a* as the number of times with random overlap counts equal or larger than the real number of overlapping genes between two gene sets. The empirical *P*-value is defined as $P = a/100$. If a is 0 after 100 times of simulations, then $P < 0.01$. Fisher's exact test *P*-values were also calculated.

Histone modification enrichment test

The histone modification data were derived from published profiles of H3K4me3 and H3K27me3 in mES cells (Mikkelsen et al. 2007). The percentage of modified genes in the whole-genome profile of H3K4me3 or H3K27me3 was defined as control percentage (p0) and the percentage of modified genes in SMAD1/5 and SMAD4 targets as p1. Then a ratio of p1/p0 >1 means more enriched than the whole genome average whereas <1 means less enriched than the whole genome average. Fisher's exact test *P*-values of the enrichment were also calculated.

Gene Ontology analysis

The gene functional enrichment analysis for the SMAD1/5 and SMAD4 target genes was carried out with Cytoscape v2.3.2 plug-in BiNGO 2.0 (Maere et al. 2005).

Noggin and BMP4 cDNA microarray data analysis

The raw data in CEL files were processed using the Affy package in the BioConductor library (http://www.bioconductor.org/) with the "invariantset" normalization method and the "liwong" summarization method (Li and Wong 2001a; Li and Wong 2001b). Differentially expressed genes were determined by the RankProd package (Hong et al. 2006) of BioConductor using "pval" method with 100 permutations, treating the three samples as of the same origin. Noggin or BMP4 treated R1 ES cells were compared individually to the control untreated cells.

SMAD targets and H3K4me3/H3K27me3 enrichment (Mikkelsen et al. 2007) *P* values were determined by Fisher's exact test.

Gene expression profile analysis of SMAD targets during ES cell to EB differentiation

The ES cell to EB differentiation microarray data of R1 cells (Hailesellasse Sene et al. 2007) was used in the expression analysis for the SMAD1/5 and SMAD4 target genes. The dataset contains the time of 0 hours, 6 hours, 12 hours, 18 hours, 24 hours, 36 hours, 48 hours, 4 days, 7 days, 9 days and 14 days of ES cell differentiation to EB. The probe signals were processed as described (Boyer et al. 2006). The signals of the arrays were scaled to a mean value of 200 over all the probes on the chip. We excluded the probes which did not show a max/min value ratio of ≥2 or a max-min value difference of ≥100. The remaining signal values of the probe sets were normalized to each probe set average expression level among all the samples and log2 transformed. Log signal values were then row normalized to a mean of 0 and a variance of 1 among all the samples. Probe sets were ordered by the difference between the average signal value of the differentiated ES cells and the average signal value of undifferentiated ES cells, which were visualized using Java TreeView (Saldanha 2004).

Motif Analysis

The *de novo* motif discovery was performed using DME (Smith et al. 2005). The sequences described below were extracted from UCSC Genome database (http://genome.ucsc.edu/). All the SMAD-binding sites and their flanking sequences (starting from right outside of each binding site to 796 bp away from the site) were used as positive (real) and negative (background) sequence input for DME, respectively. The appearance frequencies of the top 10 motifs in experimentally identified SMAD1/5 binding sites ranged from 17% to 37%,

which were 1.37 to 2.14 times those in the flanking background sequences of the same length as the binding site (Table S5). Frequencies of the top 10 motifs for SMAD4 ranged from 17% to 27%, and are 1.22 to 1.4 times of the background. Before running DME, simple repeats (single repeats >5 bases such as CCCCCC, and double repeats >9 bases such as ACACACACAC) were removed from the ChIP-enriched sequences (Ben-Yehuda et al. 2005). Then DME was run with default parameters to find the top ranked motifs with width of 8 nucleotides. The final motif logos were created by WebLogo (Crooks et al. 2004).

Supplemental References

Ben-Yehuda, S., M. Fujita, X.S. Liu, B. Gorbatyuk, D. Skoko, J. Yan, J.F. Marko, J.S. Liu, P. Eichenberger, D.Z. Rudner, and R. Losick. 2005. Defining a centromere-like element in Bacillus subtilis by Identifying the binding sites for the chromosome-anchoring protein RacA. *Mol Cell* **17:** 773-782.

Boyer, L.A., K. Plath, J. Zeitlinger, T. Brambrink, L.A. Medeiros, T.I. Lee, S.S. Levine, M. Wernig, A. Tajonar, M.K. Ray, G.W. Bell, A.P. Otte, M. Vidal, D.K. Gifford, R.A. Young, and R. Jaenisch. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441:** 349-353.

Chen, X., E. Weisberg, V. Fridmacher, M. Watanabe, G. Naco, and M. Whitman. 1997. Smad4 and FAST-1 in the assembly of activin-responsive factor. *Nature* **389:** 85-89.

Crooks, G.E., G. Hon, J.M. Chandonia, and S.E. Brenner. 2004. WebLogo: a sequence logo generator. *Genome Res* **14:** 1188-1190.

De Santa, F., M.G. Totaro, E. Prosperini, S. Notarbartolo, G. Testa, and G. Natoli. 2007. The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. *Cell* **130:** 1083-1094.

Hailesellasse Sene, K., C.J. Porter, G. Palidwor, C. Perez-Iratxeta, E.M. Muro, P.A. Campbell, M.A. Rudnicki, and M.A. Andrade-Navarro. 2007. Gene function in early mouse embryonic stem cell differentiation. *BMC Genomics* **8:** 85.

Hong, F., R. Breitling, C.W. McEntee, B.S. Wittner, J.L. Nemhauser, and J. Chory. 2006. RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* **22:** 2825-2827.

Li, C. and W.H. Wong. 2001a. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* **98:** 31-36.

Li, C. and W.H. Wong. 2001b. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol* **2:** RESEARCH0032.

Li, R., Y. Li, K. Kristiansen, and J. Wang. 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24:** 713-714.

Maere, S., K. Heymans, and M. Kuiper. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21:** 3448-3449.

Mikkelsen, T.S., M. Ku, D.B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T.K. Kim, R.P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E.S. Lander, and B.E. Bernstein. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448:** 553-560.

Mimura, F., S. Yamagishi, N. Arimura, M. Fujitani, T. Kubo, K. Kaibuchi, and T. Yamashita. 2006. Myelin-associated glycoprotein inhibits microtubule assembly by a Rho-kinase-dependent mechanism. *J Biol Chem* **281:** 15970-15979.

Ritchie, M.E., J. Silver, A. Oshlack, M. Holmes, D. Diyagama, A. Holloway, and G.K. Smyth. 2007. A comparison of background correction methods for two-colour microarrays. *Bioinformatics* **23:** 2700-2707.

Saldanha, A.J. 2004. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20:** 3246-3248.

Smith, A.D., P. Sumazin, D. Das, and M.Q. Zhang. 2005. Mining ChIP-chip data for transcription factor and cofactor binding sites. *Bioinformatics* **21 Suppl 1:** i403-412.

Smyth, G.K. and T. Speed. 2003. Normalization of cDNA microarray data. *Methods* **31:** 265-273.

Vincent, P., Y. Collette, R. Marignier, C. Vuaillat, V. Rogemond, N. Davoust, C. Malcus, S. Cavagna, A. Gessain, I. Machuca-Gayet, M.F. Belin, T. Quach, and P. Giraudon. 2005. A role for the neuronal protein collapsin response mediator protein 2 in T lymphocyte polarization and migration. *J Immunol* **175:** 7650-7660.

Wang, X. and B. Seed. 2003. A PCR primer bank for quantitative gene expression analysis. *Nucleic Acids Res* **31:** e154.

Yang, Y.H., S. Dudoit, P. Luu, and T.P. Speed. 2001. Normalization for cDNA microarray data. In *Microarrays: Optical Technologies and Informatics, Vol. 4266 of Proceedings of SPIE* eds M.L. Bittner Y. Chen A.N. Dorsel, and E.R. Dougherty), pp. pp.141-152. Spie Press.

Ying, Q.L., M. Stavridis, D. Griffiths, M. Li, and A. Smith. 2003. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol* **21:** 183-186.

Zang, C., D.E. Schones, C. Zeng, K. Cui, K. Zhao, and W. Peng. 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25:** 1952-1958.

Figure S1

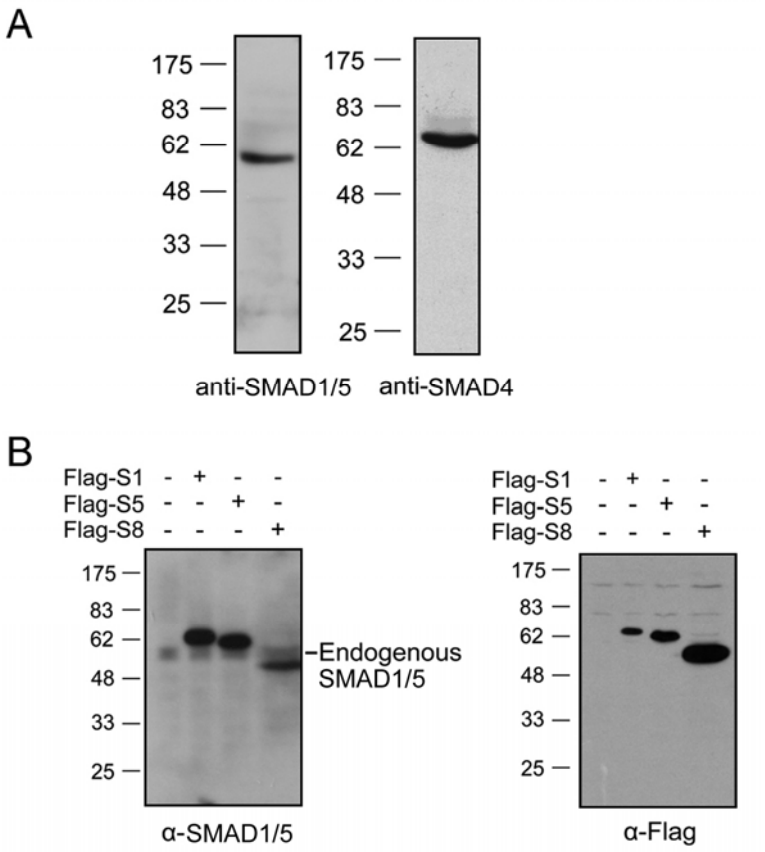
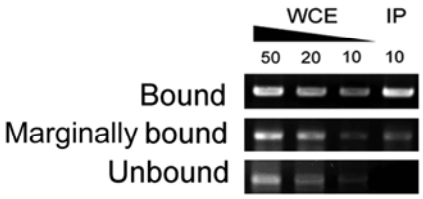


Figure S1. Antibody characterization of SMAD1/5 and SMAD4 for ChIP experiment. (*A*) The specificity of the two affinity-purified antibodies was examined by immunoblotting of R1 cell lysates. (*B*) Anti-SMAD1/5 antibody mainly recognize both SMAD1 and SMAD5. Flag-SMAD1，Flag-SMAD5 and Flag-SMAD8 were overexpressed in HEK293T cells and cell lysates were immunoblotted by anti-SMAD1/5 antibody and anti-Flag antibody. The same amount of samples was loaded for immunoblotting with different antibodies.

Figure S2



Bound rate (%)	SMAD1/5	SMAD4	Total
Bound	33/41 (80.5%)	39/50 (78.0%)	72/91 (79.1%)
Marginally bound	4/41 (9.8%)	8/50 (16.0%)	12/91 (13.2%)
Unbound	4/41 (9.8%)	3/50 (6.0%)	7/91 (7.7%)

Figure S2. Summary of binding validation by ChIP-site specific PCR. The upper panel shows a representative ChIP-site specific PCR result. Immunoenriched (anti-SMAD IP) DNA (10ng) from R1 cells and a range of unenriched whole cell extract (WCE) DNA amounts (10, 20, 50ng of DNA) are used for PCR with primer pairs which were designed according to the predicted bound regions. We grouped the binding state into three categories: Bound (PCR product yield of 10ng IP more than that of 20ng input DNA), Marginally bound (PCR product yield of 10ng IP fallen between those of 10ng and 20ng input DNA) and Unbound (PCR product yield of 10ng IP less than that of 10ng input DNA). The table is a summary of different binding states of all randomly tested putative SMAD binding events.

Figure S3

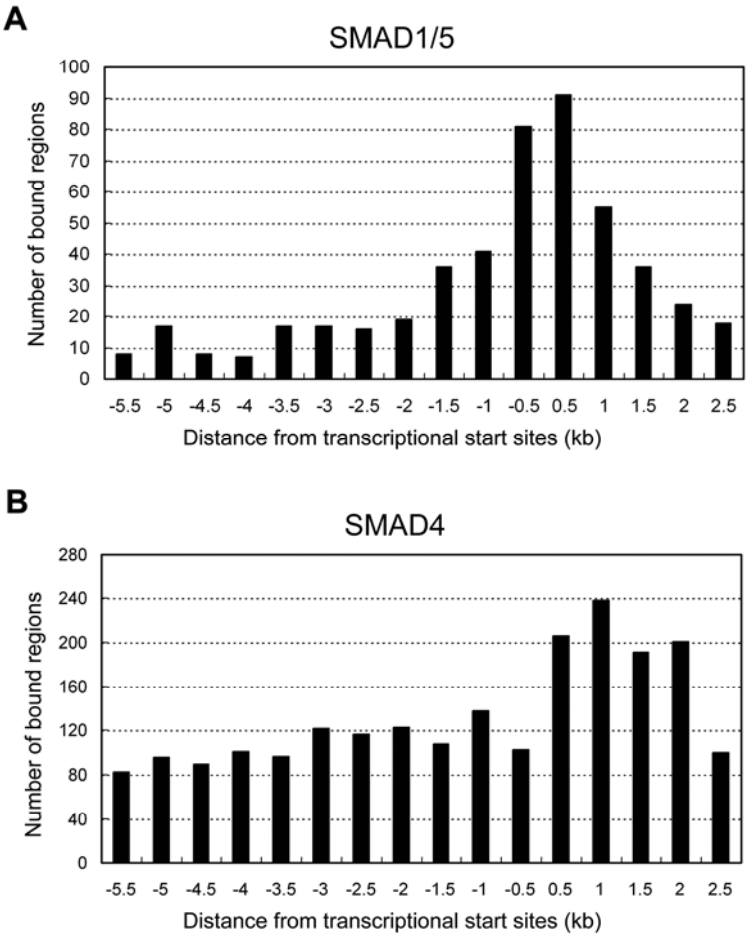


Figure S3. Distribution of the SMAD-bound regions along promoter regions. (*A*) Histogram of SMAD1/5-bound regions' distance to TSS along the interrogated promoter region. (*B*) Histogram of SMAD4-bound regions' distance to TSS along the interrogated promoter region.

Figure S4

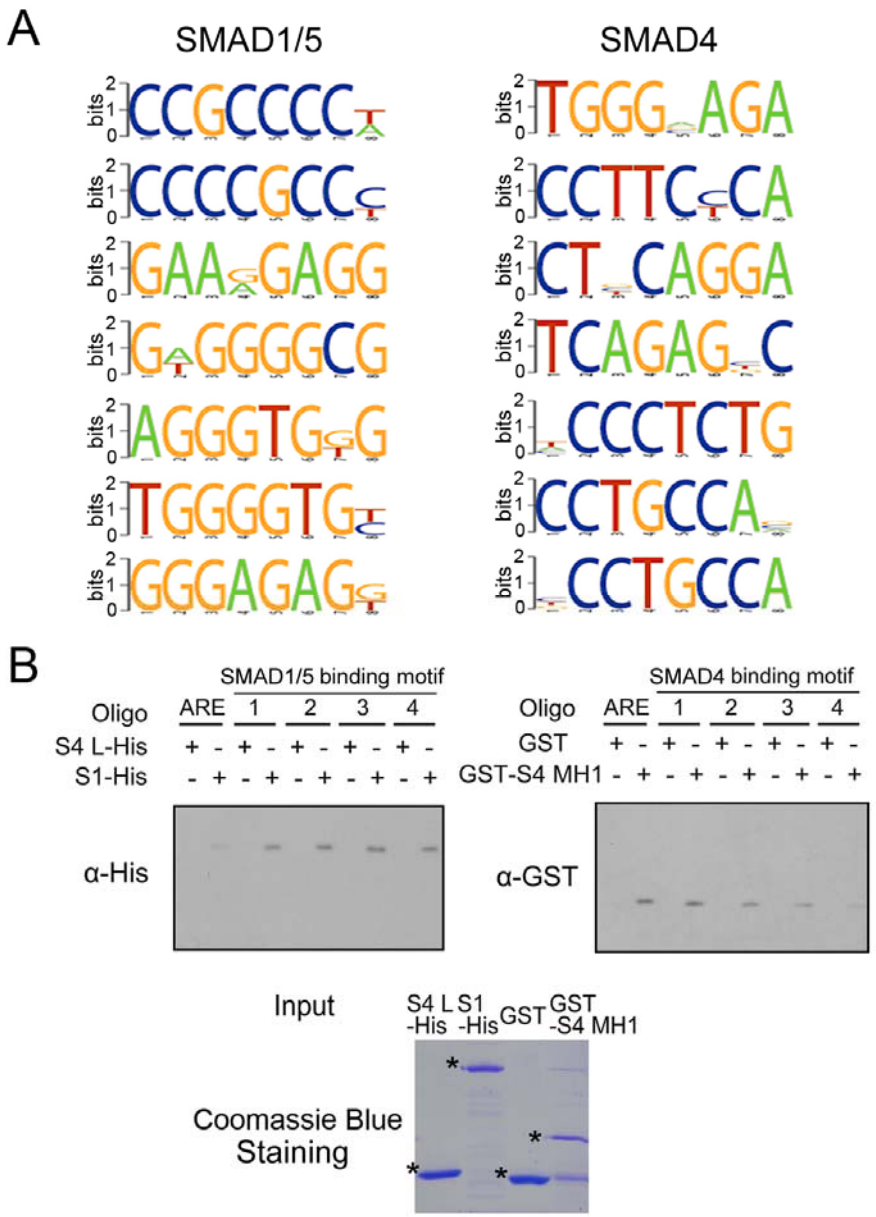


Figure S4. *De novo* prediction and validation of SMAD binding motifs. (*A*) Seven of the top ten scored DNA binding motifs of SMAD1/5 and SMAD4 identified by DME. (*B*) Validation of the top four motifs for SMAD1/5 and SMAD4 respectively in Table S5. For each 8-mer motif, four tandem repeat oligonucleotides were synthesized and labeled with biotin. Biotinylated oligonucleotides were incubated with *Escherichia coli* purified proteins

before streptavidin bead precipitation. The pull-down proteins were examined by immunoblotting. The Coomassie blue staining indicates protein input. S4 L-His and GST serve as negative controls. S1-His: His-tagged full-length SMAD1; S4 L-His: His-tagged SMAD4 linker region (aa144-294); GST-S4 MH1: GST-tagged SMAD4 MH1 domain (aa1-144). ARE (activin response element) derived from the *Xenopus Mix.2* promoter (Chen et al. 1997) served as a positive control for SMAD4 binding. Asterisk (*) indicates the specific band.

Figure S5

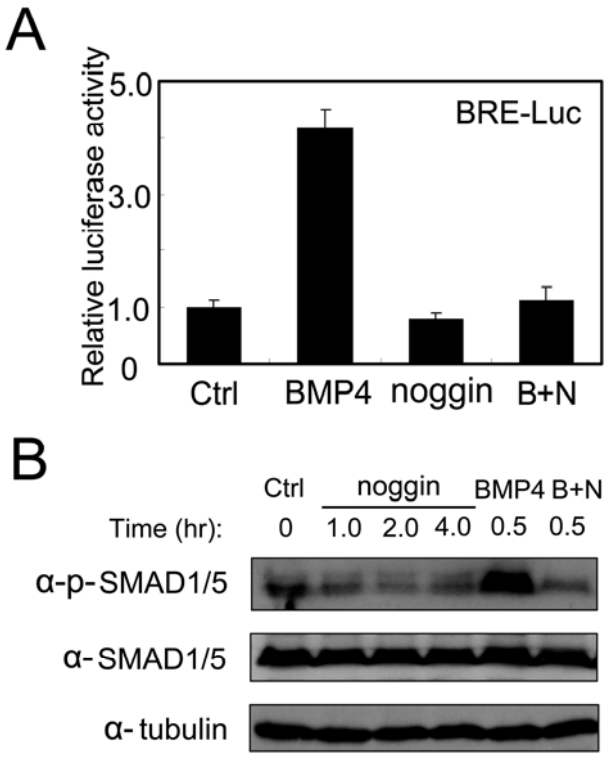


Figure S5. Noggin inhibits BMP4 signaling in R1 ES cells. (A) R1 cells were transfected with the BRE-luciferase reporter (0.1μg) and Renilla (10ng) and treated with 20ng/ml BMP4 and/or 100ng/ml noggin at 24 hours post-transfection. After 12 hours, cells were then harvested and luciferase activity was measured. (B) Cell lysates from R1 cells treated with 20ng/ml BMP4 and 100ng/ml noggin for indicated time were immunoblotted for anti-phospho-SMAD1/5, anti-SMAD1/5 antibodies. Tubulin served as a loading control. B: BMP4; N: noggin.

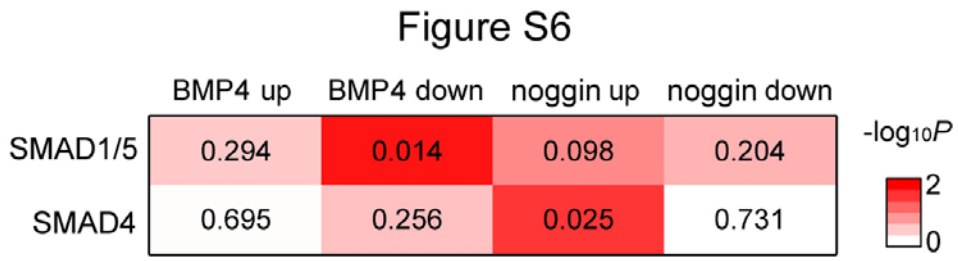


Figure S6. The significance of overlaps between SMAD bound targets with BMP4 and noggin responsive genes. The significance of *P* values are indicated inside each cell of the intersecting groups of genes and each cell is color coded according to the color legend.

Figure S7

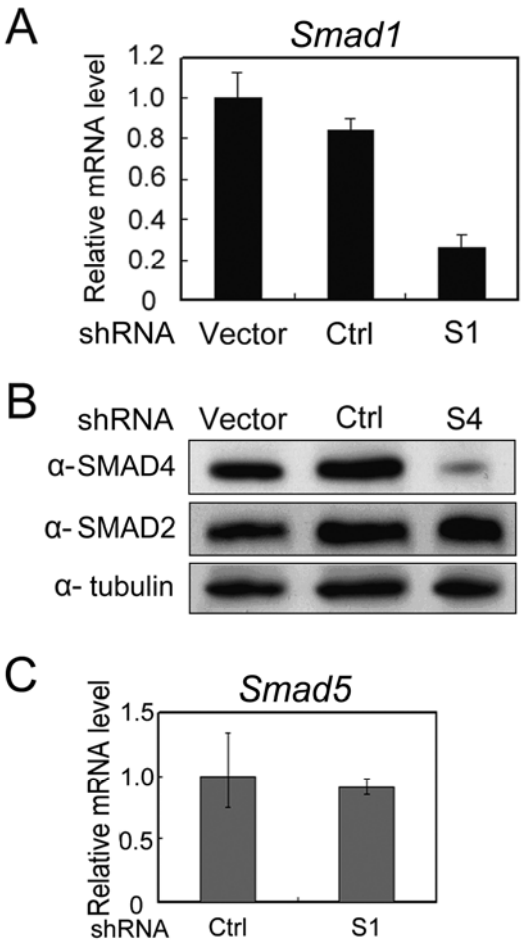


Figure S7. Knockdown of SMAD expression in mouse ES cells. (*A*) *Smad1* expression was determined by qRT-PCR in R1 cells stably expressing vector, control shRNA or *Smad1* shRNA. S1: *Smad1*. (*B*) Cell lysates from R1 cells stably expressing indicated shRNA plasmid were immunoblotted with anti-SMAD4, anti-SMAD2 antibodies. Tubulin served as a loading control. S4: SMAD4. (*C*) *Smad1* shRNA does not influence *Smad5* expression. The *Smad5* mRNA level was determined by qRT-PCR in R1 cells stably expressing control shRNA or *Smad1* (S1) shRNA.

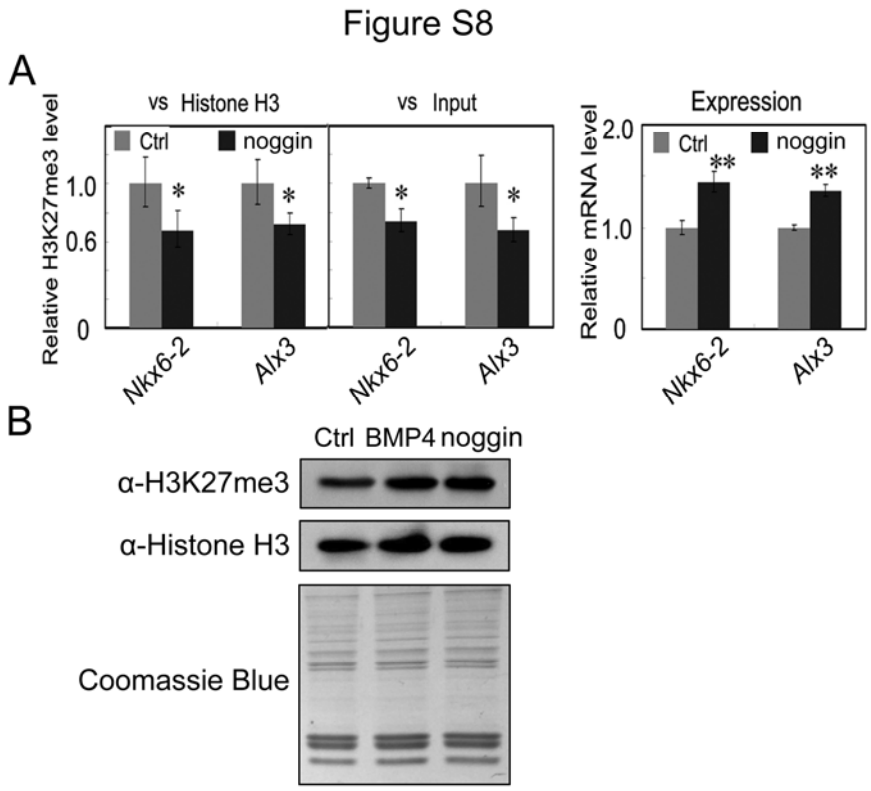


Figure S8. Correlation of H3K27me3 modification with noggin-activated expression of the *Nkx6-2* and *Alx3* genes. (*A*) The H3K27me3 levels in the *Nkx6-2* and *Alx3* promoters were analyzed and their corresponding mRNA levels were examined upon 100 ng/ml noggin treatment for 4 hours. Relative H3K27me3 level was determined by ChIP assay and qPCR and normalized by histone H3 enrichment level or input signal intensity. The mRNA levels were determined by qRT-PCR. The significance was analyzed by Student *t* test, and the data are presented as mean ± SEM (n=3, ** $P<0.01$, * $P<0.05$). (*B*) Total histone H3K27me3 levels did not change in BMP4- or noggin-treated R1 ES cells. Cells were treated with 20 ng/ml BMP4 or 100 ng/ml noggin for 4 hours before histone extraction. H3K27me3 and total histone H3 antibodies were used for immunoblotting. Coomassie blue staining of samples served as a loading control.

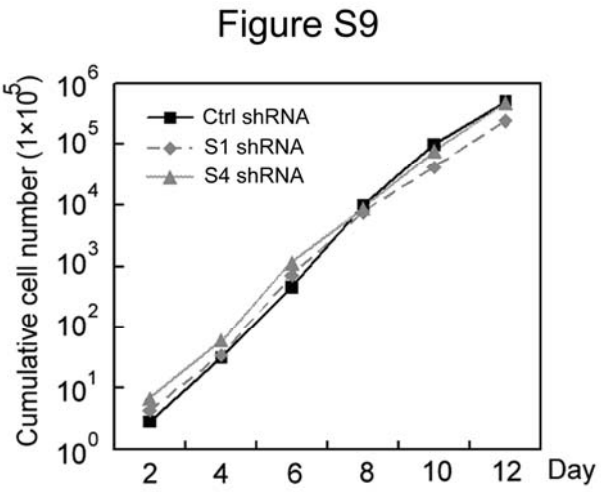


Figure S9. Knockdown of *Smad1* and *Smad4* does not affect mouse ES cell proliferation. Cells were plated in feeder-free KO-SR ES cell culture medium and every two days the numbers of cells were counted. Experiments were done in duplicates with similar results and a representative one is shown here. S1: *Smad1*; S4: *Smad4*.

Figure S10

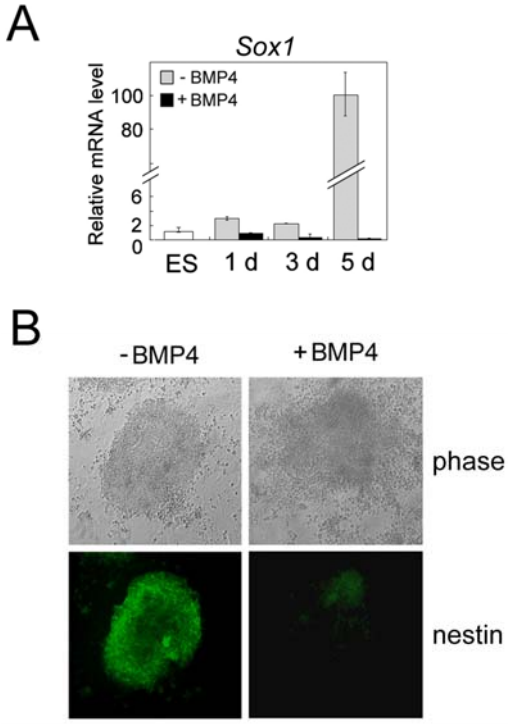


Figure S10. BMP4 inhibits early neural differentiation. (*A*) BMP4 potently inhibits the expression of the neural precursor marker *Sox1* during neural differentiation. R1 cells were cultured in neural differentiation medium in the presence or absence of 10 ng/ml BMP4 for 5 days. Samples were collected at day 1, 3, 5 of neural differentiation and qRT-PCR was performed to detect *Sox1* expression. (*B*) BMP4 represses nestin expression. R1 cells were cultured in neural differentiation medium in the presence or absence of 10 ng/ml BMP4 for 5 days. Nestin expression of was determined by immunofluorescence.

Figure S11

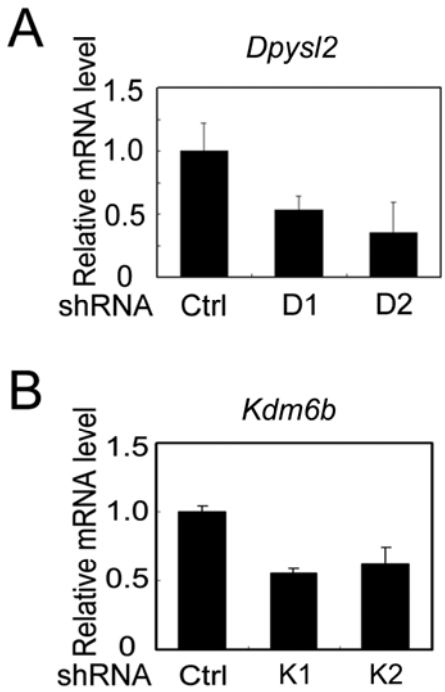


Figure S11. Knockdown of *Dpysl2* and *Kdm6b* in mouse ES cells. (*A*) *Dpysl2* expression was determined by qRT-PCR in R1 cells stably expressing control shRNA or two *Dpysl2* shRNA constructs (D1 and D2). (*B*) *Kdm6b* expression was determined by qRT-PCR in R1 cells stably expressing control shRNA or two *Kdm6b* shRNA constructs (K1 and K2).

Figure S12

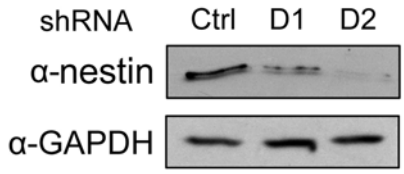


Figure S12. *Dpysl2* knockdown reduces nestin expression. Nestin expression was determined by immunoblotting in *Dpysl2* shRNA D1- or D2-expressing cells at day 5 of neural differentiation. GAPDH served as a loading control.

Figure S13

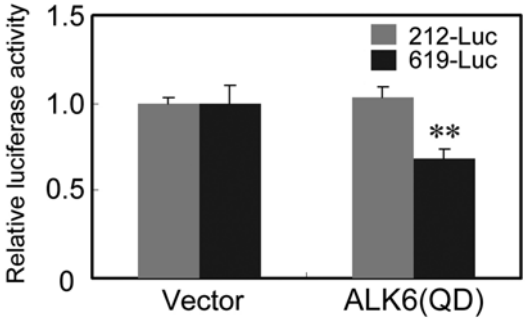


Figure S13. Constitutively active BMP receptor ALK6 (QD) represses the *Dpysl2* promoter activity. Reporter plasmid (0.1μg), Renilla plasmid (10ng), empty vector or ALK6(QD) plasmid (0.2μg) were cotransfected into HEK293T cells. At 36 hour post-transfection, cells were harvested for luciferase assay. The experiment was performed in triplicate, and the data are presented as mean ± SEM of three independent experiments after normalization to Renilla activity. (Student *t* test, ** $P<0.01$)

Figure S14

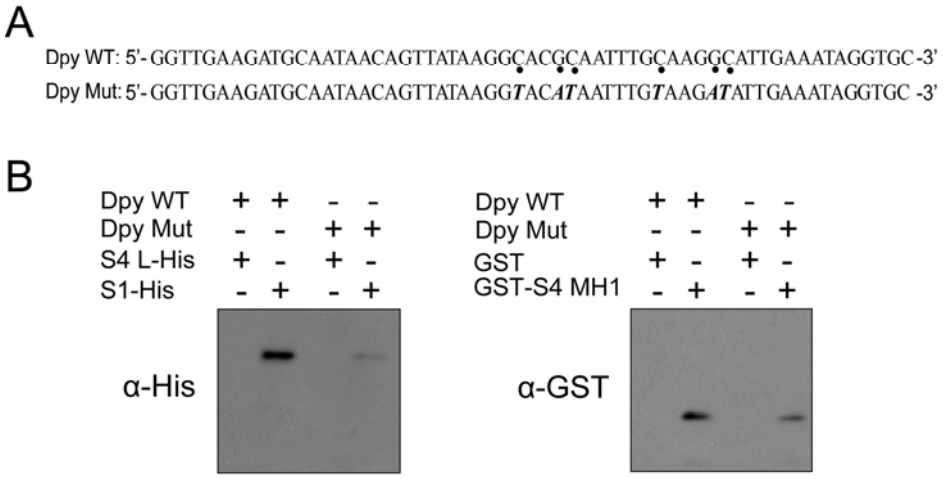


Figure S14. SMAD1 and SMAD4 can directly bind to the *Dpysl2* promoter. (*A*) The nucleotide sequence containing the predicted SMAD binding sites of the *Dpysl2* promoter. Dpy WT: wild-type sequence; Dpy Mut: mutant sequence (mutation sites are indicated by black dots). (*B*) Direct association of SMAD1 and SMAD4 with the *Dpysl2* promoter determined by oligonucleotide pull-down assay. S4 L-His and GST served as negative control protein. The protein input is the same as Fig. S4B. S4 L-His: His-tagged SMAD4 linker region; S1-His: His-tagged full-length SMAD1; GST-SMAD4 MH1: GST-tagged SMAD4 MH1 domain.

Figure S15

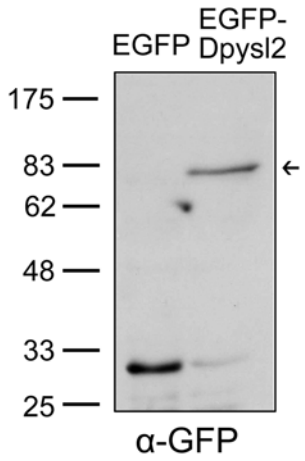


Figure S15. Overexpression of DPYSL2 in R1 ES cells. Cells were transiently transfected with pEGFP-Dpysl2 or pEGFP-C3 empty vector respectively. At 48 hours post-transfection, cell lysates were immunoblotted with anti-GFP antibody. The arrow indicates the band of EGFP-Dpysl2 fusion protein.

Figure S16

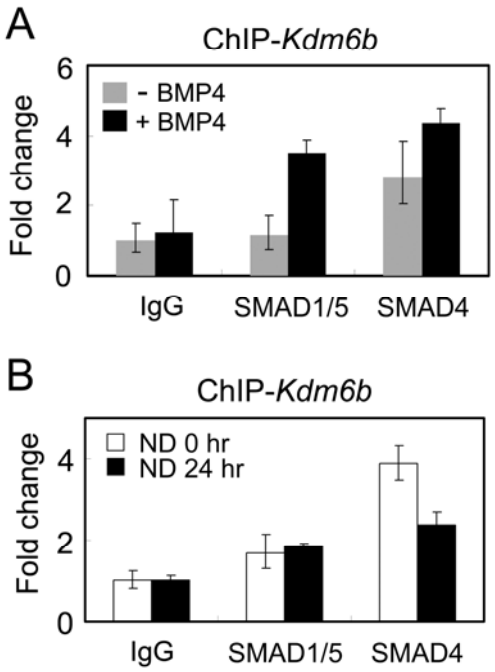


Figure S16. SMAD binding to the *Kdm6b* promoter. (*A*) R1 cells were harvested for ChIP after 12 hour treatment in the presence (+) or absence (-) of 10ng/ml BMP4 in neural differentiation medium. Control (IgG) and SMAD antibody were used to pull down predicted the *Kdm6b* promoter segment, and relative enrichment fold was determined by qPCR and normalized to input signal. (*B*) R1 cells were harvested for ChIP at 0 hour and 24 hour time points during neural differentiation. SMAD1/5 and SMAD4 binding to the *Kdm6b* promoter before and after differentiation was compared. ND: neural differentiation.

Figure S17

	NANOG (1233)	SOX2 (786)	DAX1/NR0B1 (1692)	NAC1/NACC1 (769)	OCT4/POU5F1 (753)	KLF4 (1702)	ZFP281 (578)	REX1/ZFP42 (1480)	MYC (3414)
SMAD1/5(562)	45 (30.3)	34 (18.7)	83 (40.4)	35 (18.9)	36 (17.0)	82 (41.2)	27 (14.1)	69 (35.0)	140 (81.0)
SMAD4(2518)	142 (132.6)	107 (83.5)	245 (182.3)	96 (81.4)	86 (82.4)	214 (182.3)	74 (61.4)	169 (157.9)	370 (365.7)

Figure S17. SMAD-associated genes intersect with the core self-renewal network. Numbers of genes bound by both SMAD proteins and each of core self-renewal transcription factors are indicated in each cell. The core self-renewal transcription factors-bound targets were obtained from Kim et al (2008). The gene numbers analyzed are indicated in the parenthesis following each factor in the column and row headers. The average number of genes overlapped between two random gene groups of the same sizes as the target groups are listed inside the parenthesis. Grey-highlight indicates significant overlap with an empirical *P*-value <0.01.

Supplementary Tables

Table S1 SMAD1/5 binding sites identification (provided as Dataset 1)

Table S2 SMAD4 binding sites identification (provided as as Dataset 2)

Table S3 SMAD1/5 target genes and ChIP-seq verification (provided as Dataset 3)

Table S4 SMAD4 target genes and ChIP-seq verification (provided as Dataset 4)

Table S5 Top 10 SMAD1/5 and SMAD4 binding motifs and their statistics

			BG count	Corrected BG count	FG count	Score	FG%	BG %	FG / Corrected BG
SMAD1/5	Motif 1	GRGTGTGG	126	61	124	928.5	0.214	0.105	2.03
	Motif 2	GAGGYGGG	194	94	154	875.89	0.266	0.162	1.64
	Motif 3	GGCGGGGY	223	108	168	867.08	0.29	0.187	1.56
	Motif 4	CCGCCCCW	145	70	129	855.64	0.223	0.121	1.84
	Motif 5	CCCCGCCY	325	158	216	844.22	0.373	0.273	1.37
	Motif 6	GAARGAGG	135	65	122	841.95	0.211	0.112	1.88
	Motif 7	GWGGGGCG	102	49	105	810.46	0.181	0.085	2.14
	Motif 8	AGGGTGKG	146	71	123	770.16	0.212	0.123	1.73
	Motif 9	TGGGGTGY	101	49	100	754.14	0.173	0.085	2.04
	Motif 10	GGGAGAGK	196	95	144	722.41	0.249	0.164	1.52
SMAD4	Motif 1	TGGGVAGA	1055	513	669	2246.59	0.245	0.188	1.3
	Motif 2	CTTCHCTG	1058	514	668	2208.59	0.245	0.188	1.3
	Motif 3	TCTGCHCT	922	448	591	2052.59	0.216	0.164	1.32
	Motif 4	CCTTCYCA	683	332	465	1991.53	0.17	0.122	1.4
	Motif 5	CTBCAGGA	1008	490	628	1983.54	0.23	0.179	1.28
	Motif 6	TCAGAGBC	776	377	515	1982.97	0.189	0.138	1.37
	Motif 7	CAGAGCCH	1251	608	739	1883.1	0.271	0.223	1.22
	Motif 8	HCCCTCTG	877	426	557	1882.51	0.204	0.156	1.31
	Motif 9	CCTGCCAV	771	375	504	1865.2	0.185	0.137	1.34
	Motif 10	BCCTGCCA	791	384	512	1839.96	0.188	0.141	1.33

FG stands for foreground and BG stands for background

FG%=FG count / Number of binding sites

Corrected BG count is calculated by DME, which is the number of background occurrences if the two sets had the same size

Corrected BG % = Corrected BG count / Number of binding sites

Degenerate consensus codes: R=A or G; Y=C or T; W=A or T; K=G or T; B=C,G or T; V=A,C or G; H=A,C or T

Table S6. Differentially expressed genes in response to BMP4 or noggin (provided as

Dataset 5)

Table S7 GO terms (biological processes) enriched among SMAD1/5-binding genes predicted from ChIP-chip and verified by ChIP-seq

GO Term (Biological Processes)	P-value	Odd Ratio
GO:0035235|ionotropic glutamate receptor signaling pathway	3.07E-005	37.6854545
GO:0060179|male mating behavior	0.00039466	50.2472727
GO:0019725|cellular homeostasis	0.00047994	2.70146628
GO:0042592|homeostatic process	0.00072999	2.12584615
GO:0048878|chemical homeostasis	0.00076842	2.58119552
GO:0007215|glutamate signaling pathway	0.00114281	13.7038017
GO:0055082|cellular chemical homeostasis	0.00126389	2.80449894
GO:0055065|metal ion homeostasis	0.00140586	4.18727273
GO:0065008|regulation of biological quality	0.00149505	1.74394005
GO:0007631|feeding behavior	0.00156124	5.84270613
GO:0050801|ion homeostasis	0.00257151	2.57678322
GO:0006873|cellular ion homeostasis	0.00303224	2.65730769
GO:0007040|lysosome organization	0.00360318	9.42136364
GO:0007275|multicellular organismal development	0.00421685	1.38503553
GO:0006875|cellular metal ion homeostasis	0.00485785	3.81624856
GO:0048856|anatomical structure development	0.00507009	1.42235067
GO:0007033|vacuole organization	0.00509843	8.37454545
GO:0030823|regulation of cGMP metabolic process	0.00561482	16.7490909
GO:0009653|anatomical structure morphogenesis	0.00600934	1.5919928
GO:0031323|regulation of cellular metabolic process	0.00728897	1.34248439
GO:0007399|nervous system development	0.00743474	1.68615009
GO:0006511|ubiquitin-dependent protein catabolic process	0.00773356	3.08535885
GO:0055080|cation homeostasis	0.00884023	2.75327522
GO:0044249|cellular biosynthetic process	0.0099846	1.29790003

Table S8 GO terms (biological processes) enriched among SMAD4-binding genes predicted from ChIP-chip and verified by ChIP-seq

GO Term (Biological Processes)	P-value	Odd Ratio
GO:0048522|positive regulation of cellular process	2.34E-006	1.63269148
GO:0048518|positive regulation of biological process	3.56E-006	1.57287548
GO:0031325|positive regulation of cellular metabolic process	1.19E-005	1.85445874
GO:0010604|positive regulation of macromolecule metabolic process	2.81E-005	1.82182071
GO:0009893|positive regulation of metabolic process	3.04E-005	1.79218214

GO:0042127|regulation of cell proliferation	8.97E-005	1.8473262
GO:0006357|regulation of transcription from RNA polymerase II promoter	0.00011669	1.73408015
GO:0045604|regulation of epidermal cell differentiation	0.00015802	18.473262
GO:0031328|positive regulation of cellular biosynthetic process	0.00023753	1.7535326
GO:0048731|system development	0.00024396	1.34370234
GO:0045935|positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0.00025725	1.78917792
GO:0009891|positive regulation of biosynthetic process	0.00026129	1.74582476
GO:0010557|positive regulation of macromolecule biosynthetic process	0.00036375	1.74516992
GO:0051254|positive regulation of RNA metabolic process	0.00039848	1.83162655
GO:0032504|multicellular organism reproduction	0.00043128	2.92868788
GO:0048609|reproductive process in a multicellular organism	0.00043128	2.92868788
GO:0006810|transport	0.00047598	1.2858419
GO:0051239|regulation of multicellular organismal process	0.0004881	1.58846521
GO:0031323|regulation of cellular metabolic process	0.00049625	1.26915541
GO:0045944|positive regulation of transcription from RNA polymerase II promoter	0.00051082	1.86480879
GO:0051234|establishment of localization	0.0005732	1.28049413
GO:0019222|regulation of metabolic process	0.00060309	1.25582513
GO:0045893|positive regulation of transcription, DNA-dependent	0.00076494	1.78434917
GO:0006812|cation transport	0.00077767	1.6685527
GO:0065009|regulation of molecular function	0.00080872	1.71634219
GO:0030048|actin filament-based movement	0.00087371	6.15775401
GO:0048856|anatomical structure development	0.00090985	1.29725139
GO:0051179|localization	0.00091271	1.24588841
GO:0048513|organ development	0.00092055	1.34755382
GO:0008284|positive regulation of cell proliferation	0.00102719	1.98815376
GO:0045941|positive regulation of transcription	0.00147388	1.68791227
GO:0008037|cell recognition	0.00190472	3.43688596
GO:0010628|positive regulation of gene expression	0.00199625	1.65844746
GO:0007215|glutamate signaling pathway	0.00206969	6.71754983
GO:0045682|regulation of epidermis development	0.00206969	6.71754983
GO:0006091|generation of precursor metabolites and energy	0.00222039	1.87863682
GO:0048511|rhythmic process	0.00223444	2.70941176
GO:0019219|regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0.00254492	1.24759126
GO:0048535|lymph node development	0.00282433	4.86138475
GO:0060349|bone morphogenesis	0.00282433	4.86138475
GO:0050790|regulation of catalytic activity	0.00285151	1.67938746
GO:0044057|regulation of system process	0.00309943	2.09363636
GO:0009889|regulation of biosynthetic process	0.00317847	1.23438848
GO:0048732|gland development	0.00335609	2.35443536
GO:0050730|regulation of peptidyl-tyrosine phosphorylation	0.00341445	3.14438503
GO:0046631|alpha-beta T cell activation	0.00360019	4.61831551

GO:0031326|regulation of cellular biosynthetic process	0.00398016	1.2281393
GO:0001958|endochondral ossification	0.00411143	5.68408063
GO:0050793|regulation of developmental process	0.00414991	1.38977087
GO:0015837|amine transport	0.00423767	2.63903743
GO:0050732|negative regulation of peptidyl-tyrosine phosphorylation	0.00469214	7.9171123
GO:0060255|regulation of macromolecule metabolic process	0.00471335	1.21034392
GO:0006813|potassium ion transport	0.00491057	2.05258467
GO:0032502|developmental process	0.00497162	1.1938268
GO:0051252|regulation of RNA metabolic process	0.00526773	1.23735088
GO:0006090|pyruvate metabolic process	0.00559028	4.19846864
GO:0048519|negative regulation of biological process	0.00559361	1.32306263
GO:0042592|homeostatic process	0.0061358	1.49207116
GO:0030001|metal ion transport	0.00614623	1.57022727
GO:0048523|negative regulation of cellular process	0.00647637	1.33519617
GO:0006006|glucose metabolic process	0.00653866	2.26203209
GO:0007275|multicellular organismal development	0.00653926	1.21040352
GO:0051336|regulation of hydrolase activity	0.00681134	1.98370599
GO:0022612|gland morphogenesis	0.00719895	4.92620321
GO:0000188|inactivation of MAPK activity	0.00720718	6.92747326
GO:0009261|ribonucleotide catabolic process	0.00720718	6.92747326
GO:0006355|regulation of transcription, DNA-dependent	0.00729065	1.22765349
GO:0006811|ion transport	0.00730617	1.42320971
GO:0065008|regulation of biological quality	0.0079514	1.34241709
GO:0060443|mammary gland morphogenesis	0.00846375	12.315508
GO:0060056|mammary gland involution	0.00846375	12.315508
GO:0046426|negative regulation of JAK-STAT cascade	0.00846375	12.315508
GO:0042532|negative regulation of tyrosine phosphorylation of STAT protein	0.00846375	12.315508
GO:0033275|actin-myosin filament sliding	0.00846375	12.315508
GO:0070252|actin-mediated cell contraction	0.00846375	12.315508
GO:0045449|regulation of transcription	0.00875423	1.21596155
GO:0006112|energy reserve metabolic process	0.0090378	3.25998742
GO:0006865|amino acid transport	0.00905159	2.68701993
GO:0007595|lactation	0.00919372	4.61831551
GO:0006066|cellular alcohol metabolic process	0.00985491	1.66676048

Table S9 PCR primers

Gene	Forward Primer	length(bp)	Reverse Primer	length(bp)
For real time PCR				
Accn4	AGGAGGCAGGGGATGAACA	19	TGAGGTGAGTAGGGCCAGTG	20
Alx3	GCTACCAGTGGATTGCCGAG	20	GCTCCCGAGCATACACGTC	19

Kank2	GGAGGAAATTCGGATGGATCT	22	ACTTTCAGTTCTCGCTCTGTGA	22
Camk1	CTCCTTATTGGGACGACATCTCT	23	AGCTGTATCTCCTGCAATCCA	21
Cul1	TTCAGGCCATTGAATAAACAGGT	23	TCTCTGTGTCAGCCAAAAATTGA	23
Dpysl2	CAGAATGGTGATTCCCGGAGG	21	CAGCCAATAGGCTCGTCCC	19
Fosb	GAGCTGACAGATCGACTTCAGG	22	CCGTCTTCCTTAGCGGATGTT	21
Gapdh	CATGGCCTTCCGTGTTCCTA	20	CCTGCTTCACCACCTTCTTGAT	22
Kdm6b	CCCCCATTTCAGCTGACTAA	20	CTGGACCAAGGGGTGTGTT	19
Llgl1	AGAGAGCTGCGAAGGAAGC	19	GCCTCACCCCAGAAAATCCTC	21
Msx2	ATACAGGAGCCCGGCAGATA	20	CGGTTGGTCTTGTGTTTCCT	20
Myh7	AGCAGGAGCTGATTGAGACC	20	TGTGATAGCCTTCTTGGCCT	20
Myst2	CTCGAACTCCAACCGGAAATG	21	CCGGCGTAGGACACTTCAT	19
nestin	GGAGTGTCGCTTAGAGGTGC	20	TCCAGAAAGCCAAGAGAAGC	20
Ngef	CCTGATGAACTGACGCTGGAAC	22	CGCTGAGGGTCTTCCATCTTA	21
Nkx6-2	AAGTCTGCCCCGTCTCAAC	19	GGTCTGCTCGAAAGTCTTCTC	21
Ntrk3	GTCAAGTTCTATGGGGTGTGTG	22	GGCGATGTGGAGCATCTGAG	20
Ptk2b	ATCTTGACCACCCTCACATCG	21	TAGTGTCCCAGCTCCCCATAA	21
Rela	GCGCGGGGACTATGACTTG	19	GCCCGGTTATCAAAAATCGGA	21
Rtn4rl1	TCGAGTACCTCCAAGATGACAT	22	GTAGCTGGTTCTCATGCAGCA	21
Scarf2	ATGAGTGTGGGATAGCGGTGT	21	GGCACTTTGTGTCGCAGTT	19
Sesn2	AGTGTTCTTACCTGGTGGGTT	21	GTAACTTGTTGACCTCGCTGA	21
Shank3	CCTGAAGGTTCTCCGCAAC	19	GTCCAGCAGGGTCGTCAATG	20
Smad1	ACTGAAGCCTCTGGAATGCT	20	GCGGTTCTTATTGTTGGACG	20
Smad5	ATGCCCAGCATATCCAGCAG	20	CAGAAGAAATGGGGTTCAGC	20
Snx2	ACAGATCCAGGGACGAGATTG	21	GGTGGCACGATGTAACCAACA	21
Sox1	TTACTTCCCGCCAGCTCTTC	20	TGATGCATTTTGGGGGTATCTCTC	24
Tal1	CTGATGGTCCTCACACCAAAG	21	CCTCCTGGTCATTGAGTAACTTG	23
Tmem108	GTGGTTCGACTCGCACTCTC	20	CCCCGTAAAGTTACCCTTGTAGA	23
Wwc2	TCATCTGGGAGCAGTCTAGGG	21	TGATAGTCTGTGTCCATCTGGTC	23
Fgf5	CTGTATGGACCCACAGGGAGTAAC	24	ATTAAGCTCCTGGGTCGCAAG	21
Eomes	CCTGGTGGTGTTTTGTTGTG	20	TTTAATAGCACCGGGCACTC	20
Cdx2	CCTGCGACAAGGGCTTGTTTAG	22	TCCCGACTTCCCTTCACCATAC	22
T	CTCTAATGTCCTCCCTTGTTGCC	23	TGCAGATTGTCTTTGGCTACTTTG	24
Gsc	AAACGCCGAGAAGTGGAACAAG	22	AAGGCAGGGTGTGTGCAAGTAG	22
Sox17	AAGAAACCCTAAACACAAACAGCG	24	TTTGTGGGAAGTGGGATCAAGAC	23
Afp	GCCACCGAGGAGGAAGTG	18	AGTCTTCTTGCGTGCCAGC	19
For regular RT-RCR				
Gapdh	ACCACAGTCCATGCCATCAC	20	TCCACCACCCTGTTGCTGTA	20
Nanog	ACCTGAGCTATAAGCAGGTTAAGAC	25	GTGCTGAGCCCTTCTGAATCAGAC	24
Pou5f1	CACGAGTGGAAAGCAACTCA	20	AGATGGTGGTCTGGCTGAAC	20
Zfp42	GATTCACATCCTAACCCACGCA	22	TATCCCCAGTGCCTCTGTCATT	22
For ChIP assay				
Alx3	TTTGGGTCGAGGAGACTCACCTTT	24	TGGGTCAGCTATGGAAGAGCAGAA	24
Ascl2	GGGATGCAGCTGTGAAGAAGA	21	GTGCTGGAGGCGGGGTGAGTTT	22

Camk1	TGGCAGACTAATGCTGTAGGGC	22	AGAAAGTAGCAGAAAGTGAGGC	22
Acap1	CAAGCGAGCCTAGAGAAGAGAG	22	TGAGCATCTAAAGGTCGAGGA	21
Cspg5	GTCCATGATAAGCCTAGTGTA	21	CTTCGGTATCAGGGTGATTTTC	22
Dpysl2	GCTACCCAAGGCTACCTCCAT	21	TCCACGCATCACGGTAAGTTTG	22
Fosb	GCCTAAGAGTTGAGAGCATCCC	22	GATTCAGCCCTTCCCTCGCCA	21
Kdm6b	CATGAGACTGGTCGGAAGAA	20	TTTGTTAGAGACCAGGCTGCC	21
Mycn	CTGGTTACTTGCTCTATATGCC	22	TGGAATCGTGAAGTGAGTGCGT	22
Neurod6	GGTGGGCTTTTGTGTGAGTGAG	22	AGGCACACACCCACAAGTTAC	22
Nkx6-2	TCAGAACGGGAGGACAGCATAGG	23	GAACGAAAAAGGGTGGACTCTAGG	25
Ntrk3	GACCCTTTGGCGCAGCTATG	20	TTCTCAAGTATCCATCCTCCG	21
Sesn2	TTCACACCGCCTTCCACCCT	20	GAATCAAGGGTAAAGTGCTG	20
Sox30	TTATCAGGCTCTGCGGGGTTA	21	ACCCACCATTATCCTGCTTA	20
Spink2	TGATGACCACAACCAGAACCCC	22	GAATGTGGTGGAGGGATGGTGG	22
Tacr2	GCTGAAGATGGGGTGAAGGAAG	22	AAAGTCTATCGTTCAGCAGTG	21
Tmem108	GCGGTTGCTGAGAATACTTGT	21	CACAAACTACGTGGAAACTT	20
Zfp526	CCAGAAACCTACAACTCAACTC	22	TTAGGAGGGGCATCTGGGTG	20