**Stem Cell Reports, Volume *6***

## Supplemental Information

# BMP-SMAD Signaling Regulates Lineage Priming, but Is Dispensable for Self-Renewal in Mouse Embryonic Stem Cells

**Maria Gomes Fernandes, Ruben Dries, Matthias S. Roost, Stefan Semrau, Ana de Melo Bernardo, Richard P. Davis, Ramprasad Ramakrishnan, Karoly Szuhai, Elke Maas, Lieve Umans, Vanesa Abon Escalona, Daniela Salvatori, Dieter Deforce, Wim Van Criekinge, Danny Huylebroeck, Christine Mummery, An Zwijsen, and Susana M. Chuva de Sousa Lopes**

**Supplemental Figures**

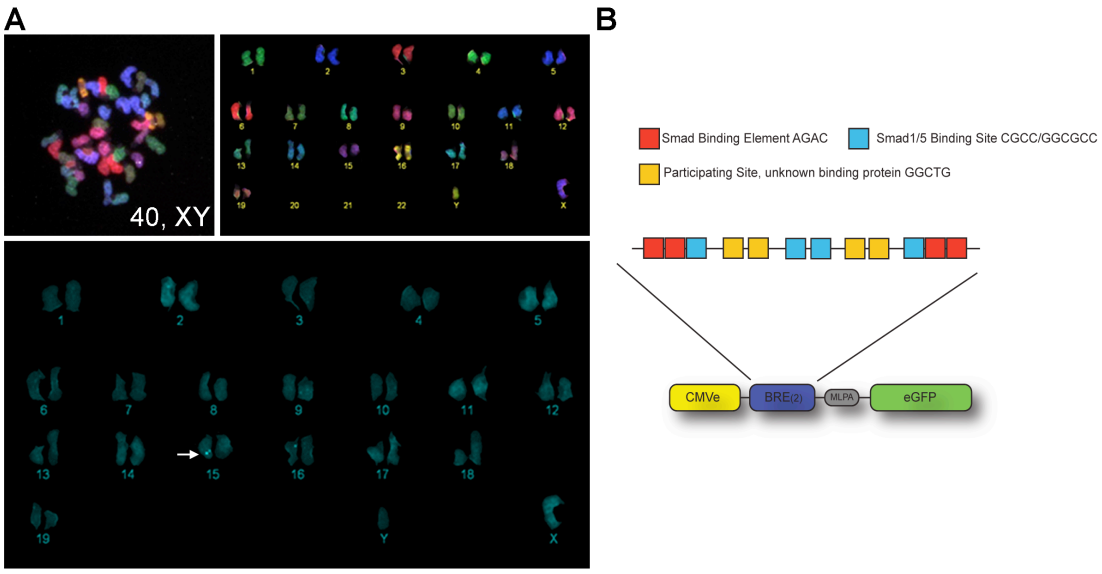



**Figure S1. The *BRE:gfp* construct, related to Figure 1.**
**(A)** Representative karyogram of a XY *BRE:gfp* mESC line. The arrow identifies the subtelomeric region of the Chromosome 15, where the *BRE:gfp* construct was mapped by DNA-FISH.
**(B)** Schematic representation of the *BRE:gfp* construct. Multiple binding elements (red and light blue boxes) are arranged in tandem, both in forward and reverse orientations and placed downstream of the cytomegalovirus (*CMV*) enhancer (*CMVe*) and upstream of a minimal promoter (*MLPA*) in order to drive the expression of *eGFP*.

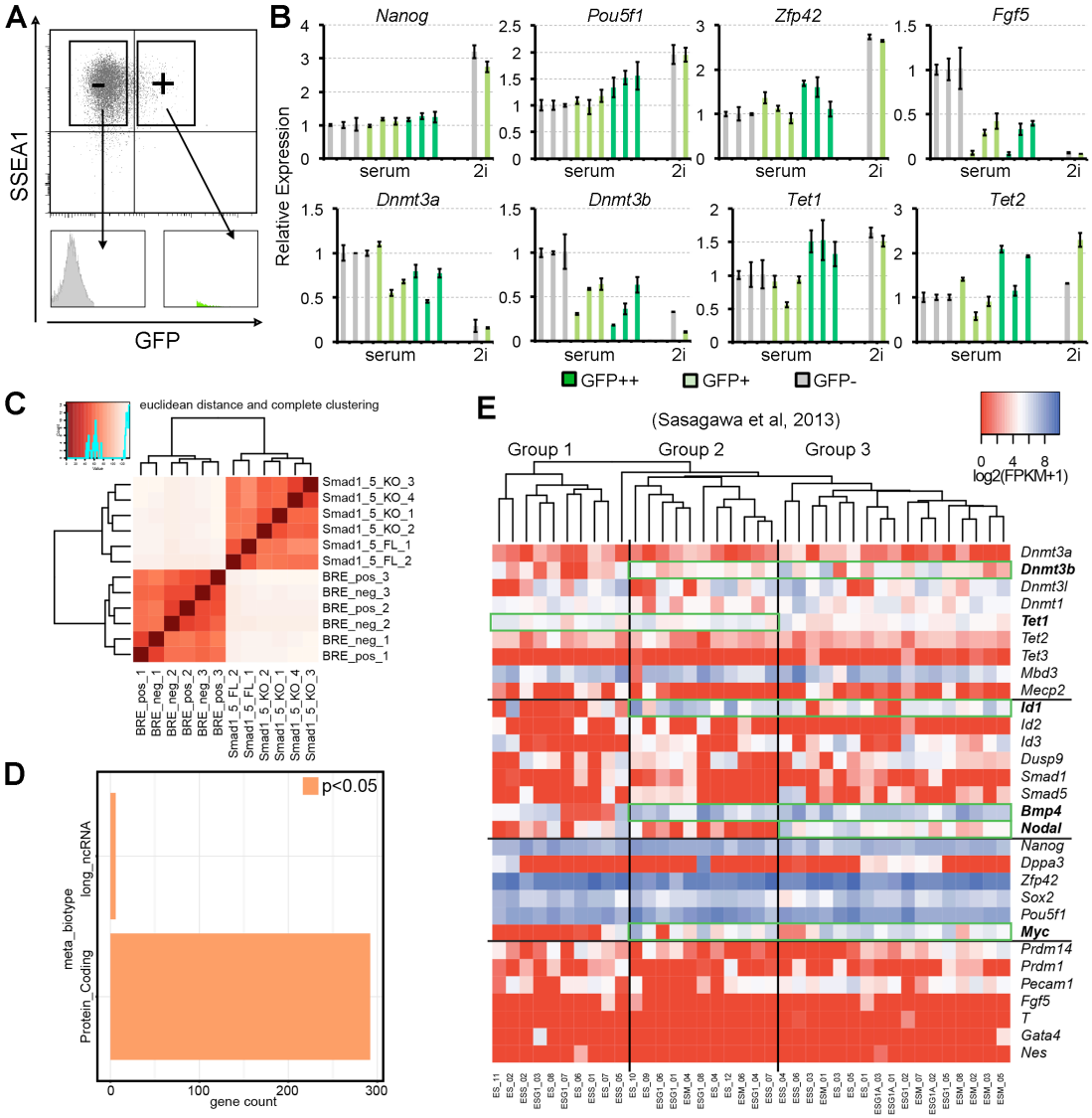



**Figure S2. Characterization and comparison of "serum" and "2i" mESCs, related to Figure 2.**
**(A)** Gatings used to FACS sort two subpopulations (GFP- and GFP+) of "2i" *BRE:gfp* mESCs and the profile of the individual cell groups.
**(B)** Relative expression of several genes in the subpopulations (GFP-, GFP+, GFP++) "serum" and (GFP- and GFP+) "2i" *BRE:gfp* mESCs compared to the GFP- "serum" cells. Each bar represents mean±standard deviation of technical triplicates.
**(C)** Hierarchical clustering of independent "2i" $S1^{fl/fl}S5^{fl/fl}$ mESCs (FL), "2i" $S1^{-/-}S5^{-/-}$ mESCs (KO), "serum" GFP++ (pos) and GFP- (neg) fraction of *BRE:gfp* mESCs.
**(D)** Barplot depicting the number of the significantly differentially expressed genes (DEGs) between "serum" GFP- and GFP++ fraction of *BRE:gfp* mESCs per biotype. In orange DEGs with P <0.05.
**(E)** Heatmap of the log2 fragments per kilobase of transcript per million reads (FPKM) values of 30 genes of interest in 38 individual naïve mESCs.

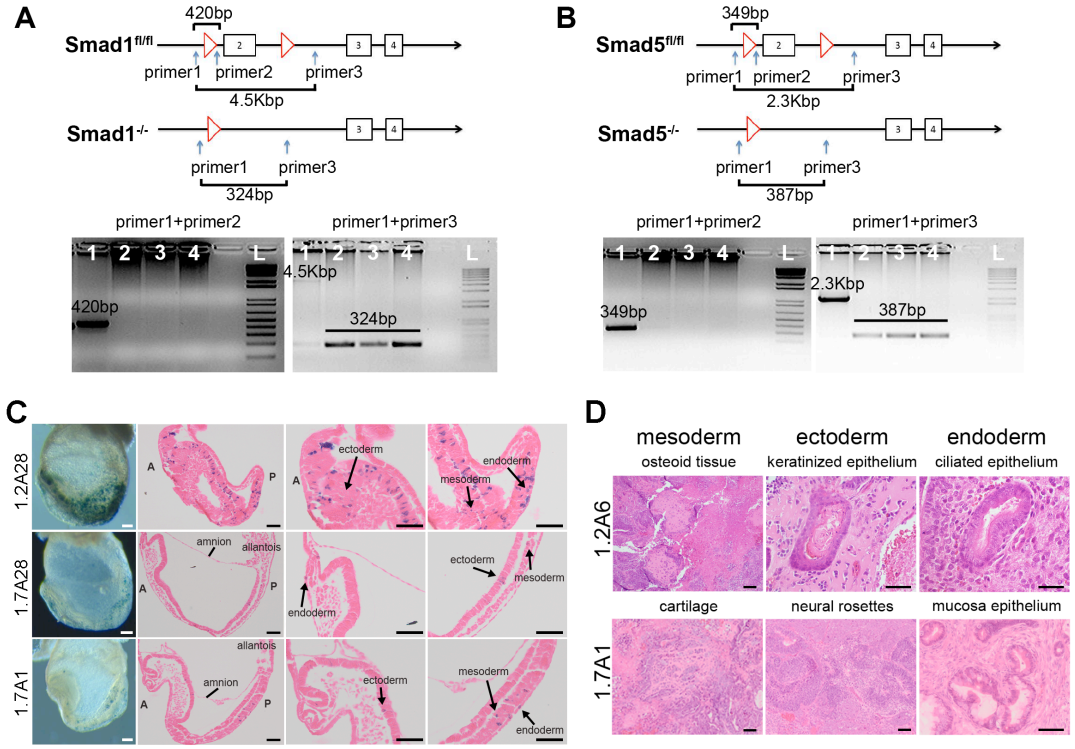
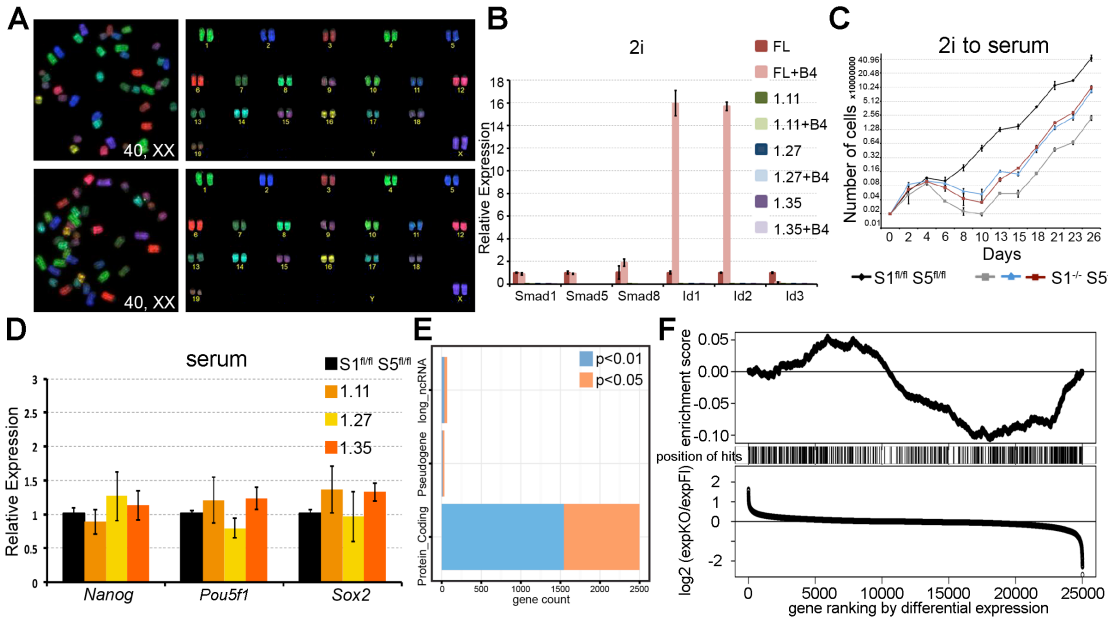



**Figure S3. Derivation and pluripotency state of "2i" *S1*$^{-/-}$*S5*$^{-/-}$ mESCs, related to Figure 3.**
**(A, B)** Scheme representing the location of the primers and the size of the respective genotyping PCR bands before and after excision of the exon 2 of *Smad1* and respective representative PCR gel (A) and of the exon 2 of *Smad5* and representative PCR gel (B). Lane numbers: 1, *S1*$^{fl/fl}$*S5*$^{fl/fl}$ mESCs; 2, 3 and 4, *S1*$^{-/-}$*S5*$^{-/-}$ mESCs; L, DNA ladder 100 Kbp+.
**(C)** β-Galactosidase staining of chimeric embryos isolated at E8.5 generated using 3 different *S1*$^{-/-}$*S5*$^{-/-}$ mESC lines (1.2A28, 1.7A28 and 1.7A1). Whole mount embryos are shown in the left side, and selected paraffin sections show contribution of *S1*$^{-/-}$*S5*$^{-/-}$ mESCs to ectoderm, endoderm and mesoderm (black arrows). A, anterior; P, posterior. Scale bars are 50 µm.
**(D)** Hematoxylin and eosin-stained paraffin sections of teratomas formed after subcutaneous injection of *S1*$^{-/-}$*S5*$^{-/-}$ mESCs (1.2A6 and 1.7A1). The teratomas obtained contained tissues from three embryonic germ layers: mesoderm (osteoid tissue, cartilage); ectoderm (keratinized epithelium, neural rosettes); endoderm (ciliated epithelium, mucosa epithelium). Scale bars are 50 µm.




G

**GGCGCC motif - UP and DOWN - p<2.2e-16**

| | motif detected | motif not detected |
|---|---|---|
| **DEGs*** | 961 | 674 |
| **all genes*** | 12335 | 22236 |
| **DEGs - UP** | 443 | 338 |
| **DEGs - DOWN** | 518 | 336 |

* genes include protein-coding genes, pseudogenes and long non-coding RNAs

**GCCG motif - UP and DOWN - p<2.2e-16**

| | motif detected | motif not detected |
|---|---|---|
| **DEGs*** | 1590 | 45 |
| **all genes*** | 30182 | 4389 |
| **DEGs - UP** | 762 | 19 |
| **DEGs - DOWN** | 828 | 26 |

* genes include protein-coding genes, pseudogenes and long non-coding RNAs

**Figure S4. Characterization of *S1*$^{-/-}$*S5*$^{-/-}$ mESCs, related to Figure 3.**

**(A)** Representative karyogram for independent *S1*$^{-/-}$*S5*$^{-/-}$ mESC lines showing a normal karyotype (40, XX).

**(B)** Relative expression of several *Smad* and *Id* genes in "2i" *S1*$^{-/-}$*S5*$^{-/-}$ mESCs compared to *S1*$^{fl/fl}$*S5*$^{fl/fl}$ mESCs, before and after 1 hour stimulation with 25 ng/ml of BMP4 (+B4) Bars represent mean±standard deviation of relative expression of technical triplicates.

**(C)** Proliferation rate of *S1*$^{fl/fl}$*S5*$^{fl/fl}$ mESCs and three independent *S1*$^{-/-}$*S5*$^{-/-}$ mESCs lines in "serum" during 26 days. Mean±standard deviation is depicted.

**(D)** Relative expression of pluripotency genes in "serum" *S1*$^{-/-}$*S5*$^{-/-}$ mESCs (1.11, 1.27, 1.35) compared to *S1*$^{fl/fl}$*S5*$^{fl/fl}$ mESCs.

**(E)** Number of significantly differentially expressed genes (DEGs) between the "2i" *S1*$^{fl/fl}$*S5*$^{fl/fl}$ and *S1*$^{-/-}$ *S5*$^{-/-}$ mESCs per biotype. In blue DEGs with P <0.01 and in orange DEGs with P <0.05.

**(F)** Enrichment score for SMAD1/5 targets (n=562) identified in mESCs (Fei et al., 2010) calculated using standard gene enrichment analysis (Subramanian et al., 2005) (top panel). Mid panel depicts the position of the SMAD1/5 target hits in the ranked gene list. Genes were ranked by log2 fold-change of expression between *S1*$^{-/-}$*S5*$^{-/-}$ (KO) and *S1*$^{fl/fl}$*S5*$^{fl/fl}$ (FL) mESCs (bottom panel). A gene with a low rank is more highly expressed in the KO.

**(G)** Presence of the putative binding motifs of SMAD1/5 (GGCGCC and GCCG) in the promoter region, defined as ± 2Kb from the transcription start site (TSS) of DEGs (P<0.01) between *S1*$^{-/-}$*S5*$^{-/-}$ and *S1*$^{fl/fl}$*S5*$^{fl/fl}$ mESCs; and, as comparison, presence of the same binding motifs in the promoter region of all genes genome-wide.

**Table S1. Differentially expressed genes between GFP++ and GFP- *BRE:gfp* mESCs and between *S1*$^{fl/fl}$*S5*$^{fl/fl}$ and *S1*$^{-/-}$*S5*$^{-/-}$ mESCs. Related to Figure 2 and 3.**

**Table S2. Differential DNA methylation between GFP++ and GFP- *BRE:gfp* mESCs and between *S1*$^{fl/fl}$*S5*$^{fl/fl}$ and *S1*$^{-/-}$*S5*$^{-/-}$ mESCs. Related to Figure 2 and 3.**

**Table S3. Counts of SMAD binding motifs GGCGCC/GCCG in the promoters of differentially expressed genes between *S1*$^{fl/fl}$*S5*$^{fl/fl}$ and *S1*$^{-/-}$*S5*$^{-/-}$ mESCs. Related to Figure 3.**

**Table S4. List of primers used for qPCR and genotyping of *Smad1* and *Smad5.* Related to Figure 2 and 4.**

| Gene | Forward primer | Reverse primer | Melting temperature Tm | Reference |
|---|---|---|---|---|
| *Nanog* | CTTTCACCTATTAAGGTGCTTGC | TGGCATCGGTTCATCATGGTAC | 60ºC | Kurimoto et al, 2007, Nature Protocols, 2, 739-752 |
| *Pou5f1* | AGAGGGAACCTCCTCTGAGC | TTCTAGCTCCTTCTGCAGGG | 60ºC | Wahlestedt et al, 2013, Blood, 121, 4257-4264 |
| *Zfp42* | TCCATGGCATAGTTCCAACAG | TAACTGATTTTCTGCCGTATGC | 60ºC | Kurimoto et al 2006, Nucleic Acids Research, 34, e42(1-17) |
| *Sox2* | AGCTCGCAGACCTACATGAA | CCCTGGAGTGGGAGGAA | 60ºC | Yellajoshyula et al 2011, PNAS, 108, 3294-3299 |
| *Dppa3* | GACCCAATGAAGGACCCTGAA | GCTTGACACCGGGGTTTAG | 60ºC | - |
| *Esrrb* | AACCGAATGTCGTCCGAAGAC | GTGGCTGAGGGCATCAATG | 60ºC | - |
| *Fgf5* | ACTGAAAAGACAGGCCGAGA | TGAACCTGGGTAGGAAGTGG | 60ºC | - |
| *T* | GTCTAGCCTCGGAGTGCCT | CCATTGCTCACAGACCAGAG | 60ºC | Yellajoshyula et al 2011, PNAS, 108, 3294-3299 |
| *Sox17* | CTGCACAACGCAGAGCTAAG | GCTTCTCTGCCAAGGTCAAC | 60ºC | - |
| *Pax6* | CTTCATCCGAGTCTTCTCCG | CGGGACTTCAGTACCAGGG | 60ºC | - |
| *Sox1* | AGACTTCGAGCCGACAAGAG | AACTGTGCAAACAGGTGCAG | 60ºC | Guttman et al, 2011, Nature, 477, 285-300 |
| *Dnmt3a* | GACTCGCGTGCAATAACCTTAG | GGTCACTTTCCCTCACTCTGG | 60ºC | Kurimoto et al, 2007, Nature Protocols, 2, 739-752 |
| *Dnmt3b* | CTCGCAAGGTGTGGGCTTTTGTAAC | CTGGGCATCTGTCATCTTTGCACC | 60ºC | Kurimoto et al, 2007, Nature Protocols, 2, 739-752 |
| *Tet1* | GATGCCATGAGTGTCACCAC | AAAGATGGTGGGTTCTGCAC | 60ºC | - |
| *Tet2* | GGGGTTGGAGCAAGTACAAA | CGGGTGTGTGTCATTTGAAG | 60ºC | - |
| *Id1* | CCCTGAACGGCGAGATCA | AAAAAACCTCTTGCCTCCTGAA | 60ºC | Moya et al, 2012, Developmental Cell, 22, 501-514 |
| *Id2* | CTCCAAGCTCAAGGAACTGG | AGGCTGACGATAGTGGGATG | 60ºC | - |
| *Id3* | ATCCTGCAGCGTGTCATAGACT | AGGCGTTGAGTTCAGGGTAAGT | 60ºC | - |
| *Smad1* | AACAGCAGCTACCCCAACTCTC | CGTAAGCAACTGCCTGAACATC | 60ºC | - |
| *Smad5* | CCAGCCGTGAAGGCGATTG | GCCTTTTCTGCCCATTTCTCT | 60ºC | - |
| *Smad8* | GAACCCCTCATGCCGCACAA | GAACACCAGTGCTGGGGTTCCT | 60ºC | - |
| *Gfp* | AACCACTACCTGAGCACCC | ACCTCTACAAATGTGGTATG | 60ºC | Hezroni et al, 2011, Nucleus, 2, 300-309 |
| *Gapdh* | CGTCCCGTAGACAAAATGGT | TTGATGGCAACAATCTCCAC | 60ºC | Oh et al, 2013, Exp Mol Med, 45, e23(1-8) |
| *Actb* | ACCATGTACCCAGGCATTG | TACTTGCGCTCAGGAGGAG | 60ºC | Egorova et al, 2011, Developmental Dynamics, 240, 1670-1680 |
| | | | | primers without reference were designed and tested in this study |

| Gene | Forward primer | Reverse primer | Melting temperature Tm | Reference |
|---|---|---|---|---|
| *Smad1* | (P1) TCCTCATCAGAAAAGAACCCTAAGG | (P2) AAGATTCCTCATTCCCAAGTCCC | 58ºC | Tremblay et al., 2001, Development, 128: 3609-3621 |
| | (P1) TCCTCATCAGAAAAGAACCCTAAGG | (P3) TGGAGATAAATGTGCGATTCCG | 58ºC | primer sequences provided by Liz Robertson |
| *Smad5* | (P1) GAGCGTCTTCCTTAGCTAATGTG | (P2) CACTGGCAAAGCAGAGGTTCAGA | 58ºC | Umans et al, 2003, Genesis, 37, 5-11 |
| | (P1) GAGCGTCTTCCTTAGCTAATGTG | (P3) AAAAATCAGCGCTCGACACG | 58ºC | |

## Supplemental Experimental Procedures

### Blastocyst collection

All animal procedures here described were approved by the local animal ethical committee. Blastocysts from CBA/Bl6 females crossed with *BRE:gfp* heterozygous CBA/Bl6 males were isolated on embryonic day (E)3.5 using M2 with HEPES (Sigma-Aldrich) containing 75 µg/ml of bovine serum albumin (BSA, Life Technologies). E0.5 was considered the noon of the day of the plug. The blastocysts were cultured for 24 hours in a drop (30 µl) of EmbryoMax KSOM+AA with phenol red (Chemicon) under mineral oil (Sigma-Aldrich) at 37ºC in humidified air. Next day, the blastocysts were washed in Dulbecco's phosphate buffered saline (DPBS, Life Technologies) without calcium and magnesium, treated with acid Tyrode's solution for 3 minutes at room temperature (RT) to remove the zona pellucida and placed individually in separate organ culture dishes (Fisher Scientific) for mESC derivation.

### Derivation of *BRE:gfp* mESCs, $S1^{fl/fl}S5^{fl/fl}$ mESCs and Cre-recombination to obtain $S1^{-/-}S5^{-/-}$ mESCs

Conditional knockout mESCs for *Smad1* and *Smad5* ($S1^{fl/fl}S5^{fl/fl}$) mESCs were derived by crossing homozygous mice carrying both the *Smad1* conditional allele ($Smad1^{RobPC}$) (Tremblay et al., 2001) and the *Smad5* conditional allele ($Smad5^{tm1Huy2}$) (Umans et al., 2003) and were hemizygous for the R26R Cre-reporter transgene (Soriano, 1999). During derivation of *BRE:gfp* mESCs and $S1^{fl/fl}S5^{fl/fl}$ mESCs, blastocysts were cultured for 3 days in either "2i" medium [N2B27 medium (1:1 mixture of Dulbecco's Modified Eagle Medium DMEM/F12 1:1 nutrient mix (Life Technologies) and Neurobasal (Life Technologies), with 1x non-essential aminoacids (NEAA) (Life Technologies), 50 µg/ml BSA, 0.1 mM 2-mercaptoethanol (Life Technologies), 50 U/ml penicillin and 50 µg/ml streptomycin (Life Technologies), 1x N2 (Life Technologies) and 1x B27 (Life Technologies)) and 2000 U/ml mouse leukemia inhibitory factor (LIF) (Millipore), 1 µM PD0325901 (Axon) and 3 µM CHIR99021 (Axon)] on 0.1% gelatin-coated organ dishes; or in "serum" medium [DMEM+glutamax (Life Technologies) with 15% fetal calf serum (FCS) (Life Technologies), 1x NEAA, 0.1 mM 2-mercaptoethanol, 50 U/ml penicillin and 50 µg/ml streptomycin and 1000 U/ml LIF] in organ dishes coated with FCS. Thereafter, individual ICM outgrowths were isolated mechanically, washed 3x in DPBS, placed in a drop of 0.25% trypsin/EDTA (Life Technologies) for 5 minutes at RT and disrupted mechanically by pipetting. The cell clumps were placed directly in either "2i" on gelatin or "serum" on MEFs and cultured for an additional 3-5 days. Emerging mESCs colonies (passage 1, P1) were passed using 0.05% trypsin/EDTA.

The excision of the Smad1 and Smad5 floxed alleles was achieved by homologous recombination using a Cre recombinase-expression vector (pEFBOS-CreIRESpuro) as described (Davis, 2008). Briefly, $8x10^6$ cells were suspended in 750 µl phosphate buffered saline (PBS) and 10 µg of the vector added directly to the cell suspension. Electroporation was performed as described (Barnett and Köntgen, 2001). The electroporated cells were plated in "2i" medium. Selection with 2 µg/ml puromycin was started 48 hours after plating and maintained for 48 hours. From the resulting colonies, 96 clones were manually isolated, grown in "2i" conditions and genotyped as described (Tremblay et al., 2001; Umans et al., 2003).

Karyotyping of the mESC lines and DNA-FISH for GFP in *BRE:gfp* mESCs was performed as described (Szuhai and Tanke, 2006).

### Generation and analysis of chimeric embryos

Blastocysts where obtained by superovulation of CD1(HsD) females. Mouse chimeric embryos were produced by injection of "2i" $S1^{-/-}S5^{-/-}$ mESCs into the blastocoel cavity of blastocysts. Per line, around 20 to 30 injected blastocysts were obtained. Those were transferred into uteri of E3.5 pseudo-pregnant females and 4 days later the embryos were recovered, fixed 2 hours at RT in 25% gluteraldehyde/2% formaldehyde in PBS, washed in PBS and incubated overnight (o/n) at 30°C in an humidified chamber in freshly made staining solution (1 mg/ml of X-gal, 2 mM of $MgCl_2$, 5 mM of $K_3Fe(CN)_6$ and 5 mM of $K_4Fe(CN)_6$x3H20 in PBS) previously heated to 50°C to avoid precipitation. Thereafter, the embryos were postfixed o/n with 4% paraformaldehyde (PFA) at 4°C, individually embedded in 2% low melting point agarose (Life Technologies), followed by inclusion in paraffin, sectioned (7 µm) and eosin stained following standard procedures.

### Teratoma formation assay

For teratoma formation assays, "2i" $S1^{-/-}S5^{-/-}$ mESC were trypsinized and $1x10^6$ cells (per injection) were resuspended in 300 µl ice cold 1:1 culture medium and Matrigel growth factors reduced

(Corning) and drawn into 1 ml syringe immediately before the injection. NOD.CB17-Prkdc$^{scid}$/NcrCr mice were injected in the right dorso-lateral area. Per mESCs line, 3 mice were injected. Animals were monitored for weight and health, and sacrificed once the tumor reached 1 cm$^3$. Teratomas were surgically removed, fixed o/n in 4% PFA, paraffin embedded, sectioned (5 µm) and stained for hematoxiline and eosin by standard procedures.

**Immunofluorescence and alkaline phosphatase activity**

Cells were fixed with 4% PFA for 15 minutes at RT, permeabilized with 0.1% Triton-X (Sigma-Aldrich) in PBS for 8 minutes at RT, blocked with 100 µg/ml BSA in 0.05% Tween 20 (Millipore) in PBS (blocking solution) for 1 hour at RT and incubated o/n at 4ºC with the primary antibodies. Primary antibodies used were rabbit αNANOG (1:200, ab80892, Abcam), goat αPOU5F1 (1;100, sc8628, Santa Cruz), rabbit αID1 (1:100, sc488, SantaCruz), goat αBrachyury T (1:100, sc17743, SantaCruz), goat αSOX17 (1:100, AF1924, R&D Systems) and goat αSOX1 (1:100, AF3369, R&D Systems). Next day, cells were incubated with the secondary antibodies diluted in blocking solution for 1 hour at RT. Secondary antibodies were Alexa Fluor 488 donkey αgoat (1:500, A-11055, Life Technologies), Alexa Fluor 594 donkey αgoat (1:500, A-11058, Life Technologies), Alexa Fluor 594 donkey αmouse (1:500, A-21203, Life Technologies) and Alexa Fluor 555 donkey αrabbit (1:500, A-31572, Life Technologies). Thereafter, cells were treated with DAPI (Life Technologies) 1:1000 in PBS, washed and mounted using ProLong Gold (Life Technologies). The assay for phosphatase activity was performed as described (Lawson et al., 1999).

**Imaging and quantification**

Bright field images were made with a Nikon eclipse Ti-S inverted microscope coupled to a Nikon Digital Sight DS-2 MBW (Nikon) operating under the NIS-elements BR version 3.0 software (Nikon). Confocal images were made on a Leica TCS SP8 confocal microscope (Leica, Mannheim) operating under the Leica Application Suite Advanced Fluorescence software (Leica, Mannheim).
Quantification of NANOG heterogeneity in "serum" *BRE:gfp* mESC was determined in the maximum intensity projection of z-stack imaging covering the entire volume of each colony using the SP8 confocal. NANOG-positive cells in each colony (total of n=16 colonies from N=3 independent experiments) were manually counted 3 times and averaged. Statistical analysis was performed using a Student's *t*-test (two-tailed, unequal variance), *P≤0.05.

**Quantitative reverse-transcription polymerase chain reaction (qPCR)**

RNA isolation was performed using RNeasy Micro Kit (Qiagen) for a maximum of 45 µg of RNA or RNeasy MiniKit (Qiagen) for a maximum of 100 µg RNA, following the manufacturer's instructions. The cDNA was obtained using iScript™ cDNA Synthesis Kit (BioRad) following manufacturer's instructions. qPCR was performed using iQ SYBR Green Supermix (Biorad) on the CFX96$^{TM}$ Real-time system, C1000$^{TM}$ Thermal Cycler (Biorad). All the samples were analysed in technical triplicates. The primers used are listed in Table S4. The qPCR conditions were 1x (95°C, 3 minutes), 40x (95°C, 15 seconds; 60°C, 30 seconds; 72°C, 45 seconds) and 1x (95°C, 10 seconds; 65°C, 5 seconds; 95°C, 50 seconds). Expression was normalized to the housekeeping genes *Gapdh* and *Actb* using the ΔΔCt method. Statistical analysis was performed using a Student's *t*-test (two-tailed, unequal variance), *P≤0.05, **P≤0.01.

**FACS sorting and analysis**

Pre-plated "FCS" *BRE;gfp* mESCs were resuspended in FACS buffer (100 µg/ml BSA in DPBS), incubated with mouse αSSEA1 IgM (1:50, sc21702, Santa Cruz) diluted in FACS buffer 20 minutes on ice, washed with FACS buffer and incubated with secondary antibody Alexa Fluor 647 goat αmouse IgM (1:500, A-21238, Life Technologies) diluted in FACS buffer 20 minutes on ice and resuspended in FACS buffer for FACS analysis on a LSR II Flow Cytometer (BD BioSciences) or FACS sorting on a FACSAria III Flow Cytometer (BD BioSciences). Results were processed using FACSDiva version 6.0 software (BD BioSciences).

**Western blotting**

"2i" mESCs were washed twice with ice cold DPBS and scraped in lysis buffer [(50 mM Tris/HCl pH7.5, 170 mM NaCl, 0.5% NP40 (ICN Biomedicals), 400 mM sodium orthovanadate (Sigma-Aldrich), 45 mM sodium pyrophosphate, 1 mM sodium floride (Sigma-Aldrich), 10 mM EDTA and 1:100 protease inhibitor cocktail (Sigma-Aldrich)]; "FCS" mESCs were first pre-plated 45 minutes at RT and lysed in lysis buffer for 30 minutes on ice, with pipetting every 10 minutes. After centrifugation at 4ºC for 10 minutes at 24.000 G, the supernatant was collected. Protein concentration

was measured using Pierce BCA Protein Assay Kit (Thermo Scientific) according to manufacturer's instructions. Samples were run on 10% or 8% acrylamide gels, detection was done using Western Lighting Ultra (Perkin Elmer), according to manufacturer's specifications and imaging was made using a Fuji LAS 3000 mini (R&D Systems). Primary antibodies were rabbit αPSMAD5 (1:1000, ab76296, Abcam, antibody cross-reacts with PSMAD1/8, personal communication EM), mouse αSMAD1 (1:500, LS-C184471, Lifespan Biosciences, antibody cross-reacts with SMAD5, personal communication EM), mouse αGFP (1:500, sc9996, Santa Cruz), rat αTubulin (1:1000, MAB1864, Millipore) and secondary antibodies were donkey αmouse HRP (1:25000, 715-035-150, Jackson Immuno Research), donkey αrabbit HRP (1:25000, 711-035-152, Jackson Immuno Research) and donkey αrat HRP (1:25000, 712-035-150, Jackson Immuno Research).

**RNA-sequencing (RNAseq) and reduced representation bisulfite sequencing (RRBS)**

For RNA-seq and RRBS, "serum" BRE:gfp mESC were preplated, immunostained for SSEA1 as described above and GFP- and GFP++ subpopulations from 3x different passages were FACS sorted; and two independent $S1^{fl/fl}S5^{fl/fl}$ mESC lines and four independent $S1^{-/-}S5^{-/-}$ mESC clones grown in "2i" were collected.

RNA was isolated using RNeasy MiniKit (Qiagen), RNA integrity number (RIN) was measured using a 2100 Bioanalyzer (Agilent Technologies). The sequencing libraries were generated using TruSeq Stranded Total RNA kit (Illumina) and sequenced on an Illumina Hiseq 2000 sequencer.

DNA was isolated using Wizard Genomic DNA purification kit (Promega) and 1 µg gDNA was used for digestion by MSP1 enzyme. Following o/n incubation at 37°C, digestion reaction were terminated by adding 0.5 M EDTA and the DNA was further purified on a GeneJET PCR purification column. Libraries were prepared using NEBNext Ultra DNA library preparation kit (Illumina) and methylated adapters added. Subsequently, adapter ligated fragments were bisulfite converted using EZ DNA Methylation Gold kit (Zymo Research). After 14 PCR cycles, the product was purified using AMPure XP beads. Quality of libraries was checked on a High sensitivity DNA chip (Agilent) and sequencing was done on an Illumina HiSeq2500 PE 2x50bp.

**RNAseq and RRBS data analysis**

*RNAseq expression data:* To map the sequenced reads, a STAR (version 2.4.1d) index was created based on the mouse mm10 transcriptome (Ensemble build GRCm38) and paired-end reads were directly aligned to this index. A count table for annotated genes was produced using featureCounts version 1.4.6 and genes were further classified in different biotypes based on Vega gene and transcript annotation (http://vega.sanger.ac.uk/info/about/gene_and_transcript_types.html). The raw counts were imported in the R package DESeq2 for differential expression. To determine significantly DEGs between GFP++ and GFP- or $S1^{-/-}S5^{-/-}$ and $S1^{fl/fl}S5^{fl/fl}$ mESCs we made use of a design matrix to block respectively for time and strain specific effects and applied a cut-off of 0.01 and/or 0.05 on the p-values (P) adjusted for multiple testing hypothesis. For intuitive visualization and comparison of gene expression levels, we calculated Transcript Per Million (TPM) values.

*RNAseq hierarchical clustering:* Unsupervised hierarchical clustering of all samples was performed on the DESeq2 based variance normalized counts using Euclidean distance and complete linkage.

*RNAseq gene ontology:* Enrichment analysis for gene ontology (GO) terms was made with the R package topGO based on DEGs ($P < 0.05$) and utilizing Fisher's exact test.

*RNAseq motif sequence analysis:* To perform simple motif analysis, we defined promoter regions as ± 2Kb from the transcription start site (TSS) and counted the occurrences for putative binding sites of SMAD1/5 (GCCG and GGCGCC) for all (up and down) DEGs ($P<0.01$) between $S1^{-/-}S5^{-/-}$ and $S1^{fl/fl}S5^{fl/fl}$ mESCs; and, as comparison, the promoter region of all genes belonging to gene biotypes: protein-coding, pseudogenes and long non-coding RNAs. One-sided Fisher's Exact was used to determine significant overrepresentation of these motifs in promoter regions of DEGs relative to the genome wide promoter regions.

*RRBS Genome specific region assignment:* Sequencing reads were mapped to mouse genome mm10 using bismark version 0.14.1 and analysed further with the R package methylKit. In brief, we considered only CpGs located in regions with a depth of coverage of at least five reads and filtered out the top 0.01% CpGs. To normalize for read coverage between samples we used a median-based scaling factor. Only CpGs covered in all compared samples were retained for further analysis. The genome was binned in 600bp tiles as these were considered optimal for robust detection of Differentially Methylated Regions (DMRs) based on a pairwise comparison analysis of a range of tiles (100bp to 1000bp with 100bp increments). To visualize global methylation changes we pooled sample replicates. The methylation level of each sampled tile was estimated as the number of reads reporting a C, divided by the total number of reads reporting a C or T within that tile. Furthermore, tiles were annotated to the

closest gene based on the distance to its TSS. To assign tiles to genes we used the Ensemble GRCm38 transcriptome. To assign tiles to enhancers we used mESC mm9 enhancers regions available for download at http://chromosome.sdsc.edu/mouse/download.html and converted these to mm10 coordinates using CrossMap version 0.1.8. Tiles were assigned to promoters if they overlap within the ± 2Kb region around a TSS. We used a CpG observed/expected ratio of 0.325 to distinguish low- and high-CpG density promoters as described (Etchegaray et al., 2015). Regions that do not belong to any of the aforementioned regions (e.g. intergenic space) are described as "no annotation" for simplification.
*RRBS global methylation profile:* To quantitatively assess global DNA methylation changes, we created histograms for tiles (methylation change > 20%) and performed a one-sided two-sample Kolmogorov-Smirnov test to determine significant distribution differences between populations.

**Analysis of published single cell RNAseq data and SMAD1/5 ChIP data**
*Single cell RNAseq:* Expression levels of DNA methyltransferases, 5-methylcytosine hydroxylases, BMP responsive genes, BMP signaling pathway genes, pluripotency genes and early differentiation genes were extracted from the transcriptomes of 38x "serum" mESCs single cells deposited in Gene Expression Omnibus under accession number GSE42268 (Sasagawa et al., 2013). Data was analysed and visualized using R statistics version 3.0.1.
*SMAD1/5 ChIP-on-chip data*: To calculate the enrichment of SMAD1/5 targets identified by ChIP-on-chip (Fei et al., 2010), we used gene set enrichment analysis as described (Subramanian et al., 2005). Hits were not weighted and p-values were calculated by permuting genes.

**Supplemental references**

Reference List


Barnett, L., and Köntgen, F. (2001). Gene Targeting in a Centralized Facility. In Gene Knockout Protocols, M. Tymms, and I. Kola, eds. (Humana Press), pp. 65-82.

Davis, R.P. (2008). Targeting a GFP reporter gene to the MIXL1 locus of human embryonic stem cells identifies human primitive streak-like cells and enables isolation of primitive hematopoietic precursors. Blood *111*, 1876-1884.

Etchegaray, J.P., Chavez, L., Huang, Y., Ross, K.N., Choi, J., Martinez-Pastor, B., Walsh, R.M., Sommer, C.A., Lienhard, M., Gladden, A.*, et al.* (2015). The histone deacetylase SIRT6 controls embryonic stem cell fate via TET-mediated production of 5-hydroxymethylcytosine. Nature cell biology *17*, 545-557.

Fei, T., Xia, K., Li, Z., Zhou, B., Zhu, S., Chen, H., Zhang, J., Chen, Z., Xiao, H., Han, J.D.*, et al.* (2010). Genome-wide mapping of SMAD target genes reveals the role of BMP signaling in embryonic stem cell fate determination. Genome research *20*, 36-44.

Lawson, K.A., Dunn, N.R., Roelen, B.A., Zeinstra, L.M., Davis, A.M., Wright, C.V., Korving, J.P., and Hogan, B.L. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. Genes & development *13*, 424-436.

Sasagawa, Y., Nikaido, I., Hayashi, T., Danno, H., Uno, K.D., Imai, T., and Ueda, H.R. (2013). Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. Genome biology *14*, R31.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. Nature genetics *21*, 70-71.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America *102*, 15545-15550.

Szuhai, K., and Tanke, H.J. (2006). COBRA: combined binary ratio labeling of nucleic-acid probes for multi-color fluorescence in situ hybridization karyotyping. Nat Protocols *1*, 264-275.

Tremblay, K.D., Dunn, N.R., and Robertson, E.J. (2001). Mouse embryos lacking Smad1 signals display defects in extra-embryonic tissues and germ cell formation. Development *128*, 3609-3621.

Umans, L., Vermeire, L., Francis, A., Chang, H., Huylebroeck, D., and Zwijsen, A. (2003). Generation of a floxed allele of Smad5 for cre-mediated conditional knockout in the mouse. Genesis *37*, 5-11.