Please cite this article in press as: Gomes Fernandes et al., BMP-SMAD Signaling Regulates Lineage Priming, but Is Dispensable for Self-Renewal in Mouse Embryonic Stem Cells, Stem Cell Reports (2016), http://dx.doi.org/10.1016/j.stemcr.2015.11.012

Stem Cell Reports
Article
ISSCR
OPEN ACCESS

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

Maria Gomes Fernandes,[1] Ruben Dries,[2,3] Matthias S. Roost,[1] Stefan Semrau,[4] Ana de Melo Bernardo,[1] Richard P. Davis,[1] Ramprasad Ramakrishnan,[1] Karoly Szuhai,[5] Elke Maas,[6] Lieve Umans,[2,3,6] Vanesa Abon Escalona,[6] Daniela Salvatori,[1,7] Dieter Deforce,[8] Wim Van Criekinge,[9] Danny Huylebroeck,[2,3] Christine Mummery,[1] An Zwijsen,[6] and Susana M. Chuva de Sousa Lopes[1,10,*]

[1]Department Anatomy and Embryology, Leiden University Medical Center, Leiden 2333 ZC, the Netherlands
[2]Department Development and Regeneration, Laboratory of Molecular Biology (Celgen), KU Leuven, Leuven 3000, Belgium
[3]Department of Cell Biology, Erasmus University Medical Center, Rotterdam 3015 CN, the Netherlands
[4]Leiden Institute of Physics, Leiden University, Leiden 2333 CA, the Netherlands
[5]Department Molecular Cell Biology, Leiden University Medical Center, Leiden 2333 ZC, the Netherlands
[6]Department Human Genetics, VIB Center for the Biology of Disease, KU Leuven, Leuven 3000, Belgium
[7]Center Laboratory Animal Facility, Leiden University Medical Center, Leiden 2333 ZC, the Netherlands
[8]Laboratory of Pharmaceutical Biotechnology, Faculty of Pharmaceutical Sciences, Ghent University, Ghent 9000, Belgium
[9]Mathematical Modelling, Statistics and Bio-informatics, Faculty Bioscience Engineering, Ghent University, Ghent 9000, Belgium
[10]Department Reproductive Medicine, Ghent University Hospital, Ghent 9000, Belgium
*Correspondence: lopes@lumc.nl
http://dx.doi.org/10.1016/j.stemcr.2015.11.012

SUMMARY

Naive mouse embryonic stem cells (mESCs) are in a metastable state and fluctuate between inner cell mass- and epiblast-like phenotypes. Here, we show transient activation of the BMP-SMAD signaling pathway in mESCs containing a BMP-SMAD responsive reporter transgene. Activation of the BMP-SMAD reporter transgene in naive mESCs correlated with lower levels of genomic DNA methylation, high expression of 5-methylcytosine hydroxylases *Tet1/2* and low levels of DNA methyltransferases *Dnmt3a/b*. Moreover, naive mESCs, in which the BMP-SMAD reporter transgene was activated, showed higher resistance to differentiation. Using double *Smad1;Smad5* knockout mESCs, we showed that BMP-SMAD signaling is dispensable for self-renewal in both naive and ground state. These mutant mESCs were still pluripotent, but they exhibited higher levels of DNA methylation than their wild-type counterparts and had a higher propensity to differentiate. We showed that BMP-SMAD signaling modulates lineage priming in mESCs, by transiently regulating the enzymatic machinery responsible for DNA methylation.

INTRODUCTION

Culture conditions affect features of mouse embryonic stem cells (mESCs), such as their proliferation, gene expression, epigenetic status, self-renewal, and capacity for multilineage differentiation (Marks et al., 2012; Tesar et al., 2007). In culture medium with fetal calf serum, naive mESCs grown on mouse embryonic fibroblast feeder cells (here abbreviated as "serum") transit between inner cell mass (ICM)-like and epiblast-like pluripotency states (Sasai et al., 2013; Trott and Martinez Arias, 2013). However, when cultured in serum-free conditions with inhibitors of mitogen-activated protein kinase and glycogen synthase kinase 3 signaling, also called "2i" medium, mESCs become more homogeneous and adopt the more ICM-like or "ground" state (Marks et al., 2012; Nichols et al., 2009; Ying et al., 2003). The observation that naive mESCs interconvert between pluripotent states while remaining uncommitted has raised the suggestion that such heterogeneity may allow the cells to respond differently to environmental cues. In agreement, subpopulations of naive mESCs show different potentials to differentiate (Graf and Stadtfeld, 2008; Hanna et al., 2009; Hayashi et al., 2008). How the metastable transcriptional and epigenetic diversity of cultured mESCs is regulated and maintained has remained elusive.

The two notable characteristics of mESCs are their capacity to self-renew and differentiate into all embryonic lineages (Niwa et al., 1998). In mESCs, pluripotency is maintained by a core network of regulatory transcription factors, including *Pou5f1*, *Sox2*, and *Nanog* (Kashyap et al., 2009; Kim et al., 2008; Marson et al., 2008; Navarro et al., 2012); the balance between self-renewal and differentiation is regulated by protein-encoding genes that include *Id1* and *Dusp9*, both downstream targets of the bone morphogenetic protein (BMP) signaling pathway (Li and Chen, 2013). Moreover, it has been shown that both the BMP and TGFβ (via NODAL) SMAD-mediated signaling pathways are involved in maintaining heterogeneity of NANOG in naive mESCs (Galvin-Burgess et al., 2013). Conversely, NANOG may attenuate BMP signaling via a feedback loop that involves titration of phosphorylated

Please cite this article in press as: Gomes Fernandes et al., BMP-SMAD Signaling Regulates Lineage Priming, but Is Dispensable for Self-Renewal in Mouse Embryonic Stem Cells, Stem Cell Reports (2016), http://dx.doi.org/10.1016/j.stemcr.2015.11.012

(P)SMAD1 by direct NANOG-SMAD1 interaction (Suzuki et al., 2006). However, the functional role of BMP-SMAD signaling in the metastable state of naive pluripotency has not been investigated.

Here, we report the derivation and characterization of transgenic mESCs that allow a real-time readout of SMAD-mediated BMP signaling activity. This transgenic *BRE:gfp* reporter mESC line expresses a well-characterized BMP responsive element (BRE) containing several PSMAD1/5 DNA-binding sites isolated from the *Id1* promoter to drive GFP expression (Korchynskyi and ten Dijke, 2002; Monteiro et al., 2008). Activation of the BMP-SMAD reporter transgene was heterogeneous in serum mESCs (±50% GFP + cells) and 2i mESCs (±4% GFP + cells). By genetic abrogation of the core BMP pathway components SMAD1 and SMAD5, we demonstrated that BMP-SMAD signaling is dispensable for the maintenance and self-renewal of mESCs both in serum and 2i states, but that it regulates the levels of DNA methylation (via *Dnmt3a/b* and *Tet1/2*) and hence lineage priming in pluripotent mESCs.

RESULTS

BMP-SMAD Signaling Is Activated during the Acquisition of Pluripotency

BMP signaling plays key roles in patterning of post-implantation mouse embryos (Kishigami and Mishina, 2005; Tam and Loebel, 2007). However, a role during pre-implantation development has been less evident because genetic ablation of single members of the BMP-SMAD pathway showed no evidence of a phenotype during the pre-implantation period (Goumans and Mummery, 2000; Graham et al., 2014; Reyes de Mochel et al., 2015; Zhao, 2003). We investigated whether the BMP-SMAD signaling pathway was active in pre-implantation embryos by examining *BRE:gfp* blastocysts at E3.5. We were unable to detect GFP at this stage (data not shown). As the BMP-SMAD pathway has been shown to play dual roles in self-renewal and differentiation of mESCs (Li and Chen, 2013), we monitored GFP during the derivation of mESCs from *BRE:gfp* blastocysts into the naive state (serum) and the ground state (2i). One day after plating (D1), GFP was still undetectable in blastocysts in either culture condition (Figure 1A); however, by D4, GFP+ cells were evident within the ICM-like cells of *BRE:gfp* blastocyst outgrowths in both serum and 2i (Figure 1A). This suggested that the BMP-SMAD pathway was activated during the acquisition of pluripotency in vitro.

BMP-SMAD Signaling Activation in Serum and 2i mESCs

Once *BRE:gfp* mESCs lines had been established (Figures 1A and 1B) and karyotyped (Figure S1A), a striking difference was observed between the two conditions: serum *BRE:gfp* mESCs exhibited an heterogeneous pattern of GFP expression with about 50% of the cells being GFP+, whereas in 2i *BRE:gfp* mESCs less than 4% of cells were GFP+ (Figure 1B). In serum *BRE:gfp* mESCs, the GFP+ cells produced ID1 (Figure 1C), confirming that GFP expression corresponded to the activation of BMP-SMADs. The promoter of *Id1* contains the PSMAD1/5 DNA-binding sites that were used to generate the *BRE:gfp* transgene (Figure S1B). Most 2i *BRE:gfp* mESCs showed no GFP and consequently no/low ID1 (Figure 1C). POU5F1 and NANOG were detected in both serum and 2i *BRE:gfp* mESCs. Quantification of NANOG suggested that it was more homogeneously expressed in GFP− cells per colony (Figure 1D) and this difference was statistically significant ($n = 16$; $p < 0.05$).

To measure BMP-SMAD signaling activation, we investigated the levels of PSMAD1/5/8, which were low in 2i medium in serum mESCs and high in 2i after 1 hr of stimulation with 25 ng/ml BMP4; in agreement, faint GFP was observed in 2i compared with serum *BRE:gfp* mESCs (Figure 1E). In addition, we examined the number of GFP+ cells present in 2i and showed that this increased in response to BMP4 but not to Activin A (which activates the NODAL pathway) (Figure 1F), and that *BRE:gfp* mESCs could be interconverted to adopt the GFP pattern associated with each culture medium within four cell passages (Figure 1G).

In Serum, GFP+ *BRE:gfp* mESCs Correlated with Low Levels of *Dnmt3b* and Lower DNA Methylation

To further understand the role of BMP-SMAD signaling activation in pluripotency, fluorescence-activated cell sorting (FACS) sorted subpopulations of serum (GFP++, GFP+, GFP−) and 2i (GFP+, GFP−) *BRE:gfp* mESCs (Figures 2A and S2A) were analyzed by qPCR (Figures 2B and S2B). In serum, the sorted GFP++ mESCs ($N = 3$) exhibited lower levels of *Dnmt3a/b*, in particular *Dnmt3b*, and higher levels of *Tet1/2*, but similar high transcriptional levels of pluripotency genes (Figure 2B). A direct comparison between 2i and serum is provided in Figure S2B. Comparing whole transcriptome RNA sequencing (RNAseq) data of three independent serum GFP++ and GFP− mESC samples, we confirmed that *Dnmt3b* as well as *Tet1/2* were among the few statistically significant differentially regulated genes observed ($n = 315$; $p < 0.05$), mostly protein-coding genes (Figures 2C, S2C, and S2D; Table S1). Next, using available single-cell RNAseq data (Sasagawa et al., 2013), we performed a hierarchical clustering of 38 individual cells from naive mESCs based on the expression of 30 selected genes. Interestingly, the cluster with the lowest transcriptional levels of *Dnmt3b* and high levels of *Tet1* (Group 1) did not correlate with the cell clusters showing high transcriptional levels of *Id1*/*Bmp4* (group 2/3) (Figure S2E). This is in agreement with our qPCR (Figure 2B) and RNAseq results (*Id1* is not

Please cite this article in press as: Gomes Fernandes et al., BMP-SMAD Signaling Regulates Lineage Priming, but Is Dispensable for Self-Renewal in Mouse Embryonic Stem Cells, Stem Cell Reports (2016), http://dx.doi.org/10.1016/j.stemcr.2015.11.012

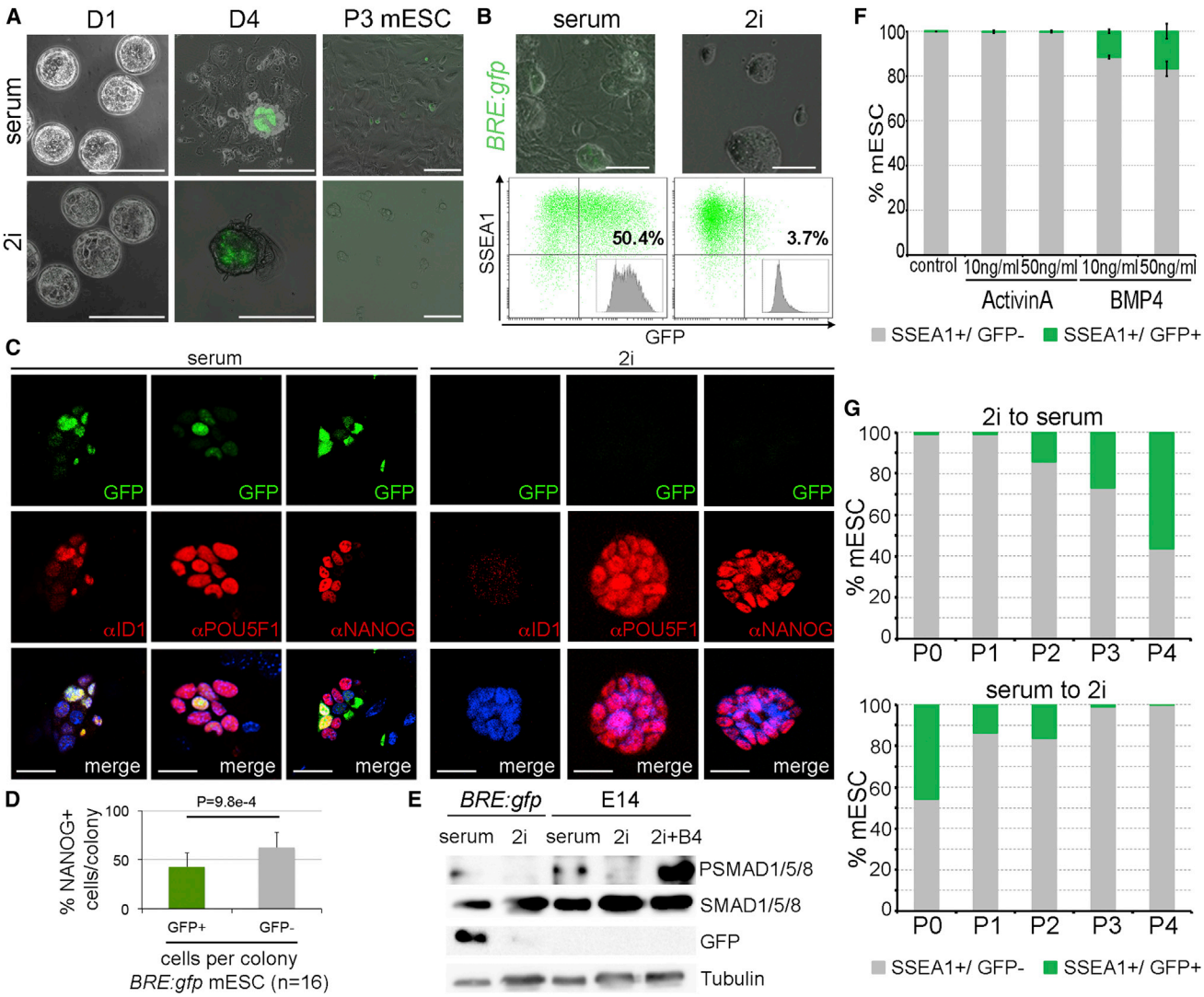


Figure 1. BMP-SMAD Signaling Activation in Serum and 2i Culture Conditions

(A) Derivation of *BRE:gfp* mESCs in serum and 2i conditions. D1, 1 day after blastocyst collection; D4, D1 plus 3 days after blastocyst plating, P3 mESCs, passage 3 of the derived mESCs. Scale bars represent 100 μm.

(B) Established serum and 2i *BRE:gfp* mESCs and their respective GFP expression profiles by FACS analysis. Scale bars represent 100 μm.

(C) Immunofluorescence of serum and 2i *BRE:gfp* mESCs for ID1, POU5F1, and NANOG. Scale bars represent 20 μm.

(D) Percentage (%) of NANOG-positive cells in the GFP+ and GFP− cells per colony *BRE:gfp* mESCs.

(E) Western blotting for PSMAD1/5/8, SMAD1/5/8, GFP and Tubulin in serum and 2i *BRE:gfp* and E14 mESCs as well as 2i E14 stimulated 1 hr with 25 ng/ml BMP4.

(F) Percentage (%) of GFP+ and GFP− cells in 2i *BRE:gfp* mESCs after 1 hr treatment with Activin A or BMP4. Bars represent the mean ± SD (N = 3).

(G) Percentage (%) of GFP+ and GFP− cells in 2i *BRE:gfp* mESCs switched to serum and serum *BRE:gfp* mESCs switched to 2i and cultured for four consecutive passages (P1–P4). See also Figure S1.

differentially expressed) (Figure 2C; Table S1) and suggests a clear discrepancy between the cells expressing ID1 protein (and GFP protein) and *Id1* transcript. This discrepancy in the co-expression of proteins and transcripts is a well-known confounding but intrinsic property of cells, including mESCs (Torres-Padilla and Chambers, 2014).

We performed reduced-representation bisulfite sequencing (RRBS) of GFP++ and GFP− *BRE:gfp* mESCs and observed that DNA methylation levels were in general lower in mESCs with activation of the BMP-SMAD reporter transgene than in mESCs without reporter activity, as illustrated by the significant shifts toward lower DNA methylation at

Please cite this article in press as: Gomes Fernandes et al., BMP-SMAD Signaling Regulates Lineage Priming, but Is Dispensable for Self-Renewal in Mouse Embryonic Stem Cells, Stem Cell Reports (2016), http://dx.doi.org/10.1016/j.stemcr.2015.11.012

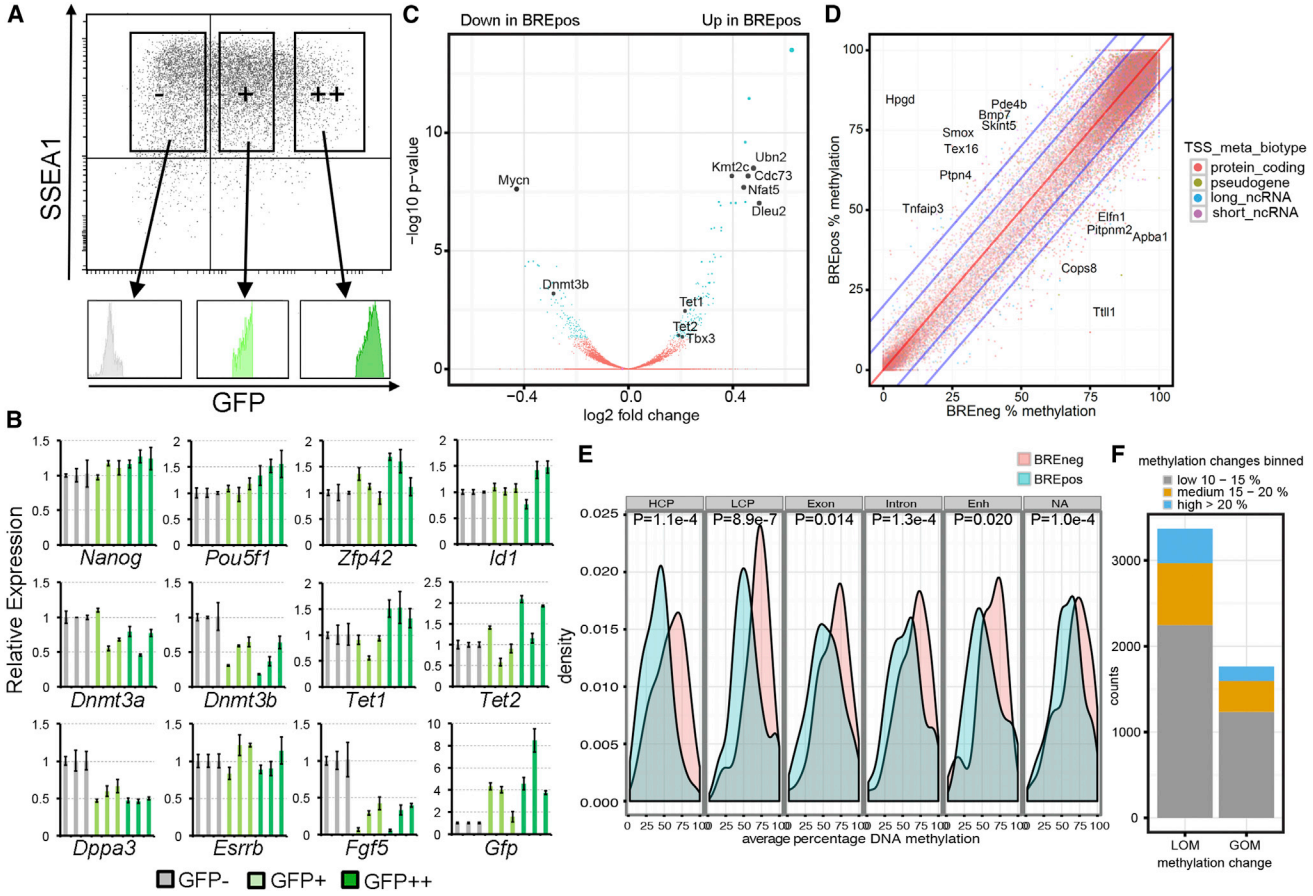


Figure 2. Transcriptome and Methylome in Subsets of Serum *BRE:gfp* mESCs

(A) Gatings used to FACS sort three subpopulations (GFP−, GFP+, GFP++) of serum *BRE:gfp* mESCs and the profile of the individual cell groups.

(B) Relative expression of several genes in the three subpopulations (GFP−, GFP+, GFP++) of serum *BRE:gfp* mESCs compared with the GFP− cells. Each bar represents the mean ± SD of technical triplicates and the three bars of the same color represent independent experiments ($n = 9$, $N = 3$).

(C) Volcano plot showing −log10 p values versus log2 fold transcriptional changes between GFP++ and GFP− fractions of serum *BRE:gfp* mESCs. Differentially expressed genes (DEGs) with $p < 0.05$ are blue, and genes with $p > 0.05$ are red; some highlighted DEGs are black.

(D) Scatterplot depicting a comparison of the percentage of DNA methylation in each 600-bp tile (dot) between GFP++ and GFP− fractions of serum *BRE:gfp* mESCs. Each tile was classified into a biotype category according to the nearest TSS. The red line represents no difference; the inner and outer blue lines represent borders for 10% and 20% change in methylation levels, respectively.

(E) Distribution of DNA methylation at specific genomic regions in GFP++ (in blue) and GFP− fractions of serum *BRE:gfp* (in red) mESCs. p Values were calculated with the two-sample Kolmogorov-Smirnov test. HCP, high CpG-content promoters; LCP, low CpG-content promoters; Enh, enhancers; NA, no annotation.

(F) Number of (600-bp tile) counts showing loss of methylation (LOM) or gain of methylation (GOM) in GFP++ compared with GFP− serum *BRE:gfp* mESC. See also Figure S2 and Tables S1 and S2.

all genomic regions in GFP++ cells (Figures 2D–2F; Table S2). This is in agreement with the reduced levels of *Dnmt3b* expression in GFP++ cells.

BMP-SMAD Signaling Is Dispensable for Self-Renewal of mESCs

To clarify the role of BMP-SMAD signaling in the maintenance of the naive and ground state, we derived *Smad1* and *Smad5* double-knockout ($S1^{-/-}S5^{-/-}$) mESC lines in 2i from double homozygous floxed *Smad1;Smad5* mESC lines ($S1^{fl/fl}S5^{fl/fl}$) (Tremblay et al., 2001; Umans et al., 2003) that were hemizygous for the R26R Cre-reporter transgene (Soriano, 1999) using Cre recombinase (Figures S3A and S3B). We derived the $S1^{-/-}S5^{-/-}$ mESC in 2i because BMP-SMAD signaling activation was less prominent in 2i and therefore the chance of deriving pluripotent

Please cite this article in press as: Gomes Fernandes et al., BMP-SMAD Signaling Regulates Lineage Priming, but Is Dispensable for Self-Renewal in Mouse Embryonic Stem Cells, Stem Cell Reports (2016), http://dx.doi.org/10.1016/j.stemcr.2015.11.012

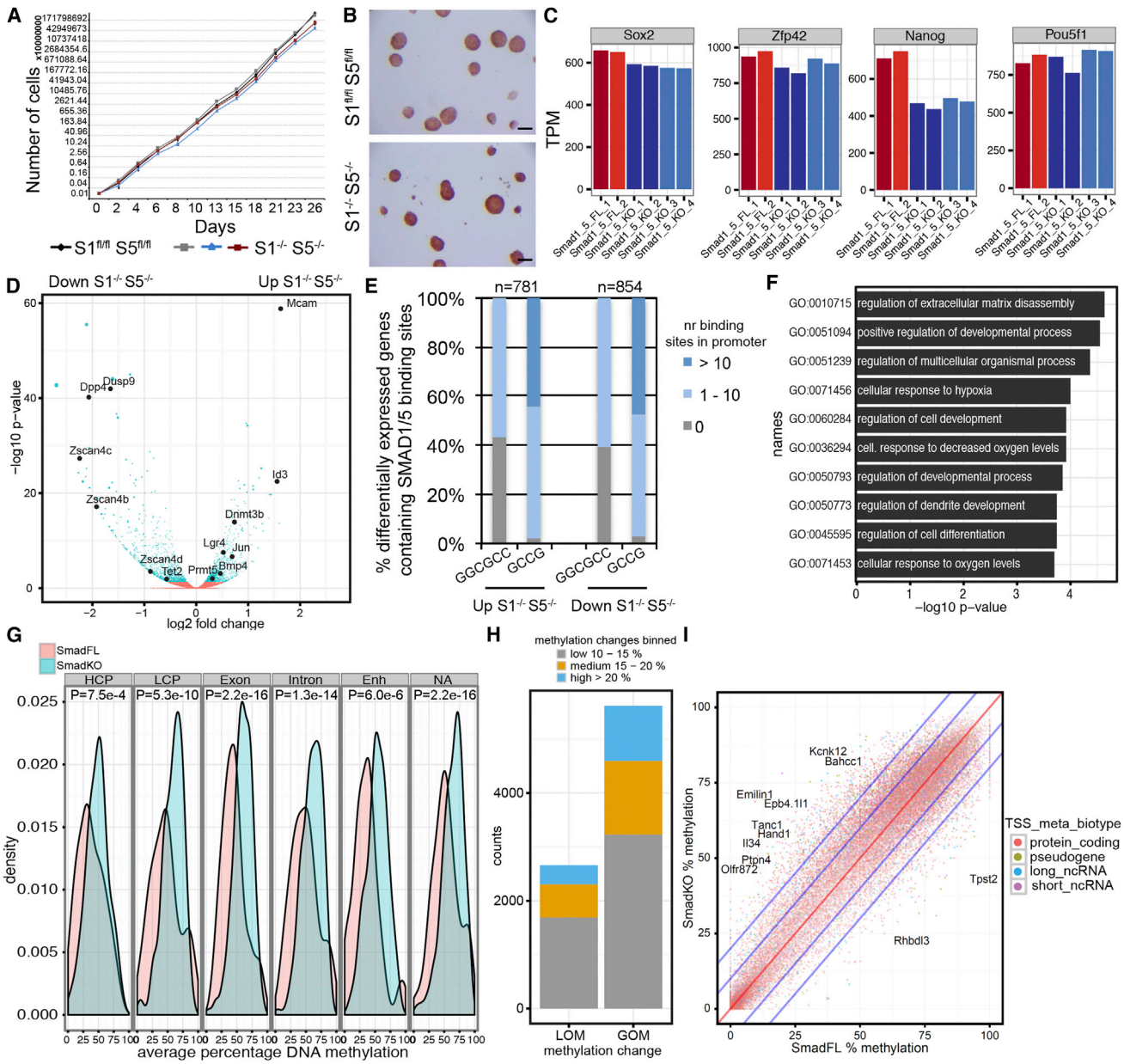


Figure 3. Transcriptome and Methylome in *S1*$^{-/-}$*S5*$^{-/-}$ versus *S1*$^{fl/fl}$*S5*$^{fl/fl}$ mESCs

(A) Growth of *S1*$^{fl/fl}$*S5*$^{fl/fl}$ mESCs and three independent *S1*$^{-/-}$*S5*$^{-/-}$ mESCs lines in 2i during 26 days. Means ± SD are depicted.

(B) Alkaline phosphatase activity in 2i *S1*$^{fl/fl}$*S5*$^{fl/fl}$ and *S1*$^{-/-}$*S5*$^{-/-}$ mESC. Scale bars represent 100 μm.

(C) Expression of *Sox2*, *Zfp42*, *Nanog*, and *Pou5f1* in transcripts per million (TPM) in 2i *S1*$^{fl/fl}$*S5*$^{fl/fl}$ (FL) and *S1*$^{-/-}$*S5*$^{-/-}$ (KO) mESC.

(D) Volcano plot showing −log10 p values versus log2 fold transcriptional changes between *S1*$^{fl/fl}$*S5*$^{fl/fl}$ and *S1*$^{-/-}$*S5*$^{-/-}$ mESCs in 2i. DEGs with $p < 0.05$ are blue, and genes with $p > 0.05$ are red; some highlighted DEGs are black.

(E) Percentage of DEGs ($p < 0.01$) (n = 781 upregulated in 2i *S1*$^{-/-}$*S5*$^{-/-}$; n = 854 downregulated in 2i *S1*$^{-/-}$*S5*$^{-/-}$) showing putative SMAD1/5 binding sites (GGCGCC/GCCG) in the promoter region.

(F) Top ten GO terms associated with biological processes ($p < 0.05$) in DEGs in 2i *S1*$^{-/-}$*S5*$^{-/-}$ mESCs.

(G) Distribution of DNA methylation levels at specific genomic regions in 2i *S1*$^{fl/fl}$*S5*$^{fl/fl}$ (in red) and *S1*$^{-/-}$*S5*$^{-/-}$ mESCs (in blue). p Values were calculated with two-sample Kolmogorov-Smirnov test. HCP, high CpG-content promoters; LCP, low CpG-content promoters; Enh, enhancers; NA, no annotation.

(H) Number of (600-bp tile) counts showing LOM or GOM in 2i *S1*$^{fl/fl}$*S5*$^{fl/fl}$ compared with *S1*$^{-/-}$*S5*$^{-/-}$ mESCs.

(legend continued on next page)

Please cite this article in press as: Gomes Fernandes et al., BMP-SMAD Signaling Regulates Lineage Priming, but Is Dispensable for Self-Renewal in Mouse Embryonic Stem Cells, Stem Cell Reports (2016), http://dx.doi.org/10.1016/j.stemcr.2015.11.012

$S1^{-/-}S5^{-/-}$ mESCs was higher. The pluripotency of the $S1^{-/-}S5^{-/-}$ mESCs was confirmed by showing its contribution to the three germ layers in $S1^{-/-}S5^{-/-}$ <> wild-type chimeric embryos (Figure S3C), as well as in teratoma formation assays (Figure S3D) in independent lines with a normal karyotype (Figure S4A). Moreover, we showed that *Smad8* was not upregulated in response to the deletion of *Smad1* and *Smad5*, and that *Id1* and *Id2* were upregulated after stimulation with BMP4 only in the $S1^{fl/fl}S5^{fl/fl}$ parental line, as expected (Figure S4B). The 2i $S1^{-/-}S5^{-/-}$ mESCs self-renewed at the same rate as the parental $S1^{fl/fl}S5^{fl/fl}$ mESCs (Figure 3A) and showed comparable alkaline phosphatase activity (Figure 3B). Unexpectedly, when $S1^{-/-}S5^{-/-}$ mESCs were switched from 2i to serum, after an initial period of adaptation the cells continued to self-renew at similar rates as the parental $S1^{fl/fl}S5^{fl/fl}$ mESCs (Figure S4C) instead of differentiating. In general, the expression level of pluripotency genes remained high in the parental $S1^{fl/fl}S5^{fl/fl}$ and $S1^{-/-}S5^{-/-}$ mESCs in 2i (Figure 3C) and serum (Figure S4D). Our results demonstrated that BMP-SMAD signaling is dispensable for self-renewal of mESCs.

$S1^{-/-}S5^{-/-}$ mESCs Have High Levels of *Dnmt3b* and High Levels of DNA Methylation

Next, we investigated the SMAD1/5-responsive genes using RNAseq (Figure 3D) and found that most differentially expressed genes (DEGs) between $S1^{-/-}S5^{-/-}$ and $S1^{fl/fl}S5^{fl/fl}$ mESCs were protein-coding genes (Figure S4E). Interestingly, about half of the DEGs (including protein-coding, pseudogenes, and long non-coding RNAs) were upregulated ($n = 781$; $p < 0.01$) and half of the genes were downregulated ($n = 854$; $p < 0.01$) in $S1^{-/-}S5^{-/-}$ mESCs (Figure 3E; Table S1).

To investigate whether the observed expression changes were consistent with direct transcriptional regulation, we integrated our RNAseq dataset with a list of direct SMAD1/5 targets ($n = 562$) identified by ChIP (Fei et al., 2010). Using gene set enrichment analysis, we found a significant enrichment of SMAD1/5 targets in genes that were downregulated in $S1^{-/-}S5^{-/-}$ mESCs ($p < 1 \times 10^{-4}$) (Figure S4F).

Moreover, the great majority of the DEGs contained the sequence motifs GCCG and/or GGCGCC, well-characterized SMAD1/5 binding sites (Korchynskyi and ten Dijke, 2002), in their promoters, defined as ±2 kb from the transcriptional start site (TSS) (Figure 3E; Table S3). By contrast, genome-wide occurrence of GGCGCC and GCCG motifs at such promoters (including protein-coding, pseudogenes, and long non-coding RNAs) was not, or much less, enriched (Figure S4G), and significantly different from the enrichment observed at DEGs ($p < 2.2 \times 10^{-16}$). As an example, *Dnmt3b* was significantly upregulated in $S1^{-/-}S5^{-/-}$ mESCs and contained 21x GCCG and 5x GGCGCC in the promoter region, suggesting direct (co-)regulation by DNA-binding BMP-SMADs. The DEGs were significantly enriched for gene ontology (GO) categories such as "regulation of developmental process," "regulation of cell development," and "regulation of cell differentiation" (Figure 3F), compatible with BMP-SMAD signaling not being involved in self-renewal of mESC, but rather predisposing mESCs to differentiate. The downregulation of *Dnmt3b* and enrichment in developmental genes in $S1^{-/-}S5^{-/-}$ mESCs, led us to investigate the levels of DNA methylation by RRBS on several independent $S1^{fl/fl}S5^{fl/fl}$ and $S1^{-/-}S5^{-/-}$ mESC lines (Table S2). $S1^{-/-}S5^{-/-}$ mESCs displayed a significant shift toward higher levels of DNA methylation at all genomic regions analyzed when compared with $S1^{fl/fl}S5^{fl/fl}$ mESCs (Figures 3G–3I), suggesting that the enrichment in developmental genes is caused by the higher levels of DNA methylation.

mESCs Differentiated More Efficiently to Mesendoderm or Neurectoderm in the Absence of BMP-SMAD Signaling

Finally, we examined the differentiation capacity of $S1^{-/-}S5^{-/-}$ mESCs in both serum and 2i and found that they formed endoderm (*Sox17*), mesoderm (*T*), and ectoderm (*Pax6* and *Sox1*) more efficiently than the parental line (Figures 4A–4C) in the monolayer using differentiation protocols for either the mesendoderm (ME) or neuroectoderm (NE) lineages (Thomson et al., 2011). In addition, we investigated the capacity of the FACS-sorted subpopulations of serum *BRE:gfp* mESCs to differentiate to ME and NE and showed that GFP++ mESCs had lower levels of ME and NE early differentiation markers than GFP− mESCs (Figure 4D), demonstrating that GFP++ mESCs were less prone to differentiate. In agreement, GFP++ mESCs retained higher levels of pluripotency markers, at least after 4 days of differentiation to ME (Figure 4E). Our data showed that transient BMP-SMAD signaling activation tilted mESCs to a less differentiation-prone state, whereas in the absence of BMP-SMAD signaling the balance was shifted toward an increased predisposition to differentiate.

(I) Scatterplot depicting a comparison of the percentage of DNA methylation in each 600-bp tile (dot) between 2i $S1^{fl/fl}S5^{fl/fl}$ and $S1^{-/-}S5^{-/-}$ mESCs. Each tile was classified into a biotype category according to the nearest TSS. The red line represents no difference, and the inner and outer blue lines represent borders for 10% and 20% change in methylation levels, respectively. See also Figures S3, S4 and Tables S1, S2, and S3.

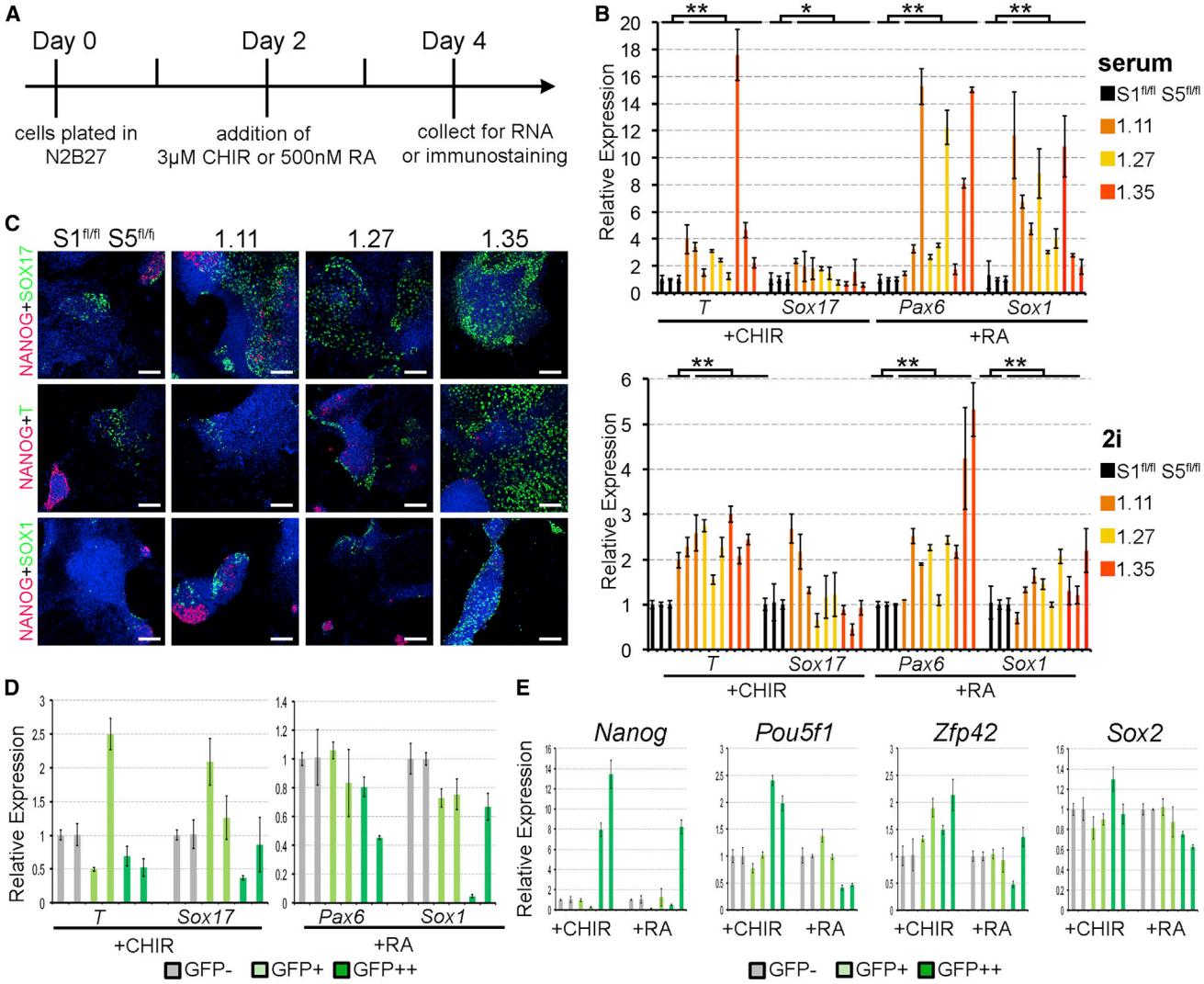


Figure 4. BMP-SMAD Signaling during mESC Differentiation to Mesendoderm and Neurectoderm

(A) Schematic representation of the protocol to differentiate mESCs to mesendoderm (3 μM CHIR) or neurectoderm (500 nM retinoic acid [RA]).

(B) Relative expression of early lineage markers in differentiated serum and 2i $S1^{-/-}S5^{-/-}$ and $S1^{fl/fl}S5^{fl/fl}$ mESCs.

(C) Immunofluorescence of differentiated serum $S1^{fl/fl}S5^{fl/fl}$ and $S1^{-/-}S5^{-/-}$ mESCs for NANOG, SOX17, T, and SOX1. Scale bars represent 100 μm.

(D) Relative expression of early lineage markers in differentiated subpopulations (GFP−, GFP+, GFP++) of serum *BRE:gfp* mESCs compared with GFP− cells.

(E) Relative expression of pluripotency genes in differentiated subpopulations (GFP−, GFP+, GFP++) of serum *BRE:gfp* mESCs compared with GFP− cells.

Each bar represents the mean ± SD of technical triplicates and bars of the same color represent independent experiments (n = 9, N = 3) in (B) and independent experiments (n = 6, N = =2) in (D) and (E). Statistical analysis was performed on technical triplicates of independent experiments (n = 9, N = 3), *$p \leq 0.05$, **$p \leq 0.01$.

DISCUSSION

A recent study reported the absence of *Bmp4* and *Id1* in (embryonic day) E3.5 ICMs and a high transient upregulation in E4.5 epiblasts, followed by downregulation of *Bmp4* and *Id3* expression during the next 6 days of the derivation of mESCs and their further maintenance in 2i (Boroviak et al., 2014). We now show this in real-time using *BRE:gfp* blastocysts to derive mESCs. Moreover, we demonstrated that BMP-SMAD signaling is not functionally implicated

in self-renewal, in agreement with studies that have mapped genome-wide the genes that are directly regulated by SMAD1/5 (Chen et al., 2008; Fei et al., 2010). They showed that the genes regulated by SMAD1/5 were involved in fate determination, rather than self-renewal. Here, we provide functional evidence that SMAD1/5 are not necessary for mESC self-renewal in either naive (serum) or ground (2i) state.

Specific levels of DNA methylation and associated enzymes have been associated with the different pluripotency states (ground, naive, primed) (Habibi et al., 2013; Hackett et al., 2013; Smallwood et al., 2014), as well as with different levels of GFP in *Nanog:gfp* naive mESCs (Ficz et al., 2013). This reflects faithfully the rapid loss of genomic DNA methylation that the embryo undergoes in vivo during pre-implantation development, and the gain of DNA methylation during the transition between ICM and epiblast (Smith et al., 2012). Therefore, it is perhaps not surprising that the machinery to regulate rapid switches in genomic DNA methylation is present in pluripotent stem cells derived from ICM and epiblast. A role for BMP-SMAD signaling in LIF-dependent conversion between EpiSCs and ESCs has been reported (Onishi et al., 2014), but the association with changes in DNA methylation between EpiSCs and ESCs remains to be investigated.

Finally, it has been suggested that the epigenetic variation observed in pluripotent cells is stochastic and results in a diversity of predispositions to acquire specific cell fates when the cells are triggered to differentiate (Lee et al., 2014). Our data provide evidence that the cellular diversity of both serum and 2i mESCs regarding DNA methylation and associated enzymes is not a stochastic process as previously thought, but is in fact regulated by cell-cell signaling interactions involving the BMP-SMAD signaling pathway.

EXPERIMENTAL PROCEDURES

mESCs Derivation and Culture

Derivation of *BRE:gfp* mESCs in 2i and serum and the conditional knockout mESCs for *Smad1* and *Smad5* ($S1^{fl/fl}S5^{fl/fl}$) in 2i, as well as the Cre-recombination of $S1^{fl/fl}S5^{fl/fl}$ mESCs, are described in the Supplemental Experimental Procedures. Genotyping of the *BRE:gfp* mESCs was performed as described (Monteiro et al., 2008). E14 mESCs were cultured in either 2i or serum. Stimulation (1 hr) with BMP4 (R&D Systems) or Activin A (R&D Systems) was followed by FACS analysis or western blotting (see Supplemental Experimental Procedures). Details about generation of chimeric embryos, the teratoma formation assay, RNAseq, and RRBS are provided in the Supplemental Experimental Procedures.

mESCs Differentiation and Proliferation

mESCs were differentiated to ME or NE as described (Thomson et al., 2011). Briefly, mESCs (10,000 cells/cm^2) were grown in N2B27 medium without supplements for 48 hr, after which either 3 μM CHIR99021 or 500 nM all-*trans* retinoic acid (RA) (Sigma-Aldrich) was added to the N2B27 medium for an additional 48 hr. Cells were then collected for immunofluorescence or qPCR (see Supplemental Experimental Procedures). For the proliferation assay, the total number of serum and 2i mESCs was monitored during each passage for 26 days of culture. Serum mESCs were pre-plated prior to counting.

Statistics

Quantification of NANOG-Positive Cells

Whole *BRE:gfp* mESC colonies (total n = 16) from three independent experiments (N = 3, 5–6 colonies per experiment) were manually counted three times and averaged. N refers to the number of independent experiments; n refers to total number or colonies counted. Statistical analysis was performed using the Student t-test (two-tailed, unequal variance), *$p \leq 0.05$.

qPCR

In qPCR, each bar represents the average of technical triplicates. N refers to the number of independent experiments; n refers to total replicates. Statistical analysis was performed using the Student t-test (two-tailed, unequal variance), *$p \leq 0.05$; **$p \leq 0.01$.

RNAseq Expression Data

To determine significantly DEGs between GFP++ and GFP− or $S1^{-/-}S5^{-/-}$ and $S1^{fl/fl}S5^{fl/fl}$ mESCs, we applied a cut-off of 0.01 and/or 0.05 on the p values adjusted for multiple testing hypothesis. N refers to the number of independent experiments; n refers to the number of genes.

RNAseq GO

Enrichment analysis for GO terms was done with the R package topGO based on DEGs ($p < 0.05$) and utilizing Fisher's exact test.

RNAseq Motif Sequence Analysis

One-sided Fisher's exact was used to determine significant over-representation of the analyzed motifs in promoter regions of DEGs relative to the genome-wide promoter regions. n refers to the number of genes.

SMAD1/5 ChIP-on-chip Data

To calculate the enrichment of SMAD1/5 targets identified p values were calculated by permuting genes. n refers to the number of genes.

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.

ACCESSION NUMBERS

The GEO accession number for both the transcriptomics and methylomics data reported in this paper is GEO: GSE71556.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2015.11.012.

AUTHOR CONTRIBUTIONS

M.G.F. designed and performed the experiments, analyzed the data, and wrote the manuscript. R.D., M.S.R., S.S., A.dM.B., R.P.D., R.R., K.S., E.M., L.U., D.S., V.A.E., D.E., W.V.C., and D.H. performed the experiments and/or analyzed the data. A.Z., C.M., and S.C.dS.L. designed the experiments, analyzed the data, and wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We acknowledge S. Kobayakawa and M. Bialecka for discussions, S. Mendes and C. Visseren for technical support, M. Bouma for technical help with teratomas, and Z. Zhang for generation of mouse chimeras using equipment provided by InfraMouse (KU Leuven-VIB) through a Hercules type 3 project (ZW09-03). This manuscript is dedicated to Cheryl Visseren, one of our most talented students, who carried out the initial experiments but tragically passed away on July 4, 2014. This work was supported by the Interuniversity Attraction Poles-Phase VII [IUAP/PAI P7/14] (to S.C.dL., C.M., A.Z., and D.H.) and individual grants by the Fundação para a Ciência e Tecnologia (FCT) [SFRH/BD/78689/2011 and SFRH/BD/94387/2013] to M.M.G.F. and A.dM.B., respectively, the Netherlands organization of Scientific Research (NWO) [ASPASIA 015.007.037] to S.M.C.dS.L., and the Bontius Stichting [PANCREAS] to M.S.R. The bio-informatics work in the D.H. team was also supported by Fund for Scientific Research-Flanders (FWO-V) GA.0941.11 and G.0782.14.

Received: August 24, 2014
Revised: November 16, 2015
Accepted: November 18, 2015
Published: December 17, 2015

REFERENCES

Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nat. Cell Biol. *16*, 516–528.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

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 Res. *20*, 36–44.

Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.L., Walter, J., et al. (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. Cell Stem Cell *13*, 351–359.

Galvin-Burgess, K.E., Travis, E.D., Pierson, K.E., and Vivian, J.L. (2013). TGF-β-superfamily signaling regulates embryonic stem cell heterogeneity: self-renewal as a dynamic and regulated equilibrium. Stem Cells *31*, 48–58.

Goumans, M.J., and Mummery, C. (2000). Functional analysis of the TGFbeta receptor/Smad pathway through gene ablation in mice. Int. J. Dev. Biol. *44*, 253–265.

Graf, T., and Stadtfeld, M. (2008). Heterogeneity of embryonic and adult stem cells. Cell Stem Cell *3*, 480–483.

Graham, S.J., Wicher, K.B., Jedrusik, A., Guo, G., Herath, W., Robson, P., and Zernicka-Goetz, M. (2014). BMP signalling regulates the pre-implantation development of extra-embryonic cell lineages in the mouse embryo. Nat. Commun. *5*, 5667.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360–369.

Hackett, J.A., Dietmann, S., Murakami, K., Down, T.A., Leitch, H.G., and Surani, M.A. (2013). Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency. Stem Cell Reports *1*, 518–531.

Hanna, J., Markoulaki, S., Mitalipova, M., Cheng, A.W., Cassady, J.P., Staerk, J., Carey, B.W., Lengner, C.J., Foreman, R., Love, J., et al. (2009). Metastable pluripotent states in NOD-mouse-derived ESCs. Cell Stem Cell *4*, 513–524.

Hayashi, K., Lopes, S.M.C.d.S., Tang, F., and Surani, M.A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell Stem Cell *3*, 391–401.

Kashyap, V., Rezende, N.C., Scotland, K.B., Shaffer, S.M., Persson, J.L., Gudas, L.J., and Mongan, N.P. (2009). Regulation of stem cell pluripotency and differentiation involves a mutual regulatory circuit of the Nanog, OCT4, and SOX2 pluripotency transcription factors with polycomb repressive complexes and stem cell microRNAs. Stem Cells Dev. *18*, 1093–1108.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049–1061.

Kishigami, S., and Mishina, Y. (2005). BMP signaling and early embryonic patterning. Cytokine Growth Factor Rev. *16*, 265–278.

Korchynskyi, O., and ten Dijke, P. (2002). Identification and functional characterization of distinct critically important bone morphogenetic protein-specific response elements in the Id1 promoter. J. Biol. Chem. *277*, 4883–4891.

Lee, H.J., Hore, T.A., and Reik, W. (2014). Reprogramming the methylome: erasing memory and creating diversity. Cell Stem Cell *14*, 710–719.

Li, Z., and Chen, Y.-G. (2013). Functions of BMP signaling in embryonic stem cell fate determination. Exp. Cell Res. *319*, 113–119.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Francis Stewart, A., Smith, A., et al. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590–604.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521–533.

Monteiro, R.M., de Sousa Lopes, S.M.C., Bialecka, M., de Boer, S., Zwijsen, A., and Mummery, C.L. (2008). Real time monitoring of BMP Smads transcriptional activity during mouse development. Genesis *46*, 335–346.

Navarro, P., Festuccia, N., Colby, D., Gagliardi, A., Mullin, N.P., Zhang, W., Karwacki-Neisius, V., Osorno, R., Kelly, D., Robertson, M., et al. (2012). OCT4/SOX2-independent Nanog autorepression modulates heterogeneous Nanog gene expression in mouse ES cells. EMBO J. *31*, 4547–4562.

Nichols, J., Silva, J., Roode, M., and Smith, A. (2009). Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development *136*, 3215–3222.

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev. *12*, 2048–2060.

Onishi, K., Tonge, P.D., Nagy, A., and Zandstra, P.W. (2014). Local BMP-SMAD1 signaling increases LIF receptor-dependent STAT3 responsiveness and primed-to-naive mouse pluripotent stem cell conversion frequency. Stem Cell Reports *3*, 156–168.

Reyes de Mochel, N.S., Luong, M., Chiang, M., Javier, A.L., Luu, E., Toshihiko, F., MacGregor, G.R., Cinquin, O., and Cho, K.W. (2015). BMP signaling is required for cell cleavage in preimplantation-mouse embryos. Dev. Biol. *397*, 45–55.

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 Biol. *14*, R31.

Sasai, M., Kawabata, Y., Makishi, K., Itoh, K., and Terada, T.P. (2013). Time scales in epigenetic dynamics and phenotypic heterogeneity of embryonic stem cells. PLoS Comput. Biol. *9*, e1003380.

Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat. Methods *11*, 817–820.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature *484*, 339–344.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat. Genet. *21*, 70–71.

Suzuki, A., Raya, A., Kawakami, Y., Morita, M., Matsui, T., Nakashima, K., Gage, F.H., Rodriguez-Esteban, C., and Izpisua Belmonte, J.C. (2006). Nanog binds to Smad1 and blocks bone morphogenetic protein-induced differentiation of embryonic stem cells. Proc. Natl. Acad. Sci. USA *103*, 10294–10299.

Tam, P.P.L., and Loebel, D.A.F. (2007). Gene function in mouse embryogenesis: get set for gastrulation. Nat. Rev. Genet. *8*, 368–381.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196–199.

Thomson, M., Liu, Siyuan, J., Zou, L.-N., Smith, Z., Meissner, A., and Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell *145*, 875–889.

Torres-Padilla, M.E., and Chambers, I. (2014). Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. Development *141*, 2173–2181.

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.

Trott, J., and Martinez Arias, A. (2013). Single cell lineage analysis of mouse embryonic stem cells at the exit from pluripotency. Biol. Open *2*, 1049–1056.

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.

Ying, Q.-L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281–292.

Zhao, G.-Q. (2003). Consequences of knocking out BMP signaling in the mouse. Genesis *35*, 43–56.

STEMCR, 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

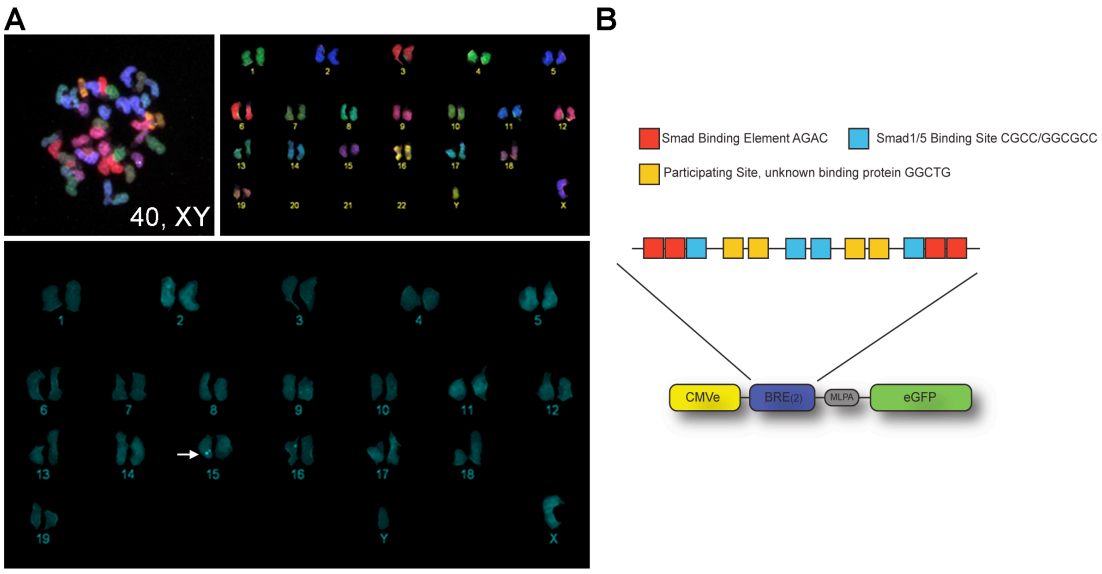


Figure S1. The *BRE:gfp* construct, related to Figure 1.
(A) Representative karyogram of a XY *BRE:gfp* mESC line. The arrow identifies the subtelomeric region of the Chromosome 15, where the *BRE:gfp* construct was mapped by DNA-FISH.
(B) Schematic representation of the *BRE:gfp* construct. Multiple 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*.

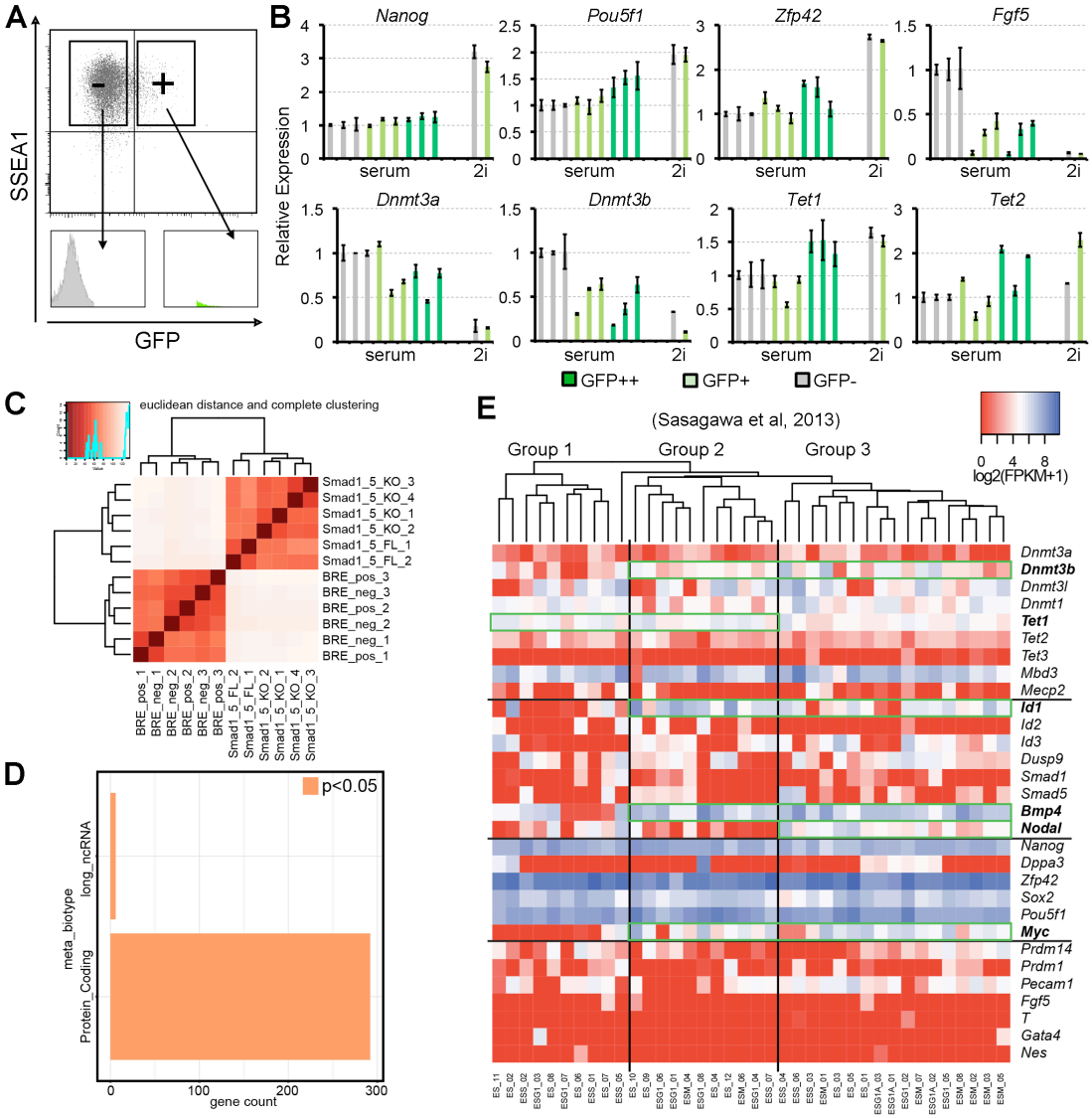


Figure S2. Characterization and comparison of "serum" and "2i" mESCs, related to Figure 2.
(A) Gatings used to FACS sort two subpopulations (GFP- and GFP+) of "2i" *BRE:gfp* mESCs and the profile of the individual cell groups.
(B) Relative expression of several genes in the subpopulations (GFP-, GFP+, 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.

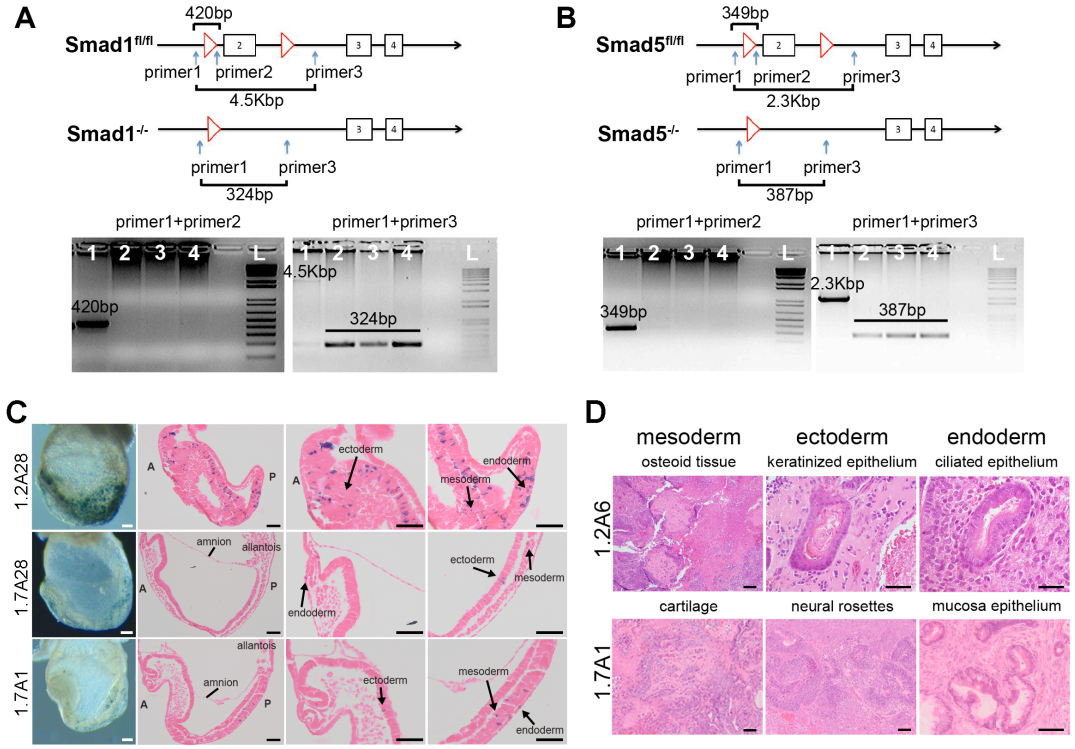
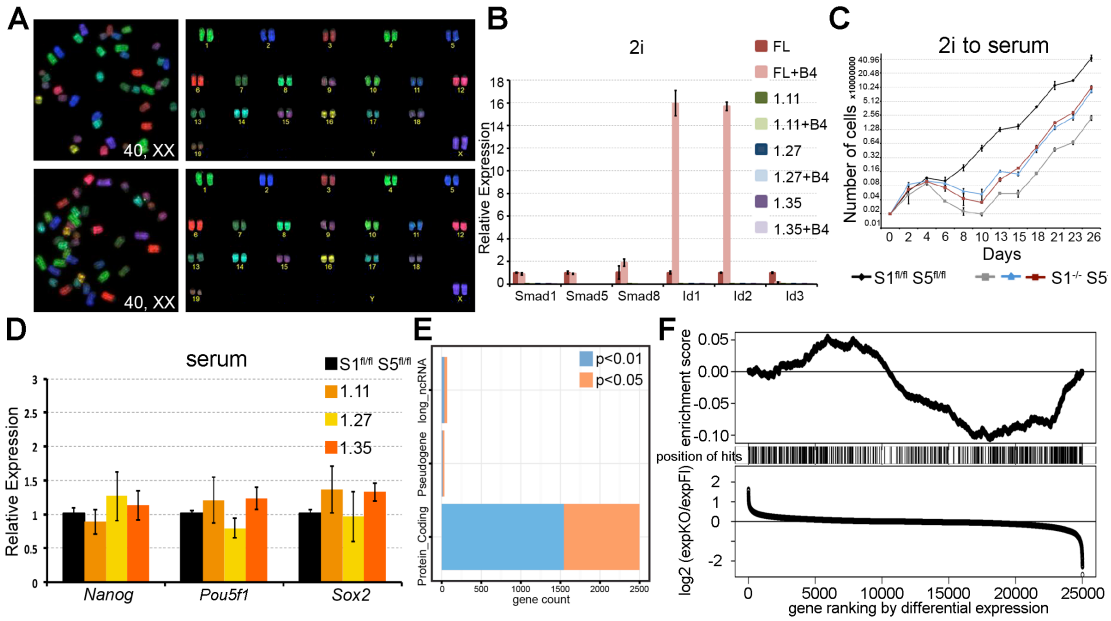


Figure S3. Derivation and pluripotency state of "2i" *S1*$^{-/-}$*S5*$^{-/-}$ mESCs, related to Figure 3.
(A, B) Scheme representing the location of the primers and the size of the respective genotyping PCR bands before and after excision of the exon 2 of *Smad1* and respective representative PCR gel (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.