Supplementary Material for:

Master Transcription Factors Determine Cell-Type-Specific Responses to TGF-β Signaling

Alan C. Mullen, David A. Orlando, Jamie J. Newman, Jakob Lovén, Roshan M. Kumar, Steve Bilodeau, Jessica Reddy, Matthew G. Guenther, Rodney DeKoter, and Richard A. Young*

*To whom correspondence should be addressed: young@wi.mit.edu

Figure S1. Sites bound by Smad3 are enriched for the motifs of cell-type specific master transcription factors. Related to Figure 1

(A) DNA motif discovery was performed using bound sites for Smad3 (top) and master transcription factors (bottom) in human ES cells (hESC), murine ES cells (mESC), myotubes, and pro-B cells. The top five motifs identified using MEME are displayed in order of increasing E-value (Bailey and Elkan, 1994).

(B) E-value for each motif in (A).

(C) TGF-β signaling pathway is active in standard mES cell culture conditions. Cytoplasmic (Cyt) and nuclear (Nuc) fractions were prepared from mES cells grown in standard mES culture conditions. The fractions were probed with pSmad3 (top), TBP (center) and GAPDH (bottom).

Figure S2. Oct4, Sox2 and Nanog are highly associated with Smad3 sites. Antibodies against Smad3 and Smad2/3 show co-occupancy with Oct4 in mES cells. Related to Figure 2

(A) Oct4 is the transcription factor most highly associated with Smad3. The percent of total Smad3 sites (y-axis) occupied by each transcription factor (x-axis) was determined for each factor for which there was ChIP-seq data available. Dark blue shading represents the fraction of sites that are co-occupied by Smad3 and Oct4. Gray shading represents the fraction of sites co-occupied by Smad3 and the indicated factor in the absence of Oct4 binding (Oct4-Independent). Less than 10% of sites are occupied by any single transcription factor in the absence of Oct4.

(B) Smad3 and Smad 2/3 co-occupy DNA sites with the mES cell master transcription factors. Gene tracks represent binding of Oct4, Sox2, Nanog (Marson et al., 2008), Smad3, and Smad2/3 at *Pou5f1*, the gene encoding Oct4 (left) and at *Lefty1* (right). Smad2/3 ChIP-seq was performed on cells treated with Activin for 1 hr to increase the TGF-β signal. The track floor is set at 2 counts.

(C) Smad3 and Smad2/3 co-occupy the genome with Oct4. Binding plots show the distribution of Smad2/3- (left) and Smad3- (right) bound sites relative to sites bound by Oct4. For each Oct4-bound site (y-axis) the presence of Smad3-bound sites (left) and Smad2/3-bound sites (right) are displayed within a 5kb window centered on the Oct4-bound site. Intensity at 0 indicates that bound sites overlap.

(D) The Oct4 motif was the most enriched, identifiable motif enriched in sites bound by Smad2/3. Motif discovery was performed using Smad2/3-bound regions identified by ChIP-seq.

Figure S3. Smad3 co-occupies sites with cell-type-specific master transcription factors. Reduction of Myod1 or PU.1/Spib results in reduced Smad3 occupancy at sites normally co-occupied with Myod1 or PU.1. Master transcription factors occupy cell-type specific sites of relative nucleosome depletion. Related to Figure 4

(A) Over 70% of Smad3 sites are also bound by cell-type-specific master transcription factors. The percentage of Smad3-bound regions (y-axis) that are co-occupied by Oct4, Myod1 and PU.1 are indicated for ES cells, myotubes, and pro-B cells. Percentages were calculated using the 1000 highest-confidence Smad3-bound regions for each cell-type (Table S1). P-values are shown above each bar (Table S1).

(B) Smad3 co-occupies cell-type specific sites with master transcription factors. The percentage of all Smad3 sites (totals in parenthesis) in each cell type that occupy sites also occupied by master transcription factors are shown. Smad3 occupies sites bound by Oct4 in mES cells (bottom left), Myod1 in myotubes (bottom center) and PU.1 in pro-B cells.

(C) Knockdown of Myod1 results in reduced Smad3 binding at sites normally co-occupied by Smad3 and Myod1. Myoblasts were transfected with control siRNA and siRNA against Myod1. Myoblasts were differentiated into myotubes 48 hr after transfection. 94 hr after transfection, myotubes were crosslinked following a 2 hr treatment with TGF-β. ChIP was performed against Smad3 and IgG. RT-PCR was performed to quantify the fold enrichment of Smad3 relative to IgG (y-axis) at the indicated genes. Myod1 mRNA remained at 48% of controls when cells were crosslinked (left).

(D) Deletion of *Sfpi1*(PU.1) and *Spib* result in reduced Smad3 binding at sites normally co-occupied by Smad3 and PU.1. Smad3 ChIP was performed in wildtype (WT) and *Sfpi1*$^{-/-}$ *Spib*$^{-/-}$ pro-B cells. Double knockout cells were analyzed due to the high level of homology and redundancy of these two factors in pro-B cells (DeKoter et al., 2002; Schweitzer and DeKoter, 2004). RT-PCR was performed to calculate the fold enrichment of Smad3 under both conditions (y-axis, log scale). *Tnfrsf19, Vpreb2, Rag2*, and *Jund* are co-occupied by PU.1 and Smad3. *Il2ra* is occupied by Smad3 alone, and Smad3 occupancy increases with loss of PU.1 and Spib at *Il2ra*.

(E) Nucleosomes are depleted at sites occupied by master transcription factors. ChIP-seq was performed to map genome-wide H3 occupancy in mES cells, myotubes and pro-B cells. Meta-regions were created by summing the relative histone density at all sites occupied by Oct4 in mES cells, Myod1 in myotubes and PU.1 in pro-B cells. The x-axis indicates the distance from the master transcription factor binding in kb, and the y-axis represents relative enrichment of histone H3. Oct4 occupies sites of nucleosome depletion in mES cells, while Myod1 occupies sites of nucleosome depletion in myotubes and PU.1 occupies sites of nucleosome depletion in pro-B cells.

(F) Smad3 is directed to sites already depleted of nucleosomes. ChIP-seq was performed to map genome-wide H3 occupancy in mES cells under normal conditions (nl) and after treatment with SB431542 (SB) for 24 hr to block TGF-β signaling. Relative H3 density (y-axis) is shown for normal conditions and for SB431542 treatment at sites normally co-occupied by Oct4 and Smad3 in mES cells. The high degree of overlap between these conditions suggests that nucleosome depletion at these sites is

independent of TGF-β signaling. Thus, Smad3 is directed to sites occupied by Oct4 that are already depleted of nucleosomes.

Extended Experimental Methods

Growth conditions for cells

Human embryonic stem cells

Human embryonic stem (hES) cells (BGO3) were maintained in feeder free conditions using defined media (Ludwig et al., 2006). Cells were grown in a monolayer on tissue culture plates coated with matrigel at the dilution recommended by the manufacturer (BD, 354277). Cells were maintained in mTESR1 media and supplement (Stemcell Technologies, 05850) with 100 U/mL penicillin and 100 mg/mL streptomycin and passaged as described in the Stemcell Technologies protocol. Briefly, cells were washed two times with DMEM/F12 (Invitrogen, 11320) before treatment with dispase (Stemcell Technologies, 07913) for seven minutes at 37°C. Cells were then washed three times with DMEM/F12 before being scraped from the plate using cell lifters (Corning, 3008) and resuspended in mTESR1. Cells used for these experiments were between passage 40-60 and had been maintained off feeders for four passages. Cells were grown to 80% confluence before crosslinking. No additional TGF-β or Activin was added to hES cells prior to crosslinking.

Murine embryonic stem cells

V6.5 murine (m) ES cells were maintained on irradiated murine embryonic fibroblasts (MEFs) and expanded for two passages on gelatinized-tissue culture plates without MEFs prior to all experiments. Cells were grown under standard mES cell conditions as described previously (Marson et al., 2008). Briefly, cells were grown on 0.2% gelatinized (Sigma, G1890) tissue culture plates in mES cell media composed of DMEM-KO

(Invitrogen, 10829-018) supplemented with 15% fetal bovine serum (Hyclone, characterized SH3007103), 1000 U/mL LIF (ESGRO, ESG1106), 100 mM nonessential amino acids (Invitrogen, 11140-050), 2 mM L-glutamine (Invitrogen, 25030-081), 100 U/mL penicillin, 100 mg/mL streptomycin (Invitrogen, 15140-122), and 8 nL/mL of 2-mercaptoethanol (Sigma, M7522). Cells were grown to 80-90% confluence before crosslinking. For microarray expression analysis, cells were cultured in normal conditions and in the presence of 10 μM SB431542 (Tocris, 1614) for 24 hr before RNA was extracted using TRIzol reagent (Invitrogen, 15596-026). mES cells used for Smad2/3 ChIP-seq and for the replicate of Smad3 ChIP-seq were treated with 10 ng/ml Activin (R&D, 338-AC) to activate the TGF-β pathway for 1 hr prior to crosslinking. The presence of Activin expanded the number of Smad3-bound regions compared to normal culture conditions and increased peak height by 20 percent, but did not change the percent co-occupancy with Oct4 (Table S1). This treatment was necessary for a successful Smad2/3 ChIP due to a weaker antibody.

Loss of Oct4 expression in mES cells

ZHBTc4 mES cells were treated with doxycycline to repress Oct4 expression as previously described (Niwa et al., 2000). Briefly, both alleles of *Pou5f1* (the gene encoding Oct4) are disrupted in ZHBTc4 cells. These cells express a tetracycline regulated transactivator protein (tTA) and a tetracycline repressible *Pou5f1* transgene. Oct4 protein is expressed in the absence of doxycycline, but is repressed with addition of doxycyline. Cells were treated with and without doxycyline (2.0 μg/ml) for 24 hr before crosslinking or protein extraction.

Induction of Myod1 in mES cells

ES[MC1R(20)]:tetMyod1 cells were used to induce Myod1 expression as previously described (Nishiyama et al., 2009). This line of MC1 mES cells contains a tetracycline-repressible Myod1 cassette inserted into the ROSA26 locus. Briefly, mES cells were cultured in mES cell media as described in *Murine embryonic stem cells* except that penicillin and streptomycin were replaced with puromycin (1.0 μg/ml for continued selection) and doxycycline (0.2 μg/ml). Cells were expanded on MEFs and then passaged two times on gelatinized plates in the absence of MEFs prior to analysis. Myod1 expression was induced by doxycycline withdrawal as described (Nishiyama et al., 2009). Media was changed daily, and cells were analyzed at the specified time points relative to induction of Myod1. Cells analyzed on day 5 were cultured in mES cell media as described without the addition of exogenous TGF-β or Activin.

Murine C2C12 cells

C2C12 myoblasts (ATCC, CRL-1772) were expanded in C2C12 growth media and differentiated into myotubes as previously described (Caretti et al., 2004; Yaffe and Saxel, 1977). Briefly, cells were cultured in DMEM (Invitrogen, 11965) supplemented with 20% fetal bovine serum, 2 mM L-glutamine, 100 U/mL penicillin, 100 mg/mL streptomycin. To induce differentiation, C2C12 cells were grown to confluence and shifted to differentiation media containing DMEM, 2% horse serum (GIBCO, 26050-070), 1x transferrin/selenium/insulin (GIBCO, 51300-044), 2 mM L-glutamine, 100 U/mL penicillin, 100 mg/mL streptomycin. Multinucleated myotubes were visible after 48 hr of

culture in differentiation media. Approximately 1x 10^8 myotubes were treated with 2.5 ng/ml TGF-β (R&D Systems, 240-B) for 2 hr prior to crosslinking for ChIP-seq analysis. For genome-wide expression analysis, C2C12 myotubes were maintained in normal differentiation media or treated for 12 hr with 2.5 ng/ml TGF-β before RNA extraction.

Myod1 knockdown experiments

siRNA transfections were performed as previously described (Blais et al., 2007). Briefly, siIMPORTER reagent (Millipore, 64-101) was used according to manufacturer's protocol and the siRNA:siIMPORTER reagent ratio was adjusted for optimal knockdown. The specificity of knockdown was checked using two independent siRNA constructs that targeted Myod1 (Ambion, 4390816 siRNA ID# s232595 and ID# s232596). C2C12 myoblasts were seeded onto 10 cm plates (Corning, 430167) and transfected the following day using siRNA against Myod1 (ID# s232595) and a non-targeting siRNA control (Ambion, 4390844). For each plate, 20μM siRNA was diluted to a final concentration of 2.5 μM in serum free DMEM containing 120 μl siRNA diluent (Millipore, 20-272), and 84 μl siIMPORTER was diluted 8.6-fold in serum free DMEM. The diluted siRNA and siIMPORTER were combined and incubated at room temperature for 30 minutes. Cells were treated with1.68 mL siIMPORTER/siRNA solution diluted to 6 mL in serum-free DMEM for 4 hr at 37°C before an equal amount of growth media was added. Growth media was replaced 24 hr later. Cells were induced to differentiate into myotubes 48 hr after transfection. Cells were treated with 2.5 ng/ml of TGF-β 94 hr after transfection and crosslinked 2 hr later. Approximately 1x10^8 cells were used for ChIP-seq analysis.

Murine 38B9 pro-B cells

38B9 cells (Ramakrishnan and Rosenberg, 1988) were grown in suspension in RPMI-1640 (Invitrogen, 22400), 10% fetal calf serum, 2 mM L-glutamine 100 U/mL penicillin, 100 mg/mL streptomycin, and 8 nL/mL of 2-mercaptoethanol. Cells were treated with 5 ng/ml of TGF-β for 1 hr prior to crosslinking for ChIP-seq analysis. Cells were crosslinked at a concentration of approximately $1x10^6$ cells/ml. For genome-wide expression analysis, cells were either maintained in normal media or treated with 5 ng/ml TGF-β for 12 hr prior to RNA extraction.

Growth of wildtype and Sfpi1(PU.1)$^{-/-}$ Spib$^{-/-}$ pro-B cells

Wildtype and *Sfpi1*$^{-/-}$ *Spib*$^{-/-}$ IL-7-dependent cultured pro-B cells were generated from murine fetal liver progenitor cells using procedures previously described (DeKoter et al., 2002; Schweitzer and DeKoter, 2004). Pro-B cells were maintained in early log phase growth by continuous passage in cultures containing complete Iscove's Modified Dulbecco's medium, 5% conditioned medium from an IL-7-secreting J558 plasmacytoma cell line (Winkler et al., 1995), and mitotically arrested ST2 stromal cells. ST2 cells were maintained in complete IMDM media. ST2 cells were mitotically arrested by incubation with mitomycin C (20 mg/ml, Sigma) for 2 hr, then washed twice with complete medium and re-plated at 2.5×10^4 cells/cm^2. Pro-B cells were grown to approximately 10^8 cells at a maximum concentration of 10^6 cells/ml. 1 hr prior to crosslinking TGF-β (5 ng/ml) was added to the culture media. Pro-B cells were removed from ST2 monolayers by gentle pipetting prior to crosslinking for ChIP-seq analysis.

Chromatin immunoprecipitation

A summary of the bound sites determined for all ChIP-seq data is contained within Table S1.

For ChIP-seq experiments the following antibodies were used: Oct4 (Santa Cruz Bio, sc8628), Smad3 (Abcam, ab28379), Smad2/3 (Gift from D. Wotton), Myod1 (Santa Cruz Bio, sc760), PU.1 (Santa Cruz Bio, sc352), H3 (Abcam, ab1791), p300 (Santa Cruz Bio, sc-585), and IgG (Miilopore, 12-370).

The protocol for Chromatin Immunoprecipitation (ChIP) was performed as previously described (Lee et al., 2006). Cells were chemically crosslinked by the addition of one-tenth volume of fresh 11% formaldehyde. BG03, V6.5 and myotubes are adherent and were crosslinked for 10 minutes at room temperature. Cells were rinsed twice with 1X PBS and harvested using a silicon scraper and flash frozen in liquid nitrogen. 38B9 cells, wild-type pro-B cells and *PU.1*$^{-/-}$ *Spib*$^{-/-}$ pro-B cells were grown in suspension and were crosslinked for 20 minutes at room temperature. Crosslinking was quenched with one-twentieth volume 2.5 M glycine. Cells were washed twice with 1X PBS and flash frozen in liquid nitrogen. Cells were stored at -80^{o}C prior to use. Cells were resuspended, lysed and sonicated to shear and solubilize crosslinked DNA. Appropriate sonication conditions vary depending on cells, culture conditions, crosslinking and equipment.

Sonication was performed on approximately $1x10^8$ cells in sonication buffer (10 mM Tris-HCl pH8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate and 0.5% N-lauroylsarcosine) using a Misonix Sonicator 3000. BGO3 cells, myotubes and pro-B cells were sonicated at 21 watts for 8 x 20 second pulses (60 second pause between pulses). V6.5 mES cells were sonicated under the same conditions for 9 x 20 second pulses. After sonication, samples were divided in half. 1% Triton-X was added for samples to be precipitated with Oct4, Myod1, PU.1, and p300. Sonicated samples were centrifuged at 20,000g for 10 minutes and the soluble whole cell extracts were incubated overnight with 50 μl of Dynal Protein G magnetic beads that had been pre-incubated with 5 mg of the appropriate antibody. Beads were washed 1X with 20mM Tris-HCl pH8, 150 mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100, 1X with 20 mM Tris-HCl pH8, 500 mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100, 1X with 10 mM Tris-HCl pH8, 250 nM LiCl, 2 mM EDTA, 1% NP40 and 1X with TE containing 50 mM NaCl. Smad3 and Smad2/3 ChIPs were performed as described above except that immunoprecipitation was performed without addition of 1% Triton-X, and beads were washed 2X with TE containing 50 mM NaCl. All cells for H3 ChIPs were sonicated at 24 watts for 8 x 30 second pulses (80 second pause between pulses). Immunoprecipitation was performed after addition of 1% Triton-X. Beads were washed 4X with RIPA buffer (50 mM HEPES, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Na-Deoxycholate) and 1X with TE containing 50 mM NaCl. IgG ChIPs were performed under the same conditions as the experimental ChIPs to which they are compared.

Bound complexes were eluted from the beads (50 mM Tris-Hcl, pH 8.0, 10 mM EDTA and 0.5-1% SDS) by heating at 65°C for 45 minutes and vortexing every 5 minutes. Crosslinking was reversed by incubating samples at 65°C for 6 hr. Whole cell extract DNA reserved from the sonication step was treated in the same way to reverse crosslinking. After reversal of crosslinking, samples were treated with RNAse and proteinase K prior to preparation for sequencing.

ChIP-seq sample preparation and analysis

All protocols for Illumina/Solexa sequence preparation, sequencing and quality control are provided by Illumina (http://www.illumina.com/pages.ilmn?ID=203). A summary of the technique and minor protocol modifications are described below.

Sample Preparation

DNA was prepared for sequencing according to a modified version of the Illumina/Solexa Genomic DNA protocol. Fragmented DNA was prepared for ligation of Solexa linkers by repairing the ends and adding a single adenine nucleotide overhang to allow for directional ligation. A 1:100 dilution of the Adaptor Oligo Mix (Illumina) was used in the ligation step. A subsequent PCR step with limited (18 cycle) amplification cycles added additional linker sequence to the fragments to prepare them for annealing to the Genome Analyzer flow-cell. The amplified samples were separated on a 2% agarose gel and products between 150-350 bp (representing shear fragments between 50 and 250 nucleotides in length and ~100 bp of primer sequence) were excised. The

DNA was purified from the agarose and diluted for loading on the flow cell. Human and mES cell samples were prepared using the Illumina/Solexa Genomic DNA Kit. Myotube, pro-B cell and all H3 ChIP samples were prepared in a similar manner using individually purchased reagents with the following differences. End repair of fragmented DNA was performed using the End-It-DNA Repair Kit (Epicentre, ER0720). DNA was purified with the Qiaquick PCR Purification Kit (21806). Purified DNA was treated with Klenow fragment (NEB, M0212) and 1 mM dATP for 37°C for 30 minutes to add an A tail. DNA was purified by Qiaquick MinElute Purification Kit (28006). Adapters from the Illumina/Solexa Kit were ligated onto the fragmented DNA with DNA ligase (Promega, M8221) during a 15 minute incubation at room temperature. DNA was purified by MinElute Purification Kit. PCR was performed using Phusion NZ (NEB, F531) and primers from the Illumina/Solexa Kit.

Polony Generation and Sequencing

The DNA library (2-4 pM) was applied to the flow-cell (8 samples per flow-cell) using the Cluster Station device from Illumina. The concentration of library applied to the flow-cell was calibrated such that polonies generated in the bridge amplification step originate from single strands of DNA. Multiple rounds of amplification reagents were flowed across the cell in the bridge amplification step to generate polonies of approximately 1,000 strands in 1 μm diameter spots. Double stranded polonies were visually checked for density and morphology by staining with a 1:5000 dilution of SYBR Green I (Invitrogen) and visualizing with a microscope under fluorescent illumination. Validated flow-cells were stored at 4^{o}C until sequencing.

Flow-cells were removed from storage and subjected to linearization and annealing of sequencing primer on the Cluster Station. Primed flow-cells were loaded into the Illumina Genome Analyzer II. After the first base was incorporated in the Sequencing-by-Synthesis reaction the process was paused for a quality control checkpoint. A small section of each lane was imaged and the average intensity value for all four bases was compared to minimum thresholds. Flow-cells with low first base intensities were re-primed and if signal was not recovered the flow-cell was aborted. Flow-cells with signal intensities meeting the minimum thresholds were resumed and sequenced for 26, 32, or 36 cycles.

ChIP-seq data analysis

Images acquired from the Illumina/Solexa sequencer were processed through the bundled Solexa image extraction pipeline, which identified polony positions, performed base-calling and generated QC statistics. Bowtie (version 0.12.2) (Langmead et al., 2009) was used to align sequences to NCBI Build 36 (UCSC mm8) of the murine genome and NCBI Build 36 (UCSC hg18) of the human genome. Alignments were performed using the following criteria: -n2, -e70, -m2, -k2, --best. Only sequences that aligned to a unique location were used to determine enriched regions as described below. Previously published ChIP-seq quality score (FASTQ) files profiling the genomic occupancy of Oct4, Sox2, Nanog (Marson et al., 2008), Ronin (Dejosez et al., 2010), Zfx, and c-Myc (Chen et al., 2008) in mES cells were aligned using Bowtie as described above.

Analysis methods were derived from previously published methods (Marson et al., 2008). Each read (reads from biological replicates were combined) was extended 200 bp, towards the interior of the sequenced fragment, based on the strand of the alignment. Across the genome the number of ChIP-seq reads was tabulated in 25 bp bins. The 25 bp genomic bins that contained statistically significant ChIP-seq enrichment were identified by comparison to a Poissonian background model. Assuming background reads are spread randomly throughout the genome, the probability of observing a given number of reads in a 1kb window can be modeled as a Poisson process in which the expectation can be estimated as the number of mapped reads multiplied by the number of bins (40) into which each read maps, divided by the total number of bins available. Enriched bins within 200 bp of one another were combined into regions.

The Poissonian background model assumes a random distribution of background reads. However, significant deviations from this expectation have been observed. Some of these non-random events can be detected as sites of apparent enrichment in negative control DNA samples creating false positives. To remove these false positive regions, negative control DNA from whole cell extract (WCE) or IgG ChIPs were sequenced for each cell type. Enriched bins and enriched regions were required to have greater than five-fold density in the experimental sample compared with the control sample when normalized to the total number of reads in each dataset. This served to filter out genomic regions that are biased to having a greater than expected background density

of ChIP-seq reads. For mouse, the complete set of RefSeq genes was downloaded from the UCSC website (http://hgdownload.cse.ucsc.edu/goldenPath/mm8/database/refGene.txt.gz) on March 5, 2010. For human, the complete set of RefSeq genes was downloaded from the UCSC table browser (http://genome.ucsc.edu/cgi-bin/hgTables?command=start) on March 1, 2009. Genes with enriched regions within 10kb upstream of the transcription start site or within the body of the gene were called bound.

A summary of the bound sites (Table S1) and bound genes (Table S3) for each antibody is provided. Sites were considered bound by a factor below a p-value of $<10^{-9}$, except for c-Myc, Ronin, and Smad3 day5 no Myod1 where a p-value of $<10^{-5}$ was used. The peak heights for each binding site are contained in Table S1. These data represent merged biological replicates for myotubes and pro-B cells. Data for Smad3 in mES cells contains separate analysis for biological replicates. These samples were not combined because the second sample was analyzed after treatment with Activin for 1 hr as described above. Data for BG03 hES cells and Smad2/3 in mES cells represent single experiments.

Analysis of enriched sites defined by ChIP-seq

ChIP-Seq Enriched Sites Maps (Fig 1C, H, 2B, 4E, S2B)

The visualization shows the location of enriched sites (P-val $<10^{-9}$, Table S1) in a collection of datasets (query datasets, indicated on the top) in relation to the enriched sites of another dataset (base dataset, indicated on the y-axis). For example, the

visualization in Fig 1C shows the location of Oct4 and Smad3 binding (query datasets, x-axis) in relation to Oct4 enriched sites (base dataset, y-axis). For each of the enriched sites in the base dataset, corresponding genomic sites were calculated as +/- 2.5kb from the center of that enriched site (one genomic site per enriched site row). For each of these genomic sites, the location and length of any enriched sites in the query datasets were drawn.

Determination of co-occupied regions

Table S1 contains the genomic coordinates of enriched sites (P-val $<10^{-9}$) co-occupied by the indicated pair of factors. These coordinates are the union of all overlapping enriched sites of the two factors. Overlapping is defined as having at least one base of shared sequence. It is possible for an enriched site of one factor to span (or bridge a gap between) two separate enriched sites of the other factor. In those cases, only one enriched site would be reported and it would be the union of all three enriched sites.

Calculation of p-values for enriched region overlaps:

We developed a conservative statistical method for determining the probability of observing the overlap of a particular number of enriched regions. The method needed to account for the fact that the whole genome is not accessible (or mappable), and therefore, there is not an equal probability of finding an enriched region across the genome. Furthermore, all enriched regions cannot be enumerated, which makes calculation of expectation difficult.

An empirical method was used to determine the probability of observing an overlap between two enriched regions. To perform this calculation, a large set of enriched regions (Table S2) was compiled for mES cells. The following datasets were used: Oct4, Sox2, Nanog, Tcf3, Suz12 (GSE11724), c-Myc, CTCF, E2F1, Essrb, Klf4, n-Myc, Tcfcp2l1, Zfx (GSE11431), Pol2 (GSE12680), TBX3 (GSE19219), TBP (GSE22303), Smc1, Smc3, Med1, Med12, Nipbl (GSE22562), Ronin (Dejosez et al., 2010) and Smad3. Two regions were randomly selected from this collection and it was determined if they overlapped. This process was repeated 1,000,000 times to calculate the expected probability that two enriched regions overlapped. This probability is defined as pO. By calculating the expected probability empirically from a collection of actual datasets, we are intrinsically accounting for the accessibility of the genome, and because the collection is large, we are exploring the space available for enriched regions. However, because many of the datasets used are of factors that may co-occupy sites in the genome, our empirical estimate of pO is probably an overestimate (since there is a bias in our collection towards overlap), and therefore makes the following p-value calculations more conservative than they may be in reality.

The binomial distribution was used to determine the p-value for observing at least an overlap of N regions between a dataset of size A and a dataset of size B, using the R command:

```
1-pbinom(N,A,B*pO)
```

For comparison of regions in murine cells outside of mES cells, the value of pO was assumed to be constant across cell-types and is therefore applicable. For comparisons in hES cells a new empirical pO was calculated using a collection of human enriched regions (CTCF, Med1, Med12, Oct4, Pol2, Smad3, Stat3, Suz12, TCF7l1, TBP).

The top 1000 enriched regions of a dataset may have a different expected probability of overlap because these regions are the least noisy and may have a slightly different genomic distribution. To account for this possibility, pO was calculated using a collection of the top 1000 regions in both murine and human ES cells and this value of pO was used where appropriate.

Gene transcription status and pie charts (Fig 1A, 1F)

Each gene in the HG18 or MM8 genome builds were classified into active, bivalent, or silent groups based on the presence of co-occupancy of H3K4me3 (GSM575281,GSE11724), H3K79me2 (GSM602673,GSE11724), and H3K27me3 (GSM602672,GSM307619). See the table below for a description of the classification rules.

	Classification rules		
	H3K4me3 within +/- 2kb of the TSS	H3K79me2 within first 5kb of gene body	H3K27me3 within +/- 5kb of the TSS
Active	Required	Required	-
Bivalent	Required	-	Required
Silent	Absent	Absent	-

Smad binding element (SBE) scanning (Fig 1E, J, 4H)

Sequence selection

Repeat masked sequence was downloaded from UCSC (build HG18 for Fig 1E and MM8 for Fig 1J and 4H) for the regions +/-2.5kb around the peak of each bound region (Table S1).

Motif counting

For a given set of repeat masked sequences (see above), a 250 bp window was advanced across the sequences in 50 bp increments. Within each window the average number of occurrences of the AGAC quadruplet or its reverse complement, GTCT, was calculated.

Motif discovery

DNA motif discovery was performed as previously described (Marson et al., 2008). Briefly, a modified version of the ChIP-seq read mapping algorithm was used. The genomic bin size was reduced to 10 bp and the read extension placed greater weight towards the middle of the 200 bp extension. Greater weight was placed toward the middle of the extension to increase the precision of the peak for each site and worked by placing 1/3 count in the 8 bins from 0-40 and 160- 200 bp, 2/3 counts in the 8 bins from 40-80 and 120-160 bp and 1 count in the 4 bins from 80-120 bp. 100 bp of genomic sequence, centered at the 500 largest peaks of ChIP-seq density, were submitted to the motif discovery tool MEME (Bailey and Elkan, 1994) to search for over-represented DNA motifs. Motifs were restricted to a maximum of 15 nucleotides and

could be present on the + or - strand. The top 5 motifs enriched at sites bound by each factor are shown in Fig S1 with the associated E-value.

Distance from Smad3 site to nearest TF site (Fig 2D)

The distance (in bp) from each Smad3 site in mES cells (Table S1) to the nearest site of each TF was calculated. The distance from the center of each Smad3 site to the center of the nearest site bound by the indicated transcription factor was determined. These distances were grouped into bins (x-axis), with the x-axis value indicating the upper limit of each bin. The sum in each bin is shown (y-axis).

High-resolution placement of Oct4, Sox2, and Smad3 binding sites (Fig 3H):

The same method described in the *ChIP-seq data analysis* was used to resolve the binding sites of Oct4, Sox2 and Smad3 at high resolution with the exception that 5bp bins were used to map regions bound by each factor instead of 25bp bins. The set of Oct4-bound genomic regions that were co-occupied by Sox2 and Smad3 were selected for analysis. The distance between the peak of Oct4 (5bp bin with the most reads) to the peak of either Sox2 or Smad3 was calculated. The relative ordering of Oct4 and Sox2 peaks is conserved (Chen et al., 2008; Marson et al., 2008) and can be used to correct for strand. Thus, the distance between Oct4 and Sox2 peaks was defined as positive and the distance between Oct4 and Smad3 peaks could be positive (on the same side of Oct4 as Sox2) or negative (on the opposite side of Oct4 from Sox2). Figure 3H shows the histogram of the distances between the peaks of Oct4 and Sox2 (black) or the peaks of Oct4 and Smad3 (red).

H3 density plots (Fig 3I, 4I, S3E,F)

The average read density within the H3 dataset was calculated in 10 bp bins +/- 1kb around the peaks of the bound regions indicated for each cell type. The average density across the whole region was then mean normalized and plotted. The background dataset used in Fig 3I was produced by randomly generating a set of enriched regions that had the same distribution in size and distance from the nearest TSS as the Oct4 dataset to which it was compared. For Fig S3, the average H3 density around the Oct4/Smad3 co-occupied sites (Table S1) was determined for normal mES cells (dark blue) and mES cells treated with SB431542 for 24 hr (light blue). The data were calculated, normalized, and plotted as described above.

Expression arrays

Genomic expression analysis was measured using Agilent Whole-Mouse Genome Microarrays (Agilent, G4122F). Total RNA was isolated from cells using TRIzol reagent following the protocol for cells grown in monolayer and suspension as appropriate (Invitrogen, 15596-026). RNA was further purified with RNeasy columns (Qiagen, 74104) after DNase treatment (Invitrogen, 18068-015) following the manufacturers' protocols. RNA samples from two biological replicates were used for duplicate microarray expression analysis. 2 μg of RNA were labeled for each sample using the two-color low RNA Input Linear Amplification Kit PLUS (Agilent, 5188-5340). Briefly, double-stranded cDNA was generated using MMLV-RT enzyme and an oligo-dT based primer. *In vitro* transcription was performed using T7 RNA polymerase and either Cy3-

CTP or Cy5-CTP, directly incorporating dye into the cRNA. Labeled cRNA was hybridized overnight at 65°C. The hybridization cocktail consisted of 825 ng cy-dye labeled cRNA for each sample, Agilent hybridization blocking components, and fragmentation buffer. The hybridization cocktails were fragmented at 60°C for 30 minutes, and then Agilent 2X hybridization buffer was added to the cocktail prior to application to the array. The arrays were hybridized for 16 hr at 60°C in an Agilent rotor oven set to maximum speed. The arrays were treated with Wash Buffer #1 (6X SSPE / 0.005% n-laurylsarcosine) on a shaking platform at room temperature for 2 minutes, and then Wash Buffer #2 (0.06X SSPE) for 2 minutes at room temperature. The arrays were then dipped briefly in acetonitrile before a final 30 second wash in Agilent Wash 3 Stabilization and Drying Solution.

Arrays were scanned using an Agilent DNA microarray scanner. Array images were quantified and statistical significance of differential expression was calculated for each hybridization using Agilent's Feature Extraction Image Analysis software with the default two-color gene expression protocol.

Determination of change in gene expression

Converting probe level expression to gene-level

Biological replicates of each expression array were generated. To calculate an average dataset from the biological replicates the log10 ratio values for each feature were averaged and the log ratio p-values were multiplied. Then, for each gene we selected the set of probes that map to that gene using the Agilent provided probe-to-gene

mapping. From this set, we used the probe with the median level of expression change as the representative for that gene. For genes with an even number of probes, the average of the two middle probes was used. The gene level expression data, as well as the representative probe(s) used can be found in Table S4. A gene was determined to be significantly affected by TGF-β signaling if its gene-level expression changed by at least 1.5 fold (absolute log10 ratio of greater than 0.176) with a p-value less than or equal to 0.05. Genes without annotated features are reported as NA and were excluded from any expression analysis. Genes that changed in expression by more than 50%, but had a p-value greater than 0.05, were classified as unaffected and excluded from the analysis.

Expression heat maps (Fig 6D)

Fig 6D is an expression heatmap of the 2079 genes (one per row) affected by TGF-β signaling in only one of the three cell-types. 358 genes were excluded from this analysis because they were affected in more than one cell type. Genes affected in mES cells were placed at the top, genes affected in myotubes in the middle and genes affected in pro-B cells at the bottom. Each gene is represented only once.

P-value calculation of enrichment of bound genes in TGF-β affected genes (Fig 6B, D)

The p-value of the significance of the overlap between genes occupied by the factor listed on the y-axis and genes significantly affected by TGF-β signaling in each cell-type was calculated using the hyper-geometric test. P-values were then corrected for multiple hypothesis testing using the Benjamini-Hochberg method. The corrected p-

values are plotted (x-axis). In order for a transcription factor to be considered significantly binding affected genes, a corrected p-value at or below $1x10^{-5}$ was required.

ChIP-re-ChIP

The initial ChIP was performed using $4x10^{8}$ cells and 400 ul of beads bound to the Oct4 antibody. Cells were washed in low salt, high salt, LiCl and TE with 50 mM NaCl as described. The IP was then incubated at room temperature for 30 minutes in 100 μl of LB3 containing 100ug of the peptide used to generate the Oct4 antibody (Santa Cruz). The sample was vortexed every 5 minutes. After 30 minutes the eluted material was removed and elution was repeated with fresh LB3 and peptide. A second round of ChIP was performed using antibodies against Smad3 and IgG under the conditions described. Enrichment was quantified by quantitative (q)PCR.

ChIP-PCR

qPCR was performed using templates of purified genomic DNA after ChIP as described in above in the section titled “Chromatin immunoprecipitation”. Syber Green real-time PCR was carried out using the 7000 ABI Detection System according to the manufacturer’s instructions (Applied Biosystems). Data were normalized to whole cell extract and internal control regions for ChIP-re-ChIP, and data were normalized to IgG ChIPs and internal control regions for knockdown. PCR was performed in triplicate and error bars show standard deviation. Actin primers were used for internal control regions for mES cells and myotubes. Myod1 primers were used for internal control regions for

pro-B cells. Data shown are representative experiments. Experiments in mES cells were performed in biological duplicate, myotubes in singlicate and pro-B cells in triplicate. Primers are listed below.

Lefty1
5’-ATCCCCAATCCACATTCACT-3’
5’-GTAGCCAGCAGACAGGACAA-3’
Lefty2
5’-GCAATCTGCCCACTGTAAAA-3’
5’-TCGATCTTCCCAAGACTCCT-3’
Nodal
5’-TAGATAGCCACGCCCTCTCT-3’
5’-AGGACGCTGTCTCCTGAACT-3’
Pou5f1 (Oct4)
5’-ACGGCAGATGCATAACAAAG-3’
5’-TTAAGGAAGGGCTAGGACGA-3’
Smad7
5’-ACTTTGTTTTCGGAGGGATG-3’
5’-TTTGCGGTCATTGCTAACTC-3’
Desmin
5’-TGATGTCAGGAGGGCTACAA-3’
5’-GTGGATGTGAAGATGGGTGA-3’
Id1
5’-CCTAGGTCTCCTGCCTTGTC-3’
5’-CCTAACCCTGCCAATTTCAT-3’
Junb
5’-CAGGAAGGCTGCAGTTACTCT-3’
5’-AGCGGCTTTCAACAGACC-3’
Myod1
5’- GCTTCTTTCGGCCAAGTATC-3’
5’- TGTTGTGAGTCACGGGTTTT-3’

Pnmt

5’-CCCTAAGGCCTGTCAGTTGT-3’

5’-GCAGAGGTCACAGACAGGA-3’

Il2ra

5’- CCTCACAATCTTTGCTTCCA-3’

5’- TCAATCTCTCGATCCAGCAC-3’

Jund

5’-GCGCTTCATTTCCTCATTTT-3’

5’-GTCCTGTTTTCTGGGTTGCT-3’

Rag2

5’-TCTGTCTCCCTCAACCATCA-3’

5’-AGGCTGCAGGGTAGAGTGTT-3’

Tnfrsf19

5’-TCACTCCACAGCTTCTGTCC-3’

5’-TTCCGACCACAGTGCTTTAC-3’

Vpreb2

5’-GGGAGTACGTGTTCCTGGAT-3’

5’-CAGGAAATGGATGCACAAAG-3’

Actin, beta upstream

5’-GGCACCACACCTTCTACAAT-3’

5’-GTTACCCGGGATACTGACCT-3’

RNA Extraction, cDNA preparation, and TaqMan expression analysis

RNA utilized for RT-PCR was extracted with RLT buffer according to the manufacturer’s protocol (Qiagen, 74104) with in-column DNAse digestion (Qiagen, 79254). Purified RNA was reverse transcribed using Superscript III (Invitrogen, 18080) with random primers (Invitrogen, 48190) to prime first-strand synthesis according to the manufacturer’s protocol.

qPCR was carried out on the 7000 ABI Detection System using the following Taqman probes according to the manufacturer's protocol (Applied Biosystems). All experiments shown were performed at least in biological in triplicate. Expression was normalized to GAPDH, and fold change in expression was calculated relative to the indicated conditions. Error bars represent standard deviation. The following Taqman probes were used for expression analysis:

Oct4	Mm00658129_gH
Lefty1	Mm00438615_m1
Lefty2	Mm00774547_m1
Nodal	Mm00443041_m1
Id3	Mm00492575_m1
Gapdh	Mm99999915_g1
Myod1	Mm00440387_m1

Western Blots and Co-Immunoprecipitation

For Western Blot analysis, cells were harvested using CelLytic (Sigma, C2978) and protease inhibitors per the manufacturer's protocol. For Co-IPs using anti-Smad3, mES cells were harvested in cold PBS and extracted for 30 min at 4°C in TNEN250 (50 mM Tris pH 7.5, 5 mM EDTA, 250 mM NaCl, 0.1% NP-40) with protease inhibitors. After centrifugation, supernatant was mixed with 2 volumes of TNENG (50 mM Tris pH 7.5, 5 mM EDTA, 100 mM NaCl, 0.1% NP-40, 10% glycerol). Protein complexes from mES cells were immunoprecipitated overnight at 4°C using 5 µg of Smad3 (Abcam, ab28379) and Rabbit IgG (Upstate, 12-370) bound to 50 µl of Dynabeads®. Co-IPs using anti-pSmad3 were performed on cells in standard culture conditions and on cells cultured with SB431542 for 24 hr. pSmad3 Co-IPs were performed as above except Tris was

replaced with HEPES as a buffer, 100 ug of nuclear protein was loaded for each sample, and immunoprecipitation was performed at 4°C for 2.5 hr. All buffers contained protease inhibitor and HALT phosphatase inhibitors (Pierce, 78420). Myotubes and pro-B cells were treated with TGF-β as described in Growth Conditions for Cells and harvested as described for mES cells in the presence of protease and phosphatase inhibitors. Protein complexes from myotubes and pro-B cells were incubated at 4°C for 3 hr with 5 µg of Smad3 and IgG bound to 50 µl of Dynabeads. Immunoprecipitates were washed three times with TNEN125 (50 mM Tris pH 7.5, 5 mM EDTA, 125 mM NaCl, 0.1% NP-40). Beads were boiled for 10 minutes in XT buffer (Biorad, 161-0788) containing 100mM DTT to denature proteins. After SDS-PAGE, Western blots were performed using antibodies against Oct4 (Santa Cruz Bio, sc-5279), Smad3 (Abcam, ab28379), pSmad3 (Cell Signaling, 9520), Myod1 (Abcam, ab3106) and PU.1 (Santa Cruz Bio, sc-352) as indicated. Smad4 (Santa Cruz Bio, sc-7966) was used as a positive control for co-immunoprecipitation. TBP (Abcam, ab818) and GAPDH (Abcam ab9484) was used as a loading control.

Electrophoretic Mobility Shift Assay (EMSA)

To prepare nuclear extracts, V6.5 mES cells cells were plated in 15-cm plates (as described). Cells were treated with SB431542 for 24 hr and 1 hr prior to collecting nuclear fractions, cells were washed and then treated with 10 ng/ml Activin to activate the TGF-β pathway. Cells were then washed twice in cold PBS and resuspended in 5 ml ice-cold hypotonic lysis buffer (20 mM HEPES, pH 7.9, 20% glycerol, 10 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.1% Triton X-100, 0.5 mM dithiothreitol, 1 mM

phenylmethylsulfonyl fluoride) in the presence of HALT protease and phosphatase inhibitors (Pierce, 78440). After 10 min incubation on ice, nuclei were spun down, resuspended in 0.5 ml of nuclear extraction buffer (hypotonic lysis buffer and 420 mM NaCl) and rotated for 1 hr at 4°C. Supernatants were then clarified by centrifugation, aliquoted and stored at -80°C. Protein concentrations were determined using the Bradford protein assay (Biorad).

Oligonucleotide probes containing adjacent Smad and Oct4 binding sites were generated by labeling one DNA strand (forward strand) with T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P]-ATP (Perkin Elmer). Unincorporated [γ-^{32}P]-ATP was removed by G-25 spin columns (Roche). Labeled oligonucleotides were subsequently mixed with a 10-fold excess of unlabeled complementary (reverse stand) oligonucleotides. After heating for 5 min at 95°C, the oligonucleotides were placed at room temperature and allowed to anneal.

EMSAs were performed as follows: DNA binding reactions (20 µl) containing 20 mM HEPES (pH 7.9), 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 5% glycerol, 500 ng poly(dI-dC) and 10 µg nuclear extract were preincubated with or without specific competitor (100 fold excess) at room temperature for 20 minutes. Following preincubation, 50,000 cmp of radiolabeled DNA probe was added to the reaction mixtures and further incubated for 20 minutes. Each reaction mixture was then loaded onto a native 5% polyacrylamide gel (acrylamide:bis, 39:1) containing 0.5 X TBE and 1% glycerol and elecrophoresed in 0.5 X TBE at 300 V for 2 hr at 4°C. After elecrophoresis, the gels were fixed for 15 minutes in gel fixing solution (10% acetic acid, 20% methanol, 70%water), dried at 80°C for 1 hr and exposed to phosphorimager

screen (Fuji). For antibody supershift assays, nuclear extracts were first incubated with radiolabeled DNA probe before 1µl of each antibody was added to the reaction mixtures.

Murine Lefty1 enhancer:

5'-GGTGGGAGGGAGACTGCCCTTTGTCATGTAGAAGGAGCTT-3'

3'-AAGCTCCTTCTACATGACAAAGGGCAGTCTCCCTCCCACC-5'

Competitor for murine Lefty1 enhancer (with scrambled ends in italics):

5'-*ACGTG*CAGGGAGACTGCCCTTTGTCATGTAGAAG*CAGTTG*-3'

3'-*CAACTG*CTTCTACATGACAAAGGGCAGTCTCCCT*GCACGT*-5'

Supplementary References

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol *2*, 28-36.

Blais, A., van Oevelen, C.J., Margueron, R., Acosta-Alvear, D., and Dynlacht, B.D. (2007). Retinoblastoma tumor suppressor protein-dependent methylation of histone H3 lysine 27 is associated with irreversible cell cycle exit. J Cell Biol *179*, 1399-1412.

Caretti, G., Di Padova, M., Micales, B., Lyons, G.E., and Sartorelli, V. (2004). The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. Genes Dev *18*, 2627-2638.

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.

Dejosez, M., Levine, S.S., Frampton, G.M., Whyte, W.A., Stratton, S.A., Barton, M.C., Gunaratne, P.H., Young, R.A., and Zwaka, T.P. (2010). Ronin/Hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells. Genes Dev *24*, 1479-1484.

DeKoter, R.P., Lee, H.J., and Singh, H. (2002). PU.1 regulates expression of the interleukin-7 receptor in lymphoid progenitors. Immunity *16*, 297-309.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc *1*, 729-748.

Ludwig, T.E., Bergendahl, V., Levenstein, M.E., Yu, J., Probasco, M.D., and Thomson, J.A. (2006). Feeder-independent culture of human embryonic stem cells. Nat Methods *3*, 637-646.

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.

Nishiyama, A., Xin, L., Sharov, A.A., Thomas, M., Mowrer, G., Meyers, E., Piao, Y., Mehta, S., Yee, S., Nakatake, Y.*, et al.* (2009). Uncovering early response of gene regulatory networks in ESCs by systematic induction of transcription factors. Cell Stem Cell *5*, 420-433.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet *24*, 372-376.

Ramakrishnan, L., and Rosenberg, N. (1988). Novel B-cell precursors blocked at the stage of DJH recombination. Mol Cell Biol *8*, 5216-5223.

Schweitzer, B.L., and DeKoter, R.P. (2004). Analysis of gene expression and Ig transcription in PU.1/Spi-B-deficient progenitor B cell lines. J Immunol *172*, 144-154.

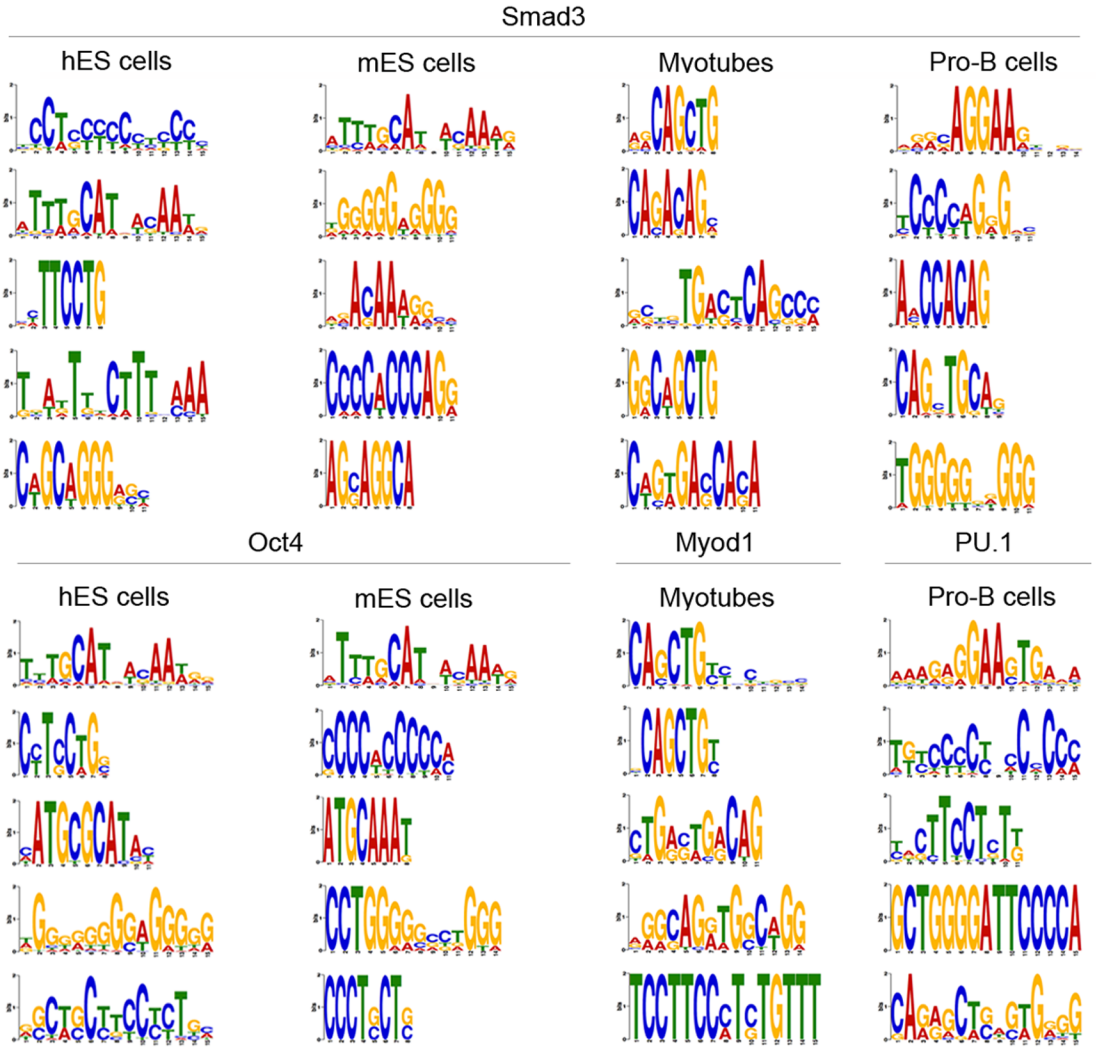
Winkler, T.H., Melchers, F., and Rolink, A.G. (1995). Interleukin-3 and interleukin-7 are alternative growth factors for the same B-cell precursors in the mouse. Blood *85*, 2045-2051.

Yaffe, D., and Saxel, O. (1977). Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature *270*, 725-727.

Supplemental Figure 1

Figure S1

A

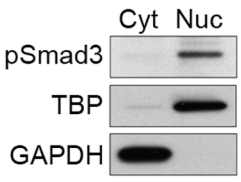


B

Smad3			
hES cells	mES cells	Myotubes	Pro-B cells
1.6e-90	**1.4e-125**	**1.0e-241**	**2.7e-197**
2.7e-57	4.8e-84	7.9e-26	2.4e-75
1.3e-30	1.4e-29	2.9e-6	6.9e-10
5.7e+3	3.4e+2	5.1e-5	6.3e-13
1.2e+4	4.3e+2	2.0e+3	1.4e-4

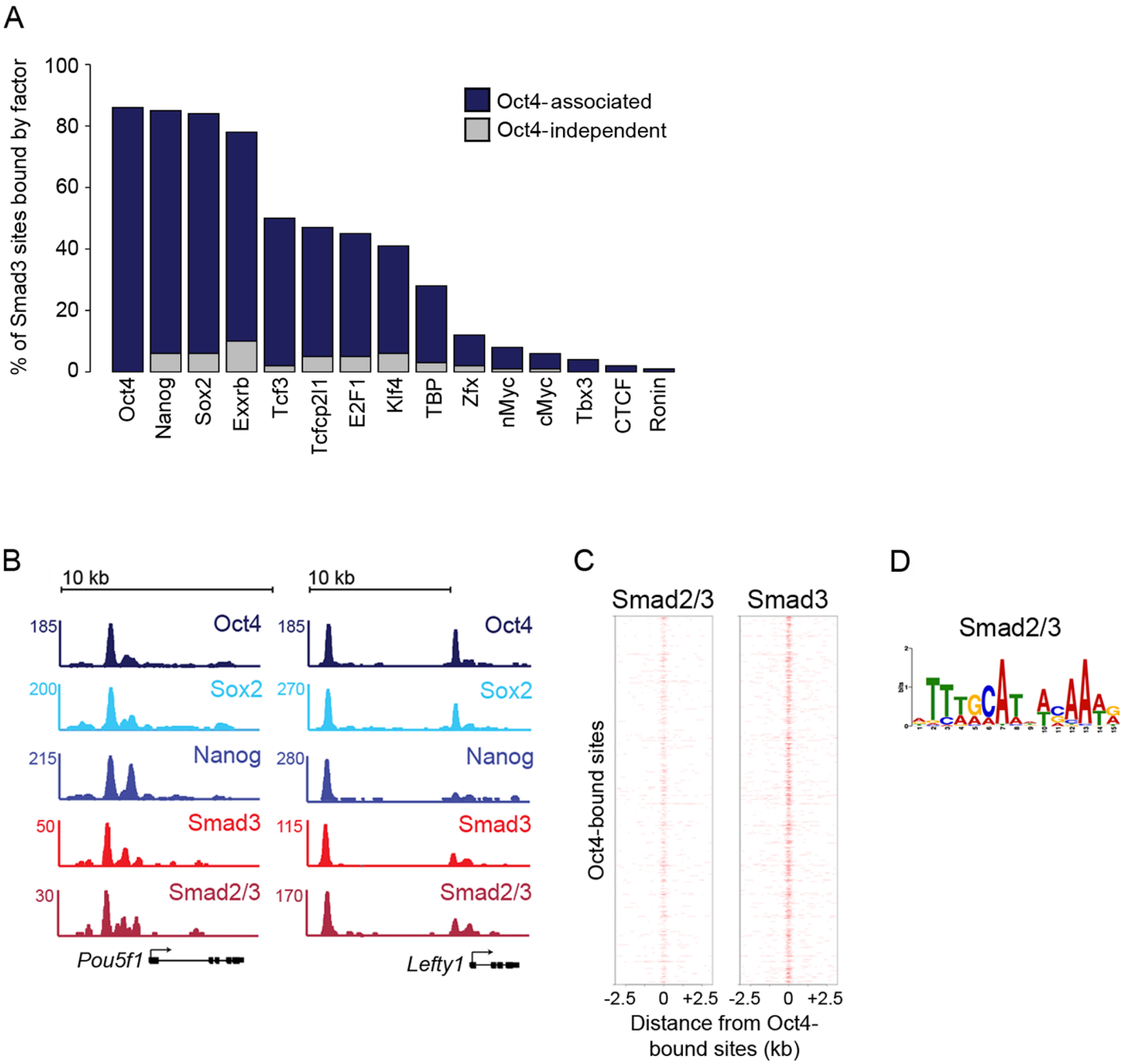
Oct4		Myod1	PU.1
hES cells	mES cells	Myotubes	Pro-B cells
1.4e-369	**2.7e-651**	**1.1e-584**	**1.9e-498**
3.0e-19	4.2e-26	8.7e-68	5.2e-3
1.2e-18	2.9e-5	4.6e-14	2.9e+1
7.9e-15	1.2e+3	8.1e+1	2.6e+6
2.4e+4	1.7+4	1.4e+6	4.3e+5

C



Supplemental Figure 2

Figure S2



Supplemental Figure 3

Figure S3