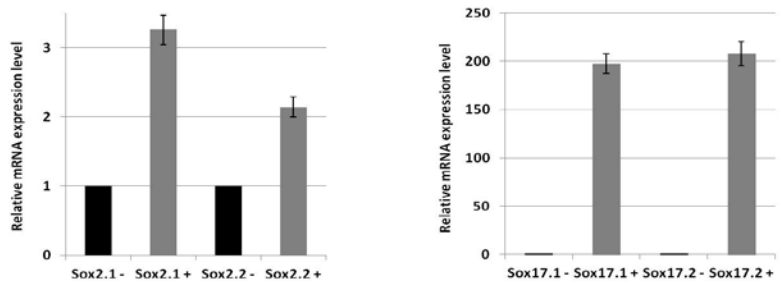
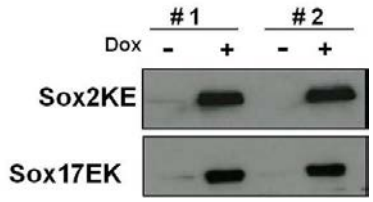
Supplementary Figure 1:



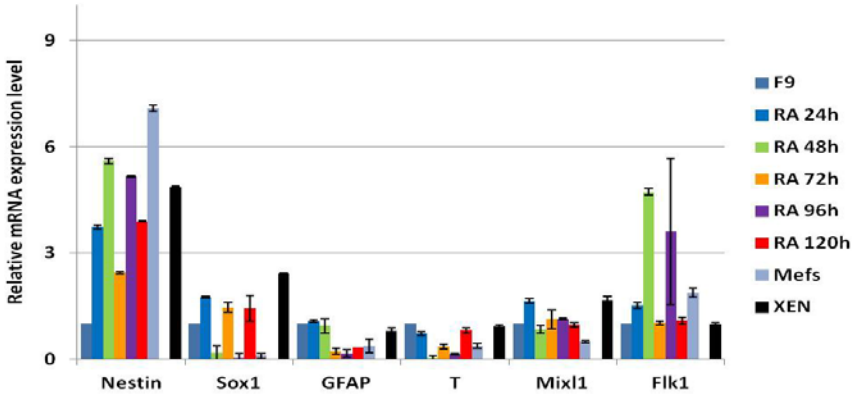
Quantitative RT-PCR of *Sox2* and *Sox17* in Sox2 and Sox17 inducible ESC lines treated with doxycycline for 48 hours in two independent clones. Expression levels are relative to the untreated controls.

Supplementary Figure 2:



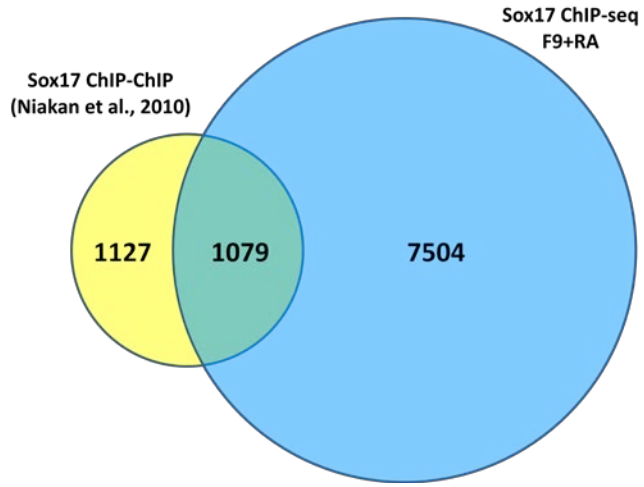
Western-blot showing the expression of the V5 tagged Sox2KE and Sox17EK proteins in 2 independent cell clones treated with doxycycline for 48 hours (related to Figure 4).

Supplementary Figure 3:



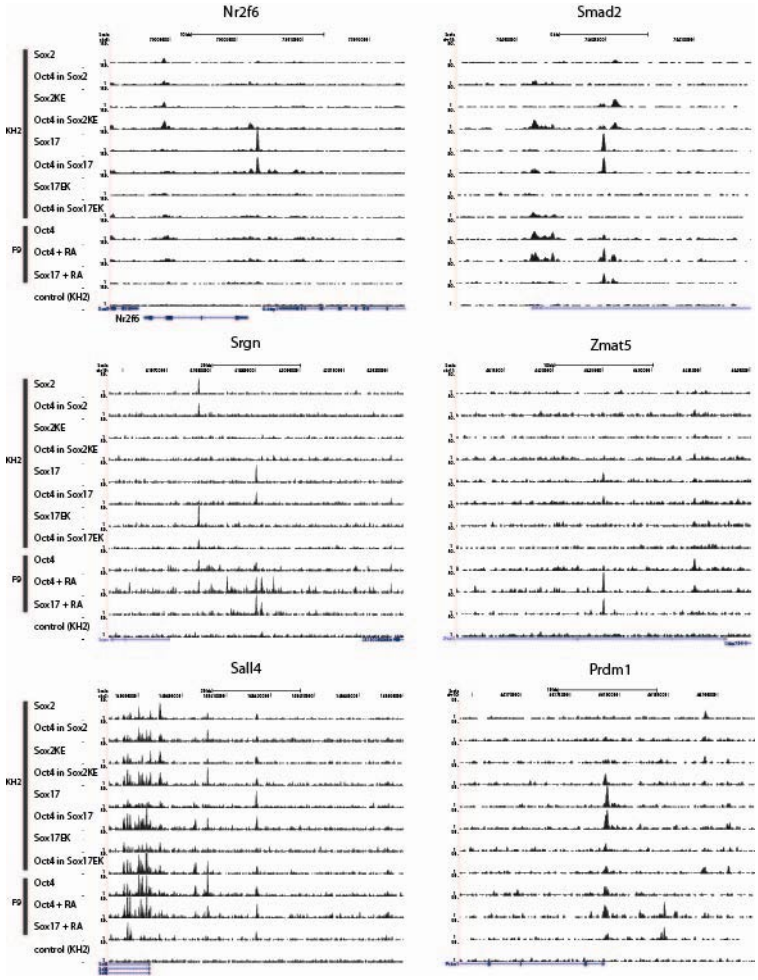
Relative mRNA expression as determined by Q-RT-PCR for Ectodermal markers, *Nestin, Sox1, GFAP and Mesodermal markers, T, Mixl1 and Flk1* in F9 cells treated with RA for 24 to 120 hours, Mefs and XEN cells.

Supplementary Figure 4:



Overlapping genes between Sox17 ChIP-seq data obbtained from F9+RA cells and Sox17 ChIP-chip data obtained from XEN cells (Niakan et al., 2010). For this analysis, the 6195 Sox17 binding sites found in F9+RA cells were used to collect a list of genes with a TSS (transcription start site) within 50kb of the peak summit. The collected genes were intersected with the list of Sox17 target genes identified in XEN cells by ChIP-chip (Niakan et al., 2010). 1079 genes were found to intersect between the two studies, but only 347 (+/- 22) genes were expected to overlap after random permutation using a Monte Carlo simulation (Hutchins et al., 2013).

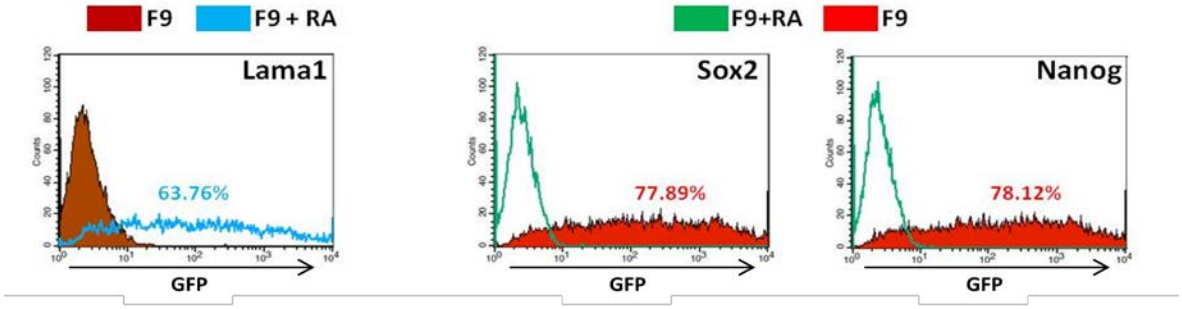
Supplementary Figure 5:



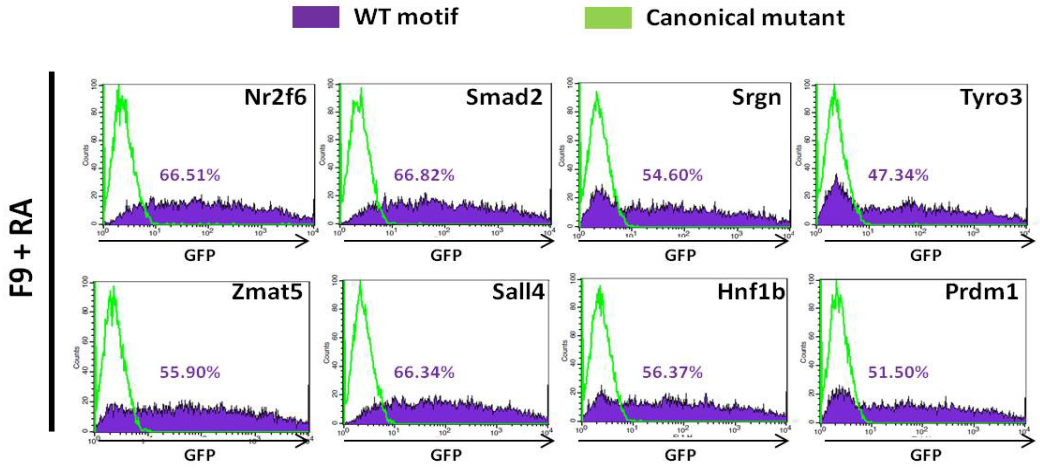
Binding profiles of Sox factors and Oct4 in KH2 and F9 cells at Nr2f6, Smad2, Srgn, Zmat5, Sall4 and Prdm1 genomic loci are shown (related to Figure 7).

Supplementary Figure 6:

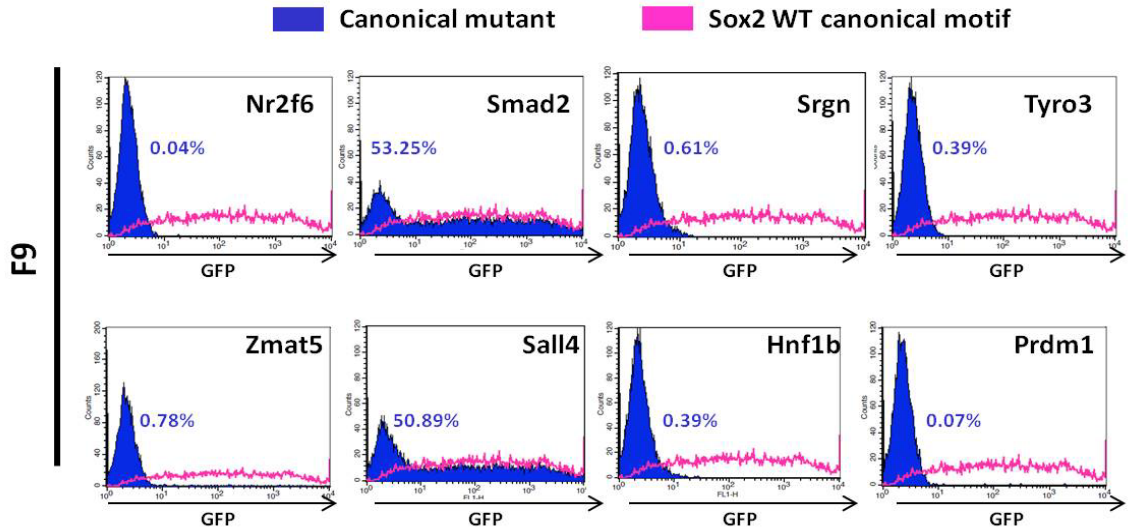
Controls:



WT and Conical mutant motifs in differentiated F9+RA cells:



Canonical mutants in undifferentiated F9 cells:



Quantification of the GFP reporter assay by flow cytometry in F9 cells or F9+RA cells using wild-type compressed motif (WT motif) orcompressed motifs mutated into canonical motifs (Canonical mutants) (related to Figure 8C).

Supplementary Experimental Procedure:

Computational analysis:

ChIP-Seq Data Analysis

ChIP-Seq reads were aligned using the Eland software to NCBI build 37 (mm9) of the mouse genome. The MACS program (version 1.4) (Zhang et al., 2008) was used to detect peaks of ChIP enrichment. It was run with the default settings using a tag size of 35 and band width 280. For KH2 cells MACS was run by comparison to the Input DNA control, it was run separately for each ChIP replica and only those peaks were kept that were found in both replicas. The intersection was done by running windowBed (Quinlan and Hall, 2010) with window size of 50. Intersecting peaks were additionally filtered by retaining only peaks with a p-value $< 10^{-10}$. This resulted in eight stringent lists of high-confidence ChIP-seq peaks that were used for further analysis. Peaks from F9 cells were called without using controls by applying a stringent p-value $< 10^{-10}$ cut-off.

De novo motif discovery

The findMotifsGenome.pl program from software suite HOMER (Heinz et al., 2010) was used to find motifs de novo in intersected (using windowBed (http://code.google.com/p/bedtools/) with window size 50) Sox/Oct peak lists in the same cell lines. The program was run with mm9

genome, motif length 12, 14, 16, region size 50, 3 permitted mismatches allowed in global optimization phase and 5 motifs were requested to find.

Finding instance of specific motifs

The findMotifsGenome.pl with the "-find" option and the annotatePeaks.pl (another program from HOMER software suite) with the "-m" option were both run to find genomic positions of motifs within intersected Sox/Oct peak lists in the same cell lines. The non-redundant information from the outputs of those two programs was than combined into one well-annotated BED file (Supplementary Table 3). The canonical and compressed motif files used as inputs to findMotifsGenome.pl and annotatePeaks.pl were the files obtained earlier from de novo motif discovery in intersected Sox2/Oct4 in Sox2 OE KH2 and Sox17/Oct4 in Sox17 OE KH2, respectively. The motif score in both motif files was set to 7.5 and the fragment size to use for motif finding was set to 50 before the programs were run.

Venn diagrams

Venn diagrams, each showing a Sox protein intersecting with Oct4 in two different cell environments, were created in Matlab. Intersections were performed for Oct4 peak lists of equal size – in order to get such lists, Oct4 peaks were sorted according to their p-values and peaks with the highest p-values were rejected. The intersections were performed by using windowBed on peak summits with window size 50.

Pie charts

The pie charts represent the fraction of total peaks that intersect with compressed, canonical, or both compressed and canonical (‘overlapping’) motifs. The intersections were performed by using windowBed on peak and motif summits with window size 50.

Histograms

Histograms showing motif counts dependent on the distance of the motifs to closest peak summit were done based on the values obtained from the software findMotifsGenome.pl. It was run with the fragment size to use for motif finding parameter set to 100.

Box plot comparison

The peak height distributions were plotted in box-plots and comparison intervals were drawn using notches. Two medians are significantly different at the 5% significance level if their intervals do not overlap. Interval endpoints are the extremes of the notches. Outliers are shown as red crosses. Logarithmic scale was applied for the box plots.

Pileup analysis

Selected enriched regions were aligned with each other according to the position of either canonical or compressed motifs using glbase (https://bitbucket.org/oaxiom/glbase/wiki/Home). Each read was previously extended 245bp towards the interior of the sequenced fragment, based on the strand of the alignment. Read density profiles for each experiment were normalized to the read density per million total reads around the center of the region. These regions were sorted in descending order according to the motif score. Heatmaps were generated using Matlab software.

Evolutionary analysis

In evolutionary analysis, we identified the motif regions in mouse genome that overlap either the 30-Way PhastCons or 30-Way PhyloP Elements of either 11 euarchontoglires or 20 placental mammals tracks from the UCSC Genome Browser using centered windows extended either +/-15 or +/-1000 bp from the motif.

Quantitative real-time PCR (Q-RT-PCR).

Total RNA was extracted in Trizol (Invitrogen), and purified using the RNeasy Mini Kit with DNaseI treatment (Qiagen) to remove contaminating genomic DNA. Synthesis of cDNA was performed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). qRT-PCR was performed using specific primers (listed in Supplemental table 5) and SYBR Green mix on the ABI 7900HT Real-Time PCR System.

Western Blotting

Protein extracts from doxycycline treated and non-treated Sox-V5 ES cells were prepared using RIPA lysis buffer (Sigma) supplemented with protease inhibitors (Roche). Western blot analysis was performed using an anti-V5 antibody (Invitrogen), an anti-Sox2 (Santa Cruz), an anti-Oct4 (Santa Cruz) and an anti-Sox17 (R&D systems).

Immunostaining

B6CBAF1 mice used were bred and raised at the Biological Resource Center in Singapore. Prior to fixation, the zona pellucida was briefly treated with Acid Tyrode's solution (Sigma, T1788). Embryos were washed and then fixed with 2.5% paraformadehyde in phosphate-buffered saline (PBS) for 15 min at room temperature. Fixed embryos were permeabilized in 0.5% Triton X-100 and blocked with 0.1% BSA in PBS at room temperature. The embryos were stained one hour at room temperature with the primary antibodies (anti-Oct4 SC-5279, Santa Cruz Biotechnology, 1:100, anti-Sox2 AB5603, Chemicon 1:100 and anti-Sox17, R&D Sytems, AF1924 1:100).

Embryos were then incubated with Alexa-488 secondary antibody (Molecular Probes, 1:300) for one hour at room temperature. To visualize nuclei, embryos were stained with TOTO-3 (Molecular Probes, Carlsbad, 1:400). Images were captured with a confocal microscope (LSM 510 META; Zeiss).

Chromatin Immunoprecipitation (ChIP) assay

Chromatin Immunoprecipitation experiments were performed on 2 independant Sox2/Sox2KE/Sox17/Sox17EK expressing KH2 ES clones and one replicate for F9 cells using protocols from (Yu et al, 2009). Briefly, cells were cross-linked with 1% formaldehyde for 10 min at room temperature, and formaldehyde was inactivated by the addition of 125 mM glycine. Chromatin from each cell line was sonicated to a length of around 250-500bp and immunoprecipitated using the following antibodies: V5 antibody (Invitrogen, R96025), Oct4 antibody (Santa Cruz, SC8628) and Sox17 antibody (R&D systems, AF1924) and sequenced using the SOLEXA platform. ChIP-seq computational analysis procedures are detailed in Supplementary method. The data may be obtained via Gene Expression Omnibus (GSE43275).

RNA interference

siRNAs targeting *Oct4, Nanog* and non-targeting control were purchased from Dhamarcon as siGenome-Smart Pool. F9 cells were seeded at $5 * 10^4$ cells per well of a 6-well plate. 200 pmol of siRNAs was complexed with 10 µl of Lipofectamine RNAiMAX reagent. Re-transfection was performed once more 72 h after the initial transfection. Cells were harvested from d3 to d6.

Protein immunoprecipitation and Western blots

F9 cells treated with RA were first incubated with 2 mM of the cross-linking agent DSP (Pierce Biotechnology) for 2 hr at 4°C and then lysed in RIPA buffer (50mM Tris pH 7.5, 150mM NaCl, 5mM EDTA and 0,2-0,5% NP40) supplemented with EDTA-free pellet protease inhibitor (Roche Diagnostics) and 1mM of PMSF. The lysates were then spun and supernatants were used for immunoprecipitation. Lysates were incubated with 5µg of Oct4 antibody (Santa Cruz Biotechnology SC-8628) overnight at 4ºC and then mixed with 25µl of protein G beads (Pierce Crosslink co-IP kit #88805). Protein/Antibody/Bead complexes were finally washed three times with the lysis buffer and kept in protein loading buffer for western-blot. Sox2 and Sox17 western blots were performed using the anti-Sox2 from Santa Cruz Biotechnology (SC17201) and the anti-Sox17 from R&D systems (AF1924).

Supplementary Table 5:

Primers for RT-PCR

mOct4-F	TCTTTCCACCAGGCCCCCGGCTC
mOct4-R	TGCGGGCGGACATGGGGAGATCC
mNanog-F	AGGGTCTGCTACTGAGATGCTCTG
mNanog-R	CAACCACTGGTTTTTCTGCCACCG
mGata4-F	TCTCTGCATGTCCCATACCA
mGata4-R	GTGTGAAGGGGTGAAAAGGA
mGata6-F	GTGGAGATGGCAGCAAAAAT
mGata6-R	GAAGCACATGATTTGGAGCA
mLama1-F	CAACTGCTCGCAGAATACCA
mLama1-R	GGCAGCGGTCACATTTATCT
mFoxA2-F	GACCTCAAGGCCTACGAACA
mFoxA2-R	CCACTCAGCCTCTCATTTCC
mCol4a2-F	CCACGTTGGTGCTTATCCTT
mCol4a2-R	GAATGAGTGCTGGTGCAGAA
mCxcr4-F	GGGTCATCAAGCAAGGATGT
mCxcr4-R	CTGGAGTGAAAACTGGAGGA
mActin-F	TGAAACAACATACAATTCCATCATGAAGTGTGA
mActin-R	AGGAGCGATAATCTTGATCTTCATGGTGCT

Primers for the reporter assays

Hnf1b-SalI-F	ACGCGTCGACGTCGGCCATAGCGGCCGCGGAACCTGTCTGGGGCACTTAAAA
Hnf1b-EcoRV-R	TGTCTGTC gatatcGCTGCCTCAGGTGATAGGAG
Smad2-SalI-F	ACGCGTCGACGTCGGCCATAGCGGCCGCGGAACTGTGGCTAGTCACCCCACT

Smad2-BamHI-R	CGCGGATCCGCGAGAGGTCAGGGACAATGGTG
Sall4-SalI-F	ACGCGTCGACGTCGGCCATAGCGGCCGCGGAACAGAGAAATGGGGAGATGGA
Sall4-BamHI-R	CGCGGATCCGCGCCAGGTGGACTCTCCAGGTA
Zmat5-SalI-F	ACGCGTCGACGTCGGCCATAGCGGCCGCGGAAATGGTTAGGACAGGCACAGG
Zmat5-BamHI-R	CGCGGATCCGCGAGGAAGGAAACGACAGCAGA
Nr2f6-SalI-F	ACGCGTCGACGTCGGCCATAGCGGCCGCGGAAGTCCCCCTCTCTCCTTTCAG
Nr2f6-BamHI-R	CGCGGATCCGCGCGCGTTCTCTGGTGTGTCTA
Srgn-SalI-F	ACGCGTCGACGTCGGCCATAGCGGCCGCGGAAGCCCATACACGTCCTCTGTT
Srgn-BamHI-R	CGCGGATCCGCGTGGTACCACTGCCTTTGTGA
Tyro3-SalI-F	ACGCGTCGACGTCGGCCATAGCGGCCGCGGAACACAGGGGGTGTTAGCAGTT
Tyro3-BamHI-R	CGCGGATCCGCGGGTGACAGGAAGGAGAGCAG
Prdm1-SalI-F	ACGCGTCGACGTCGGCCATAGCGGCCGCGGAACTGCCGCAGACTTCTTTACC
Prdm1-EcoRV-R	TGTCTGTC gatatcGCAGGGACAACAAAACCAGT

Primers for mutagenesis

The motif sequence is shaded in gray and the mutated bases are underlined.

The core wild-type motif related to the sequence CATTGTATGCAAT is giving as reference.

WT Smad2:

TATTGTATGCAAAT

>Smad2_tocan

5'-gttccagtgaaaggag**tattgtcatgcaaat**agttaagccaaga-3'

5'-tcttggcttaactatttgcatgacaatactcctttcactggaac-3'

>Smad2_mut

5'-gagtcatttgggttccagtgaaaggagtagggtacataaatagttaagccaagaatacagactggg-3'

5'-cccagtctgtattcttggcttaactatttatgtaccctactcctttcactggaacccaaatgactc-3'

WT Nr2f6:

CTTTGTATGCAGGA

>Nr2f6_tocan

5'-attaagcacccttttgtcatgcaggaccccccag-3'

5'-ctgggggtcctgcatgacaaagggtgcttaat-3'

>Nr2f6_mut

5'-tctactggagcaagcatttattaagcaccctgggtacataggacccccagggca-3'

5'-tgccctgggggtcctatgtacccagggtgcttaataaatgcttgctccagtaga-3'

WT Srgn

TATTGTATGTAAAA

>Srgn_tocan

5'-gaaggctcgagagtctttttacatgacaataggactgtcttc-3'

5'-gaagacagtcctattgtcatgtaaaaagactctcgagccttc-3'

>Srgn_mut

5'-tggagtagagaaggctcgagagtctttttatgtaccctaggactgtcttccgg-3'

5'-ccggaagacagtcctagggtacataaaaagactctcgagccttctctactcca-3’

WT Tyro3

CTTTGTATGCAAAT

>Tyro3_tocan

5'-cctgtggccgctttgtcatgcaaatcgggaagc-3'

5'-gcttcccgatttgcatgacaaagcggccacagg-3'

>Tyro3_mut

5'-gatagccctgtggccgctgggtacacaaatcgggaagcctagc-3'

5'-gctaggcttcccgatttgtgtacccagcggccacagggctatc-3'

WT Zmat

CATTGTATGTAAAA

>Zmat_tocan

5'-ccatgtgagcctaacttcattgtcatgtaaaacagacataacaatag-3'

5'-ctattgttatgtctgttttacatgacaatgaagttaggctcacatgg-3'

>Zmat5_mut

5'-gttactgaccatgtgagcctaacttcagggtacataaaacagacataacaataggggaagg-3'

5'-ccttcccctattgttatgtctgttttatgtaccctgaagttaggctcacatggtcagtaac-3'

WT Sall4

CACTGTATGCTAAG

>Sall4_tocan

5'-tcagctgcctctaacttagcatgacagtggctgctt-3'

5'-aagcagccactgtcatgctaagttagaggcagctga-3'

>Sall4_mut

5'-tcagctgcctctaacttagagtaccgtggctgcttggttctgg-3'

5'-ccagaaccaagcagccacggtactctaagttagaggcagctga-3'

WT Hnf1b

CTTTGTATGCAGAC

>Hnf1b_tocan

5'-ccagcccaggtctgcatgacaaagacactgg-3'

5'-ccagtgtctttgtcatgcagacctgggctgg-3'

>Hnf1b_mut

5'-ggccagcccaggtctgtctacccagacactgggcagggc-3'

5'-gccctgcccagtgtctgggtagacagacctgggctggcc-3'

WT Prdm1 WT

GATTGTATTCTAAG

>Prdm1_tocan

5'-acccacactacgattgtcattctaagtcctccgga-3'

5'-tccggaggacttagaatgacaatcgtagtgtgggt-3'

>Prdm1_mut

5'-ccctccatgtttacccacactacgagggtcctctaagtcctccggatcg-3'

5'-cgatccggaggacttagaggaccctcgtagtgtgggtaaacatggaggg-3'

Pdgfra WT

TTTTGTGTGCAGAT

>Pdgfra_tocan

5'-cctgctttgatctgcacgacaaaagaagtgacctg-3'

5'-caggtcacttcttttgtcgtgcagatcaaagcagg-3'

>Pdgfra_mut

5'-tctgaacgaagcccctgctttgatctatgcacccaagaagtgacctgaaataacatggcg-3'

5'-cgccatgttatttcaggtcacttcttgggtgcatagatcaaagcaggggcttcgttcaga-3'

References

Hutchins AP, Diez D, Takahashi Y, Ahmad S, Jauch R, Tremblay ML, Miranda-Saavedra D (epub, Januray 7 2013) Distinct transcriptional regulatory modules underlie STAT3's cell type-independent and cell type-specific functions. Nucl. Acids Res.

Niakan KK, Ji H, Maehr R, Vokes SA, Rodolfa KT, Sherwood RI, Yamaki M, Dimos JT, Chen AE, Melton DA, McMahon AP, Eggan K (2010) Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes Dev* 24(3): 312-326

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W*., et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.