

**Supplemental Information: A *Sox2* distal enhancer cluster regulates embryonic stem cell differentiation potential.**

Harry Y Zhou<sup>1\*</sup>, Yulia Katsman<sup>1\*</sup>, Navroop K Dhaliwal<sup>1</sup>, Scott Davidson<sup>1</sup>, Neil N Macpherson<sup>1</sup>, Moorthy Sakthidevi<sup>1</sup>, Felicia Collura<sup>1</sup> and Jennifer A. Mitchell<sup>1,2</sup>

1) Department of Cell and Systems Biology, University of Toronto, Toronto, ON, M5S 3G5, Canada.

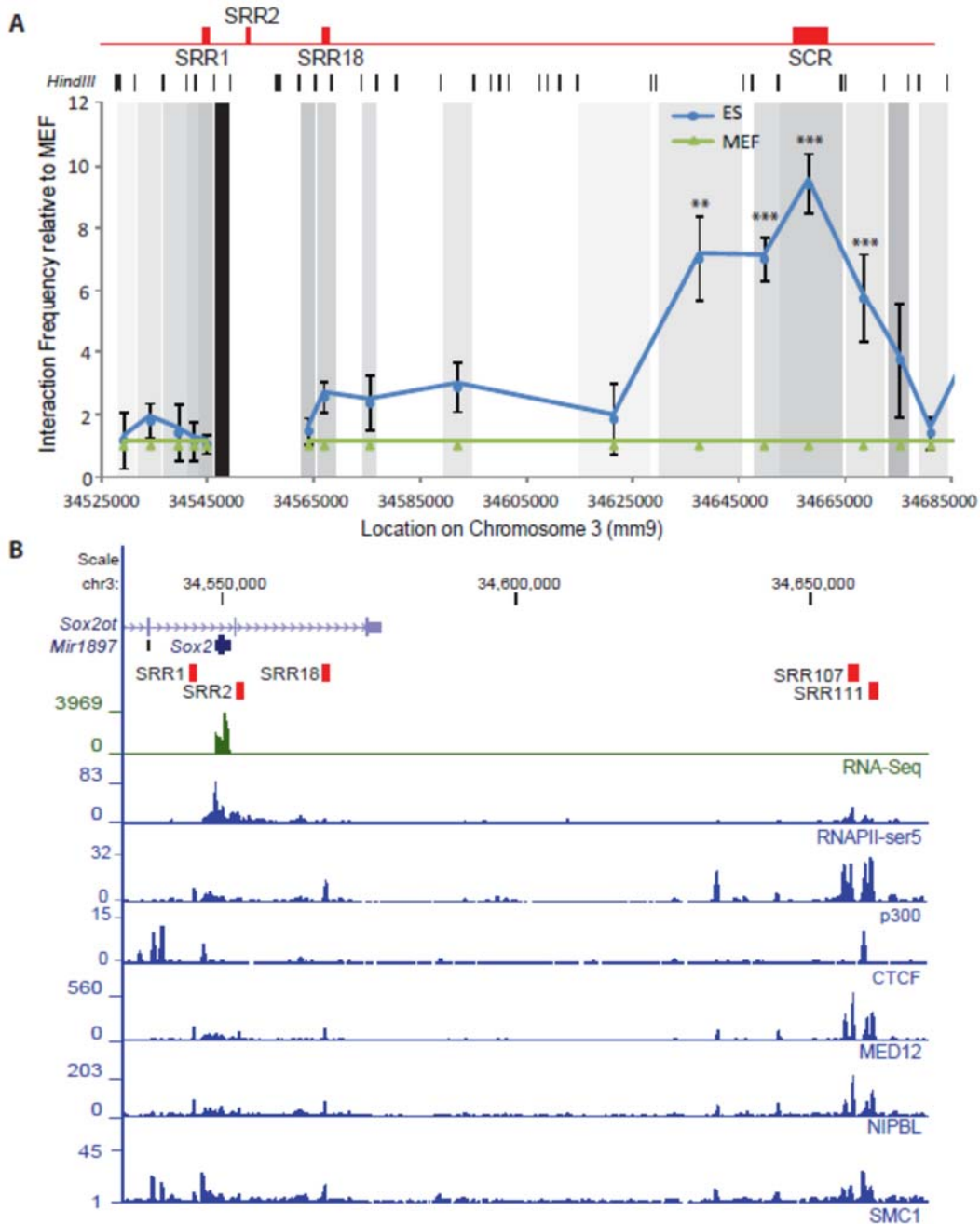
2) Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, ON, M5S 3G5, Canada.

\* These authors contributed equally.

**Contact:** Jennifer A Mitchell, email: [ja.mitchell@utoronto.ca](mailto:ja.mitchell@utoronto.ca). Department of Cell and Systems Biology, University of Toronto, 25 Harbord Street, Toronto, Ontario M5S 3G5, Canada. Phone: (416)978-6711, Fax: (416)978-8532.

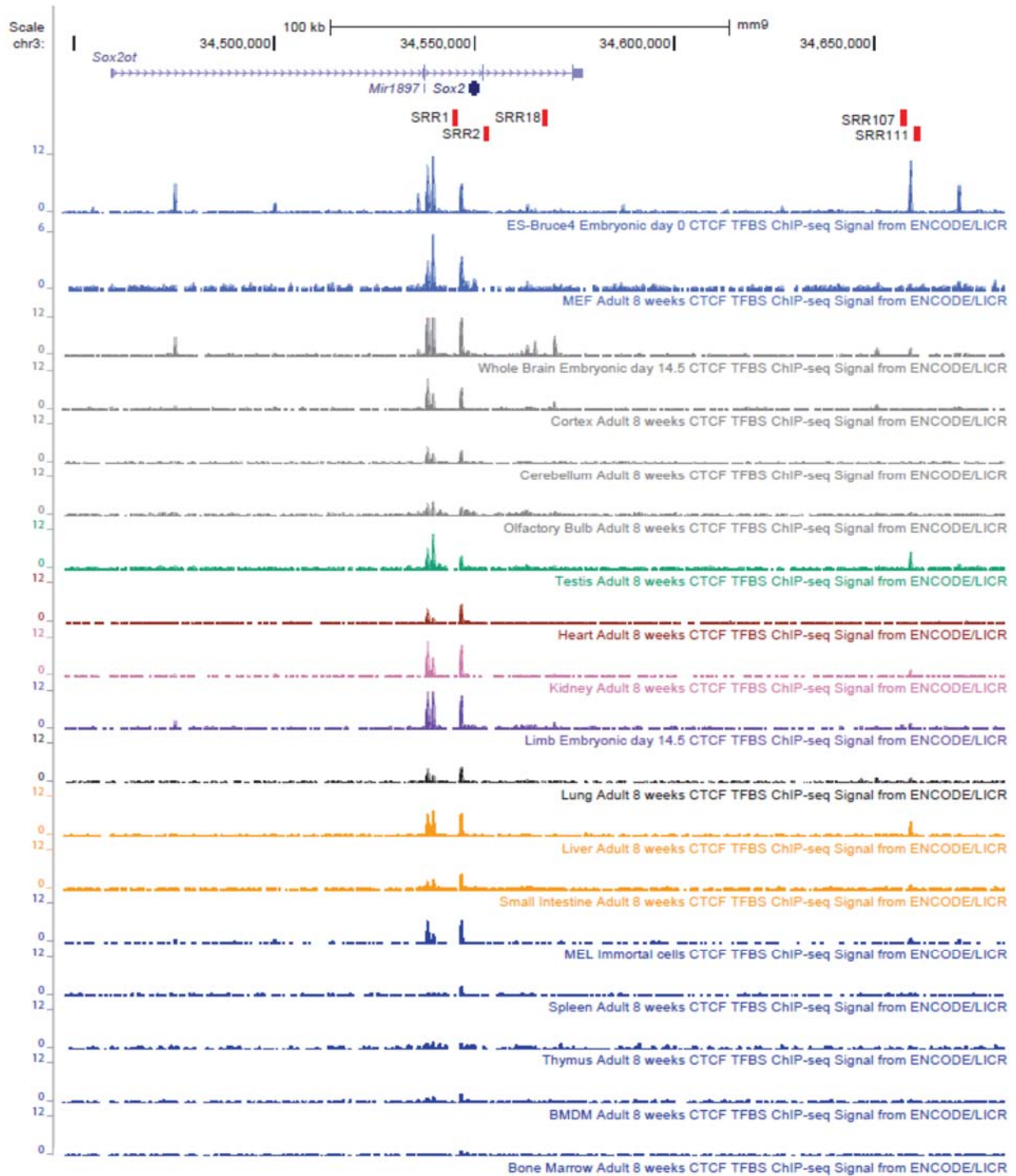
**Table S1: Transcription factors and other chromatin associated proteins bound at the Sox2 regulatory regions.** Light green indicates binding within the gene proximal and SCR enhancers. Dark green indicates binding only within the SCR enhancers.

ChIP-Seq	Enhancers					Function
	SRR1	SRR2	SRR18	SRR107	SRR111	
OCT4	Light Green	Light Green	Light Green	Light Green	Light Green	Transcription Factors
SOX2	Light Green	Light Green	Light Green	Light Green	Light Green	
NANOG	Light Green	Light Green	Light Green	Light Green	Light Green	
SMAD1	Light Green	Light Green	Light Green	Light Green	Light Green	
TCF3	Light Green	Light Green	Light Green	Light Green	Light Green	
TCFCP2L1	Light Green	Light Green	Light Green	Light Green	Light Green	
ESRRB	Light Green	Light Green	Light Green	Light Green	Light Green	
KLF4	Light Green	Light Green	Light Green	Dark Green	Dark Green	
NR5A2	Light Green	Light Green	Light Green	Dark Green	Dark Green	
STAT3	Light Green	Light Green	Light Green	Dark Green	Light Green	
E2F1	Light Green	Light Green	Light Green	Dark Green	Light Green	
EP300	Light Green	Light Green	Light Green	Light Green	Light Green	
CREBBP	Light Green	Light Green	Light Green	Light Green	Light Green	
NCOA3	Light Green	Light Green	Light Green	Dark Green	Dark Green	
STAG2	Light Green	Light Green	Light Green	Dark Green	Light Green	Cohesin Associated
NIPBL	Light Green	Light Green	Light Green	Light Green	Light Green	
RNAPII	Light Green	Light Green	Light Green	Light Green	Light Green	RNAPII Associated
TAF3	Light Green	Light Green	Light Green	Light Green	Light Green	
CDK8	Light Green	Light Green	Light Green	Light Green	Light Green	
CDK9	Light Green	Light Green	Light Green	Light Green	Light Green	
Aff4	Light Green	Light Green	Light Green	Light Green	Light Green	
BRD4	Light Green	Light Green	Light Green	Light Green	Light Green	
ELL2	Light Green	Light Green	Light Green	Light Green	Light Green	
ELL3	Light Green	Light Green	Light Green	Light Green	Light Green	
SUPT5	Light Green	Light Green	Light Green	Dark Green	Light Green	
MED1	Light Green	Light Green	Light Green	Light Green	Light Green	Mediator Complex
MED12	Light Green	Light Green	Light Green	Light Green	Light Green	
PRDM14	Light Green	Light Green	Light Green	Light Green	Light Green	Chromatin Repressive Complex
KDM1A	Light Green	Light Green	Light Green	Light Green	Light Green	
CHD4	Light Green	Light Green	Light Green	Light Green	Light Green	
HDAC2	Light Green	Light Green	Light Green	Light Green	Light Green	

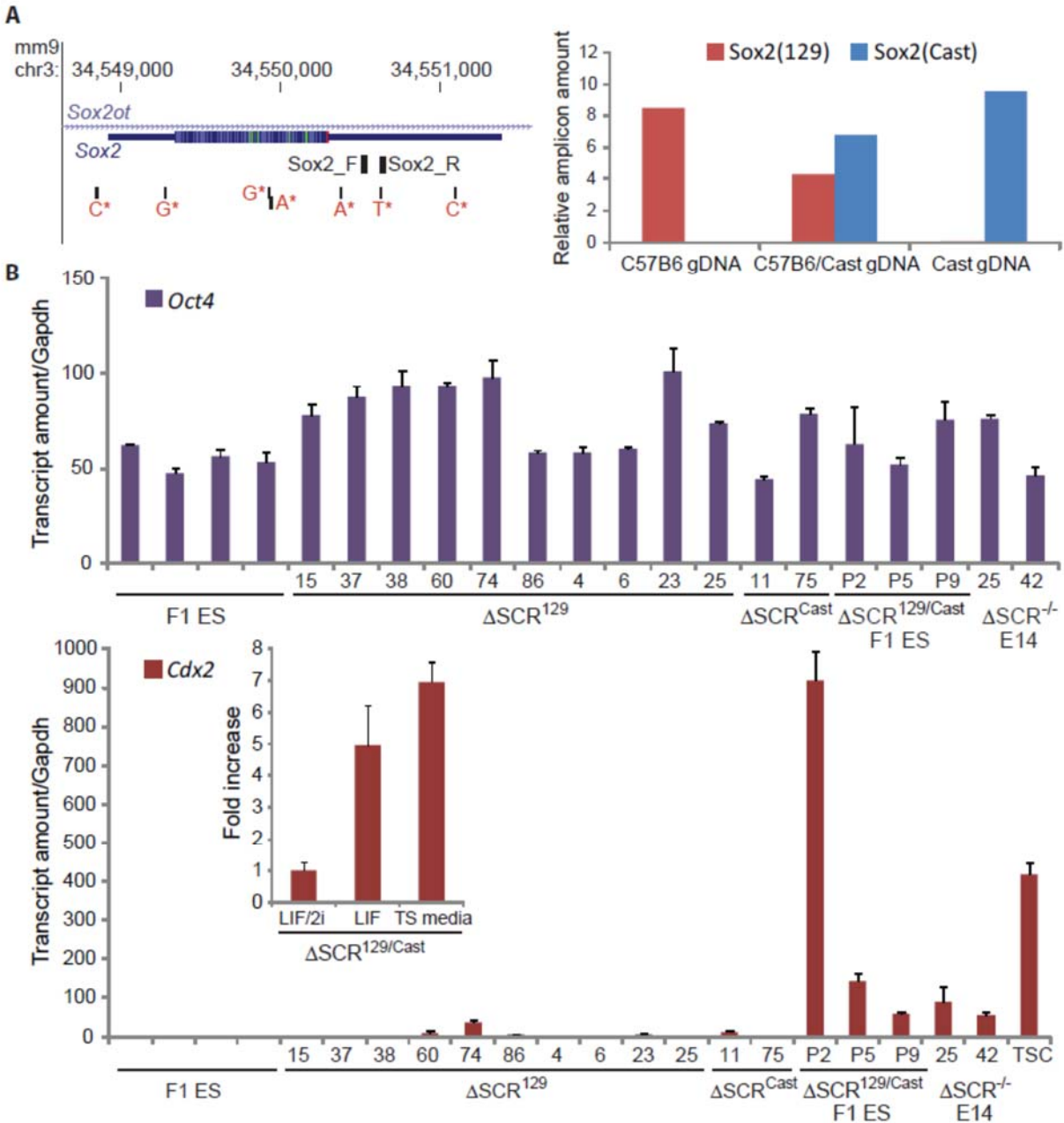


**Figure S1 Related to Figure 2: Chromatin interactions and Chromatin associated proteins bound at the *Sox2* locus in ES cells.**

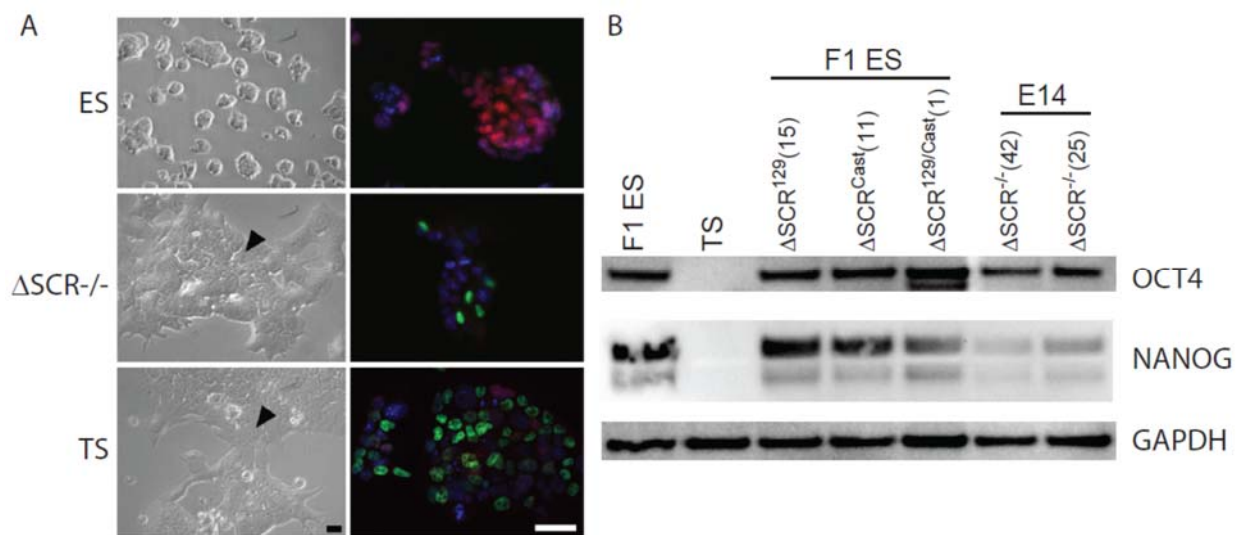
A) Chromosome conformation capture (3C) was performed on mouse ES cells and MEFs. The frequency of interaction between the *Sox2* promoter (anchor) and the surrounding chromatin was normalized to that between adjacent fragments at the *Alpha aortic actin* ( $\alpha$ -Act) locus. Black bar represents the anchor fragment and shaded bars represent the interacting fragments. Values are an average of three independent experiments, with each experiment performed in triplicate. Interaction data for ES cells is represented relative to MEF interactions. Error bars represent standard deviation. A significant difference between ES cells and MEFs is indicated by \* (p < 0.05), \*\* (p < 0.01) or \*\*\* (p < 0.001). B) *Sox2* regulatory regions (SRR) are shown as red bars. RNA-Seq data for mouse ES cells is shown in green. The ENCODE ChIP-seq data for p300, and CTCF in mouse ES cells are displayed on the UCSC browser. Also shown are, RNA polymerase II phospho-S5 (RNAPII-ser5), Mediator 12 (MED12), the cohesin complex component SMC1, and the cohesin complex loading factor NIPBL ChIP-Seq data for mouse ES cells.



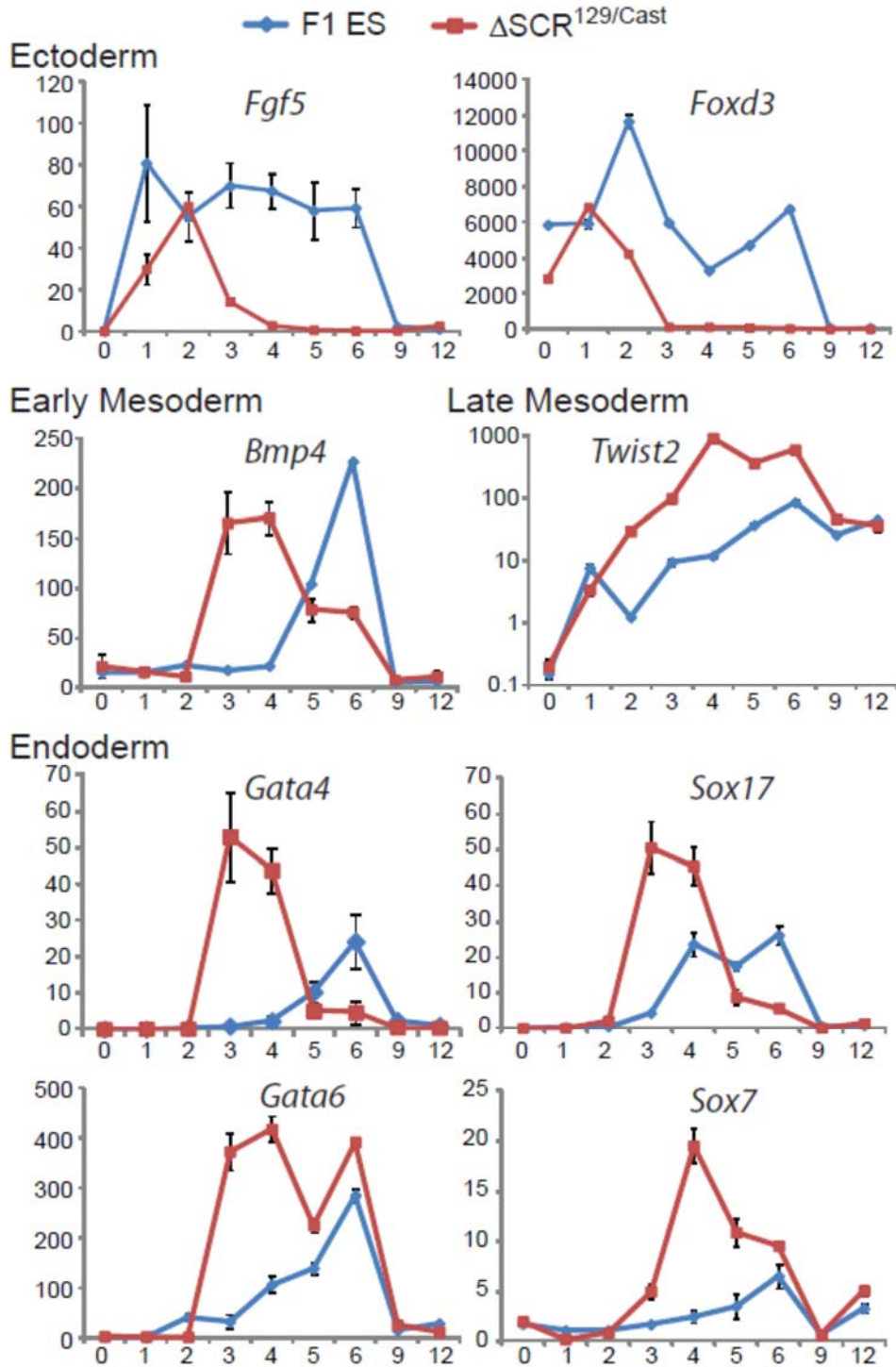
**Figure S2 Related to Figure 2: CTCF bound regions surrounding *Sox2* in various cell types.** *Sox2* regulatory regions (SRR) are shown as red bars. The available ENCODE/LICR ChIP-seq data for CTCF in various cell types are displayed on the UCSC browser.



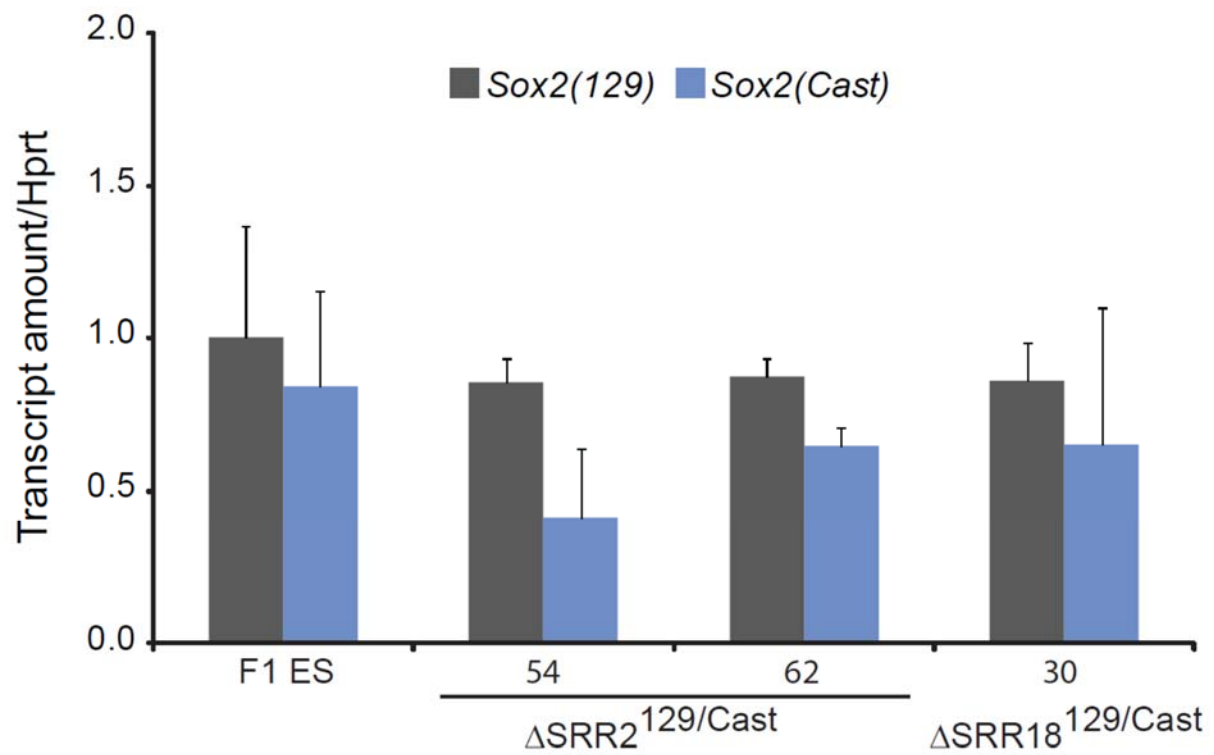
**Figure S3 Related to Figure 3: Allele specific primer validation and gene expression profiling of SCR deleted clones by real-time quantitative PCR.** A) Allele specific primers were designed and validated by qPCR. Left panel shows the location of the primers and the Cast SNP located at the 3' end of the reverse primer. Right panel shows the relative amplicon amount for 129 and Cast specific primers tested on C57B6 (same genotype as 129 at this SNP) and Cast genomic DNA as well as a 50:50 mixture of C57B6 and Cast genomic DNA. B) Transcript amount relative to *Gapdh* is shown for *Oct4*, and *Cdx2*. Levels of *Cdx2* were normalized to F1 ES cell levels. *Cdx2* inset graph shows the fold increase in *Cdx2* expression when ΔSCR<sup>129/Cast</sup> (clone 1) was maintained in ES media with LIF/2i, LIF or TS media. Clone number indicated above genotype. ΔSCR<sup>129/Cast</sup> (clone 1) RNA was obtained at passage 2, 5 and 9 indicated by P2, P5 and P9. For other clones, RNA was isolated from cells at passage 2 after colony selection. Error bars represent standard deviation.



**Figure S4 Related to Figure 4: SCR deletion in E14 ES cells causes differentiation to a trophectoderm cell-like phenotype.** A) Left panel displays bright-field images of F1 ES and E14  $\Delta SCR^{-/-}$  cells (clone 42), which showed features consistent with differentiation to TS-like cells, and F1 TS cells. Arrow heads indicate similar features in  $\Delta SCR^{-/-}$  (clone 42) and TS cells. Scale bar, 30  $\mu m$ . Right panel displays SOX2 and CDX2 immunofluorescence in F1 ES, E14  $\Delta SCR^{-/-}$  (clone 42) and F1 TS cells. Scale bar, 10  $\mu m$ . B) OCT4 and NANOG protein levels in F1 ES cells, TS cells,  $\Delta SCR^{129}$  (clone 15),  $\Delta SCR^{Cast}$  (clone 11),  $\Delta SCR^{129/Cast}$  (clone 1),  $\Delta SCR^{-/-}$  in E14 ES cells (clones 25 and 42). GAPDH levels reveal relatively equal protein loading in all samples.

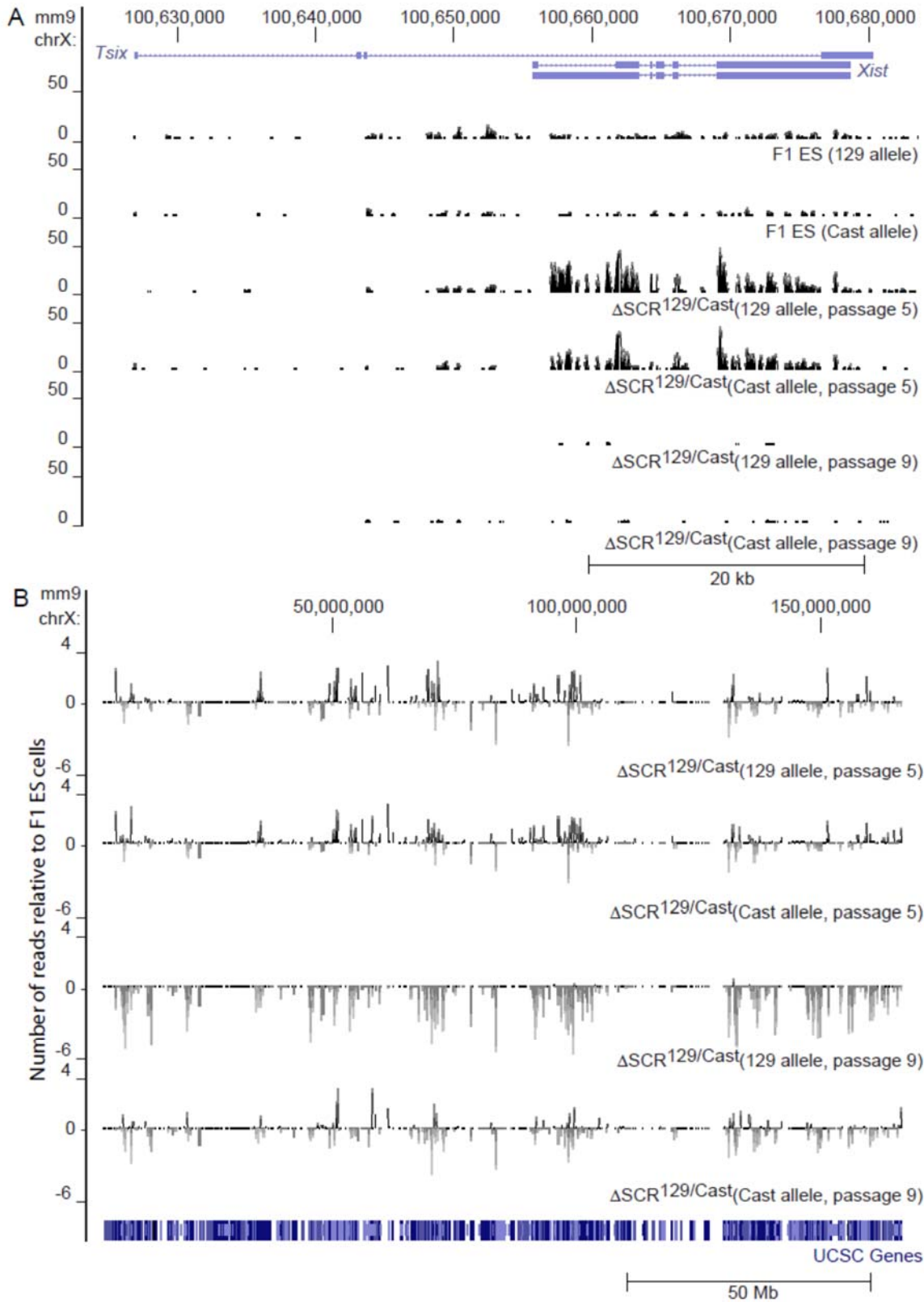


**Figure S5 Related to Figure 5: Embryoid body formation reveals impaired differentiation to neuroectoderm and increased mesoderm formation in  $SCR^{129/Cast}$  cells.** Embryoid body formation was induced by the hanging drop method and changes in gene expression were monitored over 12 days by RT-qPCR. Y axis shows transcript levels relative to *Gapdh*. *Twist2* is displayed on a  $\log_{10}$  scale. X axis shows days of embryoid body formation. Error bars represent standard deviation.



**Figure S6 Related to Figure 6: Homozygous deletion of SRR2 or SRR18 does not affect Sox2 transcription.** Allele specific RT-qPCR revealed that SRR2 and SRR18 are all dispensable for Sox2 expression in ES cells. Sox2 mRNA levels shown are relative to the levels in F1 ES cells. Clone number is shown below the graph. Error bars represent standard deviation.





**Figure S7 Related to Figure 7: X inactivation in  $\Delta$ SCR<sup>129</sup>/Cast.** A) *Xist* expression is increased on both alleles in  $\Delta$ SCR<sup>129</sup>/Cast at passage 5. B) Decreased expression on the X chromosome in  $\Delta$ SCR<sup>129</sup>/Cast relative to F1 ES cells indicates X inactivation is more prominent on the 129 X chromosome. The number of reads is expressed on a Log2 scale relative to the F1 ES cell read count.