Cell Reports, Volume *40*

Supplemental information

SOX transcription factors direct TCF-independent WNT/β-catenin responsive transcription to govern cell fate in human pluripotent stem cells

Shreyasi Mukherjee, David M. Luedeke, Leslie McCoy, Makiko Iwafuchi, and Aaron M. Zorn

Figure S1 - related to Figure 1

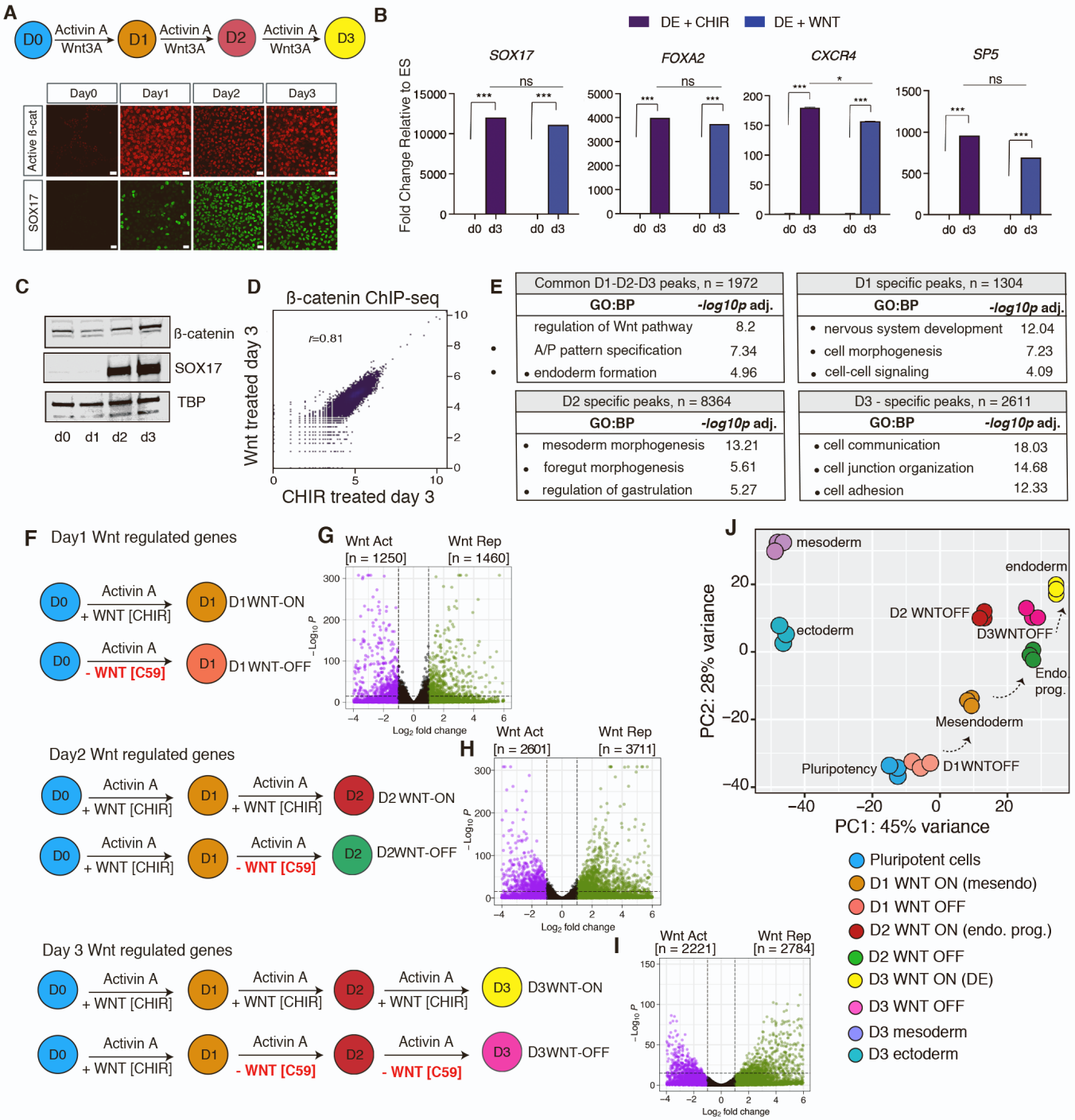
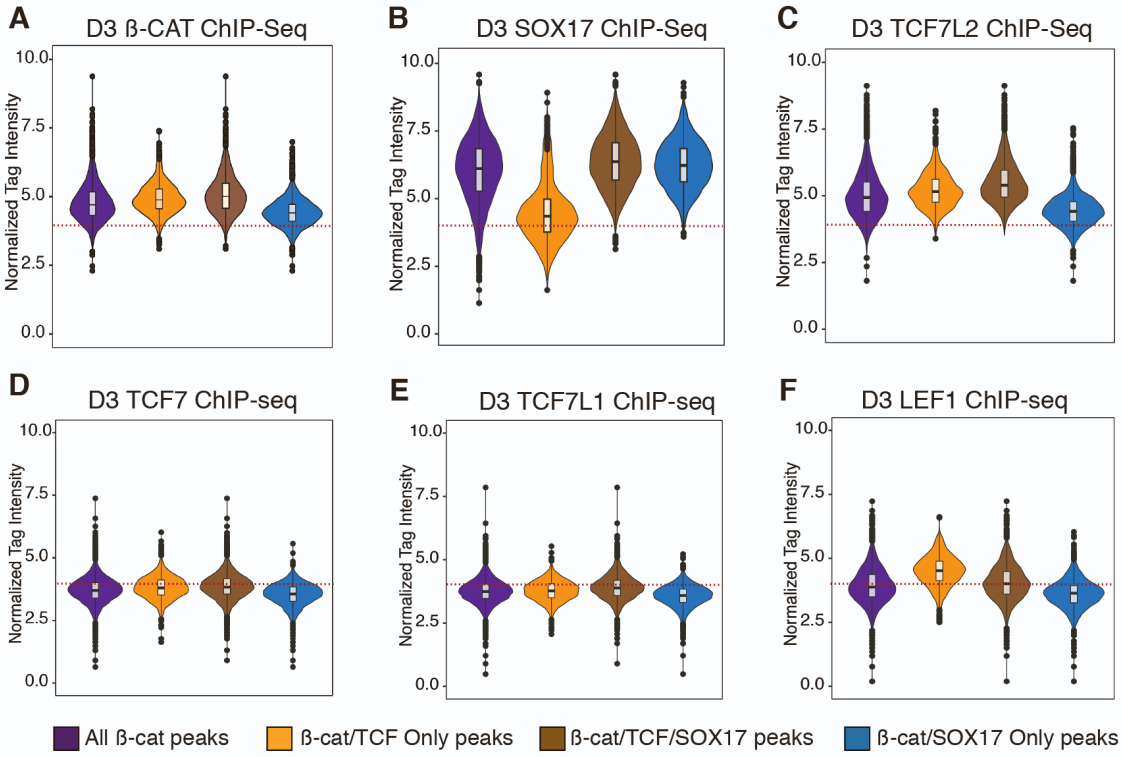


Figure S1 – Related to Figure 1. Characterization of ß-catenin chromatin binding and Wnt-target gene expression during DE differentiation. A. Schematic of the 3-day DE differentiation protocol with ACTIVIN A and WNT3A treatment instead of CHIR. Immunostaining shows that WNT3A stimulates DE differentiation and activates ß-catenin and SOX17 expression in a manner indistinguishable from CHIR in Fig. 1B (scale bar = 100 μm, 3 biological replicates). **B.** RT-qPCR validating that expression levels of DE markers *SOX17, FOXA2, CXCR4* and WNT target gene *SP5* are comparable in day 3 DE differentiated with either CHIR99021 (in purple) or WNT3A (in blue). The difference in expression levels between DE differentiated with CHIR or WNT3A, or between Day0 (d0) and Day3 (d3) was calculated by two tailed student's T-test, ***p<0.001, *p<0.05, ns = not significant (p>0.05); n = 3 **C.** Western blots of nuclear extracts showing ß-catenin, SOX17 and TBP (loading control) protein levels (n = 3). **D.** Scatterplot showing the high degree of similarity between Day 3 BCAT ChIP-Seq datasets generated with CHIR or WNT3A, based on Spearman's rank correlation coefficient (r=0.81). **E.** GO enrichment analysis of four clusters of ß-catenin bound genomic regions: common peaks, Day 1 specific peaks, Day 2 specific peaks and Day 3 specific peaks. For each category, the most enriched GO terms and the adjusted -log10 p-values (Fisher's exact test, FDR 5%) are shown. **F.** Schematic showing the experimental strategy to identify Wnt-target genes at different days of DE differentiation. Cultures were treated with ACTIVIN A and either the WNT-agonist CHIR or the WNT-antagonist C59 at the indicated times followed by RNA-seq analysis in biological triplicate and differential expression analysis. **G – I.** Volcano plots of differentially expressed transcripts in WNT-ON versus WNT-off samples (log2 fold change >1, p-adj<0.05) showing WNT-responsive transcripts at days 1, 2 and 3 of differentiation. **J.** Principal component analysis (PCA) plot showing distribution of WNT-ON and WNT-OFF RNA-seq samples during endoderm differentiation, relative to stage-matched mesoderm and ectoderm differentiation data obtained from Gifford et al, *Cell,* 2013 (GSM1112846, GSM1112844, GSM1112835, GSM1112833). **K – N.** Relative expression levels of direct ß-catenin bound and WNT activated genes plotted as a *loess* smoothed trendline (individual transcript data is shown in light grey) for the following categories: common genes bound and activated by ß-catenin on all days of differentiation **(F)**, Day 1 specific genes **(G)**, Day 2 specific genes **(H)** and Day 3 specific genes **(I)** and GO enrichment analysis of each category (the top 5 enriched terms and associated -log10 p-values (Fisher's exact test, FDR 5%) are showed.

Figure S2 - Related to Figure 2



G

All D3 ß-CATpeaks

Rank	Motif	E-val
1.	GATA	1.7e-1080
2.	TCF	1.5e-559
3.	SOX	4.6e-230

D3 ß-CAT/SOX17 only peaks

Rank	Motif	E-val
1.	GATA	3.1e-421
2.	TBX	6.2e-31
3.	SOX	3.6e-038

D3 ß-CAT/TCF only peaks

Rank	Motif	E-val
1.	TCF	9.1e-179
2.	GATA	3.5e-74
3.	SMAD	4.9e-12

D3 ß-CAT/SOX17/TCF peaks

Rank	Motif	E-val
1.	GATA	6.9e-541
2.	TCF	3.8e-297
3.	FOXA	6.5e-83

Figure S2 – Related to Figure 2. Dynamic chromatin co-occupancy of ß-catenin, SOX17 and TCFs. A – F. Quantification of Day 3 DE ChIP-Seq tag density across 2 biological replicates for **A.** ß-catenin, **B.** SOX17, **C.** TCF7L2, **D.** TCF7, **E.** TCF7L1 and **F.** LEF1 at the following peak categories: All ß-catenin peaks, only ß-catenin/TCF peaks, peaks co-occupied by ß-catenin, SOX17 and at least one TCF and peaks bound by only ß-catenin and SOX17 but not TCFs. Dotted lines represent the approximate tag density corresponding to the peak calling threshold. **G.** *De-novo* DNA-binding motif analyses of the above-described peak categories. The top 3 enriched motifs and their associated E-values are shown. E-values were calculated through DREME (Bailey., 2011), representing the likelihood of finding similar DNA-binding motifs by chance if the input sequences were shuffled. The enrichment for GATA and FOX TFs is expected for most DE enhancers as GATA4-6 and FOXA are known endoderm regulators.

Figure S3 - Related to Figure 3

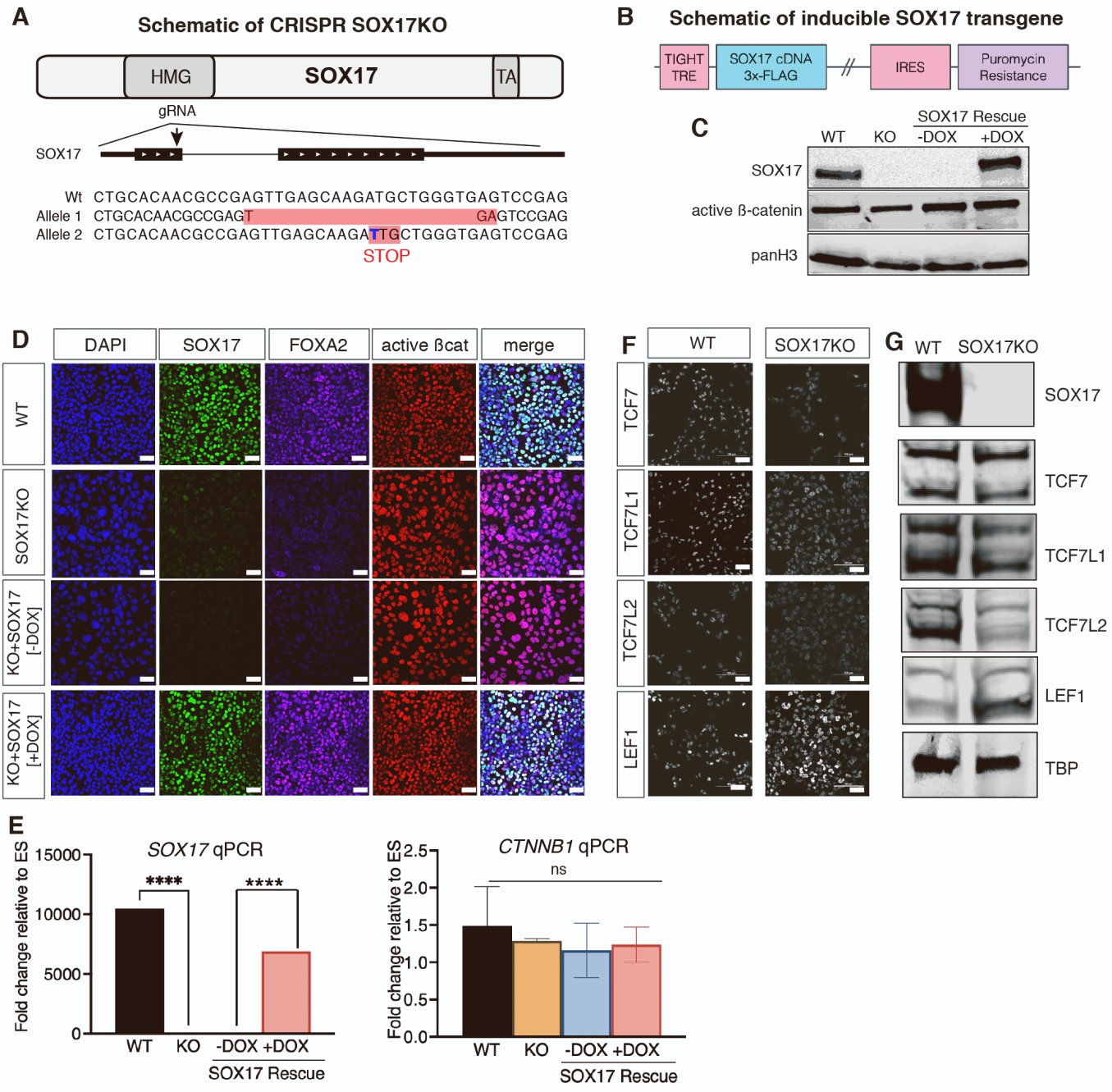
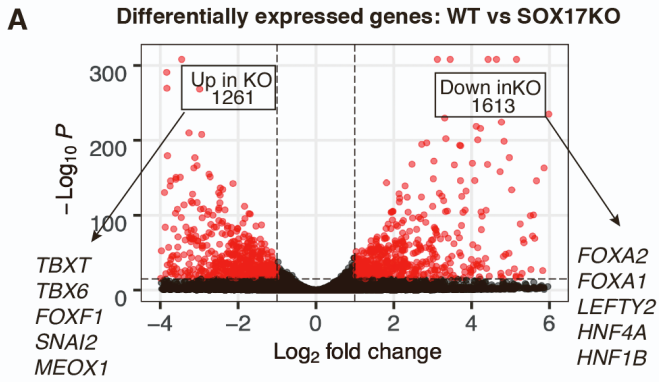
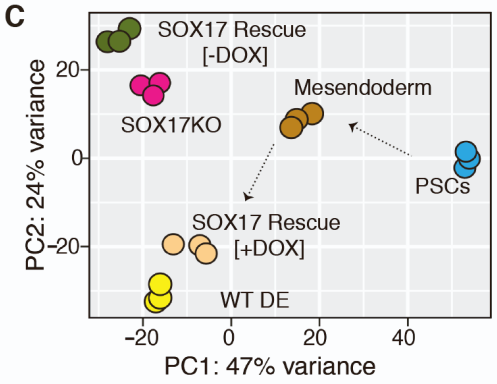


Figure S3 – Related to Figure 3. Characterization of *SOX17KO* and the inducible *SOX17* Rescue cell lines. **A.** Schematic showing the CRISPR-Cas9 targeting strategy to generate the homozygous indel mutant SOX17 knockout (KO) line with stop codons introduced into the DNA-binding HMG domain of the 1st intron. (TA= transactivation domain. **B.** Schematic of the doxycycline (DOX)-inducible SOX17-3xFLAG transgene introduced into the *SOX17KO* cell line for rescue experiments. **C.** Western blots of confirming that the *SOX17KO* line has no detectable SOX17 protein and that DOX inducible SOX17-3xFLAG transgene rescues SOX17 expression to wildtype levels. Furthermore, the expression of active ß-catenin protein does not change in *WT*, *SOX17KO* or *KO+SOX17* rescue conditions (n = 3). **D.** Immunostaining of SOX17, FOXA2 and active ß-catenin in *WT*, *SOX17KO* and KO+*SOX17* rescue cells with or without DOX induction (scale bar = 100 μm), showing that the SOX17KO cell line fails to differentiate into SOX17+/FOXA2+ DE, and that DOX-induced SOX17 transgene rescues DE differentiation (n = 3). **E** RT-qPCR of relative *SOX17* and *CTNNB1* expression levels in *WT*, *SOX17KO* and *KO+SOX17* rescue cells +/-DOX at Day 3 of differentiation. mRNA levels were normalized to the housekeeping gene *PPIA,* and then normalized to expression levels in PSCs. Significance based on student's t-Test, ****p<0.001, n = 3 **F.** Immunostaining and **G.** western blots showing protein expression levels of TCF7, TCF7L1, TCF7L2 and LEF1 in *WT* and *SOX17KO* cells (scalebar = 100 μm). There is a modest increase of LEF1 levels *SOX17KO* cells.

Figure S4 - Related to Figure 3





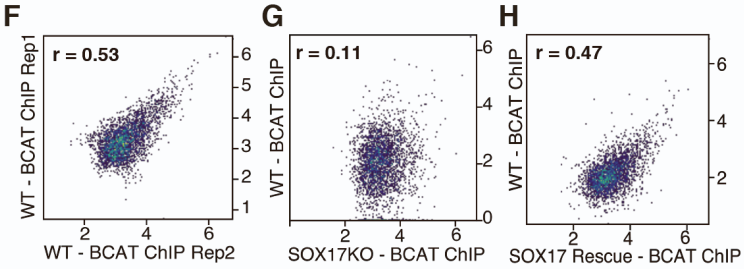
B

tranctripts up in SOX17KO

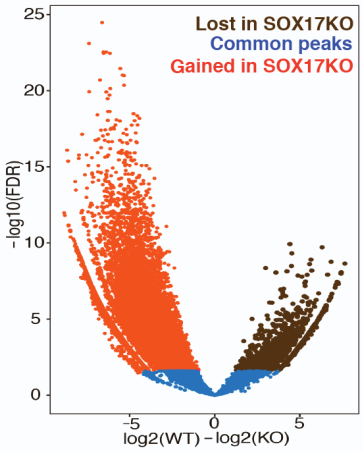
GO:BP	*-log10p*	adj.
hindgut morphogenesis	10.42	
notochord development	9.38	
planar cell polarity pathway	8.68	
outflow tract morphogenesis	7.40	
somite specification	7.10	

transcripts down in SOX17KO

GO:BP	*-log10p*	adj.
lung morphogenesis	4.49	
endoderm differentiation	3.74	
pancreas development	3.56	
endoderm differentiation	2.65	
epithelial development	2.07	



D Differential ß-catenin binding in WT vs SOX17KO



E

ß-catenin peaks lost in SOX17KO, n = 4337

Rank	Motif	*E-val*	GO:BP	*-log10p* adj.
1.	GATA	7.97e-03	• primary germ layer formation	6.10
2.	SOX	8.27e-07	• digestive tract morphogenesis	5.03
3.	FOXA	5.65e-04	• pancreatic differentiation	3.49

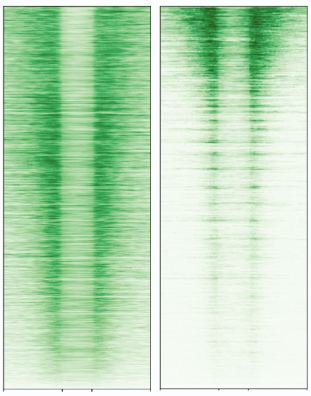
unchanged ß-catenin, n = 3894

Rank	Motif	*p-val*	GO:BP	*-log10p* adj.
1.	GATA	8.93e-03	• endothelial tube morphogenesis	5.69
2.	TCF	1.09e-03	• sprouting angiogenesis	4.53
3.	TBX	5.15e-04	• VEGF signaling pathway	4.13

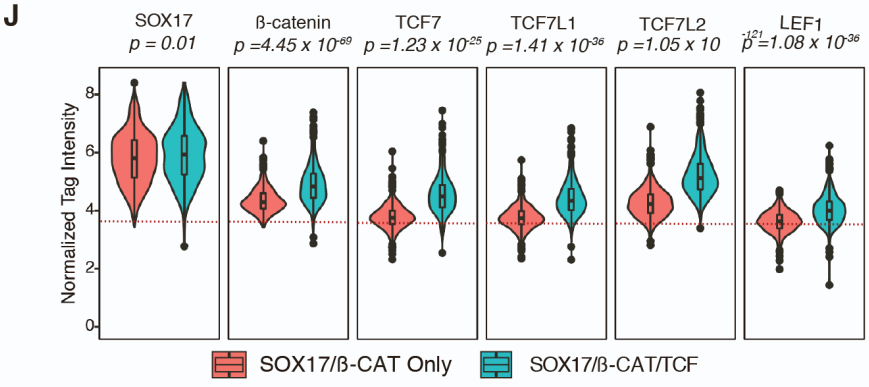
ß-catenin peaks gained in SOX17KO, n = 24096

Rank	Motif	*p-val*	GO:BP	*-log10p* adj.
1.	GATA	4.57e-03	• cardiac septum morphogenesis	75.38
2.	TCF	4.31e-06	• E-M transition	42.26
3.	SMAD	2.11e-04	• catenin import into nucleus	22.44

I peaks co-occupied and coregulated by SOX17 & ß-CAT



J



K SOX17/ ß-CAT cobound and coregulated genes

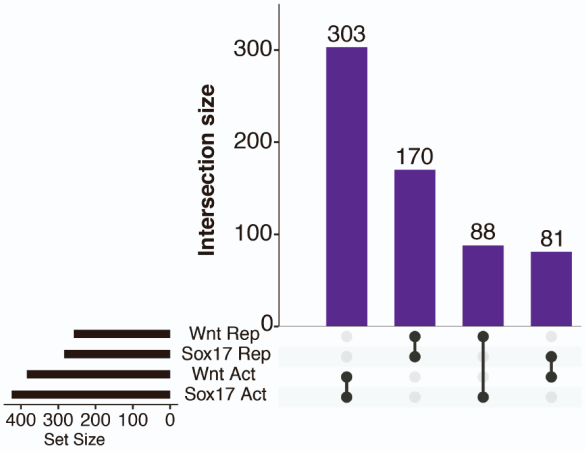


Figure S4 – Related to Figure 3. Identification of SOX17 regulated transcripts and differential ß-catenin binding in *WT* and *SOX17* KO cells. **A.** Volcano plot showing differentially expressed genes between *WT* and *SOX17KO* cells at Day 3 (n = 3, Log2 foldchange >1, FDR p< 0.05). Typical endoderm marker genes are down in the *SOX17KO*, whereas typical mesendoderm or mesoderm markers are upregulated. **B.** GO enrichment analysis confirms that terms associated mesoderm development are upregulated in the SOX17KO, whereas terms associated with endoderm are down (adjusted -log10 transformed p-values, Fisher's exact test, FDR 5%). **C.** Principal component analysis (PCA) of RNA-seq samples (in biological triplicates) at Day 3 of differentiation in *WT*, *SOX17KO* and *KO+SOX17* rescue cells +/- DOX, relative to Day 1 mesendoderm and Day 0 pluripotent stem cells (PSCs). **D.** Volcano plot showing differential ß-catenin ChIP-seq binding events in *WT* vs. *SOX17KO* (fold change >1, FDR p< 0.05). Brown dots represent ß-catenin peaks lost in *SOX17KO*, orange dots are ß-catenin peaks gained in *SOX17KO*, and blue dots are ß-catenin peaks that did not change. **E.** *De-novo* DNA-binding motif analysis and GO term enrichment of genes associated with ß-catenin peaks in each of the above categories. **F-H** Scatterplots and Spearman's rank correlation coefficient comparing the SOX17-dependent ß-catenin ChIP-Seq peaks in **F.** *WT* vs. *WT* replicates, **G.** *WT* vs. *SOX17KO* and **H.** *WT* vs. *KO+SOX17* (DOX) rescue showing that adding back SOX17 to the *SOX17KO* cells rescues the ß-catenin binding pattern at SOX17-dependent ß-catenin bound loci. **I.** ChIP-Seq density plots of H3K4me1 and H3K27ac epigenetic histone marks from Day 3 DE indicating that loci co-bound and coregulated by SOX17/ß-catenin bear hallmarks of active enhancers (H3K4me1 data from GSM772971). **J.** Quantification of ChIP-seq read density of SOX17, ß-catenin, TCF7, TCF7L1, TCF7L2 and LEF1 binding in Day3 *WT* cells identifying enhancers with significant binding of only SOX17/ß-catenin but not any TCF (orange) and enhancers significant for the binding of SOX17, ß-catenin and at least one TCF (green), as identified through differential binding analysis (log2FC > 1, FDR p<0.05). The dotted line represents the tag counts threshold for significant peak. P values were calculated via the Wilcoxon rank sum test. **K.** UpSET plot showing the distribution of SOX17/ß-catenin co-bound and coregulated genes, indicating whether a gene is activated (Act) or repressed (Rep) by SOX17 or WNT.

Figure S5- Related to Figure 4

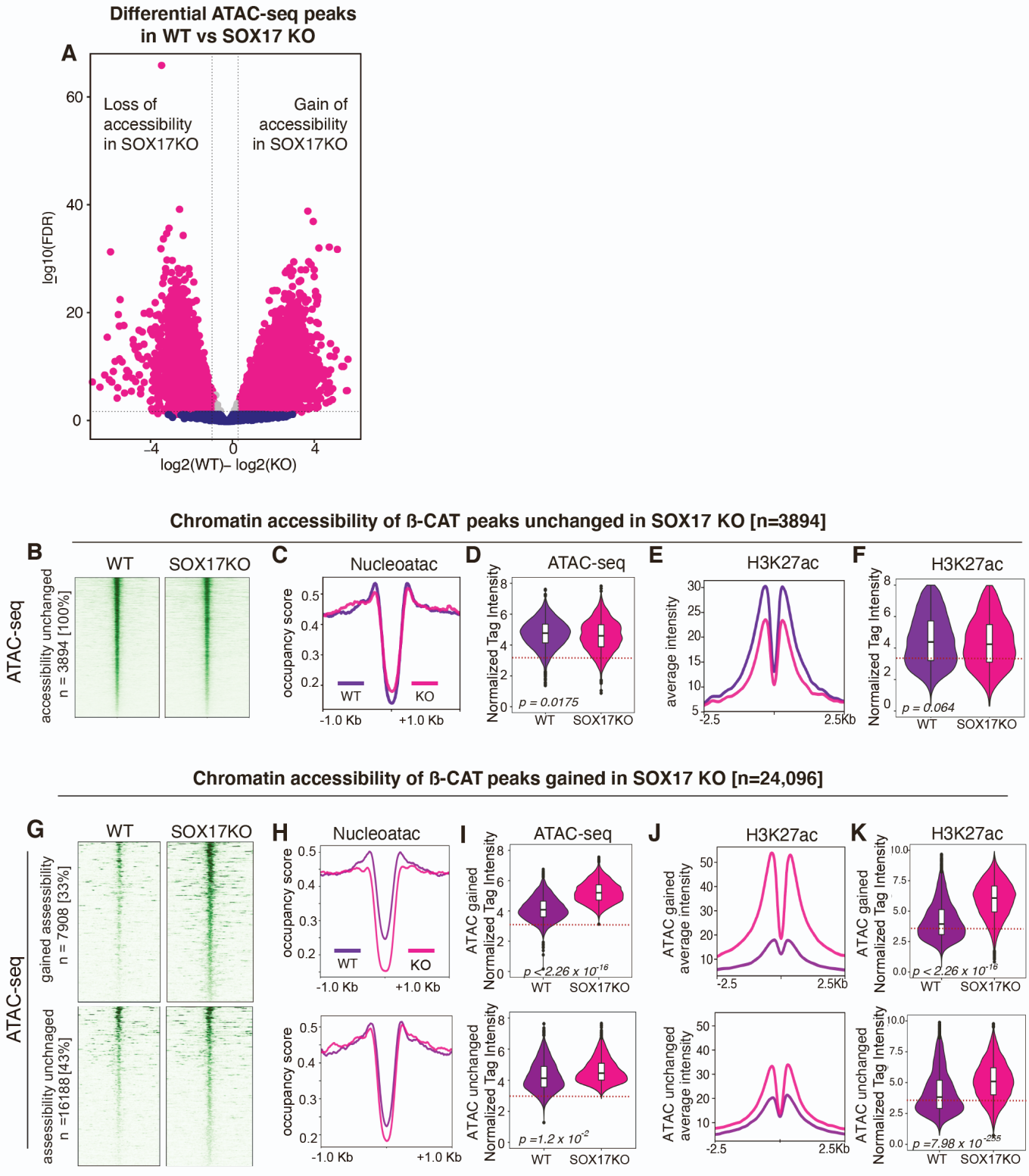
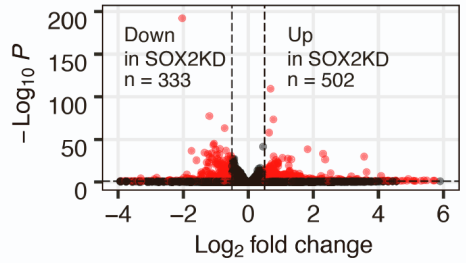


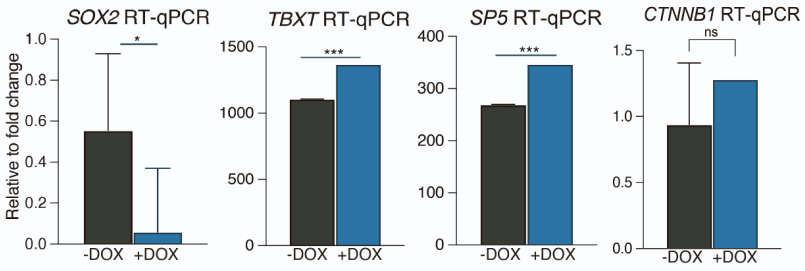
Figure S5 – Related to Figure 4. ATAC-seq and epigenetic analysis of ß-catenin bound loci. A. Volcano plot showing differential SOX17-dependent chromatin accessibility in WT and SOX17KO cells (n = 2, fold change > 1.5, FDR p< 0.05) **B.** Heatmap of ATAC-seq signal for ß-catenin peaks that remain unchanged in SOX17KO. **C.** ATAC-seq metaplot showing average nucleosome occupancy signal at unchanged ß-catenin peaks and **D.** quantification of ATAC-Seq read densities. **E.** Metaplots and **F.** violin plots showing H3K27ac ChIP-seq signal at unchanged ß-catenin peaks. P values calculated by the Wilcoxon rank sum test. The dotted lines in **D.** and **F.** represent the tag count threshold for significant ATAC-Seq peaks. **G.** Heatmaps of ATAC-seq signal for regions that gain *de-novo* ß-catenin peaks in SOX17KO cells. Heatmaps show ATAC signal intensity at two classes of new ß-catenin peaks; those that gain accessibility in SOX17KO cells and those where accessibility is unchanged in both WT and SOX17KO. **H.** Metaplot of nucleoatac signal showing average nucleosome occupancy and **I.** quantification of ATAC-Seq signal at both classes of *de-novo* ß-catenin binding regions. **J.** Metaplot and **K.** quantification of H3K27ac ChIP-seq read intensity at *de-novo* ß-catenin peaks. P values were calculated via the Wilcoxon rank sum test. The dotted lines in **I.** and **K.** represent the tag count threshold for significant ATAC-Seq peaks.

Figure S6- Related to Figure 5

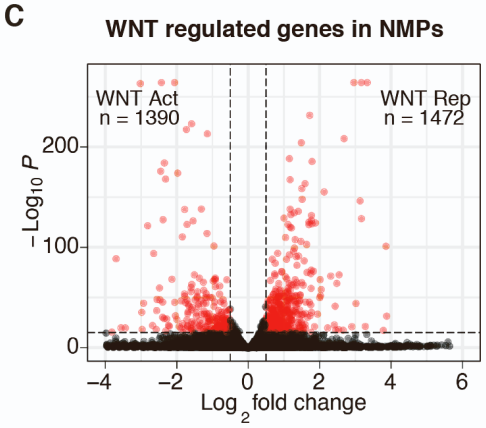
A



B



C



D

SOX2 Activated

GO:BP	*-log10p* adj.
VEGF pathway	9.75
epidermis development	9.29
mammary gland development	8.36
skeletal muscle development	7.51
neurogenesis	2.42

SOX2 Repressed

GO:BP	*-log10p* adj.
endoderm development	40.78
Wnt signaling pathway Wnt, PCP pathway	23.20 12.74
A/P axis specification	6.16
mesenchyme differentiation	3.46

E

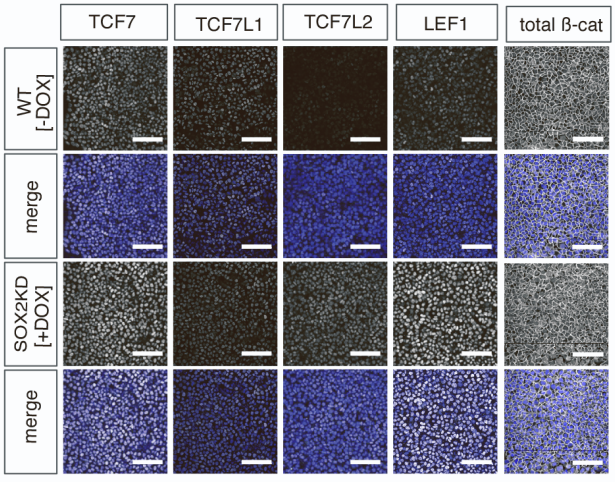
Wnt Activated

GO:BP	*-log10p* adj.
Wnt signaling	36.47
blastoderm segmentation	16.21
A/P axis specification	12.85
heart looping	9.16
mesoderm formation	8.89

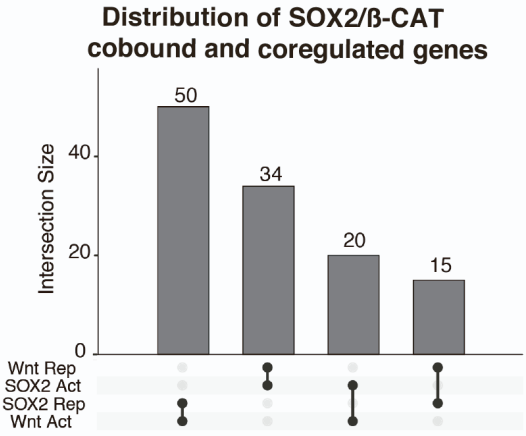
Wnt Repressed

GO:BP	*-log10p* adj.
synapse organization	6.75
negative reg. of FGF signaling	6.26
axonogenesis	3.70
negative reg. of muscle development	3.68
long-term memory	3.68

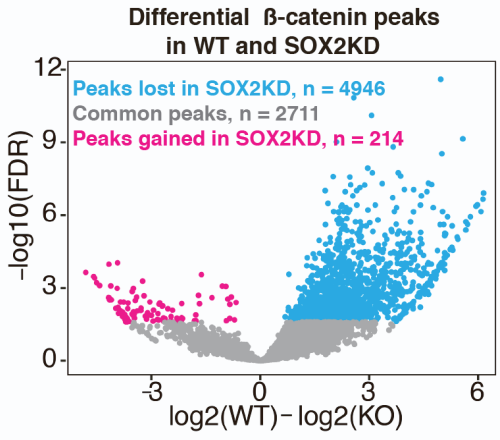
F



G



H



I

ß-catenin peaks lost in SOX2KD n = 4946

Rank	Motif	*E-val*	GO:BP	*-log10p* adj.
1.	TCF	1.1e-255	catenin import into nucleus	12.04
2.	SOX	1.3e-147	ganglion development	10.3
3.	POU	6.1e-119	nervous system development	4.09

Common ß-catenin peaks n = 2711

Rank	Motif	*E-val*	GO:BP	*-log10p* adj.
1.	TCF	8.5e-418	gastrulation	32.9
2.	POU	3.2e-073	mesoderm morphogenesis	24.67
3.	FOX	1.1e-037	segmentation	23.5

Figure S6 – Related to Figure 5. Characterization of SOX2/ß-catenin interactions during NMP differentiation. A. Identification of SOX2-regulated transcripts. Volcano plot showing differential gene expression (log2 fold change >1, FDR p<0.1) of *SOX2KD* cells +/-DOX. **B** RT-qPCR validating expression levels of *SOX2, TBXT,* WNT target *SP5* and *CTNNB1* in *SOX2KD* cells +/-DOX at Day3 of NMP differentiation. mRNA levels were normalized to that of the housekeeping gene *PPIA,* and then normalized to expression levels in PSCs. Significance based on student's t-Test, *p<0.05, **p<0.01, ***p<0.001, p>0.05 is ns (not significant); n = 3. **C.** Identification of WNT-regulated transcripts in NMPs. Volcano plot showing differential gene expression comparing Day3 NMPs differentiated with the WNT-agonist CHIR or the WNT-antagonist C59 in biological triplicates. **D-E.** GO term enrichment analysis of **D.** SOX2 regulated and **E.** WNT regulated genes showing top five terms with adjusted -log10 p-values (Fisher's exact test, <FDR 5%). **F.** Immunostaining showing TCF7, TCF7L1, TCF7L2, LEF1 and total ß-catenin protein levels in *SOX2KD* cells +/-DOX. (n = 3, scalebar = 100 μm). **G.** UpSET plot showing distribution of coregulated ß-catenin and SOX2 genes. **H.** Volcano plot showing differentially bound ß-catenin ChIP-seq peaks in *SOX2KD* cells +/-DOX (fold change >1, p<0.1, n = 2). **I.** Table showing *de-novo* DNA-binding motif analysis of ß-catenin peaks that are lost or unchanged in *SOX2KD*+DOX NMPs. The top 3 most enriched motifs for each category, and their associated E-values are shown.

Figure S7 - Related to Figure 7

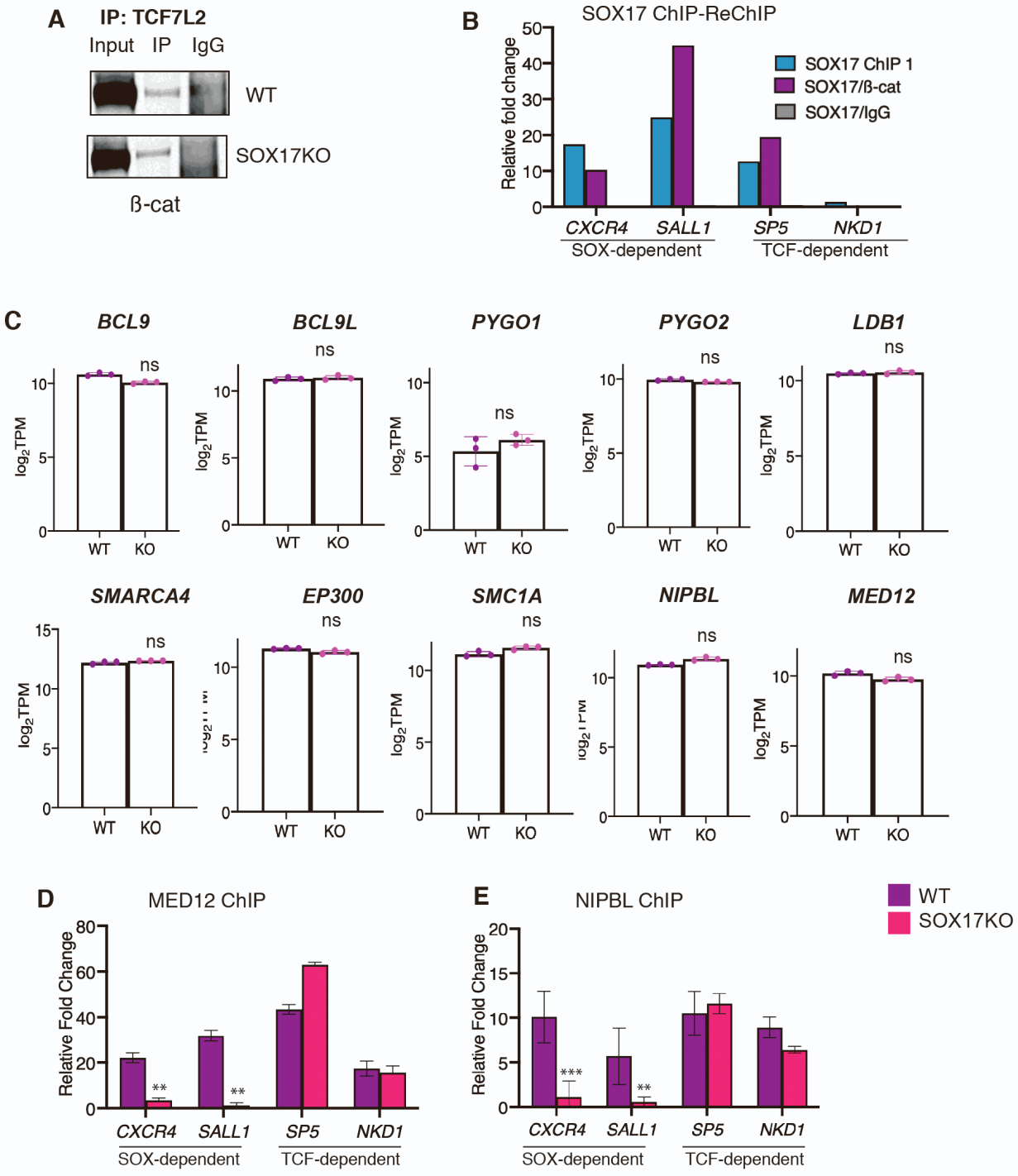


Figure S7 – Related to Figure 6. Analysis of WNT-enhanceosome complex components. A. Western blot confirming that loss of SOX17 does not disrupt TCF7L2 and ß-catenin interactions based on coimmunoprecipitation of TCF7L2 Day 3 WT and *SOX17KO* DE. **B.** Representative SOX17 ChIP-reChIP-qPCR showing relative enrichment following a 1st SOX17 ChIP and subsequent 2nd reChIP with ß-catenin or IgG at SOX-dependent or TCF-independent enhancers. **C.** mRNA expression levels (log_2TPM) of Wnt-enhanceosome components or epigenetic interactors of ß-catenin in Day 3 *WT* DE and *SOX17KO* cells. $p > 0.05$, ns = not significant in two-tailed student's T-test. **D – E.** ChIP-qPCR (n = 3) of MED12 and NIPBL showing relative enrichment at SOX-dependent or TCF-dependent enhancers in Day 3 *WT* DE and *SOX17KO* cells. **=*p<0.01,* ***=p<0.001 based on two-tailed student's T-test.

Table S1 – Related to Figure 6 and STAR Methods: Details of luciferase reporter assays. List of wild-type and SOX or TCF binding site mutated sequences used for enhancer reporter constructs and luciferase assays.

Key: SOX17 sites, TCF sites, ***SOX or TCF sites mutated***

***WT* SALL1: chr16:51187339-51187650**	GATGTTGAAGAATGAAGATAATAATGTTTCCATAGGTGGTGC TTCAAATGCCATTATTTCTCACTGAATATTTAAAGAGATCCCT CGGCAAAGATGGATCTGCGCACTCCTGGGGTGTGAGCGGC TCGCATCTCCCAGACCCCGGGGAACGTGTACGGGAGCACG TGTAAATCCCGCACCCGCCTCCATCTCACAAACAG
Δ*Sox17* SALL1: chr16:51187339-51187650	GATGTT***TATACCAGAG***A***ACGTACTGG***TTTCCATAGGTGGTG CTTCAAAT***ACGTCGGTT***ACTCACTGAATATTTAAAGAGATCC CTCG***CCTTGCTGCAT***TCTGCGCACTCCTGGGGTGTGAGCGG CTCGCATCTCCCAGACCCCGGGGAACGTGTACGGGAGCAC GTGTAAATCCCGCACCCGCCTCCATCTCACAAACAG
***WT* CXCR4: chr2:136936488-136936880**	TGACTATAATAATAACCTTCTTTTACGATAGCTGGGTCCTCTC CTCAGAGCTTCAGGCTTTCTGAATTCAGAGCCTTCCTTAAGG TAAGGAAATGGCAGTTTGCTTGAATACAATGGTCTTTTTAAG GCAGCTCAGATGTTTATCTTATCCCAATATTTACCCTATTGAA TAGGATTTGGGTGGCTTATTTTTCTAATTCT
Δ*Sox17* CXCR4: chr2:136936488-136936880	TGACT***AGTGACAGCT***CCTTCT***TCCGTGTTAT***CTGGGTCCTCT CCTCAGAGCTTCAGGCTTTCTGAATTCAGAGCCTTCCTTAAG GTAAGGAAATGGCAGTTTGCTTGAA***GGCATACGCGCATCGA*** AGGCAGCTCAGATGTTTATCTTATCCCAATATTTAC***TGGTAG CTCTCT***GATTTGGGTGGC***TCGTACTTCG***AATTC
***WT* SP5: chr2:171571253-171571933**	TTCCCATCCCCCCTAATAATCAGTTCTTTTATCCAGACCAACA AACACACCATAGGAGCTTTGTGGATTCAAAGGATTTGCTTTC GCTTCTGAAAGAGCCGCTATTCTTTGATGATTGGGTAGCGG CAAACTTCAAAGCCATAAATCTTCCCTCTGACTGGCTGGCGG CCCAGCAAAGTCCTTATCAAATTCTTGGAGGT
Δ*Sox17* SP5: chr2:171571253-171571933	TTCCCATCCCCCCTAATAATCAGTTCTTTTATCCAGACCAACA AACACA***AGGTACACCGTTCGT***AGATTCAAAGGATTTGCTTTC GCTTCTGAAAGAGC***TGGTACTTCGC***GATGATTGGGTAGCGG CAAACTTCAAAGCCATAAATCTTCCCTCTGACTGGCTGGCGG CCCAGCAAAGTCCTTATCAAATTCTTGGAGGT
Δ*TCF* SP5: chr2:171571253-171571933	TTCCCATCCCCCCTAATAATCAGT***CAGCCCGCA***CAGACCAA CAAACACACCATAGGAGCTTTGTGAGA***GTGTCCTACG***TTGC TTTCGCTTCTGAAAGAGCCGCT***CCTCGCGAGTGCGT***GGGTA GCGGCCCG***ACGCGACCGTG***TAAATCTT***CCGGCCAGTGCTG T***GGCGGCCCAGCAAAGTCCT***AGCCTGAT***TCTTGGAGGT
***WT* AXIN2: chr17:63556520-63556873**	CAAGCGGGGGGCGCCGCTACCCTTTCATCCCACCTCCCAAA GGCACTCCCGCGCGCACTCACACGCCGACTCACATCCATAA CCGCGACAGCGACGCGAC
Δ*TCF* AXIN2: chr17:63556520-63556873	CAAGCGGGGGGCGCCGCT***GGGAGTGACGAGCA***CCTCCCA AAGGCACTCCCGCGCGCACTCACACGCCGACTCACATCCAT AACCGCGACAGCGACGCGAC