An Enhancer-driven Stem Cell-like Program Mediated by SOX9 Blocks Intestinal Differentiation in Colorectal Cancer

Xiaoyan Liang, Gina N. Duronio, Yaying Yang, Pratyusha Bala, Prajna Hebbar, Sandor Spisak, Pranshu Sahgal, Harshabad Singh, Yanxi Zhang, Yingtian Xie, Paloma Cejas, Henry W. Long, Adam J. Bass, Nilay S. Sethi

INVENTORY

Supplementary Material contains 10 Figures, 5 Tables, Additional Methods and Discussion

Supplementary Figure 1 is related to Figure 1
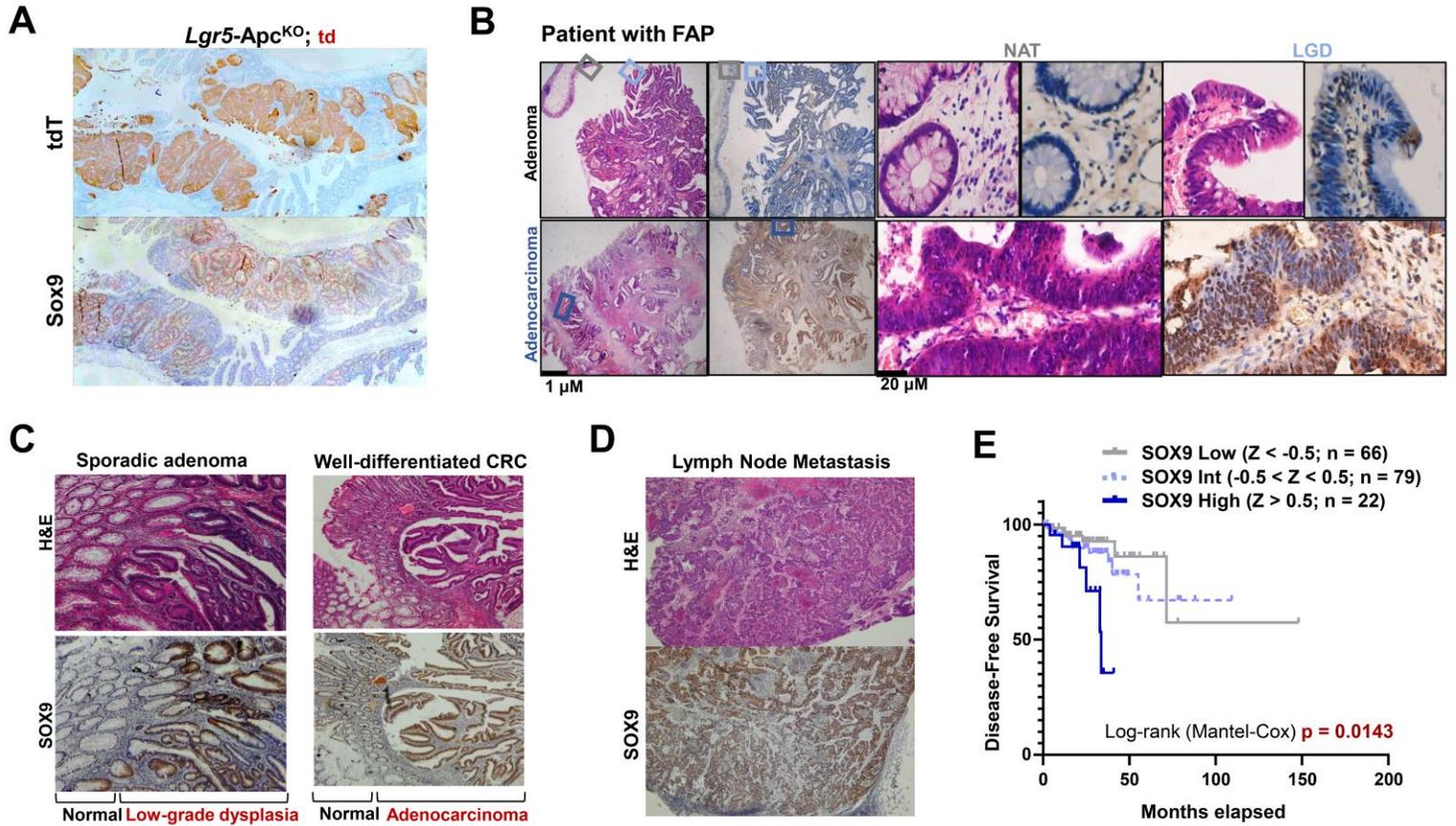Supplementary Figure 2 and 3 is related to Figure 2.
Supplementary Figure 4 is related to Figure 3.
Supplementary Figure 5 and 6 is related to Figure 4.
Supplementary Figure 7 and 8 is related to Figure 5.
Supplementary Figure 9 is related to Figure 6.
Supplementary Figure 10 is related to Figure 7.

Supplementary Figure 1. Elevated expression of SOX9 in murine and human CRC

(A)Histopathology of intestinal lesions from Lgr5-Apc$^{f/f}$;td mice treated with tamoxifen 28 days prior to harvesting tissue. td-RFP and Sox9 immunohistochemistry (IHC) images are shown.

(B) Representative histopathology of paired adenomas and adenocarcinoma from patients with familial adenomatous polyposis (FAP). H&E and SOX9 IHC are shown.

(C) Representative histopathology of human adenomas (left) and CRC (right). H&E and SOX9 IHC are shown.

(D) Histopathology of human CRC that metastasized to the lymph node. H&E and Sox9 IHC are shown.

(E) Kaplan-Meier curve indicated disease-free survival of patients with CRC that have low, intermediate, or high expression of SOX9.
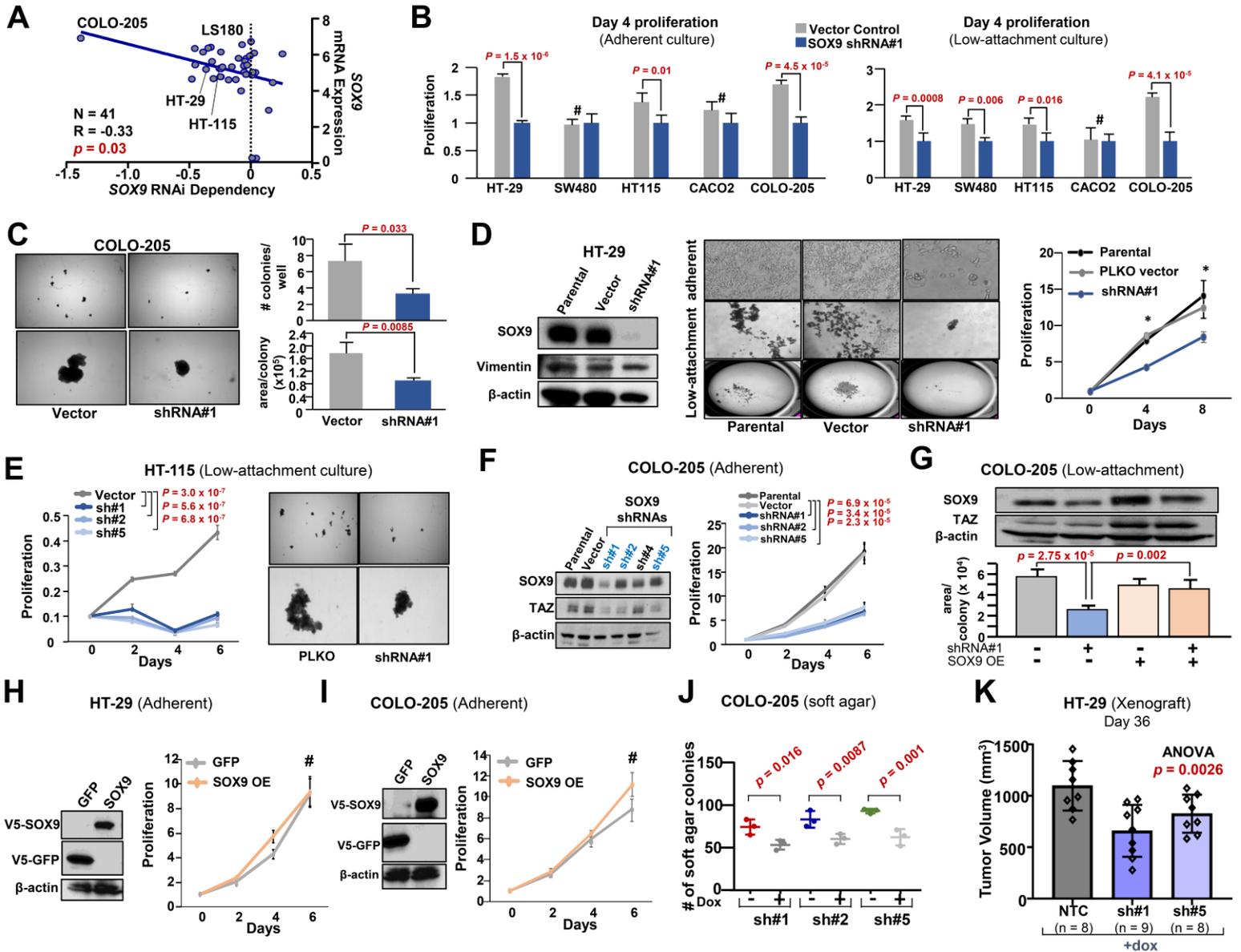
Supplementary Figure 2. WT SOX9 is necessary for CRC proliferation and growth

(A) Cell line dependency for SOX9 RNAi plotted against *SOX9* mRNA expression in CCLE CRC cell lines; regression line indicated in blue; *P*-value calculated by Pearson correlation.

(B) Adherent proliferation and low-attachment colony formation of five CRC cell lines stably expressing vector control and SOX9 shRNA#1 as determined by a luminescent cell viability assay measured at four days. Data expressed as mean ± S.D and *P value*s calculated by two-sided Student's t-test; # indicates no significant difference.

(C) Phase contrast images of COLO205 vector control and shRNA#1 colonies under low-attachment culture conditions. Quantification of colony number and size expressed as mean ± S.D. of three cell culture replicates; *p*-values calculated by two-sided Student's t-test.

(D) Immunoblot (left panel), phase contrast images (middle panel), and proliferation (right panel) of HT-29 parental, vector control, and shRNA#1 cells. Immunoblot shows SOX9 and Vimentin expression relative to β-actin loading control. Phase contrast images of isogenic HT-29 cells growing under low-attachment or adherent conditions. Normalized proliferation determined by a cell viability assay. Data expressed as mean ± S.D. and *p*-values calculated by two-sided Student's t-test, * = *P value* < 0.05.

(E) Normalized proliferation by a cell viability assay and phase contrast images of HT-115 control and SOX9-shRNAs under ultra-low attachment culture. Data presented as mean ± S.D. of three cell culture replicates; *P*

values calculated by two-sided Student's t-test. Representative phase contrast images of HT-115 control or shRNA#1 colonies under ultra-low attachment culture conditions.

(F) Immunoblot and proliferation of COLO-205 cells expressing control or indicated SOX9 shRNAs. Immunoblot showing SOX9, TAZ, and β-actin (loading control) expression. Proliferation determined by a cell viability assay. Data presented as mean ± S.D. of three cell culture replicates; *p-value*s calculated by two-sided Student's t-test.
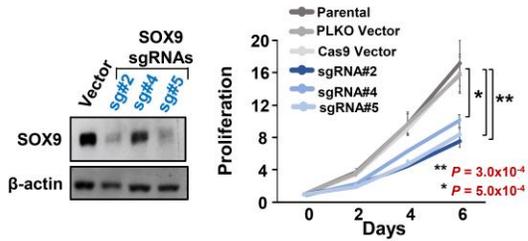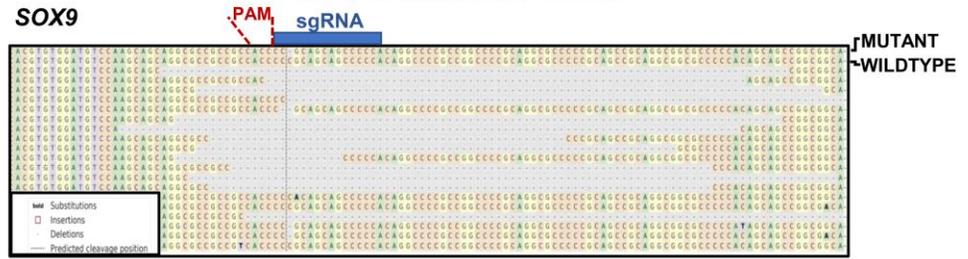
(G) Immunoblot and colony area quantification of COLO-205 cells expressing control or SOX9 sh#1 targeting 3'UTR and overexpression of GFP control or WT SOX9. Immunoblot showing protein levels of SOX9, TAZ and β-actin (loading control). Quantification of colony size presented as mean ± S.D. of three cell culture replicates; *p*-value calculated by two-sided Student's t-test.

(H) Anti-V5 Immunoblot and proliferation of V5-tagged GFP control and SOX9 overexpression in HT-29 cells as determined by a luminescent cell viability assay. Data expressed as mean ± S.D. of three cell culture replicates; *p*-values calculated by two-sided Student's t-test, # = not significant difference.

(I) Anti-V5 Immunoblot and proliferation of V5-tagged GFP control and SOX9 overexpression in COLO-205 cells as determined by a cell viability assay. Data expressed as mean ± S.D. of three cell culture replicates; *p*-values calculated by two-sided Student's t-test, # = not significant difference.

(J) Quantification of soft agar colony formation in COLO-205 cell lines expressing doxycycline-inducible shRNA#1, shRNA#2, or shRNA#5 in the presence and absence of 1 µg/ul doxycycline. Data expressed as mean ± S.D. of three cell culture replicates; *p-value*s calculated by two-sided Student's t-test

(K) Quantification of xenograft tumor volumes of HT-29 cells expressing a control or indicated inducible SOX9 shRNAs in nude mice given doxycycline 625ppm chow for 36 days. Data expressed as mean ± S.D. of indicated tumor number; *p*-value calculated by one-sided ANOVA.

Supplementary Figure 3. SOX9 KO impairs human CRC proliferation

(A) Immunoblot and proliferation of COLO-205 cells stably expressing control and indicated sgRNAs targeting SOX9 under adherent culture. Immunoblot showing protein levels of SOX9 and β-actin (loading control). Data presented
as mean ± S.D. of three cell culture replicates; *p-value*s calculated by two-sided Student's t-test.
(B) Most frequent DNA amplicon sequencing reads from LS180 CRC cells harboring an endogenous *SOX9* heterozygous mutation nucleofected with an inactivating SOX9 sgRNA complexed with recombinant Cas9.
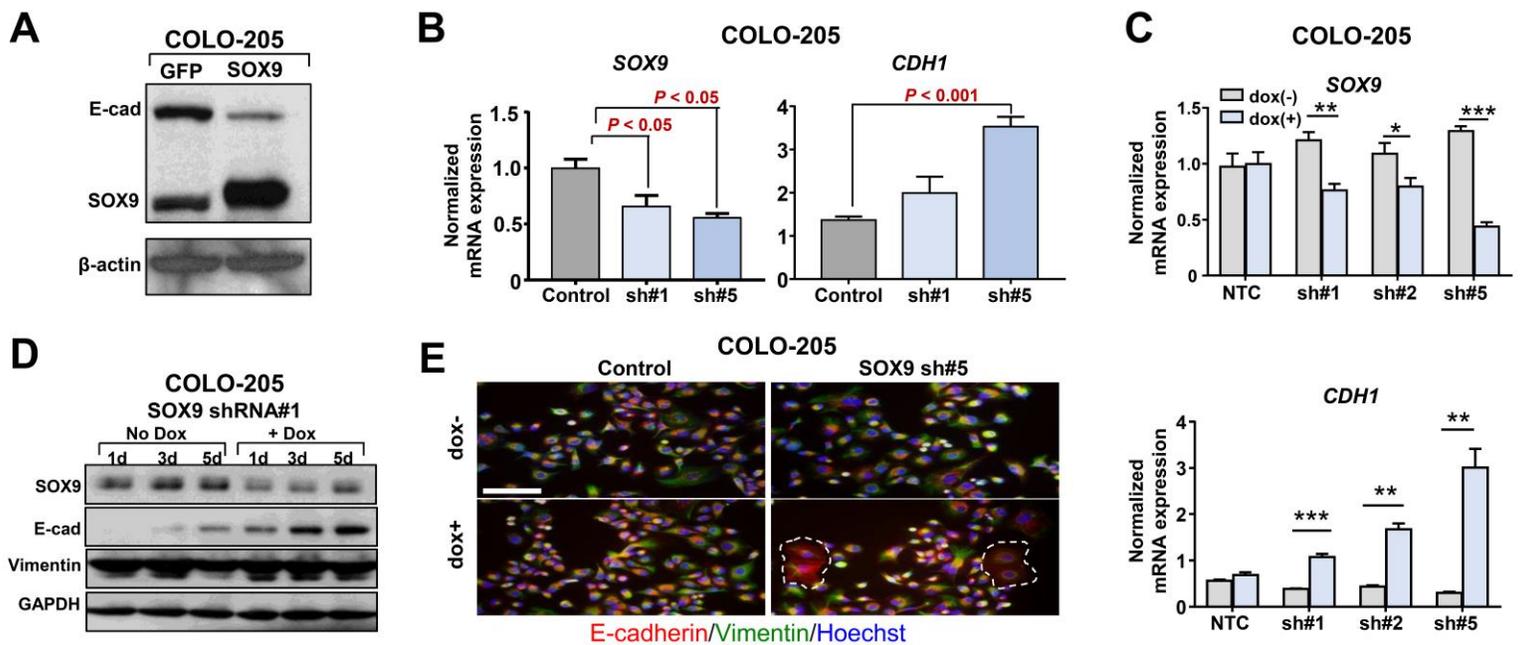
Supplementary Figure 4. SOX9 expression is inversely correlated with intestinal differentiation in human CRC

(A) E-cadherin, SOX9 and β-actin (loading control) protein expression in COLO-205 cells stably expressing GFP or SOX9.

(B) mRNA expression levels of *SOX9* and *CDH1* in COLO-205 cells stably expressing vector control or indicated SOX9 shRNAs by qRT-PCR. Data expressed as mean ± S.D of three technical replicates; *p-value*s calculated by two-sided Student's t-test.

(C) SOX9 and *CDH1* mRNA expression in COLO-205 cells expressing nontargeting control (NTC) or indicated inducible SOX9 shRNAs in the presence and absence of 0.5 μg/ul doxycycline for 13 days by qRT-PCR. Data expressed as mean ± S.D of three technical replicates; *p*-values calculated by two-sided Student's t- test where *p < 0.05, ** p < 0.01, and *** p < 0.005.

(D) SOX9, E-cadherin (*CDH1*), Vimentin and GAPDH (loading control) protein expression following inducible SOX9 KD with shRNA#1 over time in the presence and absence of 0.5 μg/ul doxycycline by immunoblot.

(E) Immunofluorescence of E-cadherin (red), Vimentin (green), and Hoechst (blue) in HT-115 cells following inducible SOX9 KD.
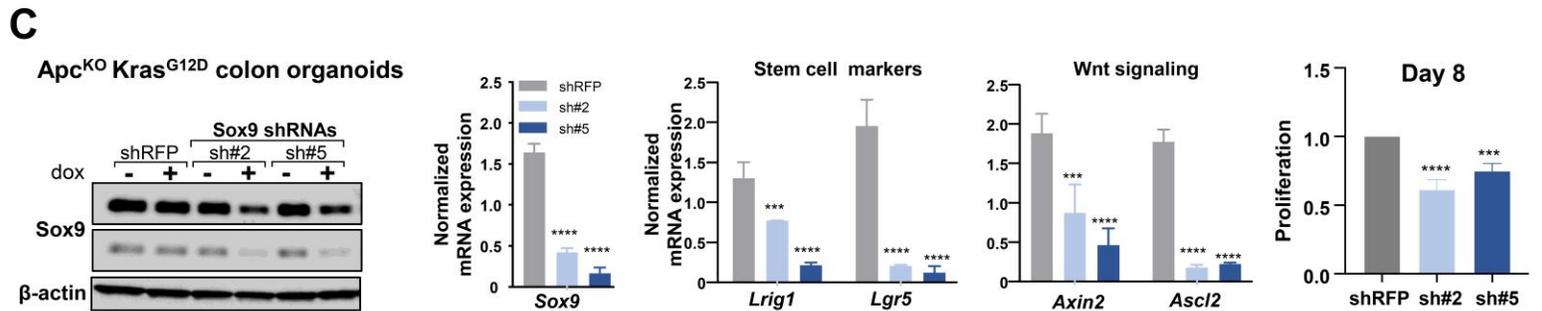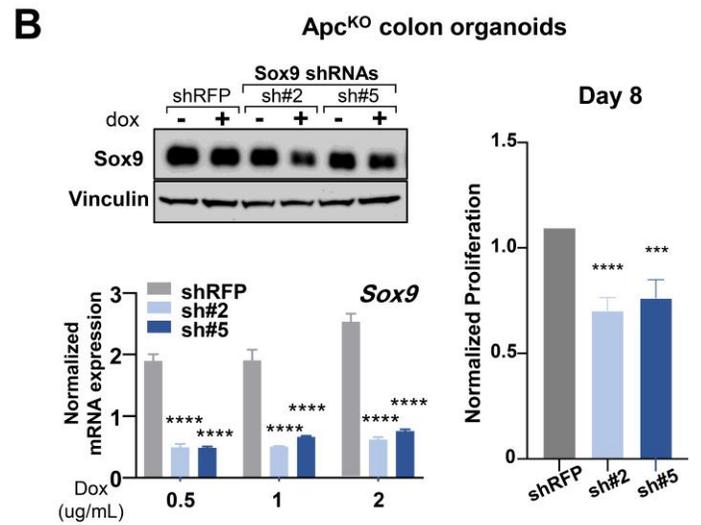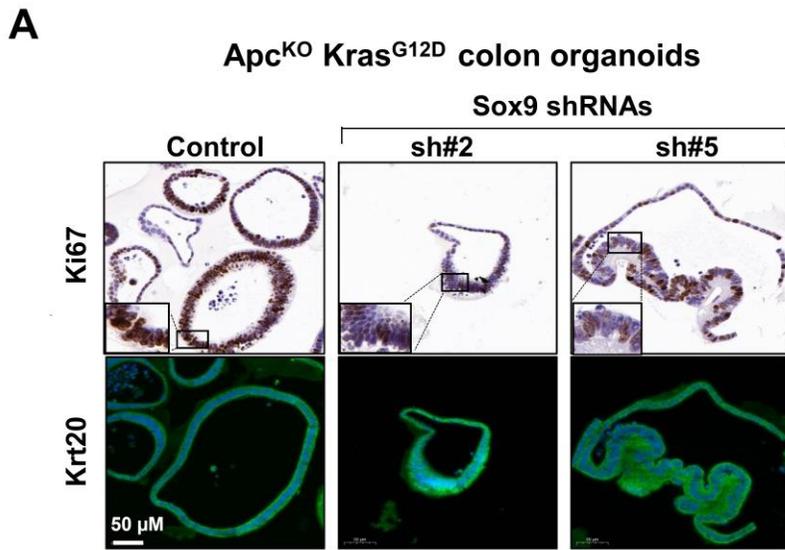
Supplementary Figure 5. SOX9 suppression promotes intestinal differentiation in murine neoplastic organoid models

(A) Ki67 immunohistochemistry and KRT20 immunofluorescence (green) in fixed mouse Apc^{KO};Kras^{G12D} colon organoids stably expressing control or indicated SOX9 shRNAs. Scale bar = 50 µM.
*(B)* Immunoblot showing Sox9 and Vinculin (loading control) expression levels in mouse Apc^{KO} colon organoids engineered to express inducible shRNAs targeting RFP control or SOX9; mRNA expression levels of *Sox9* in engineered Apc^{KO} colon organoids treated with 0.5, 1, or 2 µg/mL of doxycycline by qRT-PCR; Normalized proliferation of Apc^{KO} colon organoids expressing shRFP, shSox9 #2, or shSox9 #5 determined by a cell viability assay after 8 days in culture. Data expressed as mean ± S.D of three technical replicates for qRT-PCR and three cell culture replicates for proliferation; *P value*s calculated by two-sided Student's t-test where *** $p < 0.005$ and **** $p < 0.001$.
(C) Immunoblot showing Sox9 and β-actin (loading control) expression levels in mouse Apc^{KO};Kras^{G12D} colon organoids engineered to express inducible shRNAs targeting RFP control or SOX9; mRNA expression levels of *Sox9* in engineered Apc^{KO};Kras^{G12D} colon organoids treated with 0.5 µg/mL of doxycycline by qRT-PCR; mRNA expression levels of stem cell and WNT pathway markers *Lrig1, Lgr5, Axin2,* and *Ascl2* in engineered Apc^{KO};Kras^{G12D} colon organoids treated with 0.5 µg/mL of doxycycline by qRT-PCR; Normalized proliferation rate of Apc^{KO};Kras^{G12D} colon organoids expressing shRFP, shSox9 #2, or shSox9 #5 determined by a cell viability assay after 8 days in culture. Data expressed as mean ± S.D of three technical replicates for qRT-PCR and three cell culture replicates for proliferation; *P value*s calculated by two-sided Student's t-test where *** $p < 0.005$ and **** $p < 0.001$.
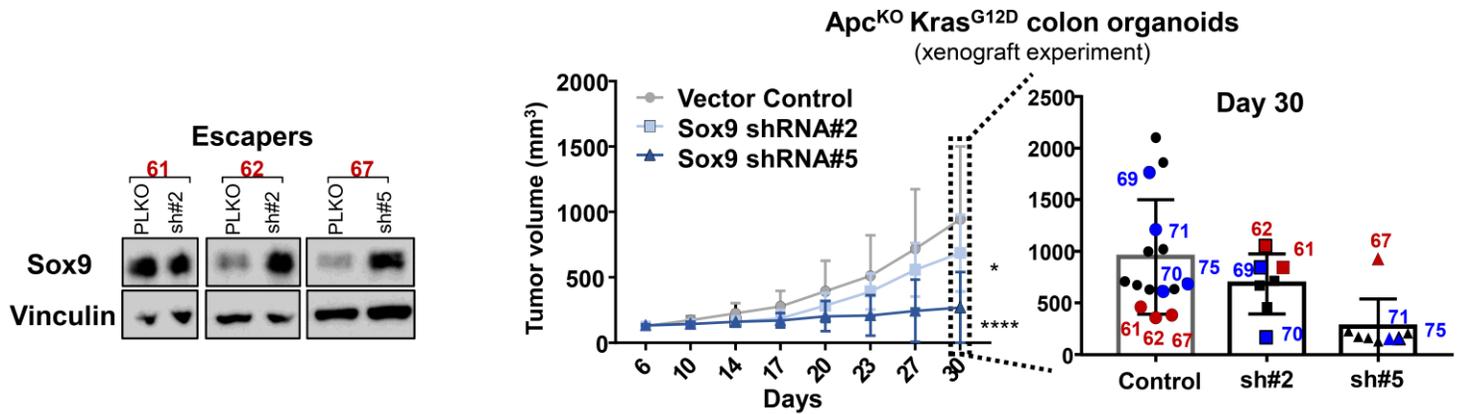
Supplementary Figure 6. SOX9 KD leads to impaired xenograft growth

Primary tumor growth of Apc^{KO};Kras^{G12D} colon organoids expressing control or indicated Sox9 shRNAs injected into the flanks of nude mice (each mouse carried a control tumor on one flank and one of two Sox9 shRNAs in the other flank). Individual dots represent tumor volumes from indicated mice (using last two digits of mouse #) at day 30. Immunoblot using protein extracted from tumors at the experiment end-point. There was a subset of mice in which Sox9 KD was not successful ("escapers") indicated in red and a majority of mice in which Sox9 KD was confirmed ("validated") indicated in blue (please see main figure for immunoblot). Data expressed as mean ± S.D of control (n = 15), shRNA#2 (n = 7), and shRNA#5 (n = 8) groups; *P value*s calculated by two-sided Student's t-test where *p < 0.05 and **** p < 0.001.
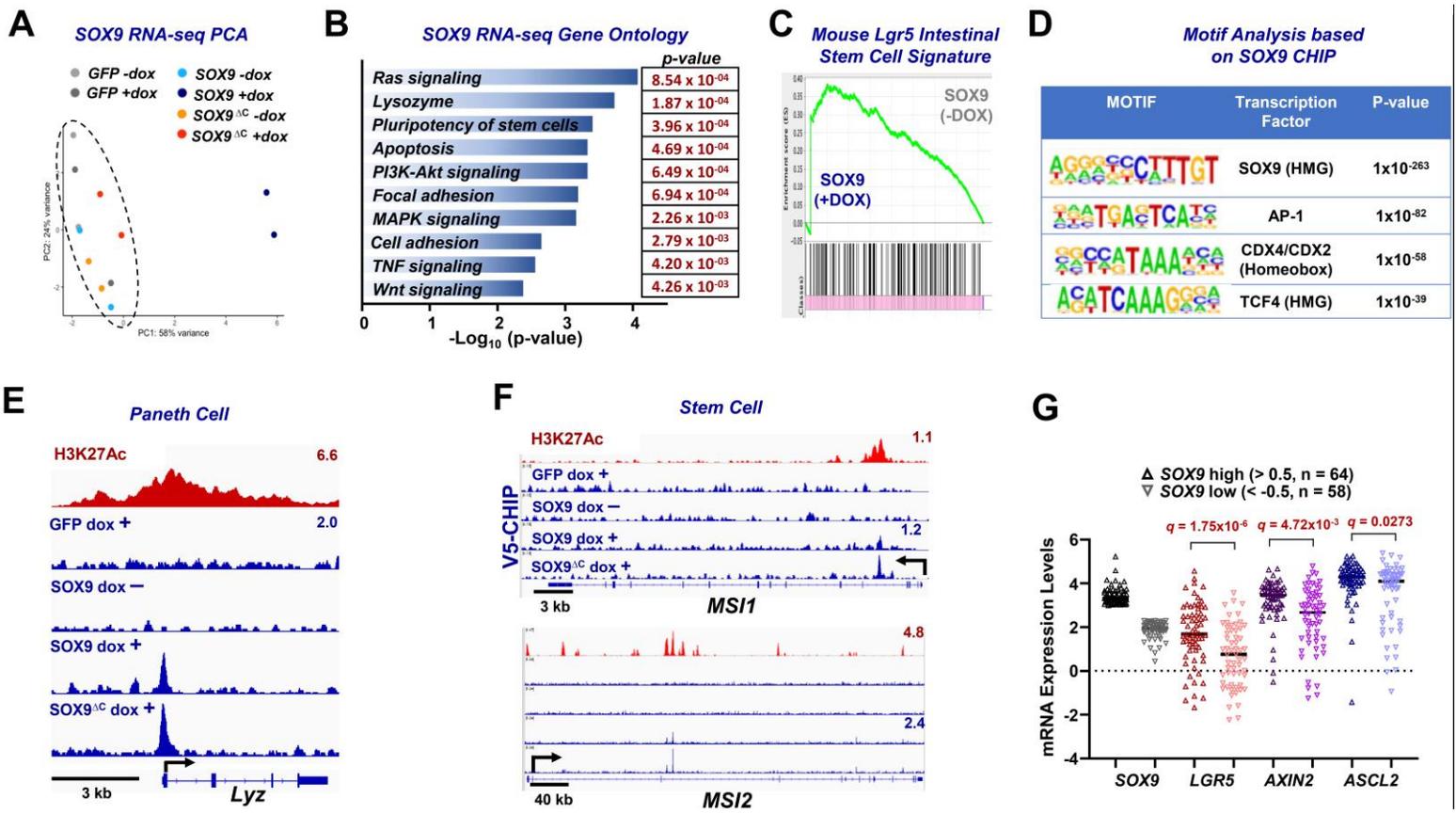
Supplementary Figure 7. SOX9 directly regulates crypt-restricted transcriptional programs

(A) Principal component analysis of RNA-seq transcriptional profiles from LS180 CRC cells expressing GFP control, WT SOX9, or SOX9$^{\Delta C}$ in the presence or absence of doxycycline.

(B) Gene ontology analysis and associated *p*-values based on RNA-seq data from (A).

(C) Gene-set enrichment analysis (GSEA) using 200 gene Lgr5+ intestinal stem cell signature.

(D) Monomer and dimer transcription factor motif predictions using HOMER de novo motif analysis in the V5-SOX9-ChIP-seq data set. Statistical analysis was performed using the Fisher exact test.

(E) Lyz gene track using integrative genomics viewer (IGV) snapshot showing signal from H3K27ac-ChIP-seq data (red) and the V5-ChIP-seq data (blue) derived from HT-115 cells inducibly overexpressing GFP, WT SOX9 and SOX9$^{\Delta C}$ with the indicated doxycycline conditions

(F) MSI1 and MSI2 gene tracks using IGV snapshot showing signal from H3K27ac-ChIP-seq data (red) and the V5-ChIP-seq data (blue) derived from HT-115 cells inducibly overexpressing GFP, WT SOX9 and SOX9$^{\Delta C}$ with the indicated doxycycline conditions

(G) mRNA expression levels of indicated stem cell genes in human CRC samples expressing high and low SOX9 levels from TCGA cohort. Each point represents the expression value of an individual patient and the mean is indicated by the black line.
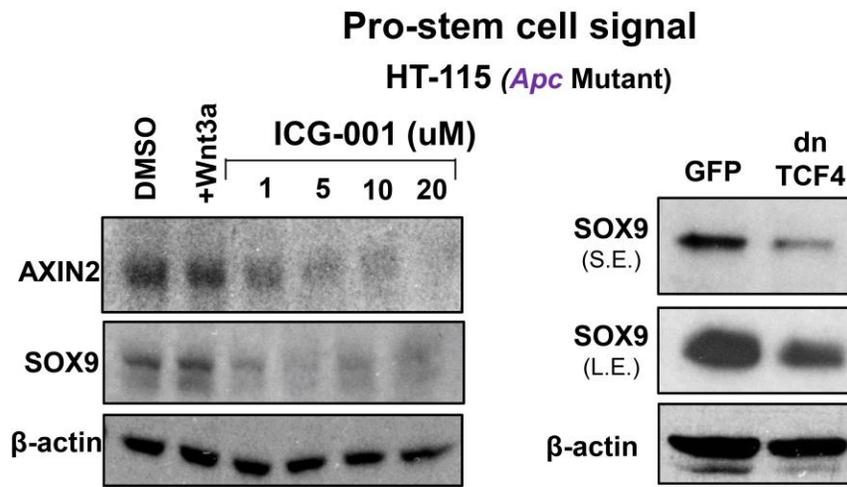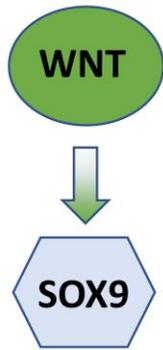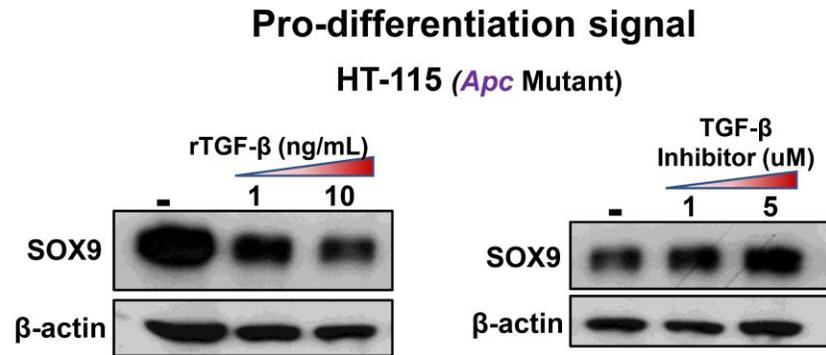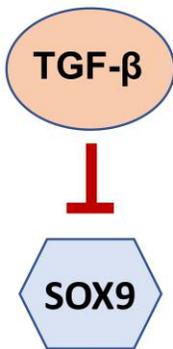
Supplementary Figure 8. SOX9 is regulated by positive signals from Wnt and negative cues from TGF-β signaling pathways

(A) Schematic of Wnt pathway regulation of SOX9 based on our findings. AXIN2, SOX9 and β-actin (loading control) protein expression in HT-115 (*Apc* Mutant) cells treated with DMSO, rWNT3A or indicated concentrations of WNT- inhibitor ICG-001 by immunoblot (left panel). SOX9 and β-actin (loading control) protein expression in HT-115 cells expressing GFP or dominant-negative (dn) TCF4 (right panel); S.E. = short exposure, L.E. = long exposure.
(B) Schematic of TGF-β pathway regulation of SOX9 based on our findings. Protein expression of SOX9 and β-actin (loading control) in HT-115 cells treated with indicated concentration of rTGF-β or TGF-β inhibitor.
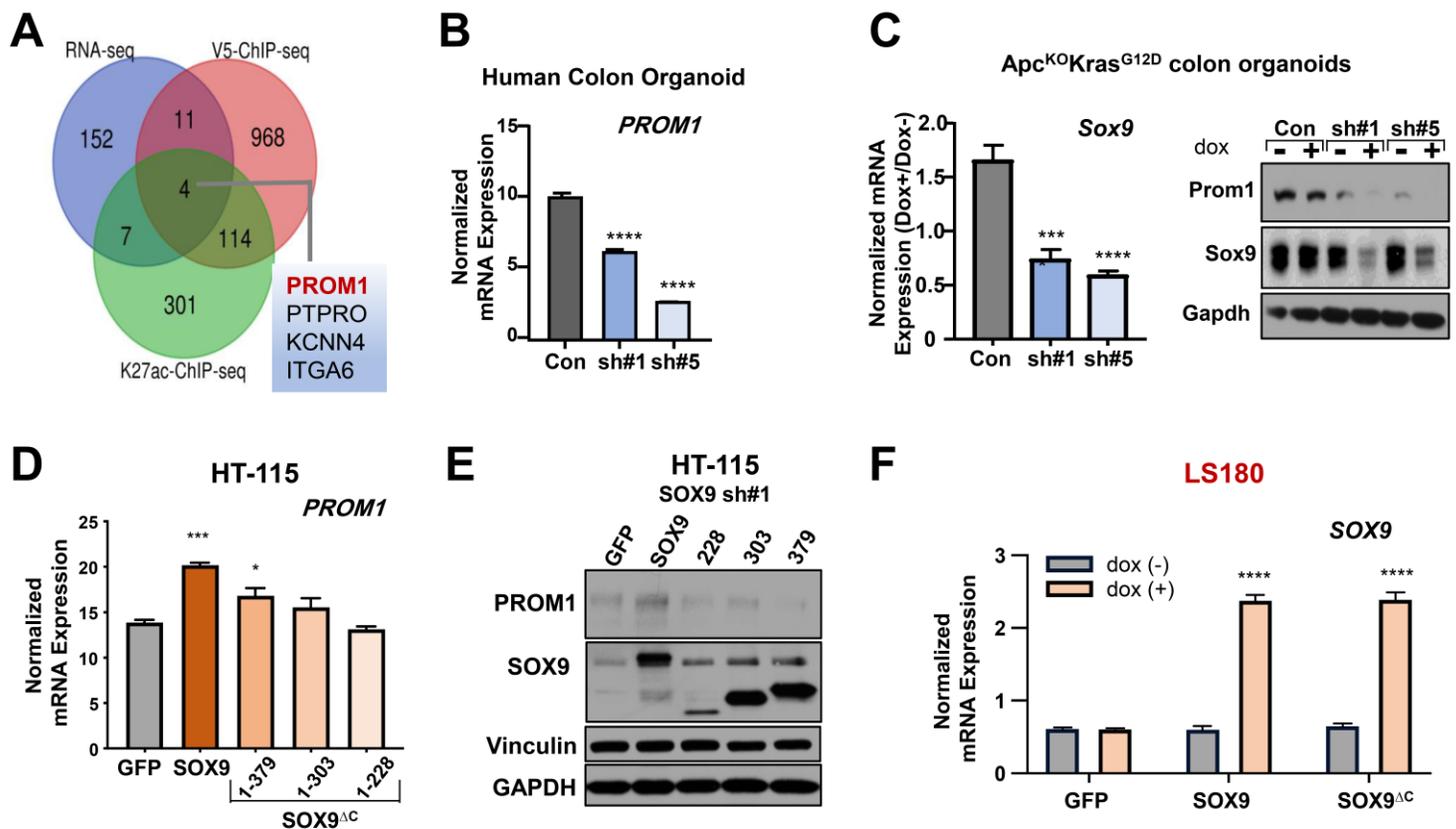
Supplementary Figure 9. SOX9 directly activates intestinal stem cell gene *PROM1* in CRC

(A) Venn diagram summarizing genes potentially regulated by SOX9 based on RNA-seq and indicated CHIP-seq data sets from HT-115 cells inducibly overexpressing GFP control or SOX9.

(B) mRNA expression levels of *PROM1* in control or SOX9 shRNA expressing normal human colon organoids by qRT-PCR. Data expressed as mean ± S.D of three technical replicates; *p*-values calculated by two-sided Student's t-test
where **** $p < 0.001$.

(C) mRNA expression levels of *Sox9* in control or Sox9 shRNA expressing Apc^KO^;Kras^G12D^ colon organoids by qRT-PCR. (left). Data expressed as mean ± S.D of three technical replicates; *p*-values calculated by two-sided Student's t-test where **** $p < 0.001$. Prom1, Sox9, and Gadph (loading control) protein expression in Apc^KO^;Kras^G12D^ colon organoids expressing vector control or inducible Sox9 shRNA in the presence and absence of 0.5 µg/mL doxycycline by immunoblot (right).

(D) mRNA expression levels of *PROM1* in HT-115 CRC cells overexpressing GFP control, WT SOX9, or each of three truncated SOX9 proteins by qRT-PCR. Data expressed as mean ± S.D of three technical replicates; *p*-values calculated by two-sided Student's t-test where *$p < 0.05$ and *** $p < 0.005$.

(E) PROM1, SOX9, Vinculin (loading control) and GAPDH (loading control) protein expression in HT-115 SOX9 shRNA#1 (targets 3'UTR) expressing GFP control, WT SOX9, or each of three truncated SOX9 proteins by immunoblot.

(F) mRNA expression levels of *SOX9* in LS180 CRC cells inducibly overexpressing GFP control, WT SOX9, or mutant SOX9^ΔC^ in the presence or absence of 0.5µg/mL doxycycline by qRT-PCR. Data expressed as mean ± S.D of three technical replicates; *p*-values calculated by two-sided Student's t-test where **** $p < 0.001$.
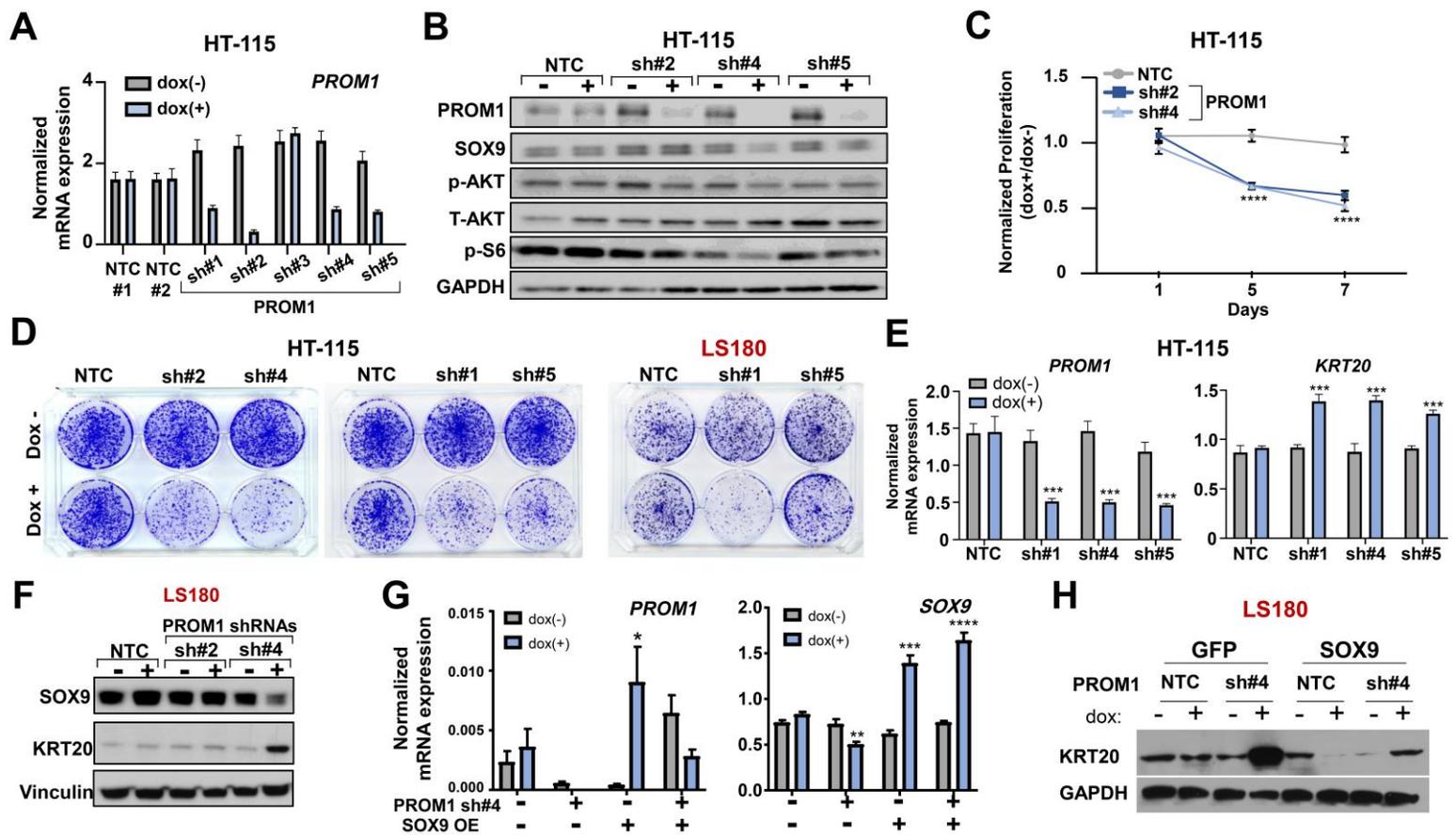
Supplementary Figure 10. PROM1 is a functional regulator of a SOX9-mediated stem cell program in CRC

(A) mRNA expression levels of *PROM1* in HT-115 CRC cells expressing inducible control or *PROM1* targeting shRNAs in the presence or absence of 0.5 μg/mL doxycycline by qRT-PCR. Data expressed as mean ± S.D of three
technical replicates.
(B) PROM1, SOX9, p-AKT, T-AKT, p-S6, and GAPDH (loading control) protein levels in HT-115 CRC expressing inducible control or *PROM1*-targeting shRNAs in the presence or absence of 0.5 μg/mL doxycycline by immunoblot.
(C) Normalized proliferation of HT-115 CRC cell lines inducibly expressing control or *PROM1* targeting shRNAs. Data points shown as cells cultured in presence of 0.5 μg/mL doxycycline divided by cells cultured in the absence of doxycycline. Proliferation determined by a cell viability assay. Data expressed as mean ± S.D. of three technical replicates; *P value*s calculated by two-sided Student's t-test where **** p < 0.001.
(D) Crystal violet staining of HT115 and LS180 CRC cell lines inducibly expressing non-targeting control (NTC) or *PROM1* targeting shRNAs in the presence or absence of 0.5 μg/mL doxycycline.
(E) mRNA expression levels of *PROM1* and *KRT20* in HT-115 CRC cells inducibly expressing non-targeting control or *PROM1* targeting shRNAs and constitutively over-expressing in the presence or absence of 0.5 μg/mL doxycycline measured by qRT-PCR. Data expressed as mean ± S.D. of three technical replicates; *P value*s calculated by two- sided Student's t-test where *** p < 0.005.
(F) SOX9, KRT20 and Vinculin (loading control) protein levels in LS180 cell lines inducibly expressing non-targeting control, PROM1-sh#2, or PROM1-sh#4 in the presence or absence of 0.5 μg/mL doxycycline by immunoblot.
(G) mRNA expression levels of *PROM1* and *SOX9* in LS180 CRC cell lines inducibly overexpressing control, SOX9, or *PROM1*-shRNA#4 in the presence or absence of 0.5 μg/mL doxycycline by qRT-PCR. Data expressed as mean ± S.D. of three technical replicates; *P value*s calculated by two-sided Student's t-test where ** p < 0.01, *** p < 0.005.
(H) KRT20 and GAPDH (loading control) protein levels in LS180 cell lines inducibly expressing non-targeting control or PROM1- shRNA#4 and GFP control or SOX9 in the presence or absence of 0.5 μg/mL doxycycline by immunoblot.

Supplementary Methods | Reporter Assay

Cell Culture, Lentivirus Packing, and Transduction. All cell lines were maintained at 37 °C with 5% CO2. The human colorectal cancer cell lines were obtained from the CCLE core facility and used at early passage for the experiments. HEK293T, HT-115, HT-29, and COLO-205 cells were maintained in DMEM medium supplemented with 10% FBS and 1% penicillin/streptomycin. LS180 cells were cultured in RPMI-1640 containing 10% FBS and 1% penicillin/streptomycin. Genomic information of cell lines are available in Supplementary Table 3.

To generate lentiviruses, expression vectors were co-transfected into HEK293T cells with the lentiviral packaging constructs psPAX2 and pMD2.G (VSV-G) in a 1:1:1 ratio using X-tremeGENE 9 DNA Transfection Reagent (Roche) according to the manufacturer's instructions. Cell culture media was changed the following day and lentiviral supernatant was harvested 48 h and 72 h later and filtered through a 0.45 μm filter (Millipore).

To transduce colonic organoids, cells were resuspended in 500 μl lentiviral supernatant with 8 μg/mL polybrene and 10 μM Y-27632, centrifuged at 600g 32 °C 1 hours, and incubated for 6 hours in cell culture incubator. The infected cells were suspended in 30-50 ul of Matrigel and cultured with Wnt/R-spondin-deprived medium containing 10 μM Y-27632 and 10 μM SB431542.To perform lentiviral infection, the CRC cells were plated in a 6-cm dishes and infected with 0.5-1 mL virus in media containing 8 mg/mL polybrene overnight.

Cell Proliferation Assays. Cell viability was quantified by measuring cellular ATP content using the CellTiterGlo® Cell Viability assay (Promega) according to the manufacturer's instructions. All experiments were performed in triplicate in 96-well plates. Area measurements and quantification of low-attachment colonies was measured using ImageJ software.

RNA isolation, qPCR, and RNA-seq. Total RNA was isolated using the RNeasy Mini Kit (Qiagen, Germantown, MD, USA) and cDNA was synthesized using the iScript™ Reverse Transcription Supermix for RT-qPCR (Bio-Rad). Gene-specific primers for SYBR Green real-time PCR were synthesized by Integrated DNA Technologies or ETON biosciences. Real-time PCR was performed and analyzed using CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories, Inc., Hercules, CA) and using Power SYBR Green PCR Master Mix (Thermo Fisher Scientific). Relative mRNA expression was determined by normalizing to *GAPDH* expression. See Supplementary Table 4 for primers used for qPCR. For RNA-seq analysis, read alignment, quality control and data analysis were performed using VIPER [1]. RNA-seq reads is mapped by STAR [2]. Read counts for each gene is generate by Cufflinks [3]. Differential expression is called by Deseq2[4].

Immunoblot. Immunoblot analysis was performed as previously described[5]. Briefly, cells were lysed in RIPA buffer supplemented with a protease inhibitor cocktail (Roche). Whole cell extracts were resolved by SDS-PAGE, transferred to PVDF membranes, and probed with indicated primary antibodies. Bound

antibodies were detected with horseradish peroxidase (HRP)-conjugated secondary antibodies and chemiluminescent HRP substrate. Primary antibodies listed in Supplementary Table 5.

The following agents were used: recombinant WNT3A as a Wnt activator and ICG-001 as a Wnt pathway inhibitor, all of which were employed at indicated concentrations and durations.

Histopathology. Paraffin-embedded intestines, organoids or xenograft tumors were serially sectioned and mounted on a slide. For morphological analysis, sections were serially dehydrated in xylene and ethanol, stained with H&E for histological assessment, AB-PAS to identify goblet cells and mucus, and immunostaining.

Immunofluorescence and immunohistochemistry. For immunostaining, antigen retrieval was performed using a sodium citrate buffer. Slides were permeabilized using a 0.2% Triton X100 for 30 minutes at room temperature and blocked with donkey serum for 1 hour. Primary antibodies for immunohistochemistry and immunofluorescence are in Supplementary Table 1. Binding of primary antibody was detected with 3,39-diamino-benzidine-tetrahydrochloride-dihydrate and counterstained with hematoxylin. Hoechst 33342 (Invitrogen) was incubated at room temperature for 10 min to stain the nuclei. Slides were counterstained with 4',6-diamidino-2-phenylindole (DAPI) and mounted using the Prolong Gold anti-fade mounting media (Invitrogen).

Reporter Assay. A 537-bp region in intron 1 of PROM1 (hg19, ch4:16,068,931-16,069,504, negative strand) was identified as a putative SOX9-regulated enhancer based on V5-SOX9 and H3K27Ac CHIP data. PCR-amplified products or gene blocks (gBlocks, IDT) consisting of the putative PROM1 enhancer, a minimal promoter, eGFP, and PGK-hygromycin resistance cassette was cloned into the lentiviral LV 1-5 backbone vector using Gibson modular assembly platform[6]. Briefly, PROM1 enhancer was generated as a gBlock by adding the following primer sequencing to the 5' and 3' end of the 537-bp region, respectively: 5'→3' Forward (site 1), GATCAGTGTGAGGGAGTGTAAAGCTGGTTT and 5'→3' Reverse, CTAACTCGAACGCTAGCTGTGCGATCGTTT (site 2); the minimal promoter was PCR amplified from the 7TFP WT CDH1 reporter plasmid using the following primer sequences: 5'→3' Forward (includes site 2), CTAACTCGAACGCTAGCTGTGCGATCGTTTAGTCAGTTCAGACTCCAGCCC and 5'→3' Reverse, AGGCCTCGGGATTcctaggAACAGCGGTTTGTGGCTTTACCAACAGTACCGGAA (includes site 3); eGFP was generated by adding the following primer sequences to the 5' and 3' end, respectively: 5'→3' Forward (site 3), AAACCGCTGTTcctaggAATCCCGAGGCCT and 5'→3' Reverse, GACCCGACATTagcgctACAGCTTAAGCGG (site 4); and PGK-hygro was PCR amplified from plenti-PGK-hygro-gateway destination vector using the following primers: 5'→3' Forward (includes site 4), GACCCGACATTagcgctACAGCTTAAGCGGACTAGTGATCTCTCGAGGTTAACGAAT and 5'→3' Reverse, AGAGTAATTCAACCCCAAACAACAACGTTTTCAGCTCTTGTTCGGTCGGCA

(includes site 5). PCR products were extracted from agarose gels after electrophoresis using the QIAquick Gel Extraction kit (Qiagen) according to the manufacturer's protocol. PCR products and gBlocks were then adjusted to 57 nM with Tris EDTA buffer (pH = 8.0). 5µL of PCR products and gBlocks (5.7 x 10^{-2} pmol each) were added to 15 µl of isothermal master mix, incubated at 50°C for 20 min, and transformed into competent bacteria. Isothermal master mix was prepared by combining 320µL of 5X isothermal assembly reaction buffer, 1.2µL of T5 exonuclease (NEB), 20µL of Phusion polymerase (NEB), 160µL of Taq ligase (NEB) and 700 µL of water. 5X isothermal assembly reaction buffer was prepared by combining 3 ML of 1M Tris-HCL (pH = 7.5), 300 uL of 1M MgCl2, 600 µl of 10mM dNTPs, 300 µL of 1M DTT, 1.5g of PEG-8000, 20mg of NAD and water to a total of 6mL. The resultant reporter vector was transiently transfected into HEK293T cells and either treated with Wnt3a and/or concomitantly transfected with indicated expression plasmids.

Genetic Dependencies. To evaluate genetic dependencies, we interrogated the genetic dependency combined RNAi dataset and CRISPR Avana dataset at DepMap portal (https://depmap.org) (19Q4 data release), which is a public catalog of essential genes and dependencies for cell line proliferation as determined by genome-wide RNAi (three publicly available dataset: Achilles, Novartis and Marcotte) and CRISPR screening. Higher negative values of combined RNAi or CRISPR dependency score represent higher dependency for that gene.

ChIP-seq analysis. The ChiLin pipeline 2.0.0[7] was used for quality control and pre-processing of the data. Burrows-Wheeler Aligner (BWA Version: 0.7.17-r1188) was used to map reads and Model-based Analysis of ChIP-Seq (MACS2)[8](v2.1.0.20140616) for peak calling using default parameters. Briefly, to find the nearest gene for ChIP-seq peaks, we used 'bedtools closest' to get the closest (it can be overlapping or non-overlapping) gene between peak file and reference gene file (+/- 1kbTSS). We used macs2(v2.1.2) to call narrow peaks using default parameters, cut offs: fdr="0.01", keepdup="1", extsize="146"[Macs2]. CEAS analysis is used to annotate resulting peaks with genome features. Differential analysis of peaks was determined by DESeq[9]. Super-enhancers were called by ROSE [10] in H3K27ac ChIP-seq data. Cistrome toolkit was used to probe which factors might regulate the user-defined genes. Genomic Regions Enrichment of Annotations Tool (GREAT) [11] was used to annotate peaks with their biological functions. Conservation plots were obtained with the Conservation Plot (version 1.0.0) tool available in Cistrome.

ChIP-seq data visualization. Normalized profiles corresponding to read coverage per 1 million reads were used for heatmaps and for visualization using the integrative genomics viewer (IGV) [12]. Wiggle tracks were visualized using the integrative genomics viewer. Heat maps were prepared using deepTools (version 2.5.4)[13]

Statistical Analysis and Reproducibility. Experiments were performed in triplicate. Data are represented as mean ± s.d unless indicated otherwise. For each experiment, either independent biological or technical

replicates are as noted in figure legends. Statistical analysis was performed using Microsoft Office or Prism 7.0 (GraphPad) statistical tools. Pairwise comparisons between groups (that is, experimental versus control) were performed using an unpaired two-tailed Student's *t*-test or Kruskal–Wallis test as appropriate unless otherwise indicated.

Supplementary Discussion

Like SOX9, the apical transmembrane protein PROM1/CD133 is expressed on Lgr5+ stem cells and transit amplifying cells in the intestines[14]. We provide evidence that PROM1 activation can block intestinal differentiation via SOX9, which in turn activates PROM1 transcription, reinforcing a positive feedback circuit. A deeper look into how PROM1 outer membrane modifications (i.e. glycosylation), membrane-bound and intracellular partners, and downstream pathway mediators impact signal transduction could yield greater insight into this regulation.

Supplementary References

1. Cornwell M, Vangala M, Taing L, et al. VIPER: Visualization Pipeline for RNA-seq, a Snakemake workflow for efficient and complete RNA-seq analysis. BMC Bioinformatics 2018;19:135.
2. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15-21.
3. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010;28:511-5.
4. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
5. Wong GS, Zhou J, Liu JB, et al. Targeting wild-type KRAS-amplified gastroesophageal cancer through combined MEK and SHP2 inhibition. Nat Med 2018;24:968-977.
6. Akama-Garren EH, Joshi NS, Tammela T, et al. A Modular Assembly Platform for Rapid Generation of DNA Constructs. Sci Rep 2016;6:16836.
7. Qin Q, Mei S, Wu Q, et al. ChiLin: a comprehensive ChIP-seq and DNase-seq quality control and analysis pipeline. BMC Bioinformatics 2016;17:404.
8. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008;9:R137.
9. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11:R106.
10. Whyte WA, Orlando DA, Hnisz D, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 2013;153:307-19.
11. McLean CY, Bristor D, Hiller M, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 2010;28:495-501.
12. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2013;14:178-92.
13. Ramirez F, Dundar F, Diehl S, et al. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res 2014;42:W187-91.
14. Snippert HJ, van Es JH, van den Born M, et al. Prominin-1/CD133 marks stem cells and early progenitors in mouse small intestine. Gastroenterology 2009;136:2187-2194 e1.