Supplementary methods

TCGA data analysis. The Cancer Genome Atlas (TCGA) colonic adenocarcinoma and rectal adenocarcinoma (TCGA-COAD and TCGA-READ respectively) annotated mutation and methylation files were extracted from the GDC portal (https://portal.gdc.cancer.gov/) for 527 cases only, as the remaining cases did not have complete data for mutation and methylation. We determined that there were 50 *BRAF*^{V600E} CRC cases within this cohort and used *BRAF*^{V600E} to mark the serrated CRC subtype. We do not examine *KRAS* mutant serrated CRC herein, as *KRAS* mutation also marks the conventional pathway to CRC. We then used the analysis pipeline described below to examine alteration (both DNA mutation and hypermethylation) to candidate genes involved in regulation of the stem cell niche, senescence and DNA mismatch repair including; the Wnt-pathway, Bone Morphogenic Protein (BMP)/transforming growth factor-beta (TGFβ) pathways, *P16 INK4A*, *MLH1*. TCGA methylation data generated using two different platforms, the Infinium Human Methylation 27K and 450K, were combined. No batch effects were observed. The median beta-value across multiple microarray probes corresponding to the same gene was used to summarise methylation for each gene, with Beta-value for hyper-methylation cut off set to 0.47. All data processing was undertaken using software R, version 3.4.1 Patched. Waterfall plots (Figure 1) were generated with R package oncoPrint() function in ComplexHeatmap R package version 1.12.0 [1]. To examine whether Sphingolipid pathway genes were differentially regulated in the TCGA *BRAF*^{V600E} CRC cohort and examine the effect on patient survival we analysed RNA expression and patient survival by downloading 622 cases with clinical and RNA expression data using R Bioconductor TCGAbiolinks package, version 2.7.5. Maximal survival time analysed was 5 years. RNA expression values were transformed using the 'voom' method limma package, version 3.30.13. Normalised, transformed expression values were used to classify patients into high and low expressor groups for each gene of interest using the surv_cutpoint() function in survminer R, the minimal proportion of observations was set at 20% for each gene [2]. The prognosis of each group was examined using Kaplan-Meier survival estimators with the survminer R package, version 0.4.1, with survival outcome compared by log-rank tests.

Genome engineering & Organoid culture (cont). Organoids were dissociated to single cells and small clumps of cells with TrypLE express, washed and transfected in the presence of basal medium without antibiotics and supplemented with 1uM Jagged-1 (Anaspec) using LF2000 and plasmid DNA in a 96 well plate for 3 hours at 37˚C. Empty px458-GFP plasmid was used to visualise transfection efficiency and used as a kill test control for selection medium conditions. Cells were washed and plated onto 50% matrigel in non-selective culture medium containing 1uM Jagged-1. The following day, media was gently aspirated and top layer of 50% matrigel added. Plate was spun 20 minutes at 200xg at room temperature followed by 30 minute incubation at 37 degrees C before addition of non-selective culture medium containing 1uM Jagged-1. 72 hours after transfection, media was changed to selection medium, with media changes subsequently 2-3 times weekly.

Single organoids were handpicked and expanded for DNA sequence analysis. For novel gRNAs, genomic DNA was isolated from pooled, polyclonal organoids and subjected to PCR amplification of target genomic region, followed by PAGE analysis of duplex formation to measure indel formation. Genomic DNA from single organoid lines was also isolated and target genomic regions amplified by PCR. PCR products were cloned into a p-JET1.2 cloning vector

according to the manufacturer's instructions (ThermoFisher). Plasmid DNA from at least 6 clones was analysed by Sanger sequencing to determine biallelic target gene alteration.

We found optimal recovery and expansion of organoids from single cells as required for the transfection protocol using a matrigel "sandwich" and used this method for all cultures described herein [3]. To facilitate visualisation of the entire 48-well in a flat plane we modify the "sandwich" method to include the addition of 60ul 50:50 ADMEM:Matrigel mixture to a 48-well, centrifuge this bottom matrigel layer for 40minutes at 200xg at room temp, followed by a 30 minute incubation at 37 degC to solidify matrigel before plating organoids in growth medium on top. The following day we remove the media and any non-attached organoids, cover with 40ul 50:50 ADMEM:Matrigel mixture and centrifuge 100xg for 20 minutes. After a 30 minute incubation at 37degC, organoids in matrigel are covered with 200-250ul growth medium. The basal culture medium for mouse colon organoids was Advanced Dulbecco's modified Eagle medium/F12 (Life Technologies) supplemented with 1x gentamicin/antimycotic/antibiotic (Life Technologies), 10mM HEPES, 2mM GlutaMAX, 1xB27 (Life Technologies), 1xN2 (Life Technologies). The following niche factors were used: 50ng/ml mouse recombinant EGF (Peprotech), 100 ng/ml mouse recombinant noggin (Peprotech), 20% R-spondin-2 conditioned medium, 50% Wnt-3A conditioned medium. To select mutant organoids the following reagents were used: 1uM 4-OHT for 24h (Sigma-Aldrich), 10ng/ml human recombinant TGF-β1 (Peprotech), 1uM EGFR inhibitor (EGFRi, Calbiochem #324840), 20uM 5-FU (Sigma-Aldrich). Following initial isolation and immediately after each split, organoids were cultured in 10uM Y-27632 (In Vitro Technologies), 3uM iPSC (Calbiochem Cat #420220), 3uM GSK-3 inhibitor (XVI, Calbiochem, # 361559) for the first 3 days. Prior to transfection, organoids were cultured with 10mM Nicotinamide (Sigma-Aldrich) and 10uM GSK-3 inhibitor for 2 passages [4].

Kras$^{G12D/+}$*; Apc* $^{\Delta 580l/\Delta 580}$ (AK) colonic organoids were generated by treating organoids from Kras$^{\text{LSL-G12D/+}}$;*Apc*$^{CKO/CKO}$ mice with Ad5CMV::Cre, followed by selection in medium without Wnt ligands [5].

Orthotopic Injection (cont). A customised needle (Hamilton Inc. part number 7803-05, removable needle, 33 gauge, 12 inches long, point 4, 12 degree bevel) was used. Colonoscopy was performed using a Karl Storz Image 1 Camera System comprised of: Image1 HDTV HUB CCU; Cold Light Fountain LED Nova 150 light source; Full HD Image1 3 Chip H3-Z Camera Head; Hopkins Telescope, 1.9mm, 0 degrees. A sealed luer lock was placed on the working channel of the telescope sheath to ensure minimal air leakage (Coherent Scientific, part number 14034-40). In each mouse up to 3 injections of 20µl were performed. Injection sites were monitored by weekly colonoscopy and the videos were viewed offline using QuickTime Player for analysis. Tumour growth of the largest tumour visualised was scored as previously described using the Becker scale [6]. A Clinical Record Score was used to determine humane endpoint, scores were obtained by one point being given for the presence of each of the following observations: weight loss >15%; hunched/ruffled coat; inability to pass stool; dehydrated; absence of movement; or facial grimace. Once a score of 3 was reached the mice were euthanased. Statistical analysis performed using Prism.

Western blot analysis. To reduce endogenous MAPK signalling organoids were cultured without EGF for 4 days and with 1uM EGFRi for one day prior to treatment with 1uM 4-OHT for 16h. Cells were harvested using cold PBS containing 1x Halt™ Protease and Phosphatase

Inhibitor Cocktail (PPI, Thermo Scientific), pelleted and lysed using RIPA buffer (150mM sodium chloride, 1% NP40, 0.5% sodium deoxycholate, 0.1% SDS, 50mM Tris, pH8.0) containing 1x PPI. Protein was quantitated using the DC™ Protein Assay Kit II (Bio-Rad) and analysed using 4-20% (Bio-Rad) SDS-PAGE. Standard western blot analyses were performed using the following primary antibodies: anti- Phospho-p44/42 MAPK (Erk1/2) (#4370, Cell Signaling Technology, 1:2000), total p44/42 MAPK (Erk1/2, #137F5, Cell Signaling Technology, 1:2000), anti-Hsp90 loading control (sc-7947, Santa Cruz Biotechnology, 1:1000). Primary antibody binding was detected using HRP-conjugated secondary anti-Rabbit antibody (Novex, #A24537, 1:10,000) and visualised with ECL Western Blotting Detection Reagents (Amersham, RPN2209) and Gel Doc™ XR+ Gel Documentation System (Bio-Rad).

Immunohistochemistry and histology. Tissues were fixed in 10% formalin, paraffin embedded and sectioned. Consecutive sections were prepared using a rotary microtome. H&E and alcian blue stains were preformed according to standard procedures. The $Braf^{V600E}$ staining was performed on a Ventana Benchmark Ultra, using a prediluted Ventana mouse primary antibody. Antigen retrieval was for 64 min using Ventana retrieval solution CC1. The primary antibody was incubated for 32 min at 36 degrees C. The Ventana Optiview DAB detection system was used with Ventana Haematoxylin II as the nuclear counterstain. The alpha-smooth muscle actin (aSMA) staining was completed using a rabbit polyclonal aSMA antibody (1:500, Abcam ab5694) and Keratin 20 using a rabbit polyclonal (1:200, Sigma SAB4502249). Antigen retrieval was performed in Vector Antigen Unmasking solution (Vector Laboratories, H-3300) and carried out in a Biocare Medical pressurised decloaker (Metagene). A goat anti-rabbit HRP secondary antibody (1:400, Life Technologies A24537) followed by DAB (BioLegend 926901) was used to visualise signal. Histological evaluation of mouse tumours was undertaken by a pathologist with expertise in human serrated pathway lesions. In line with current WHO standards mucinous adenocarcinoma was not graded (low versus high), and mucinous adenocarcinoma was defined as >50% mucinous differentiation [7]. Tumour budding was evaluated using H&E stained tumour sections in a hotspot area of $0.785mm^2$ by counting the number of buds as described for diagnostic practice [8]. 0 = Absent; 1-4 = low level; intermediate 5-9 and 10 or more is high level.

Nucleic acid analysis. Total RNA and gDNA was extracted from organoids, tumours and normal control colon tissue using AllPrep DNA/RNA Mini or RNeasy Mini Kit (QIAGEN) or UltraClean® Tissue & Cells DNA Isolation Kit (Mo Bio Laboratories). For qRT-PCR, 300-1000 ng RNA was reverse transcribed to cDNA using Superscript IV according to the manufacturer's instructions (Thermo Scientific). cDNA was diluted 1:5 and level of transcripts of interest evaluated using Primer:probe assays (IDT) with KAPA PROBE FAST qPCR Master Mix (KAPABiosystems) master mix and run on a QuantStudio 7 Flex Real-Time PCR System (Thermo Scientific). For RNAseq analysis, 1ug of total RNA was used for sequencing library preparation using a BIOO stranded Poly-A kit (PerkinElmer) according to the manufacturer's instructions and sequenced on Illumina HiSeq to generate 1 x 100bp single-end reads.

RNAseq data processing. Images generated by HiSeq™ 2500 were converted into nucleotide sequences by a base calling pipeline and stored in fastq.gz format, and examined QC plots with FastQC version 0.11.3. Raw reads with low quality were removed using Trim Galore prior to analysing the data. The criteria of removing low quality reads were set as: quality Phred score less than 28, reads contains adaptor sequences. FastQC was performed again, all low quality reads were removed. All the subsequent analyses were based on trimmed reads, which were

then mapped to reference mouse genome mm10 GRCm38, using STAR 2.4.2a modified. No more than 1 base mismatch were allowed. Only uniquely mapped reads were retained. quantMode was enabled to generate gene level quantification. The counts files were then merged into a table for downstream differential expression analysis. All data preprocessing were completed in shell HPC command line environment.

Differential expression analysis. Differentially expressed genes (DEG) were analysed using the glmFit() glmLRT() function in R Bioconductor edgeR Package version 3.16.5 [9]. A negative binomial generalized log-linear model was employed, with Benjamini-Hochberg correction using false discovery rate (FDR) of *P* values. Hierarchical clustering was performed on DEGs only, using Euclidean distance and complete linkage clustering option with Heatmap() function in ComplexHeatmap R package version 1.12.0 [1]. MDS plots were generated using the plotMDS() function in Limma R package version 3.30.13 [10]. Differentially expressed genes were defined as FDR < 0.05, and absolute value of Log_2 (fold change) > 2.

Gene set enrichment analysis (GSEA). GSEA methodology was developed using guidelines for RNAseq datasets [11] . Transcript counts generated from STAR aligner were sub-grouped into files with pairwise comparison groups of normal colon to tumour. The PreprocessReadCounts (V1 Beta) module of GenePattern was used to normalise and transform counts. Briefly, lowly expressed genes were removed from the analysis if they had <1 read/million in four of the samples, as four is the smallest group of biological replicates [12]. The remaining data was then normalised using Trimmed Mean of M-values [13]. Mean-variance transformation to approximate a normal distribution was performed subsequently using voom [10]. Gene set enrichment analyses were performed using online module GSEA (v18).

Microsatellite instability. MSI status was assessed using a panel of five markers (*m-Bat-26, mBat-67, m-Bat-37, GA29, TG27*-see Supplementary Methods Table 1 for sequence) [14, 15, 16]. These included 3 mononucleotide and 2 dinucleotide repeats. Fragment sizes were assayed using PCR coupled with fluorescent detection on an ABI3100 Capillary Sequencer. Differences in fragment sizes were assessed using the GeneMarker software.

Supplementary methods table 1-Microsatellite marker primer information

Marker	Type	Annealing Tm	Forward Primer	Reverse Primer	Reference
mBat-26	Mononucleotide	59	TCACCATCCATTGCACAGTT	CTGCGAGAAGGTACTCACCC	Bacher et al, 2005
mBat-37	Mononucleotide	59	TCTGCCCAAACGTGCTTAAT	CCTGCCTGGGCTAAAATAGA	Bacher et al, 2005
mBat-67	Mononucleotide	59	CCGACTGCTCTTCCGAAGGTC	TTGCCCATTTATCATCTAGTTCAT	Bacher et al, 2005
TG27	Dinucleotide	60	GGATCACTCGATGTACGGCTACTC	CCAGGCAGGCAAAGCATTTAT	Kabbarah et al, 2003
GA29	Dinucleotide	59	CAGGAGGTCAAGGTCATCCTAAG	CCACCATGGTAGGAGCTTGCTA	Kabbarah et al, 2003

Supplementary methods table 2- gRNA sequence

Target gene	gRNA sequence
Rnf43	CCACCAGGAGGTACCAAGCCGGC
Znrf3	GGGTCATCCCTTGTACTCATCGG
Tgfbr2	CCTGTGGCCGCTGCATATCGTCC
p16 Ink4a	CCCAACGCCCCGAACTCTTTCGG
Mlh1	TAGTGAACCGCATAGCGGCGGGG

Supplementary Methods References

1 Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;**32**:2847-9.

2 Hothorn T, Lausen B. On the exact distribution of maximally selected rank statistics. Computational Statistics & Data Analysis 2003;**43**:121-37.

3 Onuma K, Ochiai M, Orihashi K, Takahashi M, Imai T, Nakagama H*, et al.* Genetic reconstitution of tumorigenesis in primary intestinal cells. Proc Natl Acad Sci U S A 2013;**110**:11127-32.

4 Schwank G, Andersson-Rolf A, Koo BK, Sasaki N, Clevers H. Generation of BAC transgenic epithelial organoids. PLoS One 2013;**8**:e76871.

5 Roper J, Tammela T, Cetinbas NM, Akkad A, Roghanian A, Rickelt S*, et al.* In vivo genome editing and organoid transplantation models of colorectal cancer and metastasis. Nat Biotechnol 2017;**35**:569-76.

6 Becker C, Fantini MC, Wirtz S, Nikolaev A, Kiesslich R, Lehr HA*, et al.* In vivo imaging of colitis and colon cancer development in mice using high resolution chromoendoscopy. Gut 2005;**54**:950-4.

7 Bosman FT, Carneiro, F., Hruban, R.H., Theise, N.D. WHO Classification of Tumours of the Digestive System, 2010.

8 Lugli A, Kirsch R, Ajioka Y, Bosman F, Cathomas G, Dawson H*, et al.* Recommendations for reporting tumor budding in colorectal cancer based on the International Tumor Budding Consensus Conference (ITBCC) 2016. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc 2017;**30**:1299-311.

9 McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 2012;**40**:4288-97.

10 Smyth GK. limma: Linear Models for Microarray Data. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S, eds. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. New York, NY: Springer New York, 2005:397-420.

11 Rahmatallah Y, Emmert-Streib F, Glazko G. Gene set analysis approaches for RNA-seq data: performance evaluation and application guideline. Brief Bioinform 2016;**17**:393-407.

12 Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W*, et al.* Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc 2013;**8**:1765-86.

13 Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 2010;**11**:R25.

14 Bacher JW, Abdel Megid WM, Kent-First MG, Halberg RB. Use of mononucleotide repeat markers for detection of microsatellite instability in mouse tumors. Mol Carcinog 2005;**44**:285-92.
15 Kabbarah O, Mallon MA, Pfeifer JD, Edelmann W, Kucherlapati R, Goodfellow PJ. A panel of repeat markers for detection of microsatellite instability in murine tumors. Mol Carcinog 2003;**38**:155-9.
16 Bond CEL, C.; Kawamata, F.; McKeone,D.M.; Fernando, W.; Jamieson, S.; Pearson, S.; Kane, A.; Woods, S.L.; Lannagan, T.R.M.; Somashekar, R.; Lee, Y.; Dumenil, T.; Spring, K.J.; Borowsky, J.; Fennell, L.; Bettington, M.; Lee, J.; Worthley, D.L.; Leggett, B.A.; Whitehall, V.L.J. Oncogenic BRAF Mutation Induces DNA Methylation Changes in a Murine Model for Human Serrated Colorectal Neoplasia. Epigenetics 2017;**In Press**.