SI Text

SI Methods

ChIP-qPCR. ChIP assays were performed as described previously with minor modifications (1). Briefly, 6 replicates of SC and CC were plated into 3D type-1 collagen at 20,000 cells/well and cultured for 2 weeks. Collagen plugs were then snap frozen in liquid nitrogen and stored at -80 °C. For ChIP, plugs were pooled together and crushed into a fine powder in liquid nitrogen with mortar and pestle and the powder immediately transferred to pre-warmed PBS containing 1% paraformaldehyde. Suspensions were rotated at 37 °C for 10 min, and the fixative was quenched by transfer of samples to ice and addition of glycine to a final concentration of 125 mM. Samples were then taken through the remainder of the ChIP assay using 5 µg of the indicated antibodies (Abcam: H3K9Me3, H3K9Me2, H3K27Ac, H3K36Me3; Millipore: H3K27Me3) and Dynabeads protein A or G (Life Technologies) to capture antigen-antibody complexes. Two hundred and fifty picograms of paired input and immunoprecipitated (IP) DNA were amplified in duplicate by real-time PCR (Roche LightCycler), and IP:Input fold-enrichments were calculated and plotted as the mean ± SD.

Exome sequencing. CC and SC 3D cultures were collagenase treated for 1 h at 37 °C; colonies were pelleted, washed, and DNA was isolated using the GeneJET Genomic DNA Purification Kit (Fisher Scientific, Suwanee, GA). Samples were submitted to HudsonAlpha Institute for Biotechnology Genomic Services Lab for Illumina methyl 450 analysis. One hundred ng of genomic DNA was used to produce exome-captured sequencing libraries using NimbleGen SeqCap EZ v3.0 following manufacturer's instructions applying pre-capture multiplexing. The exome-captured sequencing libraries were quality-controlled using an Agilent 2100 Bioanalyzer and quantified using the KAPA qPCR NGS Library Quantification Kit prior to cluster generation on an Illumina cBot. Sequence data (101 bp PE, 100X coverage) was generated on Illumina's HiSeq 2500 platform. Reads were mapped to the hg19 reference genome with BWA (version 0.7.5a) and sorted and indexed with SAMtools (2). Duplicated reads were marked by Picard (http://picard.sourceforge.net/). SNVs and INDELs were called simultaneously on HCA-7, SC, and CC samples by SAMtools with base quality≥30, reads with mapping quality ≥30, and mapping quality downgrading coefficient of 50 (2). SNVs and INDELs with strand bias $p<0.01$, or base quality bias $p<0.01$, or mapping quality bias $p<0.01$, or end distance bias $p<0.01$ were filtered out. Furthermore, SNVs within 3 bp around a gap were removed. SNVs and INDELs were annotated, and their effects were predicted by snpEff and snpSift. VarScan 2 was used to perform copy number analysis using its recommended workflow (3). That is, raw copy numbers variations between SC and CC samples were first obtained by running “copynumber” function. To account for overall differences in the amount of sequencing depth between SC and CC samples, a data ratio was included based on the uniquely mapped reads and the read length in the SC and CC samples. The candidate CNV regions were adjusted for GC content, filtered, and categorized as gain, loss, and neutral by the “CopyCaller” function. Finally, CNV regions were smoothed and segmented by the DNAcopy R package.

Karyotypic Analysis. Mitotic arrest in three-day old CC and SC plastic cultures was induced with colchicine (0.5 µg/ml, 20 min), followed by hypotonic treatment (0.075 M KCl, 15 min, 37 °C) and fixation with methanol/acetic acid (3:1). The cell suspension was dropped onto cold slides and Giemsa stained after short-term trypsin digestion. The karyotype was described according to the International System for Human Cytogenetic Nomenclature (ISCN). *Chromosome microarray analysis:* Genomic DNA (gDNA) was extracted using the Gentra Puregene Blood Kit (Qiagen, Valencia, CA) from CC and SC cultures. CytoScan 750K (Affymetrix, Santa Clara, CA) was used to assess gDNAs of CC and SC. Data were analyzed

using the Chromosome Analysis Suite (ChAS) software 3.0 (Affymetrix), considering at least 25 markers for losses and 50 markers for gains. The CNVs detected were compared with the Database of Genomic Variants (DGV, http://dgv.tcag.ca/dgv/app/home, version: July 2015).

SI Figures

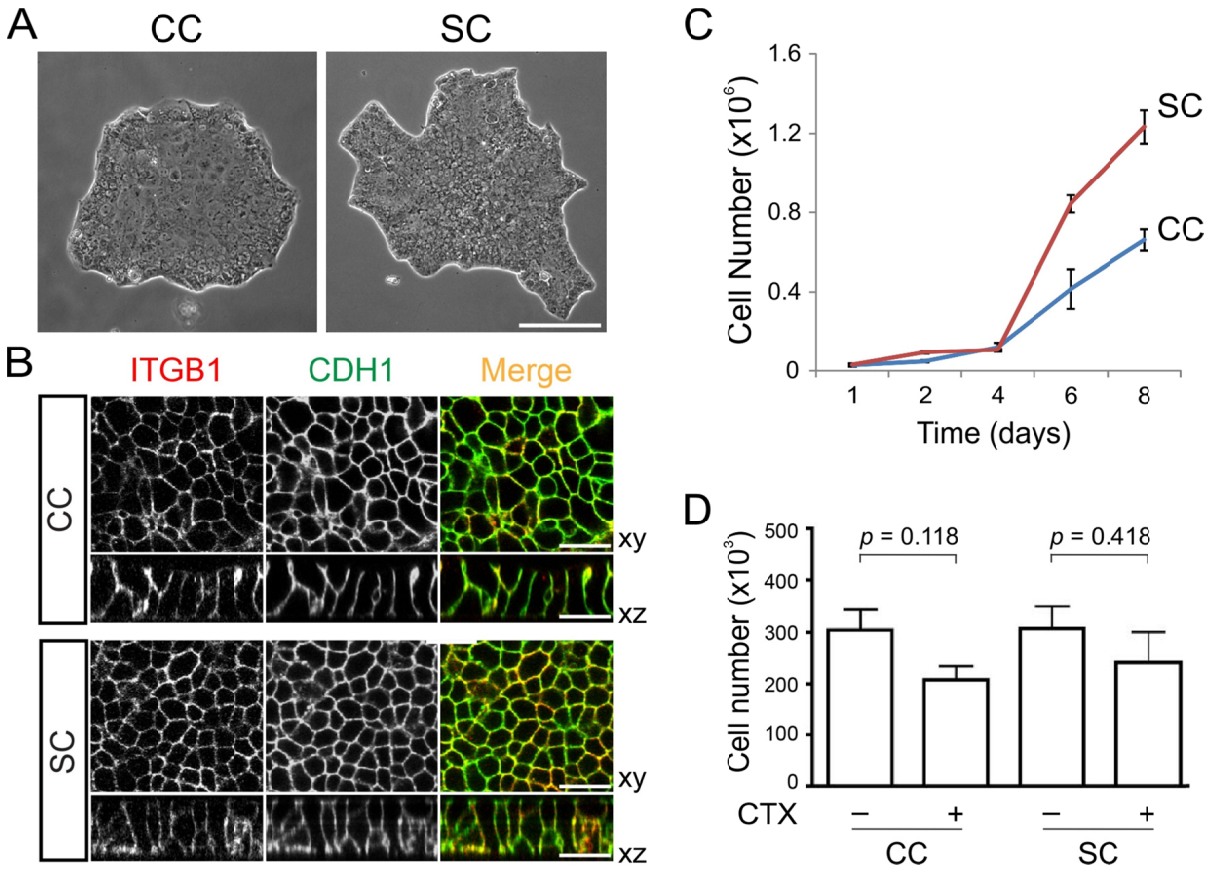


Fig. S1. CC and SC characteristics in 2D and Transwell cultures. (A) Representative DIC images of CC and SC cultured on plastic. (B) Confocal micrographs of CC and SC cultured on Transwell filters. ITGB1 and CDH1 localized to lateral membrane of polarized cells. (C) Fifty thousand cells were seeded on plastic, counted at indicated time points, and plotted as mean ± SEM. (D) Cetuximab sensitivity of CC and SC in 2D cultures. Fifty thousand CC and SC were seeded on plastic in the presence or absence of 3 µg/ml cetuximab (CTX). Cell counts at day six were plotted as mean ± SEM (n=3). *p*-values were determined using a two-tailed unpaired *t*-test.

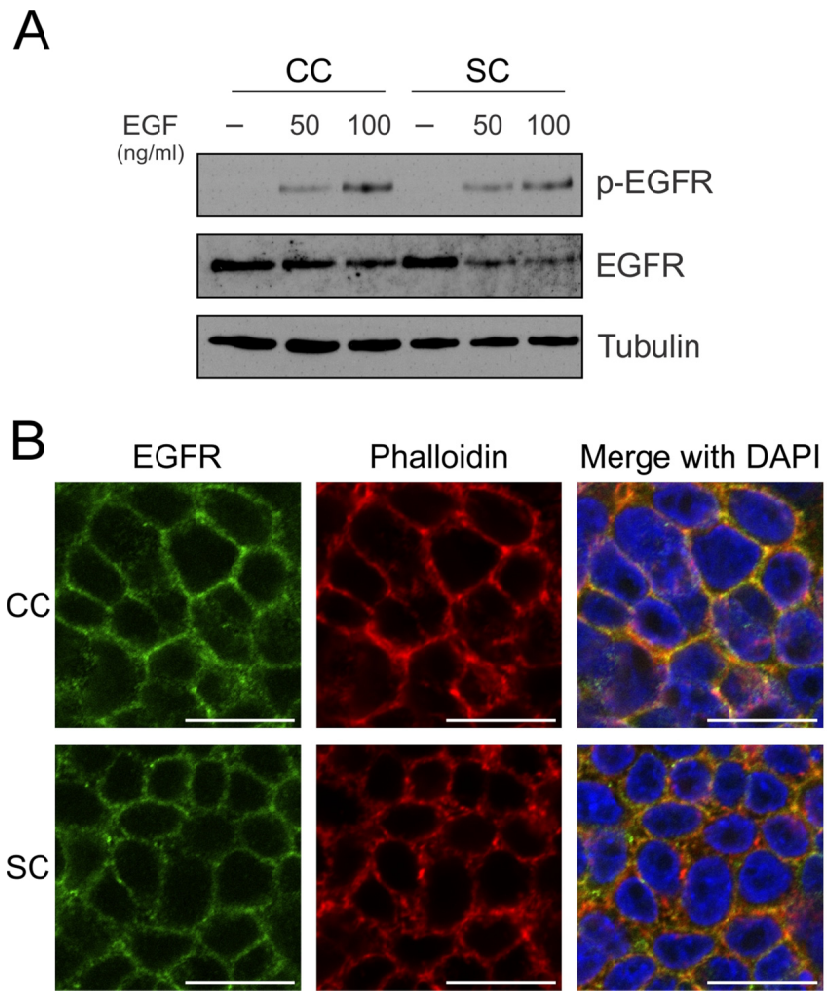


Fig. S2. Equivalent EGFR activation and cell-surface EGFR in CC and SC. (A) CC and SC cells cultured on plastic were stimulated with 50 or 100 ng/ml EGF for 5 min. Cells were then lysed and immunoprecipated for EGFR and subsequently probed with phospho-tyrosine, 4G10 (p-EGFR) and EGFR by immunoblotting. Tubulin immunoblot shows comparable loading between CC and SC. (B) CC and SC cells grown on Transwell filters were incubated with Alexa-488-labeled monoclonal antibody (C225) directed against the extracellular domain (EGFR) at 4 °C and then fixed, permeabilized, and stained with Phalloidin and DAPI. Scale bars: 10 μm.

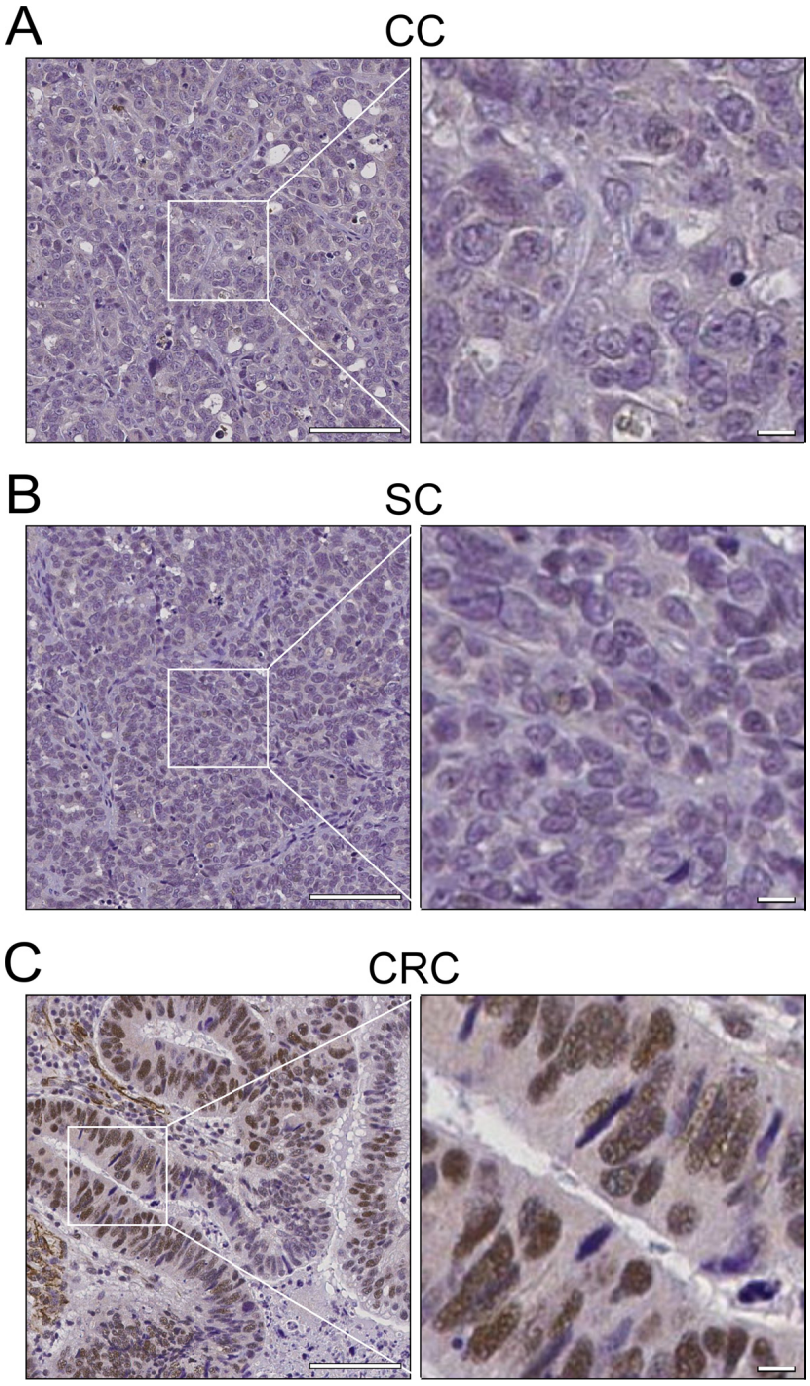


Fig. S3. Representative MLH1 immunoreactivity in CC and SC xenografts. (A) MLH1 staining in CC. (B) MLH1 staining in SC. (C) MLH1 staining in a moderately differentiated CRC. Scale bars: left panels, 100 μm; magnified views on right, 10 μm.

A

CC	Hetero	Homo	Total
SNVs	1,555	36	1,591
Misssense SNVs (gene)	180 (156)	6 (7)	186 (163)
INDELs	151	3	154
Missense INDELs (gene)	5 (5)	0 (0)	5 (5)

SC	Hetero	Homo	Total
SNVs	1,936	43	1,979
Misssense SNVs (gene)	188 (172)	14 (14)	202 (186)
INDELs	143	13	156
Missense INDELs (gene)	3 (3)	1 (1)	4 (4)

B

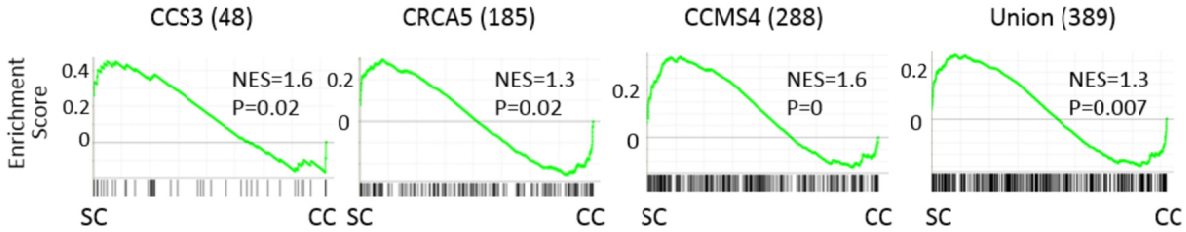


Fig. S4. (A) Unique SNVs/INDELs in CC (left) and SC (right). The table lists the total number of SNVs/INDELs, the number of heterozygous and homozygous SNVs/INDELs, and the number of missense SNVs/INDELs. (B) Gene set enrichment analysis (GSEA) of gene signatures associated with poor prognosis subtypes linked to CC and SC gene expression from RNA isolated from 15 day 3D cultures.

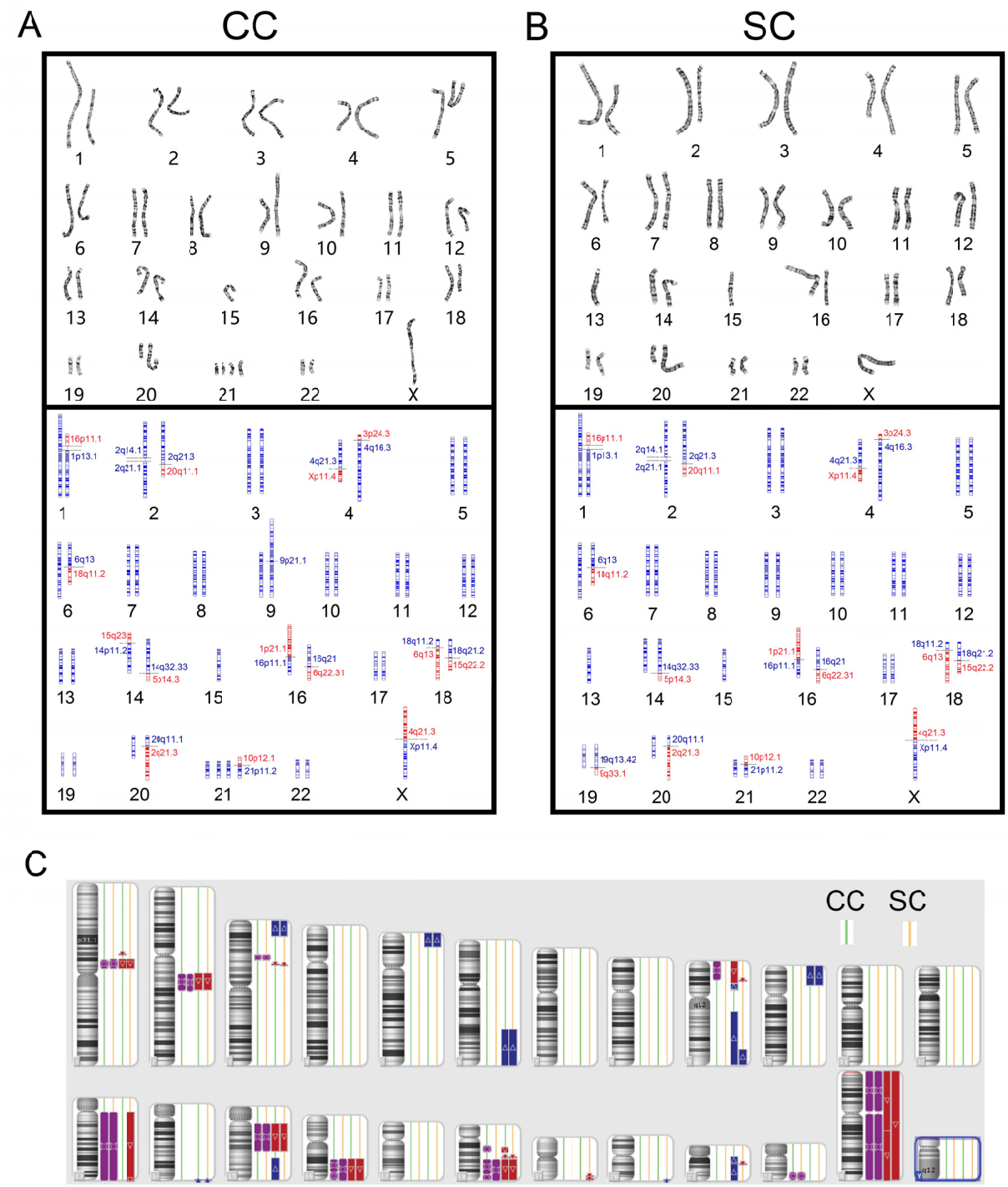


Fig. S5. Chromosomal and digital karyotyping analysis of CC and SC. (A) CC karyogram in the top panel shows near diploid, 46 chromosomes with -X, -15, +21, and +21. (B) SC karyogram in the top panel shows 43 chromosomes -X, -13, and -15. The lower panels in A and B highlight chromosomal rearrangements in CC and SC, respectively. (C) In this schematic of CC and SC chromosomal microarray, each chromosome is associated with two pairs of vertical lines (CC,

green; SC, orange). The first pair on the right depicts loss or gain of chromosomal regions in red and blue, respectively. The second pair depicts regions of loss of heterozygosity in purple. The genomic differences between CC and SC are mainly observed on chromosome 9, 13, 15, 18, and 21. CC chromosome 9 shows isodicentric rearrangement. One copy of chromosome 13 is lost in SC but is homozygously duplicated in CC. CC have an extra fragment of the long arm of chromosome 15 added to the short arm of chromosome 14. CC have an approximately 7 Mb deletion on 18q12.1. CC have two more copies of chromosome 21 than SC. The results are further summarized in Table S3.

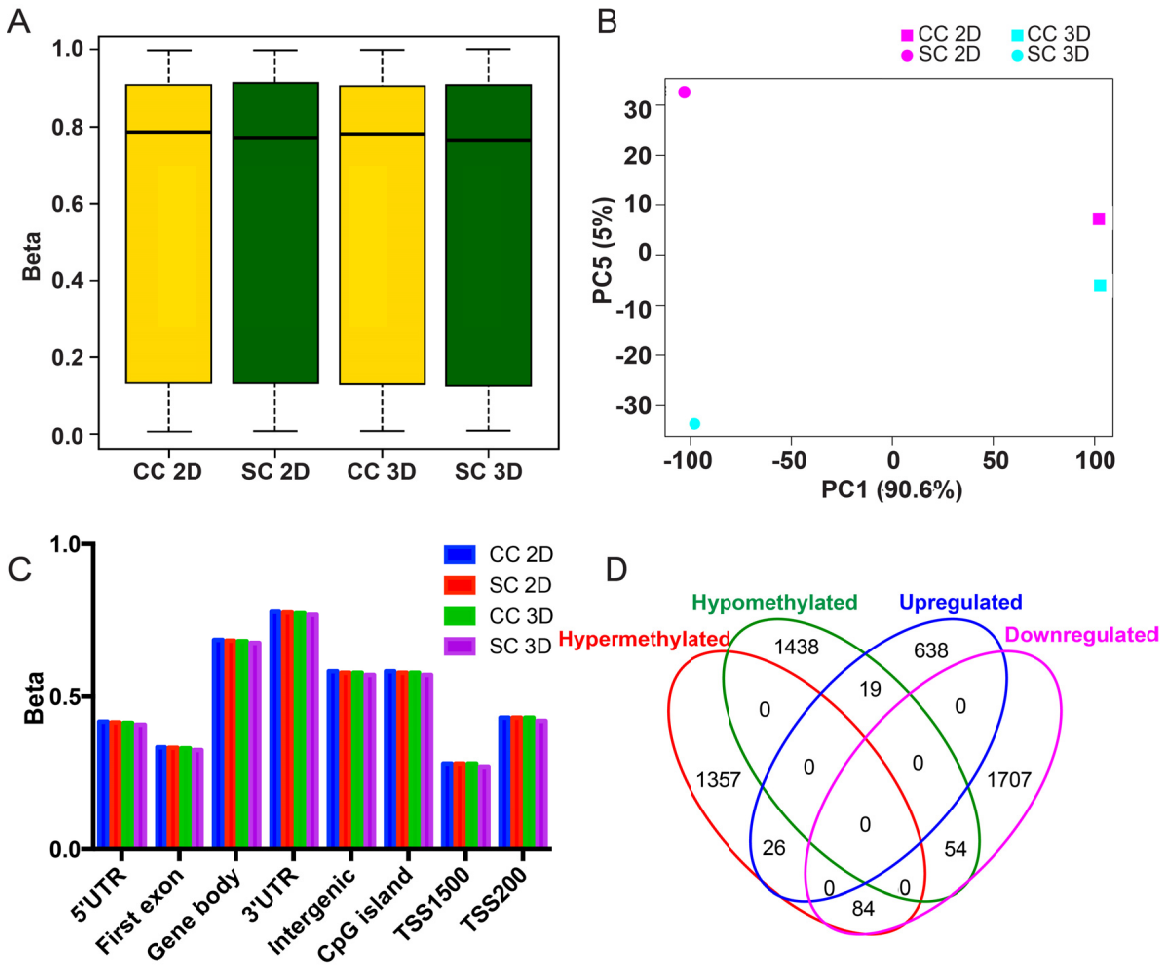


Fig. S6. Characterization of DNA methylation in CC and SC. (A) The distribution of DNA methylation values (beta) in CC (yellow) and SC (green) under 2D plastic culture conditions (2D) or in 3D culture with type-1 collagen (3D). (B) Principal component analysis (PC) of DNA methylation status of CC and SC in 2D or 3D culture. (C) Average DNA methylation levels at different genomic regions for SC and CC in 2D or 3D culture (UTR: translated region; TSS200: 200 nt upstream of transcriptional start site [TSS]; TSS1500: 200-1500 nt upstream of TSS. (D) Venn diagram showing correlation analysis between genes with differentially methylated regions and gene expression levels in CC and SC cells.

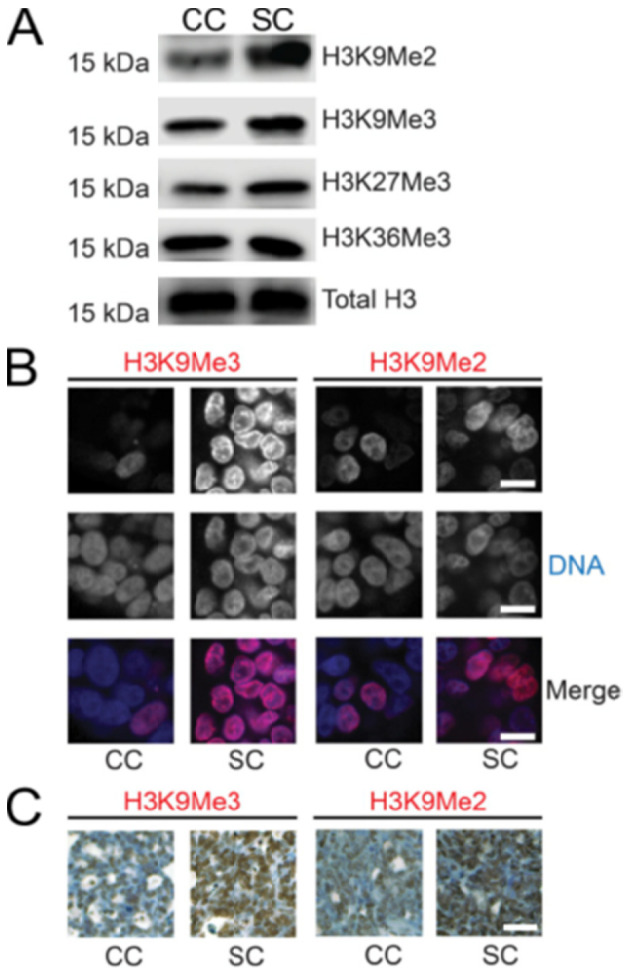
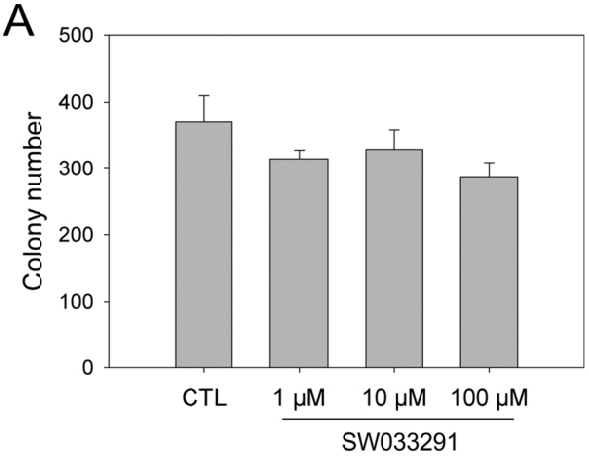
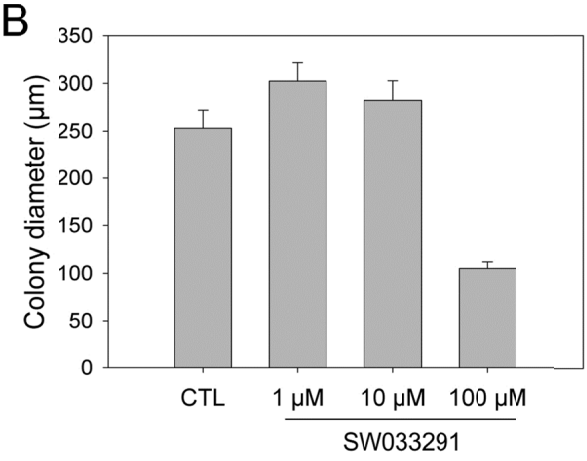


Fig. S7. Comparison of histone methylation between CC and SC. (A) CC and SC cultured for 15 days in 3D in type-1 collagen were lysed and subjected to immunoblotting with the indicated antibodies on the right. Total H3 levels confirm equal loading. (B) Global histone methylation in CC and SC 3D cultures stained with indicated antibodies are displayed in red and with DAPI in blue. Scale bars: 10 μm. (C) Global histone methylation in CC and SC athymic nude mouse xenografts as assessed by immunohistochemical staining for H3K9Me3 and H3K9Me2. Scale bars: 50 μm.

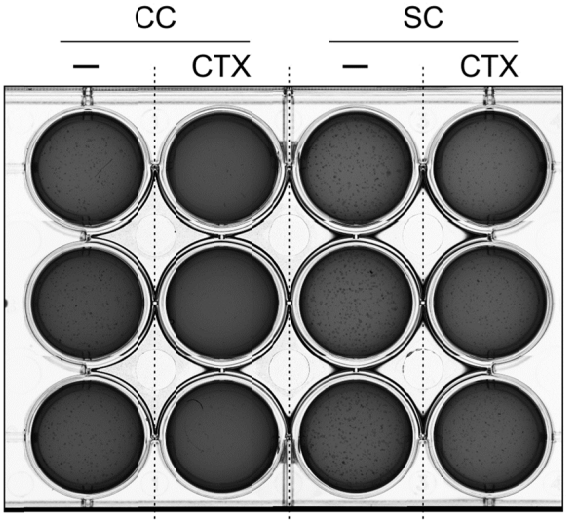
A



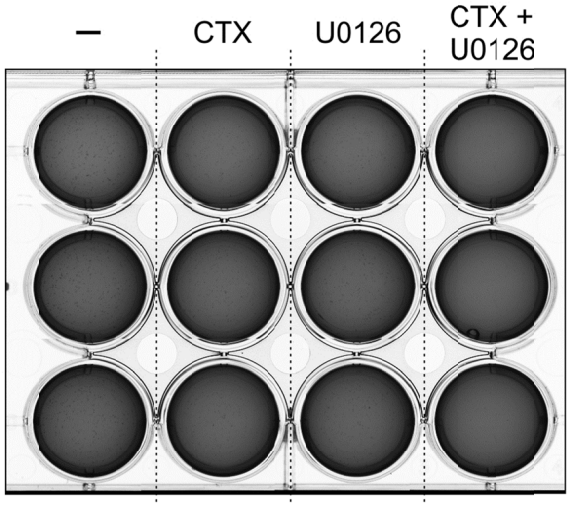
B



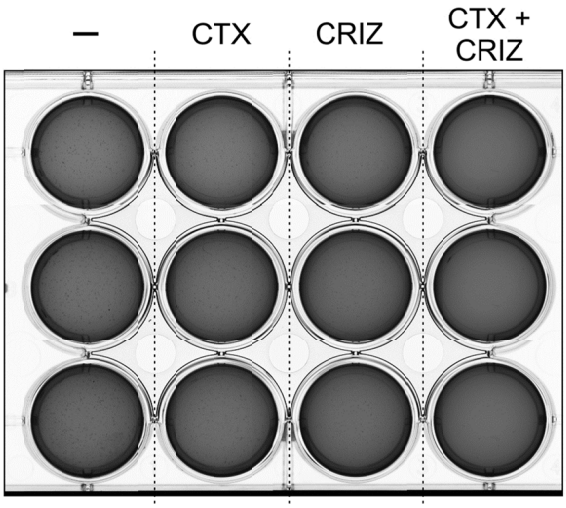
C: Raw data Fig. 1F



D: Raw data Fig. 2B



E: Raw data Fig. 2D



F: Raw data Fig. S10A,B

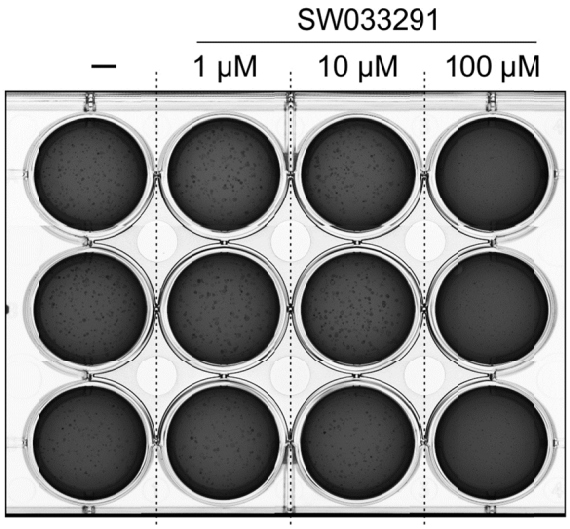


Fig. S8: Effect of HPGD inhibitor on CC cells in 3D. Two thousand CC cells were cultured in type-1 collagen for 17 days. Fresh medium was added with or without the HPGD inhibitor, SW033291, at indicated concentrations every two to three days. Colony count (A) and diameter (B) were determined using a GelCount™ plate reader. Results are plotted as mean ± SEM. (C-F) Raw GelCount™ plate reader images of the collagen cultures used for quantification of colony number and size throughout the manuscript; each subfigure is labeled with matching figure number in the manuscript.

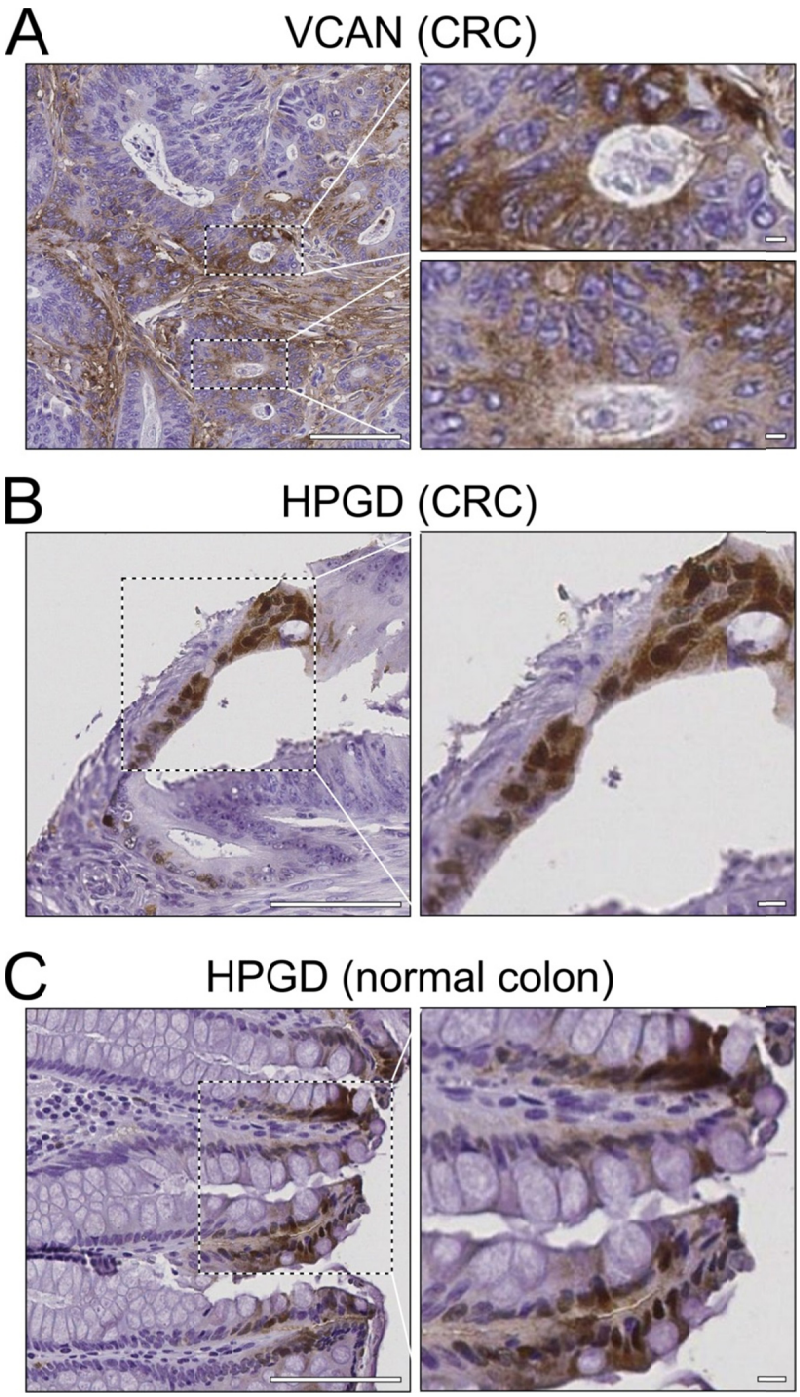


Fig. S9. Representative VCAN and HPGD immunoreactivity in colorectal tissue. (A) VCAN staining in CRC. Note epithelial VCAN staining. (B) HPGD staining in CRC. (C) HPGD staining in normal colon; staining is confined to differentiated luminal compartment. Scale bars: left panels, 100 µm; magnified views on right, 10 µm.

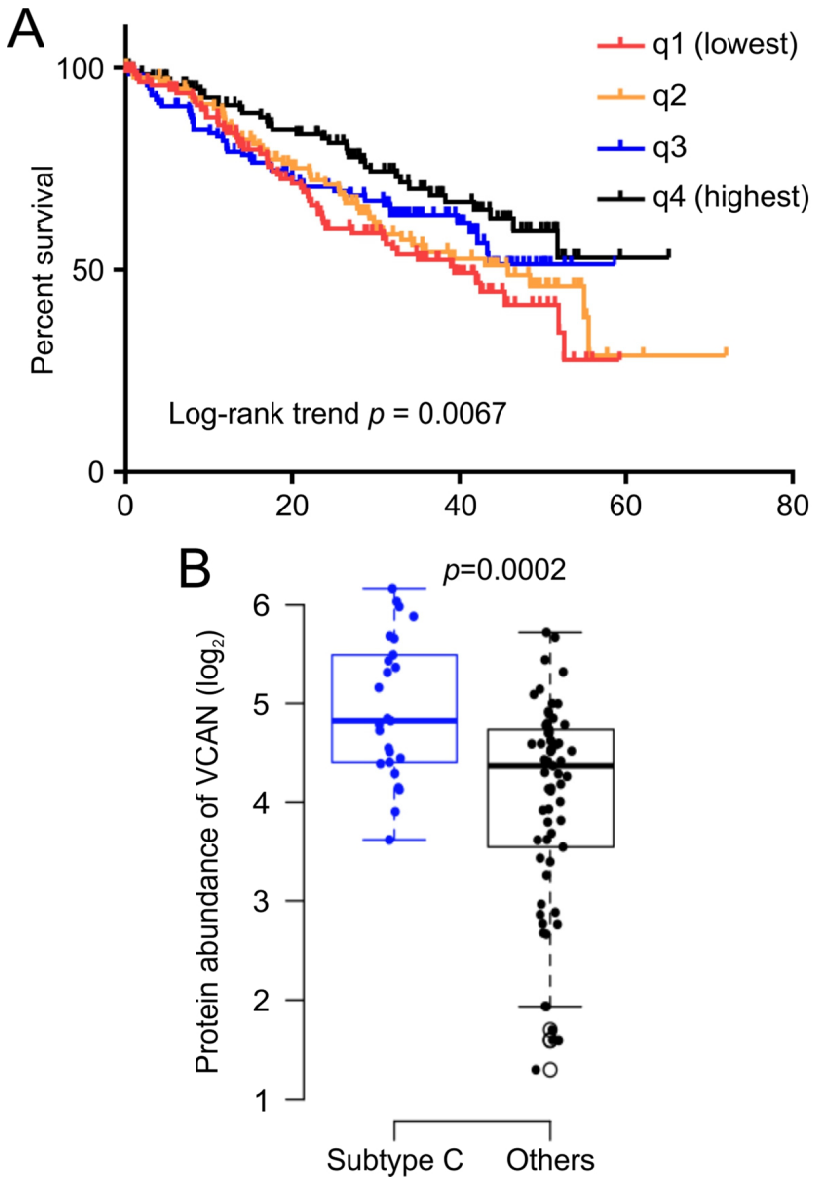


Fig. S10. (A) Kaplan-Meier plots of overall survival of 458 CRCs stratified by quartiles based on HPGD mRNA expression. The samples were from the Moffitt 468 dataset; 10 samples without appropriate microarray data were excluded (5, 6). Comparison among four quartiles was performed with the log-rank test for trend in survival. (B) VCAN is upregulated in human epithelial-to-mesenchymal-like subtype C compared to the other four subtypes based on a proteogenomic-based CRC classification proposed by B. Zhang *et al* (4).

Table S1. Mutations in CC and SC.
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=cxqzmcumhnehvcx&acc=GSE76352

Table S2. Comparison of CC and SC gene expression in 3D.
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=cxqzmcumhnehvcx&acc=GSE76210

Chromosome	CC	SC
der(X)t(X;4)(p11.4;q21.3)	1	1
1	1	1
der(1)t(1;16)(p13.1;p11.1)	1	1
del(2)(q14.1q21.1)	1	1
der(2)t(2;20)(q21.3;q11.1)	1	1
3	2	2
der(4)t(3;4)(p24.3;p16.3)	1	1
der(4)t(X;4)(p11.4;q21.3)	1	1
5	2	2
6	1	1
der(6)t(6;18)(q16.2;q21.1)	1	1
7	2	2
8	2	2
9	1	2
idic(9)(p21.1)	1	-
10	2	2
11	2	2
12	2	2
13	2	1
14	-	1
der(14)t(14;15)(p11.2;q23)	1	-
der(14)t(5;14)(p14.3;q32.33)	1	1
15	1	1
der(16)t(1;16)(p21.1;p11.1)	1	1
der(16)t(6;16)(q22.31;q21)	1	1
17	2	2
der(18)t(6;18)(q16.2;q21.1)	1	1
der(18)t(15;18)(q22.2;?q21.2)	1	1
19	2	1
der(19)t(9;19)(q33.1;q13.42)	-	1
20	1	1
der(20)t(2;20)(q21.3;q11.1)	1	1
21	3	1
der(21)t(10;21)(p12.1;p11.2)	1	1
22	2	2
mar	1	1

Total	43	46

Table S3. Karyotypic comparison of CC and SC. In compliance with International System for Human Cytogenetic Nomenclature (ISCN), the following abbreviations are used: der = derivative chromosome; t = translocation; idic = isodicentric chromosome; mar = marker chromosome.

References

1. McDonald OG, Wamhoff BR, Hoofnagle MH, & Owens GK (2006) Control of SRF binding to CArG box chromatin regulates smooth muscle gene expression in vivo. *J Clin Invest* 116(1):36-48.
2. Li H*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
3. Koboldt DC*, et al.* (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22(3):568-576.
4. Zhang B*, et al.* (2014) Proteogenomic characterization of human colon and rectal cancer. *Nature* 513(7518):382-387.
5. Schell MJ*, et al.* (2016) A Composite Gene Expression Signature Optimizes Prediction of Colorectal Cancer Metastasis and Outcome. *Clinical cancer research : an official journal of the American Association for Cancer Research* 22(3):734-745.
6. Schell MJ*, et al.* (2016) A multi-gene mutation classification of 468 colorectal cancers reveals a prognostic role for APC. *Nature Communications* In presss.