SUPPLEMENTAL MOVIE, TABLE LEGENDS, METHODS and FIGURES

Supplemental Movie 1. SW620 Pool of cells infected with H2BeGFP lentivirus and filmed upon a DOX pulse-chase. H2BeGFP accumulated in the nucleus was diluted every cell division, whereas non-dividing cells retained the label.

Supplemental Table 1. Cancer cell models used to study SCCC. Tumor type, colorectal (CRC), melanoma (MEL) or glioblastoma (GBM), and culture system, cell line or patient-derived xenograft (PDX) in a single-cell clone or pool are indicated for each model. Basic information on patients is also provided. Models were molecularly characterized with targeted sequencing (Amplicon Seq) as previously described or by Exome sequencing (1). Information on microsatellite instability is also provided for CRC models. MSS: Microsatellite stable; MSI: Microsatellite instable. Proportion of cells transduced with the H2BeGFP lentivirus is indicated for all models upon fresh infection alone or combined with FACS for enriching in positive cells. The exact position of each insertion site of the pSIN-TRE-H2BeGFP-rtTA2 lentivirus in the genome of all single-cell derived clones is shown. We indicate the number of replicates analyzed by microarray for each cancer model. n. d.: not determined.

Supplemental Table 2. Genes differentially expressed in SCCC versus RCCC across different tumor types define a PanCancer-SCCC signature. PanCancer (PanC)-SCCC expression profile was generated from the expression profiles of 29 replicates of SCCC and their corresponding 29 paired RCCC isolated from colorectal (CRC-SW1222 and CRC-T70), melanoma (MEL-MMLN9 and MEL-MMPG3) and glioblastoma (GBM-e216 and GBM-e225) models evaluated by microarray. From this list, the 100 genes higher and lower expressed in SCCC versus RCCC were used to build the PanC-SCCC signature. Combined list of genes differentially expressed in SCCC versus RCCC isolated from CRC, MEL and GBM models is shown. FC, fold change; FDR, false discovery rate.

The CRC-SCCC signature correspond to the median expression of top 100 genes upregulated minus median expression of 100 genes downregulated in CRC samples. Univariate and multivariate Cox survival analyses of disease-free survival (DFS) according to CRC-SCCC signature score in a cohort of 151 high-risk stage II or stage III tumors from CRC patients that received adjuvant chemotherapy. HR, hazard ratio; CI, confidence interval.

Supplemental Table 3. TET2 determines global gene expression in SCCC. Combined list of genes differentially expressed in SCCC versus RCCC isolated from CRC-SW1222 knock-down control (shCTRL) and *TET2* (sh*TET2*) cells is shown. FC, fold change; FDR, false discovery rate. List of 121 genes included in TET2-SCCC signature. Univariate and multivariate Cox survival analyses of disease-free survival (DFS) according to TET2-SCCC signature in a cohort of 151 high-risk stage II or stage III tumors from CRC patients that received adjuvant chemotherapy. HR, hazard ratio; CI, confidence interval.

Supplemental Table 4. Landscape of 5hmC across multiple tumor types. Histopathological features of patients analyzed for 5hmC levels are shown. 19 different types of cancer were evaluated. 5hmC levels are indicated as HIGH, LOW or NEG based on immunohistochemistry and immunofluorescence scores. In the case of CRC patients, disease-free survival (DFS) is shown as months from diagnosis to relapse or death due to any cause. This cohort include CRC patients from VHIO cohort (n = 61) and from the HD-PC-TMA (High Density PanCancer Tissue Microarray) (n = 47). Multivariate analysis was calculated including all CRC patients (n = 108) in the analysis regardless of treatment type. Univariate analysis was calculated for both, all (n = 108) and treated (n = 87) CRC patients. n. r., patients who did not relapse. Adjuvant chemotherapy used is indicated: 5FU, 5-fluorouracil; CAPE, Capecitabine; FOLFOX, Folic Acid + 5-fluorouracil + Oxaliplatin. In the other cancer types, overall survival (OS) is shown as months from diagnosis to death due to any cause. n.a., non-annotated. n.d., not determined. HR, hazard ratio. CI,

confidence interval. N, none. Y, yes. Log-rank Mantel Cox tests were utilized for analyzing significance in OS and DFS.

Supplemental Table 5. 5hmC positive cells and content are higher in liver metastases than in paired CRC primary tumors. Histopathological features and mutational status of clinically relevant oncogenes in paired primary tumors and liver metastases, from 197 CRC patients analyzed for 5hmC and Ki67 levels, are shown. 5hmC and Ki67 levels are shown as percentage of positive tumor cells measured by immunofluorescence and immunohistochemistry respectively in primary tumors and paired metastases. Synchronous (s) and metachronous (m) cases are indicated. n. a., non-annotated. n.d., not determined.

Supplemental Table 6. Primers and antibodies. List of DNA oligos used for qPCR, shRNA, cloning *H2BeGFP* cDNA into a lentiviral vector, CRISPR sgRNA plasmids generation, *TET2* CRISPR screening and lentiviral integration screening. Details about all primary and secondary antibodies used for immunofluorescence, immunohistochemistry and WB are also shown.

SUPPLEMENTAL METHODS

Cell lines

Colorectal cancer (CRC) cell lines SW480, SW620, and DLD1 were purchased from ATCC. CRC-SW480-H2BeGFP-C2, CRC-SW620-H2BeGFP-C3, and CRC-DLD1-H2BeGFP-C7 single cell clones were obtained by dilution of the corresponding parental pools of infected cells. All cell lines were transduced with lentiviruses expressing H2BeGFP protein (pSIN-TRE-H2BeGFP-hPGK-rtTA2-inducible lentivirus described in Methods). Cell lines were maintained in Dulbecco's modified eagle medium (DMEM; Biowest) supplemented with 10% Fetal bovine serum (FBS) (complete DMEM) and 1% penicillin/streptomycin (P/S) (Life Technologies).

Cell Cultures

Colorectal (CRC) SW1222 cell line was maintained in Dulbecco's modified eagle medium (DMEM; Biowest) supplemented with 10% Fetal bovine serum (FBS) (complete DMEM) and 1% P/S (Life Technologies). Minitumors (MTs) embedded in Matrigel from CRC-SW1222 cell line were generated from single-cells resuspended in complete DMEM and mixed 1:1 with Corning Matrigel Basement Membrane Matrix (BD Bioscience). Drops of 25 µl of the mixture were incubated for 30 minutes at 37ºC without medium to allow the matrix solidification. Then, complete DMEM was added and changed twice weekly. MTs embedded in Matrigel from CRC-T70 patient-derived cells were mixed 1:1 with Matrigel and drops of 50 µl were seeded as described above. Finally, Organoids Medium (Advanced DMEM/F12, 5% FBS, 2mM Glutamax, B27 Supplement without vitamin A 1X, N2 Supplement 1X, 0.1 µg/ml R-Spondin 1, 100 ng/ml Noggin, 10 mM Nicotinamide, 1 mM N-Acetyl-L-Cysteine, 1 µg/ml Gastrin I, 500 nM LY2157299, 10 µM SB202190, 0.01 µM Prostaglandin E2, 10 mM Hepes, 50 ng/ml hEGF) was added and changed thrice weekly. Rock inhibitor Y-27623 (10 µM) was included into the culture medium the first 7 days after seeding. Primary melanoma (MEL) cells were cultured in complete DMEM

supplemented with 5 ng/ml of EGF (Preprotech) and 4 µg/ml of insulin (Sigma-Aldrich). MTs in suspension from MEL lines were cultured on Ultra-Low attachment multiwell plates or flasks (Corning Costar) in DMEM/F12 culture medium supplemented with N2-Supplement (Life Technologies), 20 µg/ml of human insulin, 10 ng/ml of human EGF, 10 ng/ml of human FGF, 0.2% of Fungizone and 0.2% of P/S. MTs in suspension from primary glioblastoma (GBM) cells were seeded in Neurobasal medium with B27 supplement (Life Tecnologies), 20 ng/ml of human EGF, 20 ng/ml of human FGF, 0.2% of Fungizone and 0.2% of P/S. All cell cultures were incubated at 37ºC in a humidified 5% CO2 environment.

Plasmids

For generation of the lentiviral vector pSIN-TRE-H2BeGFP-rtTA2, the coding sequence of human Histone2B fused to the eGFP (H2BeGFP) (Addgene Plasmid ID: 11680) (2) was amplified by PCR with specific primers (Supplemental Table 6) containing XbaI and NheI restriction sites. The resulting PCR product was then subcloned into the NheI cut pRRL-cPPT-hPGKTMPrtTA-WPRE plasmid (3, 4) to obtain the final pSIN-TRE-H2BeGFP-rtTA2 construct.

shRNA knock down control or against transcript of *TET2* (NM_017628) were performed using the Mission® shRNA Lentiviral vector system (pLKO1-puro TRC1 and TRC2; Sigma-Aldrich). The specific shRNA constructs (TRC number) are included in Supplemental Table 6.

pCMV6-Entry and pCMV6-*TET2* (NM_001127208) plasmids were from Origene (ID: PS100001 and RC226438, respectively). Generation of catalytically inactive TET2 (pCMV6-*TET2 HxD*) was performed by site-directed mutagenesis of pCMV6-*TET2* using QuickChange Mutagenesis Kit (Strategene). HxD represents mutation H1382Y/D1384A of TET2 described previously as a catalytically inactive TET2 (5).

TET2-specific sgRNA oligos were cloned into the pSpCas9(BB)-2A-GFP (px458) expression vector (Addgene, plasmid ID: 48138) (6), which bicistronically expresses sgRNA and Cas9 nuclease. Four sgRNA sequences were determined by the CRISPR Design Tool (http://crispr.mit.edu/). SgRNA

scramble sequence was obtained from OriGene. Oligos sequences are annotated in Supplemental Table 6. Between the four sgRNA constructs targeting the *TET2* locus, the pSpCas9-sgRNA*TET2*guide2-2A-GFP was selected as the best guide sequence by DNA mismatch-specific (T7E1) endonuclease assay (7).

Lentiviral infections

To infect cell lines and patient-derived cells with H2BeGFP and shRNA (shCTRL and sh*TET2*) lentiviral construct, lentiviruses were produced in 293T cells using standard procedures and psPAX2 and pMD2.G (Addgene, plasmids ID: 12260 and 12259, respectively) packaging vectors. 48h after transfection, the supernatant was collected and filtered. This supernatant was then used to infect cells directly (CRC-SW1222, CRC-SW480, CRC-SW620, CRC-DLD1, MEL-MMLN9, and MEL-MMPG3 cell lines), or concentrated (CRC-T70, GBM-e216, and GBM-e225).

Lentiviral integration sites analysis

Genomic DNA was extracted by incubating cell lines O/N at 56ºC with a lysis buffer (50mM TrisHCl pH 8, 100mM EDTA pH 8, 100mM NaCl, 1% SDS and proteinase K (20mg/ml)). Saturated NaCl buffer was added five minutes and DNA was precipitated with isopropanol and washed twice with 70% ethanol. DNA was fragmented using the ultra-sonicator S220 (Covaris^{TM}). Library preparation was performed following the standard Illumina protocol (Genomic Sample Prep). Genomic regions flanking the lentiviral integration site were amplified by PCR using Phusion Hot Start II High-Fidelity DNA Polymerase (ThermoFisher Scientific) following manufacturer's instructions. The primers designed were: one for the LTR region in the lentivirus and the other in adaptors used in the previous library preparation (Supplemental Table 6). A second library from the PCR product was generated and loaded onto MiSeq platform (Illumina).

Western blotting

For transient TET2 overexpression, 293T cells were transfected with pCMV6-Entry or pCMV6-*TET2* plasmids (described above) using PEI. Cells were lysed with 1% SDS lysis buffer containing protease inhibitors (Roche). 50 ug/lane of lysates were separated by 6% SDS-PAGE and protein was transferred onto a nitrocellulose membrane (BioRad). Membranes were blocked with TBS, 0.25% Tween-20, and 5% non-fat milk, incubated with TET2 or β-tubulin antibodies (Supplemental Table 6) diluted 1:1000 and 1:10,000 respectively in blocking solution, and visualized using SuperSignal West Pico Chemiluminescent Substrate (ThermoFisher Scientific).

Isolation of SCCC and RCCC by flow cytometry

To obtain SCCC and RCCC from colorectal (CRC) Matrigel-embedded MT models, cultures were maintained on 5 µg/ml Doxycycline (DOX; Sigma-Aldrich) medium for two weeks (pulse), and then grown on DOX-free medium for an additional 7 days (chase). Then, cultured MT were harvested using ice-cold Matrigel Recovery Solution (BD Bioscience) and incubated on ice for 1 hour as per manufacturer's protocol. MT were filtered through a 100 µm cell strainer (Corning) to purify MT bigger than 100 µm of diameter. MT were collected from the cell strainer surface and dissociated using trypsin-EDTA to obtain a single cell suspension. Finally, single-cells were resuspended in sorting medium: 4 mM Glutamine (Life Technologies), 20% FBS, 1% penicillin/streptomycin, 10 µM Y-27632 (Calbiochem) diluted in CO2-independent medium (Life Technologies).

To obtain SCCC and RCCC from melanoma (MEL) and glioblastoma (GBM) MT models in suspension, after DOX pulse-chase treatment (6-9 days DOX-free medium), a single cell suspension was obtained using accutase (Sigma-Aldrich). Then, cells were resuspended in each respective medium.

To isolate SCCC and RCCC from tumor xenografts, injected mice were first treated with DOX (2 mg/ml) ad libitum in drinking water containing 5% sucrose (Sigma-Aldrich) (Pulse). Then, when

tumors reached between 5 and 8 mm of minimum diameter, treatment was removed and tumors continued growing until the experiment end-point (chase). Finally, mice were euthanized and xenografts were processed to obtain a single cell suspension as previously described (8). Isolated cells were then purified using Ficoll-Paque Plus (GE Healthcare), resuspended in sorting medium (see above), and immunostained with anti-human TRA-85-allophycocyanin (APC) (R&D Systems) to discriminate human from mouse cells.

DAPI was added to exclude dead cells and cellular aggregates in all cell suspensions. Finally, live SCCC and RCCC were sorted using a MoFlo Legacy cell sorter (Beckman Coulter). The population of cells retaining an H2BeGFP signal equivalent to that observed in cells continuously exposed to DOX, were considered and sorted as SCCC. The RCCC fraction comprised around 10-20% of all cells with an H2BeGFP signal one order of magnitude lower than the SCCC fraction. SCCC and RCCC were collected directly into Trizol reagent (Life Technologies) not exceeding 10% of its volume for RNA extraction, quantitative RT-PCR or transcriptomics. SCCC and RCCC were collected in sorting medium prior being embedded back into Matrigel.

Analyses of H2BeGFP signal decay

Time courses of H2BeGFP pulse-chase experiments were performed with single-cell clones derived from three colorectal cancer (CRC) cell lines described above: SW480-H2BeGFP C2, SW620-H2BeGFP C3, and DLD1-H2BeGFP C7. Cells were treated with DOX (5 µg/ml) for three days to obtain maximum H2BeGFP expression (pulse). Then cells were washed with phosphate buffered saline (PBS) and cultured for additional 24, 48, 72, 96, or 120 h in DOX-free medium (chase). Finally, cells were synchronized with Nocodazole (Sigma-Aldrich) and H2BeGFP signal was measured using FACS-Calibur flow cytometer (BD Bioscience). In addition, cells continuously treated with DOX along the experiment (point, 0 h) and untreated cells (no DOX = ND) were analyzed as references. Average H2BeGFP signal was calculated for each time point using FCS express software (De Novo Software).

To calculate cell-doubling time for each clone, cells were trypsinized and counted after 4 days from seeding (t = 0) in presence of Trypan Blue to exclude dead cells. Then, cells were counted every 24 h during five days. A regression correlation was calculated using cell counts at each time point.

Expected mathematical dilution of H2BeGFP signal was calculated dividing by half the maximum H2BeGFP signal on continuously DOX-treated cells every doubling time. Such kinetics was compared with H2BeGFP average values detected at those different points indicated above.

The movie with SW620-H2BeGFP was acquired with a high speed multidimensional microscopy CellR TIRFM (Olympus). Fluorescent signal was quantified using Image J free software. Data is represented as mean ± SEM of triplicates from three independent experiments.

In vitro self-renewal assays

SCCC and RCCC sub-populations were isolated from MT grown embedded in Matrigel (for colorectal (CRC) models) or in suspension cultures (for melanoma (MEL) and glioblastoma (GBM) models) as described in Methods. After sorting, SCCC and RCCC isolated from Matrigel-embedded MT were resuspended in a 1:1 mixture of Matrigel and complete DMEM, and re-seeded as a drop in 48-multiwell plates. SCCC and RCCC obtained from suspension MT cultures were seeded at a density of 1-4 cells/µl in 96-multiwell plates using the corresponding medium for suspension cultures. Matrigel-embedded MT cultures were maintained for up to 4 weeks, and fixed in 4% PFA. Structures larger than 400 µm were counted. Suspension MT cultures were re-feed twice a week, and after 2-3 weeks, MTs were counted under the microscope. MT-forming units (%) were estimated according to the formula: number of MT/number of plated live cells x 100. Data is represented as mean ± SEM of triplicates from three independent experiments.

In vivo cancer-initiation potential

SCCC and RCCC were isolated by flow cytometry from subcutaneous CRC-SW1222-H2BeGFP (n=16) xenografts growing in NOD-SCID mice and treated with a doxycycline (DOX) pulse-chase (as described in Methods). SCCC and RCCC-isolated cells were injected at limiting dilution series into the kidney capsule of NOD-SCID mice. Once tumors were palpable, all mice were euthanized and tumor formation analyzed.

Cell cycle analysis

$6x10^6$ single CRC-SW1222-H2BeGFP cells pre-treated with 5 µg/ml of DOX (pulse), were seeded in ten 75 cm^2 Ultra-low attachment surface flasks to obtain sphere cultures. After seven days of H2BeGFP dilution (chase), cells were trypsinized and resuspended in sorting medium (described in Methods). SCCC and RCCC were isolated as detailed above. Immediately after sorting, an equal number of cells from each SCCC and RCCC were centrifuged and resuspended in 0.5 ml of cold PBS. Then, 1.5 ml of fridge-cold absolute ethanol was added while gently vortexing the cell pellet and fixed for at least 30 minutes on ice. Cells were washed three times with ice-cold PBS and resuspended in 1 ml of ice-cold PBS. 100 µl of ribonuclease (100 µg/ml, DNase free, Sigma-Aldrich) were added to the samples at room temperature for 5 minutes. Then, 400 µl of propidium iodide (50 µg/ml in PBS) was added and incubated at 4ºC for 1 hour. Finally, cells were centrifuged at 2000 rpm for 5 minutes and pellets were resuspended in 500 µl of ice-cold PBS. Cell cycle profile was analyzed by flow cytometry using a Navios Flow Cytometer (Beckman Coulter) and FCS express software (De Novo Software). Experiments were performed in triplicates.

Apoptosis assays

For in vivo assays, the proportion of apoptotic cells was determined using the Annexin V-APC kit (Bender MedSystems) or the PE Annexin V Apoptosis Detection Kit I (BD Pharmingen). Dead cells

were detected as DAPI positive (2 µg/ml, Roche). Anti-human TRA-85-APC staining discriminated cells between human and mouse species when analyzing xenograft tumors. Finally, cells were analyzed by flow cytometry using a Navios Flow Cytometer (Beckman Coulter).

For in vitro assays, CRC-SW1222-H2BeGFP sh*TET2*, *TET2* KO or overexpressing TET2 and their respective control cell lines were grown as Matrigel-embedded MT and treated with 5 µg/ml of Doxycycline (DOX) (pulse) for 9 days. After 12 days without DOX (chase), MT were recovered from Matrigel cultures and SCCC and RCCC single-cells were evaluated for apoptosis as described above. When cells were treated with TFMB-2-HG (500 µM) or Pomalidomide (10 µM, Sigma-Aldrich), the corresponding treatment was added 5 days before the end of the experiment. Data is represented as mean ± SEM of triplicates from three independent experiments.

In vitro proliferation assay

A single cell suspension in complete DMEM was mixed 1:1 with Matrigel. Drops of 25 µl of mixture containing 1000 cells were placed on 48-well plates. Three wells per cell line and time point were set up (triplicates). Matrigel-embedded cells were grown with DOX (5µg/ml). At day nine, three wells per cell line were incubated with XTT labeling mixture (Cell proliferation Kit II (XTT, Roche) for 4 hours at 37ºC according to the manufacturer's protocol. Then, the optical absorbance of the mixture was measured using an ELISA reader (Epoch Bio-Tek). This first measure was considered time 0 h. In parallel, DOX-containing medium was removed from the rest of wells by rinsing with PBS, and fresh complete DMEM was added. Data is represented as mean ± SEM of triplicates from three independent experiments.

Paraffin-embedded minitumors (MT)

DOX pulse-chase MT were removed from Matrigel using the ice-cold Matrigel Recovery Solution following the manufacturer's protocol. MT were rinsed three times with cold PBS and centrifuged at 2000 rpm for 5 minutes without disturbing the pellet. The resulting MT pellets

were fixed with 4% paraformaldehyde (Santa Cruz biothecnology) at 4ºC overnight. Then, the fixing solution was replaced with 70% ethanol and MT pellets were dehydrated and embedded in paraffin blocks. Finally, paraffin-embedded MT sections were cut 4 µm thick and analyzed by immunofluorescence (IHC-F).

Exogenous TET2 and 5hmC immunofluorescence

HEK293T cells were transiently transfected with empty vector, WT or catalytically inactive (HxD) TET2 expression vectors using polyethylenimine (PEI 25000, Polysciences, Inc.). Transfected HEK293T cells were fixed with 4% paraformaldehyde for 15 minutes. To denature de DNA, cells were treated with 2N HCl for 30 minutes, followed by neutralization with 100 mM borate buffer (pH 8.5). After permeabilize and blocking with 0.5% Triton X-100 and 5% normal donkey serum, cells were stained with rabbit anti-5hmC (Active Motif) and mouse anti-Flag (Sigma-Aldrich) antibodies. Finally, cells were incubated with the corresponding secondary antibody. Nuclei were conterstained with Hoechst 33342. For microscope image acquisition, a Nikon C2+ Confocal Microscope was used to visualize fluorescence and images were acquired using the NIS-Elements Advanced Research software.

Immunofluorescence and immunohistochemical staining

Matrigel embedded minitumors (MTs) were grown on coverslips. Then, cultures were fixed with 4% PFA during 1 h at room temperature (RT) and permeabilized with 1% Triton X-100/PBS for three hours. Samples were incubated with blocking solution (0.1% Triton X-100 and 3% BSA in PBS) overnight at 4ºC. Primary antibodies (Supplemental Table 6) diluted 1:100 in blocking solution were added and incubated 24 hours at RT. Then, MT were incubated with the corresponding secondary Alexa-Fluor antibody overnight at RT. When corresponding, Phalloidin-TRICT was added as counterstain during incubations with primary and secondary antibodies (1:1000; Sigma-Aldrich).

For standard double-immunofluorescence (IHC-F) analyses of FFPE sections, antigen retrieval was performed in 10 mM sodium citrate buffer (pH 6). Slides were then permeabilized with PBS-1% Tween-20 (Sigma-Aldrich) for 15 min. Then, tissue specimens were blocked for 1 h with PBS containing 3% of BSA and incubated with a corresponding couple of β-catenin (β-CAT), caspase-3 (CAPS3), chromogranin-A (CGA), cytokeratin-20 (CK20), phospho-histone H2AX (pH2AX), H3K4me2, Ki67, lysozyme (LYZ), mucin-2 (MUC2), and green fluorescent protein (GFP) antibodies (Supplemental table 6) diluted in blocking solution overnight at 4 ºC. Finally, slides were incubated with the corresponding Alexa Fluor secondary antibodies at a dilution of 1:200 for 1 hour at RT. For 5-hmC or BrdU and GFP or β-CAT double-staining, after antigen retrieval, DNA was denatured to single strands by immersing paraffin-sections in 2N HCl for 15 minutes at RT and then neutralizing with 0.1 M borate buffer pH 8.5 for 15 minutes at RT. Sections were permeabilized and blocked using 0.5% Triton X-100 and 5% normal donkey serum diluted in PBS for 1 hour at RT. Slides were then incubated with specific primary antibodies diluted in 0.1% Triton-X100 + 5% normal donkey serum overnight at 4ºC. For 5-hmC and Ki67 double-staining, slides were first incubated with anti-Ki67 antibody and then with the corresponding secondary antibody (see standard protocol for Ki67 antibody described above). After that, immunostaining was fixed incubating slides with 4% PFA for 15 minutes at RT. Slides were then incubated with 2N HCl and then neutralized with borate buffer (see above). Finally, samples were incubated with 5-hmC antibody and the corresponding secondary antibody. Detection of Edu incorporation into the SCCC DNA was performed with the Click-iT® EdU Alexa Fluor® 594 Imaging Kit (Invitrogen) according to the manufacturer's protocol and followed by a standard GFP staining. Hoechst 33342 (5 µg/ml) was used as counterstaining to detect cell nuclei in all samples. A Nikon C2+ Confocal Microscope was used to visualize fluorescence and images acquired using NIS-Elements Advanced Research software.

For immunohistochemical analyses of TET2, 5-hmC or Ki67, tissue microarray or tumor FFPE tissue sections were routine deparaffinated, rehydrated and treated with 1mM EDTA pH8 (TET2)

or 10 mM sodium citrate buffer pH6 (5-hmC and Ki67). For 5-hmC staining, sections were also treated with 2N HCl and then with borate buffer (see above) to complete the antigen retrieval. For TET2 and 5hmC antibodies, after the blocking of endogenous peroxidase activity, slides were incubated with blocking solution: 5% BSA, 3% rabbit serum, 0.25% gelatin from cold water fish skin and 0.1% Triton X-100 diluted in PBS1X for TET2 antibody or 5% rabbit serum and 0.5% Triton X-100 for 5-hmC antibody. For Ki67 staining, after the blocking of endogenous peroxidase activity, slides were blocked with 3% of BSA. Primary antibodies (Supplemental Table 6) were then incubated at 4ºC overnight. After washing, EnVision+System-HRP-labelled Polymer Anti-Rabbit (DAKO) or HRP-anti-mouse (ThermoFisher Scientific) was applied for 25 minutes at RT. After washing, Chromogen DAB/substrate reagent was added onto the slides and incubated for 10 minutes. Finally, the slides were counterstained with hematoxylin, dehydrated, and mounted. NanoZoomer 2.0-HT Digital slide scanner C9600 (Hamamatsu Photonics K.K., Japan) was used to visualize and assess TET2, 5-hmC and Ki67 expression.

High Density Pan-Cancer Tissue Microarrays (HD-PC-TMA)

In brief, formalin-fixed paraffin-embedded (FFPE) blocks and corresponding hematoxylin and eosin (H&E) slides were identified in the biobank of the pathology department at University Hospital Basel and the region of interest (ROI) was defined by the specialized pathologists according to the respective diseases of interest. The selected H&E slides were scanned by means of a P250 slide scanner (3D-Histech Ltd, Sysmex Suisse AG, Horgen, Switzerland) and further elaborated with Pannoramic Viewer Program (3DHISTECH Ltd, Hungary). The construction was performed with a TMA GrandMaster® (TMA-GM; 3D-Histech Ltd, Sysmex Suisse AG, Horgen, Switzerland), which allows overlapping the image of the donor blocks with their scanned and for ROI labelled H&E slides. This accurate machine allows to identify and to register the exact area which is plotted on the TMA. The diameter of each core was 1mm and in general for approximately one third the tumor entities a paired non-malignant adjacent tissue was also

assigned to the TMAs. TMAs represented 776 tumor and 233 normal tissue punches. Intact tumor tissues corresponding to 656 patients were finally evaluated for 5hmC levels (Figure 12). Basic pathologic and clinical data available included histological tumor subtype, grade, TNM stage, chemotherapy status, outcome, and disease-free survival (DFS) or overall survival (OS) time (Supplemental Table 4).

Image quantification

IHC-F of FFPE sections were performed as described above. Between one and nine standard confocal images were taken for each section. The percentage of proliferative H2BeGFP-positive (SCCC) and -negative (RCCC) cells was calculated counting the number of cells simultaneously expressing H2BeGFP and Ki67 (SCCC) or DAPI and Ki67 (RCCC) using Image-J software as we previously described (9). Percentages of BrdU, EdU and pH2AX positive cells in SCCC and RCCC were calculated counting the number of BrdU, EdU or pH2AX cells in each SCCC (H2BeGFP positive cells) or RCCC (H2BeGFP negative cells) population. The percentage of 5hmC-positive cells in primary tumors and liver metastases (CRC primary tumors vs liver metastases cohorts) was calculated with respect to the total number β-catenin positive cells (tumoral tissue) using the Image-J software. Tumors presenting 5hmC staining were considered as positive or 5hmC HIGH. Tumors with non 5hmC positive cells or less than 5% were considered as 5hmC negative or LOW (Figure 11E, Figure 12, C and D, Supplemental Figure 7C, and Supplemental Table 5). The percentage of Ki67-positive cells in primary and their corresponding paired liver metastases was analyzed by immunohistochemistry and quantified using the QuPath software (10). To calculate 5hmC content for each individual cell, the red channel Integrated Density Value (IDV) corresponding to the 5hmC staining was measured in each individual H2BeGFP-positive (SCCC) and -negative (RCCC) cells, stained green or blue respectively, using Image-J software. 200-1000 cells were measured for each population. The average of the amount of 5hmC per each SCCC and RCCC per image was shown as relative units (r.u.) and represented as column scatter plots

(Figures 6B, 7E and 11C). We quantified between 8 and 46 images per experimental condition from 5 to 8 tumors.

5hmC levels measured by immunohistochemistry in high density pan-cancer tissue microarrays (HD-PC-TMA) and VHIO cohort (Figure 11F and Figure 12, A and B) were considered 5hmC HIGH when at least 5% of tumor cells presented equal or higher signal than adjacent stroma. Tumors were considered 5hmC LOW when its accumulation in cancer cells was lower than in stroma cells or positive cells represented less than 5%. Negative tumors did not show any detectable 5hmC signal in cancer cells were considered 5hmC NEG. This approach helped to correct any potential difference in staining due to technical reasons of the immunohistochemistry technique between samples since the final value of 5hmC level was corrected by its own intrinsic stroma.

TET2 levels measured by immunohistochemistry and considered HIGH when cancer cell signal was higher than stromal background staining (Supplemental Figure 7A).

Total 5-hmC Quantification

Genomic DNA was extracted using QIAamp DNA Mini Kit (QIAGEN) and 5hmC was measured using MethylFlash Hydroxymethylated DNA Quantification Kit (Epigentek) following corresponding manufacturer's instructions. Data is represented as mean ± SEM of triplicates from three independent experiments.

Microarrays

Gene expression profiles were analyzed from 29 replicates of FACS-isolated SCCC and 29 paired RCCC. Samples correspond to DOX pulse-chase MT cultures from six models derived from three different cancer types: CRC-SW1222, CRC-T70, MEL-MMLN9, MEL-MMPG3, GBM-e216 and GBM-e225 (3 replicates for each model), or DOX pulse-chase xenograft tumors developed from CRC-SW1222 (n=20 xenografts; 3 replicates), CRC-T70 (n=18 xenografts; 5 replicates) and MEL-

MMPG3 (n=8 xenografts; 3 replicates). For in vivo replicates, between 2 and 4 xenografts were pooled to obtain one replicate. All cancer models were infected with H2BeGFP.

Transcriptomes were determined on a genome wide Human Gene 1.0 ST Array (Affimetrix). SCCC and RCCC were isolated directly in Trizol. Total RNA of biological replicates was isolated following manufacturer's instructions. RNA integrity was confirmed in an Agilent 2100 Bioanalyzer using RNA Nano- or Pico-chips. Total RNA was linearly amplified using Ovation® Pico WTA System V2 (Nugen). The resulting ssDNA was used to hybridize Affimetrix microarrays. Hybridization data was acquired using the Affimetrix GeneChip/GeneTitan platforms. We used Partek Genomics Suite 6.6 software (Partek Inc.) to normalize raw CEL files in different combinations as indicated using Robust Multichip Average (RMA) algorithm and to remove batch effect of the different scan dates. The software also rendered PCA analyses and the hierarchical cluster depicting probe set expression of genes significantly detected ($P < 0.05$). Normalized expression values were used to determine the fold change (FC) expression between the respective paired SCCC and RCCC and its statistical significance in parametric two-tailed paired sample *t* test (P value) and a Benjamini-Hochberg-corrected false discovery rate test (q-value; FDR). Normalized lists of differentially expressed genes between the indicated conditions were cut-off at ± 1.2 fold change between conditions and at a significance level of FDR q-value < 0.05 (Supplemental Tables 2 and 3). For the comparison of SCCC versus RCCC from SW1222 MT infected with sh*CTRL*, and sh*TET2* constructs, the gene list was cut-off at FDR < 0.05 and at a change of ± 1.25 fold in at least one of the knock down conditions.

Gene expression signatures and gene sets

PanC- and CRC- SCCC signatures

PanCancer (PanC)-SCCC expression profile was generated from the expression profiles of 29 replicates of SCCC and their corresponding 29 paired RCCC isolated from colorectal (CRC-SW1222 and CRC-T70), melanoma (MEL-MMLN9 and MEL-MMPG3) and glioblastoma (GBM-

e216 and GBM-e225) models evaluated by microarray (Supplemental Table 2). The signatures correspond to the median expression of top 100 genes upregulated minus median expression of 100 genes downregulated in CRC samples for CRC-SCCC signature and in all cancer samples for PanC-SCCC signature. This calculation generates a unique score and the upper quartile was selected to define positivity for the signature (SCCC sig High) in each condition. This threshold was selected based on the prevalence of relapse events in the clinical cohort with outcome data (GSE39582) and applied to all other clinical cohorts and cell lines models (Figure 4, D and E, Figure 5, G and H, and Supplemental Figure 4D).

TET2 signature

From the gene expression data on CRC-SW1222 sh*TET2* SCCC (Supplemental Table 3) we derived the TET2 gene set and signature. We first considered as TET2 induced genes in SCCC those higher expressed in shCTRL SCCC vs shCTRL RCCC but not in sh*TET2* SCCC vs sh*TET2* RCCC (Figure 10A). 121 of these 449 genes were higher expressed in SCCC of all cancer models analyzed (Supplemental Table 2). This 121 gene set was used to evaluate its enrichment in SCCC from CRC, MEL and GBM models (Figure 11A). The expression of these 121 genes was also evaluated as a signature in CRC (GSE39582), see below. Samples with high expression (defined as z-scores above 2 in normalized RNA sequencing data) of any gene in the signature were considered positive (TET sig High) (Figure 11B).

Clinical Cohorts

GSE39582 (public): Affymetrix Human Genome U133 Plus 2.0 Array data was downloaded from GEO website and normalized using RMA (R package *limma*). Probesets were mapped to unique gene symbols based on singular value decomposition. The CRC-SCCC and TET2 signature score was calculated and samples at the upper quartile were classified as positive (High) (Figures 5H and 11B). We restricted the survival analysis to patients that received adjuvant chemotherapy and that had outcome data available (n = 151), corresponding to high-risk stage II or stage III

cases. Disease-free survival (DFS) was defined as the time from diagnosis to relapse or death due to any cause. We performed Cox Proportional Hazards univariable and multivariable modelling using R package *survival*. Variables included in the model included all factors publicly available (age, gender, stage, tumor site, microsatellite status, *KRAS* mutation and *BRAF* mutation status) and CRC-SCCC or TET2 signatures. Details are shown in Supplemental Tables, 2 and 3.

VHIO (private): The percentage of 5hmC-positive cells was evaluated on a cohort of 61 CRC tumors by immunofluorescence on FFPE sections as described above. Tumors presenting 5hmC staining were considered as a 5hmC positive case or 5hmC HIGH. Tumors that were completely negative (0 % positive cells) were considered as a 5hmC negative case or 5hmC LOW (less than 5% of positive cells). We restricted the correlation study between 5hmC quantification and *TET2* expression (evaluated by nCounter) (Supplemental Figure 7C) to patients that received adjuvant chemotherapy and that had outcome data available (n=53), corresponding to high-risk stage II or stage III CRC cases.

VHIO (private) and HD-CRC-TMA: The 5hmC levels was evaluated by immunohistochemistry and immunofluorescence on FFPE sections. 5hmC and TET2 HIGH vs LOW were defined as indicated above (Image quantification section) including negative (NEG) cases as LOW. DFS was defined as the time from diagnosis to relapse or death due to any cause. This cohort include CRC patients from VHIO cohort (n = 61) and CRC patients from the HD-PC-TMA (High Density PanCancer Tissue Microarray) (HD-CRC-TMA) (n = 47) defined in Methods section and Supplemental Table 4. We included all CRC patients (n = 108) regardless of treatment type to calculate the multivariable model using Cox Proportional Hazards with R package *survival*. Variables included in the model were restricted to those with known clinical impact in DFS in early stage colon cancer (stage, grade, tumor site, microsatellite status, *BRAF* mutation status, adjuvant chemotherapy) and 5hmC status. Univariable PFS models were calculated for all CRC patients regardless of

chemotherapy status and for CRC patients treated with adjuvant chemotherapy. Details are shown in Supplemental Table 4.

CRC primary tumor vs Liver metastases (private): The percentage of 5hmC-positive cells were evaluated by immunofluorescence on FFPE sections as described in Methods section. Tumors presenting 5hmC staining were considered as positive cases. We compared 5hmC levels of primary tumor samples with their matched liver metastasis from the same patient using Wilcoxon paired test (n = 197) (Figure 12C and Supplementary Table 5). We then restricted the analysis to those patients with synchronous metastases and no chemotherapy exposure before liver resection (n = 96) (Figure 12D). Finally, for each primary CRC sample (n=55) and liver metastases (n=47) of some cases of this cohort, 5hmC and Ki67 co-immunofluorescence staining was performed and the percentage of 5hmC-positive cells per CRC tumor was correlated with their corresponding Ki67 staining (Figure 11E and Supplementary Table 5).

TCGA CRC and GBM (public): Normalized Illumina RNA sequencing gene expression data from colon and rectal cancer samples for CRC and glioblastoma samples for GBM were downloaded from Synapse website (https://www.synapse.org/#!Synapse:syn2812961). We calculated the PanC-SCCC signature score of every sample and compared with consensus colorectal cancer subtyping label as previously described (11) for CRC and with expression subtypes as described by Brennan *et al* (12) for GBM with a Kruskal Wallis test (scores of one subtype versus scores of the remaining subtypes). The scores were normalized (mean = 1, standard deviation = 0.2) for plotting purposes (Figure 4, D and E).

Cell line cohort with drug sensitivity data

Cancer cell line encyclopaedia (public): Affymetrix Human Genome U133 Plus 2.0 Array data was downloaded from Cancer Therapeutics Research Portal v2 website (http://portals.broadinstitute.org/ctrp/) and normalized using RMA (R package *limma*). Probesets were mapped to unique gene symbols based on singular value decomposition. The

PanC-SCCC signature score was calculated and cell lines at the upper quartile were classified as positive (Figure 5G and Supplemental Figure 4D). Drug sensitivity data (Area Under the Curve, AUC) was also obtained from the same study (13). We then assessed AUC levels of drug sensitivity of the chemotherapeutic agents of interest (topoisomerase inhibitors, microtubule inhibitors, antimetabolite agents) as per PanC-SCCC signature class in all cancer cell lines (n = 653) using the Wilcoxon test, with adjustment for multiple testing according to Benjamini-Hochberg procedure.

Functional Gene Set Enrichment Analyses (GSEA)

We performed Gene Set Enrichment Analyses (GSEA) and single sample GSEA projections (ssGSEA) (14) using the public Genepattern server that included calculation of the corresponding *P* values and enrichment scores (genepattern.broadinstitute.org) (15). We used custom gene sets and publicly available gene sets (Molecular Signatures Database v5.0; http://www.broadinstitute.org/gsea). Normalized probe set intensities and uniquely annotated to gene symbols, were used for ssGSEA projection towards gene sets from REACTOME, KEGG, Biological Process and Cellular Components Gene Ontology databases. The results were combined to calculate a fold change between Gene Set Enrichment scores in SCCC versus RCCC conditions. The significance of enrichment differences was calculated by one-way ANOVA P value and adjusted FDR q-value. Fold change and statistical significance were calculated using Partek Genomic Suite software (Partek Inc.). Significantly enriched gene sets (FDR < 0.05) that showed a fold change > ± 1.2 were selected for hierarchical clustering using an uncentered Pearson correlation distance and complete linkage (Cluster 3.0 Software) (16, 17). Redundancy among enriched gene sets was reduced using the Jaccard dissimilarity coefficient, taking advantage of tools provided by the Gene Weaver server (http://www.geneweaver.org/) (18).

Genomics

Target-directed or Exome sequencing was performed to genotype cancer cell models. High-seq Illumina technology was used as previously described (1). Microsatellite instability (MSI) was analyzed using the MSI-Analysis System (Promega).

Gene expression profiling of FFPE samples (nCounter)

Acquisition of gene expression profiles on the nCounter platform (Nanostring Technologies) was described before (8). Briefly, hematoxylin and eosin (H/E) staining was performed in each formalin-fixed paraffin-embedded (FFPE) tumor tissue of a collection of 53 colorectal cancer samples included in the VHIO cohort. Areas enriched with tumor tissue were identified and a minimum of two FFPE tumor tissue cores (1 mm diameter) were collected. RNA was purified using the Roche HighPure FFPE Micro Kit, and approximately 100 ng of total RNA was used to measure expression of *TET2* using the nCounter platform. Raw data was log base 2 transformed and normalized using five house-keeping transcripts and nSolver 2.0 software.

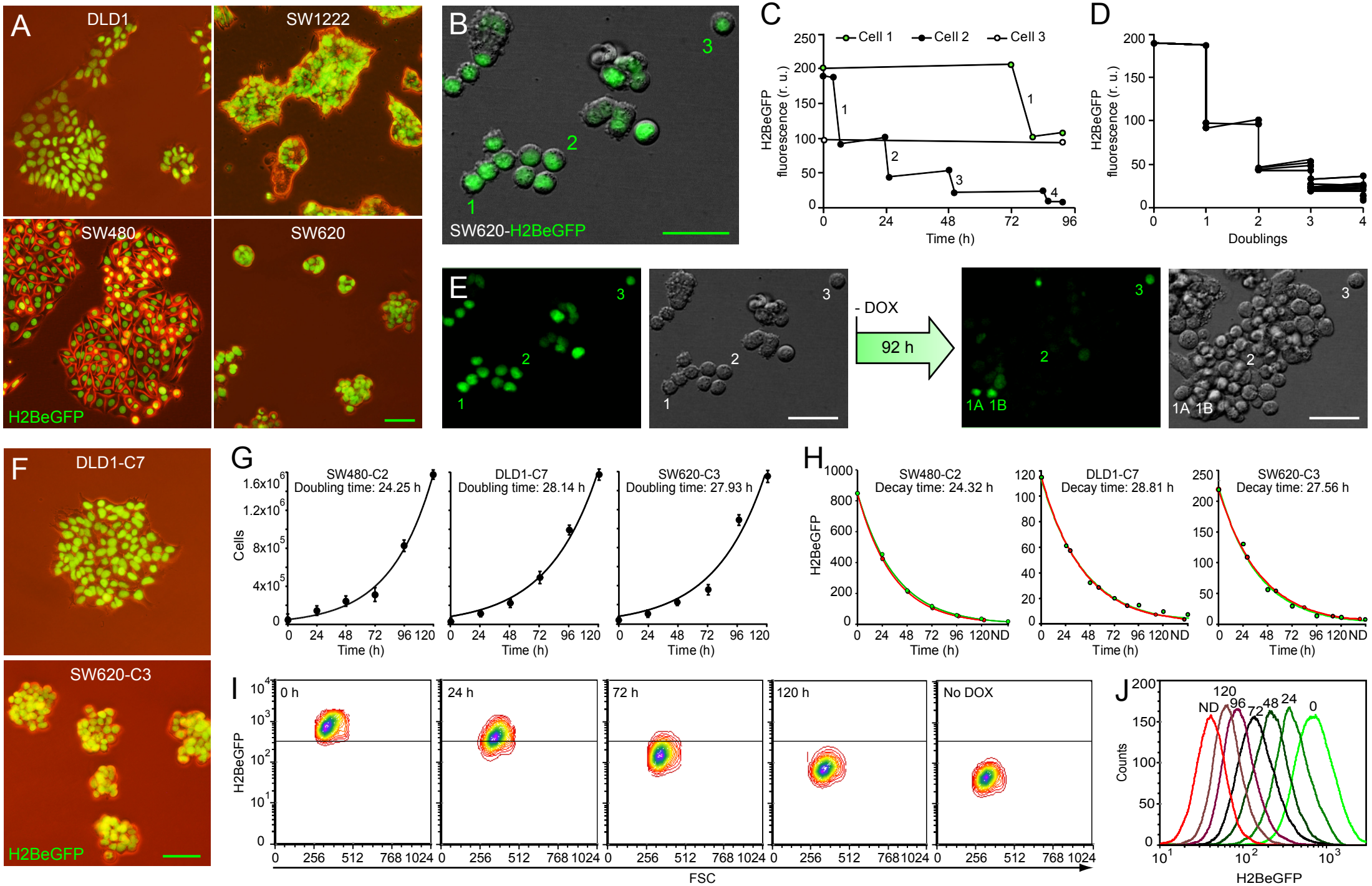
REFERENCES

1. Arques, O., Chicote, I., Puig, I., Tenbaum, S.P., Argiles, G., Dienstmann, R., Fernandez, N., Caratu, G., Matito, J., Silberschmidt, D., et al. 2016. Tankyrase Inhibition Blocks Wnt/beta-Catenin Pathway and Reverts Resistance to PI3K and AKT Inhibitors in the Treatment of Colorectal Cancer. *Clin Cancer Res* 22:644-656.
2. Kanda, T., Sullivan, K.F., and Wahl, G.M. 1998. Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Curr Biol* 8:377-385.
3. Cespedes, M.V., Espina, C., Garcia-Cabezas, M.A., Trias, M., Boluda, A., Gomez del Pulgar, M.T., Sancho, F.J., Nistal, M., Lacal, J.C., and Mangues, R. 2007. Orthotopic microinjection of human colon cancer cells in nude mice induces tumor foci in all clinically relevant metastatic sites. *The American journal of pathology* 170:1077-1085.
4. Barde, I., Zanta-Boussif, M.A., Paisant, S., Leboeuf, M., Rameau, P., Delenda, C., and Danos, O. 2006. Efficient control of gene expression in the hematopoietic system using a single Tet-on inducible lentiviral vector. *Molecular therapy : the journal of the American Society of Gene Therapy* 13:382-390.
5. Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., et al. 2010. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 468:839-843.
6. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8:2281-2308.
7. Vouillot, L., Thelie, A., and Pollet, N. 2015. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 (Bethesda)* 5:407-415.
8. Puig, I., Chicote, I., Tenbaum, S.P., Arques, O., Herance, J.R., Gispert, J.D., Jimenez, J., Landolfi, S., Caci, K., Allende, H., et al. 2013. A personalized preclinical model to evaluate the metastatic potential of patient-derived colon cancer initiating cells. *Clin Cancer Res* 19:6787-6801.
9. Ito, K., and Suda, T. 2014. Metabolic requirements for the maintenance of self-renewing stem cells. *Nat Rev Mol Cell Biol* 15:243-256.
10. Bankhead, P., Loughrey, M.B., Fernandez, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al. 2017. QuPath: Open source software for digital pathology image analysis. *Sci Rep* 7:16878.
11. Guinney, J., Dienstmann, R., Wang, X., de Reynies, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., et al. 2015. The consensus molecular subtypes of colorectal cancer. *Nat Med* 21:1350-1356.
12. Brennan, C.W., Verhaak, R.G., McKenna, A., Campos, B., Noushmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al. 2013. The somatic genomic landscape of glioblastoma. *Cell* 155:462-477.
13. Seashore-Ludlow, B., Rees, M.G., Cheah, J.H., Cokol, M., Price, E.V., Coletti, M.E., Jones, V., Bodycombe, N.E., Soule, C.K., Gould, J., et al. 2015. Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer Discov* 5:1210-1223.
14. Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. 2009. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 462:108-112.
15. Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., and Mesirov, J.P. 2006. GenePattern 2.0. *Nat Genet* 38:500-501.
16. de Hoon, M.J., Imoto, S., Nolan, J., and Miyano, S. 2004. Open source clustering software. *Bioinformatics* 20:1453-1454.
17. Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95:14863-14868.

18. Baker, E.J., Jay, J.J., Bubier, J.A., Langston, M.A., and Chesler, E.J. 2012. GeneWeaver: a web-based system for integrative functional genomics. *Nucleic Acids Res* 40:D1067-1076.

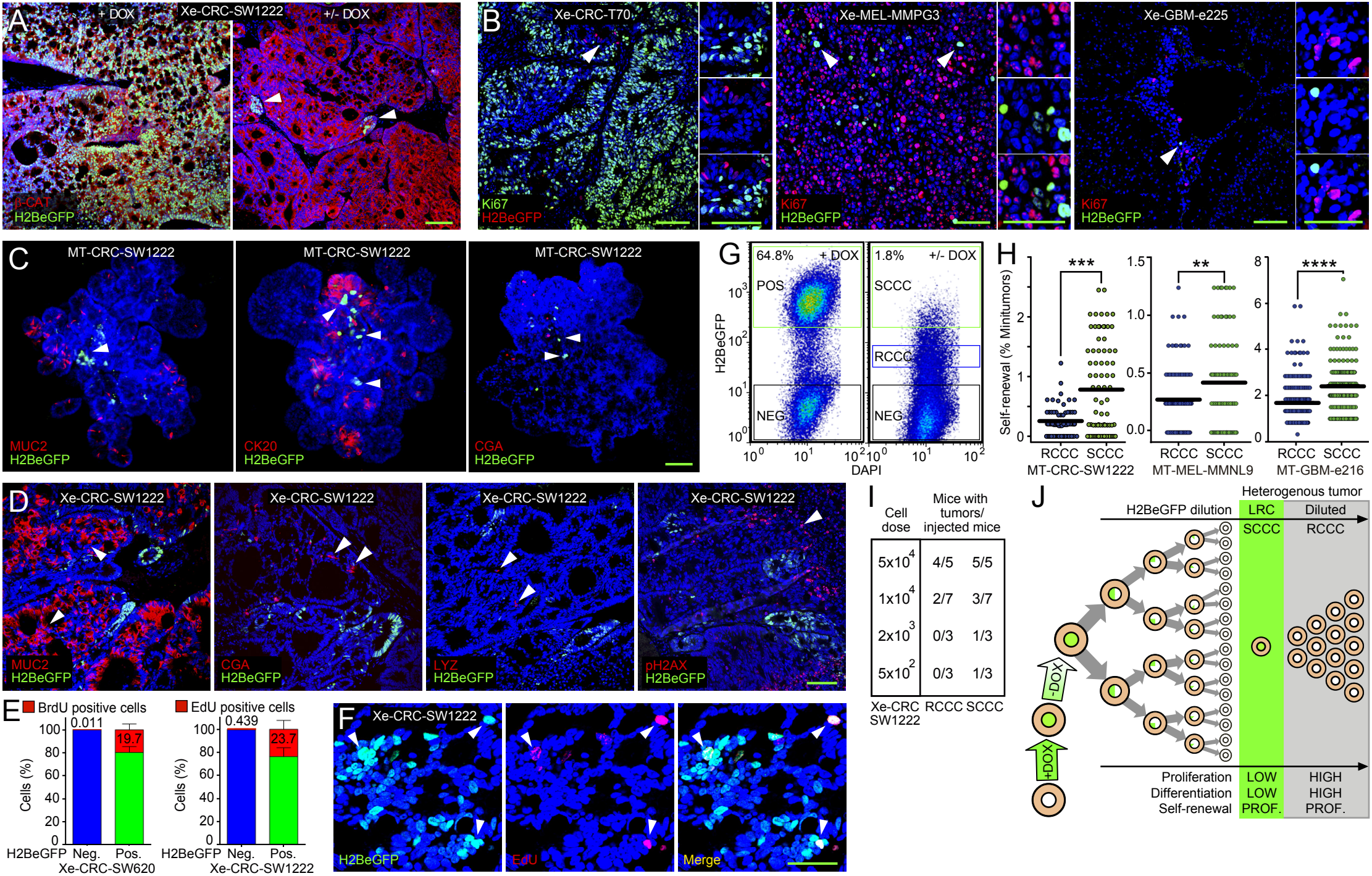
Puig et al.

Supplemental Figure 1



Supplemental Figure 1. H2BeGFP label splits with mathematical precision upon cell division. (**A**) Cultured pools of colon cancer cell lines infected with H2BeGFP lentivirus and treated with doxycycline (DOX). (**B-E**) SW620-H2BeGFP cells were filmed for 92h chase after DOX pulse (see Supplemental Movie 1). (**B**) Time zero image pointing three cells (1, 2, 3) from which H2BeGFP dilution was quantified in panel C. (**C**) Plot representing the average H2BeGFP signal of the cell offspring originated from each cell pointed at time zero (cell 1, 2, 3). Cell 1 divided once, cell 2 divided four times and cell 3 did not divide. (**D**) Plot representing H2BeGFP signal of each offspring cell originated from cell 2. (**E**) Representative pictures showing H2BeGFP signal at time zero and end point. (**F-J**) Quantification of H2BeGFP dilution dynamics from single-cell derived cultures of the indicated colon cancer cell lines clones. (**F**) Representative pictures showing nuclear accumulation of H2BeGFP after DOX treatment. (**G**) Quantification of cell doubling time by counting cell numbers each 24 hours. Data is represented as mean ± SEM of triplicates. (**H**) H2BeGFP dilution measured by flow cytometry every 24 h after DOX removal (green line). Expected H2BeGFP signal at each cell doubling time (red line) calculated by dividing by half the maximum initial H2BeGFP signal measured upon DOX treatment. (**I, J**) H2BeGFP dilution in SW620-C3 upon DOX removal shown as flow cytometry density plot (I) or histogram (J). (**A, B, E, F**) Scale bars, 100 μm.

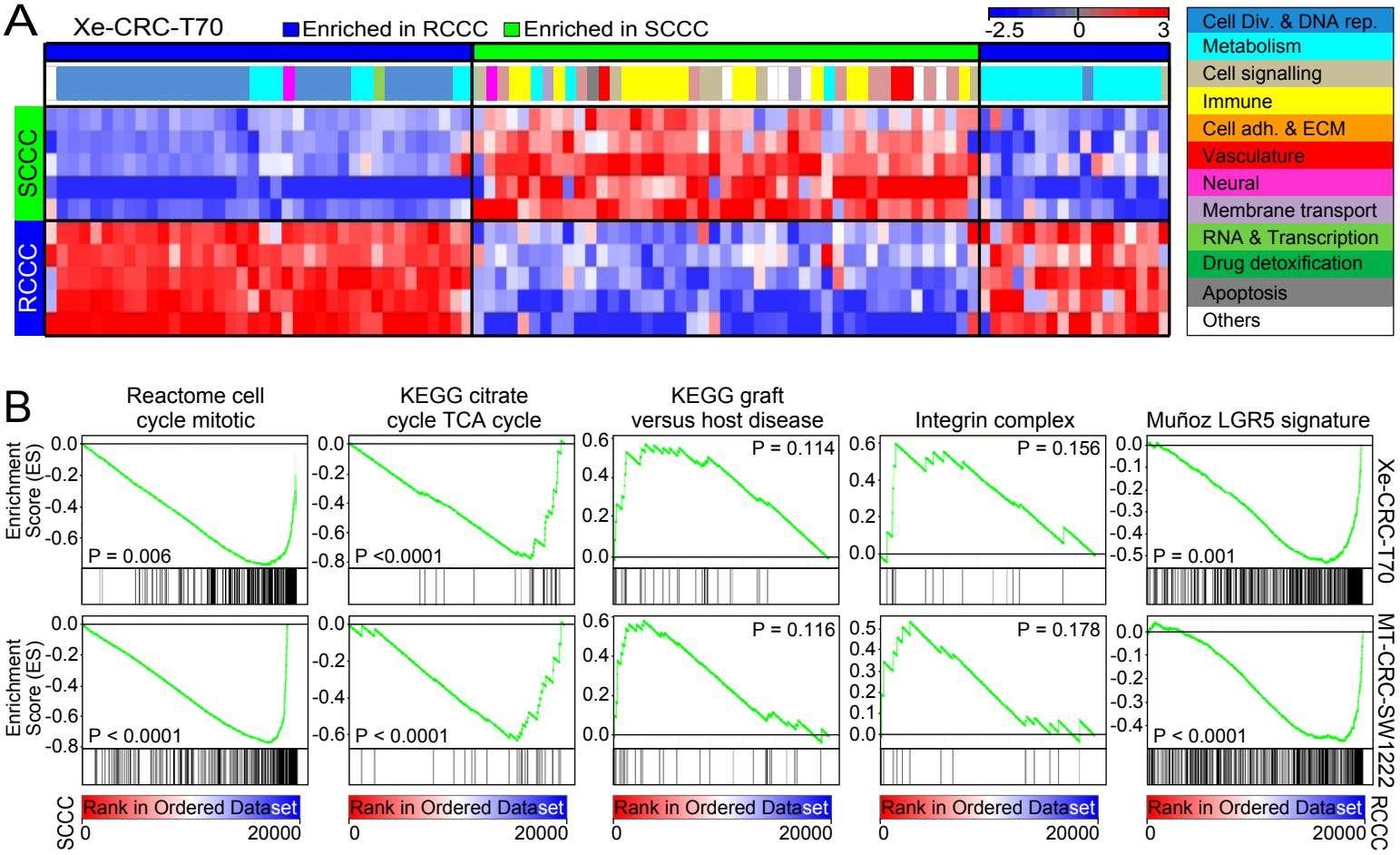
Supplemental Figure 2



Supplemental Figure 2. Immunophenotyping and cancer-initiating potential of SCCC and RCCC. (**A**) Immunofluorescence of H2BeGFP in subcutaneous tumor xenografts (Xe) growing in mice continuously treated with doxycycline (+DOX) or after a DOX pulse-chase (+/- DOX). β-catenin was used as counterstaining. (**B**) Representative immunofluorescence picture of H2BeGFP and the proliferation marker (Ki67) in DOX pulse-chase-treated (+/-DOX) tumor xenografts (Xe) from each indicated model. (**C, D**) Representative pictures of SW-CRC-SW1222 grown as minitumors (MT) (C) or xenografts (Xe) (D) and analyzed by immunofluorescence for the differentiation or senescence markers indicated. MUC2, mucin-2; CK20, cytokeratin-20; CGA, chromogranin A; LYZ, lysozyme; pH2AX, phospho-histone H2AX. (**A-C**) Arrowheads, SCCC. (**D**) Arrowheads, differentiated or senescence cells. (**E, F**) Comparative of different systems to identify label-retaining cells (LRCs). (E) Histological quantification analysis of BrdU or EdU colocalization with H2BeGFP in indicated cell models. (F) Representative pictures of H2BeGFP and EdU immunostaining. Arrowheads, SCCC label with EdU. (**A-D, F**) Scale bars, 100 μm. (**G**) Representative scatter plot of a single-cell suspension from tumor xenografts growing in mice continuously treated with DOX (+ DOX) or after a DOX pulse-chase (+/- DOX). Gates used for the isolation by FACS of RCCC, SCCC and non-infected cells (NEG). (**H**) Minitumor formation capacity was evaluated for RCCC and SCCC isolated from the indicated cell lines. Dots indicate the percentage of MT grown in each single well. **P ≤ 0.01; ***P ≤ 0.001; ****P ≤ 0.0001 of 2-tailed Student's *t* test. (**E, H**) Data are represented as mean ± SEM of triplicates. (**I**) RCCC and SCCC isolated from CRC-SW1222 xenografts were injected under the kidney capsule of NOD-SCID mice at limiting numbers. Generation of tumor xenografts were evaluated for each injected mice. (**J**) Diagram representing H2BeGFP retention-dilution dynamics in a growing and heterogeneous tumor. PROF: Proficient.

Puig et al.

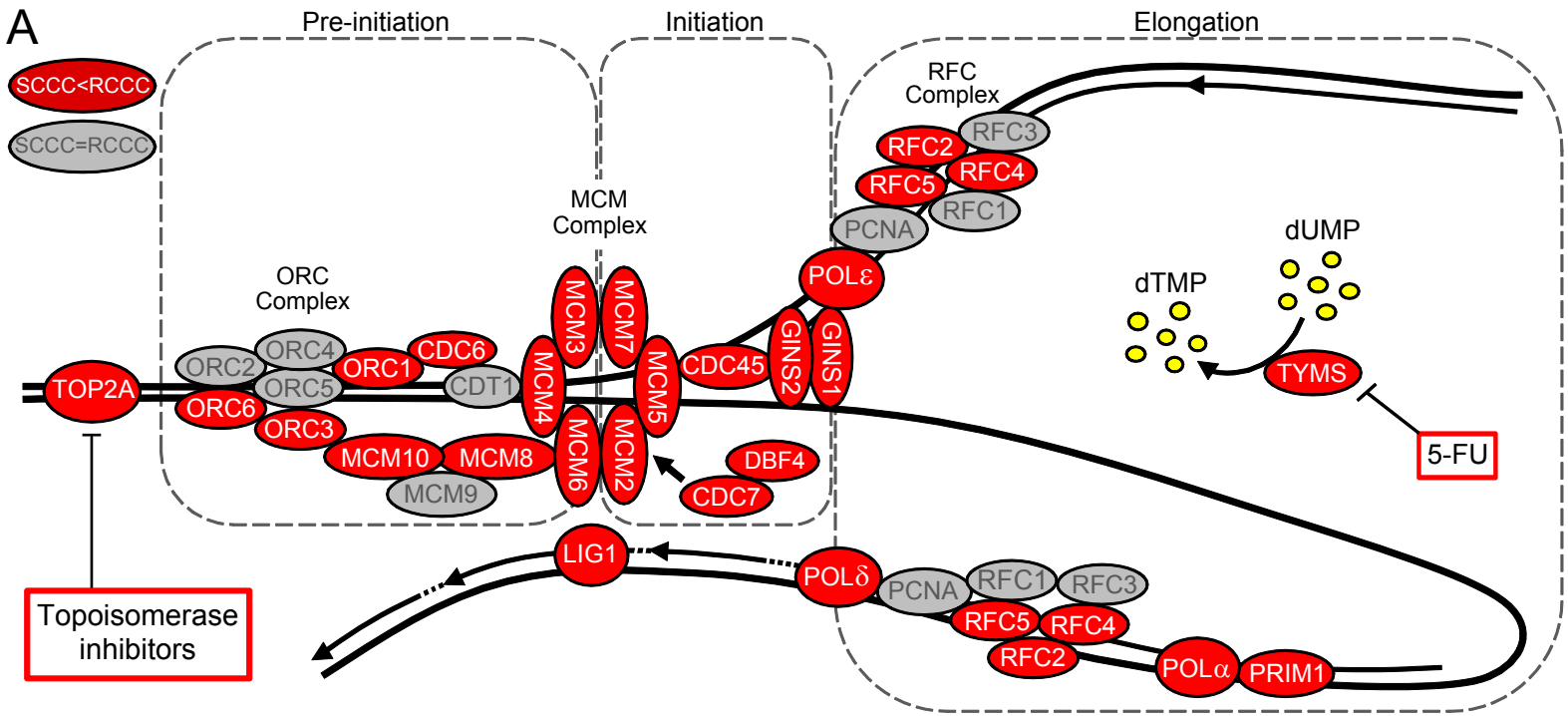
Supplemental Figure 3

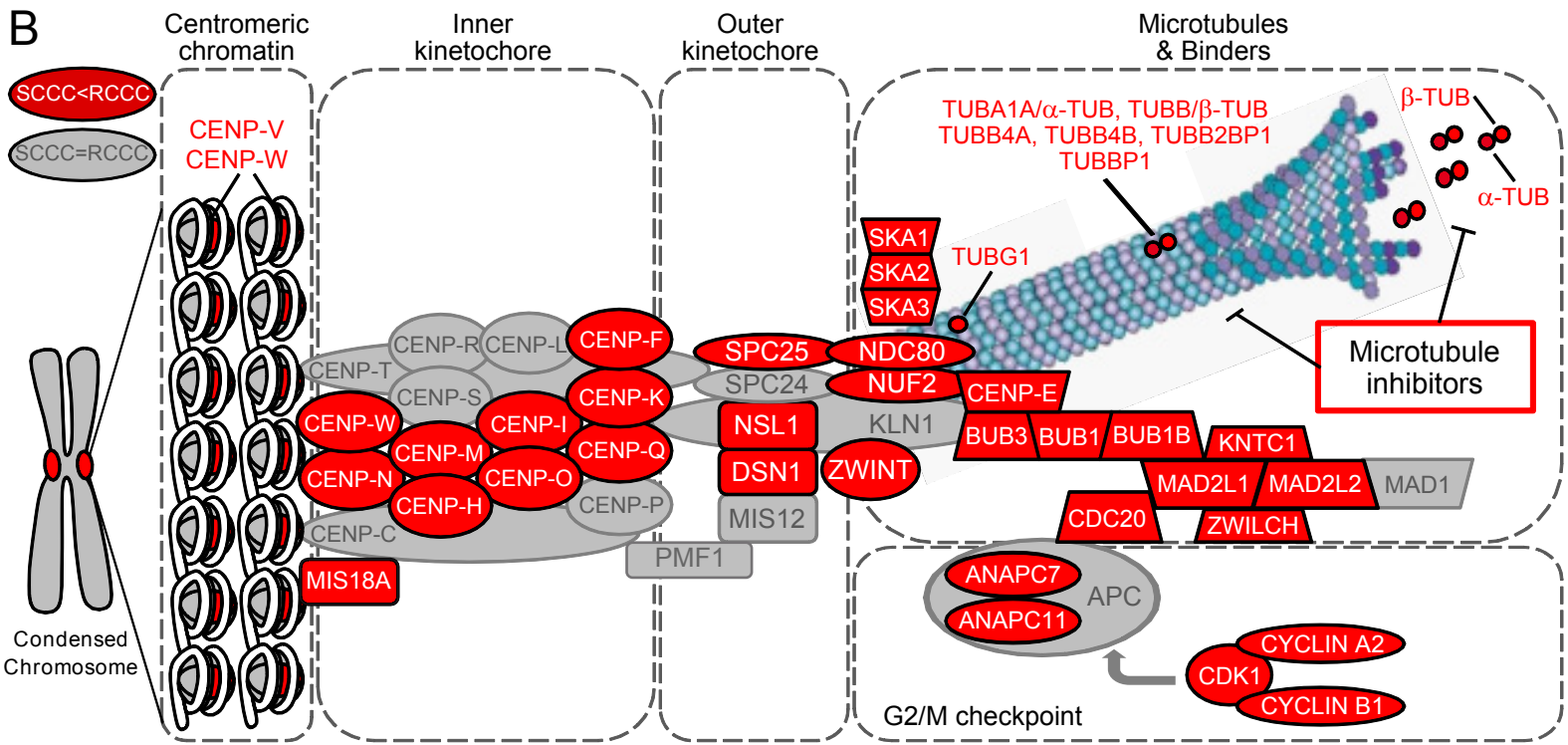


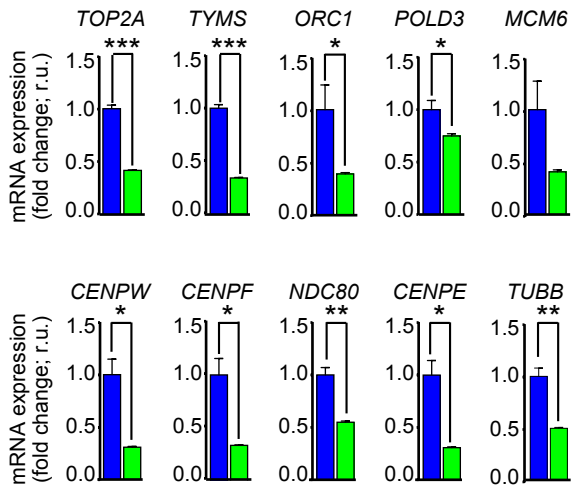
Supplemental Figure 3. SCCC present a gene expression profile related with slow proliferation and a delay in G2/M phases of the cell cycle. (**A**) Single sample GSEA projections comparing SCCC and RCCC from CRC-T70 (n = 18) xenografts cancer model. Color boxes correspond to general functions grouping differentially enriched gene sets. Color bar legend: blue, down-regulated expression; red, up-regulated expression. (**B**) Comparison of CRC-T70 (n = 18) xenograft (Xe) with CRC-SW1222 (n = 3) minitumors (MT) gene expression profiles. GSEA plots generated using the indicated gene sets are shown. P, p-value of 1-way ANOVA. (**C**) Diagram symbolizing proteins involved in the regulation of the cell cycle. Proteins codified by genes with low expression in SCCC versus RCCC across all cancer models are shown in red whereas those with equivalent expression are indicated in gray. (**D**) Gene expression in RCCC (blue bars) and SCCC (green bars) from CRC-SW1222 (n = 3) minitumors was performed by qPCR. Data is represented as mean ± SD of triplicates from three independent experiments. *P≤0.05; **P ≤ 0.01 of 2-tailed Student's *t* test. (**E**) Cell cycle analysis of RCCC and SCCC isolated from sphere cultures of CRC-SW1222-H2BeGFP cells. Scale bar, 100 μm. Frequency histograms of RCCC and SCCC showed distribution of cells in the three major phases of the cell cycle.

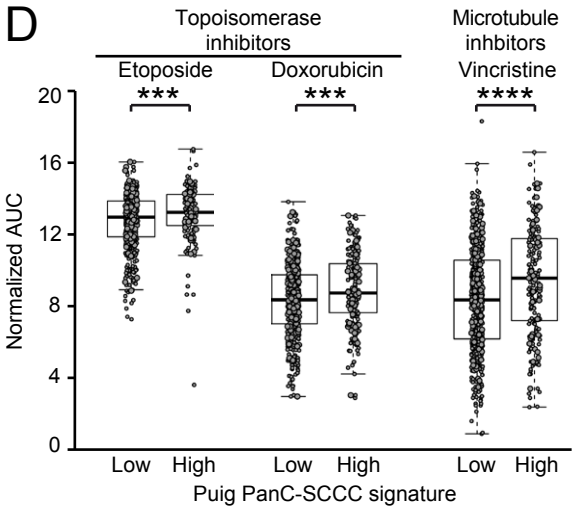
Puig et al.

Supplemental Figure 4





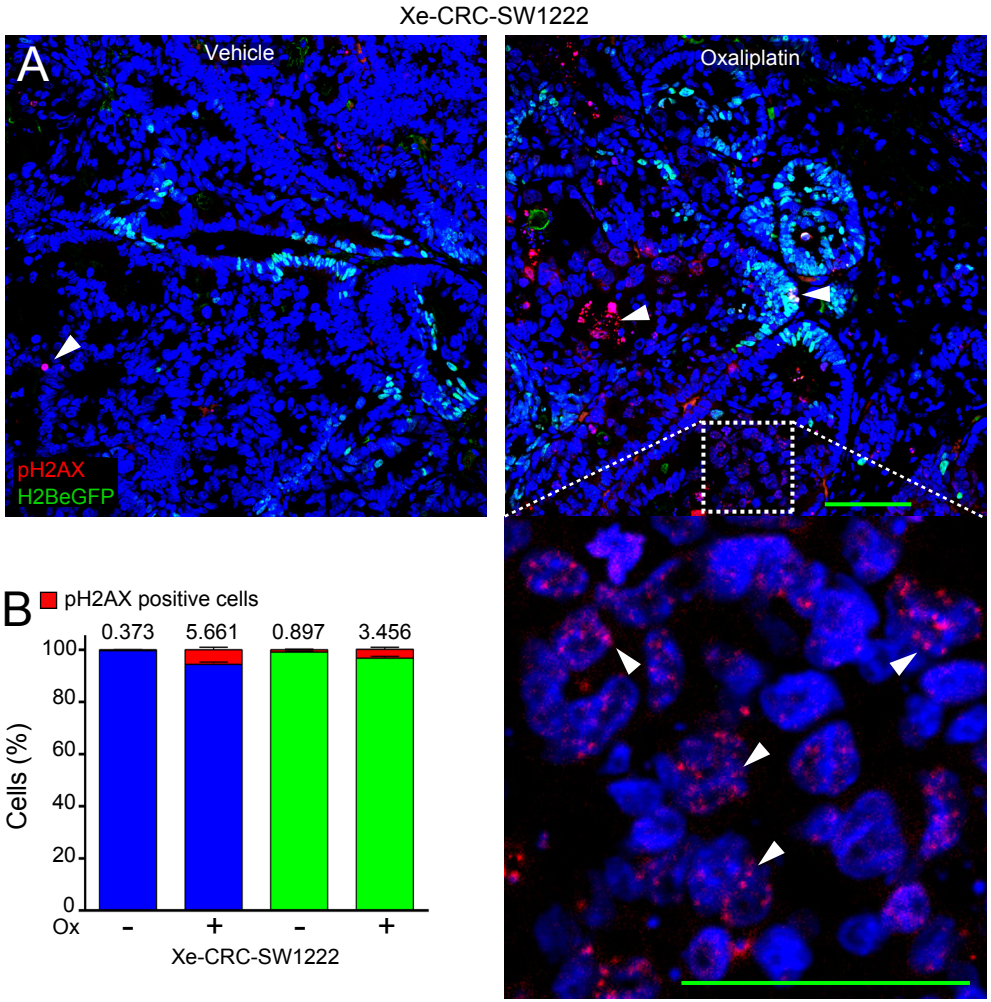




Supplemental Figure 4. **SCCC express low levels of multiple genes related with DNA replication and chromosome segregation that are direct targets of chemotherapy.** (**A, B**) Diagrams representing DNA replication fork (A) or centromeres (B) and proteins codified by genes lower (red) or equally (grey) expressed in SCCC than in RCCC across all cancer models. (**C**) Gene expression in RCCC (blue bars) and SCCC (green bars) from CRC-T70-H2BeGFP (n = 18) xenografts was performed by qPCR. Data is represented as mean ± SD of triplicates from three independent experiments. *P ≤ 0.05; **P ≤ 0.01; *** P ≤ 0.001 of 2-tailed Student's *t* test. (**D**) Sensitivity to topoisomerase and microtubule inhibitors (Area Under the Curve, AUC) of cancer cell lines (Cancer Therapeutics Research – Broad Institute) according to PanC-SCCC signature scores. ***P ≤ 0.001; ****P ≤ 0.0001 of adjusted Wilcoxon test.

Puig et al.

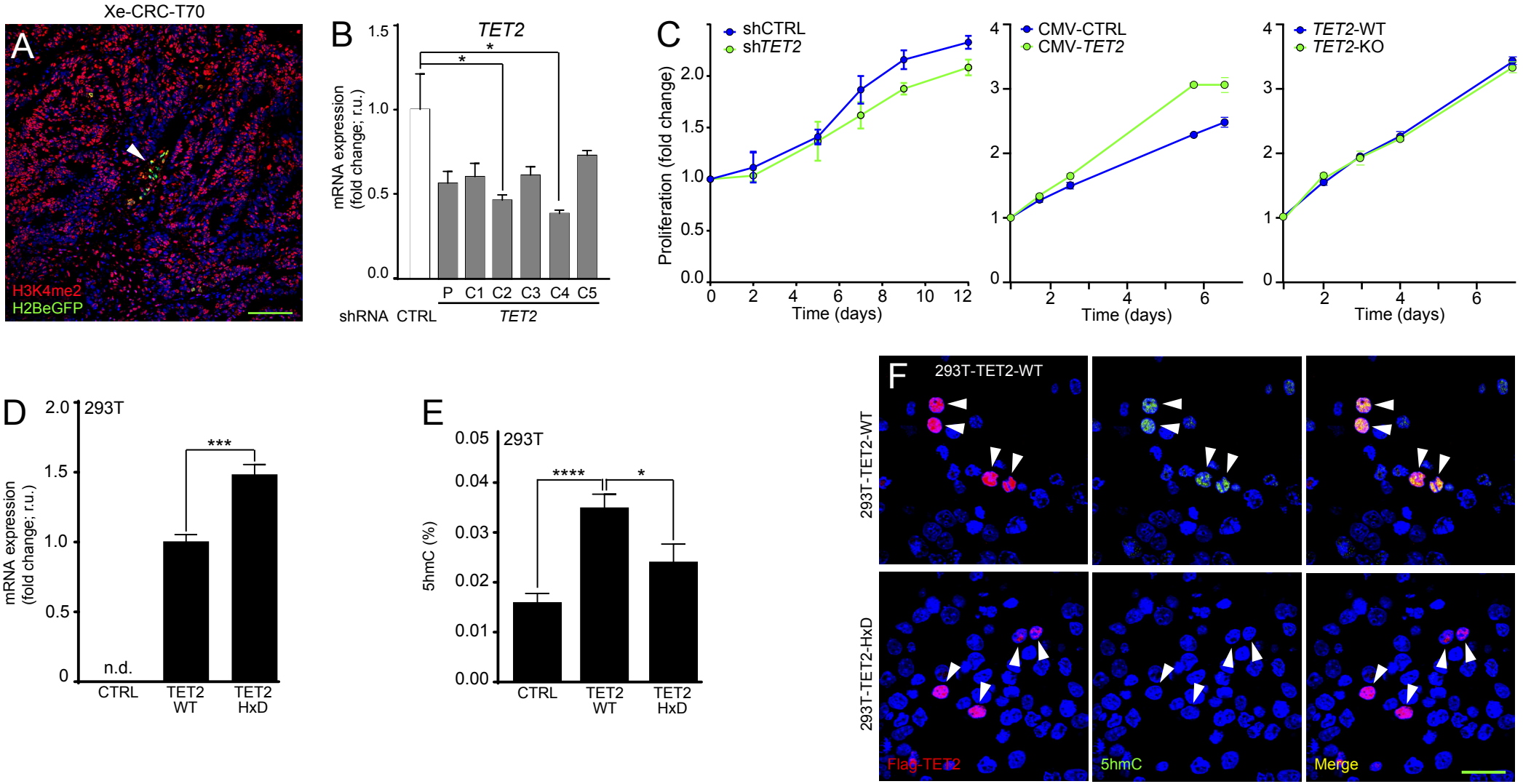
Supplemental Figure 5



Supplemental Figure 5. SCCC accumulation after oxaliplatin treatment is independent of DNA damage. (**A**) Immunofluorescence analysis of DNA damage (pH2AX) in SCCC (H2BeGFP positive cells) and RCCC (H2BeGFP negative cells) of CRC-SW1222-H2BeGFP xenografts (n = 6 per group) treated or none with oxaliplatin. Arrowhead, pH2AX positive cells. Arrowheads in magnification picture point discrete nuclear foci of pH2AX staining. Hoechst was used as counterstaining. Scale bar, 100 μm. Magnification scale bar, 20 μm. (**B**) Histological quantification analysis of pH2AX colocalization with H2BeGFP (SCCC) in xenografts represented in panel A. The percentage of pH2AX positive cells (red bars) in RCCC (blue bars) and SCCC (green bars) is indicated above each bar. Data are represented as mean ± SEM of six independent replicates.

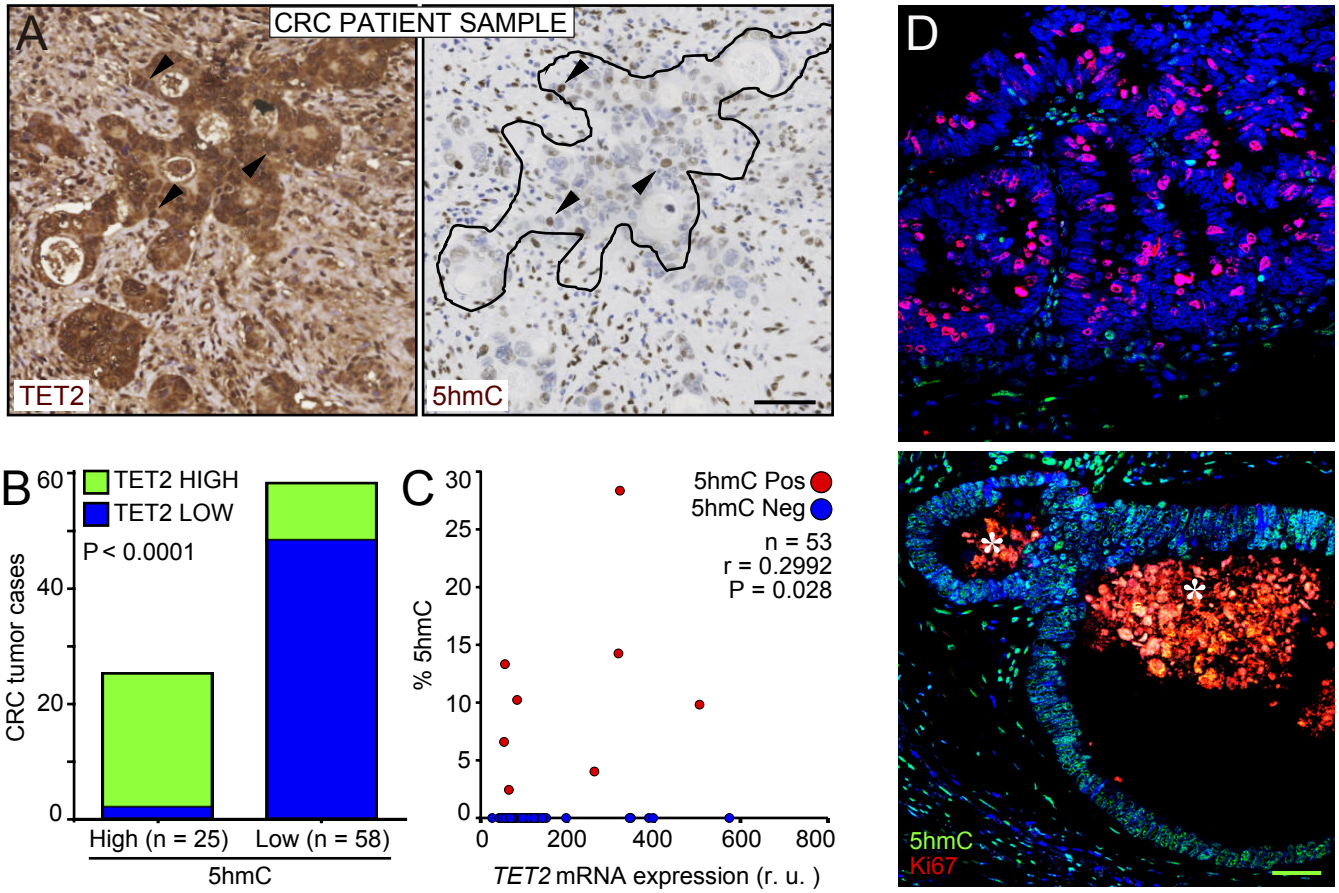
Puig et al.

Supplemental Figure 6



Supplemental Figure 6. Elimination of TET2 activity in cancer cells. (**A**) Representative immunofluorescence picture of H2BeGFP and KMT2E enzymatic product (H3K4me2) in a xenografts (Xe) generated from CRC-T70-H2BeGFP cells upon a doxycycline (DOX) pulse-chase. Arrowhead, SCCC. (**B**) Expression of *TET2* was measured by qPCR in CRC-SW1222 cells infected with five (C1-C5) different lentiviral shRNA constructs against TET2. CTRL, non-silencing (white bar); sh*TET2* (grey bars), separately (C1 to C5) or together (P, pool) shRNA constructs. (**C**) Proliferation of CRC-SW1222 cells modified for TET2 expression. WT, wild type; KO, Knock out; CTRL, control; CMV, cytomegalovirus promoter. (**D**) Expression of exogenous *TET2* was analyzed by qPCR in HEK293T cells transiently transfected with an empty vector (CTRL) or with vectors expressing WT or catalytically inactive TET2 (HxD). n.d., not detected. (**B, D**) Data is represented as mean ± SD of triplicates from three independent experiments. (**E**) Levels of 5hmC were measured in the same cell lines analyzed in panel D. (**C, E**) Data is represented as mean ± SEM of three independent replicates. (**B, D, E**) *P ≤ 0.05; ***P ≤ 0.001; ****P ≤ 0.0001 of 2-tailed Student's *t* test. (**F**) Immunofluorescence analysis of exogenous TET2 (Flag-TET2, red) colocalization with 5hmC (green) in HEK293T cells transiently transfected with WT or catalytically inactive TET2 (HxD) expression vectors. Exogenous WT or mutant (HxD) TET2 are identified using an antibody against the Flag tag. Arrowheads, exogenous WT or HxD TET2 positive cells. (**A, F**) scale bars, 100 µm.

Supplemental Figure 7



Supplemental Figure 7. TET2 and 5hmC in CRC. (**A, B**) TET2 expression and 5hmC content were evaluated by immunohistochemistry in serial tissue sections from 83 patients´ colorectal carcinomas. (A) Representative picture of a tumor presenting high expression of TET2 and its corresponding serial-section analyzed for the 5hmC content. Arrowheads, double TET2 and 5hmC positive cells. (B) Plot indicating the number of colorectal (CRC) tumor cases classified by different levels of TET2 and 5hmC contents. P, p-value. (**C**) Dispersion plot correlating the percentage of 5hmC quantified by immunofluorescence, versus normalized expression values for *TET2* mRNA analyzed by the nCounter platform in 53 patients´ colorectal carcinomas (validation cohort). n, number of samples; r, Pearson coefficient; P*,* p-value; r.u., relative units. (**D**) Co-immunofluorescence staining of 5hmC and Ki67 in CRC tumor samples reveals that Ki67-positive cells stain negative for 5hmC (up) and vice versa (bottom). Asterisk, necrotic debris. (**A, D**) scale bars, 100 µm.