Supplementary materials and methods

The Long Noncoding RNA *CCAT2* induces chromosomal instability through BOP1 - AURKB signaling

Baoqing Chen, Mihnea P. Dragomir, Linda Fabris, Recep Bayraktar, Erik Knutsen, Xu Liu, Changyan Tang, Yongfeng Li, Tadanobu Shimura, Tina Catela Ivkovic, Mireia Cruz De los Santos, Simone Anfossi, Masayoshi Shimizu, Maitri Y. Shah, Hui Ling, Peng Shen, Asha S. Multani,
Barbara Pardini, Jared K. Burks, Hiroyuki Katayama, Lucas C. Reineke, Longfei Huo, Muddassir Syed, Shumei Song, Manuela Ferracin, Eiji Oki, Bastian Fromm, Cristina Ivan, Krithika Bhuvaneshwar, Yuriy Gusev, Koshi Mimori, David Menter, Subrata Sen, Takatoshi Matsuyama, Hiroyuki Uetake, Catalin Vasilescu, Scott Kopetz, Jan Parker-Thornburg, Ayumu Taguchi, Samir M. Hanash,
Leonard Girnita, Ondrej Slaby, Ajay Goel, Gabriele Varani, Mihai Gagea, Chunlai Li, Jaffer A. Ajani, George A. Calin

Supplementary Material and Methods

Cell culture

Human colon cancer cell lines HCT116, KM12C, KM12SM, COLO320, DLD-1, HT29 and gastric cancer cell lines AGS and KATO-III (**Supplementary Table 7**) were obtained from the American Type Culture Collection. Of note, KM12SM cell is the spontaneous liver metastasis of the KM12C cells [1]. HCT116 and HT29 cells were cultured in McCoy's 5A, COLO320, DLD-1, AGS and KATO-III in RPMI-1640, KM12C and KM12SM in DMEM with 4.5 g/L glucose (10% fetal bovine serum and 1 % antibiotics) at 37°C with 5% CO_2 and 95% humidity. All cell lines were validated by the Characterized Cell Line Core at The University of Texas MD Anderson Cancer Center using STR DNA fingerprinting.

CCAT2 transgenic mice

CCAT2 transgenic mice were generated as previously described[2] . The mice used for this study were both females and males of 7-9 months of age. Briefly, a human 1.7-kb cDNA of *CCAT2* was cloned into a vector backbone which contained the CAG promoter, eGFP reporter gene and an IRES element. The *CCAT2*-vector was inserted randomly into the genome of C57BL/6N mice by pronuclear injection by the MDACC Genetically Engineered Mouse Facility. The founders were mated with WT C57BL/6N mice. Pups were selected for presence of the transgene by PCR on tail-extracted DNA according to standard protocols. All the protocols and experiments were conducted according to the guidelines of the MDACC Institutional Animal Care and Use Committee.

AOM/DSS mouse model

Because of the sex-related differences in the chemically induced colorectal cancer model [3], 7-8-week-old male mice of C57BL/6N strain [4-6], wild-type or CCAT2 transgenic, were used. Azoxymethane (AOM) was purchased from the Sigma-Aldrich (St Louis, MO). Colitis-grade Dextran Sodium Sulfate (DSS) with a molecular weight of 36000–50 000 was purchased from MP Biochemicals. All mice were initially treated with a single intraperitoneal injection of AOM (10 mg/kg of body weight). Two days after AOM administration, the mice received 3% DSS in their drinking water for 5 days, followed by a 2-week rest period without DSS. This scheme was repeated for a total of four DSS administrations. Mice were sacrificed one week after the last DSS administration. At this time, mice were dissected, the colon was excised and flushed with saline solution. The entire colon length from the cecum to rectum was cut open longitudinally exposing

the mucosa for gross examination and taking pictures with Leica camera dissecting microscope for counting and measuring the mucosal polyps. Frozen tissue samples of grossly detected polyps and of normal colon mucosa were collected. Subsequently, the colon was fixed in 10% neutral buffered formalin for histopathological examination.

Histopathologic Evaluation

Two longitudinal serial sections of each colon were prepared histologically and examined microscopically by an ACVP certified veterinary pathologist. Histopathological evaluation included counting and size measurement of polyps (colon adenomas), and grading of hyperplastic and dysplastic changes of colonic glands with a score from 1 to 4 (1 = minimal, 2 = mild, 3 = moderate, 4 = marked). The entire colon from cecum to rectum was collected at necropsy and then was open longitudinally and gently flashed with PBS solution for cleaning the feces. The colon was attached to index-card paper for exposing the mucosa and photographing the entire colonic mucosa with the Leica stereomicroscope and camera for gross counting of the polyps. Then the colon tissue was immersed in 10% neutral buffered formalin for 48 hour fixation. Fixed colon tissues were arranged in "Swiss rolls" and cut longitudinally in 2 halves which were processed and embedded in paraffin blocks, from which 4-µm thick sections were cut and stained with hematoxylin and eosin (H&E). H&E stained slides were scanned with Aperio AT2 scanner for microscopic examination and histomorphometric evaluation of colonic polyps.

Murine colon crypts isolation and murine colon organoid culture

Crypts were isolated and colon organoids were established as previously described [7-9]. Approximately 10 cm of colon was harvested from WT and CCAT2 transgenic mice, respectively. The colon was washed with cold PBS (without Ca^{2+} and Mg^{2+}), opened laterally and cut into 2-mm pieces. The pieces were washed again several times with cold PBS with penicillin/streptomycin (Gibco, 15140122) until were clear. Then, tissue fragments were subjected to enzymatic digestion in 2 ml digestion medium containing 2 mg ml^{-1} collagenase IV (Sigma-aldrich, C5138), 0.1 mg ml^{-1} dispase type II (Sigma-Aldrich, D4693), 10 µM Y-27632 (StemCell, 72304), 100 U/100 µg ml^{-1} penicillin/streptomycin and 10% FBS in DMEM medium (Gibco, 11995065), on an orbital shaker at room temperature for 30 minutes. The tissue mixture was then filtered through a 70-µm cell strainer into a 50 ml conical tube, and centrifuged at 200 x *g* for five minutes at 4°C. The pellets were resuspended in 10 ml cold DMEM medium containing penicillin/streptomycin and FBS, and centrifuged again at 200 x *g* for five minutes at 4°C.

Isolated crypts were counted using a hemacytometer with an inverted microscope, and embedded in growth factor-reduced Matrigel (Corning, catalog # 356231), diluted 3:4 in organoid culture media and seeded into 24-well plates (Corning™ 3526) at a density of 250–500 crypts in 50 µl total volume per well, and overlaid with 500 µl organoid culture media onto the Matrigel after incubation for 20 minutes at 37 °C.

Fresh medium was added every 2 or 3 days. Colon organoids were observed and treated at proper times. Outgrowing organoids were passaged every 7-days after mechanical and TrypLE™ Express (Gibco, 12604021) disruption. The organoids were washed several times with centrifugation at 200 g at 4 °C. The pellets were suspended in Matrigel with a dilution ration 1:4 and seeded as described above.

Murine colon organoids were cultured in Advanced DMEM/F12 medium (Gibco, 12634010) containing 10 mM HEPES (Invitrogen, 15630-056), 1% GlutaMAX (Invitrogen, 35050), 100 U/100 µg ml^{-1} penicillin/streptomycin, supplemented with 50 ng ml^{-1} human EGF (PeproTech, 315-09), 200 ng ml^{-1} Noggin (Peprotech, 250-38), 500 ng ml^{-1} R-spondin (BioLegend, 783606), 1 mM *N*-acetyl-L-cysteine (Sigma-Aldrich, A9165), 1× N2 (Gibco, 17502-048), 1× B27 (Gibco, 17504-044), 10 nM gastrin (Sigma-Aldrich, G9145), 10 mM Nicotinamide (Sigma-Aldrich, N0636), 10 µM Chiron (Sigma-Aldrich, SML1046), 500 nM A83-01 (StemCell, 72024), 10 µM SB202190 (Sigma-Aldrich, S7067), 10 µM Y-27632, 10 nM prostaglandin E2 (Selleck Chemicals, S3003), and 1× Primocin (InvivoGen, ant-pm-1)]. Organoids cell viability was measured every three days after treatment by using CellTiter-Glo® 3D Cell viability assay (Promega, G9683) according to manufacturer's instruction.

Gastric cancer PDX and PDO

GA-080417 cell line (referred to as #1) was established directly from patient ascites by culturing the ascites cells in RPMI (7% FBS, 1% antibiotics) and expanding them for more than 10 passages. This cell line was then injected to SCID mice by both subcutaneous ($1x10^4$ ~$1x10^5$) and orthotopic ($1x10^5$) administration to investigate their tumorigenicity. GA-080417 cells can grow into tumors after both subcutaneous and orthotopic injection.

In contrast, GA-082517 cell line (referred to as #3) was established from PDX tumor that was generated by subcutaneous injection of GA-082517 patient ascites cells. Tumor cells were disassociated from PDX tumor tissue and cultured in RPMI (7% FBS, 1% antibiotics) followed by

about 10 passages. The GA-082517 cell line was tested for tumorigenicity by both subcutaneous ($1x10^3$~$1x10^5$) and orthotopic ($1x10^5$) administration in SCID mice and was found to have tumor growth through both administration routes.

GA-051816 cell line (referred to as #2) was previously reported, for more detail refer to Song et al. [10].

Colon cancer PDX

Colon cancer PDX were established as previously described [11]. About 6-8-weeks old female NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ (NSG) mice were maintained in the MDACC animal facilities following standard animal regulation and strict health control. Rodent care and housing were in accordance with institutional guidelines and regulations as well as according to Institutional Animal Care and Use Committee approved animal protocols. Patient tumor (PX) specimens acquired at MDACC were engrafted into NSG mice. One tumor fragment (~50 mm^3) per mouse was implanted subcutaneously into the flanks of mice anesthetized under 2–4 % isoflurane/O_2 inhalation. Primary tumor xenografts (P0) growth was monitored and documented twice a week, with the date of first palpable growth noted. When tumor burden reached 1500 mm^3, mice were euthanized for tumor collection. Sections of these tumors were transplanted into new mice for PDX establishment (P1). Tumor samples at each passage were collected for histology, protein, and genomic analysis. Several PDX samples were collected in freezing media (CryoStor CS10) for storage in liquid nitrogen.

Patient sample collection

Cohort A (**TCGA colorectal cancer cohort**) consisted of 537 CRC cases with clinical and mRNA expression information. Among them 40 had mRNA data for matched normal tissues. The clinical information was retrieved from [12]. The information regarding the MSS/MSI status for 275 patients was obtained from the Cancer Genome Atlas [13]. The clinical characteristics were listed in **Supplementary Table 1**. For this cohort of patients, fragments per kilobase millions (FPKM) quantification mRNA-seq data from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/) for *BOP1*, *PES1* and *WDR12* genes were downloaded and log2 transformed.

For **Cohort B**, RNA samples from frozen cancer and matched non-neoplastic tissues of resected specimen from 100 CRC patients were obtained from the Department of Surgery and Science

(Department of Surgery II), Kyushu University Hospital. Non-neoplastic tissues were obtained from the resected specimen and were sufficiently far enough from the primary tumor. Histological diagnosis was made according to the World Health Organization criteria and pathological staging was done in accordance to the tumor-node-metastasis (TNM) classification system. Cases with stage IV were patients with synchronous distant metastasis when undergoing surgery. The details of the patient's clinical characteristics were listed in the **Supplementary Table 2**.

Cohort C consisted of RNA samples isolated from tumor and adjacent normal colon tissue samples of a total of 126 CRC patients from the Department of Comprehensive Cancer Care, Masaryk Memorial Cancer Institute, Czech Republic. Cohort C was composed of 24 MSI-H and 102 MSS (MSS/MSI-L) sporadic CRC patients. pTNM classification of the patients was done based on pathology reports and histological slides (**Supplementary Table 3**). The patients underwent standard surgical procedure and adjuvant therapy was added when necessary (stage II with risk factors or stage III). In case of advanced disease at the time of diagnosis, patients received adjuvant treatment according to the oncologist choice, following the recommendations of national guidelines.

In the **Cohort D,** a total of 206 fresh frozen tissue specimens, which encompassed primary colorectal adenocarcinoma tissues from 19 MSI-H and 187 MSS (MSI-L and MSS) cancer patients, were collected from Tokyo Medical and Dental University, Japan. Patients undergoing resection of their primary tumor that was histologically confirmed to be a stage II and III CRC and classified as MSI-H or MSS (MSI-L and MSS) were included in this study. Details of the clinical and pathological features of the included patients are shown in **Supplementary Table 4**.

Cohort E consisted of 10 paired samples, normal colon mucosa and colon tumor. The samples were obtained from the Ruder Boskovic Institute, Croatia. Tissue samples were obtained from fresh surgical specimens frozen in liquid nitrogen and stored at -80°C. The samples were histologically confirmed prior to use. Nine out of ten samples were classified as MSS and for one sample the status was not available (N/A). The *CCAT2* expression for this cohort was previously reported [14].

CIN index (CINdex) analysis

Segmented copy number data (as on hg19 reference genome), along with RNA-seq gene expression counts from TCGA Colon Adenocarcinoma (COAD) and TCGA Stomach Adenocarcinoma

(STAD) patients were downloaded from the Broad Institute Firebrowse platform (http://firebrowse.org/). An intersection was done to select only those 280 COAD tumor samples that had both copy number and gene expression data. An intersection was done to select only those 412 STAD tumor samples that had both copy number and gene expression data.

Bioconductor package CINdex (http://bioconductor.org/packages/CINdex/) [15] was applied on the segmented copy number data which enabled to characterize genome-wide DNA copy number alterations as a measure of chromosomal instability. This package calculated genomic instability at the chromosome and cytoband level. Only data from autosomes were considered for this analysis. A threshold of 2.25 and 1.75 in the un-normalized setting was used to define gains and losses respectively.

The chromosome CIN and gene expression data from 280 COAD and 412 STAD patients were shifted by 1, respectively 2 and then converted to the log base 2 scale. A Pearson correlation analysis was performed between the chromosome CIN and the gene expression data and results with a p-value less than 0.05 were short listed for further inspection. A customized correlogram table was plotted where each cell represented a correlation between two variables. Positive correlation was represented by orange color while negative correlation represented by purple colors. The cells were colored white if p-value of a correlation test was not significantly different from zero. All analyses were performed using the R statistical platform (https://www.r-project.org/).

Genomic instability analysis

Chromosome analyses were performed as previously reported [2]. Cells/mouse organoids were plated into 10 mm plates. After reaching 60-70% confluence, cells were first exposed to colcemid (0.04 mg/mL) for 1-2 hours at 37°C and to hypotonic treatment (0.075 M KCl) for 20 min at room temperature. Afterwards, cells were fixed in a methanol and acetic acid mixture (3:1 by volume) for 15 min and washed three times with the fixative. Air-dried preparations were made and the slides were stained with 4% Giemsa. The slides were analyzed for chromosomal aberrations, including chromosome and chromatid breaks, fusions, fragments and tetraploidy. A minimum of 35 metaphases were analyzed from each sample. Images were captured using a Nikon 80i microscope equipped with karyotyping software from Applied Spectral Imaging, Inc. Carlsbad, CA, USA.

Immunofluorescence staining

Cells were cultured in 8-well chamber slides (Ibidi). To check the aberrant spindles, cells were synchronized by culturing in the medium with nocodazole (100ng/ml) for 14 hours and then were released to process to G2/M phase by removal of nocodazole. Cells were fixed with 4% paraformaldehyde for 30 min at room temperature and then washed with PBS for three times, followed by blocking and permeabilization with blocking buffer, which includes 4% BSA and 0.3% Triton X-100, for 1 hour at room temperature. For the analysis of the spindle apparatus, cells were incubated overnight at 4°C with mouse anti-α-tubulin (1:16000; CST) and then with anti-mouse secondary antibodies (Alexa Fluor 647; Invitrogen) for 1 hour at room temperature. Slides were finally mounted with mounting medium containing DAPI (Abcam). For analyzing the anaphase bridges, cells were cultured, fixed, blocked, and permeabilized with same procedure and reagents. Then, the cells were stained with DAPI (Abcam 1:5000) directly, and washed three times with PBS to remove the background signal. Images were acquired with the Spin Disc Confocal microscope (Andor).

Plasmids and constructs

Mammalian expression vectors for full-length *CCAT2* and a series of mutants were constructed by subcloning the gene sequences into pCDNA3.1 (+) backbone (Life Technologies), pBabe retroviral expression vector, or MS2-24x-pCNDA vector. The full-length FLAG-tagged BOP1 human expression vector was purchased from Origene (#RC204016), and GST-tagged recombinant human BOP1 protein was purchased from Novus Biologicals. The BOP1 and CCAT2 truncated mutants were generated by using QuikChange II XL Site-Direct Mutagenesis Kit (Agilent Technologies, #200522). All constructs were confirmed by DNA sequencing at Sequencing and Microarray Facility (SMF), UT MD Anderson Cancer Center.

RNA interference

CCAT2, *BOP1*, or negative control siRNAs were purchased from Ambion. Cells were seeded into six-well plates. When the cells reached 50-70% confluence, they were transfected with 25 nM of the corresponding siRNA by Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol. RNA and proteins were collected at 48h (in the *BOP1* knock-down experiments) or 72h (in the *CCAT2* knock-down experiments) after transfection. qRT-PCR and Western blot were used to check the efficiency of knock-down.

Generation of stable clones

HCT116 cells with *CCAT2* stable overexpression ($HCT116^{CCAT2}$) were established by transfecting pcDNA 3.1 CCAT2-expression vector with Lipofectamine 2000 (Invitrogen, ThermoFisher Scientific) as previously described [16]. Empty clone was generated by transfection with the empty pcDNA3.1 vector ($HCT116^{Empty}$). In our previous papers [14, 16] these clones were termed: empty (E) and overexpressed clone 1 (OC1). DLD-1^{Empty} and DLD-1^{CCAT2} were established using the same above-mentioned method.

The cell culture supernatants with lentivirus carrying BOP1 were purchased from the MDACC shRNA and ORFeome Core of The University of Texas MD Anderson Cancer Center. HCT116, KM12SM, and HT29 cells were plated into six-well plates. Cells were infected by the cell-free supernatants containing lentivirus with 10 μg/μL polybrene (Sigma) for 48 hours and then switched to normal medium with Blasticidin S for 72 hours. Cells with successful transduction showed green fluorescence. Flow cytometry was used to select the green fluorescence-positive cells. We termed the clones with BOP1 overexpression $HCT116^{BOP1}$, $KM12SM^{BOP1}$, and $HT29^{BOP1}$.

For MYC tet-on stable clones, we first generated the HCT116 tetracycline reverse transcriptional activator (rtTA) stable clones by transducing with CMV-rtTA lentivirus and selected by neomycin (Sigma), as the activation of genes downstream of TetO induced by doxycycline is rtTA dependent. Then the stable HCT116 rtTA clones were transduced with tet-on C-MYC lentivirus and selected by puromycin (Sigma). CMV-rtTA lentivirus and inducible tet-on C-MYC lentivirus were purchased from Cellomics Technology.

RNA extraction, cDNA synthesis and quantitative real-time PCR

Total RNA was isolated using Direct-zol kit (Zymo research) following the manufacturer's protocol. Then, the cDNA was synthesized using High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. qRT-PCR was performed using SsoAdvanced™ Universal SYBR Green Supermix real-time PCR kit (Bio-Rad). Primers were synthesized by Integrated DNA Technologies (sequences are listed in **Supplementary Table 5**). The relative gene expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method. The geometric mean of *GAPDH*, *β-actin*, and U6 snRNA were used as normalizer for *in vitro* studies; *mAactb* and *Rplpo* mRNAs were used as housekeeping genes for *in vivo* experiments while for clinical samples we used *β-actin* or geometric mean of *GAPDH*, *β-actin*, and U6 snRNA as normalizer. The absolute RNA

expression of the house keeping genes showed minimal and non-statistical difference between the groups we compared.

Whole or cytosolic/nuclear protein fractionation and Western blot

Protein lysates from whole-cell pellets samples were generated using Cell Lysis Buffer (Cell Signaling Technology) that contains protease and phosphatase inhibitor cocktails (Sigma-Aldrich). Cytosolic and nuclear protein fractionations were performed by utilizing of the NER Nuclear and Cytoplasmic Extraction Reagents kit (ThermoFisher Scientific) according to the manufacturer's protocol. Protein concentration was quantified by Bradford assay (Bio-Rad). In total, 20 ug of proteins were loaded on 4-20% acrylamide CriterionTM TGXTM precast gels (Bio-Rad) and transferred to nitrocellulose membranes by semi-dry method. The membranes were incubated with the corresponding primary antibodies (listed in **Supplementary Table 6**) overnight and then incubated with the appropriate HRP-conjugated secondary antibody. Immunoreactivity was detected by incubation with ECL SuperSignal West Femto substrate (ThermoFisher Scientific), and then detected by the autoradiographic film.

Cytosolic and nuclear RNA fractionation

RNA was isolated from cellular fractions according to the previously described protocol [17]. Briefly, the cell pellet was resuspended in 380 ul Hypotonic lysis buffer (HLB: 10 mM Tris (pH 7.5), 10 mM NaCl, 3 mM MgCl2, 0.3% (vol/vol) NP-40 and 10% (vol/vol) glycerol) and the mixture was incubated on ice for 10 min. The cells were centrifuged at 1,000 x g at 4 °C for 3 min; the supernatant, which is the cytoplasmic fraction, was transferred to another tube. RNA precipitation solution (RPS: 0.5 ml of 3 M sodium acetate (pH 5.5) with 9.5 ml of ethanol) was added and the mixture was stored at -20 °C for over 1h. The remaining pellet (the nuclear fraction) was washed three times with HLB by centrifuging at 200 x g at 4°C for 2 min. TRIzol was added over the nuclear pellet and RNA was extracted using Direct-zol kit (Zymo research) following the manufacturer's protocol. The cytoplasmic fraction, after 1h, was centrifuged at 18,000 x g at 4 °C for 15 min. The pellet was washed in 70% ethanol and centrifuged again at 18,000 x g at 4 °C for 5 min. After air drying, 1 ml of TRIzol was added to the pellet and RNA was extracted. The lncRNA *NEAT1* was used as a positive control for nuclear enrichment.

γ-H2AX assay

This assay was performed with and without Bleomycin treatment. $HCT116^{Empty}$ and $HCT116^{CCAT2}$ overexpressed cells were cultured for 24 hours in 8-well chamber slides (Ibidi). In the case of Bleomycin treatment, 10 µg/ml of Bleomycin was added for two hours at 37°C. Treated and untreated cells were then washed three times with PBS and fixed by incubating them in ice cold methanol for 5 minutes at room temperature. This step was followed by blocking and permeabilization with 1% BSA-PBST at 37°C for 30 minutes. Next, the cells were incubated overnight at 4°C with the primary antibody of H2AX (rabbit polyclonal IgG) followed by one-hour incubation at room temperature with secondary antibody, goat polyclonal IgG conjugated with FITC. Finally, cells were stained with Hoechst 33342, before adding the cover slip. Images were acquired with the Spin Disc Confocal microscope (Andor).

Cellular Senescence Staining

$HCT116^{Empty}$ and $HCT116^{CCAT2}$ cells were seeded in 6-well plates to a seeding density of 3×10^5 cells per mL, and a final chamber volume of 1 mL. When the cells reached 70-80% confluence, were stained for senescence using a β-Galactosidase Staining Kit (Cell Signaling Technology #9860) following the manufactures instructions. Cells were mounted for analysis with 70% glycerol. All the cells from the circumference of three wells were analyzed for each clone.

SHAPE Analysis

RNA secondary structure probing was started by mixing stock *CCAT2* RNA and folding buffer (200 mM NaCl, 100 mM HEPES, 0.2 mM EDTA, pH 8.0), followed by incubation at 37°C with refolding buffer for 30 min (100 mM NaCl, 50 mM HEPES, 16.5 mM MgCl2, pH 8.0). The SHAPE reaction was then started by adding NMIA (+) (32.5mM, 65mM or 130mM), or DMSO as control (-). Samples were then incubated for 45 min at 37°C, precipitated with ethanol and glycogen. Reverse transcription was performed using Super Script III reverse transcriptase (Invitrogen). Electrophoresis on an 8% (vol/vol) polyacrylamide gel was then performed to separate fragments. Band-intensities were visualized by gel electrophoresis or capillary electrophoresis and were quantified using SAFA, version 1.1 Semi-Automated Footprinting Analysis [18]. SHAPE reactivity data from capillary and gel electrophoresis were incorporated as a SHAPE constraint file in the RNA structure folding program, and the 20 lowest energy structures based on those constraints were generated [19, 20]. Structures were calculated with RNAstructure default secondary structure options. Each structure image presented in this manuscript was rendered using VARNA.

Cycloheximide Chase Assay

Protein half-life studies were performed as previously described [21]. Cells were cultured in six-well plates and then transfected with *CCAT2* or empty vector for 24 hours and then cultured with cycloheximide (CHX, 100μg/ml), which blocks the translation of mRNA. After culturing with cycloheximide for 0, 4, 8, 12, and 20 hours, the cells were collected and lysed with lysis buffer (CST) for protein extraction. The degradation of BOP1 was then detected by Western blotting analysis.

Proliferation assay

Cell Counting Kit-8 (Enzo) was used to determine the cell viability according to the manufacturer's protocol. Cells were seeded at a density of 2500 cells per well in 96-well plates overnight. For the proliferation assay in knock-down experiments, cells were transfected with the corresponding siRNA and then seeded in 96-well plates. After culturing for 0, 24, 48 and 72 hours, 10ul of CCK8 was added to each well and then incubated at 37°C for 2 hours. The absorbance values at 450 nm were measured to represent the cell viability.

MTT assay – 5- Fluorouracil.

In vitro chemoresistance to 5-Fluorouracil (5-FU) of $HCT116^{CCAT2}$ clones versus $HCT116^{Empty}$ was assessed by MTT. Briefly, cells were plated 24 hours prior to treatment in 96-well microculture plates. After 24 hours, 3 different doses of 5-FU were added (2, 20 and 40 μM) to the supernatant without changing the medium. After 48 hours, the MTT reagent (Sigma) was added to each well and incubated for 3 hours at 37°C. The optical density (OD) was read at 570 nm on a microplate spectrophotometer and growth values (%) were calculated as followed (OD treated cells /OD untreated cells) x 100.

MTS assay - oxaliplatin

Cell viability of oxaliplatin treated $HCT116^{CCAT2}$ clones versus $HCT116^{Empty}$ cells was measured by The CellTiter 96® AQueous One Solution Cell Proliferation Assay (Promega). Cells were seeded in 100 μl medium into 96-well plates and incubated for 24 hours at 37°C before treatment. The cells were treated with different concentrations of oxaliplatin: 2, 20 and 40 μM (Selleckchem) for 48h. At the end of treatment, 20 μl of MTS solution was added to each well of the plate and the optical density (OD) at 490 nm was then measured using microplate reader and growth values (%) were calculated as followed (OD treated cells /OD untreated cells) x 100.

Colony formation assay

Around 500 HCT116 cells with BOP1 knock-down or overexpression and KM12SM cells with BOP1 overexpression were plated into 6-well or 12-well plates. For the knock-down experiments models, corresponding siRNAs were added and then incubated for 48 to 72 hours. Cells were then cultured with new medium for 10-14 days. The colonies were fixed with 4% paraformaldehyde solution, stained with crystal violet. Colonies with more than 50 cells were counted.

Invasion and scratch assay

Cells in serum-free media were seeded into precoated Matrigel 24-well invasion chambers (8 mm; 24 wells; BD Biosciences) according to the manufacturer's protocol. In the bottom well, 700 µl of medium with 20% FBS was added to serve as chemoattractant. Twenty-four hours after seeding, cells at the top of the insert well were removed by swabbing, the cells that penetrated the membrane and were located on the bottom side of the insert well were fixed with paraformaldehyde, stained with crystal violet, and counted. For scratch assay, cells were seeded in 6-well plates and cultured to reach around 90-100% confluency. A scratch was made by scratching a line across the bottom of each well using a 20µL pipette tip. The detached cells were then removed by washing with PBS. Migration into the open area was recorded at 48 hours after scratching.

Polysome profiling

Polysome profiling was done in accordance with previously published protocols [22, 23]. Briefly, 10^6 cells were treated 10 minutes at 37°C with cycloheximide at 100ug/ml. Cells were then trypsinized and collected under ice cold conditions. Cell lysates were prepared using dounce homogenization with 50 strokes of the dounce. The lysate was then precleared at 1200xg for 10 minutes, and equal OD254 units were loaded onto a 17-50% sucrose gradient. The sucrose gradients were fractionated on a gradient fractionator using a UA-6 detection system (Teledyne ISCO). For experiments to quantify absolute amounts of 40S and 60S subunits, a modified procedure was followed essentially as described [24]. In this procedure, the lysis buffer consisted of 20mM Tris-HCl, pH 7.5, 100mM KCl, and 5mM $MgCl_2$, and the sucrose was mixed into that buffer for pouring the sucrose gradients. Lysates were precleared as described above, and EDTA was added to a final concentration of 50mM prior to fractionating.

***In vitro* RNA-protein binding assay**

Biotin-labeled full-length and truncated fragments of *CCAT2* RNA were transcribed *in vitro* with a Biotin RNA Labeling Mix Kit (Roche) and T7 or SP6 RNA polymerase (Ambion) using the PCR products as a template, treated with RNase-free DNase I (Ambion). The reaction mix was then purified with an RNA Clean & Concentrator kit (Zymo Research) and the purified biotinylated RNA was denatured at 95°C for 2 min, cooled on ice for 2 min, and then transferred in RNA structure/folding buffer at 30°C for 30 min to allow proper RNA secondary structure formation.

The RNA-protein binding assays were performed using Pierce™ Magnetic RNA-Protein Pull-Down Kit (Thermo Scientific) according to the manufacturer's instructions. Briefly, magnetic Dynabeads M-280 Streptavidin beads were washed three times with washing buffer and then immediately subjected to capture *in vitro* transcribed RNA (2 μg) as described above. To pull down recombinant proteins, the RNA-captured beads were incubated with recombinant proteins (1 μg) in binding buffer for 1 hour at RT. Beads were washed three times and boiled in 1 x reducing sample buffer. The retrieved proteins were analyzed by Western blotting.

MS2 pull-down assay

The pCCAT2-MS2, pEmpty-MS2, and pMS2-GST vectors were generated as previously reported [14]. HCT116 cells were plated into 15mm plates. After reaching 60-70% confluence, the cells were co-transfected with pCCAT2-MS2 vector or pEmpty-MS2, and pMS2-GST vector using Lipofectamine 2000 (Invitrogen, ThermoFisher Scientific). Forty-eight hours after transfection, cells were harvested and proteins were collected and quantified. In total, 1000 μg (2 μg/μl) lysate was incubated with GSH agarose beads (GE Healthcare) for 3 hours at 4°C, followed by 3 times washing with cell-lysis buffer to remove unspecific bound proteins. Beads were then suspended in SDS buffer and heated. Bound proteins were detected by Western blotting.

RNA immunoprecipitation

EZ-Magna RIP RNA-Binding Protein Immunoprecipitation Kit (Merck Millipore) was used according to the manufacturer's protocol. Cells were harvested and lysed in the RIP buffer containing the protease inhibitor. Cell lysates were then incubated with the buffer containing magnetic beads conjugated with anti-BOP1 (Abcam), anti-PES1 (Santa Cruz), anti-WDR12 (Abcam), anti-AURKB or IgG as a negative control at 4°C overnight and then washed with washing buffer for three times to remove the unspecific bounds. RNA was then isolated using the phenol:chloroform:isoamyl alcohol method and further used for cDNA synthesis and qRT-PCR to

test the presence of *CCAT2*. For *CCAT2* segments - Aurora Kinase B interaction, UV crosslinked cells were used. The lysates were treated with RNase I for exactly 3 min at 37°C then immediately transferred on ice and centrifuged. The supernatant was incubated with magnetic beads conjugated with anti-AURKB overnight. After immunoprecipitation, RNA was purified with phenol:chloroform:isoamyl alcohol to prepare cDNA. Then, CCAT2 segment specific primers were applied to perform qRT-PCR.

Gene expression analysis (GEA)

For the mouse GEA, total RNA was extracted from bone marrow (BM) cells of WT and CCAT2 transgenic mice. Labeling and hybridization of mRNAs were performed according to Affymetrix protocols. Briefly, 5 μg of total RNA was reverse transcribed with an oligo(dT) primer that has a T7 RNA polymerase promoter at the 5' end. Second-strand synthesis was followed by cRNA production with incorporation of biotinylated ribonucleotides using the BioArray High Yield RNA Transcript Labeling Kit T3 (Enzo Life Sciences). The labeled cRNA was fragmented and hybridized to Affymetrix GeneChip Mouse Genome 230 4.0 arrays. GeneSpring GX software v.13 (Agilent Technologies) was used for probe set summarization and robust multiarray average (RMA) normalization procedures. The differentially expressed genes were selected to have a >1.5-fold change difference between the compared groups (average value), a <10% FDR using Benjamini-Hochberg corrected moderated t-test and $P < .05$.

***In silico* analysis of genomic evolution**

Genomic conservation was analyzed by downloading Multiz Alignments of 100 Vertebrates from the UCSC Genome Browser (GRCh38/hg38 Assembly, chr22:46,492,389-,46,493,270). Sequence identity was determined using the CLC Genomic Workbench 8.5.4. A phylogenetic tree was generated (Algorithm = UPGMA, Distance measure = Jukes-Cantor, Bootstrap = 100 Replicates), and identity was calculated by summarizing the distance from *Homo sapiens*.

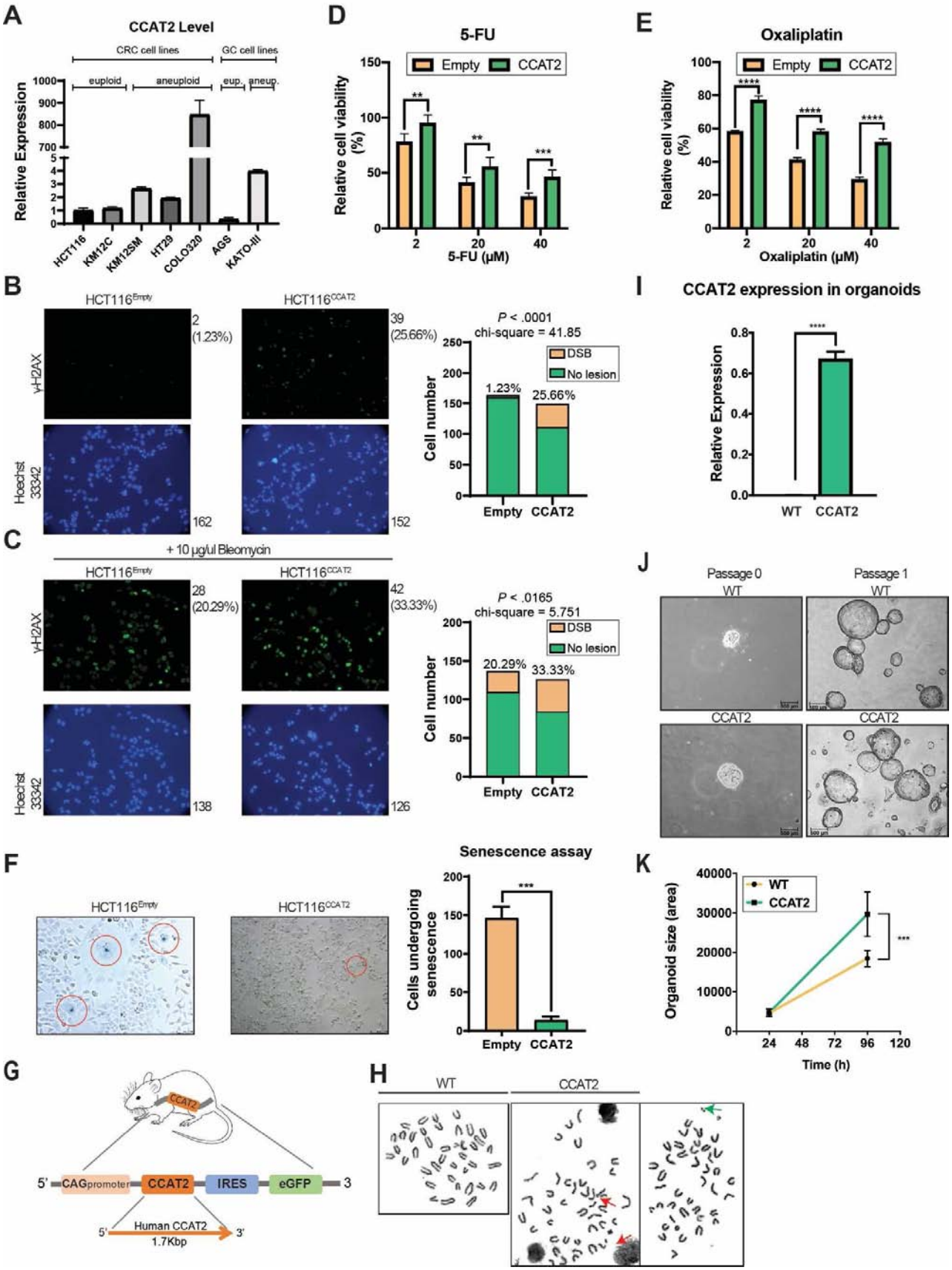
Statistical and survival analysis

Statistical analyses were carried out with GraphPad Prism 7 and SPSS software. To determine whether the data followed a normal Gaussian distribution, the Shapiro–Wilk normality test was performed. P values were determined with a paired/unpaired t-test (normal distribution) or the Wilcoxon matched-pairs signed rank test for paired non-normal distribution data and the Mann–Whitney–Wilcoxon for un-paired data (non-normal distribution). Linear correlation between gene

expressions was performed using Pearson correlation coefficient (normal distribution) or nonparametric Spearman correlation (non-normal distribution). All tests were two-sided, and P values < 0.05 were considered statistically significant.

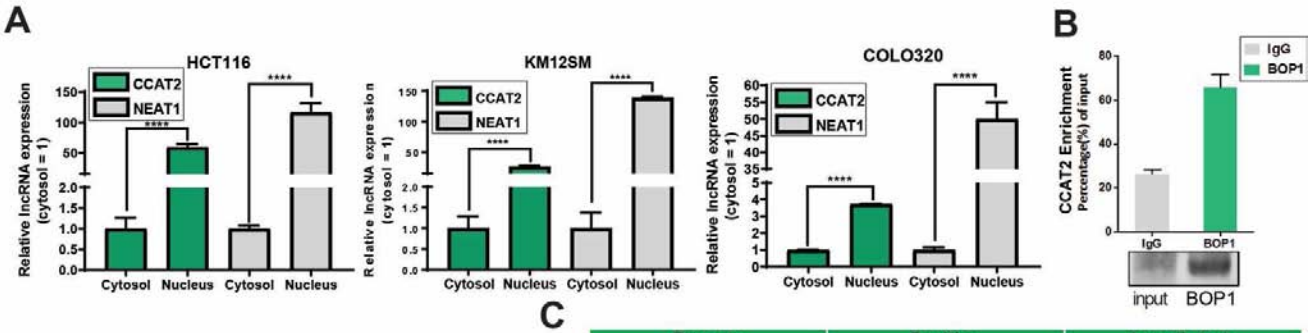
The Kaplan-Meier method was used to plot and evaluate patients' prognostic outcome. Receiver operating characteristic (ROC) curves with Youden's Index for overall survival/recurrence free survival were established to determine optimal cut-off values for *CCAT2*, BOP1, PES1 and WDR12. The log-rank test was performed for statistical univariate analysis of prognostic variables. In multivariate analyses, a Cox proportional hazard model was used to identify parameters with a statistically significant influence on survival.

Supplementary Figure 1



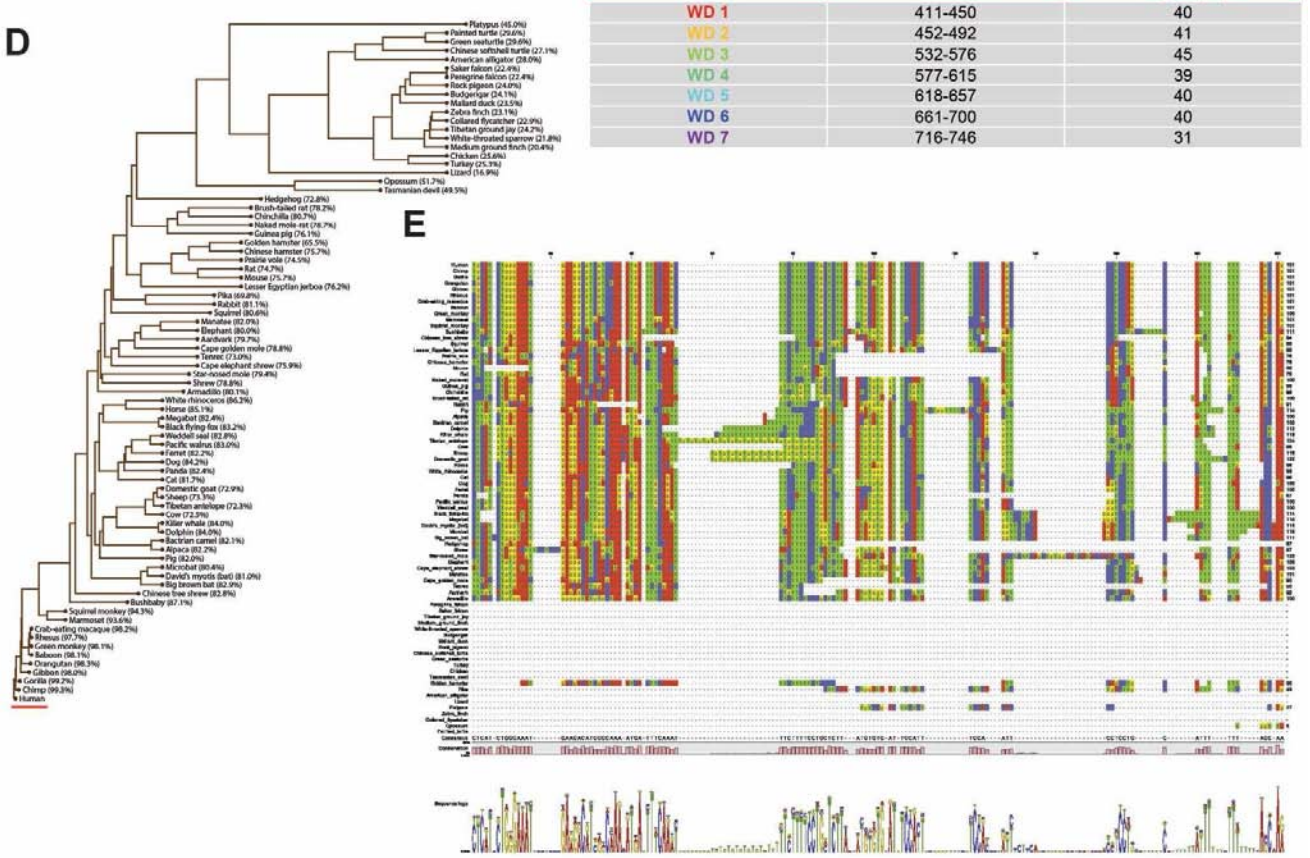
Supplementary Figure 1. Related to Figure 1. **A.** *CCAT2* expression in colorectal cancer (CRC) and gastric cancer (GC) cell lines used in the present study. **B.** Images of the γ-H2AX assay for double strand breaks (DSB) without Bleomycin treatment in HCT116^{Empty} and HCT116^{CCAT2}. **C.** Images of the γ-H2AX assay for DSB after 2 hours of exposure to 10 μg/ml of Bleomycin in HCT116^{Empty} and HCT116^{CCAT2}. **D.** Effect of 5-Fluorouracil (5-FU) on cell proliferation of HCT116^{Empty} and ^{CCAT2} clones. Cells were treated with three different concentrations of 5-FU (2, 20 and 40 μM) and after 48 hours cell viability was determined using the MTT assay. **E.** Effect of oxaliplatin on cell proliferation of HCT116^{Empty} and ^{CCAT2} clones. Cells were treated with three different concentrations of oxaliplatin (2, 20 and 40 μM) and after 48 hours cell viability was determined using the MTS assay. **F.** Representative images (in red circles) for senescence assay in HCT116^{Empty} and HCT116^{CCAT2} and the corresponding statistical analysis. **G.** Schematic representation of CCAT2-plasmid inserted into the mouse genome. **H.** Representative metaphase from the bone marrow of WT mice and CCAT2 transgenic mouse model that developed myelodysplastic syndrome. Red arrows indicate chromosomal breaks and green arrows chromosomal fragments. **I.** *CCAT2* expression in colon organoids from WT and CCAT2 transgenic mice. **J.** Representative inverted microscopy images of colon organoids of WT and CCAT2 transgenic mouse model immediately after they were established and after 1 passage. **K.** Proliferation rate of colon organoids from WT and from CCAT2 transgenic mice. Data are represented as mean values ± SD. (***$P < .001$), (****$P < .0001$). Student's t test.

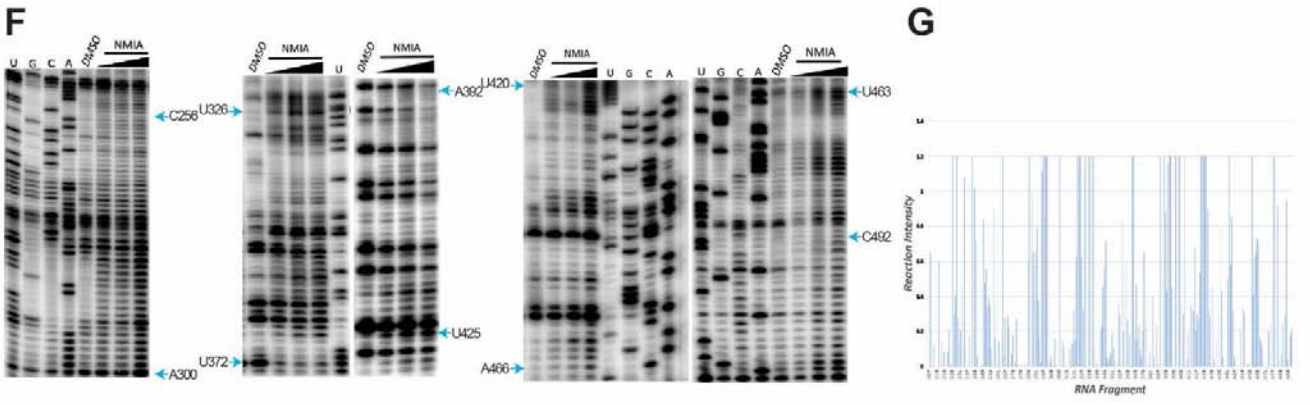
Supplementary Figure 2



C

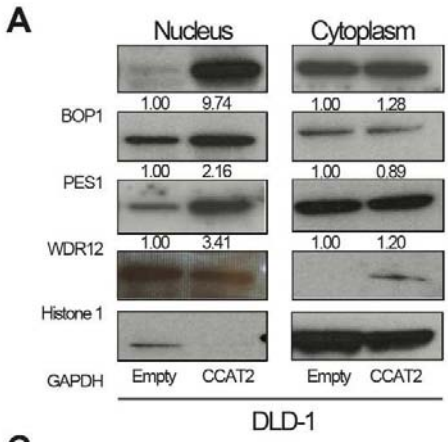
Repeats	Position	Length (aa)
WD 1	411-450	40
WD 2	452-492	41
WD 3	532-576	45
WD 4	577-615	39
WD 5	618-657	40
WD 6	661-700	40
WD 7	716-746	31

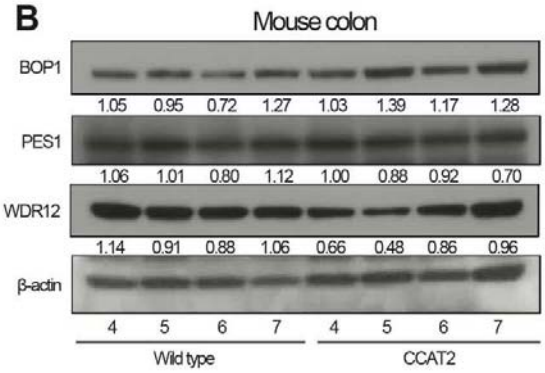


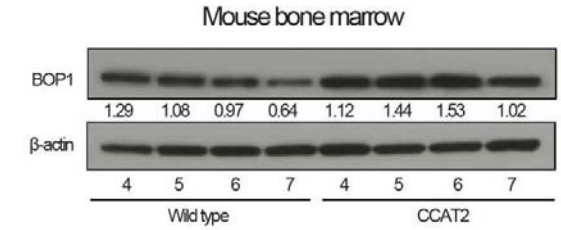


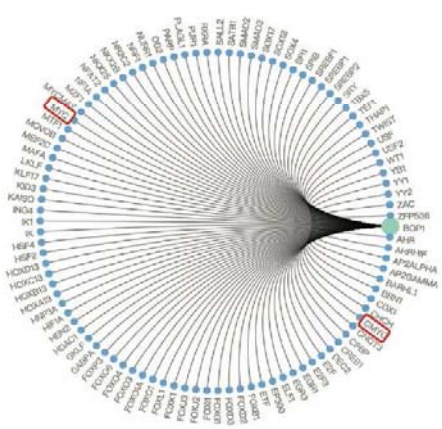
Supplementary Figure 2. Related to Figure 2. **A**. The nuclear and cytoplasmic localization of *CCAT2* and *NEAT1* (positive control) in HCT116, KM12SM and COLO320 cells as measured by qRT-PCR. **B.** RIP assays showing that BOP1 interacts with *CCAT2* in HCT116 cells. The qRT-PCR results of RIP assays are shown in the upper panel. Western-blot data, to check the immunoprecipitation efficiency, are shown in the lower panel. **C.** The positions and length of the seven WD repeats of BOP1. **D.** Evolutionary conservation of full length *CCAT2* (*Homo sapiens* is underlined in red). **E.** Conservation of the motif 3 of *CCAT2*. **F.** Representative gels for the SHAPE probing of CCAT2 330-430 fragment with different primers. Lanes from left to right were gels probed with primer 1, 2, 3, 4, 5. U, G, C, A sequencing reactions performed using dideoxy-terminating nucleotides. Control in DMSO; NMIA concentration gradient from 32.5 mmol to 130 mmol. **G.** Quantification of the SHAPE probing data gels for CCAT2 330-430 fragment. Band intensities visualized by gel electrophoresis were quantified using SAFA, version 1.1. (****$P < .0001$). Student's t test.

Supplementary Figure 3



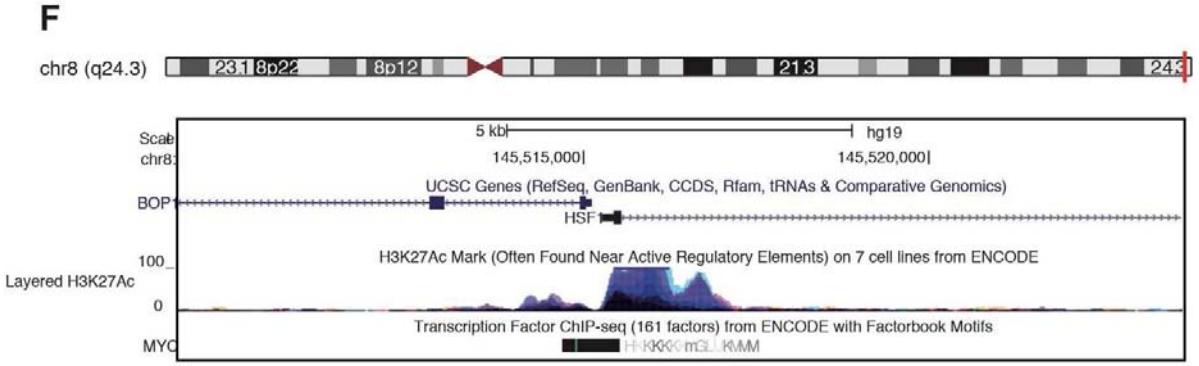


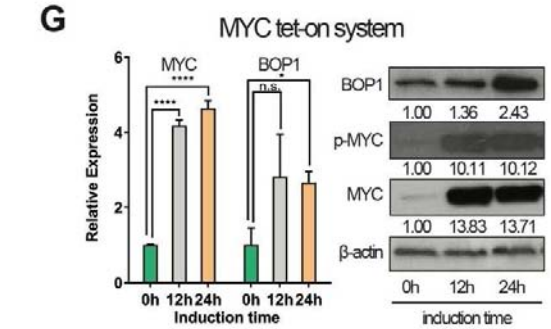




E

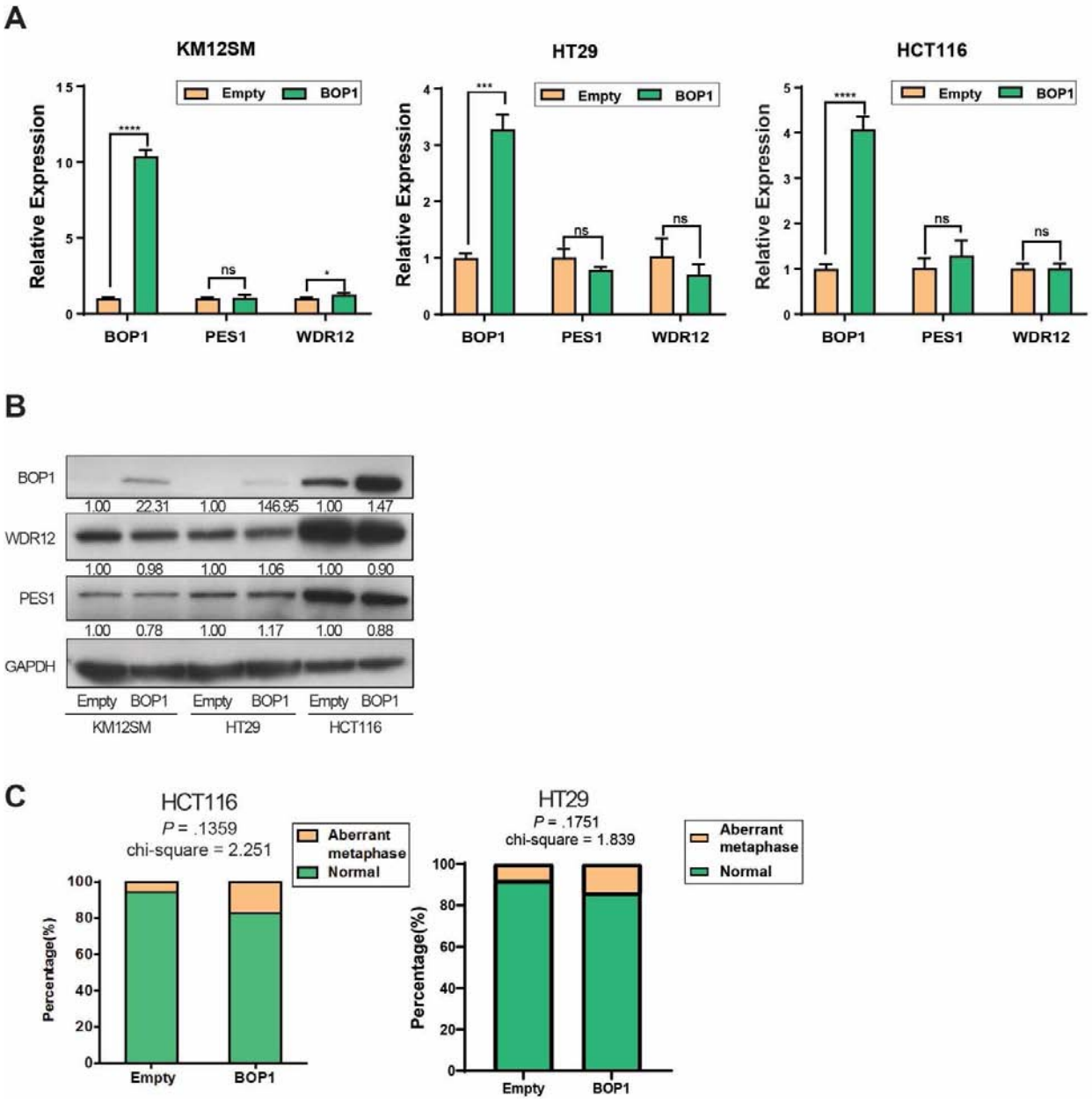
Transcript	chromosome	binding site(3's)	binding site(5's)	binding sites	prediction score(0-1)	SNP site	Methylation site
ENST00000569669.5	8	144291213	144291220	ccACGTG	1	NO	NO
ENST00000569669.5	8	144292306	144292313	gcACGTG	1	NO	NO
ENST00000569669.5	8	144291212	144291220	gcCACGTg	1	NO	NO





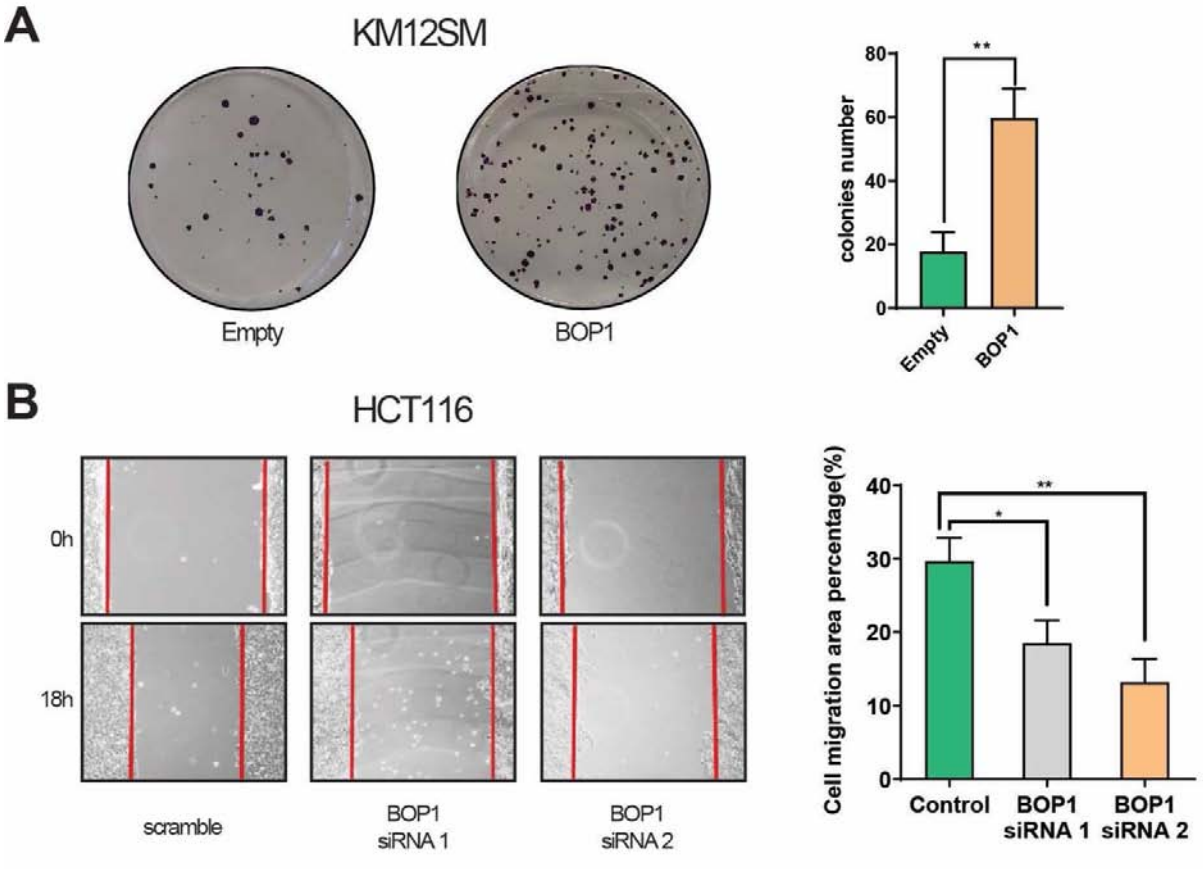
Supplementary Figure 3. Related to Figure 3. **A.** The nuclear and cytoplasmic localization of PES1, BOP1, and WDR12 (PeBoW) complex was determined by Western-blot of fractionated protein lysate from DLD-1^{Empty} and DLD-1^{CCAT2} cells. **B.** The protein expression of PeBoW complex components in the colon of CCAT2 transgenic mice (WT = 4, CCAT2 = 4). **C.** The protein expression of BOP1 in the bone marrow of *CCAT2* transgenic mice (WT = 4, CCAT2 = 4). **D.** MYC is one of the predicted transcription factors of *BOP1* (from GCBI GENERADAR)**. E.** Predicted binding sites of MYC around the genomic location of *BOP1*. **F.** CHIP-Seq data (UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly) showing MYC (Cluster Score (out of 1000): 1000, Position: chr8:145514703-145515524, Band: 8q24.3, Genomic Size: 822) binding site in the genomic region around *BOP1*. **G.** The expression of BOP1, MYC and p-MYC at 0, 12 and 24 hours in HCT116 cells with an inducible c-MYC expression system. Data are represented as mean values ± SD. (ns, not significant), (*$P < .05$), (****$P < .0001$). Student's t test.

Supplementary Figure 4



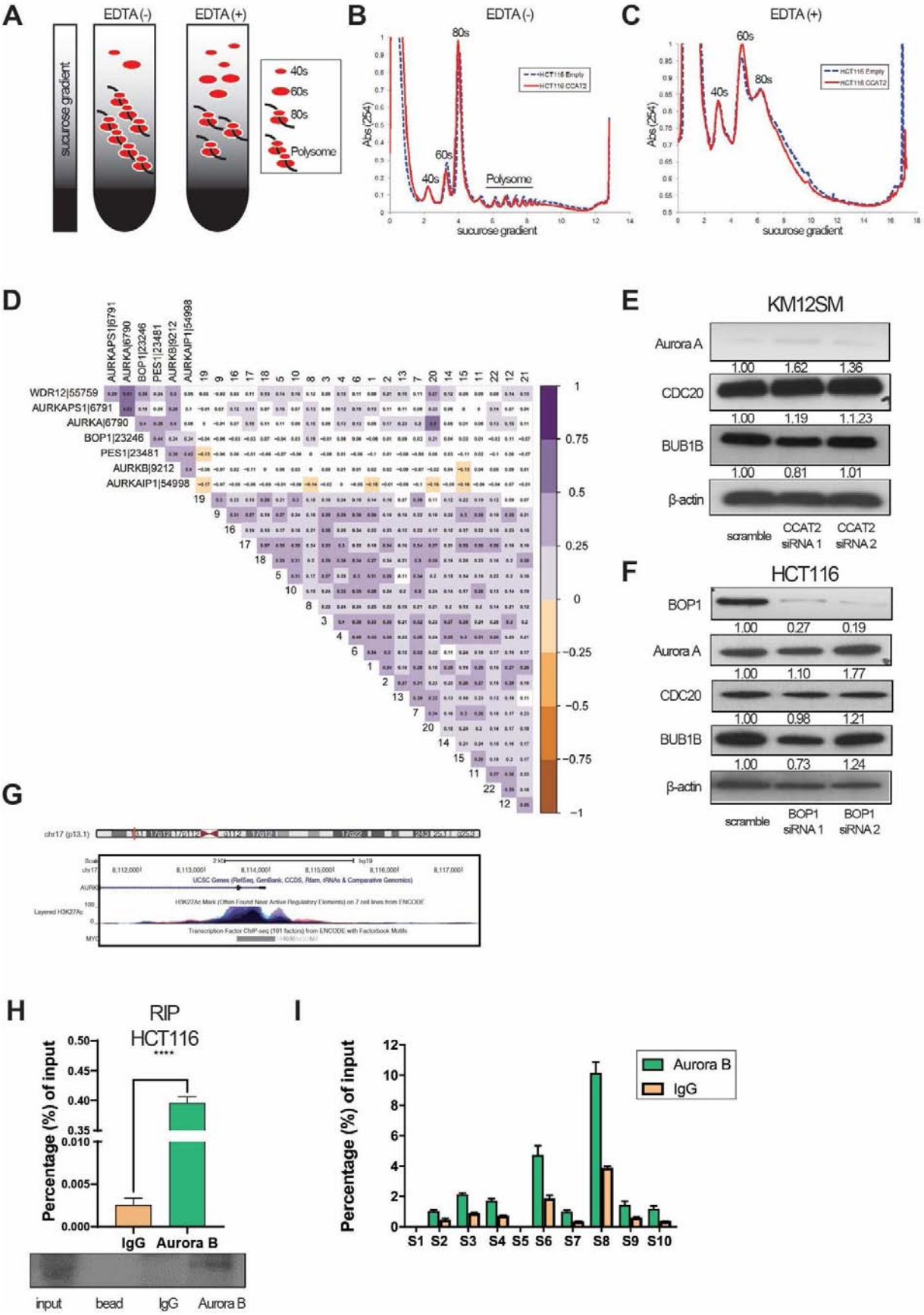
Supplementary Figure 4. Related to Figure 4. **A.** mRNA expression of PES1, BOP1, and WDR12 (PeBoW) complex members in KM12SM, HT29 and HCT116 Empty and BOP1 clones analyzed by qRT-PCR. **B.** Protein expression of PeBoW complex members in KM12SM, HT29 and HCT116 Empty and BOP1 clones analyzed by Western blot. **C.** Cytogenetic analysis showing that the frequency of cells exhibiting chromosome abnormalities (including chromosomal breaks, fusions and polyploidy) in HCT116 and HT29 Empty versus BOP1 clones after a short-term of passaging (5-10) has no significant difference. At least 35 interphase nuclei were analyzed for each clone. Data are represented as mean values ± SD. (ns, not significant), (*P < .05), (***P < .001), (****P < .0001). Student's t test.

Supplementary Figure 5



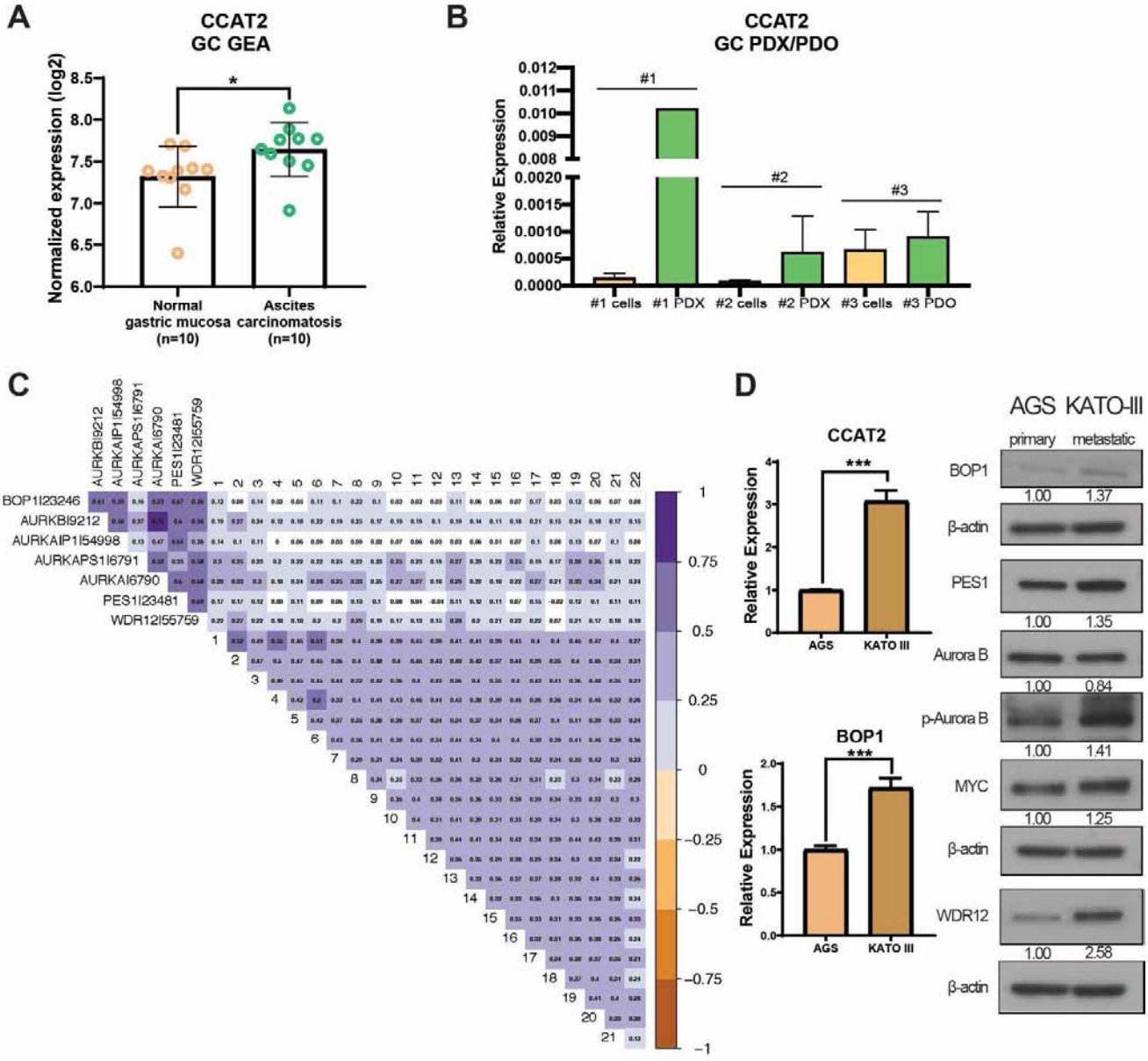
Supplementary Figure 5. Related to Figure 5. **A.** Up-regulation of *BOP1* increases the number of colonies in KM12SM^{BOP1} clones. **B.** Scratch assay showed that transient BOP1 knock-down inhibits the migration capacity of HCT116. Data are represented as mean values ± SD. (*P < .05), (**P < .01). Student's t test.

Supplementary Figure 6



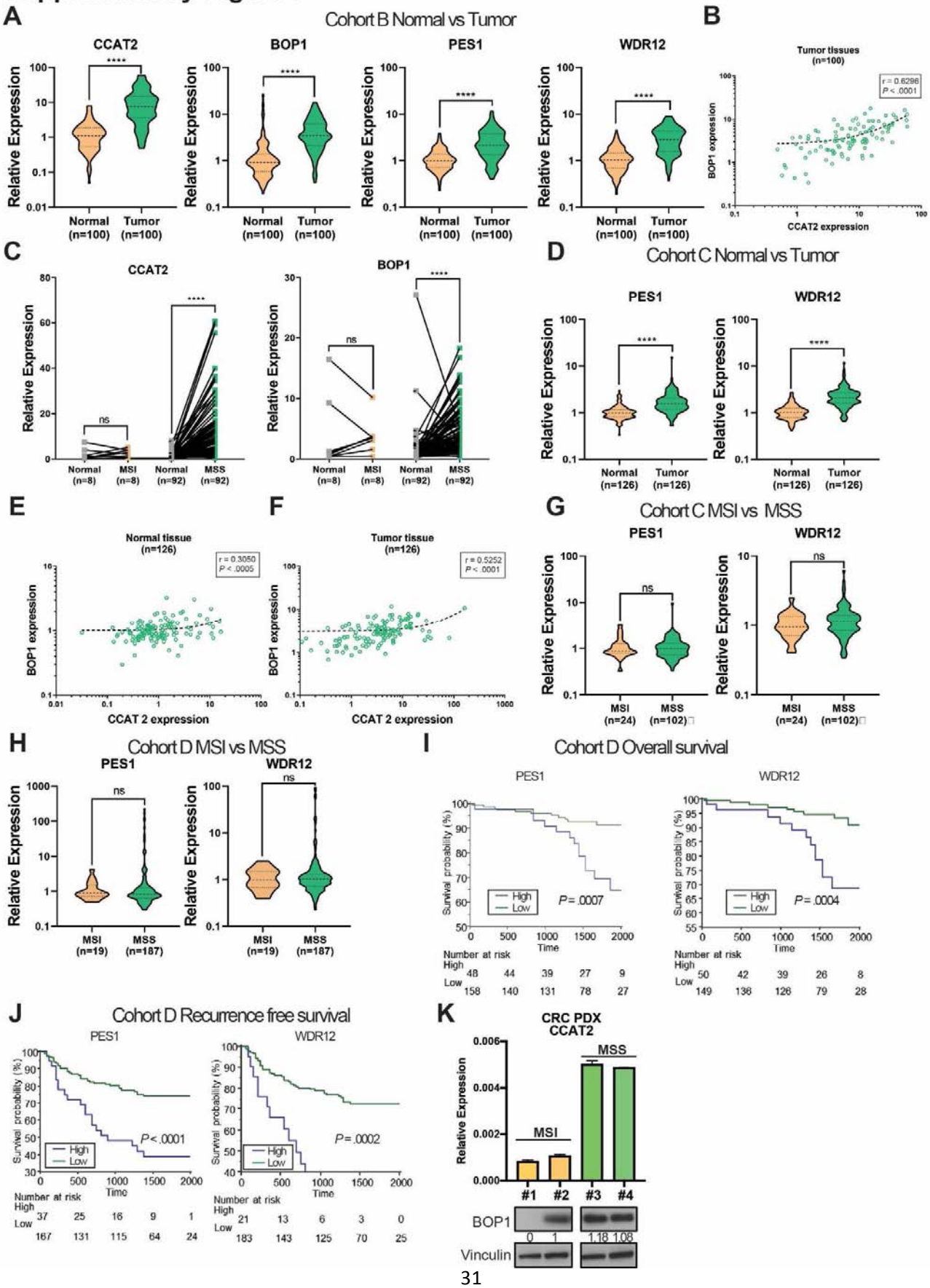
Supplementary Figure 6. Related to Figure 6. **A.** Schematic representation of Ribosome profiling analysis with or without EDTA. **B.** Ribosome profiling analysis without EDTA of $HCT116^{Empty}$ and $HCT116^{CCAT2}$ cells. **C.** Ribosome profiling analysis with EDTA of $HCT116^{Empty}$ and $HCT116^{CCAT2}$ cells. **D.** Correlation matrix between the Aurora family gene expression, PES1, BOP1, and WDR12 (PeBoW) complex components gene expression and index of CIN at whole chromosome level in the CRC cohort from TCGA database. **E.** The protein levels of key regulators of the chromosomal segregation mechanism (AURKA, CDC20, and BUB1B) were determined by Western blot in KM12SM cells after *CCAT2* knock-down. **F.** The protein levels of key regulators of the chromosomal segregation mechanism (AURKA, CDC20, and BUB1B) were determined by Western blot in HCT116 cells after BOP1 knock-down. **G.** CHIP-Seq data (UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly) showing MYC (Cluster Score (out of 1000): 595, Position: chr17:8113496-8114099, Band: 17p13.1, Genomic Size: 604) binding site in the 5' region of *AURKB*. **H.** RIP assay followed by qRT-PCR was performed to check the interaction of *CCAT2* with AURKB. HCT116 cell lysates were precipitated with AURKB antibodies and IgG was used as negative control, and were analyzed by Western blot. **I.** RNase I digestion followed by RIP assay, quantified by qRT-PCR to detect which *CCAT2* segment interacts with AURKB. (****$P < .0001$). Student's t test.

Supplementary Figure 7



Supplementary Figure 7. Related to Figure 6. **A.** Log2 normalized expression of *CCAT2* obtained from GEA comparing normal gastric mucosa with malignant ascites from GC patients. **B.** *CCAT2* expression in cells derived from GC patients and paired PDX/PDO **C.** Correlation matrix between the Aurora family gene expression, PES1, BOP1, and WDR12 (PeBoW) complex components gene expression and index of CIN at whole chromosome level in the GC cohort from TCGA database. **D.** The expression of *CCAT2*, PeBoW complex members, AURKB, pAURKB and MYC analyzed by qRT-PCR and Western blot in gastric cancer cell lines. AGS is a gastric cancer cell line derived from primary tumor and KATO III is derived from metastatic lymph nodes. Data are represented as mean values ± SD. (*P < .05), (***P < .001), (****P < .0001). Student's t test.

Supplementary Figure 8



Supplementary Figure 8. Related to Figure 7. **A.** RNA expression of *CCAT2*, *BOP1*, *PES1* and *WDR12* in 100 paired primary CRC tumor tissues and the adjacent non-tumor colon/rectum was analyzed by qRT-PCR (**Cohort B**). The expression of the RNA has been normalized to *β-actin*. Wilcoxon test for non-normal distributed paired samples was performed. **B** Spearman correlation between the expression of the lncRNA *CCAT2* and the mRNA *BOP1* in tumor tissues in **Cohort B**. **C.** *BOP1* and *CCAT2* expression in paired normal and MSI (MSI-H) or normal and MSS (MSS/MSI-L) CRC tissues. Wilcoxon test for non-normal distributed paired samples was performed. **D.** *PES1* and *WDR12* expression in adjacent normal and tumor tissues from **Cohort C**. Wilcoxon test for non-normal distributed paired samples was performed. **E.** Spearman correlation between the expression of the lncRNA *CCAT2* and the mRNA *BOP1* in normal tissues in **Cohort C**. **F.** Spearman correlation between the expression of the lncRNA *CCAT2* and the mRNA *BOP1* in tumor tissue in **Cohort C**. **G.** *PES1* and *WDR12* expression in MSI (MSI-H) and MSS (MSS/MSI-L) tumors of **Cohort C**. Mann-Whitney test for non-normal distributed un-paired samples was performed. **H.** *PES1* and *WDR12* expression in MSI (MSI-H) and MSS (MSS/MSI-L) tumors of **Cohort D**. Mann-Whitney test for non-normal distributed un-paired samples was performed. **I.** Kaplan–Meier overall survival curves and **J.** recurrence free survival curves of CRC patients from the **Cohort D**, expressing high (blue) or low (green) levels of *PES1* and *WDR12*. Time is expressed in months. **K.** The expression of the lncRNA *CCAT2* and the protein expression of BOP1 in two MSI PDX (#1 and #2) and two MSS PDX (#3 and #4). Data are represented as violin plots. (ns, not significant), (****$P < .0001$). Student's t test and Mann–Whitney–Wilcoxon test.

Supplementary Table 1. Clinical and pathological characteristics of the **Cohort A** (TCGA CRC cohort).

Variable		n
Gender	male	280
	female	257
Age	median (range)	66 (31-90)
Tumor location	right side colon	184
	left side colon	192
	rectum	49
UICC Stage	I	92
	II	201
	III	154
	IV	73
MSI/MSS status	MSS	150
	MSI-L	39
	MSI-H	29
	missing	319

Supplementary Table 2. Clinical and pathological characteristics of **Cohort B**.

Variable		n
Gender	male	66
	female	34
Age	median (range)	64 (33-85)
Tumor location	right side colon	22
	left side colon	29
	rectum	49
Differentiation grade	well	56
	moderate	36
	poor	3
	others	5
T stage	T1	17
	T2	12
	T3	63
	T4	8
Lymph node metastasis	negative	48
	positive	52
Distant metastasis	M0	91
	M1	9
UICC Stage	I-II	45
	III	46
	IV	9
Lymphatic invasion	negative	65
	positive	35
Venous invasion	negative	65
	positive	35
MSI/MSS status	MSS	82
	MSI-L	10
	MSI-H	8

Supplementary Table 3. Clinical and pathological characteristics of **Cohort C**.

Variable		n
Gender	male	76
	female	50
Age	median (range)	64 (30-84)
Tumor location	proximal	47
	distal	79
Differentiation grade	well (G1)	32
	moderate (G2)	68
	poor (G3)	19
	other (missing, G4, multiple grades)	7
T stage	T1	1
	T2	22
	T3	95
	T4	8
Lymph node metastasis	negative	69
	positive	57
Distant metastasis	M0	91
	M1	35
UICC Stage	I	17
	II	42
	III	32
	IV	35
MSI/MSS status	MSI	24
	MSS	102

Supplementary Table 4. Clinical and pathological characteristics of **Cohort D**.

Variable		n
Gender	male	120
	female	86
Age at operation	median (range)	70 (33-93)
Tumor location	right side colon	47
	left side colon	79
	rectum	80
Tumor size (mm)	median (range)	45 (10-140)
Histological type	intestinal	184
	diffuse	22
T stage	T2	9
	T3	141
	T4	56
Lymph node metastasis	negative	111
	positive	95
UICC Stage	II	111
	III	95
Lymphatic invasion	negative	92
	positive	114
Venous invasion	negative	21
	positive	185
MSI/MSS status	MSI	19
	MSS	187

Supplementary Table 5. Primers and siRNAs used in this study.

Gene	Sequence	Description
CCAT2 F1 (human)	5’ GGGCACTAGACTGGGAATTAG 3’	PCR Primer
CCAT2 R1 (human)	5’ AGGGAGCTGAGATAGGAAGAG 3’	PCR Primer
BOP1 F2 (human)	5’ GCCACAAGATGCACGTACCT 3’	PCR Primer
BOP1 R2 (human)	5’ TTCCTGGATGAAGCGTCCGTA 3’	PCR Primer
PES1 F (human)	5’ GGGCATTTATCCCCATGAACC 3’	PCR Primer
PES1 R (human)	5’ CACCTTGTATTCACGGAACTTGT 3’	PCR Primer
WDR12 F (human)	5’ CAGAGGAATGGATCTTGACTGGT 3’	PCR Primer
WDR12 R (human)	5’ CAGTGTAGGGCTTTCACTTTGT 3’	PCR Primer
BOP1-del6-264 F (human)	5’ GGCGGGTTCGCGGATGGGCTGGATCC 3’	PCR Primer
BOP1-del6-264 R (human)	5’ GGATCCAGCCCATCCGCGAACCCGCC 3’	PCR Primer
BOP1-delWD1 F (human)	5’ CCTGGTCTACAGGGTTCCCGTGGGGG 3’	PCR Primer
BOP1-delWD1 R (human)	5’ CCCCCACGGGAACCCTGTAGACCAGG 3’	PCR Primer
BOP1-delWD2 F (human)	5’ CGCTGTGTGAGGACTGTTGGCAGCACAGATCAG 3’	PCR Primer
BOP1-delWD2 R (human)	5’ CTGATCTGTGCTGCCAACAGTCCTCACACAGCG 3’	PCR Primer
BOP1-delWD3 F (human)	5’ CGGCTGCGCATCCGCAGCCACGGA 3’	PCR Primer
BOP1-delWD3 R (human)	5’ TCCGTGGCTGCGGATGCGCAGCCG 3’	PCR Primer
BOP1-delWD4 F (human)	5’ CAGAGTCCGTTCCGCCTGATGCCCAACTGC 3’	PCR Primer
BOP1-delWD4 R (human)	5’ GCAGTTGGGCATCAGGCGGAACGGACTCTG 3’	PCR Primer
BOP1-delWD5 F (human)	5’ CTCACCAAGAAGCTGATGATGCTGAGACACCACAAG 3’	PCR Primer
BOP1-delWD5 R (human)	5’ CTTGTGGTGTCTCAGCATCATCAGCTTCTTGGTGAG 3’	PCR Primer
BOP1-delWD6 F (human)	5’ ATACAGGATGCTGAGAAACCCCTTGCTGGTGC 3’	PCR Primer
BOP1-delWD6 R (human)	5’ GCACCAGCAAGGGGTTTCTCAGCATCCTGTAT 3’	PCR Primer
BOP1-del265-427 F (human)	5’ GGTGCACGCCATCAAGCTGGTTTCAGGCTCTG 3’	PCR Primer
BOP1-del265-427 R (human)	5’ CAGAGCCTGAAACCAGCTTGATGGCGTGCACC 3’	PCR Primer
CCAT2-EcoRI F (human)	5’ CCGgaattcccgaggtgatcaggtggact 3’	PCR Primer
CCAT2-SalI R (human)	5’ CGCgtcgacgtcttctgggctgatgttgc 3’	PCR Primer
CCAT2q-1 F (human)	5’ TAATACGACTCACTATAGACCCAGCAAGTTTCTCAGGA 3’	PCR Primer
CCAT2q-1 R (human)	5’ CATTTTTCAGCAATCAGGTCAA 3’	PCR Primer
CCAT2q-2 F (human)	5’ TAATACGACTCACTATAGAGACACCAAGAGGGAGGTATCA 3’	PCR Primer
CCAT2q-2 R (human)	5’ TGGCTCTTGACTTCCAGTCC 3’	PCR Primer
CCAT2q-3 F (human)	5’ TAATACGACTCACTATAGAACTTTCCCAGCCTCGTTCT	PCR Primer
CCAT2q-3 R (human)	5’ GGCTGTGGAAGTGGAATCAT	PCR Primer
CCAT2q-4 F (human)	5’ TAATACGACTCACTATAGCTCCATAGAGCCTGCAGAGG	PCR Primer
CCAT2q-4 R (human)	5’ ATTGGTCAGAGGTGGAGCTG 3’	PCR Primer

CCAT2q-5 F (human)	5’ TAATACGACTCACTATAGCAGCAGATGAAAGGCACTGA 3’	PCR Primer
CCAT2q-5 R (human)	5’ CTCCCTCCCCCACATAAAAT 3’	PCR Primer
CCAT2q-6 F (human)	5’ TAATACGACTCACTATAGACCCAGCAAGTTTCTCAGGA 3’	PCR Primer
CCAT2q-6 R (human)	5’ CACAGTTATTGCCTGGAGCA 3’	PCR Primer
CCAT2q-7 F (human)	5’ TAATACGACTCACTATAGGGCATGCCCTACGTAAGTTC 3’	PCR Primer
CCAT2q-7 R (human)	5’ TTTGGGGGTAGGTCAGGAAT 3’	PCR Primer
CCAT2q-8 F (human)	5’ TAATACGACTCACTATAGTGCATTGGTGAGCTGTGTTT 3’	PCR Primer
CCAT2q-8 R (human)	5’ ATGGTGCTGCTGGTAGCTTT 3’	PCR Primer
CCAT2q-9 F (human)	5’ TAATACGACTCACTATAGGCCATAATCATCCCTGAGGA 3’	PCR Primer
CCAT2q-9 R (human)	5’ CACCCCAGAGAGATGACACC 3’	PCR Primer
CCAT2q-10 F (human)	5’ TAATACGACTCACTATAGTTGTTGGGGTTTGATCCTTT 3’	PCR Primer
CCAT2q-10 R (human)	5’ CAAGCACTTGGAGCACACAT 3’	PCR Primer
Aurora B F (human)	5’ ACAGACGGCTCCATCTGGCCT 3’	PCR Primer
Aurora B R (human)	5’ GGCAGCTGTGGGCTGGACATT 3’	PCR Primer
c-MYC F (humane)	5’ GGCTCCTGGCAAAAGGTCA 3’	PCR Primer
c-MYC R (humane)	5’ CTGCGTAGTTGTGCTGATGT 3’	PCR Primer
β-actin F (human)	5’ CATGTACGTTGCTATCCAGGC 3’	PCR Primer
β-actin R (human)	5’ CTCCTTAATGTCACGCACGAT 3’	PCR Primer
U6 F (human)	5’ CTCGCTTCGGCAGCACA 3’	PCR Primer
U6 R (human)	5’ AACGCTTCACGAATTTGCGT 3’	PCR Primer
GAPDH F (human)	5’ CTGGGCTACACTGAGCACC 3’	PCR Primer
GAPDH R (human)	5’ AAGTGGTCGTTGAGGGCAATG 3’	PCR Primer
BOP1 F (mouse)	5’ CAGCTCTGATGAGGAGGACATTCGGAAC 3’	PCR Primer
BOP1 R (mouse)	5’ CAACCTGCTCATCAGTTAGCCG 3’	PCR Primer
mACTB F (mouse)	5’ CCTGTGCTGCTCACCGAGGC 3’	PCR Primer
mACTB R (mouse)	5’ GACCCCGTCTCTCCGGAGTCCATC 3’	PCR Primer
HPRT1 F (mouse)	5’ ACATTGTGGCCCTCTGTGTG 3’	PCR Primer
HPRT1 R (mouse)	5’ TTATGTCCCCCGTTGACTGA 3’	PCR Primer
siRNA	**Target Sequence/Producer ID**	
CCAT2 siRNA 1	5’ AGGTGTAGCCAGAGTTAAT 3’	siRNA target sequence
CCAT2 siRNA 2	5’ AGGAAGAGGTTAAGCAATT 3’	siRNA target sequence
BOP1 siRNA 1	s198254	Producer ID
BOP1 siRNA 2	s198255	Producer ID

Supplementary Table 6. Antibodies used in this study.

Antibodies	Product information
BOP1 (rabbit)	Abcam (ab86982)
BOP1 (mouse)	Santa Cruz (SC-390672)
BOP1 (IP)	Abcam (ab86652)
WDR12	Abcam (ab95070)
PES1	Abcam(ab88543)
PES1 (IP)	Santa Cruz (SC-166300)
MYC	Millipore (06-340)
pSer62 cMyc	Cell Signaling Technologies 13748
Aurora Kinase B (rabbit)	Cell Signaling Technologies 3094
Aurora Kinase B (mouse)	Santa Cruz (SC-65987)
P Thr232 Aurora Kinase B	Invitrogen PA5-38557
γ-tubulin	Cell Signaling Technologies 3873T
β-actin	Cell Signaling Technologies 3700T
GAPDH	Santa Cruz (SC51905)
Histone H1	Santa Cruz (SC34464)

Supplementary Table 7. Cell lines used in this study (colon cancer and gastric cancer cell lines).

Cell line	CCAT2 expression*	CIN	Modal chromosome no.	TP53	References
HCT116	**Low**	**-**	**45/46**	**WT**	25, 26
KM12C	**Low**	**-**	**45**	**H179R**	1, 27, 28
KM12SM	**High**	**+**	**75-82**	**H179R**	1, 27, 28
COLO320	**Very high**	**+**	**51**	**R248W**	29
HT29	**High**	**+**	**70 (3n)**	**R273H**	25, 30
AGS	**Low**	**-**	**49**	**WT**	31-33
KATO-III	**High**	**+**	**tetraploid range**	**Genomic deletion**	31, 34, 35

CIN: chromosomal instability. * - see Supplementary Figure 1A for data.

Supplementary Table 8. List of proteins identified by mass spectrometry to specifically bind to *CCAT2* after MS2 pull-down assay *

Accession	Gene	Subcellular location	# interaction hits by Mass Spec
B4DX30	ACSL5	Nucleus, Mitochondria	2
E9PRN9	**BOP1**	Nucleus, Nucleoli	1
B4DDM6	BUB3	Nucleoplasm	1
H3BTZ5	CNN2	Nucleoplasm, Actin filaments, Cytosol	1
B4DLW8	DDX5	Nucleoplasm	4
Q13347	EIF3I	Nucleoplasm, Cytosol	2
H7BY36	EWSR1	Nucleus, Nucleoli	2
Q8IW50-7	FAM219A	Nucleoplasm, Golgi apparatus, Vesicles	1
B4DY13	GTPBP4	Nucleoli	2
O94992	HEXIM1	Nucleoplasm, Vesicles	1
O60814	HIST1H2BK	Nucleoplasm, Cytosol	10
Q5T7C4	HMGB1	Nucleus	1
H7C1J8	HNRNPA3	Nucleoplasm	1
P25205	MCM3	Nucleoplasm	6
Q15233-2	NONO	Nucleoplasm	2
H0Y653	NOP56	Nucleoli fibrillar center	1
G3V1R4	NSUN2	Nucleus	3
K7ELW0	PARK7	Nucleus, Cytosol	2
E7ETC2	PPP3CA	Nucleoplasm, Cytosol	2
H0YMZ1	PSMA4	Nucleus, Vesicles, Cytosol	2
P28074	PSMB5	Nucleus, Centrosome	2
Q06124-2	PTPN11	Nucleus, Nucleoli, Actin filaments, Cytosol	1
K7EIJ4	RANBP3	Nucleoplasm	1
B4DWG0	RCBTB2	Nucleus, Golgi apparatus	1
I3L4R8	RPA1	Nucleoplasm	1
P40429	RPL13A	Nucleoli, Cytosol	1
F8VUA6	RPL18	Nucleoli, Endoplasmic reticulum, Cytosol	2
E9PKZ0	RPL8	Nucleoli, Endoplasmic reticulum, Cytosol	2
G3XAA9	RPS6KA4	Nucleus, Cytosol	1
K7EPT6	TAF15	Nucleoplasm	2
G3V448	TMX1	Nucleoli, Endoplasmic reticulum	1
Q13263	TRIM28	Nucleoplasm	10
J3KT19	USP10	Nucleoplasm, Cytosol	1
B1AHC7	XRCC6	Nucleoplasm	7
P27348	YWHAQ	Nucleoplasm, Cytosol	4
C9JN71	ZNF878	Nucleus, Nucleoli	3

* **-** MS2 Empty Pull down in all the instances gave 0 hits.

Supplementary Table 9. CINdex analysis at the chromosome level of the CRC TCGA cohort and correlation between the genes of the Aurora family, PeBoW complex and CIN at chromosomal level.

Chr. number	Gene	Correlation Test - T statistic	P-value	Correlation coefficient	Confidence interval 1	Confidence interval 2
20	AURKA	9.7	3.20E-19	0.501	0.41	0.58
20	WDR12	4.6	6.50E-06	0.266	0.15	0.37
13	AURKA	4.0	7.97E-05	0.234	0.12	0.34
20	AURKAPS1	4.0	8.80E-05	0.232	0.12	0.34
2	WDR12	3.6	0.0003271	0.213	0.10	0.32
20	BOP1	3.5	0.0005195	0.206	0.09	0.32
20	AURKC	-3.4	0.0007570	-0.200	-0.31	-0.08
7	AURKA	3.3	0.0009951	0.196	0.08	0.31
8	BOP1	3.1	0.0019504	0.184	0.07	0.30
7	AURKC	-3.1	0.0019925	-0.184	-0.29	-0.07
19	AURKAIP1	-2.9	0.0045704	-0.169	-0.28	-0.05
2	AURKA	2.8	0.0052080	0.167	0.05	0.28
17	AURKA	2.8	0.0060698	0.164	0.05	0.28
13	AURKC	-2.7	0.0067611	-0.162	-0.27	-0.05
20	AURKAIP1	-2.7	0.0069726	-0.161	-0.27	-0.04
10	WDR12	2.7	0.0070027	0.161	0.04	0.27
6	AURKAPS1	2.7	0.0078359	0.159	0.04	0.27
15	AURKAIP1	-2.7	0.0081522	-0.158	-0.27	-0.04
13	WDR12	2.5	0.0127088	0.149	0.03	0.26
8	AURKA	2.5	0.0129917	0.148	0.03	0.26
12	AURKA	2.5	0.0133298	0.148	0.03	0.26
1	AURKAIP1	-2.5	0.0137918	-0.147	-0.26	-0.03
5	AURKAPS1	2.5	0.0138759	0.147	0.03	0.26
8	AURKAIP1	-2.4	0.0160611	-0.144	-0.26	-0.03
5	WDR12	2.4	0.0164290	0.143	0.03	0.26
12	AURKC	-2.4	0.0174374	-0.142	-0.25	-0.03
17	AURKAPS1	2.4	0.0187027	0.140	0.02	0.25
22	AURKAPS1	2.3	0.0202989	0.139	0.02	0.25
12	WDR12	2.3	0.0223718	0.136	0.02	0.25
15	AURKC	-2.3	0.0241227	-0.135	-0.25	-0.02
8	AURKAPS1	2.3	0.0244076	0.134	0.02	0.25
6	AURKA	2.2	0.0252607	0.134	0.02	0.25
19	PES1	-2.2	0.0258461	-0.133	-0.25	-0.02
15	AURKB	-2.2	0.0261202	-0.133	-0.25	-0.02
4	AURKA	2.2	0.0261782	0.133	0.02	0.25
1	AURKAPS1	2.2	0.0277008	0.132	0.01	0.25
18	WDR12	2.2	0.0278869	0.131	0.01	0.24
21	WDR12	2.2	0.0310365	0.129	0.01	0.24
22	AURKA	2.1	0.0337156	0.127	0.01	0.24
4	WDR12	2.1	0.0355162	0.126	0.01	0.24
14	WDR12	2.1	0.0378901	0.124	0.01	0.24
8	AURKC	-2.1	0.0379421	-0.124	-0.24	-0.01
16	AURKAPS1	2.0	0.0417412	0.122	0.00	0.24
4	AURKAPS1	2.0	0.0429808	0.121	0.00	0.23

AURKA, aurora kinase A; AURKAIP1, aurora kinase A interacting protein 1; AURKAPS1, aurora kinase A pseudogene 1; AURKB, aurora kinase B, BOP1, BOP1 ribosomal biogenesis factor; PES1 pescadillo ribosomal biogenesis factor 1; WDR12, WD repeat domain 12.

Supplementary Table 10. CINdex analysis at the chromosome level of the GC TCGA cohort and correlation between the genes of the Aurora family, PeBoW complex and CIN at chromosomal level.

Chr. number	Gene	Correlation Test - T statistic	P value	Correlation coefficient	Confidence Interval 1	Confidence Interval 2
20	AURKA	7.39	8.05E-13	0.34	0.25	0.43
8	AURKA	7.14	4.33E-12	0.33	0.24	0.42
19	AURKA	7.01	9.73E-12	0.33	0.24	0.41
2	AURKA	6.98	1.18E-11	0.33	0.24	0.41
1	AURKAPS1	6.41	3.91E-10	0.30	0.21	0.39
3	AURKA	6.34	6.23E-10	0.30	0.21	0.38
8	WDR12	6.09	2.64E-09	0.29	0.20	0.37
13	AURKA	6.08	2.83E-09	0.29	0.20	0.37
1	AURKA	5.95	5.70E-09	0.28	0.19	0.37
6	AURKA	5.91	7.11E-09	0.28	0.19	0.37
19	AURKAPS1	5.88	8.39E-09	0.28	0.19	0.37
13	WDR12	5.81	1.24E-08	0.28	0.18	0.36
17	AURKA	5.71	2.20E-08	0.27	0.18	0.36
13	AURKAPS1	5.68	2.54E-08	0.27	0.18	0.36
10	AURKA	5.67	2.74E-08	0.27	0.18	0.36
2	WDR12	5.65	3.01E-08	0.27	0.18	0.36
2	AURKB	5.58	4.27E-08	0.27	0.17	0.35
11	AURKA	5.58	4.31E-08	0.27	0.17	0.35
20	AURKAPS1	5.45	8.49E-08	0.26	0.17	0.35
16	AURKAPS1	5.32	1.74E-07	0.25	0.16	0.34
7	AURKA	5.27	2.24E-07	0.25	0.16	0.34
10	AURKAPS1	5.26	2.33E-07	0.25	0.16	0.34
2	AURKAPS1	5.25	2.39E-07	0.25	0.16	0.34
7	AURKAPS1	5.22	2.88E-07	0.25	0.16	0.34
8	AURKB	5.20	3.21E-07	0.25	0.16	0.34
9	AURKA	5.14	4.32E-07	0.25	0.15	0.33
22	AURKA	5.10	5.16E-07	0.24	0.15	0.33
3	AURKB	5.08	5.83E-07	0.24	0.15	0.33
19	AURKB	5.00	8.52E-07	0.24	0.15	0.33
5	AURKA	4.99	8.99E-07	0.24	0.15	0.33
14	AURKAPS1	4.89	1.45E-06	0.23	0.14	0.32
9	AURKAPS1	4.85	1.74E-06	0.23	0.14	0.32
3	AURKAPS1	4.83	1.97E-06	0.23	0.14	0.32
11	AURKAPS1	4.76	2.69E-06	0.23	0.14	0.32
3	WDR12	4.67	4.05E-06	0.22	0.13	0.31
5	AURKAPS1	4.61	5.49E-06	0.22	0.13	0.31
1	WDR12	4.60	5.67E-06	0.22	0.13	0.31
16	AURKA	4.60	5.71E-06	0.22	0.13	0.31
6	AURKB	4.59	6.00E-06	0.22	0.13	0.31
15	AURKAPS1	4.58	6.22E-06	0.22	0.13	0.31
8	AURKAPS1	4.55	7.06E-06	0.22	0.13	0.31
14	AURKA	4.55	7.08E-06	0.22	0.13	0.31
8	BOP1	4.52	8.02E-06	0.22	0.12	0.31
16	WDR12	4.50	8.93E-06	0.22	0.12	0.31
22	AURKAPS1	4.50	8.99E-06	0.22	0.12	0.31
6	AURKAPS1	4.48	9.86E-06	0.22	0.12	0.31
17	WDR12	4.47	1.00E-05	0.22	0.12	0.31
21	AURKA	4.44	1.18E-05	0.21	0.12	0.30
18	AURKA	4.34	1.81E-05	0.21	0.12	0.30
15	WDR12	4.28	2.33E-05	0.21	0.11	0.30
17	AURKB	4.27	2.40E-05	0.21	0.11	0.30
19	WDR12	4.25	2.64E-05	0.21	0.11	0.30
4	AURKAPS1	4.19	3.42E-05	0.20	0.11	0.29
6	WDR12	4.13	4.45E-05	0.20	0.11	0.29
14	WDR12	4.12	4.58E-05	0.20	0.10	0.29
7	WDR12	4.06	0.0001	0.20	0.10	0.29

13	AURKB	3.98	0.0001	0.19	0.10	0.28
7	AURKB	3.95	0.0001	0.19	0.10	0.28
11	AURKB	3.89	0.0001	0.19	0.09	0.28
1	AURKB	3.88	0.0001	0.19	0.09	0.28
12	AURKAPS1	3.87	0.0001	0.19	0.09	0.28
9	WDR12	3.87	0.0001	0.19	0.09	0.28
17	AURKAPS1	3.85	0.0001	0.19	0.09	0.28
10	AURKB	3.84	0.0001	0.19	0.09	0.28
21	AURKAPS1	3.79	0.0002	0.18	0.09	0.28
4	AURKA	3.78	0.0002	0.18	0.09	0.28
22	WDR12	3.77	0.0002	0.18	0.09	0.27
5	WDR12	3.73	0.0002	0.18	0.09	0.27
21	WDR12	3.69	0.0002	0.18	0.08	0.27
15	AURKA	3.68	0.0003	0.18	0.08	0.27
16	AURKB	3.64	0.0003	0.18	0.08	0.27
20	AURKB	3.64	0.0003	0.18	0.08	0.27
12	AURKA	3.64	0.0003	0.18	0.08	0.27
4	WDR12	3.63	0.0003	0.18	0.08	0.27
5	AURKB	3.62	0.0003	0.18	0.08	0.27
18	AURKAPS1	3.60	0.0004	0.17	0.08	0.27
20	WDR12	3.58	0.0004	0.17	0.08	0.27
21	AURKB	3.58	0.0004	0.17	0.08	0.27
9	AURKB	3.56	0.0004	0.17	0.08	0.27
17	BOP1	3.50	0.0005	0.17	0.07	0.26
2	PES1	3.49	0.0005	0.17	0.07	0.26
10	WDR12	3.43	0.0007	0.17	0.07	0.26
1	PES1	3.40	0.0008	0.17	0.07	0.26
22	AURKB	3.16	0.0017	0.15	0.06	0.25
12	WDR12	3.15	0.0018	0.15	0.06	0.25
18	AURKB	3.03	0.0026	0.15	0.05	0.24
17	PES1	2.99	0.0030	0.15	0.05	0.24
14	AURKB	2.86	0.0044	0.14	0.04	0.23
1	AURKAIP1	2.82	0.0050	0.14	0.04	0.23
3	BOP1	2.77	0.0058	0.14	0.04	0.23
11	WDR12	2.60	0.0097	0.13	0.03	0.22
8	PES1	2.59	0.0100	0.13	0.03	0.22
19	AURKAIP1	2.58	0.0101	0.13	0.03	0.22
1	BOP1	2.55	0.0112	0.12	0.03	0.22
4	AURKB	2.53	0.0119	0.12	0.03	0.22
19	BOP1	2.47	0.0138	0.12	0.02	0.22
14	PES1	2.47	0.0139	0.12	0.02	0.22
19	PES1	2.45	0.0149	0.12	0.02	0.21
3	PES1	2.44	0.0152	0.12	0.02	0.21
5	PES1	2.29	0.0225	0.11	0.02	0.21
22	PES1	2.27	0.0235	0.11	0.02	0.21
13	PES1	2.27	0.0236	0.11	0.02	0.21
15	PES1	2.25	0.0249	0.11	0.01	0.20
13	BOP1	2.25	0.0250	0.11	0.01	0.20
3	AURKAIP1	2.24	0.0254	0.11	0.01	0.20
6	BOP1	2.17	0.0307	0.11	0.01	0.20
15	AURKB	2.16	0.0314	0.11	0.01	0.20
21	PES1	2.15	0.0321	0.11	0.01	0.20
17	AURKAIP1	2.13	0.0336	0.10	0.01	0.20
9	PES1	2.13	0.0337	0.10	0.01	0.20
9	BOP1	2.11	0.0354	0.10	0.01	0.20
20	PES1	2.10	0.0364	0.10	0.01	0.20
12	AURKB	2.08	0.0379	0.10	0.01	0.20
2	AURKAIP1	2.07	0.0389	0.10	0.01	0.20
21	AURKAIP1	2.05	0.0409	0.10	0.00	0.20
7	BOP1	1.99	0.0473	0.10	0.00	0.19

Supplementary Table 11. Univariate analysis for overall survival and recurrence free survival using log-rank test for patients from **Cohort D**.

Overall Survival

	Univariate analysis		
variable	HR	95%CI	P-value
Age ≧ 70	3.1	1.39-7.59	**0.005**
Male	0.83	0.39-1.84	0.65
Tumor location rectum yes/no	1.83	0.85-4.03	0.12
Tumor size > 45 mm	1.04	0.48-2.28	0.91
Poorly differentiated histology	1.53	0.45-4.01	0.45
T Stage greater than T4	1.07	0.43-2.40	0.88
Lymph node metastasis positive	2.18	1.01-5.12	**0.049**
Lymphatic invasion positive	1.53	0.70-3.58	0.3
Venous invasion positive	0.67	0.26-2.30	0.49
UICC stage 3	2.18	0.99-5.12	0.051
MSI patients yes/no	1.28	0.30-3.69	0.69
CCAT2 high expression	6.1	2.79-13.31	**<0.0001**
BOP1 high expression	3.01	1.38-6.55	**0.006**
PES1 high expression	3.52	1.62-7.74	**0.002**
WDR12 high expression	3.15	1.44-6.86	**0.004**

Recurrence Free Survival

	Univariate analysis		
variable	HR	95%CI	P-value
Age ≧ 70	1.14	0.68-1.89	0.62
Male	0.95	0.57-1.61	0.85
Tumor location rectum yes/no	2.63	1.58-4.47	**0.0002**
Tumor size > 45 mm	1.08	0.65-1.79	0.77
Poorly differentiated histology	1.05	0.44-2.16	0.9
T Stage greater than T4	1.54	0.89-2.59	0.12
Lymph node metastasis positive	1.88	1.13-3.18	**0.02**
Lymphatic invasion positive	1.38	0.83-2.36	0.22
Venous invasion positive	1.05	0.49-2.72	0.92
UICC stage 3	1.88	1.13-3.18	**0.02**
MSI patients yes/no	3.41	1.06-20.79	**0.04**
CCAT2 high expression	4.78	2.74-8.09	**<0.0001**
BOP1 high expression	2.77	1.65-4.60	**0.0002**
PES1 high expression	2.94	1.70-4.96	**0.0002**
WDR12 high expression	4.07	2.18-7.17	**<0.0001**

HR, Hazard Ratio; CI, Confidence Interval. Bold indicates a statistically significant result.

Supplementary Table 12. Multivariate analysis for overall survival and recurrence free survival in patients from **Cohort D** using Cox proportional hazard model.

***CCAT2* and other clinical factors**
Overall Survival

	Univariate analysis			Multivariate analysis		
variable	HR	95%CI	P-value	HR	95%CI	P-value
Age ≧ 70	3.1	1.39-7.59	**0.005**	2.41	1.05-6.01	**0.04**
Lymph node metastasis positive	2.18	1.01-5.12	**0.049**	2.42	1.10-5.69	**0.03**
CCAT2 high expression	6.1	2.79-13.31	**<0.0001**	5.51	2.48-12.27	**<0.0001**
Recurrence Free Survival						
	Univariate analysis			Multivariate analysis		
variable	HR	95%CI	P-value	HR	95%CI	P-value
Tumor location rectum yes/no	2.63	1.58-4.47	**0.0002**	2.37	1.41-4.08	**0.001**
Lymph node metastasis positive	1.88	1.13-3.18	**0.02**	2.03	1.21-3.45	**0.007**
UICC stage 3	1.88	1.13-3.18	**0.02**			
MSI patients yes/no	3.41	1.06-20.79	**0.04**	0.5	0.08-1.66	0.29
CCAT2 high expression	4.78	2.74-8.09	**<0.0001**	4.85	2.76-8.28	**<0.0001**

HR, Hazard Ratio; CI, Confidence Interval. Bold indicates a statistically significant result.

***BOP1* and other clinical factors**
Overall Survival

	Univariate analysis			Multivariate analysis		
variable	HR	95%CI	P-value	HR	95%CI	P-value
Age ≧ 70	3.1	1.39-7.59	**0.005**	3.06	1.36-7.49	**0.006**
Lymph node metastasis positive	2.18	1.01-5.12	**0.049**	2.6	1.18-6.12	**0.02**
BOP1 high expression	3.01	1.38-6.55	**0.006**	3.12	1.42-6.84	**0.005**
Recurrence Free Survival						
	Univariate analysis			Multivariate analysis		
variable	HR	95%CI	P-value	HR	95%CI	P-value
Tumor location rectum yes/no	2.63	1.58-4.47	**0.0002**	2.38	1.42-4.09	**0.001**
Lymph node metastasis positive	1.88	1.13-3.18	**0.02**	1.87	1.12-3.18	**0.02**
UICC stage 3	1.88	1.13-3.18	**0.02**			
MSI patients yes/no	3.41	1.06-20.79	**0.04**	0.52	0.08-1.73	0.32
BOP1 high expression	2.77	1.65-4.60	**0.0002**	2.67	1.59-4.45	**0.0003**

HR, Hazard Ratio; CI, Confidence Interval. Bold indicates a statistically significant result.

References

1. Morikawa K, Walker SM, Nakajima M, et al. Influence of organ environment on the growth, selection, and metastasis of human colon carcinoma cells in nude mice. Cancer Res 1988;48:6863-71.
2. Shah MY, Ferracin M, Pileczki V, et al. Cancer-associated rs6983267 SNP and its accompanying long noncoding RNA CCAT2 induce myeloid malignancies via unique SNP-specific RNA mutations. Genome Res 2018;28:432-447.
3. Lee SM, Kim N, Son HJ, et al. The Effect of Sex on the Azoxymethane/Dextran Sulfate Sodium-treated Mice Model of Colon Cancer. J Cancer Prev 2016;21:271-278.
4. Parang B, Barrett CW, Williams CS. AOM/DSS Model of Colitis-Associated Cancer. Methods Mol Biol 2016;1422:297-307.
5. Suzuki R, Kohno H, Sugie S, et al. Strain differences in the susceptibility to azoxymethane and dextran sodium sulfate-induced colon carcinogenesis in mice. Carcinogenesis 2006;27:162-9.
6. Gerling M, Glauben R, Habermann JK, et al. Characterization of chromosomal instability in murine colitis-associated colorectal cancer. PLoS One 2011;6:e22114.
7. Sato T, Vries RG, Snippert HJ, et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature 2009;459:262-5.
8. Sato T, Stange DE, Ferrante M, et al. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. Gastroenterology 2011;141:1762-72.
9. Miyoshi H, Stappenbeck TS. In vitro expansion and genetic modification of gastrointestinal stem cells in spheroid culture. Nat Protoc 2013;8:2471-82.
10. Song S, Wang Z, Li Y, et al. PPARdelta Interacts with the Hippo Coactivator YAP1 to Promote SOX9 Expression and Gastric Cancer Progression. Mol Cancer Res 2020;18:390-402.
11. Katsiampoura A, Raghav K, Jiang ZQ, et al. Modeling of Patient-Derived Xenografts in Colorectal Cancer. Mol Cancer Ther 2017;16:1435-1442.
12. Liu J, Lichtenberg T, Hoadley KA, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell 2018;173:400-416 e11.
13. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330-7.
14. Redis RS, Vela LE, Lu W, et al. Allele-Specific Reprogramming of Cancer Metabolism by the Long Non-coding RNA CCAT2. Mol Cell 2016;61:520-534.
15. Song L, Bhuvaneshwar K, Wang Y, et al. CINdex: A Bioconductor Package for Analysis of Chromosome Instability in DNA Copy Number Data. Cancer Inform 2017;16:1176935117746637.
16. Ling H, Spizzo R, Atlasi Y, et al. CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. Genome Res 2013;23:1446-61.
17. Gagnon KT, Li L, Janowski BA, et al. Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. Nat Protoc 2014;9:2045-60.
18. Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics 2010;11:129.
19. Das R, Laederach A, Pearlman SM, et al. SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. RNA 2005;11:344-54.
20. Deigan KE, Li TW, Mathews DH, et al. Accurate SHAPE-directed RNA structure determination. Proc Natl Acad Sci U S A 2009;106:97-102.
21. Zhou P. Determining protein half-lives. Methods Mol Biol 2004;284:67-77.
22. Pereboom TC, Bondt A, Pallaki P, et al. Translation of branched-chain aminotransferase-1 transcripts is impaired in cells haploinsufficient for ribosomal protein genes. Experimental Hematology 2014;42:394-403.
23. Reineke LC, Cheema SA, Dubrulle J, et al. Chronic starvation induces noncanonical pro-death stress granules. J Cell Sci 2018;131.
24. Dassi E. Post-Transcriptional Gene Regulation: New York: Springer Science and Business, 2016.

25. Ahmed D, Eide PW, Eilertsen IA, et al. Epigenetic and genetic features of 24 colon cancer cell lines. Oncogenesis 2013;2:e71.
26. Knutsen T, Padilla-Nash HM, Wangsa D, et al. Definitive molecular cytogenetic characterization of 15 colorectal cancer cell lines. Genes Chromosomes Cancer 2010;49:204-23.
27. Camps J, Morales C, Prat E, et al. Genetic evolution in colon cancer KM12 cells and metastatic derivates. Int J Cancer 2004;110:869-74.
28. Leroy B, Girard L, Hollestelle A, et al. Analysis of TP53 mutation status in human cancer cell lines: a reassessment. Hum Mutat 2014;35:756-65.
29. Quinn LA, Moore GE, Morgan RT, et al. Cell lines from human colon carcinoma with unusual cell products, double minutes, and homogeneously staining regions. Cancer Res 1979;39:4914-24.
30. Chen TR, Drabkowski D, Hay RJ, et al. WiDr is a derivative of another colon adenocarcinoma cell line, HT-29. Cancer Genet Cytogenet 1987;27:125-34.
31. Yao Y, Tao H, Kim JJ, et al. Alterations of DNA mismatch repair proteins and microsatellite instability levels in gastric cancer cell lines. Lab Invest 2004;84:915-22.
32. Barranco SC, Townsend CM, Jr., Casartelli C, et al. Establishment and characterization of an in vitro model system for human adenocarcinoma of the stomach. Cancer Res 1983;43:1703-9.
33. Cheng LL, Itahana Y, Lei ZD, et al. TP53 genomic status regulates sensitivity of gastric cancer cells to the histone methylation inhibitor 3-deazaneplanocin A (DZNep). Clin Cancer Res 2012;18:4201-12.
34. Sekiguchi M, Sakakibara K, Fujii G. Establishment of cultured cell lines derived from a human gastric carcinoma. Jpn J Exp Med 1978;48:61-8.
35. Yokozaki H. Molecular characteristics of eight gastric cancer cell lines established in Japan. Pathol Int 2000;50:767-77.