Supplemental Material

MIR99AHG promotes invadopodia formation in colorectal cancer metastasis by tuning PTBP1-mediated splicing modulation of SMARCA1

Danxiu Li, Xin Wang, Hui Miao, Hao Liu, Maogui Pang, Hao Guo, Minghui Ge, Xiaoni Ye, Sarah E. Glass, Stephan Emmrich, Jing Wang, Qi Liu, Taewan Kim, Jan-Henning Klusmann, Cunxi Li, Gang Ji, Yongzhan Nie, Kaichun Wu, Daiming Fan, Xu Song, Xin Wang, Ling Li, Yuanyuan Lu, Xiaodi Zhao

Supplementary Materials and methods

Human TMA

The study was approved by the Medical Ethics Committee of Xijing Hospital with written informed consent obtained from all patients. Human TMAs containing primary colorectal cancer (CRC) tissues with their adjacent normal tissues and matched lymph nodes or distant metastases were obtained from the CBSKL biobank and Outdo Biotech (Shanghai, China). The clinicopathological information of CRC patients is included in Supplementary Table 2. Serial sections of the TMA were used for *in situ* hybridization (ISH), RNAscope ISH and multispectral immunohistochemistry (IHC).

Animal studies

All the animal experiments were conducted under protocols approved by the Institutional Animal Care and Use Committee at the Fourth Military Medical University. Six- to eight-week-old female athymic BALB/c nude mice were purchased from Beijing Vital River Laboratory Animal Technology. The mice were maintained under a 12-h light/dark cycle and on a standard chow diet at the specific-pathogen-free facility at the Experimental Animal Center of the Fourth Military Medical University. For the *in vivo* tumorigenesis assays, suspensions of stable luciferase-expressing CRC cells after the indicated treatment were subcutaneously injected into the posterior flanks of mice (5×10^6 tumor cells/150 μl PBS per spot). Tumor volume was calculated using the following formula: tumor maximum diameter (L) × the right-angle diameter to that axis $(W)^2/2$. Three to four weeks after injection, the mice were sacrificed, and the tumors were paraffin-embedded for further examination. For *in vivo* metastasis assays, stable luciferase-expressing CRC cells after the indicated treatment were injected into nude mice via the tail vein (1.5×10^6 tumor cells/150 μl PBS per mouse). The mice were subjected to bioluminescence detection using an IVIS 100 Imaging System (PerkinElmer) at weekly intervals. Four to five weeks after injection, the mice were sacrificed, and the lungs were paraffin-embedded for further examination.

RNA-seq, small RNA-seq and data analysis

Total RNA from cells was isolated by TRIzol Reagent (Invitrogen) and then purified using the RNeasy Mini Kit (Qiagen). The concentration and integrity of total RNA were estimated using a Qubit 2.0 Fluorometer (Invitrogen) and an Agilent 2100 Bioanalyzer (Agilent Technologies), respectively. RNA library construction and sequencing were performed as previously described [1]. Briefly, approximately 3 μg of total RNA from each sample was used for mRNA or small RNA library construction. For RNA-seq, the NEBNext DNA Library Prep Master Mix Set for Illumina (New England BioLabs) was used to prepare a next-generation sequencing expression library following the manufacturer's protocol. Accurate quantification for sequencing applications was determined using the qPCR-based KAPA Biosystems Library Quantification Kit (Kapa Biosystems). Paired-end sequencing (75 bp) was performed on a NextSeq 500 sequencer (Illumina). RNA-Seq reads were aligned to the human genome hg19 using TopHat2, and the number of reads mapped to each gene was calculated by HTseq. For small RNA-seq, the NEBNext Small RNA Library Prep Set for Illumina (New England BioLabs) was used to prepare a small RNA library following the manufacturer's protocol. Accurate quantification for sequencing applications was performed using the qPCR-based KAPA Biosystems Library Quantification Kit. Single-end sequencing (50 bp) was performed on a NextSeq 500 Sequencer. Reads were aligned to the human genome hg19 using Bowtie. Mapped reads were annotated, and miRNA expression was quantified using ncPRO-seq (version v1.5.1) based on miRbase v19. Differentially expressed genes (DEGs), lncRNAs and miRNAs were detected by edgeR based on a negative binomial distribution. The online Database For Annotation, Visualization, Integrated Discovery (DAVID) was used to perform gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis to analyze the biological characteristics and functional annotations of the candidate DEGs [2]. Gene set enrichment analysis (GSEA) was used to determine whether DEGs were related to one phenotype or signaling pathway using GSEA software [3]. ASprofile (https://ccb.jhu.edu/) was

employed to classify alternative splicing (AS) events, and differential AS events were identified by rMATS [4].

Constructs, oligonucleotides and cell transfection

Expression vectors encoding the full-length, truncated and antisense MIR99AHG were constructed by subcloning the corresponding sequences into a pZW1-snoVector. The pZW1-snoVector was kindly provided by Dr. Ling-Ling Chen (Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences). Expression vectors encoding PTBP1 (full-length or truncation), SMARCA1-L and SMARCA1-S were constructed by subcloning the corresponding ORF regions into a pcDNA3.1 vector. The SMARCA1 minigene was constructed by subcloning the sequence region from exon 12 to exon 14 of SMARCA1 into a pcDNA3.1 vector. To establish MIR99AHG stably transfected cells, the MIR99AHG sequence was subcloned into a pLVX-IRES-Puro vector. Cells stably expressing MIR99AHG were established by lentivirus infection and selected with puromycin. Antisense oligonucleotides (ASOs) targeting MIR99AHG were purchased from RiboBio. siRNAs targeting TKS5, Cortactin, PTBP1 and U2AF2 were purchased from GenePharma. Isoform-specific siRNAs targeting SMARCA1-L and SMARCA1-S were purchased from Tsingke. The sequence information of the siRNAs used in this study is listed in Supplementary Table 3. Plasmids and oligonucleotides were transfected into the target cells using JetPRIME (Polyplus Transfection) reagent following the manufacturer's instructions.

Generation of CRISPR/Cas9 knockout cell lines

Two pairs of guide RNAs (gRNAs) targeting the start and end sequences of MIR99AHG exon 8 and two gRNAs targeting the sequences of *SMARCA1* exon 1 were designed using the CRISPRdirect web server (https://crispr.dbcls.jp). Target sequences of gRNAs are listed in Supplementary Table 4. Human codon-optimized Cas9 (hCas9) and GFP-targeting gRNA-expressing plasmid (gRNA_GFP-T1) were purchased from Addgene. The GFP-targeting sequence

in gRNA_GFP-T1 was replaced using QuickChange site-directed mutagenesis. To construct the knockout cell lines, the modified gRNA_GFP-T1, hCas9 and pEGFP-C1 vectors were cotransfected into KM12SM or DLD-1 cells. GFP-positive cells were identified by flow cytometry and sorted into single clones in a 96-well plate. Single clones were screened by the T7 endonuclease I-cutting assay. The knockout clones were confirmed by DNA sequencing.

Transwell migration and invasion assay

For the migration assays, 24-well Transwell plates with 8-μm-diameter micropore membranes without Matrigel (Millipore) were used. Invasion assays were conducted using 8-μm pore, 24-well Transwell plates with Matrigel (Corning). A total of 2×10^5 cells resuspended in serum-free medium were seeded into the upper compartment of the migration or invasion chambers, and the bottom chamber was filled with culture medium with 20% FBS as an attractant. After incubation in 5% CO_2 at 37°C for 16-48 h, cells that had migrated or invaded through the membrane were fixed with 75% ethanol, stained with crystal violet dye and then counted under a microscope (Olympus) to determine cell numbers. All assays were conducted in triplicate, and three fields of an insert were randomly chosen for quantification.

Wound healing assay

CRC cells with the indicated treatments were suspended at a concentration of $3\text{-}7\times10^5$ cells/ml, and 70 μl of the cell suspension was added to each well of a Culture-Insert 2 Well plate (Ibidi). After appropriate cell attachment, the well was gently removed, and a scratch was made across the cell monolayer. The wounded cells were cultured with serum-free culture medium and imaged at different time points after scratching.

ISH

Detection of MIR99AHG expression alone in CRC cells and TMAs was measured by using an ISH optimization kit (Roche) according to the manufacturer's instructions. The locked nucleic acid (LNA)-modified oligonucleotide probe targeting MIR99AHG was purchased from Exiqon. For pretreatment, CRC cells seeded on a Lab-Tek chamber slide (Nunc) were fixed in prechilled 4% paraformaldehyde (PFA) at 4°C for 15 min, and the TMA sections were deparaffinized with xylene, followed by rehydration in serial dilutions of ethanol. After washing three times, the sections were incubated in proteinase K for 20 min. The sections were reconstituted using hybridization solution and incubated with the digoxigenin-labeled MIR99AHG probe at 56°C overnight. After washing, the sections were blocked with 5% normal goat serum at room temperature (RT) for 1 h followed by incubation in an anti-digoxigenin alkaline phosphatase conjugate (Roche) at 4°C overnight. Colorimetric signals were obtained by incubating the sections with NBT/BCIP buffer (Roche) in the dark at RT for 4 h. Nuclear fast red was used as the counterstain.

Detection of SMARCA1-L expression in the TMAs or the colocalization of MIR99AHG and SMARCA1 pre-mRNA in CRC cells was measured by using the RNAscope Multiplex Fluorescent Kit v2 (Advanced Cell Diagnostics) according to the manufacturer's instructions. The probes for MIR99AHG, SMARCA1 pre-mRNA and SMARCA1-L were designed by Advanced Cell Diagnostics. Briefly, pretreated sections were incubated in RNAscope probe mixture at 40°C for 2 h to hybridize the probes. The sections were incubated in the detection reagents for AMP signal amplification in the following order: AMP 1 for 30 min, AMP 2 for 30 min and AMP 3 for 15 min. Horseradish peroxidase (HRP) was linked to probe channels (C1 for MIR99AHG and C2 for SMARCA1-L). HRP-C1 or HRP-C2 (4-6 drops) was added to entirely cover each slide at 40°C for 15 min. Fluorescent tyramide signal amplification was used to develop the signal. Finally, sections were counterstained with DAPI and cover slipped with ProLong Gold (Thermo Fisher Scientific). Confocal images were acquired with an Olympus FV3000 inverted confocal microscope.

IHC

Paraffin-embedded specimens were sectioned, deparaffinized and treated with 3% H_2O_2 to block endogenous peroxidase activity. Slides were immersed in antigen retrieval buffer, heated to 120°C for 20 min and then allowed to cool to RT. Incubation of primary antibody was performed at 4°C overnight, and secondary antibody was incubated at RT for 1 h, followed by visualization with diaminobenzidine and image acquisition with a light microscope. Ki-67 staining was quantified by calculating positively stained cells in five randomly chosen fields of each sample. For multispectral staining, the TMA was baked at 62°C for 3 h, and subsequent deparaffinization was performed on a Leica Bond RX followed by four sequential cycles of staining, with each round including a 30 min combined block and primary antibody incubation. Detection of the primary antibodies was performed by applying a secondary HRP-conjugated polymer (Perkin Elmer) for 10 min. The HRP-conjugated secondary antibody polymer was detected using fluorescent tyramide signal amplification using Opal dyes 540, 570, 650 and 690 (Perkin Elmer). The covalent tyramide reaction was followed by heat-induced stripping of the primary/secondary antibody complex using Perkin Elmer AR9 buffer at 100°C for 20 min preceding the next cycle. After four sequential rounds of staining, the sections were stained with DAPI to visualize the nuclei and mounted with Prolong Gold antifade reagent mounting medium (Invitrogen).

Quantification of ISH and IHC signals

The results of ISH and multispectral IHC were scored by two independent observers. Expression was visualized and classified based on the intensity of staining and the percentage of positive cells. The intensity score was divided into four grades: 0, no staining; 1, weak staining; 2, moderate staining; and 3, strong staining. The percentage score was also divided into four grades: 0, ≤1%; 1, 2%-25%; 2, 26%-50%; 3, 51%-75%; and 4 ≥75%. The histological score (H-score) was calculated by multiplying the intensity score by the percentage score. Overall scores of 0-12 were classified as negative (score: 0), weak (score: 1-4), moderate (score: 5-8) or strong (score: 9-12).

Cell viability measurement

CRC cells with the indicated treatments were seeded in a 96-well plate and allowed to adhere to the wells. At each time point, the culture medium was replaced with Cell Counting Kit-8 solution (Dojindo), and then the cells were incubated at 37°C for 2 h. The absorbance values were read at 450 nm using a Thermo Scientific Varioskan Flash multimode reader.

Cell cycle and apoptosis analysis

For cell cycle analysis, CRC cells with the indicated treatments were serum starved for 12 h to induce cell cycle synchronization. The cells were harvested, resuspended and fixed in 70% ethanol at 4°C for 24 h and stained with propidium iodide (PI, Sigma-Aldrich). For apoptosis analysis, CRC cells with the indicated treatments were harvested, resuspended in staining buffer and examined using the Annexin V-FITC Early Apoptosis Detection Kit (BD Biosciences). Cells positive for Annexin V-FITC and negative for PI were considered to have undergone apoptosis. Cell cycle and apoptosis were analyzed using a fluorescence-activated cell sorter (BD Biosciences), and the data were analyzed using EXPO32 ADC software.

Colony formation assay

CRC cells with the indicated treatments were seeded in 6-well plates at a density of 1×10^3 cells per well. After 2 weeks of incubation, the colonies were fixed with 75% ethanol and stained with crystal violet dye. The colonies in all wells were counted with a GelCount colony counter (Oxford Optronix), and each group had three replicates.

RT-PCR and RT-qPCR

Total RNA was extracted from cells using TRIzol reagent (Invitrogen). cDNA synthesis was carried out using the PrimeScript RT Reagent Kit (TaKaRa). RT-PCR was performed using KOD-Plus-Neo DNA polymerase (Toyobo). Reaction products were analyzed on 1-2% agarose gels with

ethidium bromide staining. qPCR was performed using the SYBR Premix Ex Taq Kit (TaKaRa). Fluorescence was measured, and CT values were calculated using a CFX96 Real-Time PCR Detection System (Bio-Rad). Primers for miRNAs were purchased from RiboBio. Other primers are described in Supplementary Table 3.

Immunofluorescence

CRC cells with the indicated treatment were seeded on a Lab-Tek chamber slide (Nunc) and fixed in 4% PFA for 20 min. The cells were permeabilized using 0.1% Triton X-100 in PBS for 15 min and blocked with 2% BSA in PBS for 30 min. Incubation of primary antibodies was performed at 4°C overnight, and secondary antibodies were incubated at RT for 1 h. The nuclei were stained with DAPI, and fluorescence images were taken on a Nikon A1 Confocal Laser Microscope System.

Electron microscopy

For transmission electron microscopy analysis, CRC cells after the indicated treatment were collected by trypsinization and prefixed in a mixture of 2.5% glutaraldehyde and 4% PFA at 4°C overnight and washed with phosphate buffer. Then, the cells were dehydrated using a graded ethanol immersion series and embedded in resin. Cross-sections (800 nm thick) were cut, stained with toluidine blue and observed to identify areas with cytoplasmic inclusions. Ultrathin sections (90 nm) were prepared, stained with uranyl acetate and lead citrate, and examined under a transmission electron microscope (Tecnai G2 Spirit, FEI company). For scanning electron microscopy analysis, CRC cells after the indicated treatment were prefixed with 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer for 30 min. The cells were washed with cacodylate buffer, dehydrated through a graded series of ethanol, and dried with hexamethyldisilazane. The samples were then mounted on aluminum stubs and coated with gold/palladium. The cells were examined under a scanning electron microscope (S-4800, Hitachi).

Protein extraction and immunoblotting

Proteins were extracted from cultured cells or concentrated FBS-free media using RIPA buffer (Beyotime) supplemented with protease and phosphatase inhibitors (Roche) and boiled with SDS loading buffer for 10 min. The denatured proteins were separated on a 6-12% SDS-PAGE gel for immunoblotting analysis with the indicated antibodies listed in Supplementary Table 5. The blotted bands were visualized by the Bio-Rad ChemiDoc XRS+ imaging system.

RNA pulldown

To synthesize biotin-labeled transcripts of MIR99AHG, PCR fragments were prepared using forward primers harboring the T7 RNA polymerase promoter. Following purification of the PCR products, biotinylated transcripts were synthesized using the MaxiScript T7 Kit (Ambion). Biotinylated RNA (1 μg) was first folded in RNA structure buffer (0.1 M KCl, 10 mM $MgCl_2$ and 10 mM Tris-HCl, pH 7.0) and then incubated with cell lysates at 25°C for 3 h with rotation. After incubation, RNA-protein complexes were isolated with streptavidin-coupled Dynabeads (Invitrogen). The retrieved RNA-associated proteins were fractionated by SDS-PAGE and visualized by silver staining. Protein bands present only in specific RNA samples were excised and identified by mass spectrometry. The retrieved protein samples were confirmed by immunoblotting.

Immunoprecipitation

Cells were lysed in RIPA buffer containing protease inhibitor cocktail. Protein samples with or without nuclease treatment were incubated with a specific antibody or control IgG at 4°C overnight. Subsequently, the samples were incubated with 50 μl of protein G magnetic beads (Thermo Fisher Scientific) at 4°C for 1 h and then washed five times in lysis buffer. Protein complexes were eluted with SDS buffer and then detected by immunoblotting.

RIP

For native RIP assays, RNase OUT (Invitrogen) and a protease inhibitor cocktail were added to RIPA buffer. After incubation with a specific antibody or control IgG at 4°C overnight, the cell lysates were incubated with 50 μl of protein G magnetic beads at 4°C for 1 h and then washed three times in lysis buffer. The beads were resuspended and treated with proteinase K at 45°C for 45 min. Coprecipitated RNAs were extracted using TRIzol reagent, ethanol-precipitated with Glycoblue (Invitrogen) as a carrier and then detected by RT-qPCR. The retrieved RNA data are presented as a percentage of the amount input. For RIP-based mapping assays, lysates were first mixed with RNase T1 (Thermo Fisher Scientific), after which standard native RIP assays were performed using an antibody against PTBP1. Following extraction of the coprecipitated RNA, the MIR99AHG segments bound by PTBP1, and hence protected from RNase T1 digestion and immunoprecipitated, were identified by RT-qPCR analysis using primer sets that scanned the MIR99AHG transcript at ~150-nt-long, overlapping intervals. The PCR primers used are listed in Supplementary Table 3.

ChIRP

A total of 6×10^7 cells were cross-linked by 1% glutaraldehyde at RT for 10 min, followed by three washes in cold PBS. After being snap-frozen in liquid nitrogen and stored at −80°C, crosslinked cells were resuspended in nuclear lysis buffer (10 mM EDTA, 1% SDS and 50 mM Tris-HCl, pH 7.5) supplemented with a protease inhibitor cocktail and sonicated until DNA was in the size range of 100-500 bp. Cell lysates and a set of biotin-labeled antisense probes (20 nt in length) were then incubated at 37°C for 4 h. Briefly, 18 probes unique to the MIR99AHG sequence and spanning the entire transcript were designed using the online designer at Stellaris (http://www.singlemoleculefish.com). Three probes were designed against LacZ RNA as a negative control. Probes were labeled according to their positions along the RNA and separated into two pools so that the “even” pool contained all probes numbering 2, 4, 6, etc. and the “odd” pool contained probes numbering 1, 3, 5, etc. Experiments were performed using both pools, which served as internal controls for each other. Streptavidin-coupled Dynabeads were added to pull down

the probes. After washing, the retrieved DNA was isolated using the ChIP DNA Clean & Concentrator Kit (Zymo Research) and subjected to qPCR analysis. The probes used for ChIRP are shown in Supplementary Table 3.

RAP

Cells were suspended in PBS at a concentration of 2×10^7 cells/ml in 6-well cell culture plates and incubated on ice for 15 min. The cells were irradiated for 15 min with 365-nm UV light, followed by RNA isolation using TRIzol reagent. To pull down the target RNAs, 15 pM biotin-labeled antisense probes (90 nt in length) were denatured at 85°C for 3 min and then transferred immediately on ice. Then, probes and 2 μg RNA were mixed in LiCl hybridization buffer (1 mM EDTA, 500 mM LiCl, 1% Triton X-100, 0.2% SDS, 0.1 sodium deoxycholate and 10 mM Tris-HCl, pH 7.5) and transferred to a 55°C thermomixer with shaking at 1,200 rpm for 2 h. Then, 200 μl streptavidin-coupled Dynabeads were added and incubated at RT for 2 h. After washing, the retrieved RNAs were extracted by TRIzol reagent and subjected to RT-qPCR analysis. The probes used for RAP are shown in Supplementary Table 3.

Statistics

Statistical analysis was performed by SPSS 18.0 and R (version 3.3.1). The statistical significance of differences between datasets was evaluated according to p values, and a $p < 0.05$ was considered to indicate statistical significance. The mean ± SEM (n = 3) of a representative of 3 triplicate experiments are shown. Two-tailed Student's t test, ANOVA with Dunnett's multiple comparison, nonparametric signed rank test, Chi-square test and Pearson correlation coefficients were used according to the type of experiment, and the test type is indicated in the figure legends.

Reference

1. Lu Y, Zhao X, Liu Q, Li C, Graves-Deal R, Cao Z, Singh B, Franklin JL, Wang J, Hu H, et al: **lncRNA MIR100HG-derived miR-100 and miR-125b mediate cetuximab resistance via Wnt/beta-catenin signaling.** *Nat Med* 2017, **23:**1331-1341.
2. Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: Database for Annotation, Visualization, and Integrated Discovery.** *Genome Biol* 2003, **4:**P3.
3. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102:**15545-15550.
4. Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y: **rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data.** *Proc Natl Acad Sci U S A* 2014, **111:**E5593-5601.

Figure S1

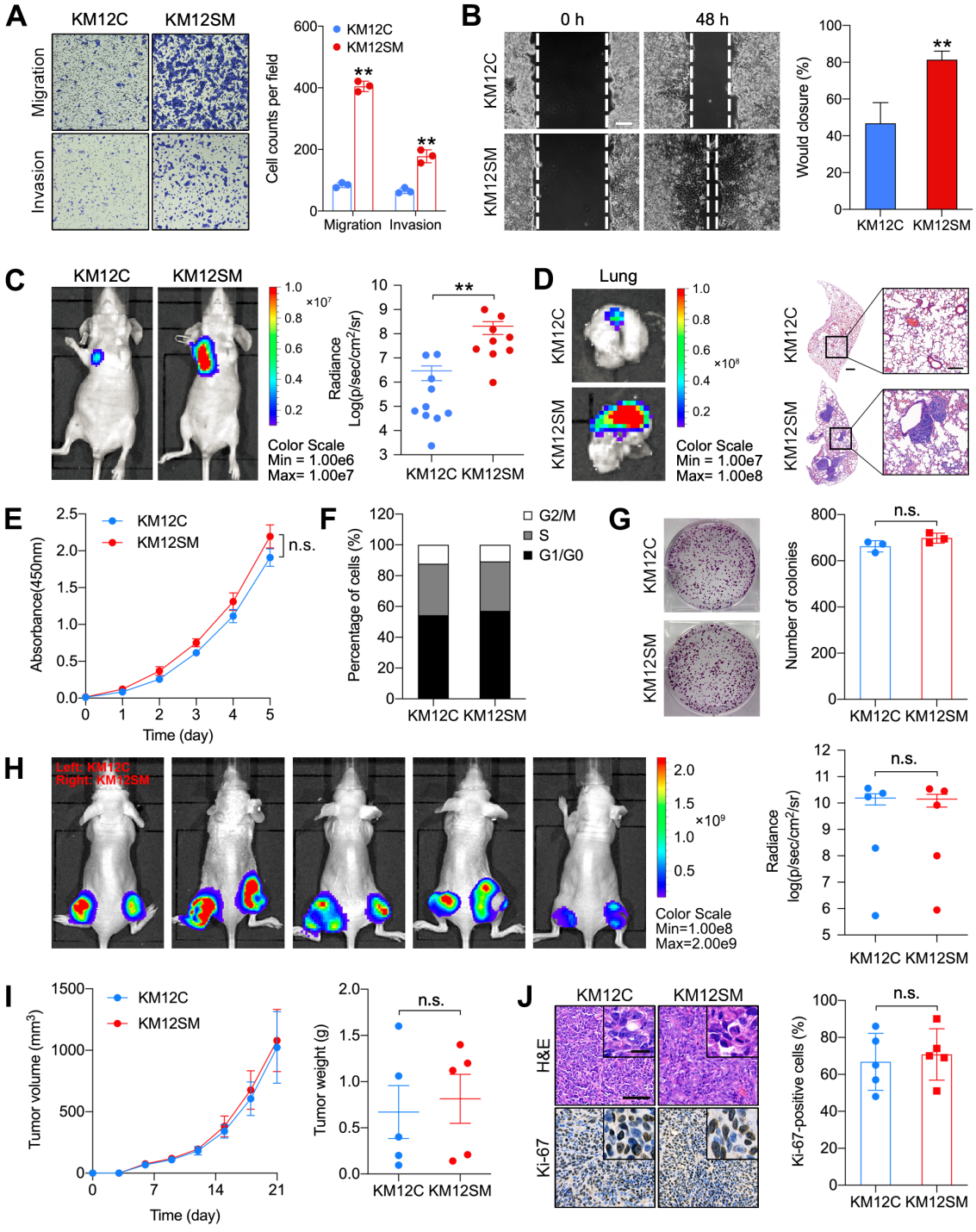
A
KM12C
KM12SM
Migration
Invasion
Cell counts per field
Migration
Invasion
B
0 h
48 h
KM12C
KM12SM
Would closure (%)
C
Radiance Log(p/sec/cm²/sr)
Color Scale Min = 1.00e6 Max= 1.00e7
D
Lung
Color Scale Min = 1.00e7 Max= 1.00e8
E
Absorbance(450nm)
Time (day)
n.s.
F
Percentage of cells (%)
G2/M
S
G1/G0
G
Number of colonies
H
Left: KM12C
Right: KM12SM
Color Scale Min=1.00e8 Max=2.00e9
I
Tumor volume (mm³)
Tumor weight (g)
J
H&E
Ki-67
Ki-67-positive cells (%)

Figure S1. KM12SM and KM12C exhibited different capacities in metastasis but not in proliferation

(A) Migration and invasion of KM12C and KM12SM cells. Representative images are shown on the left, and the quantification results for three randomly selected fields are shown on the right. ** $P < 0.01$ by Student's *t* test.

(B) Wound healing assays of KM12C and KM12SM cells. Representative images are shown on the left, and the percentage of wound closure normalized to 0 h in three randomly selected fields is shown on the right. Scale bar, 200 μm. ** $P < 0.01$ by Student's *t* test.

(C, D) Representative bioluminescence images of metastases in nude mice (n=10 mice per group) after tail vein injection of KM12C and KM12SM cells are shown in (C). The radiance values of lung metastases from the indicated groups are shown. Representative bioluminescence images and hematoxylin and eosin (H&E) staining of lung metastatic nodules from the indicated groups are shown in (D). Scale bars: main, 500 μm; inset, 200 μm. ** $P < 0.01$ by Student's *t* test.

(E-G) Growth curves (E), cell cycle analyses (F) and colony formation assays (G) in KM12C and KM12SM cells.

(H-J) Bioluminescence image of tumors in nude mice after subcutaneous injection of KM12C and KM12SM cells (n = 5 mice), and the radiance values of tumors from the indicated groups are shown (H). Quantification of growth curves and tumor weights from the indicated groups are shown (I). Xenograft tumors were stained with H&E and Ki-67, and the percentage of Ki-67-positive cells was measured (J). Scale bars: main, 100 μm; inset, 20 μm.

Figure S2

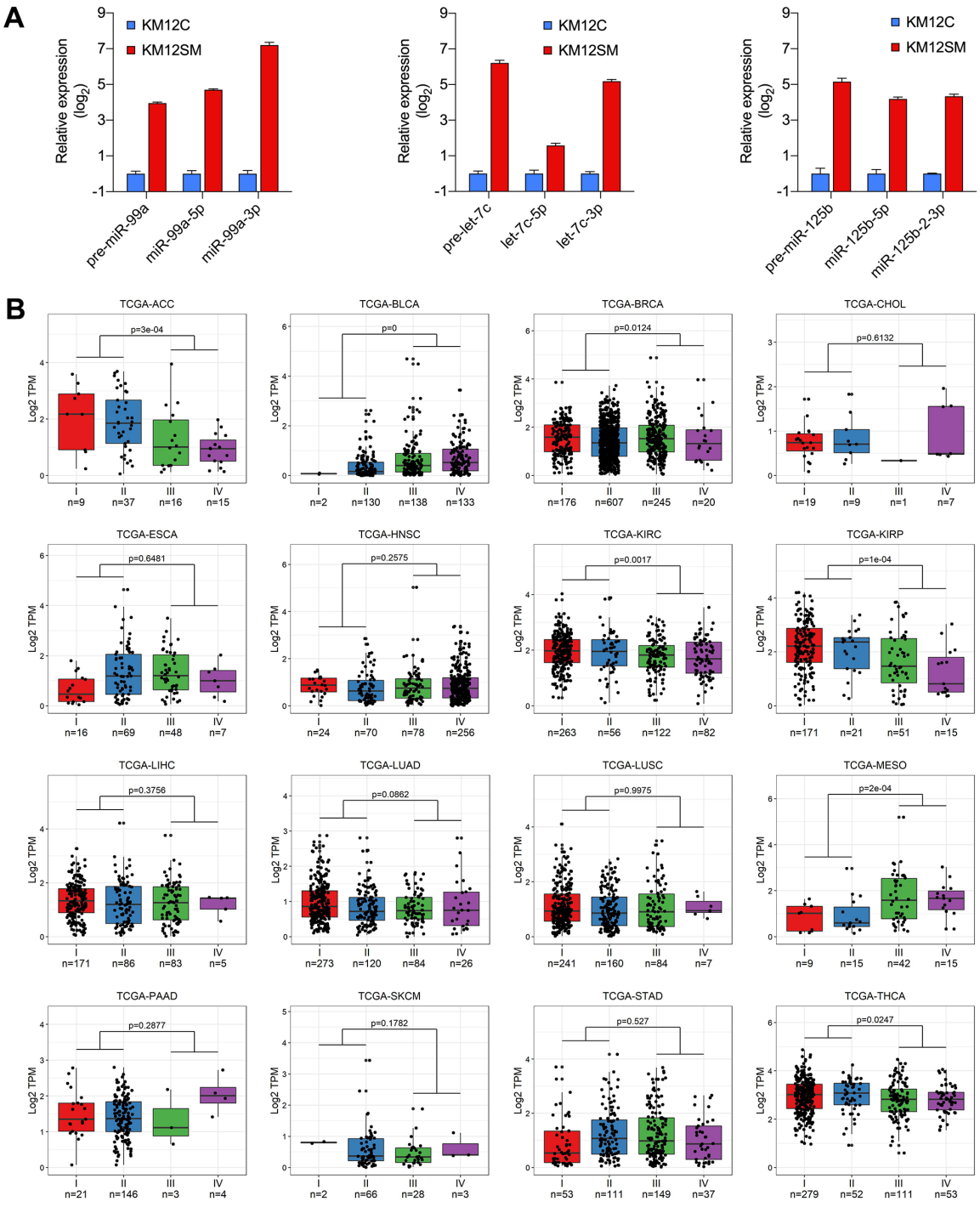
A
KM12C
KM12SM
Relative expression (log2)
pre-miR-99a
miR-99a-5p
miR-99a-3p
pre-let-7c
let-7c-5p
let-7c-3p
pre-miR-125b
miR-125b-5p
miR-125b-2-3p
B
TCGA-ACC
p=3e-04
TCGA-BLCA
p=0
TCGA-BRCA
p=0.0124
TCGA-CHOL
p=0.6132
TCGA-ESCA
p=0.6481
TCGA-HNSC
p=0.2575
TCGA-KIRC
p=0.0017
TCGA-KIRP
p=1e-04
TCGA-LIHC
p=0.3756
TCGA-LUAD
p=0.0862
TCGA-LUSC
p=0.9975
TCGA-MESO
p=2e-04
TCGA-PAAD
p=0.2877
TCGA-SKCM
p=0.1782
TCGA-STAD
p=0.527
TCGA-THCA
p=0.0247
Log2 TPM

Figure S2. Expression of MIR99AHG-derived miRNAs and MIR99AHG in CRC cells and other cancer types

(A) RT-qPCR of primary and mature transcripts of miR-125b, let-7c and miR-99a in KM12C and KM12SM cells. U6 snRNA served as the internal control. ** $P < 0.01$ by Student's *t* test.

(B) Box plots showing the expression pattern of MIR99AHG by stage from the TCGA data repository. Statistical significance was determined by the Mann-Whitney U test. ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; cholangiocarcinoma; ESCA, esophageal carcinoma; HNSC, head-neck squamous cell carcinoma; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; PAAD, pancreatic adenocarcinoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; THCA, thyroid carcinoma.

Figure S3

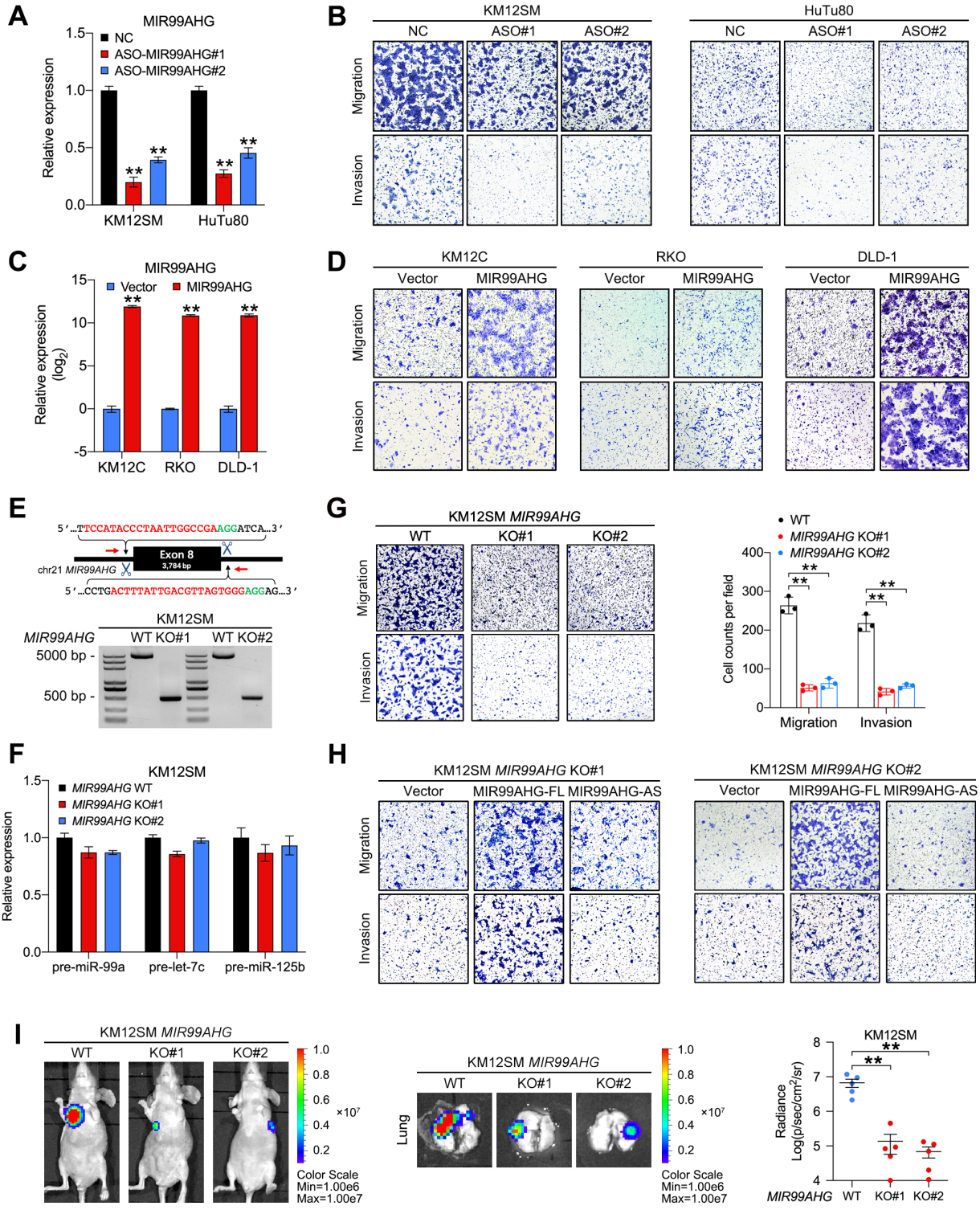
A
MIR99AHG
NC
ASO-MIR99AHG#1
ASO-MIR99AHG#2
Relative expression
KM12SM
HuTu80
B
KM12SM
HuTu80
NC
ASO#1
ASO#2
Migration
Invasion
C
MIR99AHG
Vector
MIR99AHG
Relative expression (log2)
KM12C
RKO
DLD-1
D
KM12C
RKO
DLD-1
Vector
MIR99AHG
Migration
Invasion
E
5′…TTCCATACCCTAATTGGCCGAAGGATCA…3′
Exon 8
3,784 bp
chr21 MIR99AHG
5′…CCTGACTTTATTGACGTTAGTGGGAGGAG…3′
KM12SM
MIR99AHG
WT KO#1 WT KO#2
5000 bp -
500 bp -
F
KM12SM
MIR99AHG WT
MIR99AHG KO#1
MIR99AHG KO#2
Relative expression
pre-miR-99a
pre-let-7c
pre-miR-125b
G
KM12SM MIR99AHG
WT
KO#1
KO#2
Migration
Invasion
WT
MIR99AHG KO#1
MIR99AHG KO#2
Cell counts per field
Migration
Invasion
H
KM12SM MIR99AHG KO#1
KM12SM MIR99AHG KO#2
Vector
MIR99AHG-FL
MIR99AHG-AS
Migration
Invasion
I
KM12SM MIR99AHG
WT
KO#1
KO#2
×10⁷
Color Scale
Min=1.00e6
Max=1.00e7
Lung
Radiance
Log(p/sec/cm²/sr)
KM12SM
MIR99AHG

Figure S3. MIR99AHG promotes migration, invasion and metastasis in CRC cells

(A) RT-qPCR analysis of MIR99AHG expression in KM12SM and HuTu80 cells transfected with ASOs against MIR99AHG or a negative control.

(B) Representative images of Transwell migration and invasion assays in Figure 2A.

(C) RT-qPCR analysis of MIR99AHG expression in KM12C, RKO and DLD-1 cells transfected with a MIR99AHG construct or a control vector.

(D) Representative images of Transwell migration and invasion assays in Figure 2B.

(E) Top, schematic illustration of the targeting strategy to delete exon 8 in *MIR99AHG*. Sequences targeted by CRISPR/Cas9 are shown in red. Protospacer adjacent motifs are shown in green. Red arrows indicate PCR primers used for sequencing. Bottom, representative agarose gel displaying PCR products of primers flanking exon 8 of MIR99AHG using genomic DNA isolated from wild-type and two clones of $MIR99AHG^{KO}$ KM12SM cells.

(F) RT-qPCR analysis of the expression of MIR99AHG and the miR-99a/let-7c/miR-125b-2 cluster in WT and $MIR99AHG^{KO}$ KM12SM cells.

(G) Transwell migration and invasion assays of WT and $MIR99AHG^{KO}$ KM12SM cells. ** $P < 0.01$ by one-way ANOVA (Dunnett's test).

(H) Representative images of Transwell migration and invasion assays in Figure 2C.

(I) Representative bioluminescence images of metastases in whole mice and lungs after tail vein injection of WT or $MIR99AHG^{KO}$ KM12SM cells (n = 5 mice per group). The radiance values of lung metastatic nodules from the indicated groups are shown. ** $P < 0.01$ by one-way ANOVA (Dunnett's test).

Figure S4

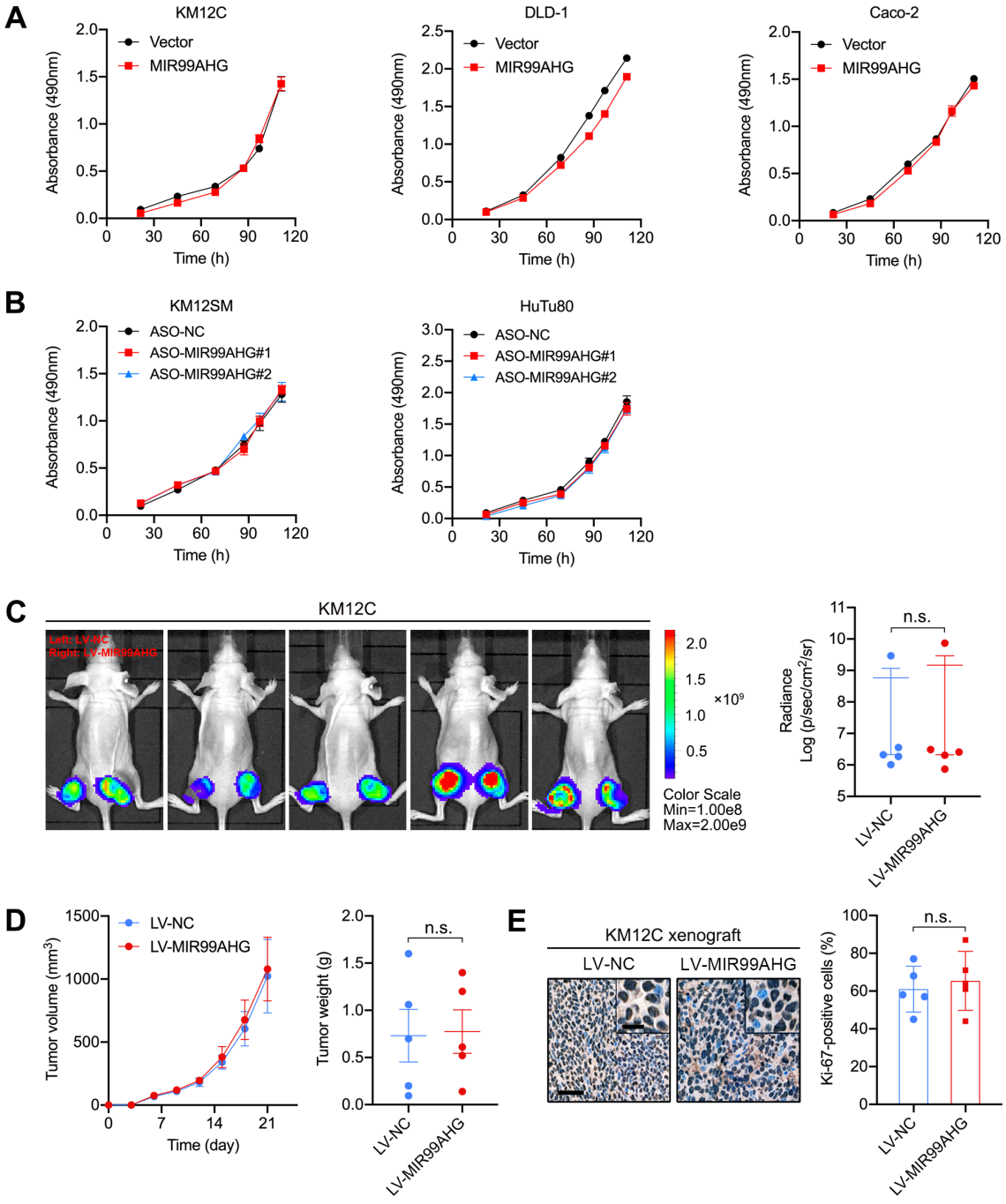
A
KM12C
DLD-1
Caco-2
Vector
MIR99AHG
Absorbance (490nm)
Time (h)
B
KM12SM
HuTu80
ASO-NC
ASO-MIR99AHG#1
ASO-MIR99AHG#2
C
KM12C
Left: LV-NC
Right: LV-MIR99AHG
×10⁹
Color Scale
Min=1.00e8
Max=2.00e9
Radiance
Log (p/sec/cm²/sr)
n.s.
LV-NC
LV-MIR99AHG
D
LV-NC
LV-MIR99AHG
Tumor volume (mm³)
Time (day)
Tumor weight (g)
E
KM12C xenograft
LV-NC
LV-MIR99AHG
Ki-67-positive cells (%)

Figure S4. MIR99AHG has no effect on proliferation and tumor growth in CRC cells

(A) Growth curves of KM12C, DLD-1 and Caco-2 cells transfected with plasmids expressing MIR99AHG or a control vector.

(B) Growth curves of KM12SM and HuTu80 cells transfected with ASOs against MIR99AHG or a negative control.

(C-E) Bioluminescence image of tumors in nude mice after subcutaneous injection of KM12C cells stably expressing MIR99AHG or control vector. The radiance values of tumors from the indicated groups are shown (C). Quantification of growth curves and tumor weights from the indicated groups are shown (D). Xenograft tumors were stained with Ki-67, and the percentage of Ki-67-positive cells was measured (E). Scale bars: main, 100 μm; inset, 20 μm.

Figure S5

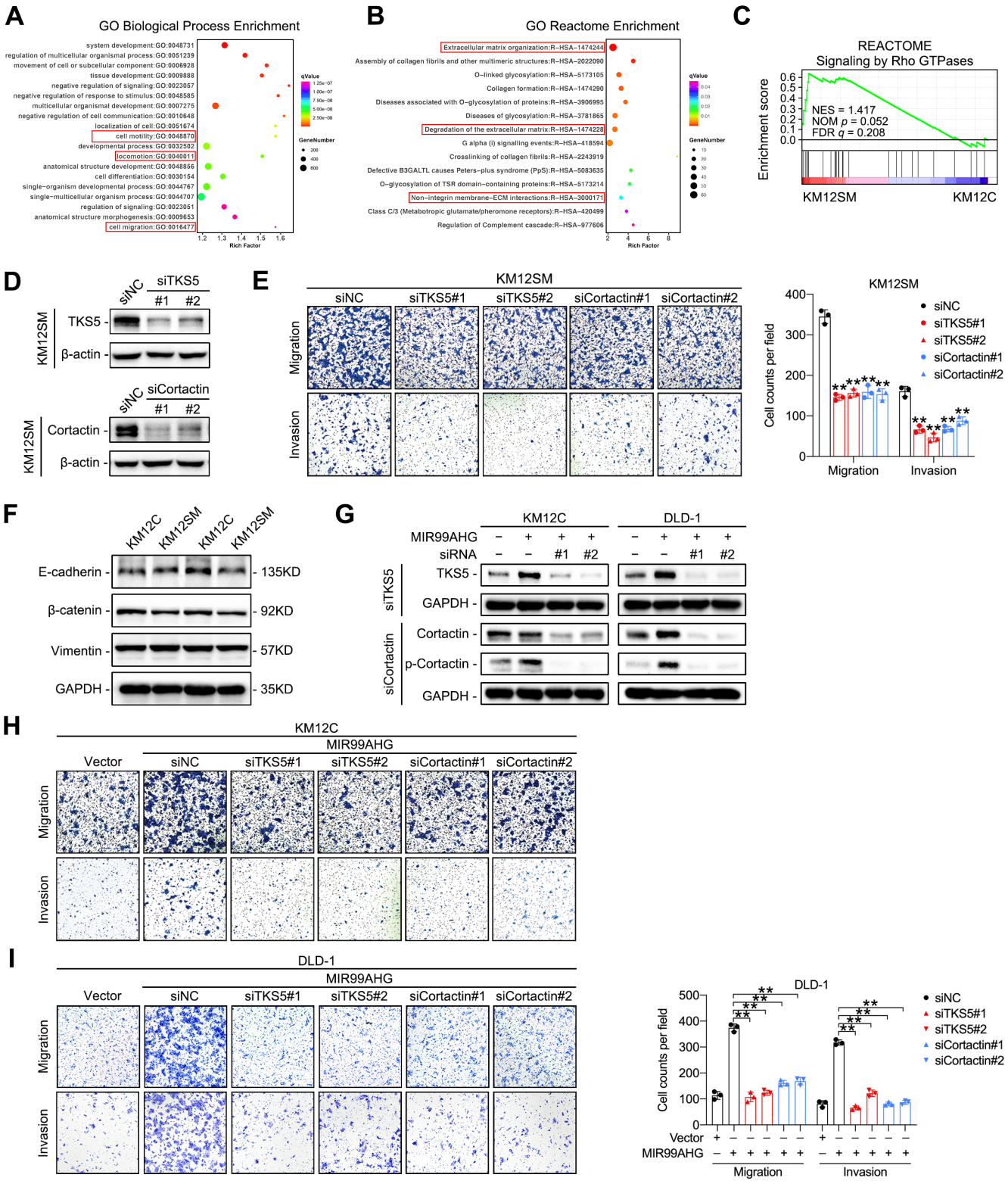
A
GO Biological Process Enrichment
system development:GO:0048731
regulation of multicellular organismal process:GO:0051239
movement of cell or subcellular component:GO:0006928
tissue development:GO:0009888
negative regulation of signaling:GO:0023057
negative regulation of response to stimulus:GO:0048585
multicellular organismal development:GO:0007275
negative regulation of cell communication:GO:0010648
localization of cell:GO:0051674
cell motility:GO:0048870
developmental process:GO:0032502
locomotion:GO:0040011
anatomical structure development:GO:0048856
cell differentiation:GO:0030154
single-organism developmental process:GO:0044767
single-multicellular organism process:GO:0044707
regulation of signaling:GO:0023051
anatomical structure morphogenesis:GO:0009653
cell migration:GO:0016477
Rich Factor
qValue
GeneNumber
B
GO Reactome Enrichment
Extracellular matrix organization:R-HSA-1474244
Assembly of collagen fibrils and other multimeric structures:R-HSA-2022090
O-linked glycosylation:R-HSA-5173105
Collagen formation:R-HSA-1474290
Diseases associated with O-glycosylation of proteins:R-HSA-3906995
Diseases of glycosylation:R-HSA-3781865
Degradation of the extracellular matrix:R-HSA-1474228
G alpha (i) signalling events:R-HSA-418594
Crosslinking of collagen fibrils:R-HSA-2243919
Defective B3GALTL causes Peters-plus syndrome (PpS):R-HSA-5083635
O-glycosylation of TSR domain-containing proteins:R-HSA-5173214
Non-integrin membrane-ECM interactions:R-HSA-3000171
Class C/3 (Metabotropic glutamate/pheromone receptors):R-HSA-420499
Regulation of Complement cascade:R-HSA-977606
Rich Factor
qValue
GeneNumber
C
REACTOME
Signaling by Rho GTPases
Enrichment score
NES = 1.417
NOM p = 0.052
FDR q = 0.208
KM12SM
KM12C
D
siNC
siTKS5
#1
#2
KM12SM
TKS5
β-actin
siNC
siCortactin
#1
#2
KM12SM
Cortactin
β-actin
E
KM12SM
siNC
siTKS5#1
siTKS5#2
siCortactin#1
siCortactin#2
Migration
Invasion
KM12SM
siNC
siTKS5#1
siTKS5#2
siCortactin#1
siCortactin#2
Cell counts per field
Migration
Invasion
F
KM12C
KM12SM
KM12C
KM12SM
E-cadherin
135KD
β-catenin
92KD
Vimentin
57KD
GAPDH
35KD
G
KM12C
DLD-1
MIR99AHG
siRNA
#1
#2
siTKS5
TKS5
GAPDH
siCortactin
Cortactin
p-Cortactin
GAPDH
H
KM12C
MIR99AHG
Vector
siNC
siTKS5#1
siTKS5#2
siCortactin#1
siCortactin#2
Migration
Invasion
I
DLD-1
MIR99AHG
Vector
siNC
siTKS5#1
siTKS5#2
siCortactin#1
siCortactin#2
Migration
Invasion
DLD-1
siNC
siTKS5#1
siTKS5#2
siCortactin#1
siCortactin#2
Cell counts per field
Vector
MIR99AHG
Migration
Invasion

Figure S5. MIR99AHG modulates invadopodia formation in CRC cells

(A, B) GO analysis showing the top 20 biological processes (A) and Reactome pathways (B) enriched in KM12SM cells compared to KM12C.

(C) GSEA plot showing gene signatures enriched in signaling by Rho GTPases in KM12C and KM12SM cells.

(D, E) Immunoblotting of TKS5 and Cortactin (D) and Transwell migration and invasion assays (E) of KM12SM cells transfected with siRNAs against TKS5 or Cortactin. ** $P < 0.01$ by one-way ANOVA (Dunnett's test).

(F) Immunoblotting of the indicated EMT-related markers in KM12C and KM12SM cells.

(G) Immunoblots of TKS5, Cortactin and p-Cortactin of the MIR99AHG-expressing KM12C and DLD-1 cells transfected with siRNA against TKS5 or Cortactin.

(H) Representative images of Transwell migration and invasion assays in Figure 2L.

(I) Transwell migration and invasion assays of the MIR99AHG-expressing DLD-1 cells transfected with siRNA against TKS5 or Cortactin. ** $P < 0.01$ by one-way ANOVA (Dunnett's test).

Figure S6

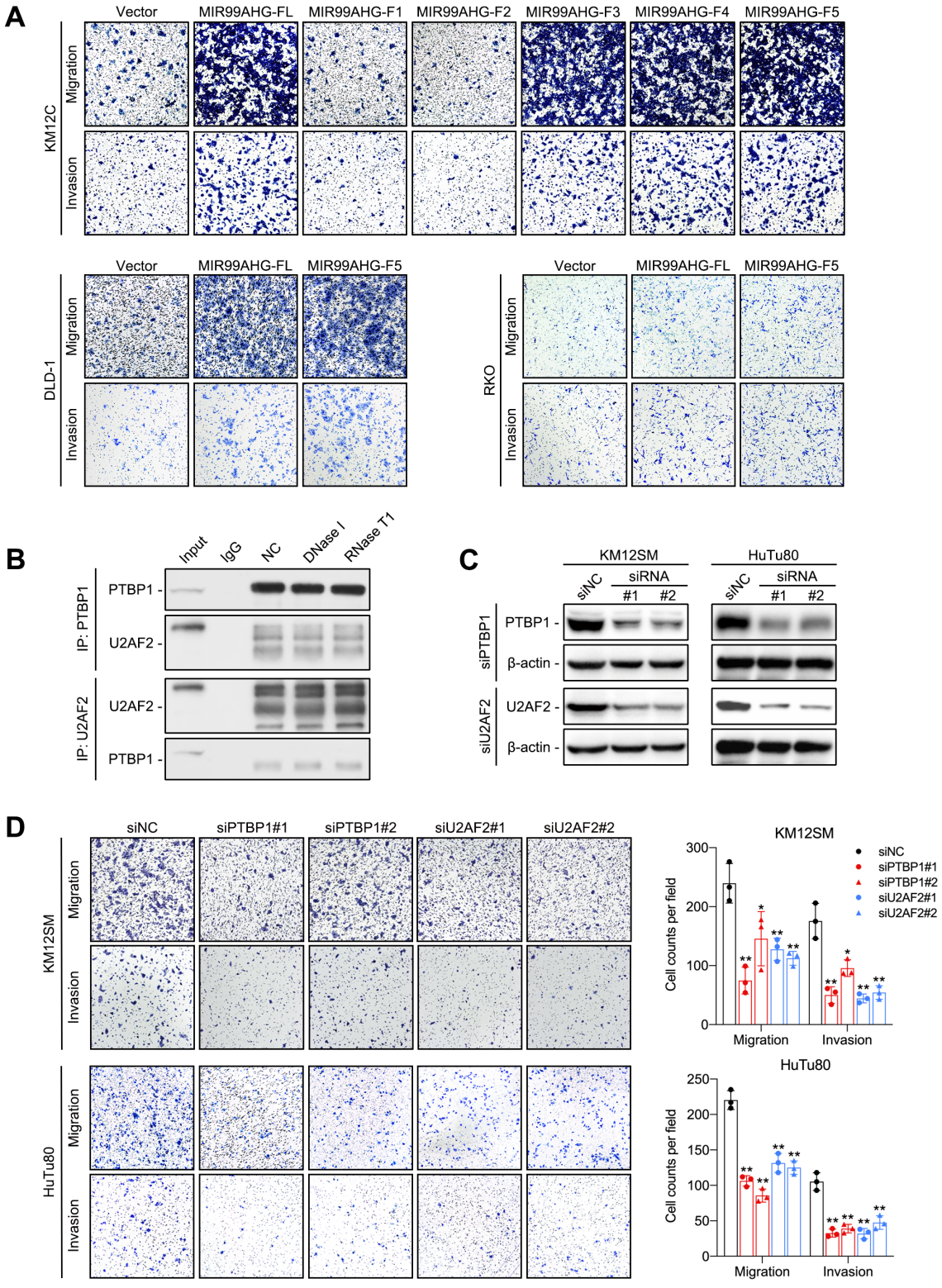
A
Vector
MIR99AHG-FL
MIR99AHG-F1
MIR99AHG-F2
MIR99AHG-F3
MIR99AHG-F4
MIR99AHG-F5
KM12C
Migration
Invasion
DLD-1
RKO
B
Input
IgG
NC
DNase I
RNase T1
IP: PTBP1
IP: U2AF2
PTBP1
U2AF2
C
KM12SM
HuTu80
siNC
siRNA
#1
#2
siPTBP1
siU2AF2
β-actin
D
siNC
siPTBP1#1
siPTBP1#2
siU2AF2#1
siU2AF2#2
Cell counts per field
Migration
Invasion

Figure S6. MIR99AHG functions via its interaction with PTBP1 in CRC cells

(A) Representative images of Transwell migration and invasion assays in Figure 3B.

(B) KM12SM cells treated with DNase I or RNase T1 were immunoprecipitated with PTBP1 or U2AF2 antibodies or normal IgG.

(C) Immunoblotting of PTBP1 and U2AF2 in KM12SM and HuTu80 cells transfected with siRNAs against PTBP1, U2AF2 or a negative control.

(D) Migration and invasion of KM12SM and HuTu80 cells transfected with siRNAs against PTBP1, U2AF2 or a negative control. ** $P < 0.01$ by one-way ANOVA (Dunnett's test).

Figure S7

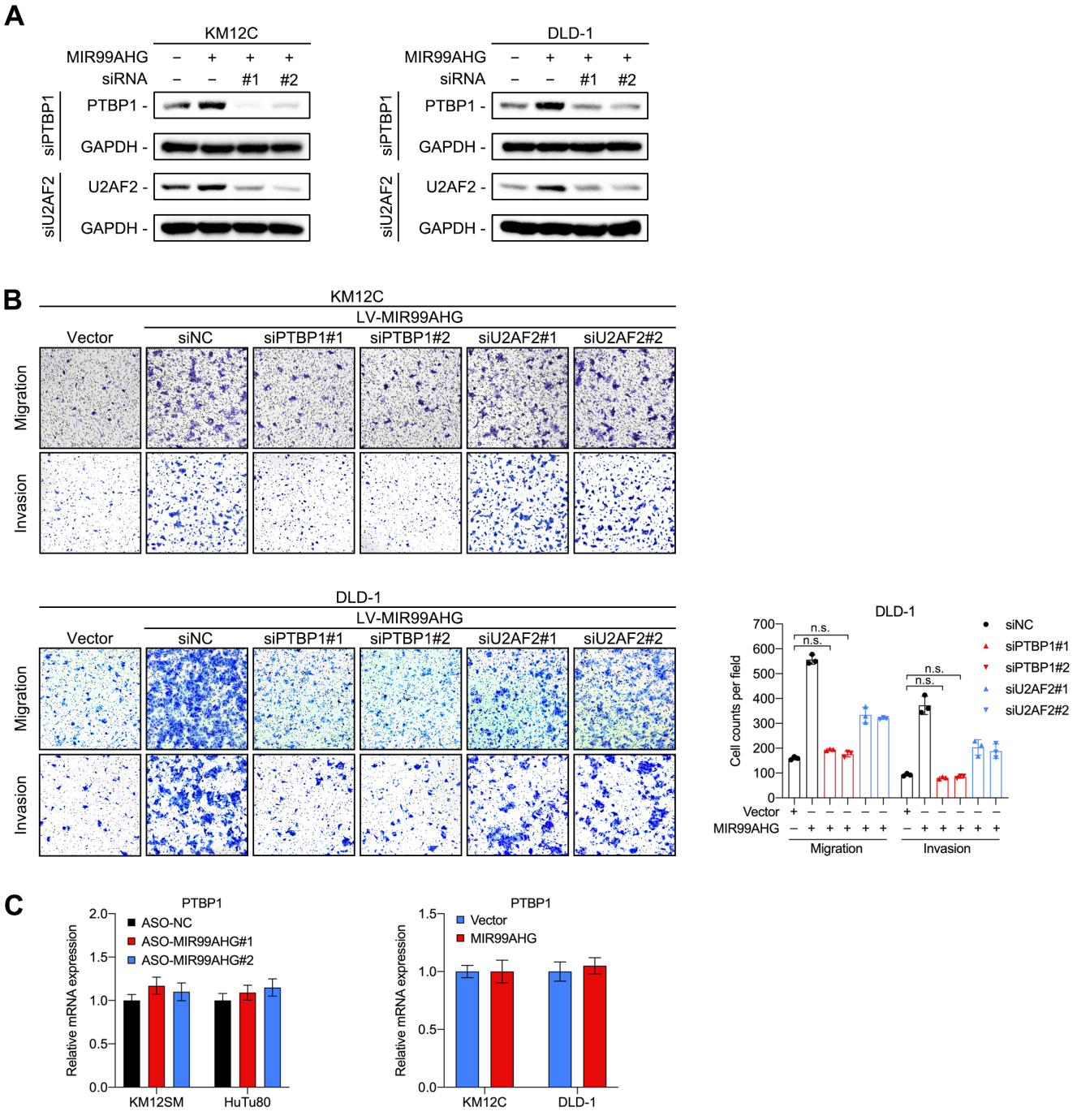
A
KM12C
DLD-1
MIR99AHG
siRNA
#1
#2
siPTBP1
siU2AF2
PTBP1 -
U2AF2 -
GAPDH -
B
KM12C
LV-MIR99AHG
Vector
siNC
siPTBP1#1
siPTBP1#2
siU2AF2#1
siU2AF2#2
Migration
Invasion
DLD-1
Cell counts per field
n.s.
siNC
siPTBP1#1
siPTBP1#2
siU2AF2#1
siU2AF2#2
C
PTBP1
ASO-NC
ASO-MIR99AHG#1
ASO-MIR99AHG#2
Relative mRNA expression
KM12SM
HuTu80
Vector
MIR99AHG
KM12C
DLD-1

Figure S7. MIR99AHG functions via its interaction with PTBP1 in CRC cells

(A) Immunoblotting of PTBP1 and U2AF2 in parental or MIR99AHG-expressing KM12C and DLD-1 cells transfected with siRNAs against PTBP1, U2AF2 or a negative control.

(B) Migration and invasion of parental or MIR99AHG-expressing KM12C (Figure 4F) and DLD-1 cells transfected with siRNAs against PTBP1, U2AF2 or a negative control. n.s., not significant.

(C) RT-qPCR analysis of PTBP1 mRNA levels after manipulation of MIR99AHG expression in the indicated CRC cells.

Figure S8

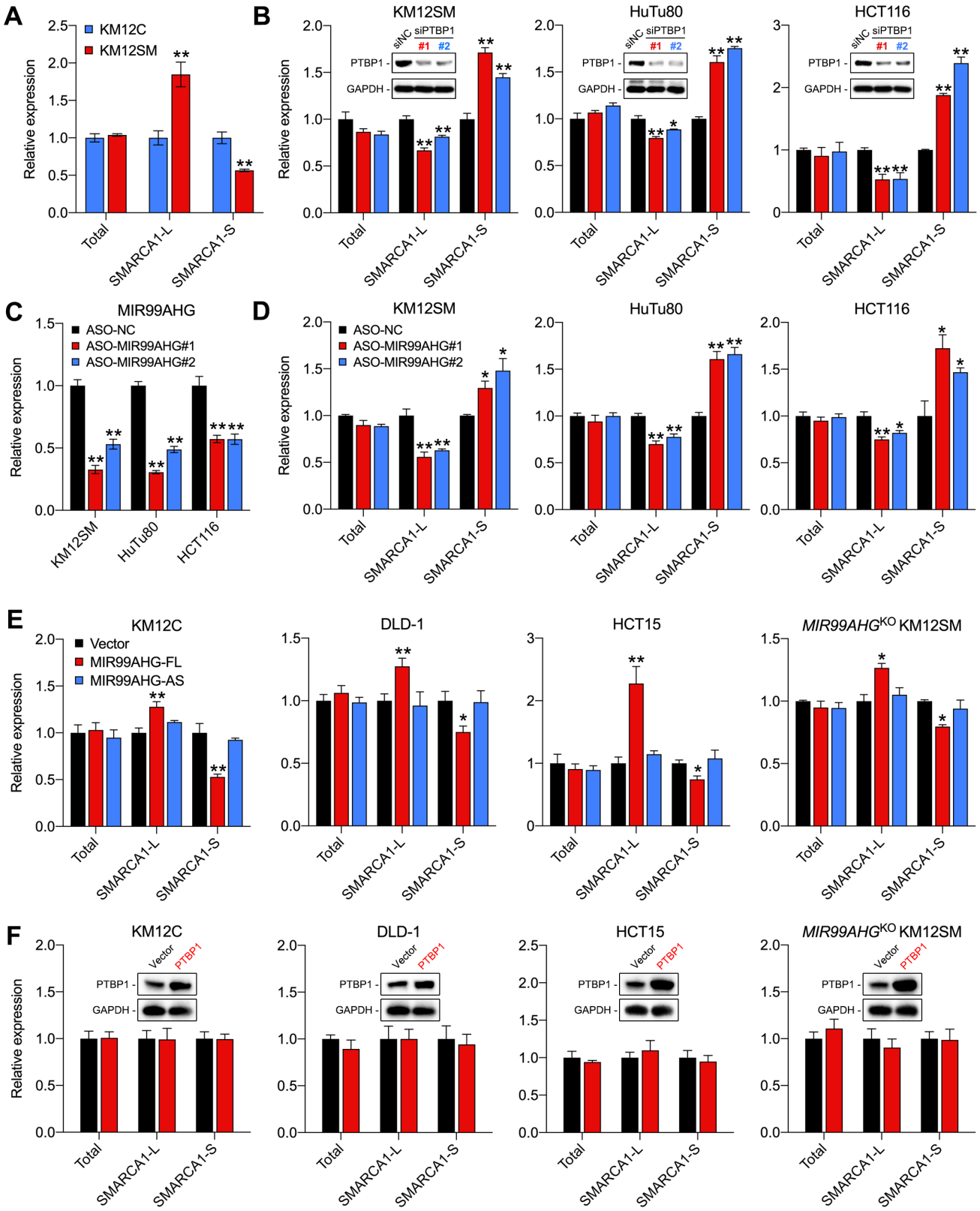
A
KM12C
KM12SM
Relative expression
Total
SMARCA1-L
SMARCA1-S
B
KM12SM
HuTu80
HCT116
siNC
siPTBP1
#1
#2
PTBP1
GAPDH
C
MIR99AHG
ASO-NC
ASO-MIR99AHG#1
ASO-MIR99AHG#2
D
E
KM12C
Vector
MIR99AHG-FL
MIR99AHG-AS
DLD-1
HCT15
MIR99AHG^KO KM12SM
F
Vector
PTBP1

Figure S8. MIR99AHG and PTBP1 regulate the AS of SMARCA1 pre-mRNA

(A) RT-qPCR analysis of the total and isoform expression of SMARCA1 in KM12C and KM12SM cells. ** $P < 0.01$ by Student's *t* test.

(B) RT-qPCR analysis of the total and isoform expression of SMARCA1 in PTBP1-silenced KM12SM, HuTu80 and HCT116 cells. Immunoblotting of PTBP1 shows the silencing efficiency of siRNAs against PTBP1 in the indicated cells. ** $P < 0.01$ by one-way ANOVA (Dunnett's test).

(C) RT-qPCR analysis of MIR99AHG expression in the indicated cells shows the silencing efficiency of ASOs against MIR99AHG in the indicated cells. ** $P < 0.01$ by one-way ANOVA (Dunnett's test).

(D) RT-qPCR analysis of the total and isoform expression of SMARCA1 in MIR99AHG-silenced KM12SM, HuTu80 and HCT116 cells. ** $P < 0.01$ by one-way ANOVA (Dunnett's test).

(E, F) RT-qPCR analysis of the total and isoform expression of SMARCA1 in MIR99AHG-FL or MIR99AHG-AS (E), or PTBP1 (F)-overexpressing KM12C, DLD-1, HCT15 and *MIR99AHG*^{KO} KM12SM cells. Immunoblotting of PTBP1 shows the efficacy of PTBP1 overexpression in the indicated cells. * $P< 0.05$, ** $P < 0.01$ by one-way ANOVA (Dunnett's test).

Figure S9

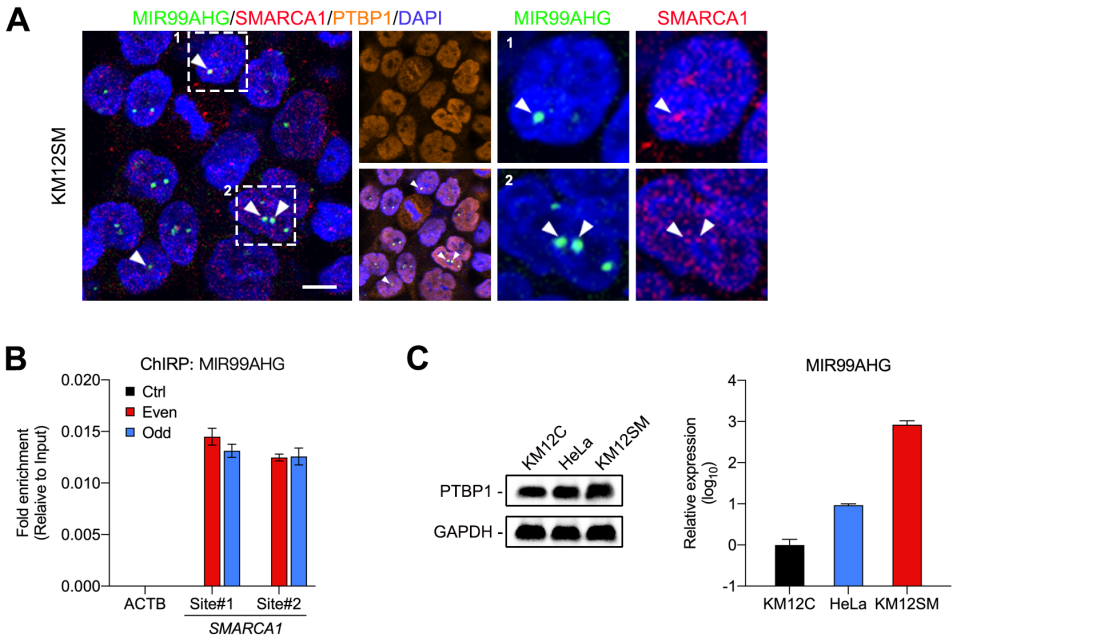
A
MIR99AHG/SMARCA1/PTBP1/DAPI
MIR99AHG
SMARCA1
KM12SM
1
2
B
ChIRP: MIR99AHG
Ctrl
Even
Odd
Fold enrichment (Relaive to Input)
0.020
0.015
0.010
0.005
0.000
ACTB
Site#1
Site#2
SMARCA1
C
KM12C
HeLa
KM12SM
PTBP1
GAPDH
MIR99AHG
Relative expression (log10)
4
3
2
1
0
-1
KM12C
HeLa
KM12SM
D
Relative enrichment (RNase/RNase inhibitor)
2.0
1.5
1.0
0.5
0.0
6U
9U
12U
15U
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
E
Relative enrichment (RNase/RNase inhibitor)
1.5
1.0
0.5
0.0
WT
Δ2
Δ15
Δ12+Δ15
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Figure S9. MIR99AHG modulates the binding position of PTBP1 on SMARCA1 pre-mRNA

(A) Confocal images of MIR99AHG and SMARCA1 pre-mRNA labeled by RNAscope ISH and PTBP1 protein concurrently stained by IF in KM12SM. Scale bar, 10 μm.

(B) ChIRP-qPCR detecting *SMARCA1* loci retrieved by a specific probe for MIR99AHG or by an unspecific control probe. ACTB was used as a negative control. SMARCA1#1 and #2 indicate different primers detecting the segments around PTBP1-binding sites #2 and #15, respectively. "Even" and "odd" indicate the two groups of biotin-labeled probe sets designed for MIR99AHG.

(C) Immunoblotting of PTBP1 and RT-qPCR analysis of MIR99AHG in KM12C, HeLa and KM12SM cells.

(D) HuTu80 cells treated with different concentrations of RNase T1 followed by an RIP-based mapping assay to identify the region on SMARCA1 pre-mRNA associated with PTBP1.

(E) qPCR validation of the binding sites of PTBP1 on the WT and mutant SMARCA1 minigene constructs by RIP-based mapping assays in KM12C cells.

Figure S10

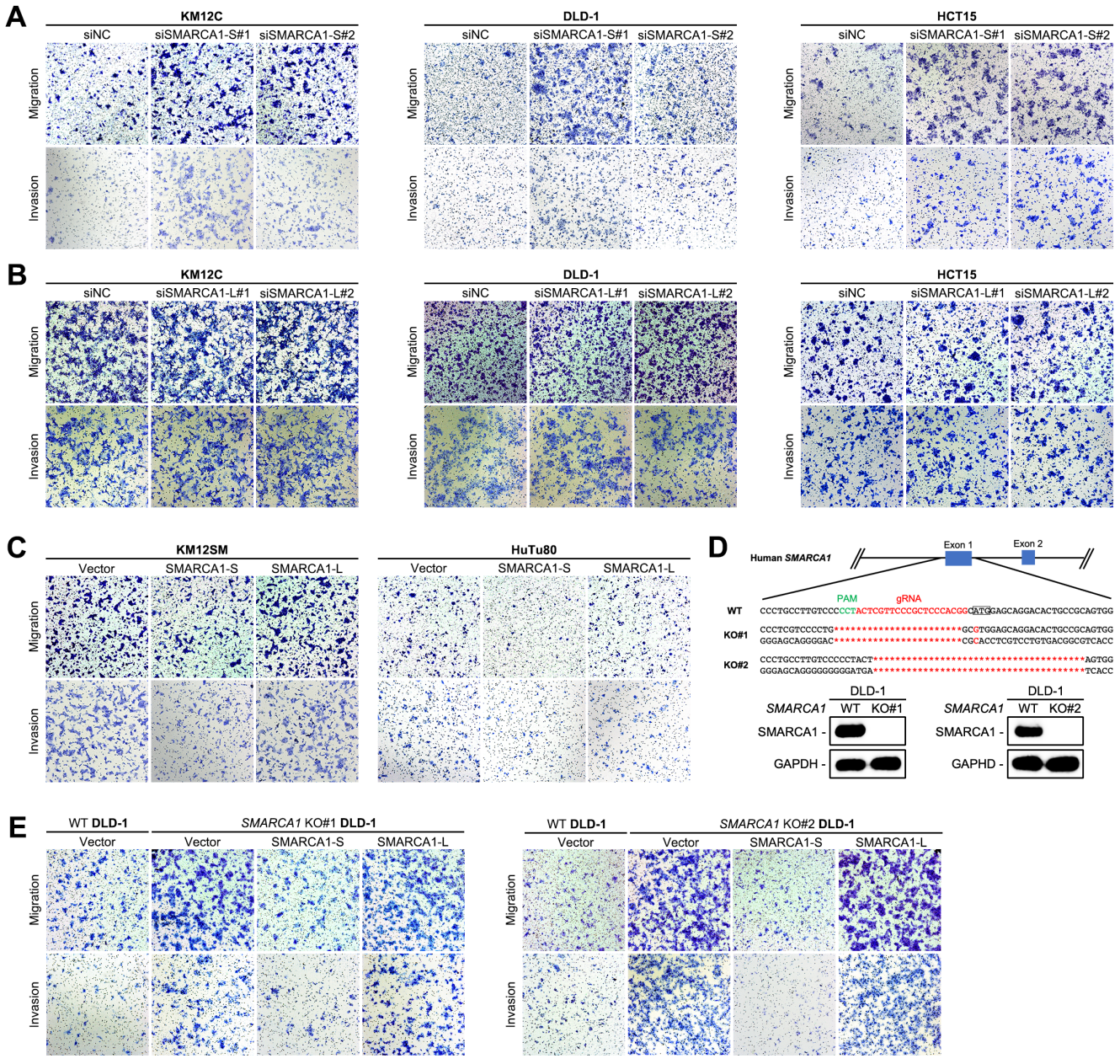
A
KM12C
DLD-1
HCT15
siNC
siSMARCA1-S#1
siSMARCA1-S#2
Migration
Invasion
B
KM12C
DLD-1
HCT15
siNC
siSMARCA1-L#1
siSMARCA1-L#2
Migration
Invasion
C
KM12SM
HuTu80
Vector
SMARCA1-S
SMARCA1-L
Migration
Invasion
D
Human SMARCA1
Exon 1
Exon 2
PAM
gRNA
WT
CCCTGCCTTGTCCCCTACTCGTTCCCGCTCCCACGGCATGGAGCAGGACACTGCCGCAGTGG
KO#1
CCCTCGTCCCCTG**********GCCTGGAGCAGGACACTGCCGCAGTGG
GGGAGCAGGGGAC**********CGCACCTCGTCCTGTGACGGCGTCACC
KO#2
CCCTGCCTTGTCCCCCTACT**********AGTGG
GGGAGCAGGGGGGGGGATGA**********TCACC
DLD-1
SMARCA1
WT
KO#1
KO#2
SMARCA1 -
GAPDH -
GAPHD -
E
WT DLD-1
SMARCA1 KO#1 DLD-1
SMARCA1 KO#2 DLD-1
Vector
SMARCA1-S
SMARCA1-L
Migration
Invasion

Figure S10. SMARCA1-S inhibits migration and invasion of CRC cells

(A, B) Representative images of Transwell migration and invasion assays in Figure 6A (A) and Figure 6B (B).

(C) Representative images of Transwell migration and invasion assays in Figure 6E.

(D) Top, *SMARCA1*^{KO} DLD-1 cells were generated by CRISPR/Cas9-mediated genome editing. Sequences of the targeted region and the two knockout alleles (KO#1 and KO#2) are shown. Bottom, immunoblotting validation of SMARCA1 expression in *SMARCA1* WT and KO DLD-1 cells.

(E) Representative images of Transwell migration and invasion assays in Figure 6G.

Figure S11

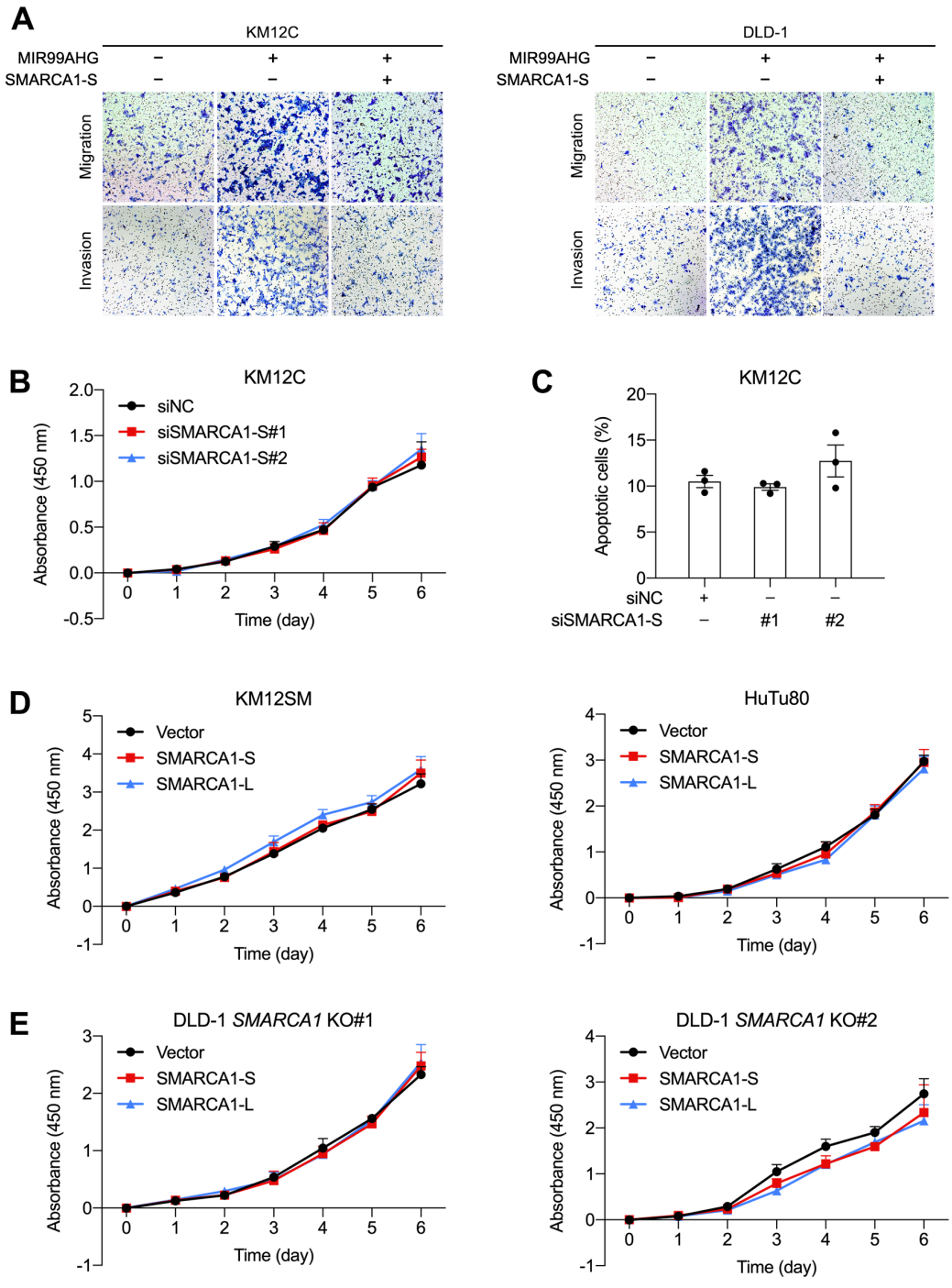
A
KM12C
DLD-1
MIR99AHG
SMARCA1-S
Migration
Invasion
B
KM12C
siNC
siSMARCA1-S#1
siSMARCA1-S#2
Absorbance (450 nm)
Time (day)
C
KM12C
Apoptotic cells (%)
siNC
siSMARCA1-S
#1
#2
D
KM12SM
HuTu80
Vector
SMARCA1-S
SMARCA1-L
Absorbance (450 nm)
Time (day)
E
DLD-1 SMARCA1 KO#1
DLD-1 SMARCA1 KO#2

Figure S11. SMARCA1 has no effect on proliferation and apoptosis of CRC cells

(A) Representative images of Transwell migration and invasion assays in Figure 6H.

(B, C) Growth curves (B) and apoptosis rates (C) of KM12C cells transfected with siRNAs against SMARCA1-S.

(D, E) Growth curves of KM12SM and HuTu80 (C) or *MIR99AHG*^{KO} DLD-1 (D) cells transfected with vectors expressing SMARCA1-S or SMARCA1-L.

Figure S12

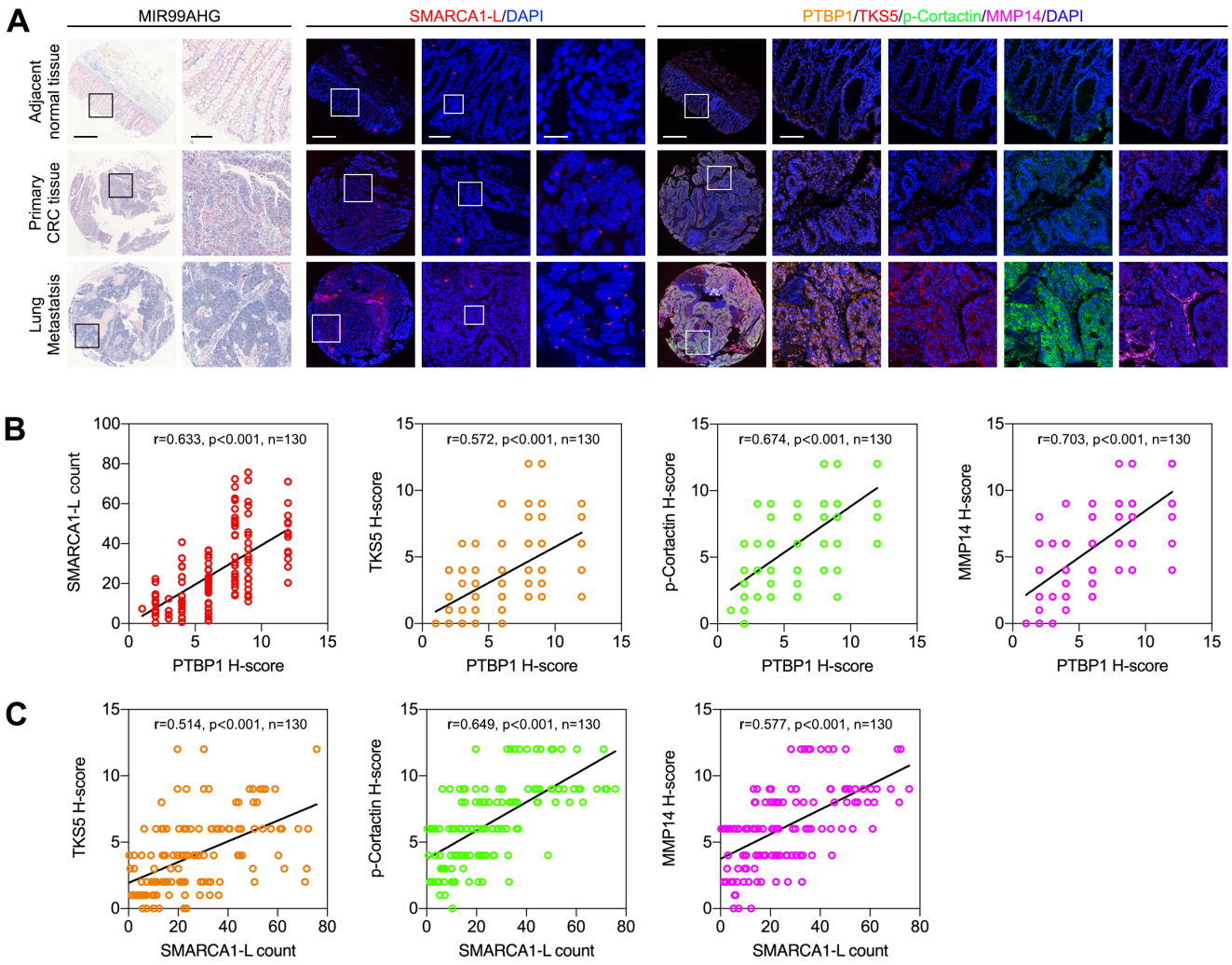
A
MIR99AHG
SMARCA1-L/DAPI
PTBP1/TKS5/p-Cortactin/MMP14/DAPI
Adjacent normal tissue
Primary CRC tissue
Lung Metastatsis
B
r=0.633, p<0.001, n=130
SMARCA1-L count
PTBP1 H-score
r=0.572, p<0.001, n=130
TKS5 H-score
PTBP1 H-score
r=0.674, p<0.001, n=130
p-Cortactin H-score
PTBP1 H-score
r=0.703, p<0.001, n=130
MMP14 H-score
PTBP1 H-score
C
r=0.514, p<0.001, n=130
TKS5 H-score
SMARCA1-L count
r=0.649, p<0.001, n=130
p-Cortactin H-score
SMARCA1-L count
r=0.577, p<0.001, n=130
MMP14 H-score
SMARCA1-L count

Figure S12. Validation of the MIR99AHG/PTBP1/SMARCA1-L axis in CRC specimens

(A) Left, representative images of ISH staining for MIR99AHG in TMAs containing 48 pairs of primary CRC tissues with their adjacent normal tissues and matched metastatic tissues. Scale bars: main, 500 μm; inset, 100 μm. Middle, confocal images of RNAscope ISH for SMARCA1-L in serial sections of TMAs. Scale bars (from left to right): 500 μm, 100 μm, 20 μm. Right, multispectral staining and imaging for PTBP1, TKS5, p-Cortactin and MMP14 in serial sections of TMAs. Scale bars: main, 500 μm; inset, 100 μm.

(B) Associations between PTBP1 levels and SMARCA1-L, TKS5, p-Cortactin or MMP14 levels in CRC tissues (n=130). Pearson correlation coefficients (r) and P values are shown.

(C) Associations between SMARCA1-L levels and TKS5, p-Cortactin or MMP14 levels in CRC tissues (n=130). Pearson correlation coefficients (r) and P values are shown.

Supplementary Table 1. Cassette exon events commonly regulated by MIR99AHG and PTBP1 in KM12SM cells.

Gene Symbol	Exon Start (ES)[1]	Exon End (EE)	Upstream ES	Upstream EE	Downstream ES	Downstream EE	MIR99AHG[2]		PTBP1[3]	
							PSI	p	PSI	p
TRPT1	64225785	64225869	64224800	64225580	64226049	64226148	-0.37	0.000	-0.44	0.000
C9orf147	112484929	112485163	112484528	112484624	112486389	112486505	-0.75	0.002	-1.00	0.001
BCS1L	218660036	218660245	218659734	218659743	218660938	218661307	-0.81	0.000	-0.5	0.001
SMARCA1	129493039	129493075	129491940	129492093	129496749	129496871	-0.66	0.002	-0.94	0.000
U2AF1L4	35744625	35744692	35744322	35744421	35745124	35745157	-0.42	0.002	-0.38	0.009
SLMAP	57865241	57865292	57864547	57864716	57871635	57871698	-0.33	0.007	-0.72	0.003
LYRM1	20902485	20902717	20900449	20900535	20915555	20915714	-0.70	0.000	-0.63	0.000
FRMD8	65393572	65393674	65389360	65389528	65394040	65394099	-0.35	0.000	-0.36	0.000
ZNF362	33270486	33270612	33256572	33256654	33276099	33276163	-0.79	0.004	-1.00	0.001
GBAP1	155217518	155217593	155217239	155217373	155217963	155218049	-0.81	0.002	-0.79	0.000
CAPS2	75390321	75390381	75359883	75361019	75390837	75390901	-0.77	0.005	-1.00	0.001
FANCI	89245194	89245311	89244016	89244067	89247628	89247731	-0.34	0.000	-0.98	0.000
CBWD5	65708901	65708958	65702969	65703014	65720040	65720092	0.85	0.000	0.74	0.000
PPARG	12301520	12301849	12287982	12288020	12312379	12312453	0.98	0.000	0.85	0.000
MX1	41449136	41449295	41445999	41446141	41451166	41451243	0.38	0.000	0.55	0.000
HCG18	30314268	30314482	30296131	30296237	30326008	30326334	0.62	0.000	0.63	0.000
ITGB1BP1	9418625	9418732	9412268	9412405	9423372	9423430	0.89	0.000	0.32	0.000
HNRNPH1	179617059	179617110	179616868	179616958	179617513	179617581	0.55	0.000	0.35	0.000
DUSP16	12554535	12554659	12521080	12521463	12560999	12561077	0.62	0.000	0.81	0.000
C6orf99	158873527	158873621	158870139	158870310	158895220	158895366	0.61	0.000	0.82	0.000

1. Data regarding the start and end positions of the exon and the upstream and downstream exons were collected from the UCSC Genome Browser (https://genome.ucsc.edu).
2. ASO-NC versus ASO-MIR99AHG.
3. siNC versus siPTBP1.

Supplementary Table 2. Clinical information of patients providing primary CRC tissues and matched adjacent normal tissues or metastatic tissues for this study.

No.[1]	Sex	Age	Primary site	Metastatic site	Differentiation
1	F	53	right colon	liver	G2
2	F	-	left colon	liver	G2
3	M	61	left colon	liver	G2-3
4	F	66	right colon	liver	G2-3
5	M	60	left colon	liver	G2
6	M	73	left colon	liver	G2
7	M	61	right colon	liver	G2
8	M	47	right colon	abdominal wall	G2
9	M	46	rectum	liver	G2
10	M	67	rectum	liver	G2
11	M	-	rectum	liver	G2
12	M	50	rectum	chest wall	G2
13	M	59	rectum	liver	G3
14	M	59	rectum	liver	G2
15	M	50	left colon	lung	G2
16	M	-	right colon	lung	G2
17	M	71	left colon	mesothelium	G2
18	-	-	left colon	anal canal	G2
19	F	50	left colon	ureter	G2
20	F	55	left colon	abdominal wall	G3
21	M	73	left colon	abdominal wall	G3
22	M	47	left colon	abdominal wall	G3
23	M	63	right colon	abdominal wall	G2
24	M	63	rectum	liver	G2-3
25	F	53	rectum	liver	G2
26	M	62	rectum	liver	G2
27	F	66	rectum	liver	G2
28	M	60	rectum	liver	G2
29	M	50	left colon	lymph node	G3
30	F	54	left colon	lymph node	G2
31	F	50	right colon	lymph node	G1-2

32	M	-	left colon	lymph node	G1-2
33	M	58	left colon	lymph node	G2-3
34	M	48	left colon	lymph node	G3
35	F	75	right colon	lymph node	G2
36	F	54	right colon	lymph node	G1-2
37	M	45	left colon	lymph node	G2
38	F	57	right colon	lymph node	G2
39	F	43	right colon	lymph node	G3
40	M	70	left colon	lymph node	G2-3
41	M	51	right colon	lymph node	G3
42	M	56	left colon	lymph node	G2-3
43	M	57	right colon	lymph node	G2-3
44	F	68	left colon	lymph node	G2
45	F	37	right colon	lymph node	G3
46	F	80	right colon	lymph node	G3
47	M	52	left colon	lymph node	G3
48	M	62	left colon	lymph node	G2-3

1. Only Patients 14-28 provided primary CRC tissues and matched metastatic tissues; the rest of the patients provided primary CRC tissues, matched adjacent normal tissues and metastatic tissues.

Supplementary Table 3. Oligonucleotides used in this study

Oligonucleotide	Sequence (5'-3')
qPCR primers for MIR99AHG	F: CTATTAGCACATAATTCAGATACATTAG
	R: AATTCATTGACAACAGAGAAAACTCAAC
qPCR primers for GAPDH	F: GCACCGTCAAGGCTGAGAAC
	R: TGGTGAAGACGCCAGTGGA
qPCR primers for PTBP1	F: GAGATGAACACGGAGGAG
	R: TGGAGAACTGGATGTAGATG
qPCR primers for 7SK	F: ATCGCCAGGGTTGATTCGGCTGAT
	R: GGATGTGTCTGGAGTCTTGGAAGC
qPCR primers #1 for MIR99AHG#5	F: TCGCAACACTTCACAAGGGTCAAGC
	R: ACTTTACCCTTCTGTCGGCCCATCC
qPCR primers #2 for MIR99AHG#5	F: TACACTCTAAGACCAGGGATGGGCC
	R: GCACAACTGGGCTGGAACAGTATTG
qPCR primers #3 for MIR99AHG#5	F: GTGCAAAACTGGTGAAATAAAAGGC
	R: TCATTAATTTCTCTTAGATGGATGCCC
qPCR primers #4 for MIR99AHG#5	F: ACAACAGTTTGGGTAACTGGATTCC
	R: TGGAGGAGGTTTGAGAGTAAATGGT
qPCR primers #5 for MIR99AHG#5	F: CCCCTCCATCTATAACGCTTTCTGC
	R: GAACAGAGAGCCCTCAAGAAGCAGA
qPCR primers #6 for MIR99AHG#5	F: ATCTGTGGAACACCAGCCAAGTC
	R: ATGTGGAATAATTCATTGACAACAGAG
qPCR primers #7 for MIR99AHG#5	F: CCTCCCCGTTGTCCAAATATTACTG
	R: CCAGGGAGTTAGGAATGAGCGAATG
qPCR primers #8 for MIR99AHG#5	F: TAACATCATTCGCTCATTCCTAACTCC
	R: GAGAAACAATAAAGGACAAATAACCCAG
qPCR primers #9 for MIR99AHG#5	F: TCATTGATGAAACTGGGTTATTTGTCC
	R: CCTAGCTCAAAGCTTACAGAAAATACAGG
PCR primers for SMARCA1-Total	F: TTTGGAAGATTATTGCATGTGGC
	R: TTGTAGATCAACCTGTGGGTTCC
PCR primers for SMARCA1-L	F: AGAAGTGGAATTTCTGGGTCAAAGG
	R: TTGTAGATCAACCTGTGGGTTCC
PCR primers for SMARCA1-S	F: CATGAAGAAAGAGAGGAAGCAATAG
	R: TTGTAGATCAACCTGTGGGTTCC

qPCR primers for MALAT1	F: GCTGACCCAGGTGCTACACAGAAGT R: AAATCGCACTGGCTCCTGGACTC
qPCR primers for U2 snRNA	F: CGGCCTTTTGGCTAAGATCAAGT R: GCTCCTATTCCATCTCCCTGCTC
qPCR primers for KCNQ1OT1	F: CCGCAGATGACAACCACTTCCTTAG R: TAGGTTGTGGGACTGGAAGTGCTTG
qPCR primers #1 for SMARCA1	F: TTTTGCCTTTTCTTTAGGTGGGTGC
(from intron 12 to intron 13)	R: AATCAATGGAATGGAAGCCACACTG
qPCR primers #2 for SMARCA1	F: AGTGTGGCTTCCATTCCATTGATTC
(from intron 12 to intron 13)	R: AATAAGGATGGGATGGTGTAGTGCC
qPCR primers #3 for SMARCA1	F: CATTTTCTTTTCCTAATTTGCCAGC
(from intron 12 to intron 13)	R: TGCTTCATGAGTCTTTAGGGTTGCT
qPCR primers #4 for SMARCA1	F: AACATTGGGACACTAAAAGGTTGGC
(from intron 12 to intron 13)	R: AACGCATTAGACCCATGACATTTGT
qPCR primers #5 for SMARCA1	F: CTCACAAATGTCATGGGTCTAATGC
(from intron 12 to intron 13)	R: GCTCTGAAGGATCACAGAAAACAGA
qPCR primers #6 for SMARCA1	F: GCAAGTCCTCACTTAATGTCGTTGC
(from intron 12 to intron 13)	R: TCAGCCTCCTGAGTAGCTGGGATTA
qPCR primers #7 for SMARCA1	F: TAATCCCAGCTACTCAGGAGGCTGA
(from intron 12 to intron 13)	R: CAGAGTCTCATTATGTCACCCAGGC
qPCR primers #8 for SMARCA1	F: CAGCCTGGGTGACATAATGAGACTC
(from intron 12 to intron 13)	R: TGGTATTACAGGCATGAGCCAACAC
qPCR primers #9 for SMARCA1	F: AGGAGGATTGCTTGAGGCTAGTCAG
(from intron 12 to intron 13)	R: TCCAGAGGAGCAAGGACTACAGGTG
qPCR primers #10 for SMARCA1	F: AGGAGTTTGAGGCTGCAGTGAGATG
(from intron 12 to intron 13)	R: GACCAGAAATATGCCACAGGAACTG
qPCR primers #11 for SMARCA1	F: GTTCCTGTGGCATATTTCTGGTCAC
(from intron 12 to intron 13)	R: CTCCAACCACTGTAATCCTGAACTG
qPCR primers #12 for SMARCA1	F: GAACTAGCCCTGGACAGAATGCC
(from intron 12 to intron 13)	R: TGTCTGTATGAGTTTTCTCTGTGCCC
qPCR primers #13 for SMARCA1	F: ATGATGTTACTTGAGGACCTGCTGT
(from intron 12 to intron 13)	R: GTGAAGTTGGTGCTAAAATTCAGGT
qPCR primers #14 for SMARCA1	F: TAGGCTTCTTTCTCTGACAATGTTG
(from intron 12 to intron 13)	R: GAGAGAGTCCTTGTAGCCTGAGATG

qPCR primers #15 for SMARCA1	F: CTACAAGGACTCTCTCTTAGCCCGC
(from intron 12 to intron 13)	R: GCTTGCTGTGAACACTATGCTATGC
qPCR primers #16 for SMARCA1	F: GCTACCTCTATGGGGAAACAGTTGC
(from intron 12 to intron 13)	R: CTCAGTAACGCTTCAAGTTGCCATC
qPCR primers #17 for SMARCA1	F: GAATATCATGGCAGAGCAGTACGTG
(from intron 12 to intron 13)	R: TCTCTGAGGGGATAACATTGTTTGG
qPCR primers #18 for SMARCA1	F: CCTGCTTTCCCTTTATGTTGTCCAC
(from intron 12 to intron 13)	R: GACGATGACCAGTGGCTTTGTTAGC
qPCR primers #19 for SMARCA1	F: CAGCAGCGCGGTAAAGTATTGGTAT
(from intron 12 to intron 13)	R: TAGCAAAAACTTGGGCATGAATCAC
qPCR primers #20 for SMARCA1	F: CTTTACCTGCTCTGCTGACTGTGAC
(from intron 12 to intron 13)	R: GCATAATCACAAAGTATGGGCAATC
qPCR primers #21 for SMARCA1	F: AATATATGACTGCTGAAGGAAGTTTTG
(from intron 12 to intron 13)	R: GTACATCCATAGGGTGAAATATTATCC
qPCR primers #22 for SMARCA1	F: TTCATCCCCCTATTCTTGATCGTTT
(from intron 12 to intron 13)	R: CCTAAGGAAATAATCAGAGATGTGCAC
qPCR primers #23 for SMARCA1	F: AGAAGTGGAATTTCTGGGTCAAAGG
(from intron 12 to intron 13)	R: CATTGCTGCTGGGAGTATAAATTGG
qPCR primers #24 for SMARCA1	F: CCCCCTCATCAACATTGAGTATTAT
(from intron 12 to intron 13)	R: ACAAAGTCCAGTAGAAAAATGAGCAG
qPCR primers #25 for SMARCA1	F: TCCCTGGATTGTCATCCTGATTTTC
(from intron 12 to intron 13)	R: GGCAACGGATAAAGGGGAAAGTAAG
qPCR primers #26 for SMARCA1	F: CCTCCTAGATATATTTCAGCTCTCACTC
(from intron 12 to intron 13)	R: AAGTTCCTAATCTTTAAAAGCTTACATCG
qPCR primers #27 for SMARCA1	F: TTATGCTAAGTACCAGGGCTGGAGG
(from intron 12 to intron 13)	R: CAACCTGTGGGTTCCAGTCTGAATC
qPCR primers for ACTB	F: GCACTCTTCCAGCCTTCCTTC R: TTGGCGTACAGGTCTTTGCGGA
qPCR primers for pre-SMARCA1#1	F: TATACTATATGTAAAGTACAGTATC R: GAAATTCCACTTCTAGGAATTTATC
qPCR primers for pre-SMARCA1#2	F: AGAAGTGGAATTTCTGGGTCAAAGG R: TCACAGAACAAATACAAATGGTTAG
scrambled siRNA control	UUCUCCGAACGUGUCACGUTT
siRNA#1 targeting TKS5	GCCAAAGCAAGGACGAGAUTT

siRNA#2 targeting TKS5	AAACCAGUGGCGACCUGAATT
siRNA#1 targeting Cortactin	CCAGGAGCAUAUCAACAUATT
siRNA#2 targeting Cortactin	GCAACUUAUUGUAUCUGAATT
siRNA#1 targeting PTBP1	CCCUCAUUGACCUGCACAATT
siRNA#2 targeting PTBP1	GCACAGUGUUGAAGAUCAUTT
siRNA#1 targeting U2AF2	GCAGCUCAACGAGAAUAAATT
siRNA#2 targeting U2AF2	GCAGUACAAGGCCAUGCAATT
siRNA#1 targeting SMARCA1-L	GAAAGAGAGGAUAAAUUCCTT
siRNA#2 targeting SMARCA1-L	GGGUCAAAGGGAAGCAAUATT
siRNA#1 targeting SMARCA1-S	CAUGAAGAAAGAGAGGAAGTT
siRNA#2 targeting SMARCA1-S	GAGAGGAAGCAAUAGAGGCTT
CHIRP probes for MIR99AHG#1	TTAAGTTTTCTGCATCCGTG
CHIRP probes for MIR99AHG#2	AGAAGGTCTGTTCAGTGGAA
CHIRP probes for MIR99AHG#3	CCGGGAACCATAAGAAGACA
CHIRP probes for MIR99AHG#4	ACACTCTACCTGGATATCTT
CHIRP probes for MIR99AHG#5	TTATCCTTGTGTACTCATGC
CHIRP probes for MIR99AHG#6	GCTTTATTCTGGATTTAGTT
CHIRP probes for MIR99AHG#7	AGGATTTGGGGAAGACTCTC
CHIRP probes for MIR99AHG#8	ATAGTCATGTCACAGGTAGG
CHIRP probes for MIR99AHG#9	CCTCATTCATTCTCCAAGAA
CHIRP probes for MIR99AHG#10	GATTATTCAGCCAATTGCTC
CHIRP probes for MIR99AHG#11	AACTGGATTGAGTCTGTAGC
CHIRP probes for MIR99AHG#12	TGTTGCTATCTTTACCTTGA
CHIRP probes for MIR99AHG#13	TCTCTTAAGATGCATTCTGT
CHIRP probes for MIR99AHG#14	CCTCGAGACTTGAGTGGAAA
CHIRP probes for MIR99AHG#15	TTAGATGGATGCCCATTTTG
CHIRP probes for MIR99AHG#16	ACCTTTTACAGATGTGCAGA
CHIRP probes for MIR99AHG#17	GAACAGAGAGCCCTCAAGAA
CHIRP probes for MIR99AHG#18	GAGTTAGGAATGAGCGAATG
CHIRP probes for LacZ#1	CGAGCTCGAATTCGTAATCA
CHIRP probes for LacZ#2	CAAGGCGATTAAGTTGGGTA
CHIRP probes for LacZ#3	TGCGTAAGGAGAAAATACCG

	CTCTTCCTCACCTCTCTTTCTTCATGCGGGGTT
RAP probes for SMARCA1#1	TGTCCATCCAGTCGACAATACTCATAACCACG
	CCACATGCAATAATCTTCCAAAATA
	ACCAGAAATATGCCACAGGAACTGGGGTCTTG
RAP probes for SMARCA1#2	TCCTGTCACACAGTTTGAAGTGCAGTGGCACA
	ATCACATCTCACTGCAGCCTCAAACT
	GCAAATTGGCATGTCTACATGGTCCCAGTGTG
RAP probes for SMARCA1#3	AGTGGTGTGTGTGCGAGCATGCCCTGTGATGG
	AATGGCATTCTGTCCAGGGCTAGTTC
	CTGTGTGCTTGCTGTGAACACTATGCTATGCTA
RAP probes for SMARCA1#4	GAAACAGGCTTCAATAAGGAGGCACACAAGG
	AACGGAGTGCCCATTGCAGCTACAGA
	GGTAACTACCACCGTCACAGTCAGCAGAGCAG
RAP probes for SMARCA1#5	GTAAAGGGGGCAGTATAGAAAGCAGCGTGGT
	GTAGTGAGAGGCACCTTGGCGTAGAGG
	CTTGTAGATCAACCTGTGGGTTCCAGTCTGAA
RAP probes for SMARCA1#6	TCATATAGTATAACCACATCAGCACTTGCCAG
	GTTAATTCCGAGACCTCCAGCCCTGG

Supplementary Table 4. gRNA sequences used in this study

No.	Target gene	Target sequence (5'-3')
gRNA#1	*MIR99AHG*	TCCATACCCTAATTGGCCGA
gRNA#2	*MIR99AHG*	GACTTGTCGCACGGATGCAG
gRNA#3	*MIR99AHG*	ACTTTATTGACGTTAGTGGG
gRNA#4	*MIR99AHG*	CCATCTATAACGCTTTCTGC
gRNA#5	*SMARCA1*	TACTCGTTCCCGCTCCCACG
gRNA#6	*SMARCA1*	ACTCGTTCCCGCTCCCACGG

Supplementary Table 5. Antibodies used in this study

Antibodies	Source	Catalog #
Rabbit anti-human TKS5	Cell Signaling Technology	16619
Rabbit anti-human SH3PXD2A (TKS5)	Proteintech	18976-1-AP
Mouse anti-human Cortactin	Abcam	Ab33333
Rabbit anti-human p-Cortactin (Tyr421)	Cell Signaling Technology	4569
Rabbit anti-human N-WASP	Cell Signaling Technology	4848
Rabbit anti-human MMP14	Abcam	ab51074
Rabbit anti-human MMP7	Cell Signaling Technology	3801
Rabbit anti-human TIMP1	Cell Signaling Technology	8946
Rabbit anti-human TIMP2	Cell Signaling Technology	5738
Rabbit anti-human GAPDH	Cell Signaling Technology	2118
Rabbit anti-human β-actin	Cell Signaling Technology	4970
Mouse anti-human E-Cadherin	Cell Signaling Technology	14472
Rabbit anti-human β-catenin	Cell Signaling Technology	8480
Rabbit anti-human Vimentin	Cell Signaling Technology	5741
Rabbit anti-human PTBP1	Proteintech	12582-1-AP
Rabbit anti-human U2AF2	Abcam	ab37530
Mouse anti-human Ki-67	Cell Signaling Technology	9449
Rabbit anti-Flag	Sigma-Aldrich	F7425
Rabbit anti-human SAMRCA1	Cell Signaling Technology	12483
anti-Rabbit HRP-linked IgG	Cell Signaling Technology	7074
anti-Mouse HRP-linked IgG	Cell Signaling Technology	7076
Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed, Alexa Fluor 488	Thermo Fisher Scientific	A-11034
Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed, Alexa Fluor Plus 594	Thermo Fisher Scientific	A32742
Rabbit IgG	Proteintech	B900610