**The lncRNA MIR17HG Directs Cooperative Epigenetic Regulation in Colorectal Cancer**

Jie Xu1, #, Qingtao Meng1, #, Xiaobo Li1, Hongbao Yang2, Jin Xu3, Na Gao1, Hao Sun1, Shenshen Wu1, Giuseppe Familiari4, Michela Relucenti5, Haitao Zhu6, Jiong Wu7, Rui Chen1,8 \*

1Key Laboratory of Environmental Medicine Engineering, Ministry of Education, School of Public Health, Southeast University, Nanjing 210009, China

2Center for New Drug Safety Evaluation and Research, China Pharmaceutical University, Nanjing 211198, China

3Department of Maternal, Child and Adolescent Health, School of Public Health, Nanjing Medical University, Nanjing 211166, China

4Department of Anatomical, Histological, Medical and Legal locomotive Apparatus, Section of Human Anatomy Via Alfonso Borelli, Sapienza University of Rome, Roma 5000161, Italia

5Department of Anatomical, Histological, Forensic Medicine and Orthopedic Science, Sapienza University of Rome, Roma 5000161, Italia

6Colorectal Cancer Center, Department of General Surgery, Jiangsu Cancer Hospital, Cancer Research Institute, Cancer hospital of Nanjing Medical University, Nanjing 210009, China

7Jiangsu Key Laboratory for Bioresources of Saline Soils, Jiangsu Synthetic Innovation Center for Coastal Bioagriculture, Yancheng Teachers University, Yancheng 224002, China

8Institute for Chemical Carcinogenesis, Guangzhou Medical University, Guangzhou 511436, China

**Running title:** **MIR17HG epigenetic regulation in colorectal cancer**

**Key words: long non-coding RNA; MIR17HG; colorectal cancer; NF-κB/RELA; WGCNA**

**#** These two authors contribute equally

**\*Correspondence author: Rui Chen**, Key Laboratory of Environmental Medicine Engineering, Ministry of Education, School of Public Health, Southeast University, 87 Dingjiaqiao, Gulou District, Nanjing 210009, China. Tel: +86 25 83272560; Fax: +86 25 83324322; Email: [101011816@seu.edu.cn](mailto:101011816@seu.edu.cn).

**Disclosure of Potential Conflicts of Interests：**

The authors declare no conflict of interest.

**SI Materials and Methods**

**RNA extraction and quantitative real-time PCR (qRT-PCR)**

Total RNA from cell and tissue samples was isolated using TRIzol Reagent (Invitrogen, USA) according to the manufacturer’s protocols, and reverse-transcribed and amplified with the SYBR Green PCR kit (Takara, Japan). The relative long non-coding RNA (lncRNA), microRNA (miRNA) and mRNA expression levels were normalized to cycA, U6, and GAPDH.

**lncRNA and mRNA microarray analysis**

A hybridization-based microarray analysis of lncRNA and mRNA expression levels was performed on the Agilent Array platform (Agilent Technologies, US) for 6 pairs of colorectal cancer (CRC) and adjacent tissues, and 6 adenoma tissues. Briefly, each mRNA sample was transcribed into double-stranded cDNA and hybridized to the Human LncRNA Array v3.0 chip (8 × 60 K). Differentially expressed mRNAs and lncRNAs were considered at a fold change (FC) ≥ 1.5 and a false discovery rate (FDR) < 0.05. The original expression data has been submitted to Gene Expression Omnibus (GEO, accession numbers GSE104364).

**RNA *in situ* hybridization (RNA-ISH)**

RNA-ISH was performed as previously described ([1](#_ENREF_1), [2](#_ENREF_2)). In brief, tissue microarray (TMA) slides were deparaffinized with xylene, rehydrated with an ethanol gradient (100%, 95% and 80%), and treated with 20µg/mL Proteinase K (Roche Diagnostics, USA) for 10 min at 37°C. Then, the slides were fixed with 4% formaldehyde (Thermo Fisher Scientific, USA) for 10 min, rinsed twice with 0.13M 1-methylimidazole and re-fixed with 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC; Thermo Fisher Scientific) for 1h. Next, 1% H2O2 was used to block endogenous peroxidase, and slides were prehybridized at 50°C for 30 min in hybridization buffer containing 50% formamide (American Bioanalytical, USA), 5 × SSC (American Bioanalytical), 50µg/mL Heparin (Sigma-Aldrich, USA), 0.1% Tween 20 (Sigma-Aldrich), 500µg/mL yeast tRNA (Invitrogen) (pH = 6). This was followed by hybridization with 200 nM Double Digoxigenin (DIG) LNA modified probes.

Oligonucleotide probes complementary to miR-17-5p, miR-17-3p, miR-19b-1-5p, and MIR17HG were artificially synthesized by Exonbio Lab (Guangzhou, China). Probes with complementarities to a fragment of MIR17HG were marked with a double digoxigenin (DIG) label. The sequence of the MIR17HG probe was as follow: 5ʹ-Dig- TAACTCATACAACCACTAAGCT -Dig-3ʹ. The sequence of the negative control (NC) was 5ʹ-Dig-TCTTACACCTAGATAAGCAAAGA -Dig-3ʹ.

Next, miRNA or lncRNA staining in TMA was scored independently by two pathologists blinded to clinical data using the following criteria. Category A: intensity of immunostaining scored from 0 to 3 (0, negative; 1, weak; 2, moderate; 3, strong). Category B: the percentage of immunoreactive cells was deemed as 1 (0%–25%), 2 (26%–50%), 3 (51%–75%), or 4 (76%–100%). The final score for MIR17HG staining was calculated by multiplying both subscores.

**Immunohistochemistry (IHC) staining**

After dewaxing, IHC was performed on mouse colon tissue sections and human colorectal TMA as previously described ([3](#_ENREF_3)). The sections were incubated overnight at 4°C with antibodies against RELA (1:500; Abcam, USA), BLNK (1:200; Thermo Fisher Scientific), PD-L1 (1:500; Thermo Fisher Scientific) and CD3E (1:100; Abcam). The antibody binding to tissue sections was visualized with biotinylated rabbit anti-mouse IgG antibodies (1:400; DAKO, Denmark) and developed with diaminobenzidine (DAB) as a substrate. For negative controls, primary antibodies were omitted. Each section was examined under a microscope by two histologists. A semi-quantitative immunoreactivity score (IRS) as previously reported ([4](#_ENREF_4)) was used to evaluate the expression levels of proteins in human or mouse colorectal tissues. IRS values for mouse colorectal tissues were obtained from 3 non-overlapping high-power fields (HPF) (× 400 magnification) in each tissue condition and section. The IRS of PD-L1 in tumor tissues were evaluated from 3-non-overlaping HPF (× 400 magnification) in each murine tumor section.

**Colony formation assay**

Colony formation assay was performed as follows. One hundred SW620 and 100 HCT116 cells were plated into 6-well plates and incubated in Dulbecco’s Modified Eagle’s Medium (DMEM) with 10% fetal bovine serum (FBS) at 37 °C. Two weeks later, the cells were fixed and stained with 0.1% crystal violet. The numbers of colonies (defined as > 50 cells/colony) were determined.

**Cell migration and invasion assays**

Cell migration and invasion were assessed using transwell chambers (8 μm; Corning, USA). SW620 and HCT116 cells (4 × 104) were suspended in 100 μL serum-free DMEM and seeded into uncoated (migration assay) or 1:8 diluted matrigel-coated (invasion assay) upper wells. The lower wells were filled with 600 μL DMEM containing 10% FBS. After incubation for 24 h, cells on the upper surface of the well were removed, and those on the lower surface were fixed and stained with 0.4% crystal violet. In each experiment, the number of transmigrated cells was determined in 5 random selected fields under a microscope.

**Luciferase reporter assay**

Wild type and mutant 3ʹUTR sequences of BLNK, PPP3R1, CBL, RELA, MALT1, NFKBIE, and MAP3K7 3ʹUTR were cloned by PCR and inserted into the pmiR-RB miRNA reporter vector (Guangzhou RiboBio Co. Ltd, China).

For the luciferase assay, cells were co-transfected with miRNA mimic or inhibitor as well as the indicated pmiR-RB vectors for 48 h. The relative luciferase activity was analyzed with a Dual luciferase reporter assay kit (Promega, USA) on Mithras LB 940 (Berthold Technologies, Germany). The relative Renilla luciferase activity was normalized to luciferase activity.

**Western blots**

Western blots was performed as previously reported ([5](#_ENREF_5)). For each treatment, 3 parallel samples were applied and equal amounts of proteins from the parallel samples were mixed and used for blots. Immunoblot assays were used to detect the PD-L1 protein level by using primary antibodies PD-L1 (1:5000; Proteintech, USA) and β-actin (1:10000; sigma). Semi-quantitative analysis of PD-L1 protein levels was performed with Image Lab 3.0 software (BIO-RAD, USA).

# References

1. Hanna JA, Wimberly H, Kumar S, Slack F, Agarwal S, Rimm DL. Quantitative analysis of microRNAs in tissue microarrays by in situ hybridization. BioTechniques. 2012;52:235-45.

2. Pena JT, Sohn-Lee C, Rouhanifard SH, Ludwig J, Hafner M, Mihailovic A, et al. miRNA in situ hybridization in formaldehyde and EDC-fixed tissues. Nature methods. 2009;6:139-41.

3. Li X, Zhang C, Zhang X, Wang S, Meng Q, Wu S, et al. An acetyl-L-carnitine switch on mitochondrial dysfunction and rescue in the metabolomics study on aluminum oxide nanoparticles. Particle and fibre toxicology. 2016;13:4.

4. Wang S, Wu X, Chen Y, Zhang J, Ding J, Zhou Y, et al. Prognostic and predictive role of JWA and XRCC1 expressions in gastric cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2012;18:2987-96.

5. Wu S, Wang S, Fu Y, Tang W, Jin H, Meng Q, et al. A novel mechanism of rs763110 polymorphism contributing to cervical cancer risk by affecting the binding affinity of C/EBPbeta and OCT1 complex to chromatin. International journal of cancer. 2017;140:756-63.

6. Mudter J, Amoussina L, Schenk M, Yu J, Brustle A, Weigmann B, et al. The transcription factor IFN regulatory factor-4 controls experimental colitis in mice via T cell-derived IL-6. The Journal of clinical investigation. 2008;118:2415-26.

Supplementary Table S1. Demographic characteristics of CRC patients recruited for microarray assay

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **Gender** | **Age** | **Diagnosis** | **Grade** | **T** | **N** | **M** | **TNM** |
| 1 | Male | 68 | CRC | 1 | 2 | 0 | 0 | 1 |
| 2 | Male | 73 | CRC | 1 | 3 | 0 | 0 | 2 |
| 3 | Female | 68 | CRC | 1 | 2 | 0 | 0 | 1 |
| 4 | Male | 62 | CRC | 2 | 1 | 0 | 0 | 1 |
| 5 | Female | 60 | CRC | 0 | 4 | 0 | 0 | 2 |
| 6 | Female | 59 | CRC | 2 | 3 | 0 | 0 | 2 |
| 7 | Female | 64 | ADE | -- | -- | -- | -- | -- |
| 8 | Male | 60 | ADE | -- | -- | -- | -- | -- |
| 9 | Male | 54 | ADE | -- | -- | -- | -- | -- |
| 10 | Male | 59 | ADE | -- | -- | -- | -- | -- |
| 11 | Male | 65 | ADE | -- | -- | -- | -- | -- |
| 12 | Female | 52 | ADE | -- | -- | -- | -- | -- |

Supplementary Table S2. Distributions of demographic and clinicopathologic characteristics of patients recruited for TMA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables** | **Testing cohort**  **(N=376)** | | **Validation cohort (N=431)** | | ***P*** |
| **n** | **%** | **n** | **%** |  |
| Age |  |  |  |  | 0.3395 |
| ≤55 | 195 | 51.9 | 209 | 48.5 |  |
| ＞55 | 181 | 48.1 | 222 | 51.5 |  |
| Gender |  |  |  |  | 0.5021 |
| Male | 226 | 60.1 | 269 | 62.4 |  |
| Female | 150 | 39.9 | 162 | 37.6 |  |
| Location |  |  |  |  | 0.4546 |
| Colon | 162 | 43.1 | 197 | 45.7 |  |
| Rectum | 214 | 56.9 | 234 | 54.3 |  |
| Grade |  |  |  |  | 0.9996 |
| Low | 123 | 32.7 | 141 | 32.7 |  |
| Intermediate/High | 253 | 67.3 | 290 | 67.3 |  |
| Depth of invasion |  |  |  |  | 0.4144 |
| T1 | 14 | 3.7 | 14 | 3.3 |  |
| T2 | 74 | 19.7 | 106 | 24.6 |  |
| T3 | 69 | 18.4 | 76 | 17.6 |  |
| T4 | 219 | 58.2 | 235 | 54.5 |  |
| Lymph node metastasis | |  |  |  | 0.0733 |
| N0 | 198 | 52.7 | 254 | 58.9 |  |
| N1 | 178 | 47.3 | 177 | 41.1 |  |
| Distant metastasis |  |  |  |  | 0.9114 |
| M0 | 329 | 87.5 | 376 | 87.2 |  |
| M1 | 47 | 12.5 | 55 | 12.8 |  |
| TNM |  |  |  |  | 0.1438 |
| I | 55 | 14.6 | 87 | 20.2 |  |
| II | 136 | 36.2 | 155 | 36 |  |
| III | 138 | 36.7 | 134 | 31.1 |  |
| IV | 47 | 12.5 | 55 | 12.8 |  |

Supplementary Table S3. Demographic information of colorectal adenoma (ADE) patients for PCR analysis

|  |  |  |
| --- | --- | --- |
| **Variables** | **Neoplasia** | |
| **n** | **%** |
| Age |  |  |
| ≤59 | 48 | 50 |
| ＞59 | 48 | 50 |
| Gender |  |  |
| Male | 47 | 49 |
| Female | 49 | 51 |
| Location |  |  |
| Left Colon | 74 | 77.1 |
| Right Colon | 14 | 14.6 |
| Rectum | 4 | 8.3 |
| Histologic types |  |  |
| Tubular | 69 | 71.9 |
| Villous | 5 | 5.2 |
| Tubulovillous | 22 | 22.9 |
| Dysplasia |  |  |
| LGIN | 83 | 86.5 |
| HGIN | 13 | 13.5 |
| LGIN means low-grade intraepithelial neoplasia, HGIN means high-grade intraepithelial neoplasia | | |

Supplementary Table S4. Demographic information of colorectal cancer (CRC) patients for PCR analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables** | | **Cancer** | | |
| **n** | **%** | |
| Age |  | |  |
| ≤58 | 54 | | 56.3 |
| ＞58 | 42 | | 43.8 |
| Gender |  | |  |
| Male | 49 | | 51 |
| Female | 47 | | 49 |
| Location |  | |  |
| Colon | 46 | | 47.9 |
| Rectum | 50 | | 52.1 |
| Grade |  | |  |
| Low | 28 | | 29.2 |
| Intermediate/High | 68 | | 70.8 |
| Depth of invasion |  | |  |
| T1 | 2 | | 2.1 |
| T2 | 20 | | 20.8 |
| T3 | 20 | | 20.8 |
| T4 | 54 | | 56.3 |
| Lymph node metastasis |  | |  |
| N0 | 49 | | 51 |
| N1 | 47 | | 49 |
| Distant metastasis |  | |  |
| M0 | 90 | | 93.8 |
| M1 | 6 | | 6.3 |
| TNM |  | |  |
| I | 18 | | 18.8 |
| II | 30 | | 31.3 |
| III | 42 | | 43.8 |
| IV | 6 | | 6.3 |

# Supplementary Figure Legends

**Supplementary Fig. S1 Construction of an immunity-associated network involved in CRC progression.**

(A) Heat map of lncRNA and mRNA expression levels in M1, M14, and M15. (B) The enriched KEGG pathways of mRNAs in M1 and M14 (*P* < 0.05). Immune-related pathways were identified in M1 and M14. (C) The expression levels of LINC00460 and XLOC\_006495 in adjacent, adenoma, and CRC tissues were analyzed by qRT-PCR (n = 96 for each group, \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001, one-way ANOVA). (D) The expression levels of RELA retrieved from TCGA were analyzed (normal colorectal cases, n = 51; CRC cases, n = 286, \*\**P* < 0.01, two-tailed t-test following log transformation). (E) Co-expression network adjusted after qRT-PCR data analysis. (F) Left panel: CRC cell lines were transfected with RELA siRNA or the negative control siRNA. The expression levels of LINC00460 and XLOC\_006495 in the indicated cells were determined by qRT-PCR. Right panel: CRC cell lines were transfected with LINC00460, XLOC\_006495 silencer or the negative control. The expression levels of RELA in the indicated cells were determined by qRT-PCR (n = 6, \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001, two-tailed t-test).

**Supplementary Fig. S2 Roles of RELA and MIR17HG in murine AOM-DSS models and CRC cohorts.**

(A) Representative H&E staining images of distal colonic sections and calculated colitis scores (n = 6 per group, compared with RELAfl/fl mice, Mann Whitney test). The scale bar is marked in each image. (B) Correlation between MIR17HG and RELA expression levels in CRC tissues (n = 96, Fisher r-to-z transformation test followed by the Pearson correlation test). (C) Upper panel: The levels of RELA in TMA were evaluated by IHC, and the staining score differences between CRC lesions and adjacent tissues (cancer - adjacent) in the testing (n = 363), validation (n = 402) and combined (n = 765) cohorts are shown. Lower panel: The levels of MIR17HG in TMA were evaluated by ISH, and the staining score differences between CRC lesions and adjacent tissues (cancer - adjacent) in the testing (n = 309), validation (n = 386) and combined (n = 695) cohorts are shown. (D) KM curves depicting overall survival in CRC patients according to tumoral RELA and MIR17HG levels respectively. Left to right: testing and validation cohorts. *P* values were calculated by the log-rank test.

**Supplementary Fig. S3 miR-17-5p promotes CRC tumorigenesis.**

(A) Expression levels of miR-17-5p, miR-17-3p and miR-19b-1-5p retrieved from TCGA were analyzed (normal colorectal cases, n = 14; CRC cases, n = 270, \**P* < 0.05, \*\**P* < 0.01, two-tailed t-test following log transformation). (B-C) Upper panel: SW620 and HCT116 cells with the indicated treatments were subjected to colony formation assay (n = 6, \*\*\**P* < 0.001, compared with NC, two-tailed t-test). Lower panel: SW620 and HCT116 cells with the indicated treatments were subjected to migration and invasion assays (n = 3, \**P* < 0.05, \*\*\**P* < 0.001, compared with NC, two-tailed t-test). (D-E) Upper panel: SW620 and HCT116 cells with the indicated treatments were subcutaneously injected into nude mice respectively (n = 6). The representative images of xenografts are shown. Lower panel: Subcutaneous tumors, liver, and lung tissues of the indicated mice were harvested, and luciferase activities were analyzed (n = 6, \**P* < 0.05, \*\**P* < 0.01, compared with NC, two-tailed t-test). (F) Upper panel: The levels of miR-17-5p in TMA evaluated by ISH and representative staining images are shown. The scale bar is marked in each image. Lower panel: The staining score differences between CRC lesions and adjacent tissues (cancer - adjacent) in the testing (n = 342), validation (n = 397) and combined (n = 739) cohorts are shown. (G) KM curves depicting overall survival in CRC patients according to tumoral miR-17-5p levels. Left to right: testing, validation and combined cohorts. *P* values were calculated by the log-rank test.

**Supplementary Fig. S4 MIR17HG promotes CRC tumorigenesis through suppressing BLNK.**

(A) The expression levels of BLNK retrieved from TCGA were analyzed (normal colorectal cases, n = 51; CRC cases, n = 286, \*\*\**P* < 0.001, two-tailed t-test following log transformation). (B) The expression levels of BLNK in cells with the indicated treatments were determined by qRT-PCR (n = 6, \*\*\**P* < 0.001, compared with NC-treated cells within each group, two-tailed t-test). (C) Left panel: Correlation of BLNK and miR-17-5p expression levels in CRC tissues. Right panel: Correlation of BLNK and MIR17HG expression levels in CRC tissues (n = 96, Fisher r-to-z transformation test followed by the Pearson correlation test). (D) Upper panel: HCT116 cells with the indicated treatments were subjected to colony formation assay (n = 6, \*\**P* < 0.01, compared with NC, two-tailed t-test). Lower panel: HCT116 cells with the indicated treatments were subjected to migration and invasion assays (n = 3, \*\*\**P* < 0.001, compared with NC, two-tailed t-test). (E) Upper panel: HCT116 cells with the indicated treatments were subcutaneously injected into nude mice (n = 6). Lower panel: Subcutaneous tumors, liver, and lung tissues of the indicated mice were harvested, and luciferase activities were analyzed (n = 6, \**P* < 0.05, \*\**P* < 0.01, compared with NC, two-tailed t-test). The representative images are shown. (F) Levels of BLNK in TMA evaluated by IHC and staining score differences between CRC lesions and adjacent tissues (cancer - adjacent) in the testing (n = 349), validation (n = 405), and combined (n = 754) cohorts. (G) KM curves depicting overall survival in CRC patients according to tumoral BLNK levels. Left to right: testing and validation cohorts. *P* values were calculated by the log-rank test.

**Supplementary Fig. S5 MIR17HG promotes CRC tumorigenesis through binding to miR-17-3p *in vitro* and *in vivo*.**

(A-B) ISH was applied to detect miR-17-3p and miR-19b-1-5p expression levels in colorectal tissues from the indicated mice. The representative images and calculated scores are shown (n = 9, \*\**P* < 0.01, \*\*\**P* < 0.001, compared with adjacent tissue in RELAfl/fl mice; &&*P* < 0.01, &&&*P* < 0.001, compared with adjacent tissues of RELA-/-mice, Kruskal-Wallis test followed by the Dunn's multiple comparison test). The scale bar is marked in each image.(C) KM curves depicting overall survival in CRC patients according to tumoral MIR17HG and miR-17-5p levels, respectively. Left to right: testing, validation, and combined cohorts. *P* values were calculated by the log-rank test.

**Supplementary Fig. S6 Regulation of MALT1, NFKBIE, PPP3R1 and MAP3K7 by miR-17-3p in CRC.**

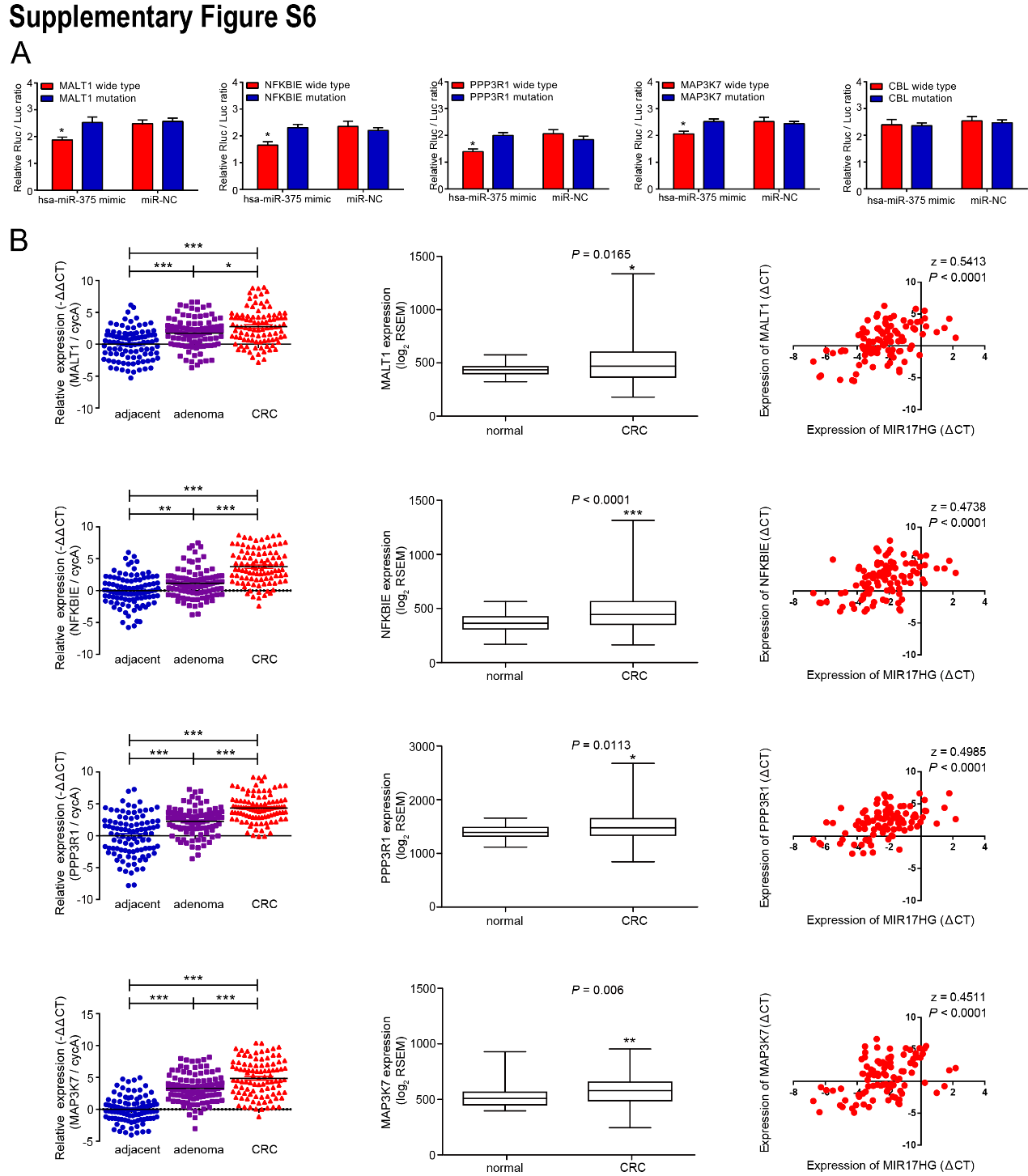
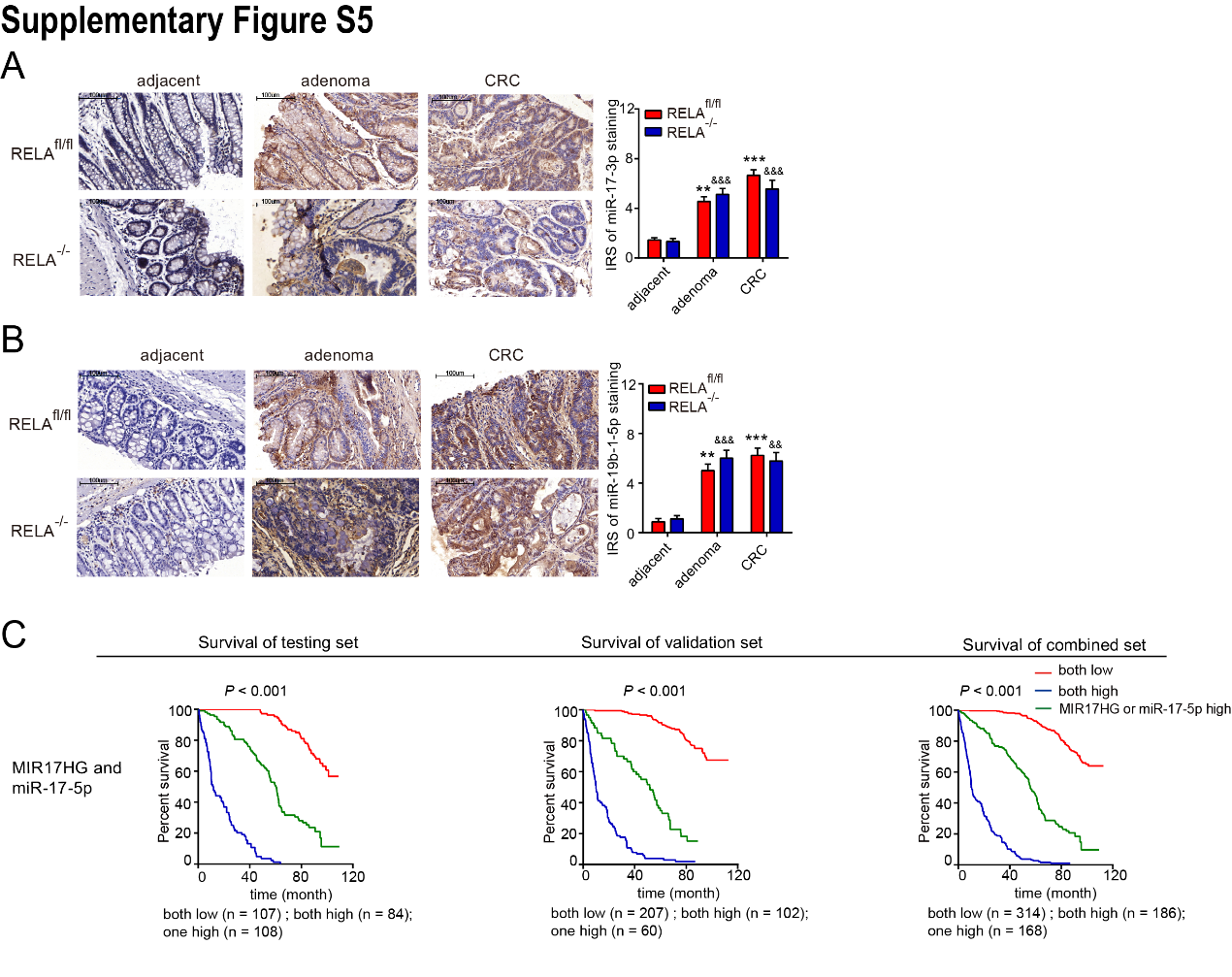
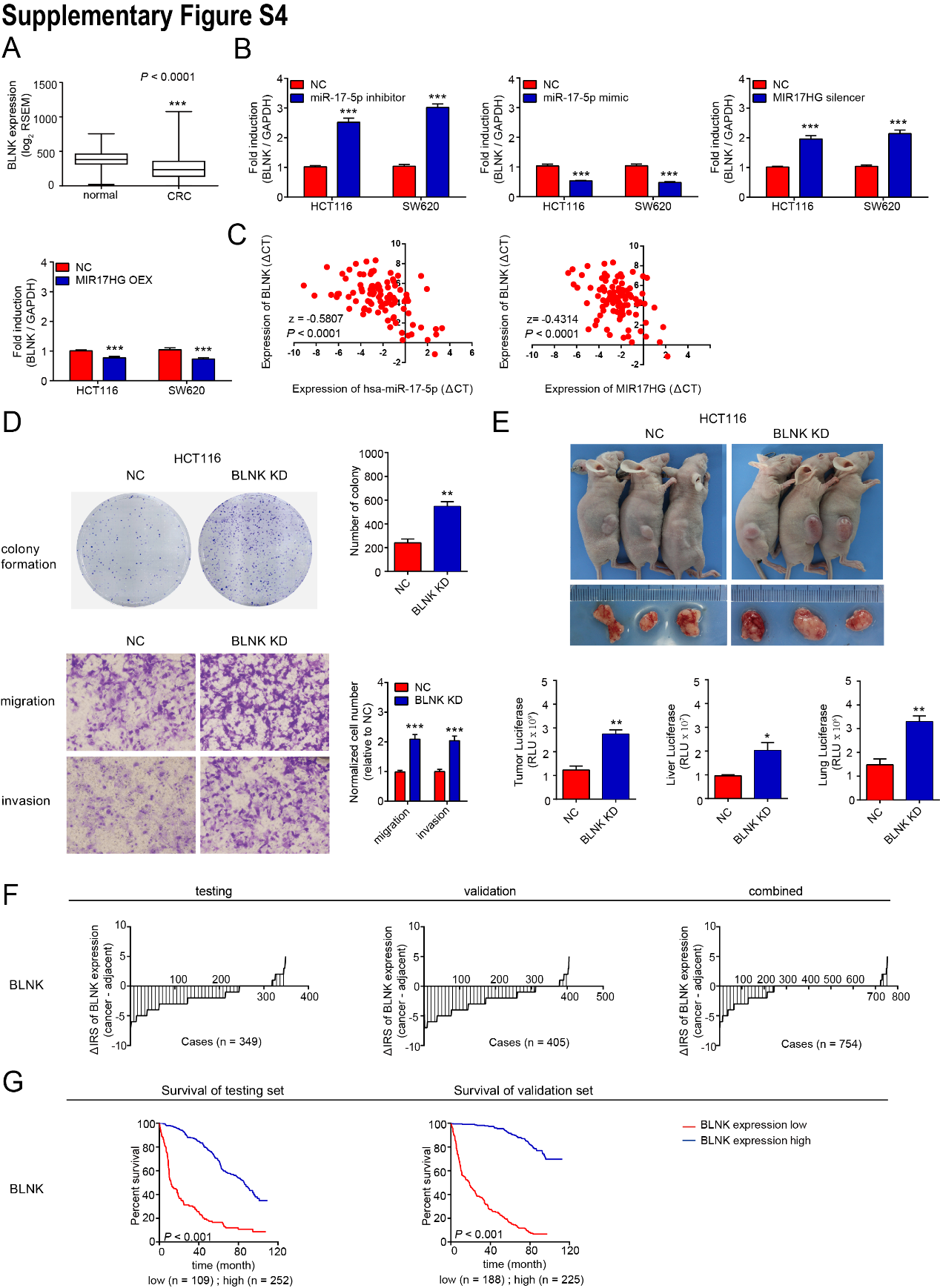
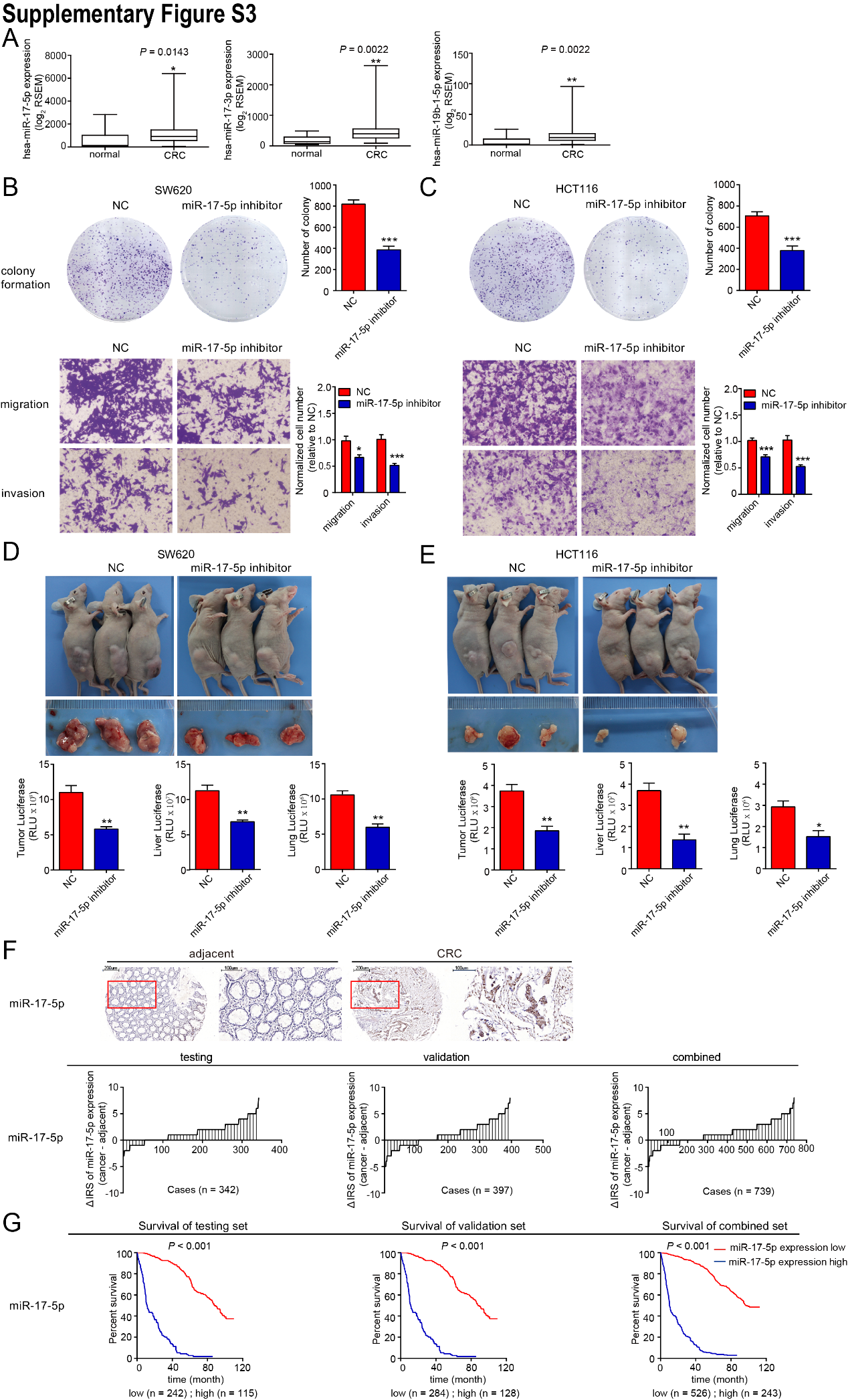
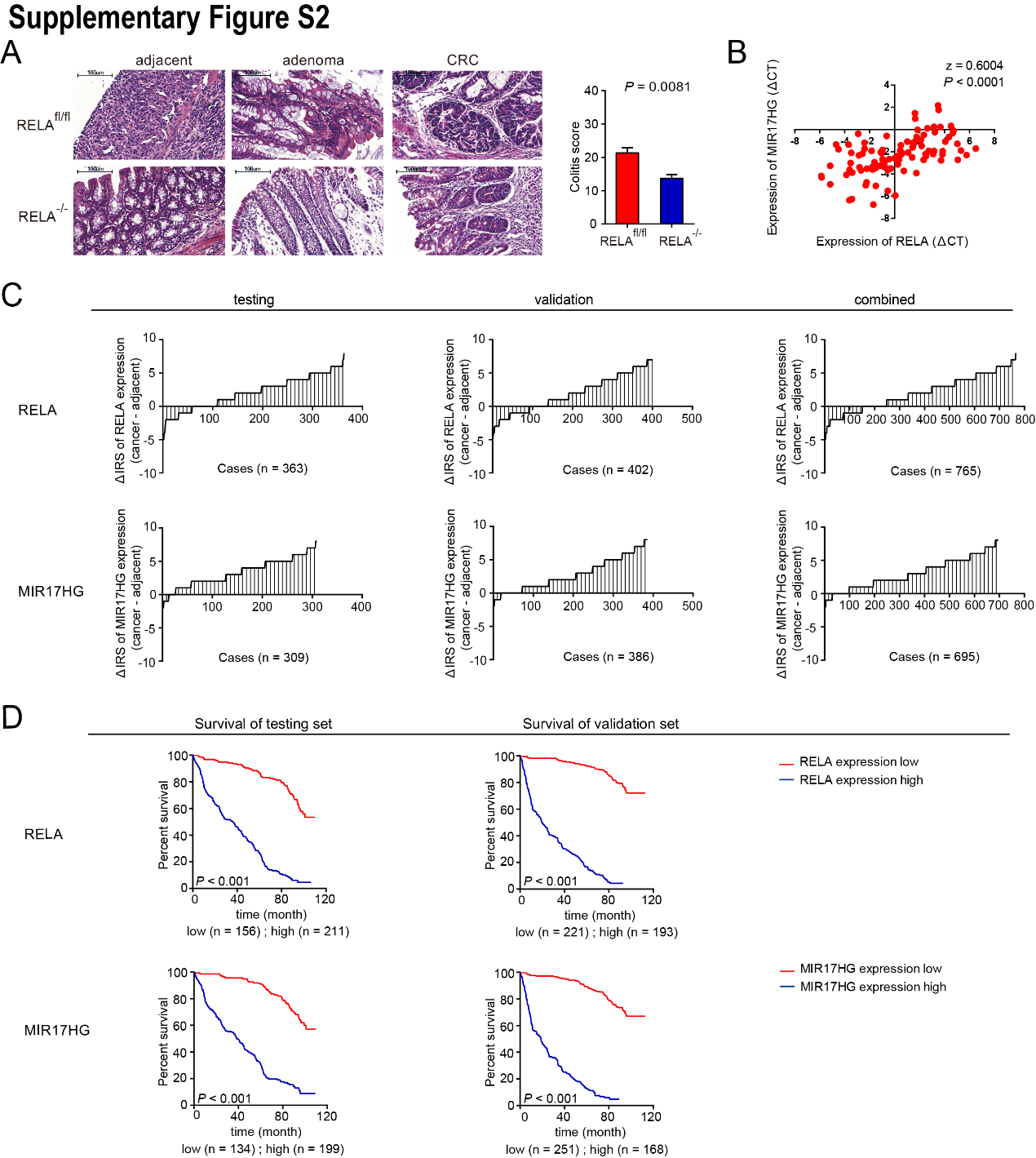
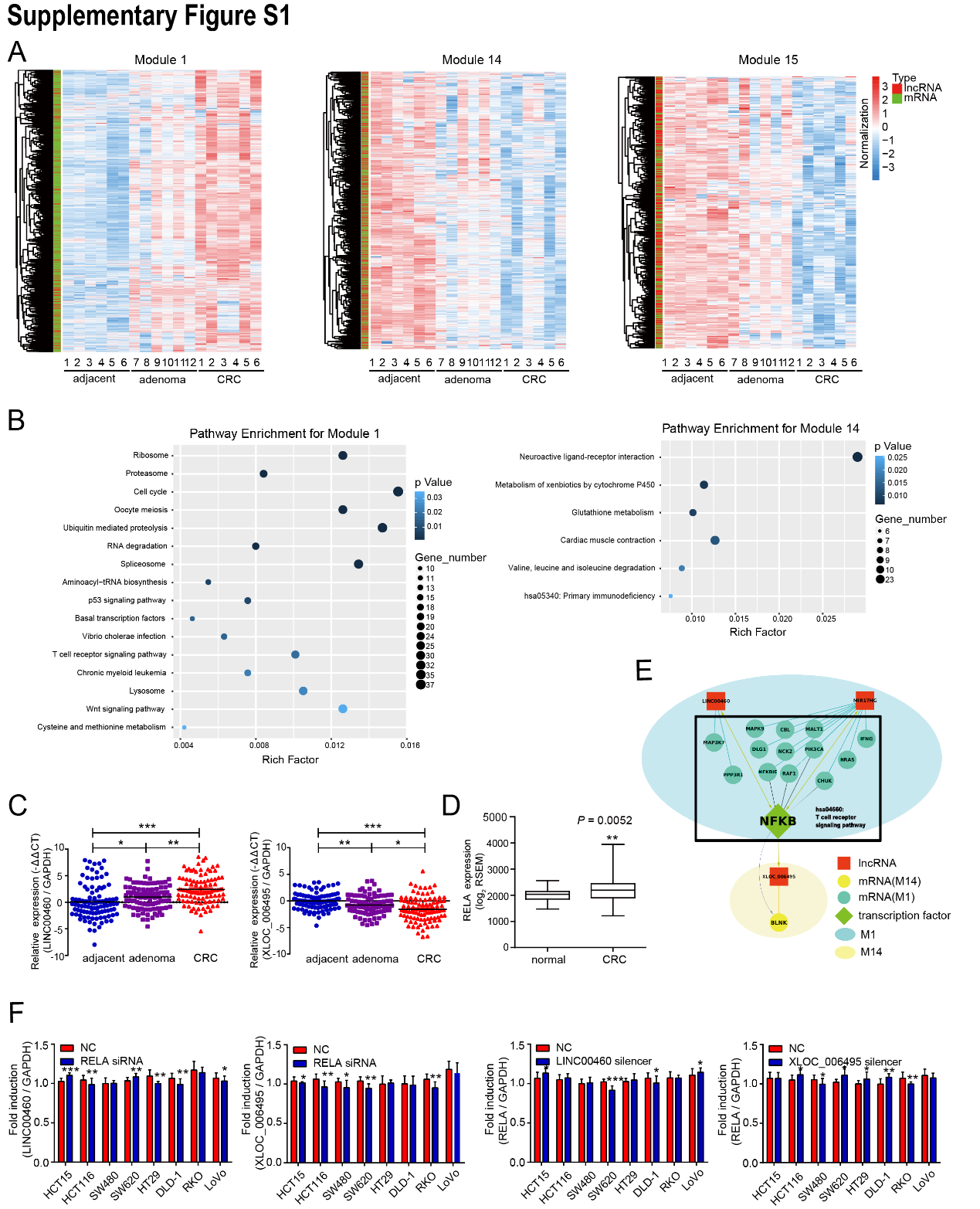
(A) Luciferase activities in cells with the indicated treatments (n = 3, \**P* < 0.05, compared with mutation and miR-NC–treated cells of each group, ANOVA followed by the Tukey’s multiple comparison test). (B) MALT1, NFKBIE, PPP3R1, and MAP3K7 expression levels in adjacent, adenoma, and CRC tissues as determined by qRT-PCR (n = 96 per group, \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001, one-way ANOVA). MALT1, NFKBIE, PPP3R1, and MAP3K7 expression levels in TCGA were analyzed by qRT-PCR (normal cases, n = 51; CRC cases, n = 286, \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001, two-tailed t-test following log transformation). The correlations of the expression levels of various targets are shown (n = 96, Fisher r-to-z transformation test followed by the Pearson correlation test).

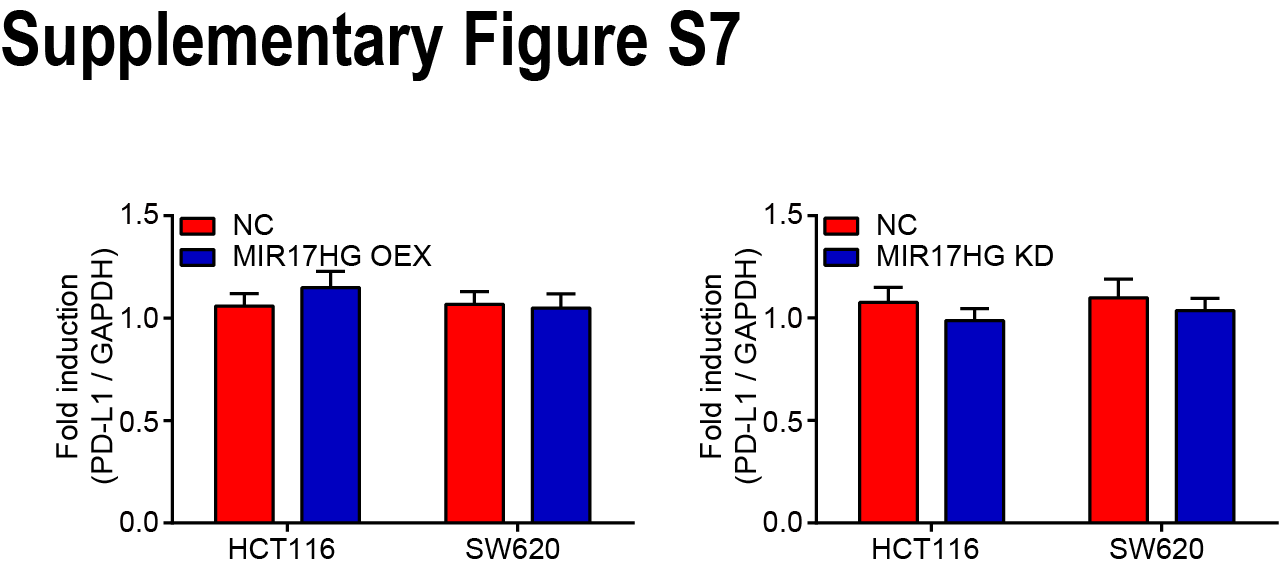
**Supplementary Fig. S7 Expression levels of PD-L1 mRNA in cells.**

Expression levels of PD-L1 mRNA in cells with the indicated treatments were determined by qRT-PCR (n = 6, compared with NC-treated cells within each group, two-tailed t-test).

**Supplementary Fig. S8 Expression levels of PD-L1 protein in tumor tissues.**

Expression levels of PD-L1 in tumor tissues formed by subcutaneous injection of (A) SW620 and (B) HCT116 cells with the indicated treatments into nude mice (n=9, compared with NC, \*\**P* < 0.01, two-tailed t-test).

****

****

**Supplementary Figure S8**

****