*tableSupplementary Appendix*

# *miR-221* targets QKI to enhance the tumorigenic capacity of human colorectal cancer stem cells.

Junko Mukohyama, Taichi Isobe, Qingjiang Hu, Takanori Hayashi, Takashi Watanabe, Masao Maeda, Hisano Yanagi, Xin Qian, Kimihiro Yamashita,Hironobu Minami, Koshi Mimori, Debashis Sahoo, Yoshihiro Kakeji, Akira Suzuki, Piero Dalerba and Yohei Shimono.

## MATERIALS and METHODS

**Ethics statements.** Human primary *colorectal carcinoma* (CRC) specimens were obtained from patients admitted to the *Division of Gastrointestinal Surgery* of the *Kobe University Hospital*. The research was pre-approved by the *Institutional Review Board* (IRB) of Kobe University (permission number: 1299) ) and was conducted in accordance with recognized ethical guidelines (Declaration of Helsinki, CIOMS). All patients included in the study provided written informed consent. All animal experiments were performed with the approval of the *Institutional Animal Care and Use Committee* (IACUC) of *Kobe University* (permission number: 150802).

**Cell lines.** All cell lines used in this study were obtained from the *American Type culture Collection* (ATCC; http://www.atcc.org) and include: HCT116 human colon cancer cells (ATCC catalog: CCL-247) and HEK293 human embryonic kidney cells (ATCC catalog: CRL-1573). All cell lines were cultured in RPMI-1640 (Sigma-Aldrich) containing 10% *fetal bovine serum* (FBS), penicillin (100 U/mL) and streptomycin (100 mg/mL; Nacalai, Japan). Early passage cells were used in all experiments. All cell lines were tested to be Mycoplasma free by PCR and authenticated using short tandem repeat profiling (BEX, Japan).

***Patient-derived xenograft* (PDX) lines.** PDX lines were established by sub-cutaneous (s.c.) xeno-transplantation of primary surgical specimens on the back of female, *non-obese, diabetic*, *severe combined immunodeficiency*, *interleukin-2 receptor gamma chain knock-out* (NOD/SCID/IL2R$\gamma^{null}$ or NSG) mice (Charles River, USA), following previously described protocols **(1-3)**.

**Preparation of single-cell suspensions and analysis by flow cytometry.** Tissue specimens from cancer patients, PDX tumors and normal colon epithelium were washed with cold (4 °C) PBS and disaggregated into single-cell suspensions, using previously published protocols, involving both mechanical and enzymatic steps **(2, 3)**. Briefly,

tumor tissues were minced into small pieces with a razor blade, resuspended in Medium199 containing 120 μg/ml penicillin, 100 μg/ml streptomycin, 20 mM HEPES, 100 unit/ml DNase-I (Sigma, USA), 200 units/ml collagenase-III (Worthington Biochemical, USA) and finally incubated at 37˚C until tissues were fully dissociated. Dissociated single-cells were re-suspended by gentle pipetting and serially filtered, first with sterile gauze and subsequently with 70-μm and 40-μm nylon meshes, in order to remove undigested tissue fragments and large cell clumps. Contaminating red blood cells were removed by osmotic lysis with an *ammonium chloride + potassium phosphate* (ACK) lysing buffer (5 min on ice).

To disaggregate murine colon, harvested colon was opened by cutting longitudinally and washed with ice cold PBS with 5 mM EDTA. The tissue fragments were incubated in PBS with 5 mM EDTA for 45 min on ice and occasionally mixed using a vortex mixer. The resulting supernatant which was enriched for crypts was further enzymatically disaggregated into single cell suspensions, using previously published protocols (2-4). Briefly, the crypts were re-suspended in RPMI-1640 medium supplemented with 2 mM L-alanine-L-glutamine (GlutaMAX, Invitrogen), 120 μg/ml penicillin, 100 μg/ml streptomycin, 0.25 μg/ml amphotericin-B (Invitrogen), 20 mM HEPES (Sigma), 1 mM sodium pyruvate (Lonza), 100 units/ml DNase-I (Worthington) and 200 units/ml Collagenase-III (Worthington), and then incubated at 37 °C for 30 minutes until tissues were fully dissociated . Dissociated single-cell suspension was serially filtered, first with sterile gauze and subsequently with 70-μm and 40-μm nylon mesh strainers, to remove undigested tissue fragments and large cell clumps. Red blood cells were removed by osmotic lysis with an ammonium chloride and potassium phosphate (ACK) hypotonic buffer (150 mM $NH_4Cl$, 1 mM $KHCO_3$) for 5 min on ice (4 °C). To prevent unspecific binding of antibodies, cells were incubated with 0.6% w/v normal human IgG (Gammagard Liquid; Baxter) for 10 min on ice (4 °C), at a concentration of 3–5 × $10^5$ cells/100 μl. Cells were subsequently washed and stained with antibodies at dilutions determined by appropriate titration experiments. To prevent non-specific binding of antibodies, cell preparations were incubated with normal mouse IgGs (1:100; Wako, Japan) for 30 min on ice (4 °C), at a concentration of 1 × $10^6$ cells/100 μl. Cells were subsequently washed and stained with antibodies at dilutions determined by appropriate titration experiments. The antibodies used for the staining of human colon cancer and human colon cancer xenograft cell suspensions are: anti-human EpCAM-AlexaFluor488 (clone 9C4; 1:100; Biolegend, USA), anti-human CD44-allophycocyanin (APC (clone IM7, 1:20; Biolegend, USA), anti-human CD45-PE-Cy7 (clone HI30, 1:40; BD Pharmigen), anti-Kit (CD117)-PE (clone 2B8, Biolegend), anti-human CD31-biotin (clone WM59, 1:40; eBioscience), anti-human CD3-biotin (clone UCHT1, 1:40; BD Pharmigen), anti-human CD16-biotin (clone 3G8, 1:40; BD Pharmigen), anti-human CD64-biotin (clone 10.1, 1:40; BD Pharmigen) anti-mouse H-2$K^d$/H2-$D^d$-biotin (clone 34-1-2S, eBioscience), and anti-mouse Cd45-biotin (clone 30-F11, BD Biosciences) antibodies. The antibodies used for the staining of murine colon cell suspensions are: anti-mouse Cd3-PE/Cy5 (clone 17A2; BD Biosciences), anti-mouse Cd45-biotin (clone 30F-11; BD Biosciences), anti-mouse Cd16/Cd32-biotin (clone 2.4G2; BD Biosciences), anti-mouse Cd31-biotin (clone 390; eBioscience), anti-mouse Epcam-FITC (clone G8.8; Biolegend), anti-mouse Cd66a-APC/Cy7 (clone Mab-CC1; Biolegend), anti-mouse Cd44-APC (clone IM7; Biolegend) and anti-mouse Kit/Cd117-

PE (clone 2B8; Biolegend). Cells stained with biotin-conjugated antibodies were visualized by secondary staining with Streptavidin-PE-Cy5 (1:200; BD Pharmigen). After 20 min on ice, stained cells were washed to remove excess unbound antibodies and resuspended in *Hank's Balanced Salt Solution* (HBSS) with 2% *heat-inactivated calf serum* (HICS), 20 mM HEPES, 5 mM EDTA, 1 mM sodium pyruvate, 120 μg/ml penicillin, 100 μg/ml streptomycin, 0.25 μg/ml amphotericin-B (Invitrogen) and 1.1 μM DAPI dilactate (Molecular Probes). Single-cell suspensions from both human and mouse samples were analyzed by flow cytometry for the differential expression of surface markers, and purified by *fluorescence activated cell sorting* (FACS) based on phenotype, using a FACSAria-III cell sorter (BD Biosciences), as previously described **(2-4)**. Forward-scatter height versus forward-scatter width (FSC-H vs. FSC-W) and side-scatter height versus side-scatter width (SSC-H vs. SSC-W) profiles were used to eliminate cell doublets. Dead cells were eliminated by excluding $DAPI^+$ cells, whereas $Lin^+$ cells were eliminated by excluding $PE/Cy5^+$ cells. $Lin^-/EpCAM^+/CD44^+$ and $Lin^-/EpCAM^+/CD44^{neg}$ human colorectal cancer cells were collected.

**Megaplex semi-quantitative real-time PCR assay.** In experiments performed on human primary colon cancer tissues, two paired sets of $EpCAM^+/CD44^+$ cancer cells (enriched in CSCs) and $EpCAM^+/CD44^{neg}$ cancer cells (depleted in CSCs) were isolated by FACS from two independent patients, as described above. A minimum of 100 cells from each population were double-sorted directly into TRIzol (Invitrogen) and used for RNA purification, following the manufacturer's protocol. Glycogen (Invitrogen) was used as a carrier for precipitation. In each sample, the expression level of 754 miRNAs was measured by *quantitative reverse transcription polymerase chain reaction* (RT-qPCR), using a previously validated protocol for the multiplex expression profiling of miRNAs (TaqMan™ Array Human MicroRNA A+B Cards Set v3.0 with Megaplex™ RT Primers, Human Pool Set v3.0; Thermo Fisher Scientific), as previously described **(5-9)**. Briefly, for each sample, the 754 target miRNAs were first reverse-transcribed into corresponding cDNAs using a multiplex mixture of 754 miRNA-specific stem-looped reverse primers. The cDNA products were subsequently pre-amplified using a multiplex pre-PCR reaction, based on a mixture of 754 independent forward primers and a universal reverse primer (i.e. complementary to a sequence shared across all reverse primers used for cDNA synthesis). Finally, the multiplexed pre-PCR product was diluted 1:8, and the expression level of each miRNA was analyzed independently by qPCR, using: 1) the miRNA-specific forward primer; 2) a miRNA-specific TaqMan probe; and 3) the universal reverse primer. Differences in miRNA expression levels between two populations were measured as fold-changes and calculated based on delta-delta Ct ($\Delta\Delta Ct$) values, after normalization to RNU48 small nuclear RNA (RNU48 snRNA) expression levels, as previously described **(10, 11)**.

**Real-time RT-qPCR analysis of *miR-221* expression.** The expression levels of *miR-221* were analyzed by RT-qPCR on paired sets of $EpCAM^+/CD44^+$ and $EpCAM^+/CD44^{neg}$ populations, isolated by FACS from human primary CRCs (n=6) and human normal colon epithelia (n=4), as described above **(3, 4)**. Briefly, RNA was extracted from

a minimum of 100 cells from each population using TRIzol (Invitrogen), following the manufacturer's protocol **(12)**. Glycogen (Invitrogen) was used as a carrier for precipitation. Reverse transcription and cDNA synthesis were performed using the *SuperScript-III* reverse transcription kit (Invitrogen). In each sample, the expression level of miR-221 was measured by *quantitative reverse transcription polymerase chain reaction* (RT-qPCR), using a previously validated protocol for the expression profiling of miRNAs (Kaiqin Lao, Thermo Fisher Scientific), as previously published **(5-8)**. Briefly, for each sample, the target miRNA was first reverse-transcribed into corresponding cDNAs using a miRNA-specific stem-looped reverse primers. The cDNA products were subsequently pre-amplified using a pre-PCR reaction, using a miRNA-specific forward primer and a universal reverse primer. Finally, the pre-PCR product was diluted 1:8, and the expression level of miR-221 was analyzed by qPCR, using: 1) the miRNA-specific forward primer; 2) a miRNA-specific TaqMan probe (**Table S1**); and 3) the universal reverse primer. RT-qPCR reactions were run on a *Thermal Cycler Dice* (TaKaRa, Shiga, Japan). Differences in miRNA expression levels between two populations were measured as fold-changes and calculated based on delta-delta Ct (ΔΔCt) values, after normalization to RNU48 small nuclear RNA (RNU48 snRNA) expression levels, as previously described **(10, 11)**. Differences in ΔΔCt values between two populations were tested for statistical significance using the Mann-Whitney Wilcoxon U-test.

**microRNA-sequencing (microRNA-seq) experiments.** Total RNA (including small RNA) was isolated from purified preparations of Epcam$^{+}$, CD66a$^{low}$, Cd44$^{+}$, Kit$^{+}$ (Kit$^{+}$) and Epcam$^{+}$, CD66a$^{low}$, Cd44$^{+}$, Kit$^{neg}$ (Kit$^{neg}$) normal murine colon epithelial cells (mouse strain: C57BL/6J, female; The Jackson Laboratory; stock #000664). The analyzed populations were sorted by FACS, as paired sets, isolated in parallel from primary tissue specimens. Total RNA was extracted using the *NucleoSpin miRNA* kit (MACHEREY-NAGEL, Germany). The microRNA-seq profiling was performed on biological triplicates (i.e. on three paired samples of Kit$^{+}$ and Kit$^{neg}$ cells) with the assistance of LC Sciences (Houston, TX). Briefly, small RNA libraries were generated using the *Truseq Small RNA Preparation Kit* (Illumina, San Diego, CA) according to the manufacturer's sample preparation guide. Purified cDNA libraries were then used for cluster generation on Illumina's *Cluster Station* and then sequenced on the Illumina *HiSeq* platform. Raw sequencing reads were obtained using Illumina's *Sequencing Control Studio* software (version 2.8). Real-time sequencing image analysis and base-calling were performed by Illumina's *Real-Time Analysis* software (v1.8.70). Gene-expression levels were evaluated using two different analytical pipelines: 1) a proprietary mapping and normalization pipeline developed by LC Sciences **(13)**; and 2) a publicly available mapping and normalization pipeline developed by *The Encyclopedia of DNA Elements* (ENCODE) *Project* international consortium (https://www.encodeproject.org/microrna/microrna-seq) **(14)**. In the first pipeline, raw reads were processed using LC Science's in-house *ACGT101-miR* program (LC Sciences, Houston, Texas, USA) to remove adapters, dimers, low quality sequences, sequences from other RNA families (rRNA, tRNA, snRNA, snoRNA), and low complexity repeats. Subsequently, unique sequences with a length of 18-26 nucleotides were

mapped to specific species precursors in miRBase (v22.0; http://www.mirbase.org), using the *National Institute for Biotechnology Information* (NCBI) *Basic Local Alignment Search Tool* (BLAST; https://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify known miRNAs and novel 3p- and 5p- derived miRNAs. Length variations at both 3' and 5' ends, as well as one nucleotide (1 nt) mismatches inside of the sequence, were allowed in the alignment. Very briefly, unique sequences mapping to the mature arm of previously annotated miRNA species were identified as known miRNAs. Unique sequences mapping to the opposite arm of previously annotated miRNA species were considered to be novel 5p-derived or 3p-derived miRNA candidates. The remaining sequences were mapped to candidate miRNA precursors in the mouse genome, and evaluated for possible hairpin RNA structure based on the sequence of flanking regions (80 nt), usng the *RNAfold* software (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi). The criteria for secondary structure prediction were: 1) the number of nucleotides in any bulge region of the stem (≤ 12 nt); 2) the number of base pairs in the stem region of the predicted hairpin (≥ 16 nt) ; 3) a low free energy of the hairpin (≤ -15 kCal/mol); 4) the total length of the hairpin (up and down stems + terminal loop ≥ 50 nt); 5) the total length of the hairpin's loop (≤ 20 nt); 6) the number of nucleotides any bulge in the mature miRNA sequence (≤ 8 nt) ; 7) the total number of biased errors in any bulge in the mature miRNA sequence (≤ 4 nt); 8) the total number of biased bulges in the mature miRNA sequence (≤ 2 nt); 9) the total number of errors in the mature miRNA sequence (≤ 7 nt); 10) the total number of base pairs in the mature miRNA region of the predicted hairpin (≥ 12 nt); and 11) the percentage of the predicted mature miRNA sequence located in stem (≥ 80%). In this first pipeline, the number of mapped miRNA reads was normalized by calculating a correction factor for each specimen, using a linear regression approach to define the relationship between the $\log_2$ values of a specific set of reference miRNAs in each specimen and those of a virtual specimen generated using the median value of all miRNAs displaying the least dispersion across the dataset (i.e. all RNAs with > 5 reads and < 4-fold change compared to the median value). In the second pipeline, raw reads were processed using the ENCODE Consortium's software platform (https://www.encodeproject.org/microrna/microrna-seq), and miRNA-expression levels were expressed as the $\log_2$ of their RPMs (*reads per million miRNA mapped reads*). Irrespective of the analytical pipeline, differences in the $\log_2$ of miR-221 expression levels between $Kit^{+}$ and $Kit^{neg}$ populations were tested for statistical significance using a one-tailed Student's t-test for paired samples.

**Association between *miR-221* expression levels and survival outcomes in human colon cancer patients.** The association between mature *miR-221* expression levels and survival outcomes was tested on a dataset of 293 colon cancer patients from *The Cancer Genome Atlas* (TCGA) database. Paired mature miRNA expression data and clinical outcome information relative to the 293 colon cancer patients was downloaded from the *Broad Institute*'s *Firehose* website (https://gdac.broadinstitute.org), at the following online addresses:

a) https://gdac.broadinstitute.org_COADREAD.miRseq_Mature_Preprocess.Level_3.2016012800.0.0
b) https://gdac.broadinstitute.org_COADREAD.Clinical_Pick_Tier1.Level_4.2016012800.0.0

Patients were divided into two groups (*miR-221*$^{high}$ vs. *miR-221*$^{low}$) using the minimum *P*-value approach, which is a method used to identify the optimal cut-off point for risk-stratification in large datasets defined by continuous variables **(15)**. *Overall survival* (OS) and *disease-free survival* (DFS) rates were estimated using Kaplan-Meier survival curves. Differences in survival rates were tested for statistical significance using the log-rank test. The association between *miR-221* expression levels and an independent contribution to survival outcomes was evaluated using univariate and multivariate analyses based on the Cox proportional hazard model. Data analyses were performed using the "R" software, version 3.3.2 (The R Project; https://www.r-project.org).

**Lentivirus backbone plasmids.** The full-length sequence of *miR-221*, inclusive of its stem-loop structure, was amplified by PCR starting from the genomic DNA of A549 cells, and cloned into the pEIZ-HIV-ZsGreen lentivirus backbone vector **(8, 16)**. The full-length coding region of the *QKI-5* mRNA (NM_001301085), devoid of its 3'UTR sequence, was amplified by PCR from the cDNA of A549 cells (**Table S1**), and cloned into the pLentiLox3.7-EF1-mCherry lentivirus backbone, a derivative of the pLentiLox3.7 construct (Addgene: #11795). All products were sequenced to exclude the presence of undesired mutations. The lentivirus vectors encoding for the *anti–miR-221* construct (miRZip-221) and a scrambled non-targeting pre-miRNA (negative control) were purchased from *System Biosciences* (USA). Lentiviruses were produced as previously described **(17)**.

**Organoid assay.** HCT116, PDX-KUC1 and PDX-KUC2 colon cancer cells were infected with either test or control lentivirus constructs, resuspended in organoid culture media, seeded on Matrigel™ in 96-well plates at a density of 3 x $10^3$ cells/well, and cultured at 37°C with 5% $CO_2$, as previously described **(18, 19)**. Briefly, the cells were resuspended in Advanced DMEM/F12 tissue-culture medium (Gibco), supplemented with 2 mM L-alanine-L-glutamine (GlutaMAX), 10 mM HEPES (Sigma), 1mM sodium pyruvate (Lonza), 10% heat-inactivated *fetal bovine serum* (FBS), 120 µg/ml penicillin, 100 µg/ml streptomycin, 0.25 µg/ml amphotericin-B (Invitrogen), ITES media supplement (Lonza), 50 ng/ml hEGF (Stem Cell Technologies), 500 ng/ml hR-Spondin1 (R&D systems), 100 ng/ml hNoggin (Peprotech) and 10 µM Y-27632 (Calbiochem). A total of 3 x $10^3$ cells (100 µl) were then plated atop a polymerized layer of Matrigel™ (BD Bioscience). The tissue-culture medium was then replaced every three days, until established organoids were formed and used for experimental purposes. The numbers of ZsGreen and/or mCherry positive organoids larger than 100 µm in diameter were counted 10 days after seeding. Differences in organoid frequency were tested for statistical significance using a two-tailed Student's t-test and, in cases where more than two conditions were tested in parallel, a two-way *analysis of variance* (ANOVA) test.

**Cell cycle analysis.** HCT116 cells were transfected with either a plasmid encoding for the *anti-miR-221* construct

(miRZip-221) or a plasmid encoding for a scrambled non-targeting pre-miRNA, and cultured for 2 days. Cells were then fixed with 70% ethanol at -20 ˚C for 2 hours, and stained with an anti-human Ki-67 antibody (clone B56; 1:20, BD Biosciences) resuspended in *PI/RNase Staining Buffer* (BD Biosciences, USA), following the manufacturer's instructions. The percentage of Ki-67$^{+}$ was evaluated using a *FACS Verse* flow cytometer (BD Biosciences). All experiments were performed in triplicate. Differences in the percentage of Ki-67$^{+}$ cells were tested for statistical significance using a two-tailed Student's t-test.

**Quantification of apoptosis.** HCT116 cells were transfected with either a plasmid encoding for the *anti-miR-221* construct (miRZip-221) or a plasmid encoding for a scrambled non-targeting pre-miRNA, and cultured for 5 days. HCT116 cells were then harvested, incubated with normal mouse IgGs (1:100, Wako, Japan) to block unspecific protein binding, and finally stained with Annexin-V (1:20, BioLegend, USA). The frequency of apoptotic cells was evaluated using a *FACS Verse* flow cytometer (BD Biosciences), by measuring the percentage of Annexin-V$^{+}$/Propidium Iodide$^{neg}$ cells. All experiments were performed in triplicate. Differences in the percentage of Annexin-V$^{+}$/Propidium Iodide$^{neg}$ cells were tested for statistical significance using a two-tailed Student's t-test.

***In vivo* xenotransplantation assay.** PDX-KUC1 cells were infected with either a lentivirus vector encoding for the *anti-miR-221* construct (miRZip-221), a lentivirus vector encoding for the *QKI-5* mRNA or an empty lentivirus vector (negative control), using a *multiplicity of infection* (MOI) titer of 20. Briefly, 1.5 x $10^5$ infected cells were washed with PBS, mixed with Matrigel, and injected sub-cutaneously (s.c.) into the back of immunodeficient NSG mouse (Charles River, USA) as previously described **(1-3, 18)**. To reduce experimental variability due to individual differences in recipient mice, the cell populations to be compared were injected on opposite flanks of the same animal. Tumor growth was periodically monitored after the injection. The tumor volume was calculated using the formula of width$^2$×length×0.5 (width < length). Differences in the frequency of animals developing sub-cutaneous tumors following the injection of equal numbers of cancer cells (1.5 x $10^5$ cells/mouse) were tested for statistical significance using Fisher's exact test. Differences in tumor volume were tested for statistical significance using a two-tailed Student's t-test.

**Plasmid vectors for luciferase reporter assays and mutagenesis.** A 415-bp fragment from the 3'UTR of the human *QKI-5* mRNA, corresponding to nucleotide positions 2395-2809 of the NM_001301085 sequence in the GenBank database, was amplified by PCR from the genomic DNA of A549 cells (**Table S1**) and cloned into the pMD20-T vector (Takara Bio, Japan). The amplification product was then cloned into the pGL3-MC vector, immediately downstream of the firefly *Luciferase* gene (*Photinus pyralis*), and sequenced to exclude the introduction of undesired mutations **(8)**. Mutations of the putative *miR-221* target sequence within the 3'UTR of the *QKI-5* mRNA were then artificially introduced using a *PrimeSTAR Mutagenesis Basal Kit* (Takara Bio, Japan) (**Table S1**).

**Luciferase reporter assay.** The capacity of *miR-221* to recognize its candidate target sequence within the 3'UTR of the human *QKI-5* mRNA was evaluated using luciferase reporter assays, in which the synthetic precursor of *miR-221* was transfected into cultured cells alongside a pGL3-MC luciferase reporter construct that harbored either wild-type (wt) or mutated (mt) versions of the relevant 3'UTR sequence form the human *QKI-5* mRNA (Supplementary table 1), immediately downstream of the cDNA encoding for the firefly (*Photinus pyralis*) *Luciferase* gene. The mutated version of the 3'UTR was designed to selectively inactivate the putative target recognition sequences of *miR-221*. Briefly, HEK293T and HCT116 cells were cultured in RPMI-1640 medium, supplemented with 10% heat-inactivated FBS, 2 mM L-alanine-L-glutamine (GlutaMAX), 120 μg/ml penicillin, 100 μg/ml streptomycin, 20 mM HEPES and 1 mM sodium pyruvate, as previously described. One day prior to transfection, either HEK293T or HCT116 cells were harvested and re-plated in 96-well plates, at a concentration of 5 x $10^4$ cells/well. All transfections were carried out with Lipofectamine 3000 (Invitrogen, USA), following the manufacturer's instructions. Cells were transfected with 25 ng of the pGL3 luciferase expression construct containing the 3'UTR of *QKI-5*, 3 ng of the pRL-TK Renilla luciferase vector (Promega, USA), and 75 ng of either the pEIZ plasmid driving constitutive *miR-221* expression or its empty backbone (negative control). Sixty hours after transfection, luciferase activity was quantified using the *Dual-Luciferase Reporter Assay System* (Promega), whereby firefly luciferase activity from the pGL3-MC reporter constructs was normalized to sea pansy *Luciferase* (*Renilla reniformis*) activity from the pRL-TK plasmid. All experiments were performed in triplicate. Differences in luciferase expression levels between equal numbers of cells transfected with different combinations of pGL3-MC reporter constructs and miRNA precursors were tested for statistical significance using both a two-tailed Student's t-test and a two-way *analysis of variance* (ANOVA) test.

**Western blotting.** HCT116 cells were seeded at 7.5 x $10^5$ cells/well in 6-well plates, transfected with 3,000 ng of either a constitutive *miR-221*-expression or a negative control plasmid, and cultured for 4 days. All transfections were carried out with Lipofectamine 3000 (Invitrogen, USA), following the manufacturer's instructions. Transfected cells were lysed using a *radio-immunoprecipitation assay* (RIPA) buffer (ThermoFisher, USA) supplemented with *cOmplete Mini* protease inhibitor cocktail (Roche, Switzerland) and left for 1 hour on ice. Cell lysates were then separated by electrophoresis on a SDS-12% polyacrylamide gel, and transferred to a *polyvinylidene difluoride* (PVDF) membrane, using the *Trans-Blot Turbo Mini PVDF Transfer* system (Bio-Rad). After blocking with 5% *bovine serum albumin* (BSA), filters were incubated with either an anti-QKI-5 (clone: A300-182A, 1:1,000, Bethyl, USA) or an anti-β-actin monoclonal antibody (clone: AC-74, 1:3,000, Sigma-Aldrich, USA) at 4°C overnight. As a final step, filters were incubated with a goat anti-mouse IgG-*horseradish peroxidase* (HRP) conjugate (1:3,000, Bio-Rad) and protein bands detected using the *Chemi-Lumi One L* bioluminescent assay (Nacalai).

**Bioinformatics analysis of miRNA and mRNA gene-expression patterns in RNA-sequencing (RNA-seq) databases.** The presence of a linear correlation between the expression levels of *miR-221* and those of the *QKI* mRNA was tested in the *colon adenocarcinoma* (COAD) dataset of the *The Cancer Genome Atlas* (TCGA) public repository (https://cancergenome.nih.gov; downloaded: March 8, 2016), which provided paired RNA-seq data for both miRNA and mRNA transcripts **(20)**. The dataset included 378 samples defined as *"colon adenocarcinoma"* and 61 samples defined as *"colon mucinous adenocarcinoma"*. Correlations between the expression levels of *miR-221* and those of the *QKI* mRNA were evaluated by calculating Pearson's correlation coefficients between the $\log_2$ TPM (*transcripts per million*) values for the *QKI* mRNA and the $\log_2$ RPM (*reads per million miRNA mapped reads*) values for *miR-221*, across the full dataset. The statistical significance of the Pearson's correlation coefficients was evaluated using a two-tailed t-test for correlation coefficients (null hypothesis: $r = 0$).

**Exome sequencing.** PDX tumors were washed with cold (4 °C) PBS and disaggregated into single-cell suspensions, using previously published protocols, involving both mechanical and enzymatic steps **(2, 3)**. Briefly, tumor tissues were minced into small pieces with a razor blade, resuspended in Medium199 containing 120 μg/ml penicillin, 100 μg/ml streptomycin, 20 mM HEPES, 100 unit/ml DNase-I (Sigma, USA), 200 units/ml collagenase-III (Worthington Biochemical, USA) and finally incubated at 37˚C until tissues were fully dissociated. Genomic DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen), following the manufacturer's instructions. The following steps for exome sequencing and analyses were performed at RIKEN Genesis, Japan. Genomic DNA sheared into approximately 200 bp fragments was used to make a library for multiplexed paired-end sequencing with the SureSelectXT Reagent Kit (Agilent Technologies). The constructed library was hybridized to biotinylated cRNA oligonucleotide baits from the SureSelectXT Human All Exon V6 Kit (Agilent Technologies) for target enrichment. Targeted sequences were purified by magnetic beads, amplified, and sequenced on an Illumina NovaSeq 6000 platform in paired-end 101 bp configuration.

**Mapping and SNV/Indel calling.** Adapter sequences were removed by cutadapt (ver.1.2.1). Mouse reads were removed using DeconSeq (ver.0.4.3) because the samples were PDXs. After quality control, reads were mapped to the reference human genome (hg19) using BWA (ver.0.7.10). Mapping result was corrected using Picard (ver.1.73) for removing duplicates and GATK (ver.1.6-13) for local alignment and quality score recalibration. SNV and Indel calls were performed with multi-sample calling using GATK, and filtered to coordinates with VQSR passed and variant call quality score ≥ 30. Annotations of SNVs and Indels were based on dbSNP150, CCDS (NCBI, Release 15), RefSeq (UCSC Genome Browser, Feb 2018), Gencode (UCSC Genome Browser, ver. 19), and 1000Genomes (phase3 release v5). Variants were further filtered according to the following criteria: with predicted functions of frameshift, nonsense, read-through, missense, deletion, insertion, or insertion-deletion.

**Supplementary Table S1**

**Table S1.** List of the oligonucleotide primers used for plasmid construction and mutagenesis.

| Amplified gene | Primer sequence | Length of product (bp) |
|---|---|---|
| **miR-221-F** (forward) | GCTCTAGACCTGCCCACGTACCTAC | 680 bp |
| **miR-221-R** (reverse) | GCGGATCCATGAACAGAAATAGAAGCCAAAAAG | |
| **QKI-5-F** | GCGAATTCTGGAATATGGTCGGGGAAATGG | 1,025 bp |
| **QKI-5-R** | GCTCTAGAGGTCAGAAGGTCATAGGTTAGTTG | |
| **QKI-5 (target site)-F** | GCCGCGGAAGCCATGCTTGCCTATTTGC | 240 bp |
| **QKI-5 (target site)-R** | GCTGCAGTCACATACTGGGCGACTGAG | |
| **QKI-5 mut-F** | ATCACCAACAGCCCAGGCTGCTCCA | n.a. |
| **QKI-5 mut-R** | CTGGGCTGTTGGTGATTTAATGTTGGC | |

## REFERENCES


1. Mukohyama J, Iwakiri D, Zen Y, Mukohara T, Minami H, Kakeji Y, et al. Evaluation of the risk of lymphomagenesis in xenografts by the PCR-based detection of EBV BamHI W region in patient cancer specimens. Oncotarget. 2016;7(31):50150-60.
2. Dalerba P, Dylla SJ, Park IK, Liu R, Wang X, Cho RW, et al. Phenotypic characterization of human colorectal cancer stem cells. Proc Natl Acad Sci U S A. 2007;104(24):10158-63.
3. Dalerba P, Kalisky T, Sahoo D, Rajendran PS, Rothenberg ME, Leyrat AA, et al. Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat Biotechnol. 2011;29(12):1120-7.
4. Rothenberg ME, Nusse Y, Kalisky T, Lee JJ, Dalerba P, Scheeren F, et al. Identification of a cKit(+) colonic crypt base secretory cell that supports Lgr5(+) stem cells in mice. Gastroenterology. 2012;142(5):1195-205 e6.
5. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, et al. Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res. 2005;33(20):e179.
6. Isobe T, Hisamori S, Hogan DJ, Zabala M, Hendrickson DG, Dalerba P, et al. miR-142 regulates the tumorigenicity of human breast cancer stem cells through the canonical WNT signaling pathway. Elife. 2014;3.
7. Lao K, Xu NL, Yeung V, Chen C, Livak KJ, Straus NA. Multiplexing RT-PCR for the detection of multiple miRNA species in small samples. Biochem Biophys Res Commun. 2006;343(1):85-9.
8. Shimono Y, Zabala M, Cho RW, Lobo N, Dalerba P, Qian D, et al. Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. Cell. 2009;138(3):592-603.
9. Tang F, Hajkova P, Barton SC, O'Carroll D, Lee C, Lao K, et al. 220-plex microRNA expression profile of a single cell. Nat Protoc. 2006;1(3):1154-9.
10. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods (San Diego, Calif. 2001;25(4):402-8.
11. Zheng G, Wang H, Zhang X, Yang Y, Wang L, Du L, et al. Identification and validation of reference genes for qPCR detection of serum microRNAs in colorectal adenocarcinoma patients. PLoS One. 2013;8(12):e83025.
12. Chomczynski P, Sacchi N. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. Nat Protoc. 2006;1(2):581-5.
13. Li X, Shahid MQ, Wu J, Wang L, Liu X, Lu Y. Comparative Small RNA Analysis of Pollen Development in Autotetraploid and Diploid Rice. Int J Mol Sci. 2016;17(4):499.
14. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 2012;22(9):1760-74.
15. Mizuno H, Kitada K, Nakai K, Sarai A. PrognoScan: a new database for meta-analysis of the prognostic value of genes. BMC Med Genomics. 2009;2:18.
16. Welm BE, Dijkgraaf GJ, Bledau AS, Welm AL, Werb Z. Lentiviral transduction of mammary stem cells for analysis of gene function during development and cancer. Cell Stem Cell. 2008;2(1):90-102.

17. Tiscornia G, Singer O, Verma IM. Production and purification of lentiviral vectors. Nature protocols. 2006;1(1):241-5.
18. Sakaguchi M, Hisamori S, Oshima N, Sato F, Shimono Y, Sakai Y. miR-137 Regulates the Tumorigenicity of Colon Cancer Stem Cells through the Inhibition of DCLK1. Mol Cancer Res. 2016;14(4):354-62.
19. Shimono Y, Mukohyama J, Isobe T, Johnston DM, Dalerba P, Suzuki A. Organoid Culture of Human Cancer Stem Cells. Methods Mol Biol. 2016.
20. Chu A, Robertson G, Brooks D, Mungall AJ, Birol I, Coope R, et al. Large-scale profiling of microRNAs for The Cancer Genome Atlas. Nucleic Acids Res. 2016;44(1):e3.