Supporting Information

Boele et al. 10.1073/pnas.1317751111

SI Materials and Methods

Overview of Sequencing Data Sources Used in This Study (Table S1). Dataset accession numbers beginning with "GEO" are from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database (www.ncbi.nlm.nih.gov/geo/), accession numbers beginning with "DRX" are from the NCBI Sequence Read Archive (www.ncbi.nlm.nih.gov/sra), and accession numbers beginning with "AI" and "DRA" are from the DNA Data Bank of Japan (www.ddbj.nig.ac.jp). For each dataset, the platform on which the data were generated is indicated. MicroRNA (miRNA) and pre-miRNA sequences were retrieved from miRBase, Release 17 (1).

Sets containing extracted sequences were accepted as is; for the remaining sets, we extracted short-RNA sequences from the raw data by removing any substring of the 5′ side of 5′-TCGTATGCCGTCTTCTGCTTG-3′ (the default adapter sequence for the Illumina Genome Analyzer) from the 3′ end of the sequenced read. We counted the number of occurrences of the 22-nt canonical miR-21 sequence, and of its respective +C (templated isomiR), +CA (adenylated isomiR) variants. The filtered reads were then mapped against all mature miRNAs in miRBase using Nexalign (2) to obtain total miRNA counts; we then used these to calculate miR-21 abundance normalized to the total miRNA expression.

Cell Culture (RIKEN MCF7). Human breast carcinoma MCF7 cells were obtained from the American Type Culture Collection (Rockville, MD). To deprive them of estrogen, they were incubated for 3 d in phenol red-free Dulbecco's modified Eagle medium with 10% (vol/vol) FCS pretreated with dextran-coated charcoal. Then, 100 nM 17β-estradiol (E2) was added to cultures for the indicated times [0, 1, 2, 3, 6, 12, 24, and 48 h for the eight-time-point (8tp) experiment; 0, 2, 4, 8, 12, and 24 h for the six-time-point (6tp) experiment].

Library Construction and Sequencing. Total RNA was isolated using ISOGEN reagent (Nippon Gene). Libraries were constructed using RNAs of sizes between 12 and 52 nt, selected to minimize the amount of rRNA and tRNA. The 6tp library was produced to be able to be sequenced from both ends of short RNA. To help reduce ligation bias, the sample was divided in two: one for the library sequenced from the 5′ end and the other for the library sequenced from the 3′ end of short RNA. Just before sequencing, the two libraries were mixed into one library.

Cetyl-trimethylammonium bromide-selective precipitation of long RNAs, as described previously (3), was carried out to separate long and short RNAs. A 5′-phosphorylated 3′ RNA/DNA chimeric oligonucleotide (5′-phosphate-UUUtcgtatgccgtcttctgcttg-biotin-3′ for sequencing from the 5′ end, and 5′-phosphate-UXXxxctgtagaactctgaacctgt-biotin-3′ for sequencing from the 3′ end, where uppercase letters indicate the RNA portion of the oligonucleotide and lowercase the DNA portion of the oligonucleotide, and XXxx the barcode sequence) and a 5′ adaptor DNA/RNA chimeric oligonucleotide (5′-biotin-acaggttcagagttctacagxxXXA-3′ for sequencing from the 5′ end, and 5′-biotin-caagcagaagacggcatacgaAAA-3′ for sequencing from the 3′ end, where uppercase and lowercase letters are defined as before, and xxXX indicates the barcode sequence) were simultaneously ligated to 1 μg short RNA with T4 RNA Ligase (Takara), as modified from ref. 4. Purified small RNAs ligated (insert of the target length plus linkers) with both linkers were separated from the adaptor dimer (49 nt) on an 8% denaturing PAGE gel.

The purified small RNAs derived from each time point were gathered into two tubes: a 5′ end sample and a 3′ end sample. Then the cDNA synthesis was carried out, using 3′ RT-PCR primer (5′ end sequence being 5′-caagcagaagacggcatacga-3′ and 3′ end sequence being 5′-acaggttcagagttctacag-3′) with M-MLV RT, RNase H Minus, and Point Mutant (Promega). The cDNA derived from small RNA were amplified by PCR using adaptor-specific primers: primer 1, 5′-caagcagaagacggcatacga-3′ and primer 2, 5′-acaggttcagagttctacag-3′. PCR was performed from 10 μL template RT product, 10 μL 5× buffer, 12 μL 2.5 mM dNTPs, 0.75 μL 100 μM primer 1, 0.75 μL 100 μM primer 2, 1.5 μL DMSO, and 0.5 μL Phusion DNA High-Fidelity DNA Polymerases (2 U/μL; Finnzyme) in a total volume of 50 μL. After incubating at 98 °C for 30 s, 10 PCR cycles were performed for 10 s at 98 °C and 10 s at 60 °C, followed by 5 min at 70 °C.

PCR products were purified by phenol and chloroform, separated on a 12% polyacrylamide gel, and excised and eluted. The eluted cDNAs were amplified by PCR using the primer 3 (5′-aatgatacggcgaccaccgacaggttcagagttctacag-3′) and primer 4 (5′-acaggttcagagttctacag-3′) and the adaptor sequences, which are necessary for Solexa sequencing, were added. The second PCR was performed from 2 μL template first PCR product, 10 μL 5× buffer, 12 μL 2.5 mM dNTPs, 0.75 μL 100 μM primer3, 0.75 μL 100 μM primer 4, 1.5 μL DMSO, and Phusion DNA High-Fidelity DNA Polymerases (2 U/μL; Finnzyme) in a total volume of 50 μL. After incubating at 98 °C for 30 s, 12 PCR cycles were performed for 10 s at 98 °C and 10 s at 60 °C, followed by 5 min at 70 °C.

Short-RNA tags were further purified on a 12% polyacrylamide gel and the desired fraction (short RNAs between 12 and 52 nt in length) was cut out of the gel and crushed and eluted. Short-RNA tags were mixed and then sequenced on a Illumina Solexa Genome Analyzer.

For the 8tp library, the RNA mix was divided into two equal parts and two different RT enzymes were used on those (Primescript and SuperScript III), but because our quality control process found no significant difference in quality between the two runs, the library variants were merged during computational analysis. The number of first and second PCR cycles for this library was 18 and 6 cycles, respectively. The number of PCR cycles was chosen based on the yield, evaluated as the depth of the band on the gel image.

miRNA Prediction. Several algorithms have been created to predict miRNAs, based on the biochemical properties of putative miRNAs and pre-miRNAs as well as mapped short RNAs (5, 6). We used the miRDeep core algorithm (5) to find possible previously unidentified miRNAs in all tags not annotated as noncoding RNAs. For each tag, we took the genomic sequence surrounding it (22 nt upstream and 52 nt downstream) in both the sense and antisense orientation and used RNAfold (7) to produce the folding information for the resulting RNA sequences as required by miRDeep. Putative miRNAs with a total count of less than 10 tags were dropped to reduce the probability of including artifacts of computational miRNA prediction rather than biologically significant hits in subsequent validation experiments. Hits overlapping ncRNAs or exons were also dropped because these were likely to be spurious hits. We validated the remaining hits by quantitative PCR (qPCR) experiments using culture samples from the 6tp experiment using the method described in ref. 8 (Fig. S2). The sequences, read counts, and genomic locations of the validated miRNAs are shown in Fig. S2.

miRNA Validation. To analyze estrogen-regulated expression of short RNA, MCF7 cells were incubated for 3 d in phenol red-free DMEM containing 5% (vol/vol) dextran charcoal-stripped FCS to deprive them of estrogen. Then, 100 nM E2 was added to the culture medium and total RNA was isolated after 0, 2, 4, 8, 12, and 24 h using ISOGEN reagent (Nippon Gene) based on a modified RNA precipitation method using ethanol instead of isopropanol. Polyadenylation and subsequent cDNA synthesis was performed using 10 μg total RNA and the RTQ primer as described in ref. 8. qPCR was performed with the RTQ-UNIr primer and each of the specific primers for short RNA using Fast SYBR Green Master Mix (Applied Biosystems) on the StepOne Real-Time PCR System (Applied Biosystems). A three-step PCR protocol (95 °C for 10 min, then 40 cycles of 95 °C for 15 s, 50 °C for 30 s, and 60 °C for 30 s) was used. The sequences of PCR primers were as follows: RTQ-UNIr, CGAATTCTAGAGCTCGAGGCAG; and GAPDH (glyceraldehyde-3-phosphate dehydrogenase) mRNA, 5′-GGTGGTCTCCTCTGACTTCAACA-3′ and 5′-GTGGTCGTTGAGGGCAATG-3′. The sequences of PCR primers for novel miRNA are the same as those described in Fig. S2. Evaluation of relative differences of PCR product amounts within a time course was carried out using the comparative cycle threshold (CT) method, using GAPDH mRNA as an endogenous control (9). GAPDH mRNA is abundant and essential in eukaryotic cells, and is unlikely to be affected by E2 treatment. It is therefore used as an internal control transcript in the qRT-PCR analysis of the other target RNAs. Because of this special role in the analysis, its absolute expression is used in the analysis and displayed in the graph, whereas for the miRNAs, relative expression levels are calculated using the absolute expression level of GAPDH mRNA (Fig. S2). The experiments were performed in triplicate from the same pool of RNA.

Microarray Expression Profiling. Expression profiling was performed for two biological replicates each for the negative control (NC), the PAPD4 (PAP associated domain containing 4) knockdown condition, and the PAPD5 (PAP associated domain containing 5) knockdown condition using the Illumina microarray system. Five hundred nanograms of RNA were labeled using the Illumina TotalPrepTM RNA Amplification Kit (Ambion), according to manufacturer's instructions. cRNA was hybridized to Illumina Human Sentrix-6 bead chips Version 3 according to standard Illumina protocols (Illumina, Catalog # BD-901-1002, Part # 11322355 Rev. A). Chips were scanned using the Illumina BeadScan and BeadStudio software packages and summarized data were generated in BeadStudio (Version 3.4.0). Next, the microarray expression data were processed using the lumi (10) and limma (11) packages. We first subtracted the background signal from the raw data, applied quantile normalization, and then transformed the data using the base-2 logarithm. We selected all probes with a detection score of at least 0.99 in both replicates in at least the NC, the PAPD5 knockdown condition, or the PAPD4 knockdown condition. We then required that the SD of the log-transformed expression data for a probe across the six conditions was at least 0.5, which leaves 1,111 probes. Next, we converted the log-transformed expression data to Z scores by subtracting their mean and dividing by their SDs. The probes were k-means clustered using $k = 3$, whereas the experimental conditions were hierarchically clustered using complete linkage. The Euclidean distance was used for both clustering calculations, which were performed using the Cluster 3.0 software (12). To calculate the difference in expression of mRNAs between the NC, the PAPD5 knockdown, and the PAPD4 knockdown, we averaged the expression after normalization and log transformation over the pair of replicates in each experimental condition. The statistical significance of differences in expression in groups of genes between the NC and each knockdown condition were calculated using the Mann–Whitney test (13) comparing target genes to nontarget genes. Target gene predictions were obtained from TargetScan Release 5.2 (14, 15). Our microarray expression data have been deposited in NCBI's GEO database (16) under accession no. GSE57939.

Statistical Analysis of Adenylation/Degradation Ratios and Overrepresentation of Degradation Products

We maximized the log-likelihood function under the negative binomial distribution assuming that the adenylation ratio of a particular miRNA is equal in the knockdown condition and the NC, and under the assumption that the adenylation ratios are different for the knockdown condition and the NC. The difference between the two values of the log-likelihood function was multiplied by 2 and compared with the χ^2 distribution with one degree of freedom to obtain the P value, which we compared with a significance threshold of 5%.

We used a similar approach to analyze differences in the abundance of degradation products by comparing the counts of sequences that match the sequence of the mature miRNA at the 5′ end but are 1, 2, 3, or 4 nt shorter at the 3′ end. For each sequence, we calculated the statistical significance of the difference in counts of the degradation product relative to the mature miRNA between the experimental condition and the NC. We then used Fisher's method to combine the individual P values for degradation products of a specific length into an overall P value for degradation products.

Nucleotidyl Transferase and Nuclease Knockdown Experiments

Cell Culture. The PAPD4 and PAPD5 knockdown samples were produced previously (see ref. 17 for full experimental procedures). In summary, THP1 cells were cultured in RPMI1640 (Invitrogen), 10% (vol/vol) FBS, penicillin/streptomycin (Invitrogen), 10 mM Hepes (Invitrogen), 1 mM sodium pyruvate (Invitrogen), and 50 μM 2-mercaptoethanol (Invitrogen). Reverse transfection of 1×10^6 cells in a 60-mm cell culture dish was performed with 20 nM (final concentration) of each siRNA (see *RNAi*), Opti-MEM (Invitrogen), and 1.6 μg/mL (final concentration) of Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions. Total RNAs were extracted 72 h after transfection with TRIzol (Invitrogen) and FastPure RNA kit (TaKaRa BIO) according to the modified manufacturer's instructions. After phase separation of TRIzol, aqueous phase was mixed with the same amount of 100% ethanol and flowed through a filter cartridge. After RNA binding to the filter cartridge, each bound total RNA was washed three times with 700 μL triple-ethanol-diluted wash buffer and eluted with elution buffer. RNA was quantified with NanoDrop (NanoDrop Technologies).

RNAi. Stealth siRNAs designed by Invitrogen were used for RNAi knockdown of nucleotidyl transferase genes PAPD4 and PAPD5 and the nuclease genes CNOT6L (CCR4-NOT transcription complex, subunit 6-like), ERI1 (exoribonuclease 1), PAN3 [PAN3 poly(A) specific ribonuclease subunit], PARN [poly(A)-specific ribonuclease], RNASEL [ribonuclease L (2′,5′-oligoisoadenylate synthetase-dependent)], EXOSC4 (exosome component 4), and EXOSC10 (exosome component 10). We attempted RNAi with several distinct stealth siRNA against the target genes, and selected RNA samples for small-RNA library construction based on the knockdown efficiency and RNA sample quality. Stealth RNAi Negative Control Medium GC Duplex #2 (Invitrogen) was used as NC siRNA. The siRNAs that were selected for each gene are listed in the table below. qRT-PCR verification confirmed that the knockdown was successful (Fig. S8*C*). Knockdown of PAPD4 and PAPD5 was also confirmed by the microarray expression profiling data for these transcripts (Fig. S8*D*).

siRNAs Used to Knock Down Nucleotidyl Transferase and Nuclease Genes. For the knockdown of the nucleotidyl transferases PAPD4 and PAPD5 and the nuclease genes CNOT6L, ERI1, EXOSC4, EXOSC10, PAN3, PARN, and RNASEL, the stealth siRNAs (Invitrogen) that were used are shown in Table S3.

Quantification of miR-21 Expression by qPCR. The expression level of miR-21 in THP1 cells upon knockdown of PAPD5 and in the corresponding NC condition was measured using the TaqMan assay (Applied Biosystems).

PAPD5 Overexpression Experiment and Analysis

cDNA and Virus Preparation. Gateway-compatible human full-length cDNA PAPD5 entry clone (OHS4559) was purchased from OpenBiosystems. The production and the titer check of lentivirus were performed as described previously (18). Briefly, the PAPD5 entry clone was recombined into pENTR lentivirus vector (CSII-EF-RfA-IRES2-VENUS) using Gateway LR clonaseII enzyme mix (Invitrogen). After transformation into OneShot Stbl3 competent *Escherichia coli* (Invitrogen), plasmids derived from a single colony were expanded and purified using PureYield Plasmid Midiprep System (Promega). For every 8.5 μg plasmid, 5 μg HIV-glycoprotein and 5 μg vesicular stomatitis virus envelope genes were cotransfected onto 4×10^6 293T cells (prepared the day before at 37 °C, 10% CO_2) using FuGeneHD (Roche) in OPTI-MEM (WAKO) medium containing 5% (vol/vol) FBS at 37 °C 5% CO_2. The virus was concentrated and stored at −80 °C for later use.

Cell Culture and Virus Transduction. Human acute monocytic leukemia cell line, THP1, was cultured in RPMI1400 (WAKO) containing 10% (vol/vol) FBS/5 M Hepes/2.5 M sodium pyruvate/3.4 μg β-mercaptoethanol/L-glutamine/antibiotics (GIBCO) at 37 °C 5% CO_2. Human breast adenocarcinoma cell line, MCF7, was cultured in DMEM containing 10% (vol/vol) FBS/L-glutamine/antibiotic (WAKO) at 37 °C 5% CO_2. The vector control and PAPD5 lentivirus were transduced at 10 multiplicity of infection containing 80 mg/mL Polybrene (SIGMA) and 2 wk posttransduction, >90% of MCF7 cells were positive for Venus expression, whereas the transduction rate for THP1 was <70%. Fluorescence-activated cell sorting (FACS) was performed on a BD FACS Aria II (BD) to sort the Venus+ THP1 cells.

RNA Isolation and qPCR. Total RNA was extracted 2 wk post-transduction using miRNAeasy kit (Qiagen) according to the manufacturer's instructions. RT of total RNA was achieved with PrimeScript Reverse Transcriptase (Takara) and random hexamers in accordance with the manufacturer's protocol. Primers targeting beta actin (ACTB) sense: CCAACCGCGAGAAGATGA; antisense: CCAGAGGCGTACAGGGATAG was used as a control for data normalization and the PAPD5 primer sense: ATGCAATGGTCCAGTGTCGTCCT and antisense: GGTTCCCAGTGGACGGACGG was used to confirm the virus transduction. PCR amplification was performed on an ABI PRISM 7500 Sequence Detection System (Applied Biosystems). For amplification, SYBR Premix Ex Taq II (Takara) was used as instructed in the manual. Changes of gene expression were determined using the $2^{-\Delta\Delta Ct}$ method.

Experimental Validation of miR-21 Isoforms by qRT-PCR. We used the method described by Kumar et al. (19) to specifically detect the relative expression level of the miR-21 and miR-21+C isomiRs. In brief, 400 ng total RNA was used as input to a T4 RNA ligase circularization reaction. Two microliters of each reaction were used for miR-specific RT. The specific RT primers used were TCTGATAAGCTAT for miR-21 and CTGATAAGCTAGT for miR-21C. PCR primers were AGCTTATCAGACTGATGTTGAT (forward) and AACATCAGTCTGATAAGCTATC (reverse) for miR-21 and AGCTTATCAGACTGATGTTGACT (forward) and AACATCAGTCTGATAAGCTAGTC (reverse) for miR-21+C. A single circulation/RT reaction was performed and two PCR reactions were performed on each RT. The specificity of the detection by this method was confirmed using synthetic miR-21 and synthetic miR-21+C.

PAPD5 Overexpression and PARN Knockdown Experiment in MCF7

To assess the combined effect of PAPD5 overexpression and PARN knockdown, we subjected MCF7 cells to the following conditions, with three replicates in each condition:

i) Negative control
ii) PARN knockdown by siRNA transfection
iii) Scrambled siRNA transfection
iv) PAPD5 overexpression by lentivirus transduction
v) Empty lentivirus transduction
vi) PAPD5 overexpression by lentivirus transduction, followed by PARN knockdown by siRNA
vii) Empty lentivirus transduction, followed by PARN knockdown by siRNA
viii) PAPD5 overexpression by lentivirus transduction, followed by scrambled siRNA transfection

In conditions *ii* and *iii*, RNA was extracted from cells 48 h after siRNA transfection. In conditions *iv* and *v*, RNA was extracted from cells 8 d after lentivirus transduction. In conditions *vi*–*viii*, the cells were subjected to siRNA transfection 8 d after the lentivirus transduction, followed by RNA extraction 48 h later. Cell culture, RNAi, and lentivirus transduction were performed as described above.

Library Preparation and Sequencing

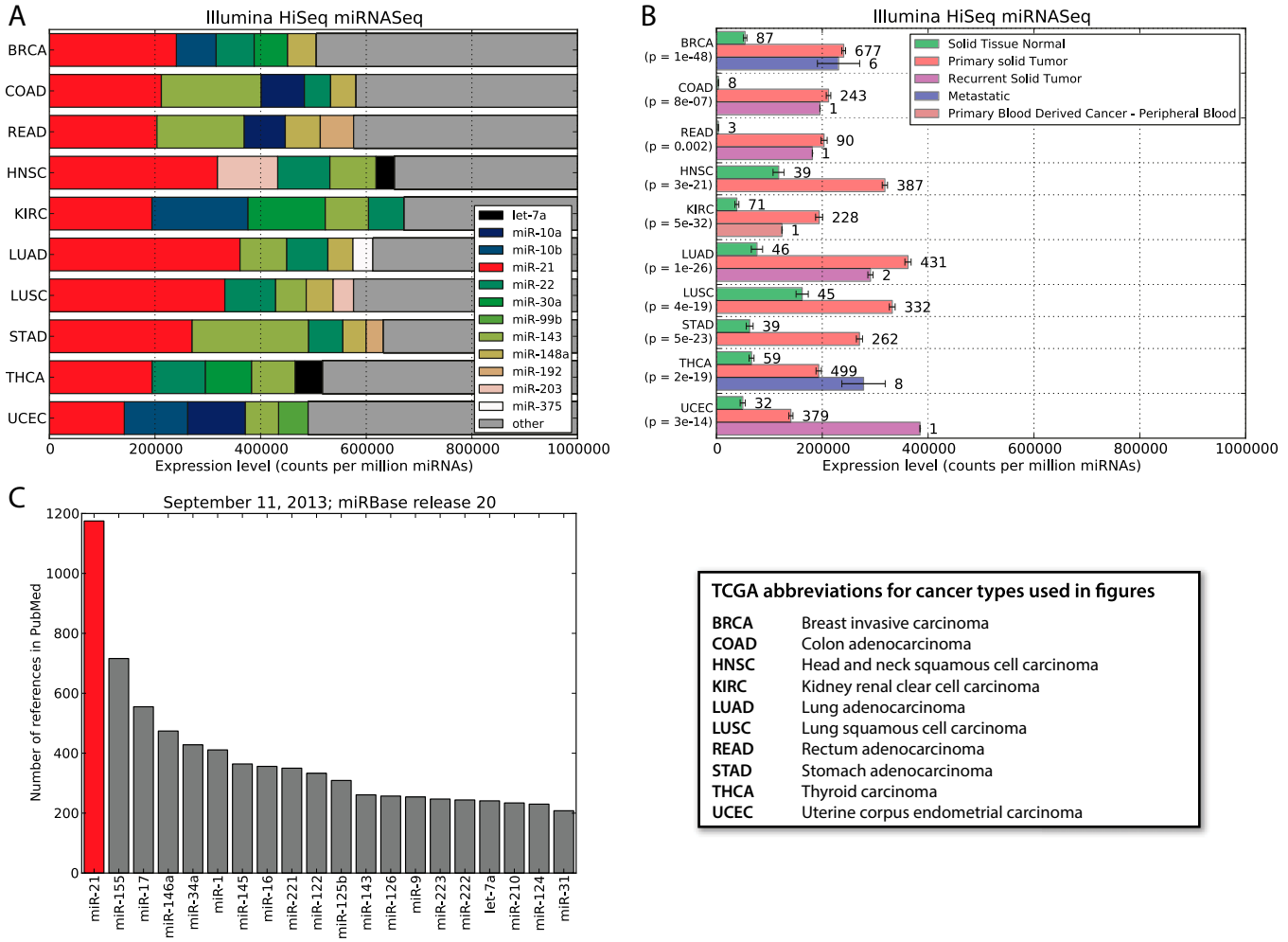
Sequencing libraries were produced using Illumina's TruSeq Small RNA Sample Preparation Kit and sequenced using Illumina's HiSeq2000 sequencer.

Evaluation of the Photoactivatable-Ribonucleoside-Enhanced Cross-Linking and Immunoprecipitation Data

In photoactivatable-ribonucleoside-enhanced cross-linking and immunoprecipitation (PAR-CLIP) sequencing data, T-to-C mismatches relative to the reference sequence are known to be enriched at cross-linking sites. To assess whether PAPD5 and miR-21 directly interact, we counted sequence reads that exactly align to the 5′ end of miR-21 as well as those with a single T-to-C mismatch in the two replicates of a previously published PAPD5 PAR-CLIP experiment in HEK293 cells (20). As a reference dataset, we used previously published (21) sequencing data of a PAR-CLIP experiment in HEK293 cells for the IGF2BP1, IGF2BP2, and IGF2BP3 proteins, which, to our knowledge, do not interact with mature miRNAs. Using the counts for the IGF2BP proteins, we calculated the background probability of finding a single T-to-C mismatch in the sequenced reads for mature miR-21 or its degradation products. We then calculated the statistical significance of the number of sequence reads with a single T-to-C mismatch in the PAPD5 conditions using the binomial distribution. The detailed calculation is shown in Table S2.

1. Kozomara A, Griffiths-Jones S (2014) miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 42(Database issue):D68–D73.
2. de Hoon MJ, et al. (2010) Cross-mapping and the identification of editing sites in mature microRNAs in high-throughput sequencing libraries. *Genome Res* 20(2):257–264.
3. Lagonigro MS, et al. (2004) CTAB-urea method purifies RNA from melanin for cDNA microarray analysis. *Pigment Cell Res* 17(3):312–315.
4. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T (2001) Identification of novel genes coding for small expressed RNAs. *Science* 294(5543):853–858.

5. Friedländer MR, et al. (2008) Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol* 26(4):407–415.
6. Hendrix D, Levine M, Shi W (2010) miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data. *Genome Biol* 11(4):R39.
7. Hofacker IL, et al. (1994) Fast folding and comparison of RNA secondary structures. *Monatsh Chem* 125(2):167–188.
8. Ro S, Park C, Jin J, Sanders KM, Yan W (2006) A PCR-based method for detection and quantification of small RNAs. *Biochem Biophys Res Commun* 351(3):756–763.
9. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25(4):402–408.
10. Du P, Kibbe WA, Lin SM (2008) lumi: A pipeline for processing Illumina microarray. *Bioinformatics* 24(13):1547–1548.
11. Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3(1):Article3.
12. de Hoon MJ, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. *Bioinformatics* 20(9):1453–1454.
13. Snedecor GW, Cochran WG (1989) *Statistical Methods* (Iowa State Univ Press, Ames, IA), 8th Ed, pp 142–144.
14. Friedman RC, Farh KK, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 19(1):92–105.
15. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. *Cell* 115(7):787–798.
16. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30(1):207–210.
17. Burroughs AM, et al. (2010) A comprehensive survey of 3′ animal miRNA modification events and a possible role for 3′ adenylation in modulating miRNA targeting effectiveness. *Genome Res* 20(10):1398–1410.
18. Shin JW, et al. (2012) Establishment of single-cell screening system for the rapid identification of transcriptional modulators involved in direct cell reprogramming. *Nucleic Acids Res* 40(21):e165.
19. Kumar P, Johnston BH, Kazakov SA (2011) miR-ID: A novel, circularization-based platform for detection of microRNAs. *RNA* 17(2):365–380.
20. Rammelt C, Bilen B, Zavolan M, Keller W (2011) PAPD5, a noncanonical poly(A) polymerase with an unusual RNA-binding motif. *RNA* 17(9):1737–1746.
21. Hafner M, et al. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141(1):129–141.



TCGA abbreviations for cancer types used in figures

BRCA	Breast invasive carcinoma
COAD	Colon adenocarcinoma
HNSC	Head and neck squamous cell carcinoma
KIRC	Kidney renal clear cell carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
READ	Rectum adenocarcinoma
STAD	Stomach adenocarcinoma
THCA	Thyroid carcinoma
UCEC	Uterine corpus endometrial carcinoma

Fig. S1. Analysis of miRNA expression in data from The Cancer Genome Atlas (TCGA) (1), and miR-21 significance in current literature. (*A*) Top-expressed miRNAs in different cancer types. For each cancer type for which Illumina HiSeq small-RNA sequencing data are available in TCGA, we calculated the average expression level (in counts per million miRNAs) of miRNAs across tumor samples, and show the top five expressed miRNAs in each cancer type. miR-21 is the most highly expressed miRNA in most cancer types. (*B*) Differential expression of miR-21 in cancer. For all 10 cancer types for which Illumina HiSeq small RNA-sequencing data of both tumor and normal samples are available in TCGA, we calculated the statistical significance of differential expression of miRNA miR-21 using the Kruskal–Wallis nonparametric test. We found that miR-21 is significantly more highly expressed in cancer compared with normal tissue in all of these cancer types. Error bars indicate the SD of the estimated mean. (*C*) Number of references in PubMed (www.ncbi.nlm.nih.gov/pubmed) per miRNA. On September 11, 2013, we counted the number of entries in PubMed that mention each miRNA in miRBase (Release 19) (2), and found that miR-21 is by far the most referenced miRNA in the scientific literature.

1. The Cancer Genome Atlas. Available at http://cancergenome.nih.gov/. Accessed August 29, 2013.
2. Kozomara A, Griffiths-Jones S (2014) miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 42(Database issue):D68–D73.

A

internal name	miRNA name	sequence of mature miRNA	matched reads	chromosome	strand	start	end	remarks	qPCR validation
s1	hsa-mir-9898	TACTTACCTGTCCCCTACCCCA	46	chr12	-	53292708	53292686	Intron of KRT8, putative 5'tailed mirtron	1x6 + 3x4
s2	hsa-mir-9899	CGGGCGCCGCGCTCCCGCCCGC	31	chr2	-	150186835	150186814	Intron of LYPD6	1x6
s3	hsa-mir-9900	ACAGGTCCTAAGAGACTGCAT	41	chr3	+	112455987	112456007	Intergenic	1x6 + 3x4
s4	hsa-mir-9901	CGGTCGCCGCGGTTCGCCGCC	67	chr16	+	33966068	33966088	Intergenic	1x6
s5	hsa-mir-9902	CCCAGAAATCTGGTATGCCAGC	125	chr12	-	123180532	123180511	Intron of GPR81/ERV1-repeat	1x6 + 3x4
s6	hsa-mir-9903	TTATCCTCCAGTAGACTAGGGA	31	chr8	-	99405916	99405895	Intergenic/MaLR-repeat	1x6
s7	hsa-mir-548aw	CGGTGCAAAAGTCATCACGGTTT	29	chr9	+	135821099	135821121	Intergenic/MADE1-repeat	1x6

B

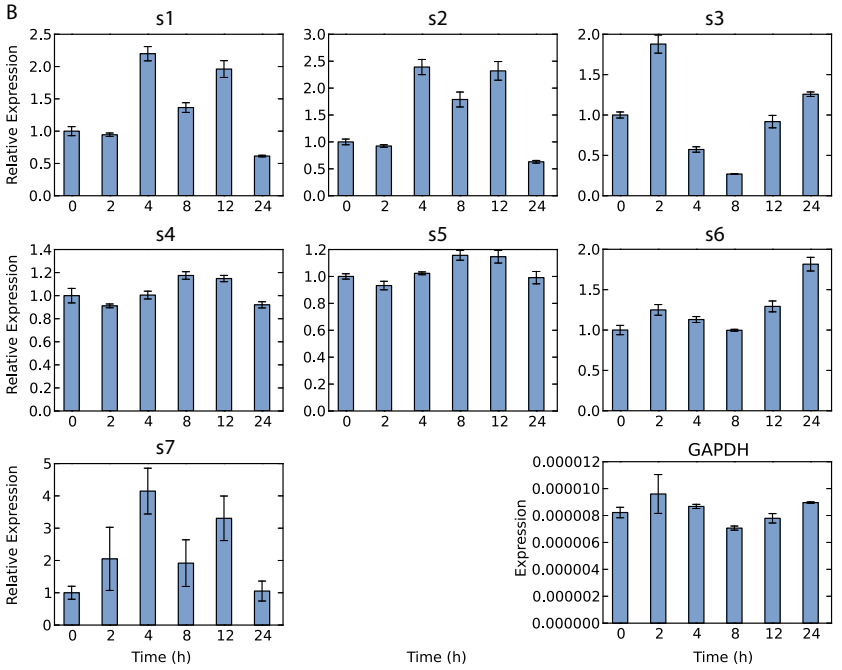


Fig. S2. Prediction of miRNAs in MCF7 and their validation by qPCR. (*A*) Deep-sequencing data of estrogen-stimulated MCF7 cells were aligned to known genomic and transcriptomic features. Subsequent analysis of the remaining reads with miRDeep lead to the discovery of 7 previously unidentified miRNAs (see *SI Materials and Methods* for details). For each of these miRNAs, its sequence and coordinates on the hg19 genome assembly are shown. In addition, information about its genomic surroundings is displayed, such as its host gene. All miRNAs in *A* were confirmed using qPCR. (*B*) For each miRNA shown in *A*, qRT-PCR experiments were performed in triplicate from the same pool of RNA, both at $t = 0$ and at five time points after stimulation with 100 nM E2. GAPDH mRNA was used for normalization, as described in ref. 1, and is shown in *Bottom Right*. Displayed values for these miRNAs are relative to those at $t = 0$. Error bars represent the standard deviation in the measured expression value.

1. Lagonigro MS, et al. (2004) CTAB-urea method purifies RNA from melanin for cDNA microarray analysis. *Pigment Cell Res* 17(3):312–315.

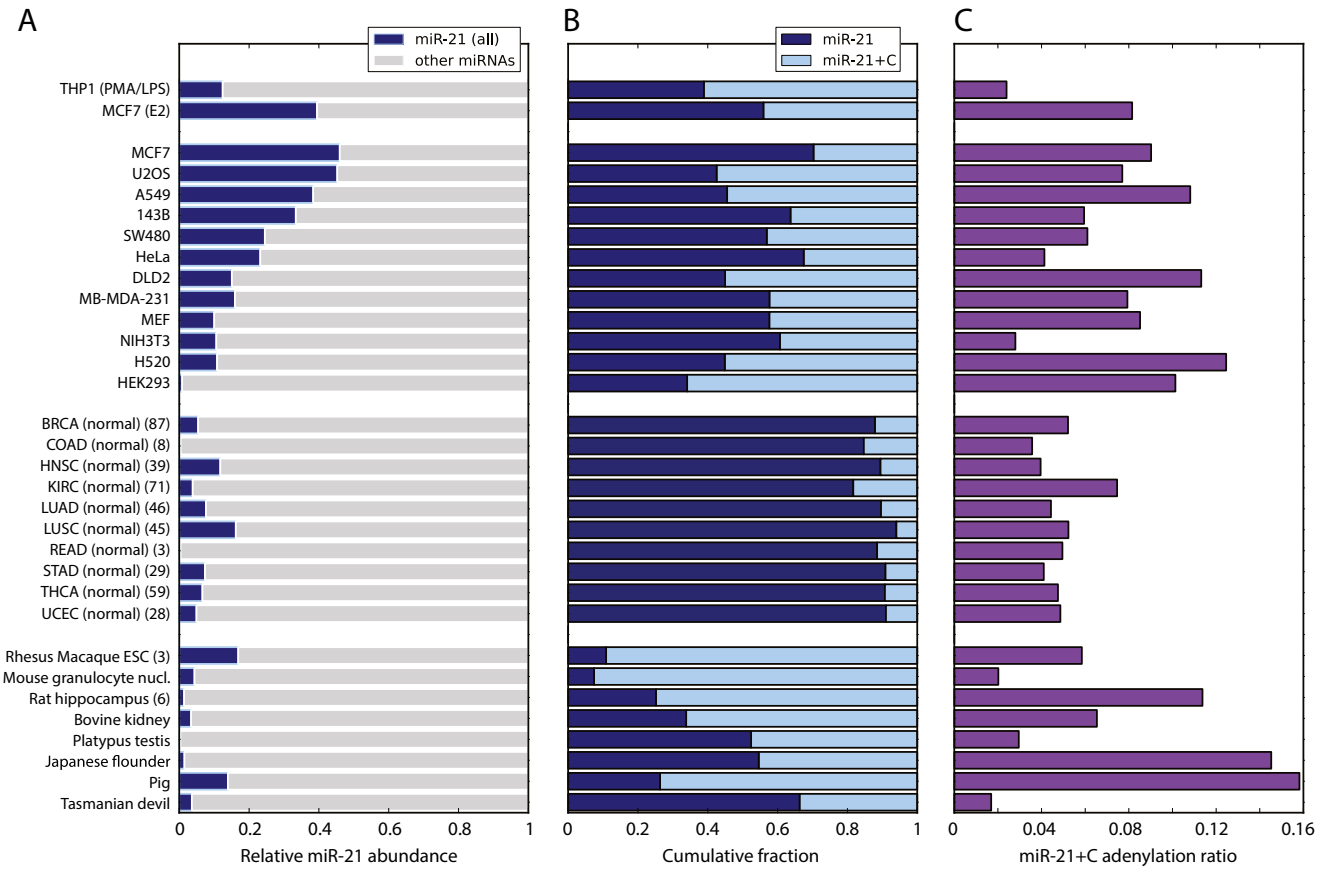


Fig. S3. Expression of miR-21 and its isomiRs across samples and organisms. miR-21 abundance in various cell types (*A*) ranges from less than 0.1% for normal rectum cells (READs) to 39% for MCF7 cells. The degradation ratio (*B*) and adenylation ratios (*C*) (as defined in Fig. 1) also vary greatly across cell types. THP1, MCF7, U2OS, A549, 143B, SW480, HeLa, DLD2, MB-MDA-231, and H520 are cancer cell lines; MEF, HEK293, and "Rhesus macaque ESC" are embryonic cells; BRCA, COAD, HNSC, KIRC, LUAD, LUSC, READ, STAD, THCA, and UCEC are normal (noncancerous) primary tissue samples; and the other samples are healthy cells. The dataset labeled "MCF7 (E2)" was sequenced after stimulation of MCF7 with 17β-estradiol, whereas the dataset labeled "MCF7" is of an unperturbed MCF7 sample. See *SI Materials and Methods* for a detailed list of the datasets used in this study and Fig. S1 for a list of abbreviations and their definitions.

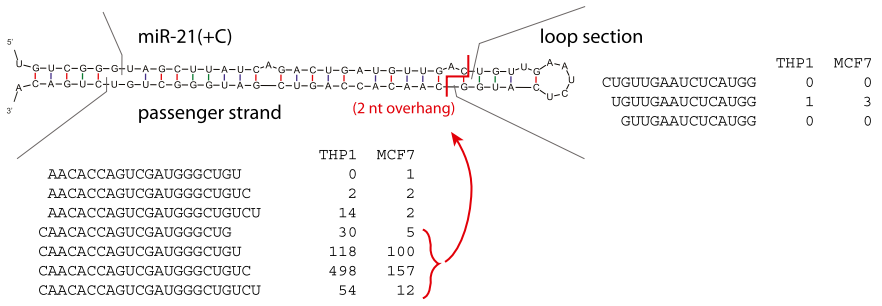


Fig. S4. The secondary structure of pre–miR-21, showing passenger strand and loop section. The loop and passenger sequences produced when the mature miRNA is excised from pre–miR-21 can be detected in deep-sequencing data, although they are not very abundant as they are degraded rapidly. In the replicate $t = 0$ (unstimulated) samples from our time course sequencing data from MCF7 cells, as well as in sequencing data from THP1 cells (1), from which read counts are shown in this figure, the loop section always starts at the uracil that follows miR-21+C, and never at the cytosine that would follow canonical miR-21 if that isomiR were excised by DICER1, suggesting that miR-21+C is excised by DICER1. The passenger strand sequences are consistent with this, as the vast majority of these (974 of 995) would result in a 2-nt overhang with respect to miR-21+C (indicated in red), whereas only 21 reads start at the adenine that would produce a 2-nt overhang compared with canonical miR-21.

1. Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W (2004) Single processing center models for human Dicer and bacterial RNase III. *Cell* 118(1):57–68.

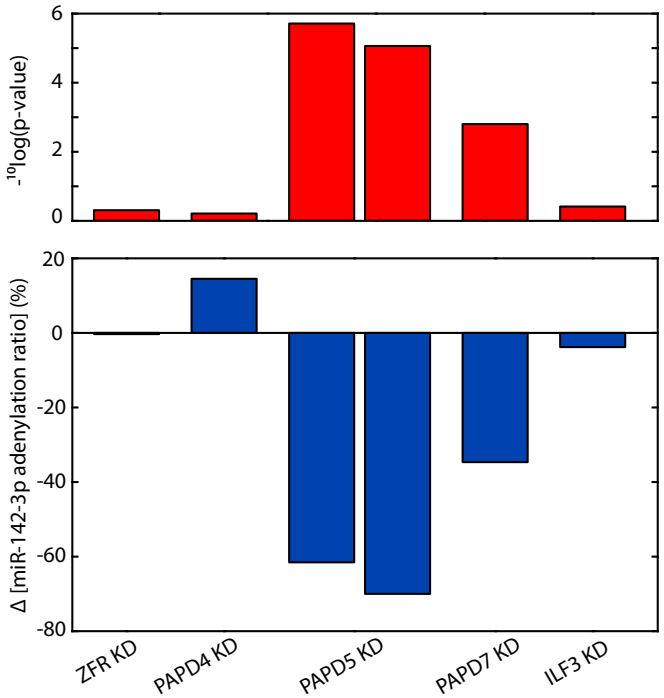


Fig. S5. miR-142-3p adenylation ratios (*Upper*) and corresponding statistical significance (*Lower*) after knockdown of candidate adenylating enzymes in THP1 cells. Besides miR-21+C, miR-142-3p was the only miRNA whose adenylation ratio was significantly decreased in both replicates of the PAPD5 knockdown experiment (P = 1.9e-6 and P = 8.7e-6).

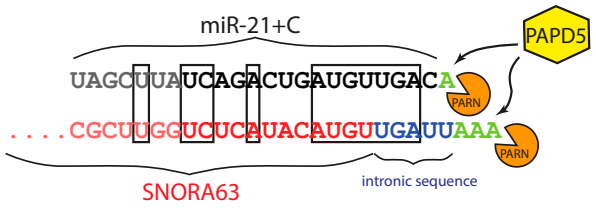


Fig. S6. Adenylation by PAPD5 and trimming by PARN of SNORA63 (small nucleolar RNA, H/ACA box 63). As part of the maturation pathway of SNORA63 and other small nucleolar RNAs (snoRNAs), oligoadenylation of the 3′ end of the snoRNA by PAPD5 is followed by 3′-to-5′ trimming by the exonuclease PARN to produce the mature snoRNA (1, 2). Interestingly, 7 nt near the 3′ end of miR-21 are identical to a stretch of nucleotides that span the exon–intron boundary at the 3′ end of the immature SNORA63 transcript. However, we did not find this sequence pattern at the 3′ end of miR-142-3p, the only other apparent miRNA target of PAPD5 in our data, nor in the other snoRNAs targeted by PAPD5.

1. Berndt H, et al. (2012) Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA* 18(5):958–972.
2. Virtanen A, Henriksson N, Nilsson P, Nissbeck M (2013) Poly(A)-specific ribonuclease (PARN): An allosterically regulated, processive and mRNA cap-interacting deadenylase. *Crit Rev Biochem Mol Biol* 48(2):192–209.

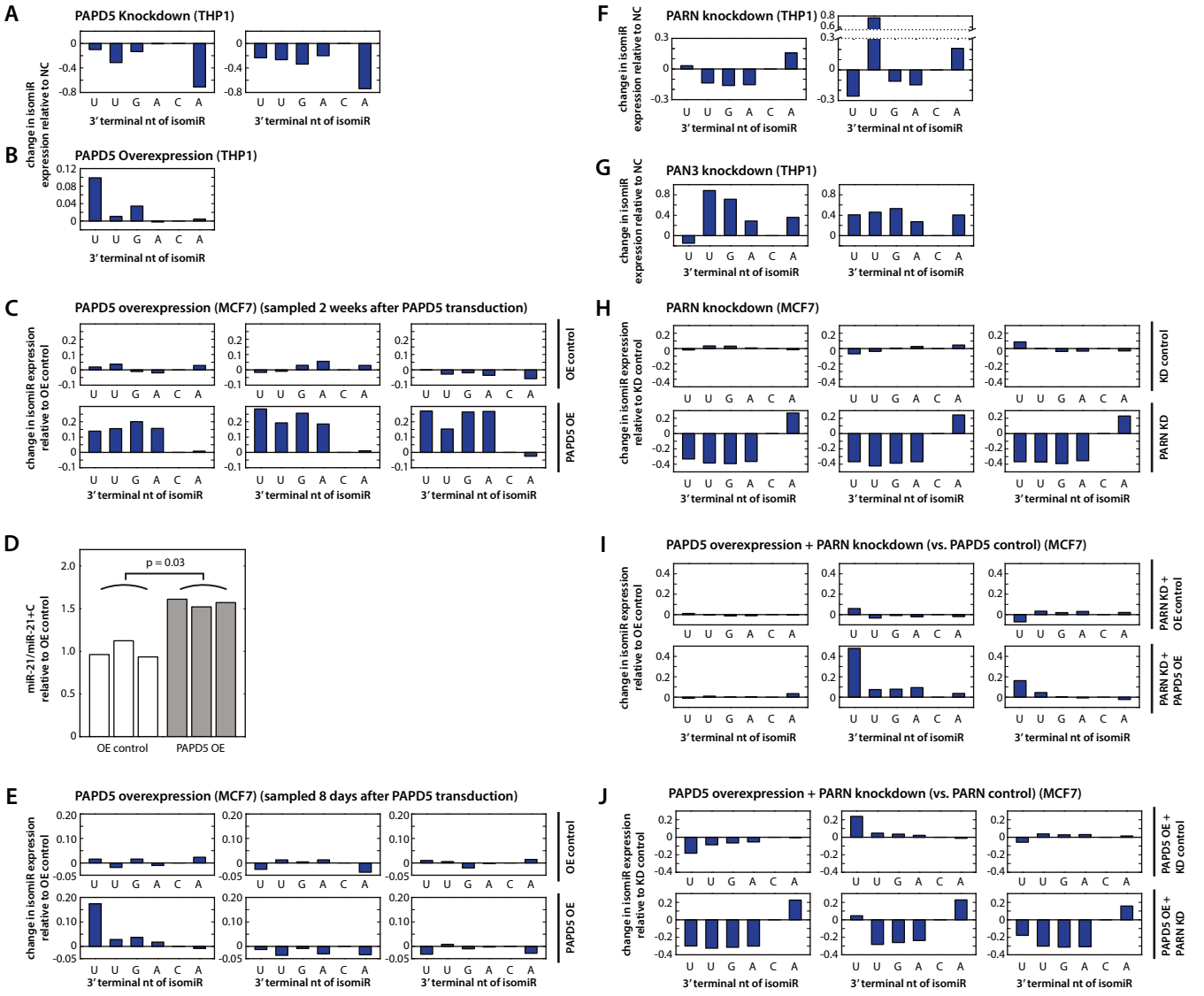


Fig. S7. Expression of miR-21 isoforms in the knockdown and overexpression experiments. (*A–J*) The expression levels of miR-21 isomiRs were normalized to the expression level of miR-21+C. Each panel shows the change in the normalized isomiR expression relative to the average of the normalized expression in the corresponding NC condition. In *A–C* and *E–J*, miR-21 isomiRs are denoted by their 3′ terminal nucleotide. *I* and *J* refer to the same experiment, in which overexpression of PAPD5 in MCF7 cells by Lentivirus transduction was followed by knockdown of PARN by siRNA. (*A*) Expression of miR-21 isomiRs as measured by short-RNA sequencing upon knockdown of PAPD5 in THP1 cells (two replicates). (*B*) Expression of miR-21 isomiRs as measured by short-RNA sequencing upon overexpression of PAPD5 in THP1 cells. (*C*) Expression of miR-21 isomiRs as measured by short-RNA sequencing upon overexpression (OE) of PAPD5 in MCF7 cells, sampled 2 wk after PAPD5 transduction (three replicates). (*D*) Expression of the 22-nt miR-21 isomiR, normalized to the expression of the miR-21+C isomiR, upon overexpression of PAPD5 in MCF7 cells (three replicates), measured by isomiR-specific qPCR. (*E*) Expression of miR-21 isomiRs as measured by short-RNA sequencing upon overexpression of PAPD5 in MCF7 cells, sampled 8 d after PAPD5 transduction (three replicates). (*F*) Expression of miR-21 isomiRs as measured by short-RNA sequencing upon knockdown of PARN in THP1 cells (two replicates). The second replicate showed an outlier in the expression of the 20-nt degradation product. This outlier was not observed in the first replicate, nor in the three replicates for PARN knockdown in MCF7 cells (*H*). (*G*) Expression of miR-21 isomiRs as measured by short-RNA sequencing upon knockdown of PAN3 in THP1 cells (two replicates). (*H*) Expression of miR-21 isomiRs as measured by short-RNA sequencing upon knockdown of PARN in MCF7 cells (three replicates). (*I*) Expression of miR-21 isomiRs as measured by short-RNA sequencing upon overexpression of PAPD5 followed by knockdown of PARN in MCF7 cells (three replicates), normalized to PAPD5 OE control to show the effect of PAPD5 overexpression. (*J*) Expression of miR-21 isomiRs as measured by short-RNA sequencing upon overexpression of PAPD5 in MCF7 cells followed by knockdown of PARN (three replicates), normalized to PARN KD control to show the effect of PARN knockdown.

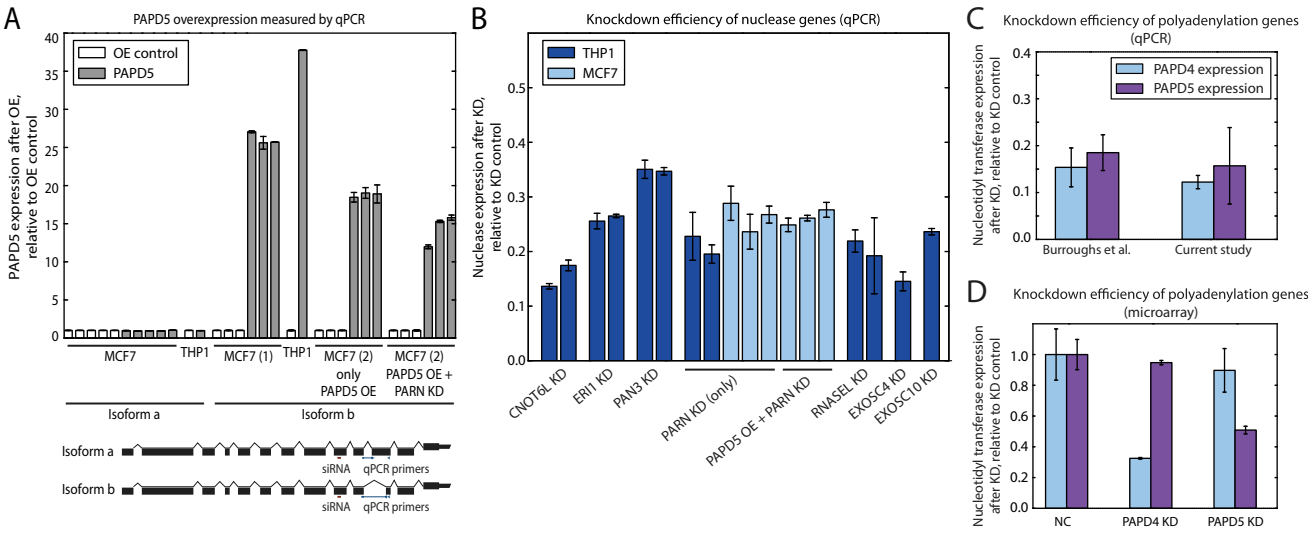


Fig. S8. qPCR validation of knockdown and overexpression experiments. (*A*) Validation of PAPD5 overexpression in MCF7 and THP1. In both MCF7 and THP1 cells, PAPD5 was overexpressed by at least 10-fold compared with mock transfected cells. Our qPCR results confirm that only the transfected "b" isoform of PAPD5 is overexpressed, whereas the expression of the "a" isoform is unchanged. A comparison between the exon structure of the two PAPD5 mRNAs is shown in *B*, along with the primers used to amplify each respective isoform and the siRNA used for the knockdown experiments. For visualization purposes, 3′ UTRs are truncated in the figure; primers spanning an exon-exon boundary are represented by thinner lines where they cross an intron. The RefSeq (www.ncbi.nlm.nih.gov/refseq/) accession for isoform a is NM_001040284 and NM_001040285 for isoform b. (*B*) qPCR validation of knockdown of candidate nucleases. Expression of the mRNAs of the nucleases CNOT6L, ERI1, PAN3, PARN, RNASEL, EXOSC4, and EXOSC10 as measured by qRT-PCR, confirmed that their knockdown was successful. Similarly, the knockdown of PARN in mock-transfected MCF7 cells and in MCF7 cells upon PAPD5 overexpression was confirmed by qPCR. In each case, the expression level of the exoribonuclease is shown relative to the corresponding NC condition. (*C* and *D*) Validation of PAPD4 and PAPD5 knockdown before mRNA expression profiling by microarray. Expression profiling was performed for two biological replicates each for the NC, the PAPD4 knockdown condition, and the PAPD5 knockdown condition. The expression of the PAPD4 and PAPD5 mRNAs, as measured by qRT-PCR (*C*) and microarray (*D*), indicated successful knockdown of these nucleotidyl transferases. In all panels, the error bars show the standard deviation in the measured expression value.

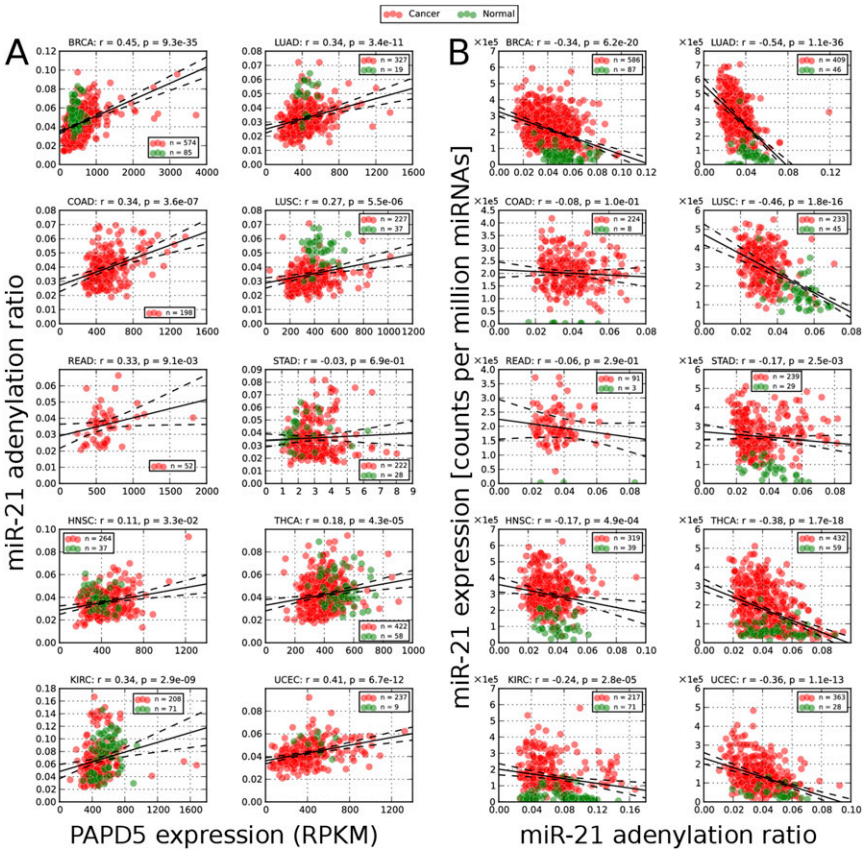


Fig. S9. Correlations between PAPD5 expression, miR-21 expression, and miR-21 adenylation in 10 cancer types from TCGA (1). (*A*) In all cancer types except STAD, a statistically significant positive correlation was found between PAPD5 expression levels and miR-21 adenylation ratio. (*B*) In all cancer types except COAD and READ, a statistically significant negative correlation was found between miR-21 adenylation ratio and miR-21 expression level. The Spearman rank correlations and their corresponding significance values are shown. Dashed lines indicate the 95% confidence region of the regression line.

1. The Cancer Genome Atlas. Available at http://cancergenome.nih.gov/. Accessed August 29, 2013.

Table S1. Overview of sequencing data sources used in this study

Source	Description/cell type	Database ID	Platform
(1)	Japanese flounder	GSE25995	IGA
(2)	Cow	GSE15450	IGA
(3)	Psoriasis	GSE31037	IGA IIx
(4)	Pig	GSE17885	IGA
(5)	Various cell lines (10 human, 2 mouse)	GSE16579	IGA
(6)	Platypus	GSE10571	IGA
(7)	Tasmanian devil facial tumor	GSE18352	IGA
(8)	Rat hippocampus	DRX000640-645	IGA IIx
(9)	Rhesus macaque embryonic stem cells	GSE27886	IGA
(10)	Mouse granulocyte nuclei	GSE20664	IGA IIx
(11)	RIKEN THP1	AIAAA0000001–AIAAT0061351	Roche FLX
(12)	Cancerous and healthy primary tissue samples obtained from TCGA	All available HiSeq miRNA sequencing data for BRCA, COAD, READ, HNSC, KIRC, LUAD, LUSC, STAD, THCA, and UCEC	HiSeq
(13)	THP1 nucleotidyl transferase knockdown samples, including the first PAPD5 KD sample	DRA000205	IGA II
This study	THP1, second PAPD5 KD sample	DRA002263	HiSeq
This study	THP1 nuclease knockdown samples	DRA002262, DRA002263	HiSeq
This study	RIKEN MCF7	DRA002268, DRA002269	IGA
This study	MCF7 and THP1 PAPD5 overexpression	DRA002260, DRA002261	HiSeq
This study	MCF7 PAPD5 overexpression and PARN knockdown samples	DRA002264, DRA002266	HiSeq

RIKEN, The Institute of Physical and Chemical Research.

1. Fu Y, et al. (2011) Identification and differential expression of microRNAs during metamorphosis of the Japanese flounder (Paralichthys olivaceus). *PLoS ONE* 6(7):e22957.
2. Glazov EA, et al. (2009) Repertoire of bovine miRNA and miRNA-like small regulatory RNAs expressed upon viral infection. *PLoS ONE* 4(7):e6349.
3. Joyce CE, et al. (2011) Deep sequencing of small RNAs from human skin reveals major alterations in the psoriasis miRNAome. *Hum Mol Genet* 20(20):4025–4040.
4. Li M, et al. (2010) MicroRNAome of porcine pre- and postnatal development. *PLoS ONE* 5(7):e11541.
5. Mayr C, Bartel DP (2009) Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138(4):673–684.
6. Murchison EP, et al. (2008) Conservation of small RNA pathways in platypus. *Genome Res* 18(6):995–1004.
7. Murchison EP, et al. (2010) The Tasmanian devil transcriptome reveals Schwann cell origins of a clonally transmissible cancer. *Science* 327(5961):84–87.
8. Shinohara Y, et al. (2011) miRNA profiling of bilateral rat hippocampal CA3 by deep sequencing. *Biochem Biophys Res Commun* 409(2):293–298.
9. Sun Z, et al. (2011) MicroRNA profiling of rhesus macaque embryonic stem cells. *BMC Genomics* 12(1):276.
10. Taft RJ, et al. (2010) Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans. *Nat Struct Mol Biol* 17(8):1030–1034.
11. Taft RJ, et al. (2009) Tiny RNAs associated with transcription start sites in animals. *Nat Genet* 41(5):572–578.
12. The Cancer Genome Atlas. Available at http://cancergenome.nih.gov/. Accessed August 29, 2013.
13. Burroughs AM, et al. (2010) A comprehensive survey of 3′ animal miRNA modification events and a possible role for 3′ adenylation in modulating miRNA targeting effectiveness. *Genome Res* 20(10):1398–1410.

Table S2. Statistical analysis of the PAR-CLIP data

Protein	T-to-C mismatched miR-21 reads	Exact matches to miR-21	Probability
IGF2BP1	2,774	197,965	
IGF2BP2	1,690	96,821	
IGF2BP3	892	53,518	
IGF2BP total	5,356	348,304	Background probability = 0.0151
PAPD5 replicate 1	13	264	$P = 3.7 \times 10^{-4}$
PAPD5 replicate 2	7	139	$P = 7.2 \times 10^{-3}$

Table S3. siRNAs used to knock down nucleotidyl transferase and nuclease genes

Gene	Replicate	siRNA sequences used
		Nucleotidyl transferases
PAPD4	Both	AUUACAUGGAGCUUGAUGUACAAGG, CCUUGUACAUCAAGCUCCAUGUAAU
PAPD5	Both	AACAACGUAGGCAUAAUCAAAGGCC, GGCCUUUGAUUAUGCCUACGUUGUU
		Nucleases
CNOT6L	First	ACACACUUGACAGCGCUGCACCUAA, UUAGGUGCAGCGCUGUCAAGUGUGU
	Second	CCUUAGUCGCAUUCCACCUGAUAUU, AAUAUCAGGUGGAAUGCGACUAAGG
ERI1	Both	GCCUCACUGUGGUCUUGAUGACUCU, AGAGUCAUCAAGACCACAGUGAGGC
EXOSC4	Both	CCCACGCUCCCAGAUUGAUAUCUAU, AUAGAUAUCAAUCUGGGAGCGUGGG
EXOSC10	Both	CAACCAGUGGAUCUGUGCCAGUUCA, UGAACUGGCACAGAUCCACUGGUUG
PAN3	Both	CCUACAGAGGUUGACAGCUACCAUA, UAUGGUAGCUGUCAACCUCUGUAGG
PARN	First	GGUGCAGGAGCUCUGUCCUAUGUAU, AUACAUAGGACAGAGCUCCUGCACC
	Second	CAGAAUCACUAUUACCGCAACAAUA, UAUUGUUGCGGUAAUAGUGAUUCUG
RNASEL	Both	GGACGCUUCGGAAUGUGGGAAAUGA, UCAUUUCCCACAUUCCGAAGCGUCC