RESOURCE

nature structural & molecular biology

TALEN-based knockout library for human microRNAs

Young-Kook Kim[1,2,6], Gabbine Wee[3,4,6], Joha Park[1,2], Jongkyu Kim[1,2], Daehyun Baek[1,2,5], Jin-Soo Kim[3,4] & V Narry Kim[1,2]

Various technical tools have been developed to probe the functions of microRNAs (miRNAs), yet their application has been limited by low efficacy and specificity. To overcome the limitations, we used transcription activator–like effector nucleases (TALENs) to knock out human miRNA genes. We designed and produced a library of 540 pairs of TALENs for 274 miRNA loci, focusing on potentially important miRNAs. The knockout procedure takes only 2–4 weeks and can be applied to any cell type. As a case study, we generated knockout cells for two related miRNAs, miR-141 and miR-200c, which belong to the highly conserved miR-200 family. Interestingly, miR-141 and miR-200c, despite their overall similarity, suppress largely nonoverlapping groups of targets, thus suggesting that functional miRNA-target interaction requires strict seed-pairing. Our study illustrates the potency of TALEN technology and provides useful resources for miRNA research.

miRNAs are phylogenetically conserved post-transcriptional regulators that induce translational repression and mRNA destabilization[1]. Canonical miRNA genes are transcribed by RNA polymerase II, yielding long primary transcripts (pri-miRNAs) that contain local hairpin structures[2]. The hairpin is cleaved initially by the nuclear RNase III protein Drosha, and this releases a pre-miRNA. After nuclear export, the pre-miRNA is processed by the cytoplasmic RNase III protein Dicer. The resulting small-RNA duplex is then loaded onto the Argonaute (also known as Ago) protein. One strand of the duplex remains stably associated with Ago while the other strand is discarded. Although miRNAs are generally produced from both the 5′ and 3′ strands of a miRNA hairpin (referred to as 5p and 3p miRNAs, respectively), one strand is usually more abundant than the other. The miRNA–Ago complex, known as a core component of RNA-induced silencing complex, acts as an effector complex in gene silencing. miRNAs generally recognize their target mRNAs through specific base-pairing at the 3′ untranslated region (UTR). The sequences 2–8 nt from the 5′ end of miRNA are critical for the interaction with the targets and hence are referred to as 'seed' sequences.

The latest release of the miRBase (release 19) has cataloged 223 miRNA loci in *Caenorhabditis elegans*, 238 loci in *Drosophila melanogaster* and 1,600 loci in *Homo sapiens*[3]. During evolution, animal miRNA genes have undergone gene duplications resulting in the formation of multiple paralogs (or 'sisters'). They are often clustered in the genome and form a single transcription unit. The miRNA sisters typically have identical or highly similar sequences in the 5′ seed region, whereas the middle and the 3′ part are more variable. The extent to which the varying sequences among miRNA sisters contribute to the targeting specificity of miRNAs remains largely unexplored.

To aid in understanding the biological and pathological significance of miRNAs, various technical tools have been introduced[4]. Chemically synthesized oligonucleotides are widely used to mimic miRNA activity in gain-of-function experiments. However, the results obtained from synthetic mimics at high doses should be interpreted with caution because the mimics may perturb the endogenous miRNA pathway[5]. Synthetic oligonucleotides complementary to miRNAs have been used to inhibit miRNAs, but they can tolerate some mismatches, thus resulting in unintended effects, especially when there are multiple related miRNAs[6]. miRNA sponges (transcripts that contain complementary sites to 'soak' miRNAs) also suffer from the specificity issue[7]. Therefore, gene knockout (KO) would be an ideal choice for deletion and study of an individual miRNA locus. Although homologous recombination–based methodology has been conventionally used in animal models[8,9], it is not widely accessible, owing to the complicated procedure and limitations in choice of organism.

Engineered nucleases cleave chromosomal DNA in a site-specific manner, thus producing a DNA double-strand break (DSB), which in turn triggers an error-prone repair process known as nonhomologous end joining; this allows targeted mutagenesis in virtually any cell type or organism[10–13]. We have previously used zinc-finger nucleases (ZFNs), transcription activator–like effector nucleases (TALENs) and CRISPR–Cas9–derived RNA-guided endonucleases (RGENs) to disrupt protein-coding genes in cell lines and animals[13–17]. Among the three types of programmable nucleases, TALENs are suitable for small loci with narrow targetable regions because TALENs are applicable to almost any DNA sequence, whereas ZFNs and RGENs are limited by the availability of zinc-finger modules and the requirement of GG dinucleotides in the target sequence, known as the protospacer-adjacent motif[18], respectively.

In this study, we used a TALEN-mediated KO technique to delete human miRNA genes with high efficiency and specificity. We generated a library of 540 TALEN pairs for 274 loci, targeting authentic

[1]Center for RNA Research, Institute for Basic Science, Seoul, Korea. [2]School of Biological Sciences, Seoul National University, Seoul, Korea. [3]National Creative Research Initiatives Center for Genome Engineering, Seoul National University, Seoul, Korea. [4]Department of Chemistry, Seoul National University, Seoul, Korea. [5]Bioinformatics Institute, Seoul National University, Seoul, Korea. [6]These authors contributed equally to this work. Correspondence should be addressed to V.N.K. (narrykim@snu.ac.kr) or J.-S.K. (jskim01@snu.ac.kr).

Received 12 August; accepted 30 September; published online 10 November 2013; doi:10.1038/nsmb.2701

miRNAs. As a case study, we generated single- and double-KO cell lines for two related miRNAs, miR-141 and miR-200c, to investigate their functional differences. The miRNA TALENs will be distributed upon request through our website (http://www.talenlibrary.net/h_miRNA/).

RESULTS

Selection of miRNA loci for TALEN library construction

One of the challenges in developing a TALEN library for miRNAs is that the current miRNA database (miRBase release 19) inevitably contains a substantial number of faulty annotations. The majority of registries, particularly the recent ones, have been annotated on the basis of high-throughput sequencing, which is sensitive enough to detect RNAs of low abundance and decay intermediates. Close examination of the data reveals that a sizable fraction of miRNA entries are supported by only small numbers of sequencing reads and that their 5′ ends are highly heterogeneous. The 5′ end of an authentic miRNA is defined by RNase III and protected by Ago with few exceptions of mirtrons[19]. miRNA genes are under strong selective pressure to preserve the 5′ end of mature miRNA because the position of the seed sequence is important for target recognition. A systematic effort to experimentally validate miRNAs in miRBase version 14 revealed that nearly one-third of the tested loci (173 of 564) lack convincing evidence that they produce authentic miRNAs[20]. Although some of the misannotated entries have been removed from the database, many more have been added without validation, thus suggesting that current estimates of the number of existing miRNAs might be inflated.

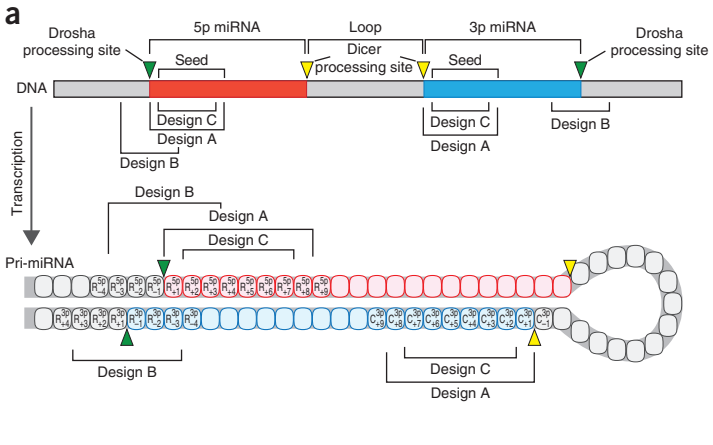
To generate a useful TALEN library, we decided to focus on miRNAs that are likely to be biologically important. To this end, we collected and examined 71 data sets from small-RNA sequencing libraries from various human tissues and cell types (**Supplementary Table 1**). We selected miRNAs whose expression ranks belonged to the top 10% (higher than 90th-percentile rank) in at least one tissue type. We also filtered out miRNAs that are heterogeneous at the 5′ termini (5′ homogeneity lower than 0.9). We defined the 5′ homogeneity as the number of reads sharing the 5′ end of the most frequent sequence, divided by total read numbers from the given strand. Because our criteria are stringent, and thus some of the tissue-specific miRNAs and alternatively processed miRNAs failed to pass the filter, we curated additional miRNAs on the basis of the literature and manual sequence examination. We eventually selected 274 miRNAs for KO study (**Supplementary Table 2**). This set includes 159 out of 187 (85%) miRNAs that are broadly conserved in vertebrates[21]. It also contains those with narrower phylogenetic distribution such as miR-498, miR-605, miR-629 and miR-1323, which are potentially primate specific.

Design strategies for TALEN constructs

A pair of TALEN enzymes form a dimer and generate a DSB between the two enzyme-binding sites. We here refer to the region between the enzyme-binding sites as the 'TALEN target region'. Error-prone repair of the DSB gives rise to small insertions or deletions (indels). Because of these small changes, the TALEN target region in the miRNA loci must be selected carefully. If the indel were to occur in a location that is not critical for miRNA biogenesis and function, the miRNA activity would remain intact even when the genomic mutagenesis itself was successful.

In the initial design (design A, **Fig. 1a**,**b**), we used the region corresponding to 1–8 nt of mature miRNA as the TALEN target region. This is because even a small indel in this region is expected to disrupt miRNA activity. In addition, because the seed region is adjacent to the processing sites, indels in this region would distort RNA structure and affect processing. Although, in principle, TALEN constructs could be designed for any genomic DNA, the efficiency of TALEN targeting can be variable depending on the target sequence and TALEN architecture. In a previous study, we reported an improved TALEN architecture with high efficiency, success rate and specificity[16]. We designed TALEN constructs, on the basis of this prior knowledge, that recognize half-site DNA sequences of 18–20 bp in length, including a conserved T base at the 5′ end, with the center of the miRNA seed region being the most preferred TALEN cleavage site (design A, **Fig. 1a**,**b**).

Although we could successfully design highly efficient TALENs for most miRNA loci in accordance with the criteria above, in some cases the seed region was not suitable for design of efficient TALENs. For those miRNA loci, we instead targeted Drosha processing sites (design B, **Fig. 1a**,**b**). Indels near Drosha processing sites would distort the



b

Design	Target region	TALE size	Predicted TALEN cleavage site		Numbers of designed TALENs
			5p miRNA	3p miRNA	
A	Seed (1–8)	18–20	$R^{5p}_{-1}\|R^{5p}_{+1} - R^{5p}_{+8}\|R^{5p}_{+9}$	$C^{3p}_{-1}\|C^{3p}_{+1} - C^{3p}_{+8}\|C^{3p}_{+9}$	476
B	Drosha processing site	18–20	$R^{5p}_{-4}\|R^{5p}_{-3} - R^{5p}_{+3}\|R^{5p}_{+4}$	$R^{3p}_{-4}\|R^{3p}_{-3} - R^{3p}_{+3}\|R^{3p}_{+4}$	37
C	Seed (2–7)	16–20	$R^{5p}_{+1}\|R^{5p}_{+2} - R^{5p}_{+7}\|R^{5p}_{+8}$	$C^{3p}_{+1}\|C^{3p}_{+2} - C^{3p}_{+7}\|C^{3p}_{+8}$	27

c

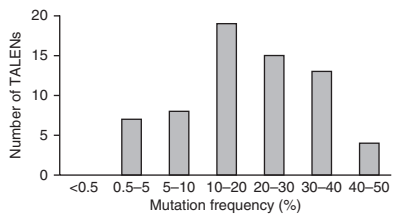


Figure 1 TALEN-mediated KO of miRNAs. (**a**) Design of TALEN constructs. Schematic structures of miRNA locus (top) and pri-miRNA (bottom) are shown. Three different design rules were applied to generate effective TALEN constructs. The TALEN target regions are indicated with brackets (design A through C). Each nucleotide position is designated with the letter 'R' or 'C' to indicate the processing site by Drosha or Dicer, respectively. The nucleotides upstream or downstream of the given processing site are indicated with negative or positive numbers, respectively, on the bottom right. On the top right, the strand of the hairpin is indicated as either 5p or 3p. (**b**) Summary of the strategies for TALEN design. (**c**) T7E1 assay of KO efficacy of the TALEN library. 66 TALEN pairs against 33 miRNA loci were tested. The number of TALENs was counted on the basis of the mutation frequency measured by the T7E1 assay (**Supplementary Fig. 1**).

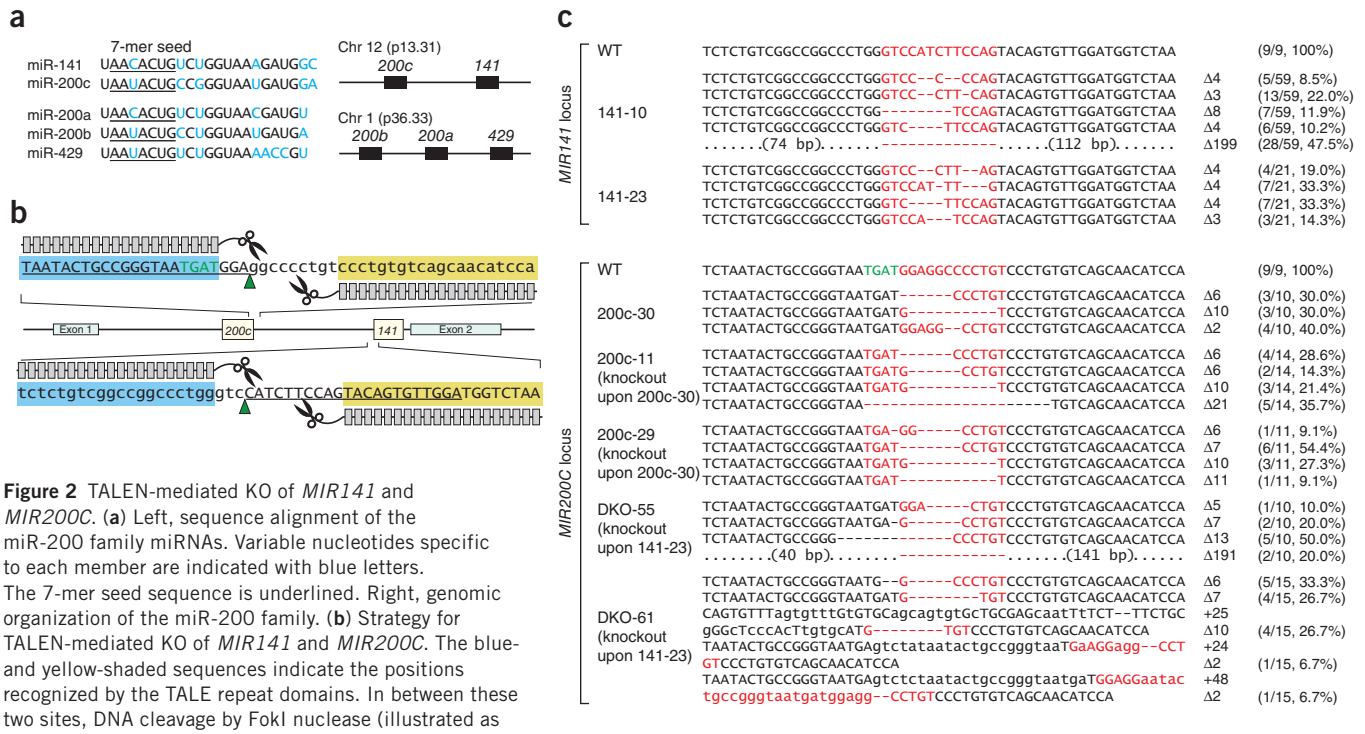


Figure 2 TALEN-mediated KO of *MIR141* and *MIR200C*. (**a**) Left, sequence alignment of the miR-200 family miRNAs. Variable nucleotides specific to each member are indicated with blue letters. The 7-mer seed sequence is underlined. Right, genomic organization of the miR-200 family. (**b**) Strategy for TALEN-mediated KO of *MIR141* and *MIR200C*. The blue- and yellow-shaded sequences indicate the positions recognized by the TALE repeat domains. In between these two sites, DNA cleavage by FokI nuclease (illustrated as scissors) occurs. The sequences corresponding to miR-200c-3p and miR-141-5p are underlined. The green nucleotides are the DNA sequences recognized only by the first *MIR200C* TALEN (L4R4) (sequence in **Supplementary Table 4**). (**c**) Genomic DNA sequences near *MIR141* and *MIR200C* loci in the indicated KO cells compared to those of wild type (WT) cells. The number of mutated nucleotides and the sequencing frequency of each allele are indicated at right. Red letters indicate the region targeted by the nuclease domain. Small letters indicate the DNA insertion.

secondary structure of a miRNA hairpin critical for processing[22]. We analyzed the combined small-RNA sequencing data to infer Drosha processing sites, assuming that the most frequent 5′ end of 5p miRNA corresponds to the Drosha processing site[22]. We extended the TALEN target region from the processing site by three nucleotides in both 5′ and 3′ directions (design B, **Fig. 1a**,**b**). When we could not effectively apply the above strategies, we allowed TALENs of shorter size, which may potentially reduce the efficiency of TALEN targeting to some extent (design C, **Fig. 1a**,**b**). We instead used a narrower region for targeting, corresponding to 2–7 nt of mature miRNA.

We designed one TALEN pair per arm, and this resulted in two pairs of TALEN constructs for each miRNA locus. We designed and constructed 540 pairs of TALEN plasmids collectively, following one of the above rules. Details of the TALEN design for each miRNA are in **Supplementary Table 3**.

To test the genome-editing activities of the TALEN library, we selected 66 TALEN pairs against 33 miRNA loci. After transfecting them into HEK293T and K562 cells, we extracted total genomic DNA to detect DNA mutation by a mismatch-sensitive T7 endonuclease I (T7E1) assay. All TALEN pairs that we tested induced mutations with a frequency above 0.5% (**Fig. 1c** and **Supplementary Fig. 1**). These results indicate that most, if not all, of our TALEN constructs can be used for miRNA KO experiments.

Knockout of the miR-200 family

MIR141 and *MIR200C* belong to the highly conserved miR-200 family (**Fig. 2a**), but interestingly, they differ in the seed region by one nucleotide. Some target-prediction algorithms that do not allow wobble pairing in the seed region classify them into two distinct groups[23,24]. However, previous functional studies suggested that miR-141 and miR-200c may have largely indistinguishable activities[25]. For instance, overexpression of miR-200a (which has the same seed motif as miR-141) and miR-200c downregulated a nearly identical set of genes[26]. Similarly, when overexpressed, miR-141 repressed *FOG2* (official symbol *ZFPM2*) mRNA, which has the miR-200c seed sites[27]. The reverse was also true: miR-200c suppressed the targets of miR-141 when overexpressed[28]. However, neither overexpression nor complementary inhibitor–based knockdown is ideal to investigate the differences between related miRNAs because the possibilities of off-targeting and cross-reactivity cannot be excluded. To address the issue of functional diversification among family members, we decided to generate single- and double-KO cell lines for the two miRNAs. To avoid complication due to other related miRNAs from the *MIR200B-MIR200A-MIR429* cluster in chromosome 1, we used a breast cancer cell line, SK-BR-3, that expresses only the *MIR200C-MIR141* cluster but not the other cluster (**Supplementary Fig. 2a**).

For the *MIR141* KO, we used TALENs that target the seed region of the 5p strand (**Fig. 2b** and **Supplementary Table 4**, TALEN L1R1). As for *MIR200C*, we chose TALENs targeting the Drosha processing site in the 3p strand (**Fig. 2b** and **Supplementary Table 4**, TALEN L4R4). To test the efficacy of the TALENs, we cotransfected the constructs with a reporter plasmid that contains the TALEN recognition sequence fused to a gene encoding a surface antigen for enrichment of the mutated cells[29] (**Supplementary Fig. 2b** and Online Methods). The result from the T7E1 assay showed that magnetic separation enriched mutated cells by three- to seven-fold (**Supplementary Fig. 2b**). We also identified potential off-target sites in the human genome that are highly homologous to on-target sites of TALENs (**Supplementary Tables 5–7**) and tested whether the TALENs induced off-target mutations in the enriched cells by T7E1 assay (**Supplementary Fig. 2c**). None of the sites showed any DNA mutation, thus indicating the high specificity of the method.

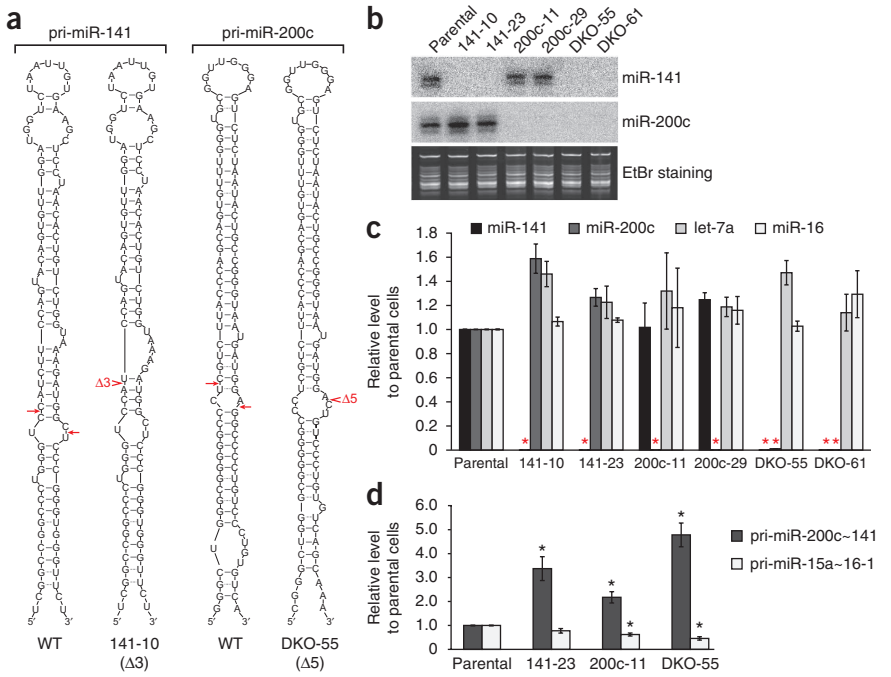


Figure 3 Confirmation of miRNA KO by TALEN. (**a**) Representative hairpin structure of miRNA precursors. Among the mutated alleles, the allele carrying the smallest deletion is shown along with the wild type. Red arrows indicate the Drosha processing sites in the wild-type hairpin. The position and number of deleted nucleotides are shown in red letters on the mutant structure. (**b**) Northern blot analysis validating miRNA KOs in the indicated cells. Ethidium bromide (EtBr) staining of the gel is shown at bottom to indicate the quantity and quality of the loaded RNA (uncropped image in **Supplementary Fig. 5**). (**c**) qRT-PCR validating the miRNA KOs. The relative level of each miRNA is normalized to that of U6 small nuclear RNA and normalized again to the miRNA level from parental cells. Data are presented as mean ± s.e. of biological replicates ($n = 4$) for miR-141 and miR-200c or as mean ± range for the other miRNAs ($n = 2$). Asterisks indicate the samples with miRNA expression level below the detection limit. (**d**) qRT-PCR to measure the relative level of pri-miR-200c~141 in the KO cells. The level of pri-miRNA was normalized against that of *GAPDH* mRNA. Data are presented as mean ± s.e. of biological replicates ($n = 3$). $^*P < 0.05$ by two-tailed t test.

After the enrichment of KO cells by magnetic separation, we separated single cells, grew them into colonies and extracted genomic DNA from individual clones for the T7E1 assay and performed fluorescent PCR analysis to identify the clones with mutated genomes (**Supplementary Fig. 2d**,**e**). We further examined selected clones by Sanger DNA sequencing to confirm the sequence alterations (**Fig. 2c**). Because the sequencing data revealed a maximum of five different kinds of indels per clone, SK-BR-3 appears to be aneuploid (which is not unusual for cancer cells) with at least five different alleles of the *MIR200C-MIR141* cluster (**Fig. 2c**). Remarkably, a single transfection cycle was sufficient to yield many complete-KO clones, thus highlighting the efficiency of the method. We also tested off-target mutations in each clone, and none of the sites showed any mutation (**Supplementary Fig. 2c**). Finally, we used two independent KO clones per miRNA locus for the following analyses. The scheme of KO procedure and the mutation efficiency at each step are summarized in **Supplementary Figure 2f**.

The secondary structure of pri-miRNA is a useful predictor of processing efficiency[22,30]. **Figure 3a** shows examples of mutant alleles possessing small deletions that critically disrupt the structure of pri-miRNA. To test whether the mutations indeed blocked processing, we carried out northern blotting and quantitative reverse-transcription PCR (qRT-PCR) with total RNA from the KO candidate cells (**Fig. 3b**,**c**). As expected, miR-141 and miR-200c were undetectable in both assays. We note that for the *MIR200C* KO we targeted the locus in two steps because a small amount of miR-200c was expressed in the initial KO (clone 200c-30) (**Supplementary Fig. 3a**). This was because the 2-nt deletion in one of the alleles failed to block pri-miRNA processing, owing to a relatively small change in the RNA structure. We therefore transfected the second TALEN pair into clone 200c-30 to induce additional indels (Online Methods), and this resulted in two null clones, 200c-11 and 200c-29. We learned from this experience that it is desirable to select the clones with the largest-sized available deletions to ensure complete miRNA depletion and that secondary structure of pri-miRNA is indeed useful to predict the outcome of miRNA biogenesis.

We also generated double-KO (DKO) clones by transfecting clone 141-23 with the first set of *MIR200C* TALENs (TALEN L4R4, **Supplementary Table 4**). In this experiment, we increased the transfection efficiency by electroporation and selected the clones with genomic changes that alter RNA structure markedly. The *MIR200C* locus was efficiently deleted (**Fig. 2c** and **Supplementary Fig. 2d**,**e**), and miRNA expression was successfully abrogated in the selected clones (**Fig. 3b**,**c**, clones DKO-55 and DKO-61).

In the KO cells, the second cluster (*MIR200B-MIR200A-MIR429*) remained silent, thus suggesting that there is no compensatory induction from the second cluster (**Supplementary Fig. 3b**). Moreover, the expression of other miRNAs including let-7a and miR-16 did not change appreciably during the KO process (**Fig. 3c** and **Supplementary Fig. 3b**). These results confirm that the KO was successful and specific. The level of pri-miR-200c~141 (tilde indicates transcript spanning the indicated gene cluster) increased in the KO cells, a result indicating that the genomic editing inhibited Drosha processing of pri-miR-200c~141 (**Fig. 3d**). Interestingly, the level of pri-miR-15a~16-1 decreased slightly in the KO cells, thus implicating a competition between pri-miRNAs for Drosha (**Fig. 3d**).

Impacts of the miR-200 family knockout

To examine the effects of miRNA depletion, we first monitored cell proliferation rates by cell counting (**Fig. 4a**). Cell numbers increased more slowly in single- and double-KO cells compared to the parental cells. To understand the mechanism behind the proliferation defect, we analyzed cell-cycle profiles by flow cytometry. In KO cells, the proportion of G1-phase cells increased modestly but reproducibly, whereas those in S and G2-M phases decreased (**Supplementary Fig. 4a**). Apoptotic cell population, which is represented by the sub-G1 phase, increased slightly in KO cells. Thus, cell-cycle retardation and apoptosis may partly explain the decrease in proliferation of the KO cell lines. None of the KO clones generated in our previous studies showed any noticeable defect in growth rate, thus ruling out the possibility that the growth defect of miRNA-KO cells might be a secondary response to TALEN transfection[16]. Taken together, the miR-200 family of miRNAs

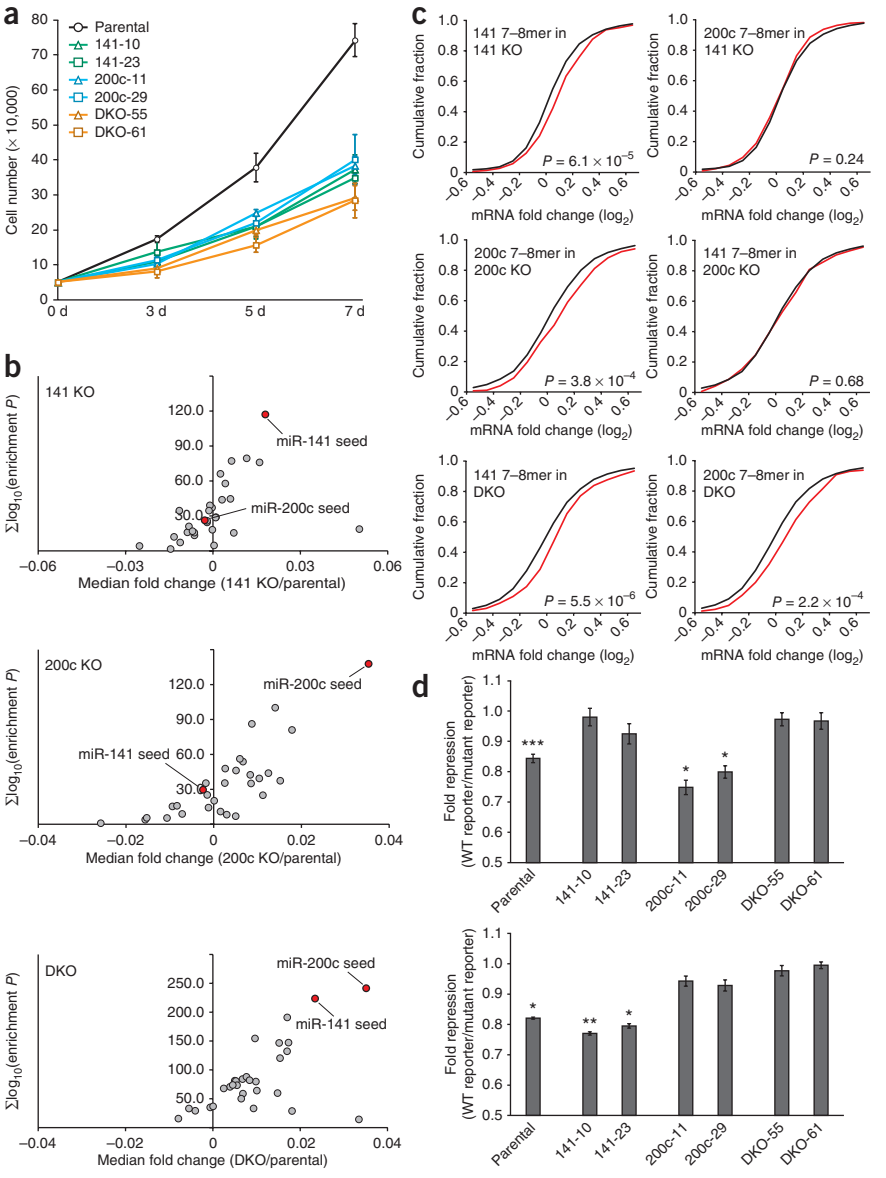


Figure 4 Cellular and molecular impact of miRNA KO. (**a**) Cell proliferation rates by cell counting. Data are presented as mean ± s.e. of biological replicates ($n = 3$). ($P = 0.0063$ for 141-10; 0.0081 for 141-23; 0.0044 for 200c-11; 0.0063 for 200c-29; 0.0031 for DKO-55; and 0.0031 for DKO-61 by two-tailed t test). (**b**) Motif enrichment analysis for miRNA seed sequences in the 3′ UTRs of the genes upregulated in response to miRNA depletion compared to the rest of the genes. Seed sequences of the 30 most highly expressed miRNAs in the SK-BR-3 cell line are shown. The y axis shows the sum of $\log_{10}(P$ value) (two-sided Fisher's exact test; Online Methods). The x axis represents the median fold change of the genes with the corresponding seed motif in the given KO cells compared to the parental cells. (**c**) Global mRNA-level change induced by KO. The fold change of mRNA levels induced by the KO is compared between the transcripts, which include miRNA 7–8mers (7-mer-m8, 7-mer-A1 or 8-mer) in their 3′ UTR (red) and the transcripts lacking a 3′ UTR seed (black). P value was calculated by two-sided Kolmogorov–Smirnov test. (**d**) Luciferase reporter assay with the *PPT2* (top) and *FOG2* (bottom) reporters to measure the effect of miR-141 and miR-200c on target repression. Data are presented as mean ± s.e. of biological replicates ($n = 4$ for *PPT2*; $n = 3$ for *FOG2*). ***$P < 0.0001$; **$P < 0.001$; *$P < 0.01$ by one-tailed t test.

may be required for G1-S transition and cell survival, and this is consistent with our previous study[27].

Next, we analyzed the molecular impact of miRNA depletion by mRNA sequencing. To determine whether the miRNA targets are selectively derepressed in KO cells, we performed motif-enrichment analyses, comparing the 3′-UTR sequences of the upregulated genes with those of the remaining genes (**Fig. 4b** and **Supplementary Fig. 4b**). In *MIR141* KO cells, the miR-141 motif (but not the miR-200c motif) was enriched in the upregulated mRNAs, thus indicating that the miR-141 targets were indeed derepressed. Similarly, in the *MIR200C* KO clones, the miR-200c motif was enriched more than were the seed motifs of other miRNAs (including miR-141). In the DKO, both motifs were enriched in the upregulated transcripts.

We also determined the fold change of the transcripts containing the seed motif over those without the motif (**Fig. 4c**). In *MIR141* KO cells, mRNAs with the miR-141 motif, compared to those without the motif, were globally upregulated. Likewise, miR-200c depletion resulted in derepression of mRNAs with the miR-200c motif. However, mRNAs with the miR-200c motif were not upregulated in the *MIR141* KO. Similarly, transcripts with the miR-141 motif were not significantly influenced by *MIR200C* KO. In DKO cells, both seed groups were globally upregulated. These data indicate that the two closely related miRNAs do not cross-react notably and may control largely nonoverlapping groups of genes.

To confirm this finding, we used luciferase reporter constructs that contained the 3′ UTRs of miRNA target genes. We noticed that the *PPT2* gene has two complementary sites for the miR-141 seed but none for the miR-200c seed (**Supplementary Fig. 4c**). In contrast, the 3′ UTR of *FOG2* contains three seed sites for miR-200c but none for miR-141 (ref. 27). We also generated, for controls, the UTR reporters with point-mutated seed sites (**Supplementary Fig. 4c**). The *PPT2* wild-type reporter was repressed specifically in cells in which miR-141 was intact (parental cells and *MIR200C* KO cells) (**Fig. 4d**). The reverse was true for the *FOG2* reporter; the reporter was suppressed only in cells in which miR-200c was expressed (parental cells and *MIR141* KO cells) (**Fig. 4d**). Thus, miR-200c cannot compensate for the absence of miR-141 and vice versa. The reporters responded to specific miRNAs that match the seed sequences, without a noticeable crossover between the sister miRNAs.

DISCUSSION

We here introduce TALENs as a general tool for the study of miRNA function. Our TALEN-based method is efficient enough to mutate multiple alleles simultaneously in a single transfection without the need for extensive screening. Owing to the high efficiency, it takes only 2–4 weeks (depending on the proliferation rate of the cells used) to obtain null clones with all alleles mutated. Therefore it is now feasible

to generate double- or even multiple-KO cells within a few months. Multiple KOs will be useful not only to examine miRNA families with paralogs but also to investigate regulatory relationships between miRNAs. Because TALENs can be modified easily for other species, multiple KOs in mouse embryonic stem cells would be particularly useful because they could be immediately used to produce mice with multiple deletions. In addition, TALEN-based KOs will be valuable for the study of narrowly conserved miRNAs (such as primate-specific miRNAs), which cannot be studied with rodent animal models. In principle, the method could be applied to any cell type of interest, including primary cells. Because the current TALENs are highly effective, transfection efficiency becomes the major determinant for successful creation of KOs. Thus, optimization of transfection in the given cell type is advisable before KO experiments. The TALENs are available upon request through our website (http://www.talenlibrary.net/h_miRNA/).

In this work, we provide general design strategies for miRNA TALENs, which can be applied to other miRNAs in humans and other species. All of the three strategies (designs A through C) used turned out to be valid (**Supplementary Fig. 1** and **Supplementary Table 3**). However, we learned from the *MIR200C* KO experiment that the target site needs to be carefully chosen in the design step because small indels in certain sites may not sufficiently disrupt pri-miRNA structure. Thus, we reviewed our initial design of the library and manually inspected the predicted secondary structure of the pri-miRNAs. Consequently, the constructs in the final library are expected to cause mutations that severely block miRNA processing even with small deletions. A recent report also showed that TALENs are a useful tool to mutate miRNA loci (*MIR155*, *MIR146A* and *MIR125B1*), although the study did not test whether the genomic deletions indeed abolished miRNA biogenesis[31]. Because a small deletion in the terminal loop may not effectively block miRNA maturation, it is desirable to target the seed region (or alternatively the processing sites) and to select clones with the largest available indels so as to ensure complete functional KO.

MIR141 and *MIR200C* belong to the miR-200 family, which is deeply conserved in bilaterian animals (**Supplementary Fig. 4d**,**e**). The family expanded in chordates, thus resulting in two subfamilies that differ by one nucleotide in the seed: (i) miR-200c, miR-200b and miR-429 (AAUACUG) and (ii) miR-141 and miR-200a (AACACUG) (**Supplementary Fig. 4d**,**e**). Earlier data suggested that the two subgroups may cross-react and that the targets may overlap[25,26]. Several groups recognized the problem and found that two subfamilies have different functions[32,33]. However, it was not possible to definitively resolve the issue, owing to the limitation of oligonucleotide-based transfection experiments that are potentially vulnerable to off-target and nonspecific effects. The miR-200 family is particularly interesting because the U base (in miR-200c) can form a wobble pair with G (in miR-141 targets). If wobble pairing is permitted, miR-200c is expected to suppress miR-141 targets. Some reports indicated that the wobble pairs in the seed region are detrimental to miRNA activity[34–36], whereas others presented contradictory results[37]. Our data are consistent with the model that functional interaction requires strict Watson–Crick–type pairing and does not allow wobble pairing in the seed, although we cannot formally exclude the possibility that wobble pairing is tolerated in rare targets with extensive complementarity to the 3′ part of miRNA. Our study implies that alteration of a single nucleotide in the seed may switch the target population and profoundly diversify the subfamily members during evolution. It is intriguing that, despite the distinct molecular effects, the gross cellular impact of miRNA KO was similar between the subfamilies (**Fig. 4a** and **Supplementary Fig. 4a**). It is plausible that the two subfamilies may still be related in their biological functions and may act on genes in the same or related pathways, even though their direct molecular targets are distinct. Further studies with the KO cells will help us dissect the functions of miRNAs in the context of a complex gene network.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Data have been deposited in the Gene Expression Omnibus database under accession code GSE51217.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank S. Kim (ToolGen) for the help with TALEN design and construction. We gratefully acknowledge S. Ryu and S. Bae for their technical help. We are grateful to members of our laboratories, particularly Y. Kim, for discussion and reading the manuscript. This work was supported by the Research Center Program of the Institute for Basic Science (EM1202) (Y.-K.K., J.P., D.B. and V.N.K.) and Basic Science Research Program of the National Research Foundation funded by the Ministry of Education, Science and Technology of Korea (20110014523) (J.K. and D.B.). J.-S.K. was supported by the National Research Foundation of Korea (2013000718).

AUTHOR CONTRIBUTIONS

Y.-K.K., G.W., J.-S.K. and V.N.K. designed the project; Y.-K.K. and G.W. performed the experiments and analyzed the data; J.P., J.K. and D.B. performed the bioinformatics analysis; Y.-K.K. and V.N.K. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Bartel, D.P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
2. Kim, V.N., Han, J. & Siomi, M.C. Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **10**, 126–139 (2009).
3. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**, D152–D157 (2011).
4. Thomas, M., Lieberman, J. & Lal, A. Desperately seeking microRNA targets. *Nat. Struct. Mol. Biol.* **17**, 1169–1174 (2010).
5. Khan, A.A. *et al.* Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat. Biotechnol.* **27**, 549–555 (2009).
6. Stenvang, J., Petri, A., Lindow, M., Obad, S. & Kauppinen, S. Inhibition of microRNA function by antimiR oligonucleotides. *Silence* **3**, 1 (2012).
7. Ebert, M.S. & Sharp, P.A. MicroRNA sponges: progress and possibilities. *RNA* **16**, 2043–2050 (2010).
8. Prosser, H.M., Koike-Yusa, H., Cooper, J.D., Law, F.C. & Bradley, A. A resource of vectors and ES cells for targeted deletion of microRNAs in mice. *Nat. Biotechnol.* **29**, 840–845 (2011).
9. Park, C.Y. *et al.* A resource for the conditional ablation of microRNAs in the mouse. *Cell Rep.* **1**, 385–391 (2012).
10. Bibikova, M., Beumer, K., Trautman, J.K. & Carroll, D. Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**, 764 (2003).
11. Urnov, F.D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646–651 (2005).
12. Miller, J.C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**, 143–148 (2011).
13. Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 230–232 (2013).
14. Kim, H.J., Lee, H.J., Kim, H., Cho, S.W. & Kim, J.S. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res.* **19**, 1279–1288 (2009).
15. Kim, S., Lee, M.J., Kim, H., Kang, M. & Kim, J.S. Preassembled zinc-finger arrays for rapid construction of ZFNs. *Nat. Methods* **8**, 7 (2011).
16. Kim, Y. *et al.* A library of TAL effector nucleases spanning the human genome. *Nat. Biotechnol.* **31**, 251–258 (2013).
17. Sung, Y.H. *et al.* Knockout mice created by TALEN-mediated gene targeting. *Nat. Biotechnol.* **31**, 23–24 (2013).

18. Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
19. Miyoshi, K., Miyoshi, T. & Siomi, H. Many ways to generate microRNA-like small RNAs: non-canonical pathways for microRNA production. *Mol. Genet. Genomics* **284**, 95–103 (2010).
20. Chiang, H.R. *et al.* Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev.* **24**, 992–1009 (2010).
21. Friedman, R.C., Farh, K.K.H., Burge, C.B. & Bartel, D.P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105 (2009).
22. Han, J. *et al.* Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**, 887–901 (2006).
23. Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
24. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nat. Genet.* **37**, 495–500 (2005).
25. Gregory, P.A., Bracken, C.P., Bert, A.G. & Goodall, G.J. MicroRNAs as regulators of epithelial-mesenchymal transition. *Cell Cycle* **7**, 3112–3118 (2008).
26. Park, S.M., Gaur, A.B., Lengyel, E. & Peter, M.E. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev.* **22**, 894–907 (2008).
27. Hyun, S. *et al.* Conserved microRNA miR-8/miR-200 and its target USH/FOG2 control growth by regulating PI3K. *Cell* **139**, 1096–1108 (2009).
28. Burk, U. *et al.* A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Rep.* **9**, 582–589 (2008).
29. Kim, H. *et al.* Magnetic separation and antibiotics selection enable enrichment of cells with ZFN/TALEN-induced mutations. *PLoS ONE* **8**, e56476 (2013).
30. Zeng, Y., Yi, R. & Cullen, B.R. Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J.* **24**, 138–148 (2005).
31. Hu, R., Wallace, J., Dahlem, T.J., Grunwald, D.J. & O'Connell, R.M. Targeting human microRNA genes using engineered Tal-effector nucleases (TALENs). *PLoS ONE* **8**, e63074 (2013).
32. Elson-Schwab, I., Lorentzen, A. & Marshall, C.J. MicroRNA-200 family members differentially regulate morphological plasticity and mode of melanoma cell invasion. *PLoS ONE* **5**, e13176 (2010).
33. Uhlmann, S. *et al.* miR-200bc/429 cluster targets PLCgγ1 and differentially regulates proliferation and EGF-driven invasion than miR-200a/141 in breast cancer. *Oncogene* **29**, 4297–4306 (2010).
34. Doench, J.G. & Sharp, P.A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
35. Brennecke, J., Stark, A., Russell, R.B. & Cohen, S.M. Principles of microRNA-target recognition. *PLoS Biol.* **3**, e85 (2005).
36. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455**, 64–71 (2008).
37. Didiano, D. & Hobert, O. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat. Struct. Mol. Biol.* **13**, 849–851 (2006).

ONLINE METHODS

Processing of small-RNA sequencing data. We collected 71 sequencing data for small RNAs from the NCBI Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/sra/) and sorted them into several groups according to their origin of tissues (**Supplementary Table 1**). The sequencing reads were preprocessed with the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The 3′ adaptor sequences were removed from the reads, and reads shorter than 17 nt in length were discarded. Quality filtering was then performed to exclude low-quality reads containing more than 5% of low-quality bases with quality scores smaller than 20. The reads from the same tissue type were then collapsed and merged. After the preprocessing, the reads were aligned to the human reference genome (hg19) by Burrows–Wheeler Aligner[38], with parameters of *–l* 19 *–k* 0.

Selection of miRNAs for knockout. To choose miRNAs for KO, we used two main criteria: the expression level and the 5′-end homogeneity. We defined the 5′-end homogeneity as the number of reads that have the same 5′ end as in the most frequent read, divided by the total number of reads aligned to the mature miRNA in the given strand. We selected miRNAs with an expression level higher than 90th-percentile rank and 5′ homogeneity more than 0.9 in at least one tissue type (**Supplementary Table 2**).

Generation of TALEN constructs. TALEN constructs are composed of TALE repeat domains and a *Fok*I nuclease domain containing Sharkey RR or Sharkey DAS heterodimer mutation[39] and are designed to target the seed region or the Drosha processing site (**Supplementary Table 3**). TALENs were synthesized by ToolGen as described previously[16].

Generation of *MIR141* and *MIR200C* knockout cell lines. For the enrichment of KO cells, the surrogate reporters were constructed as described previously[29] (**Supplementary Fig. 2b**). In this reporter vector, H-2K^k cell-surface-antigen sequence is fused downstream of the RFP sequence out of frame. However, we inserted specific DNA sequences, between the RFP-coding and H-2K^k antigen–coding sequences, that can be recognized by the cotransfected nuclease. Therefore, if the nuclease vector works properly in the cell, a small indel mutation will be made at this position, thus making the H-2K^k antigen in frame by chance. The magnetic separation system enriches the KO cells expressing this antigen and therefore increases the efficiency of TALEN-mediated KO[29].

The TALEN and reporter plasmids were transfected into SK-BR-3 cells with Lipofectamine 2000 (Life Technologies). To generate DKO cells, *MIR200C* TALEN L4R4 was transfected into 141-23 KO cells (made by the TALEN L1R1 construct) with 4D nucleofector (Amaxa). 48 h after the transfection, H-2K^k–positive cells were separated with a magnetic activated cell-sorting system (MACS, Miltenyi Biotec) according to the manufacturer's instructions. The T7E1 assay was performed to evaluate TALEN activity and enrichment efficiency of mutant cells (**Supplementary Fig. 2b**). To obtain single-cell clones, the separated cells were plated at a density of 1,000 cells per 100-mm culture dish and were incubated until the colonies formed. Individual clones were analyzed by T7E1 assay to select mutant colonies (**Supplementary Fig. 2d**), then genotyped by fluorescent PCR (**Supplementary Fig. 2e**) and DNA sequence analysis (**Fig. 2c**).

The *MIR200C* mutant clone initially generated with TALEN L4R4 (200c-30) expressed a small amount of miR-200c (**Supplementary Fig. 3a**). Therefore, a new TALEN (L3R4) was designed and transfected into the 200c-30 clone (**Supplementary Table 4**). This resulted in two *MIR200C* KO lines (200c-11 and 200c-29), which were confirmed by fluorescent PCR, DNA sequencing and expression analysis (**Figs. 2** and **3** and **Supplementary Fig. 2e**).

T7 endonuclease I assay. T7E1 assays were carried out to detect DNA mutation. After the enrichment of transfected cells and single-cell cloning, genomic DNA was extracted from each colony. The DNA fragment encompassing the targeted region was amplified by PCR (primer sequences in **Supplementary Table 8**), treated with T7E1 and size-separated by gel electrophoresis.

RNA measurement. For northern blot analysis, total RNA was prepared with TRIzol reagent (Life Technologies), separated on 15% denaturing acrylamide gel and then transferred to a Hybond-NX membrane (Amersham). The membrane was chemically cross-linked with 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC)[40] and hybridized with 5′-end-labeled oligonucleotide probe. Original images of gel and blot used in this study can be found in **Supplementary Figure 5**.

For qRT-PCR, total RNA was reverse transcribed and amplified with the miRNA assay kit (Life Technologies), and the level of miRNA was normalized to that of U6 snRNA. For the measurement of primary transcripts, the pri-miRNA analysis kit (Life Technologies) was used, and data were normalized to *GAPDH* mRNA.

Cell proliferation assay. 50,000 SK-BR-3 cells were seeded on a 35-mm dish. After 3, 5 and 7 d, the medium was aspirated to remove floating (dead) cells, and the numbers of cells were counted after trypsinization.

Luciferase assay. The *PPT2* reporter, which contains a Firefly luciferase coding sequence and the 3′-UTR sequence from the *PPT2* gene (**Supplementary Fig. 4c**), was cloned with the EZ-Cloning kit (Enzynomics). The *FOG2* reporter was made in our previous study[27]. Mutant reporters were made by site-directed mutagenesis (QuikChange Site-Directed Mutagenesis Kit, Agilent) at the seed region of the miRNA-binding site (**Supplementary Fig. 4c**).

Exponentially growing cells were seeded onto standard 12-well plates (200,000 cells per well). On the following day, wild-type or mutant reporter was cotransfected with the pRL-CMV vector expressing *Renilla* luciferase, a control for normalization. 24 h after transfection, cells were lysed, and luciferase assay was carried out with the Dual Luciferase Reporter Assay Kit (Promega) according to the manufacturer's protocol. Luciferase activity from the wild-type reporter was normalized by that from the corresponding mutant reporter.

mRNA sequencing. mRNA was enriched from total RNA by oligo(dT) Dynabeads (Life Technologies). The enriched mRNA was randomly fragmented by incubation in fragmentation buffer containing Mg^{2+} ions (New England BioLabs). The mRNA fragments ranging from 70 to 90 nt were fractionated, ligated to adaptors at both 5′ and 3′ ends, and reverse transcribed for PCR amplification. The final PCR product was checked for quality with an Agilent 2100 Bioanalyzer (Agilent), and sequencing by HiSeq 2000 for 50 cycles (Illumina) was then performed.

FASTQ sequences from the sequencer were aligned to the human reference genome (GRCh37.p10) by TopHat2 (ref. 41). After the calculation of reads per kilobase of transcript per million mapped reads (RPKM) for each isoform, representative isoforms for each gene with the highest RPKM were identified. On the basis of these data, the reference genome was rebuilt, and RPKM was recalculated with TopHat2. Fold change was calculated after quantile normalization against RPKM, and LOWESS smoothing was applied for fold change.

For the analyses of motif enrichment and global mRNA-level change, the data from two clones per miRNA KO were combined. In the motif enrichment analysis, the upregulated gene group was selected by 19 cutoffs (5%, 10%, 15%, …, 95% of total genes), and the *P* value for each cutoff was calculated by two-sided Fisher's exact test. The sum of $\log_{10}(P \text{ value})$ over the 19 cutoffs was calculated. For the analysis of global mRNA-level change, the top quartile of transcripts in terms of their context scores[42] were analyzed to enrich for a more functional subset of transcripts containing miRNA 7–8mers.

38. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
39. Guo, J., Gaj, T. & Barbas, C.F. III. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J. Mol. Biol.* **400**, 96–107 (2010).
40. Pall, G.S. & Hamilton, A.J. Improved northern blot method for enhanced detection of small RNA. *Nat. Protoc.* **3**, 1077–1084 (2008).
41. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
42. Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27**, 91–105 (2007).