In: Gene Silencing: Theory, Techniques and Applications ISBN: 978-1-61728-276-8
Editor: Anthony J. Catalano

Chapter XIII

Genome-wide Identification and Analysis of miRNAs Complementary to Upstream Sequences of mRNA Transcription Start Sites

Kenta Narikawa[1], Kenji Nishi[2], Yuki Naito[2], Minami Mazda[2] and Kumiko Ui-Tei[1,2,*]

[1]Department of Computational Biology, Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba-ken 277-8561, Japan
[2]Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan

Abstract

Small regulatory RNAs including short interfering RNAs (siRNAs) and microRNAs (miRNAs) are crucial regulators of gene expression at the posttranscriptional level. Recently, additional roles for small RNAs in gene activation and suppression at the transcriptional level were reported; these RNAs were shown to have sequences that closely or completely match to their respective promoter regions. However, no global analysis for identifying target sequences for miRNAs in the promoter region have been carried out in the human genome.

We performed a genome-wide search for upstream sequences of mRNA transcription start sites where miRNAs are capable of hybridizing with high complementarity. We identified 219 sites in the 10-kb upstream regions of transcription start sites with complete complementarity to 94 human mature miRNAs. Furthermore, the mismatched positions and nucleotides in near-completely matched sites were highly biased, and most

* Email address: ktei@bi.s.u-tokyo.ac.jp

of them appear to be possible target sites of miRNAs. The expression of downstream genes of miRNA target sites were examined following transfection of each miRNA into three different human cell lines. The results indicate that miRNAs dynamically modulates gene expression depending on the downstream genes and the cell type.

Introduction

Small regulatory RNAs including small interfering RNAs (siRNAs) and microRNAs (miRNAs) regulate gene expression at the posttranscriptional level in various organisms [1,2]. RNA silencing encompasses a group of mechanistically related pathways. However, the biogenesis and mechanism of action differ between siRNA and miRNA. In the siRNA-mediated RNA interference (RNAi) pathway, long double-stranded RNAs (dsRNAs) are processed by the cytoplasmic RNase III enzyme Dicer into 21–23-nucleotide (nt) siRNAs. The siRNA guide strand that is incorporated into the RNA-induced silencing complex (RISC) closely or completely matches their target sequences, and a core component of RISC, Argonaute (Ago) protein, mediates the sequence-specific cleavage of target mRNAs [3,4]. In the miRNA pathway, miRNAs are transcribed by RNA polymerases II and III [5]. This generates primary miRNA (pri-miRNA) transcripts, which are further processed to stem-loop-structured precursor miRNAs (pre-miRNAs) by the nuclear RNase III Drosha as part of a microprocessor complex with DGCR8 [6-8]. The pre-miRNA is exported into the cytoplasm by the Exportin-5/Ran-GTP complex [9]. Dicer cleaves the cytoplasmic pre-miRNA hairpin into the mature miRNA, which enters the miRNA-containing ribonucleoprotein (miRNP) and is capable of binding to partially complementary sequences within the 3′ UTR of the target mRNA.

In the past few years, additional roles for small non-coding RNAs to direct both gene activation and suppression at the transcriptional level in nucleus of human cells have been reported. Exogenous siRNAs with sequence complementarity to promoter regions have been shown to induce transcriptional gene silencing (TGS) [10-15]. Meanwhile, siRNA sequences with promoter complementarity also direct transcriptional activation in a process referred to as RNA activation (RNAa) [16-18]. While little is known about the mechanisms and biological effects of this aberrant siRNA behaviour, studies regarding the mechanisms of TGS and RNAa have provided some insight. TGS induced by promoter-directed siRNAs has been shown to require Ago proteins [12,13] and promoter-associated RNAs that span the targeted loci [15,19]. RNAa was shown to be associated with the 5′ end of the siRNA guide strand and Ago2 [17] and antisense transcripts [18,19]. However, siRNA targeted to the HIV-1 LTR promoter was also demonstrated to induce gene activation via an indiscriminate off-target effect [20].

Both TGS and RNAa have been shown to be inducible by endogenous miRNAs that completely or nearly completely match promoters in human cells [21,22]. miR-373 has highly complementary target sequences to the promoters of E-cadherin and CSDC2 (cold shock domain-containing protein C2), and it induces the expression of both genes [22]. Furthermore, miR-320 is encoded within the promoter region of the cell cycle gene POLR3D in the antisense orientation, and transcriptionally silences POLR3D when associated with Ago1, Polycomb-group (PcG) component EZH2, and trimethyl histone H3 lysine 27 (H3K27me3) [21]. These results suggest that endogenous miRNAs sharing complementary

sequences to promoter regions regulate gene expression at the level of transcription. In the human genome, ~700 miRNA genes are distributed in single or clustered fashion and comprise independent transcription units in intergenic regions or are embedded within introns of protein-coding genes [23]. While genome-wide computational and experimental searches for predicting miRNA regulatory targets in the 3′ UTRs but not in the upstream regions have become major challenges [24], but their intended regions have been mainly 3′ UTRs. To date, no global analyses have been carried out to identify miRNAs targeting upstream regions of mRNA coding regions. We performed genome-wide analyses searching for upstream sequences of transcription start sites (TSSs) where miRNAs share complementarity. The changes in the expression of downstream genes due to the miRNAs were examined.

Materials and Methods

Search for Upstream Regions Containing Sequences Complementary to miRNAs

The transcriptional start sites were assumed to be the 5′-terminus of the mRNAs registered in the RefSeq database in this study. The 10-kb sequences upstream of the TSSs were retrieved by matching of the 5′-terminus of mRNA sequences registered in the RefSeq release 27 database (http://www.ncbi.nlm.nih.gov/RefSeq/) [25] with human genome sequences from the UCSC Genome Browser (http://genome.ucsc.edu) [26]. Then, the 10-kb upstream sequences of human mRNA and the miRNA sequences retrieved from the miRBase release 10.0 (http://microrna.sanger.ac.uk/) [27] were aligned, and regions with perfect complementarity in both directions were selected. In addition, 1-kb upstream regions that contained incomplete complementary sequences (i.e., 1- or 2-bp mismatches) with miRNAs were also selected.

Cell Culture

Human carcinoma cell line of the uterine cervix, HeLa, and prostatic carcinoma cell line, PC-3 (Health Science Research Resources Bank, Japan), were cultured in Dulbecco's Modified Eagle's Medium (DMEM; Invitrogen) at 37˚C. Media for both cell lines were supplemented with 10 % heat-inactivated fetal bovine serum (FBS; Mitsubishi Kagaku) and antibiotics (10 units/ml of penicillin (Meiji) and 50 µg/ml of streptomycin (Meiji)). Human MCF-7 breast cancer cell line (Health Science Research Resources Bank, Japan) was cultured in Modified Eagle's Medium (MEM; Invitrogen) supplemented with 10 % heat-inactivated FBS, 100 mM sodium pyruvate (Invitrogen) and 4 mg/ml insulin (Invitrogen).

Plasmid Construction and RNA Silencing Assay

All plasmids constructed are derivatives of psiCHECK-1 (Promega, Madison, Wisconsin). Chemically synthesized double-stranded oligonucleotides containing 81-bp were

inserted into the corresponding restriction enzyme sites of psiCHECK-1 to generate psiCHECK-miR-target. Each oligonucleotide included three tandem repeats of an identical 25-bp containing completely matched target sequences of miR-378, miR-548c-5p, miR-548d-5p, miR-572, miR-574-5p and miR-606 miRNAs and cohesive XhoI/EcoRI ends (Table 1). The inserted miRNA targets were expressed as part of the 3′_UTR of the *Renilla luciferase* mRNA in the transfected cells. Cells in each well of a 24-well culture plate were transfected simultaneously with one of the psiCHECK-miR-target (100 ng), a pGL3-Control (0.5 μg; Promega) and synthetic miRNA (0.005, 0.05, and 0.5 μg; Sigma). Cells were harvested 24 h after transfection and relative *luc* activity (*Renilla luciferase* activity/firefly *luciferase* activity) was determined using the Dual-Luciferase Reporter Assay System (Promega). Unrelated siRNA against the firefly *luciferase* gene (5′-CGUACGCGGAAUACUUCGAUU-3′ and 5′-UCGAAGUAUUCCGCGUACGUG-3′) was used as a miRNA control (dsControl).

Table 1. Oligonucleotide sequences inserted into psiCHECK-miR-target constructs

oligo（-SS）	**sequence (5' ⇒ 3')**
378-SS	tcgaGGCCTTCTGACTCCAAGTCCAGTGCGGCCTTCTGACTCCAA GTCCAGTGCGGCCTTCTGACTCCAAGTCCAGTGC
548c-5p-SS	tcgaTGGCAAAAACCGCAATTACTTTTGCTGGCAAAAACCGCAA TTACTTTTGCTGGCAAAAACCGCAATTACTTTTGC
548d-5p-SS	tcgaTGGCAAAAACCACAATTACTTTTGCTGGCAAAAACCACAA TTACTTTTGCTGGCAAAAACCACAATTACTTTTGC
572-SS	tcgaGGCTGGGCCACCGCCGAGCGGACGAGGCTGGGCCACCGCC GAGCGGACGAGGCTGGGCCACCGCCGAGCGGACGA
574-5p-SS	tcgaCACACACTCACACACACACACTCACCACACACTCACACAC ACACACTCACCACACACTCACACACACACACTCAC
606-SS	tcgaGTATCTTTGATTTTCAGTAGTTTACGTATCTTTGATTTTCAG TAGTTTACGTATCTTTGATTTTCAGTAGTTTAC

oligo（-AS）	**sequence (5' ⇒ 3')**
378-AS	aattGCACTGGACTTGGAGTCAGAAGGCCGCACTGGACTTGGAG TCAGAAGGCCGCACTGGACTTGGAGTCAGAAGGCC
548c-5p-AS	aattGCAAAAGTAATTGCGGTTTTTGCCAGCAAAAGTAATTGCGG TTTTTGCCAGCAAAAGTAATTGCGGTTTTTGCCA
548d-5p-AS	aattGCAAAAGTAATTGTGGTTTTTGCCAGCAAAAGTAATTGTGG TTTTTGCCAGCAAAAGTAATTGTGGTTTTTGCCA
572-AS	aattTCGTCCGCTCGGCGGTGGCCCAGCCTCGTCCGCTCGGCGGT GGCCCAGCCTCGTCCGCTCGGCGGTGGCCCAGCC
574-5p-AS	aattGTGAGTGTGTGTGTGTGAGTGTGTGGTGAGTGTGTGTGTGT GAGTGTGTGGTGAGTGTGTGTGTGTGAGTGTGTG
606-AS	aattGTAAACTACTGAAAATCAAAGATACGTAAACTACTGAAAA TCAAAGATACGTAAACTACTGAAAATCAAAGATAC

AS and SS, respectively, indicate antisense and sense strand sequences of three copies of each mRNA target sequences. Both strands were annealed and inserted into the psiCHECK vector.

Table 2. Stem–loop RT primers specific for each miRNA

name	sequence (5' ⇒ 3')
st21	gtcgtatccagtgcagggtccgaggtattcgcactggatacgactcaaca
st378	gtcgtatccagtgcagggtccgaggtattcgcactggatacgacccttct
st548c-5p	gtcgtatccagtgcagggtccgaggtattcgcactggatacgacggcaaa
st548d-5p	gtcgtatccagtgcagggtccgaggtattcgcactggatacgacggcaaa
st572	gtcgtatccagtgcagggtccgaggtattcgcactggatacgactgggcc
st574-5p	gtcgtatccagtgcagggtccgaggtattcgcactggatacgacacacac
st606	gtcgtatccagtgcagggtccgaggtattcgcactggatacgacatcttt

st21, 378, 548c-5p, 548d-5p, 572, 574-5p and 606 are RT primers for miR-21, miR-378, miR-548c-5p, miR-548d-5p, miR-572, miR-574-5p, and miR-606, respectively.

Table 3. PCR primers used for each miRNA

name	sequence (5' ⇒ 3')
stRT-21fp	gcccgctagcttatcagactgatg
stRT-378fp	gccgccactggacttggagtc
stRT-548cfp	gccgcaaaagtaattgcggt
stRT-548dfp	gccgcgaaaagtaattgtggt
stRT-572fp	gccgcggtccgctcggcggtg
stRT-574fp	gcccgctgagtgtgtgtgtgtgagt
stRT-606fp	gcccgcaaactactgaaaatga
stRT-rp	gtgcagggtccgaggt

stRT-21fp, stRT-378fp, stRT-548c-5pfp, stRT-548d-5pfp, stRT-572fp, stRT-574-5pfp and stRT-606fp are forward primers for PCR of miR-21, miR-378, miR-548c-5p, miR-548d-5p, miR-572, miR-574-5p and miR-606, respectively. stRT-rp is a universal reverse primer.

Table 4. PCR primers for downstream genes

gene name	fp sequence (5' ⇒ 3')	rp sequence (5' ⇒ 3')	PCR product (bp)
OSBPL10	catgctggtggtgtactctgct	cgggagcttggagcactctt	144
RNPS1	tagggctccttcacctaccaaa	cgatccttctcactcgactcct	102
OSBPL9	agcgtccatcttccctacca	tgctactcggtggtgaatgg	105
POLR3A	gacctggagttgccgtgttt	tgatgtggcagcaggttttg	88
ANXA3	gcatctcatggtggcccta	cttcatttgcctgcttgtcc	136
CDC23	ctgcttattgcgggccttac	gcatccatatcctgggcatc	149
RNF6	agattatcggcttatgagagacca	ttccttgacgccatctaacc	94
C20orf43	ctcttgggaaggcagcatct	ctggaggtcatcgtgcttgt	125
actin	cacactgtgcccatctacga	gccatctcttgctcgaagtc	203

PCR primers are indicated by the gene names. fp and rp represent forward and reverse primers, respectively. Expected size of PCR product is shown in the right colume.

Detection of Endogenous miRNAs by RT-PCR

Endogenous miRNAs were detected by RT-PCR according to the procedure reported by Chen et al [28]. Briefly, 1 ml of the HeLa, PC-3, or MCF-7 cell suspensions (containing 1 x

10^5 cells/ml) was inoculated into each well of a 24-well plate 24 h prior to transfection. Cells were transfected without or with chemically synthesized miRNAs (500 ng/well) using Lipofectamine 2000 (Invitrogen, Carlsbad, California). The cells transfected with synthesized miRNAs were used as miRNA detection control. Total RNA was purified 3 days after transfection using the mirVana miRNA Isolation Kit (Ambion, Austin, Texas) according to the manufacturer′s protocol. The reverse transcriptase reactions contained total RNA (0.25 μg), 50 nmoles of stem–loop RT primer specific for each miRNA (Table 2), 1× RT buffer (RETROscript Kit; Ambion), 0.25 mM dNTPs, 3.33 U/μl of MultiScribe reverse transcriptase (Applied Biosystems, Foster City, California), and 0.25 U/μl of RNase OUT (Invitrogen). The mixture was incubated at 16 ºC for 30 min, 42 ºC for 30 min, 85 ºC for 5 min and then held at 4 ºC. The PCR reaction was performed using AmpliTaq Gold PCR Master Mix (Applied Biosystems) for each miRNA using universal reverse primers (Table 3). The mixture was incubated at 95 ºC for 10 min, followed by 40 cycles at 95 ºC for 15 s, and 60 ºC for 1 min. The reactions were fractionated on a 3 % agarose gel with TBE.

Detection of Downstream Gene Expression by Real Time RT-PCR

Using Lipofectamine 2000 (Invitrogen), HeLa, PC-3 and MCF-7 cells were cotransfected with or without 500 ng/well chemically synthesized miRNA. Total RNA was treated with RQ1 DNase (Promega) and purified using RNeasy Mini Kit (QIAGEN) 3 days after transfection. Reverse transcriptase reactions were carried out using total RNA (0.25 μg) with Anchored-oligo$(dT)_{18}$ primer by Transcriptor High Fidelity cDNA Synthesis Kit (Roche) according to the manufacturer's protocol. Real time PCR was performed using FastStart Universal SYBR Green Master (Roche) with forward and reverse primers specific for OSBPL10, RNPS1, OSBPL9, POLR3A, ANXA3, CDC23, RNF6, C20orf43 and actin genes (Table 4). The mixture was incubated at 95 ºC for 10 min, followed by 40 cycles of 95 ºC for 15 sec and 60 ºC for 1 min and monitored by ABI PRISM 7000 (Applied Biosystems). The expression level of each mRNA was first normalized to the amount of actin and then to the siRNA against the firefly *luciferase* (dsControl) transfection control.

Results

Search for miRNAs Having Perfect Complementarity within Upstream Sequences of mRNA Coding Regions

We identified 219 sites in the 10-kb upstream regions of TSSs with complete complementarity to 94 human mature miRNAs (Tables 5 and 6). In this study, the 5′-end of mRNA was assumed to be TSS (position +1). More than one site was included in some of the upstream sequences, and overlapping transcriptional variants were counted at several sites; the 219 sites corresponded to the upstream sequences of 155 genes (Table 5). Among these sites, 153 sites were complementary to pre-miRNAs which generate 88 mature miRNAs registered in the miRBase, indicating that these miRNAs are transcribed from the upstream regions as origins. As shown for miR-320 [21], these miRNAs may have multiple regulatory

functions if the antisense transcript is transcribed at the same position. Except for these cases, the 66 sites of 50 genes (labelled with blue in Table 6) appeared perfect complementarity to the miRNAs (miR-378, miR-548c-5p, miR-548d-5p, miR-572, miR-574-5p and miR-606) transcribed from the different location other than the 10-kb upstream regions.

Table 5. Results of the search for miRNAs with complementarity to upstream sequences of transcription start sites

	miRNAs	**Complementary sites*** **	**Genes**
Total number	722	24,854	18,578
Perfect match containing miRNA transcription sites	94* (20**)	219* (36**)	155* (26**)
Perfect match	6* (4**)	66* (13**)	50* (10**)
1-bp mismatch	10**	69**	53**
2-bp mismatch	44**	4453**	763**

The search was carried out using 722 mature miRNA sequences against 24,854 sequences 10-kb upstream of 18,578 genes transcription start sites. * indicates the outcome of the 10-kb upstream sequences; ** denotes the 1-kb upstream sequences. *** Note that some of the upstream sequences have multiple complementary sites. Details of completely complementary and 1-bp mismatched sites are shown in Tables 6 and 7.

Table 6. List of miRNAs and genes with perfectly matched sequences in the 10-kb upstream regions

let-7i	**gene**	**ID**	**position**		
UGAGGUAGUAGUUUGUGCUGUU	C12orf61	NM_175895	-257	+	origin

let-7i*	**gene**	**ID**	**position**		
CUGCGCAAGCUACUGCCUUGCU	C12orf61	NM_175895	-313	+	origin

miR-1	**gene**	**ID**	**position**		
UGGAAUGUAAAGAAGUAUGUAU	C20orf200	NM_152757	-2827	+	origin

miR-7	**gene**	**ID**	**position**		
UGGAAGACUAGUGAUUUUGUUGU	ISG20L1	NM_022767	-9502	-	origin

miR-7-2*	**gene**	**ID**	**position**		
CAACAAAUCCCAGUCUACCUAA	ISG20L1	NM_022767	-9462	-	origin

miR-10a	**gene**	**ID**	**position**		
UACCCUGUAGAUCCGAAUUUGUG	HOXB3	NM_002146	-5478	-	origin
	HOXB4	NM_024015	-1545	-	origin

Table 6. (Continued)

miR-10a*	gene	ID	position		
CAAAUUCGUAUCUAGGGGAAUA	HOXB3	NM_002146	-5437	-	origin
	HOXB4	NM_024015	-1504	-	origin

miR-10b	gene	ID	position		
UACCCUGUAGAACCGAAUUUGUG	HOXD4	NM_014621	-1056	-	origin

miR-10b*	gene	ID	position		
ACAGAUUCGAUUCUAGGGGAAU	HOXD4	NM_014621	-1017	-	origin

miR-15b	gene	ID	position		
UAGCAGCACAUCAUGGUUUACA	IFT80	NM_020800	-5075	+	origin

miR-15b*	gene	ID	position		
CGAAUCAUUAUUUGCUGCUCUA	IFT80	NM_020800	-5113	+	origin

miR-16	gene	ID	position		
UAGCAGCACGUAAAUAUUGGCG	IFT80	NM_020800	-5222	+	origin

miR-16-2*	gene	ID	position		
CCAAUAUUACUGUGCUGCUUUA	IFT80	NM_020800	-5265	+	origin

miR-22	gene	ID	position		
AAGCUGCCAGUUGAAGAACUGU	WDR81	NM_152348	-2601	+	origin

miR-22*	gene	ID	position		
AGUUCUUCAGUGGCAAGCUUUA	WDR81	NM_152348	-2563	+	origin

miR-25	gene	ID	position		
CAUUGCACUUGUCUCGGUCUGA	AP4M1	NM_004722	-7915	+	origin

miR-25*	gene	ID	position		
AGGCGGAGACUUGGGCAAUUG	AP4M1	NM_004722	-7877	+	origin

miR-26a	gene	ID	position		
UUCAAGUAAUCCAGGAUAGGCU	AVIL	NM_006576	-8610	-	origin

miR-26a-2*	gene	ID	position		
CCUAUUCUUGAUUACUUGUUUC	AVIL	NM_006576	-8572	-	origin

Table 6. (Continued)

miR-34b	gene	ID	position		
CAAUCACUAACUCCACUGCCAU	BTG4	NM_017589	-648	+	origin
	FLJ46266	NM_001100388	-1798	-	origin
	FLJ46266	NM_207430	-1798	-	origin

miR-34b*	gene	ID	position		
UAGGCAGUGUCAUUAGCUGAUUG	BTG4	NM_017589	-611	+	origin
	FLJ46266	NM_001100388	-1835	-	origin
	FLJ46266	NM_207430	-1835	-	origin

miR-34c-3p	gene	ID	position		
AAUCACUAACCACACGGCCAGG	BTG4	NM_017589	-1145	+	origin
	FLJ46266	NM_001100388	-1301	-	origin
	FLJ46266	NM_207430	-1301	-	origin

miR-34c-5p	gene	ID	position		
AGGCAGUGUAGUUAGCUGAUUGC	BTG4	NM_017589	-1112	+	origin
	FLJ46266	NM_001100388	-1334	-	origin
	FLJ46266	NM_207430	-1334	-	origin

miR-92b	gene	ID	position		
UAUUGCACUCGUCCCGGCCUCC	MUC1	NM_001018016	-2328	+	origin
	MUC1	NM_001018017	-2328	+	origin
	MUC1	NM_001044390	-2328	+	origin
	MUC1	NM_001044391	-2328	+	origin
	MUC1	NM_001044392	-2328	+	origin
	MUC1	NM_001044393	-2328	+	origin
	MUC1	NM_002456	-2328	+	origin

miR-92b*	gene	ID	position		
AGGGACGGGACGCGGUGCAGUG	MUC1	NM_001018016	-2287	+	origin
	MUC1	NM_001018017	-2287	+	origin
	MUC1	NM_001044390	-2287	+	origin
	MUC1	NM_001044391	-2287	+	origin
	MUC1	NM_001044392	-2287	+	origin
	MUC1	NM_001044393	-2287	+	origin
	MUC1	NM_002456	-2287	+	origin

miR-93	gene	ID	position		
CAAAGUGCUGUUCGUGCAGGUAG	AP4M1	NM_004722	-7670	+	origin

miR-93*	gene	ID	position		
ACUGCUGAGCUAGCACUUCCCG	AP4M1	NM_004722	-7709	+	origin

Table 6. (Continued)

miR-106b	**gene**	**ID**	**position**		
UAAAGUGCUGACAGUGCAGAU	AP4M1	NM_004722	-7444	+	origin

miR-106b*	**gene**	**ID**	**position**		
CCGCACUGUGGGUACUUGCUGC	AP4M1	NM_004722	-7484	+	origin

miR-132	**gene**	**ID**	**position**		
UAACAGUCUACAGCCAUGGUCG	HIC1	NM_001098202	-6360	+	origin
	HIC1	NM_006497	-5149	+	origin

miR-132*	**gene**	**ID**	**position**		
ACCGUGGCUUUCGAUUGUUACU	HIC1	NM_001098202	-6324	+	origin
	HIC1	NM_006497	-5113	+	origin

miR-138	**gene**	**ID**	**position**		
AGCUGGUGUUGUGAAUCAGGCCG	SLC12A3	NM_000339	-6703	-	origin

miR-138-2*	**gene**	**ID**	**position**		
GCUAUUUCACGACACCAGGGUU	SLC12A3	NM_000339	-6656	-	origin

miR-141	**gene**	**ID**	**position**		
UAACACUGUCUGGUAAAGAUGG	EMG1	NM_006331	-6676	-	origin

miR-141*	**gene**	**ID**	**position**		
CAUCUUCCAGUACAGUGUUGGA	EMG1	NM_006331	-6718	-	origin

miR-142-3p	**gene**	**ID**	**position**		
UGUAGUGUUUCCUACUUUAUGGA	BZRAP1	NM_004758	-3150	-	origin

miR-142-5p	**gene**	**ID**	**position**		
CAUAAAGUAGAAAGCACUACU	BZRAP1	NM_004758	-3186	-	origin

miR-146b-3p	**gene**	**ID**	**position**		
UGCCCUGUGGACUCAGUUCUGG	CUEDC2	NM_024040	-3890	+	origin

miR-146b-5p	**gene**	**ID**	**position**		
UGAGAACUGAAUUCCAUAGGCU	CUEDC2	NM_024040	-3854	+	origin

miR-181c	**gene**	**ID**	**position**		
AACAUUCAACCUGUCGGUGAGU	NANOS3	NM_001098622	-2411	-	origin

miR-181c*	**gene**	**ID**	**position**		
AACCAUCGACCGUUGAGUGGAC	NANOS3	NM_001098622	-2373	-	origin

Table 6. (Continued)

miR-181d	gene	ID	position		
AACAUUCAUUGUUGUCGGUGGGU	NANOS3	NM_001098622	-2226	-	origin

miR-190b	gene	ID	position		
UGAUAUGUUUGAUAUUGGGUU	TPM3	NM_152263	-1600	-	origin

miR-191	gene	ID	position		
CAACGGAAUCCCAAAAGCAGCUG	C3orf60	NM_199069	-947	+	origin
	C3orf60	NM_199070	-947	+	origin
	C3orf60	NM_199073	-491	+	origin
	C3orf60	NM_199417	-947	+	origin
	DALRD3	NM_001009996	-2110	-	origin

miR-191*	gene	ID	position		
GCUGCGCUUGGAUUUCGUCCCC	C3orf60	NM_199069	-989	+	origin
	C3orf60	NM_199070	-989	+	origin
	C3orf60	NM_199073	-533	+	origin
	C3orf60	NM_199417	-989	+	origin
	DALRD3	NM_001009996	-2068	-	origin

miR-195	gene	ID	position		
UAGCAGCACAGAAAUAUUGGC	BCL6B	NM_181844	-5363	+	origin

miR-195*	gene	ID	position		
CCAAUAUUGGCUGUGCUGCUCC	BCL6B	NM_181844	-5401	+	origin

miR-196a	gene	ID	position		
UAGGUAGUUUCAUGUUGUUGGG	HOXB9	NM_024017	-6080	-	origin
	HOXC9	NM_006897	-8331	-	origin

miR-196a*	gene	ID	position		
CGGCAACAAGAAACUGCCUGAG	HOXC9	NM_006897	-8294	-	origin

miR-196b	gene	ID	position		
UAGGUAGUUUCCUGUUGUUGGG	HOXA9	NM_152739	-4019	-	origin

miR-200c	gene	ID	position		
UAAUACUGCCGGGUAAUGAUGGA	EMG1	NM_006331	-7089	-	origin

miR-200c*	gene	ID	position		
CGUCUUACCCAGCAGUGUUUGG	EMG1	NM_006331	-7128	-	origin

Table 6. (Continued)

miR-208b	gene	ID	position		
AUAAGACGAACAAAAGGUUUGU	MYH6	NM_002471	-9745	-	origin

miR-210	gene	ID	position		
CUGUGCGUGUGACAGCGGCUGA	C11orf35	NM_173573	-7354	-	origin
	KIAA1542	NM_020901	-8353	+	origin

miR-212	gene	ID	position		
UAACAGUCUCCAGUCACGGCC	HIC1	NM_001098202	-6000	+	origin
	HIC1	NM_006497	-4789	+	origin

miR-219-1-3p	gene	ID	position		
AGAGUUGAGUCUGGACGUCCCG	RING1	NM_002931	-613	-	origin
	RXRB	NM_021976	-7237	+	origin
	RXRB	NM_021976	-7241	+	origin
	RXRB	NM_021976	-7245	+	origin

miR-219-5p	gene	ID	position		
UGAUUGUCCAAACGCAAUUCU	RING1	NM_002931	-654	-	origin
	RXRB	NM_021976	-7196	+	origin
	RXRB	NM_021976	-7200	+	origin
	RXRB	NM_021976	-7204	+	origin

miR-301a	gene	ID	position		
CAGUGCAAUAGUAUUGUCAAAGC	PRR11	NM_018304	-4561	+	origin

miR-320	gene	ID	position		
AAAAGCUGGGUUGAGAGGGCGA	POLR3D	NM_001722	-110	+	origin

miR-324-3p	gene	ID	position		
ACUGCCCCAGGUGCUGCUGG	DLG4	NM_001365	-3589	-	origin

miR-324-5p	gene	ID	position		
CGCAUCCCCUAGGGCAUUGGUGU	DLG4	NM_001365	-3626	-	origin

miR-345	gene	ID	position		
GCUGACUCCUAGUCCAGGGCUC	SLC25A29	NM_001039355	-1353	+	origin

miR-375	gene	ID	position		
UUUGUUCGUUCGGCUCGCGUGA	CRYBA2	NM_005209	-8270	-	origin
	CRYBA2	NM_057093	-8264	-	origin
	CRYBA2	NM_057094	-8264	-	origin

Table 6. (Continued)

miR-378	gene	ID	position		
ACUGGACUUGGAGUCAGAAGG	OSBPL10	NM_017784	-4582	-	

miR-425	gene	ID	position		
AAUGACACGAUCACUCCCGUUGA	C3orf60	NM_199069	-1420	+	origin
	C3orf60	NM_199070	-1420	+	origin
	C3orf60	NM_199073	-964	+	origin
	C3orf60	NM_199074	-254	+	origin
	C3orf60	NM_199417	-1420	+	origin
	DALRD3	NM_001009996	-1637	-	origin

miR-425*	gene	ID	position		
AUCGGGAAUGUCGUGUCCGCCC	C3orf60	NM_199069	-1461	+	origin
	C3orf60	NM_199070	-1461	+	origin
	C3orf60	NM_199073	-1005	+	origin
	C3orf60	NM_199074	-295	+	origin
	C3orf60	NM_199417	-1461	+	origin
	DALRD3	NM_001009996	-1596	-	origin

miR-483-3p	gene	ID	position		
UCACUCCUCUCCUCCCGUCUU	IGF2AS	NM_016412	-6345	+	origin

miR-483-5p	gene	ID	position		
AAGACGGGAGGAAAGAAGGGAG	IGF2AS	NM_016412	-6305	+	origin

miR-484	gene	ID	position		
UCAGGCUCAGUCCCCUCCCGAU	KIAA0430	NM_014647	-149	+	origin
	NDE1	NM_017668	-6946	-	origin

miR-497	gene	ID	position		
CAGCAGCACACUGUGGUUUGU	BCL6B	NM_181844	-5051	+	origin

miR-497*	gene	ID	position		
CAAACCACACUGUGGUGUUAGA	BCL6B	NM_181844	-5091	+	origin

miR-548c-5p	gene	ID	position		
AAAAGUAAUUGCGGUUUUUGCC	ABTB1	NM_032548	-8217	-	
	ABTB1	NM_172027	-8217	-	
	ABTB1	NM_172028	-8217	-	
	ANXA3	NM_005139	-248	+	
	C1orf94	NM_032884	-1735	+	
	C20orf43	NM_016407	-8342	-	
	CDC23	NM_004661	-6363	+	
	CDC42EP4	NM_012121	-3863	-	
	EBI2	NM_004951	-8486	+	

Table 6. (Continued)

miR-548c-5p	gene	ID	position		
	FGF23	NM_020638	-9861	+	
	FOXC2	NM_005251	-8776	+	
	GNB5	NM_006578	-3189	-	
	GPR141	NM_181791	-8087	+	
	LBP	NM_004139	-6809	-	
	MANBAL	NM_001003897	-4841	-	
	MANBAL	NM_022077	-4841	-	
	MTHFSD	NM_022764	-3325	-	
	NLGN4X	NM_020742	-9292	-	
	NLGN4X	NM_181332	-8474	-	
	NR0B1	NM_000475	-5731	+	
	NSUN5	NM_018044	-6978	+	
	NSUN5	NM_148956	-6978	+	
	OR2T8	NM_001005522	-7046	+	
	PDE6C	NM_006204	-3974	-	
	PLAC1	NM_021796	-1244	+	
	PSTPIP2	NM_024430	-798	+	
	RBP4	NM_006744	-7378	+	
	RNF6	NM_005977	-7624	-	
	RNF6	NM_183043	-8154	-	
	RNF6	NM_183044	-8292	-	
	UBQLN2	NM_013444	-1751	+	
	ZNF331	NM_001079907	-9126	-	
	ZNF418	NM_133460	-732	+	

miR-548d-5p	gene	ID	position		
AAAAGUAAUUGUGGUUUUUGCC	ADAM23	NM_003812	-7057	-	
	ALDH1L2	NM_001034173	-914	+	
	C1orf94	NM_032884	-1790	-	
	C8B	NM_000066	-333	-	
	CLEC12A	NM_138337	-525	+	
	CLEC12A	NM_201623	-525	+	
	FSTL5	NM_020116	-9873	+	
	GC	NM_000583	-6013	+	
	GJB3	NM_001005752	-3678	+	
	GJB3	NM_024009	-2557	+	
	MAPKAPK 5	NM_003668	-2843	+	
	MAPKAPK 5	NM_139078	-2843	+	
	NANOS1	NM_199461	-9261	-	
	NLRP4	NM_134444	-8220	+	
	POLR3A	NM_007055	-8654	-	
	SOX30	NM_007017	-621	+	
	SOX30	NM_178424	-621	+	
	SPAG11B	NM_058206	-5668	-	
	SPAG11B	NM_058207	-5668	-	
	TAS2R39	NM_176881	-2594	+	
	WNT5A	NM_003392	-7212	+	
	WWP2	NM_199424	-2131	-	

miR-565	gene	ID	position		
GGCUGGCUCGCGAUGUCUGUUU	SACM1L	NM_014016	-443	+	origin

miR-572	gene	ID	position		
GUCCGCUCGGCGGUGGCCCA	RNPS1	NM_006711	-24	-	

miR-574-5p	gene	ID	position		
UGAGUGUGUGUGUGAGUGUGU	JPH2	NM_020433	-1016	+	
	JPH2	NM_175913	-1016	+	
	LOC645037 (GAGE2C)	NM_001098411	-6003	+	
	OPN4	NM_001030015	-283	+	
	OPN4	NM_033282	-283	+	
	PRKCE	NM_005400	-815	+	

miR-580	gene	ID	position		
UUGAGAAUGAUGAAUCAUUAGG	SKP2	NM_005983	-4159	+	origin
	SKP2	NM_032637	-4159	+	origin

miR-584	gene	ID	position		
UUAUGGUUUGCCUGGGACUGAG	NM_001101342	NM_001101342	-923	+	origin

miR-596	gene	ID	position		
AAGCCUGCCCGGCUCCUCGGG	ARHGEF10	NM_014629	-6737	-	origin

miR-606	gene	ID	position		
AAACUACUGAAAAUCAAAGAU	OSBPL9	NM_148904	-5101	+	
	OSBPL9	NM_148905	-5101	+	
	OSBPL9	NM_148907	-5131	+	

miR-607	gene	ID	position		
GUUCAAAUCCAGAUCUAUAAC	LCOR	NM_032440	-4276	+	origin

miR-611	gene	ID	position		
GCGAGGACCCCUCGGGGUCUGAC	FEN1	NM_004111	-156	+	origin

miR-616	gene	ID	position		
AGUCAUUGGAGGGUUUGAGCAG	MBD6	NM_052897	-3672	+	origin

miR-616*	gene	ID	position		
ACUCAAAACCCUUCAGUGACUU	MBD6	NM_052897	-3632	+	origin

Table 6. (Continued)

miR-632	gene	ID	position		
GUGUCUGCUUCCUGUGGGA	C17orf75	NM_022344	-7999	+	origin

miR-636	gene	ID	position		
UGUGCUUGCUCGUCCCGCCCGCA	JMJD6	NM_001081461	-9689	-	origin
	JMJD6	NM_015167	-9689	-	origin
	MFSD11	NM_024311	-1213	+	origin

miR-649	gene	ID	position		
AAACCUGUGUUGUUCAAGAGUC	SLC7A4	NM_004173	-1654	-	origin

miR-658	gene	ID	position		
GGCGGAGGGAAGUAGGUCCGUUGGU	EIF3EIP	NM_016091	-5052	+	origin

miR-659	gene	ID	position		
CUUGGUUCAGGGAGGGUCCCCA	ANKRD54	NM_138797	-3418	-	origin
	EIF3EIP	NM_016091	-1649	+	origin

miR-661	gene	ID	position		
UGCCUGGGUCUCUGGCCUGCGCGU	PLEC1	NM_201381	-492	-	origin
	PLEC1	NM_201382	-1332	-	origin
	PLEC1	NM_201383	-2705	-	origin
	PLEC1	NM_201384	-5639	-	origin

miR-801	gene	ID	position		
GAUUGCUCUGCGUGCGGAAUCGAC	TAF12	NM_005644	-5568	+	origin

miR-923	gene	ID	position		
GUCAGCGGAGGAAAAGAAACU	NLE1	NM_001014445	-8906	-	origin
	NLE1	NM_018096	-8906	-	origin

miR-935	gene	ID	position		
CCAGUUACCGCUUCCGCUACCGC	CACNG6	NM_031897	-9926	-	origin
	CACNG6	NM_145814	-9926	-	origin
	CACNG6	NM_145815	-9926	-	origin

miR-940	gene	ID	position		
AAGGCAGGGCCCCCGCUCCCC	RNPS1	NM_006711	-4010	+	origin
	RNPS1	NM_080594	-3693	+	origin

Mature miRNAs and genes with perfectly matched sequences in regions 10-kb upstream of the TSSs. The nucleotide sequence below each miRNA designation shows the mature miRNA sequence. The gene name indicates the gene located downstream of the miRNA complementary site. ID indicates the accession ID of each transcript. Position indicates the nucleotide position relative to the 5′-end of the miRNA counted from the transcription start site (+1). The +/− shows the orientation of the miRNA target sequences; this symbol indicates the relative direction with respect to downstream mRNA transcription (+/− = same/opposite). In the

most right colume, "origin" indicates the locus from which own miRNA is originally transcribed. Blue indicates the sites appearing perfectly complementarity to the miRNAs transcribed from the different locations.

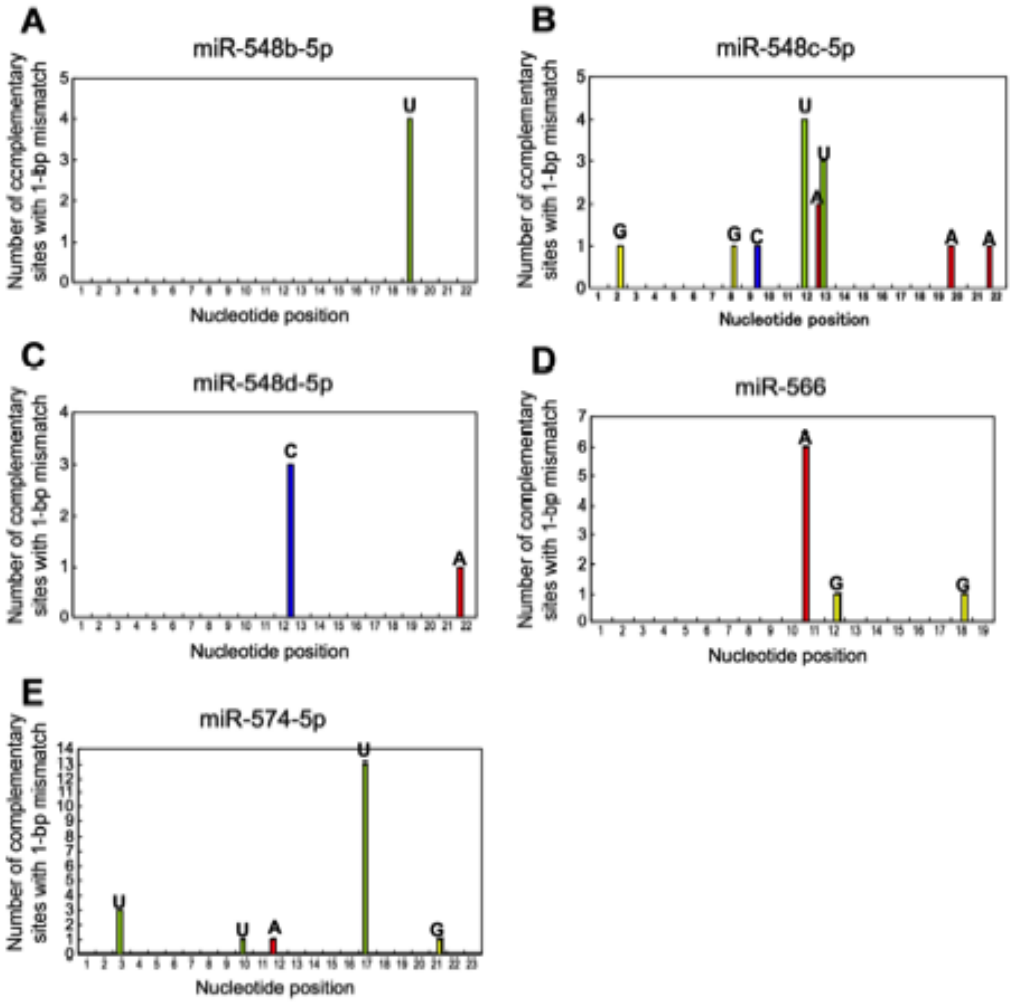

Figure 1. Positions containing 1-bp mismatches in 1-kb upstream regions of mRNA coding regions.

The mismatched positions in miR-548b-5p (A), miR-548c-5p (B), miR-548d-5p (C), miR-566 (D) and miR-574-5p (E) with target genes within the 1-kb upstream regions of the mRNA coding regions are shown from the 5′-end (+1) of the guide strand. The red bar indicates miRNA that is perfectly complementary to the target sequence when the nucleotide is changed to A. Green bar, U; Yellow, G; Blue, C.

Search for and Analysis of miRNAs with Complementary Sequences Containing 1- and 2-bp Mismatches within 1-kb of the Upstream Sequences

In a study on RNAi, target mRNAs carrying a few base mismatches with siRNAs have been demonstrated to be silenced to various degrees according to the position and the identity

of the mismatched nucleotide [29, 30]. Furthermore, miRNAs recognize target mRNA by the "seed sequence" positioned between nucleotides 2–8 from the miRNA 5′-end [31-35], but not by the entire sequence. In RNAi, the identities of off-targets are also determined by complementarity to the seed regions of the introduced siRNA guide strand [36-38]. In a study on RNAa, a sequence located at position –645 relative to the TSS in the E-cadherin promoter was reported to function, although it has a few base mismatches with miR-373 [22]. Then, we searched for 1- and 2-bp mismatched positions and nucleotides of human miRNAs with the 1-kb of the upstream sequences.

We found 69 sites that had 1-bp mismatched sequences to 10 miRNAs in the 1-kb upstream regions of 53 genes (Tables 5 and 7). The complementary sites of five miRNAs (miR-548b-5p, miR-548c-5p, miR-548d-5p, miR-566 and miR-574-5p) recognized multiple 1-bp mismatched sites; the mismatched positions and nucleotides of these miRNAs are summarized in Figure 1. The major mismatched sites of miR-548b-5p and miR-574-5p were identical at nucleotide positions 19 and 17, respectively, from the 5′-end of each miRNA (Figure 1A and E). When G is changed to U at position 19 in miR-548b-5p or A is changed to U at position 17 in miR-574-5p, the majority of these miRNAs can pair with perfect complementarity, indicating mismatched nucleotides are remarkably characteristic. The major mismatched nucleotides in miR-548c-5p, miR-548d-5p and miR-566 were found at position 11 or 12 corresponding to the central region of the miRNAs (Figure 1B–D). In total, 87 % of 1-bp mismatched nucleotides were found in the non-seed region, and the remaining 13 % of them were mismatched in the seed region. Thus, the mismatched positions and nucleotides were highly biased. Total of 4453 target sites containing 2-bp mismatches with 44 miRNAs were found (Table 5). Among them, miR-548c-3p, miR-548d-5p, miR-566 and miR-574-5p had more than ten mismatched target sites (Figure 2). The major mismatched positions in the target sequences were unevenly positioned at 11 and 18 in miR-566 (212 of 276 sites (77 %), Figure 2A). In the cases of miR-548d-5p and miR-548c-3p, the mismatched nucleotides were concentrated at position 12 and 7, respectively (Figure 2B and C). The miR-574-5p (5′-UGAGUGUGUGUGUGUGAGUGUGU-3′) is highly GU repeated sequence without the nucleotides at positions 3 and 17 where Us are changed to As. The 3753 of 4060 mismatched target sites (92 %) of miR-574-5p showed simultaneous nucleotide changes from As to Us at positions 3 and 17 (Figure 2D), suggesting that these sites might be highly distributed GU repeated sequences. In total, 2-bp mismatched positions and nucleotides are also remarkably biased similar to the 1-bp mismatched sites. In total, only 12.0 % of 2-bp mismatched nucleotide positions were found in the seed region, and the remaining 88.0 % of them were mismatched in the non-seed region, except for the miR-574-5p mismatched sites.

Changes in Gene Expression by Targeting Upstream Sequences with miRNA

Six miRNAs, which transcribed from the different locations other than the upstream regions, were found to be completely complementary at 66 sites in the 10-kb upstream regions of 50 genes (Tables 5 and 6). To investigate whether miRNAs complementary to upstream sequences of mRNAs are general regulators of downstream genes, chemically synthesized miRNAs were transfected into the cells and changes in the expression of downstream genes were measured by real-time PCR. Similar to endogenous miRNAs, the

synthetic mature miRNAs were annealed with opposite strand RNAs to form duplex RNAs (Figure 3A).

To investigate the effects of synthetic miRNAs against the completely matched sequences in the 3′ UTRs, psiCHECK-miR-target constructs containing three copies of completely matched target sequences in the 3′ UTR of the *Renilla luciferase* gene were made (Figure 3B). The synthetic miRNA and the corresponding psiCHECK-miR-target construct were transfected simultaneously into HeLa cells, and *luciferase* activity was measured. All of six miRNAs (miR-378, miR-548c-5p, miR-548d-5p, miR-572, miR-574-5p, and miR-606) reduced *Renilla luciferase* activity (Figure 3C), indicating that these synthetic miRNAs can repress gene expression when completely matched sequences are contained in the 3′ UTRs.

Next, the endogenous expression of six miRNAs labelled with blue in Table 6 were examined by RT-PCR in human HeLa, PC-3, and MCF-7 cells. For detection of miRNAs by RT-PCR, we used miRNA-specific stem–loop primers for reverse transcription according to the method reported by Chen et al. [28]. The synthetic miRNAs transfected into HeLa cells were detected using stem–loop RT primers and PCR primers specific for each miRNA. Endogenous miR-378 and miR-574-5p were detected in all three cell lines, but miR-548c-5p, miR-548d-5p, miR-572 and miR-606 were not detected in any of the cell lines tested (Figure 4).

We examined the downstream gene expression changes caused by the introduction of miRNAs; the expression of 8 genes that had sequences corresponding to the transfected miRNAs in their upstream regions were determined by real-time RT-PCR (Figure 5). The expression of OSBPL10 was reduced significantly in two of three cell lines by transfection with miR-378 (Figure 5A). The expression levels of four genes, RNF6, RNPS1, C20orf43 and ANXA3 were lowered in one of three or two cell lines by miR-548c-5p, miR-572, miR-548c-5p and miR-548c-5p, respectively (Figure 5B-E). The expression levels of OSBPL9 and CDC23 genes were significantly increased by transfection of miR-606 and miR-548c-5p, respectively, in one of three cell lines (Figure 5F,G). Thus, the mRNA expression was differentially regulated in each cell type. Furthermore, the effects of miRNA transfection varied depending on the downstream genes. In PC-3 cells, RNF6, C20orf43 were reduced by miR-548c-5p transfection, but ANXA3 and the other genes (GNB5, MANBAL, and RBP4) containing perfect complementary sites with miR-548c-5p in the upstream regions were not reduced (Figure 5, data not shown). POLR3A was decreased with miR-548d-5p, but the expression levels of the other genes (MAPKAPK5 and GJB3) were not changed in PC-3 cells (Figure 5, data not shown). In this study, when mRNA levels were significantly increased, the changes were not greater than ~30 % in any instances. Similarly, miRNA transfection resulted in ~40 % downregulation of mRNAs in all cases of reduction. Although the rate of change of gene expression was not so abundant, the effect of miRNA was obviously dose-dependent (Figure 6). Note that miR-378 and miR-574-5p are endogenously expressed in all three cell lines, which could lead to an underestimation of their effects on gene expressions.

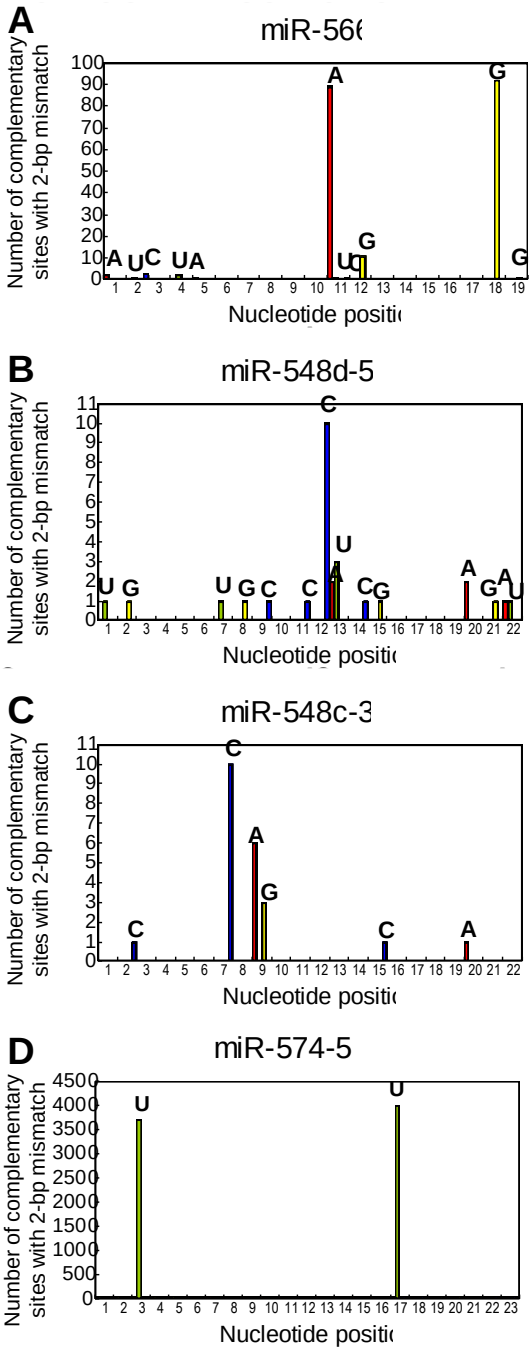

Figure 2. Positions containing 2-bp mismatches in 1-kb upstream regions of mRNA coding regions.
The mismatched positions in miR-566 (A), miR-548d-5p (B), miR-548c-3p (C) and miR-574-5p (D) with target sites within the 1-kb upstream sequences of the mRNA coding regions are shown from the 5′-end (+1) of the guide strand. The red bar indicates miRNA that is perfectly complementary to the target sequence when the nucleotide is changed to A. Green bar, U; Yellow, G; Blue, C.

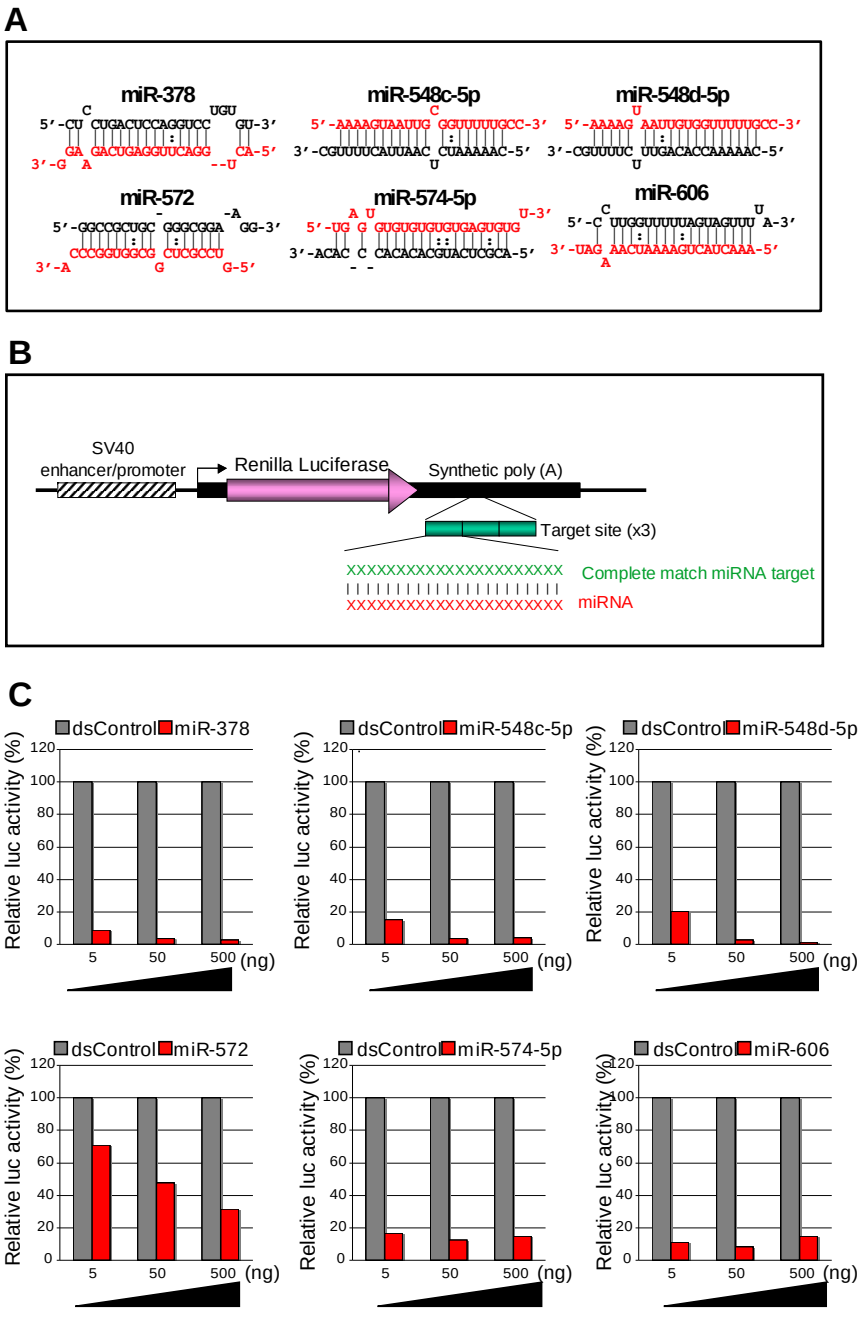

Figure 3. Effects of miRNAs against completely matched targets in 3′ UTRs.
(A) Chemically synthesized miRNAs; the red characters indicate mature miRNAs. (B) The schematic structure of the target sequence expression construct, psiCHECK-miR-target, which contains three copies of matched target sequences in the 3′ UTR of the *Renilla luciferase* gene. (C) Silencing activity of synthetic miRNAs; the activity was monitored using psiCHECK-miR-target constructs. The silencing activity is shown as the percentage of relative *luc* activity normalized to the dsControl transfection (100 %).

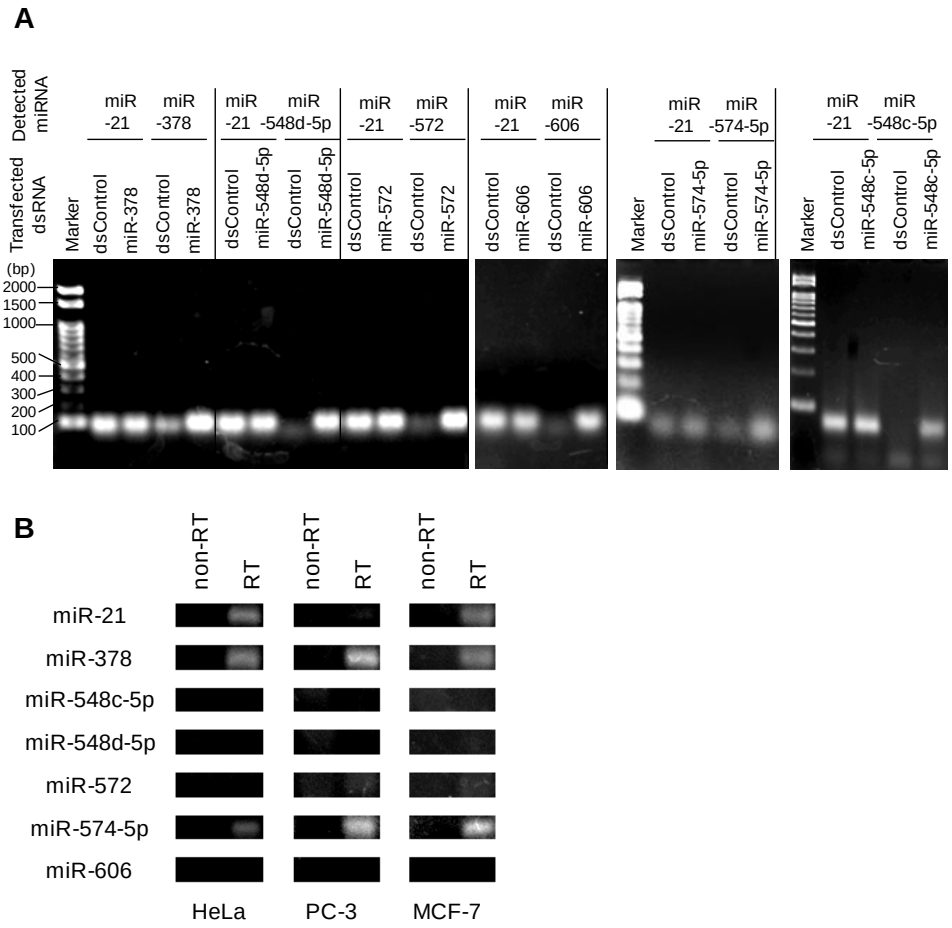

Figure 4. Detection of endogenous miRNAs by RT-PCR.
(A) Detection of miRNAs by stem-loop RT primers specific for each miRNA. The synthetic miRNAs were transfected into HeLa cells, and each miRNA was detected with specific stem=loop RT primers. miR-21 was used as transfection and detection control. Expression of each endogenous miRNA was detected when dsControl alone was transfected and intended miRNA was detected by specific primers for RT-PCR. Expected sizes of each miRNA is 64, 61, 60, 61, 61, 65 and 62 bp for miR-21, miR-378, miR-548c-5p, miR-548d-5p, miR-572, miR-574-5p and miR-606, respectively. (B) Endogenous miRNAs detected using stem-loop RT-PCR in HeLa, PC-3, and MCF-7 cells.

Table 7. List of miRNAs and genes with 1-bp mismatched complementary sequences within the 1-kb upstream regions

miR-190	gene	ID	position	sub	sequence (5' ⇒ 3')	
UGAUAUGUUUG AUAUAUUAGGU	RNF212	NM_194439	-974	7-U	UGAUAUUUUUGA UAUAUUAGGU	+

miR-297	gene	ID	position	sub	sequence (5' ⇒ 3')	
AUGUAUGUGUG CAUGUGCAUG	C12orf54	NM_152319	-151	15-A	AUGUAUGUGUGC AUAUGCAUG	+

miR-548b-5p	gene	ID	position	sub	sequence (5' ⇒ 3')	
AAAAGUAAUUG UGGUUUUGGCC	ALDH1L2	NM_001034173	-914	19-U	AAAAGUAAUUGU GGUUUUUGCC	+
	C8B	NM_000066	-333	19-U	AAAAGUAAUUGU GGUUUUUGCC	-
	CLEC12A	NM_138337	-525	19-U	AAAAGUAAUUGU GGUUUUUGCC	+
	CLEC12A	NM_201623	-525	19-U	AAAAGUAAUUGU GGUUUUUGCC	+
	SOX30	NM_007017	-621	19-U	AAAAGUAAUUGU GGUUUUUGCC	+
	SOX30	NM_178424	-621	19-U	AAAAGUAAUUGU GGUUUUUGCC	+

miR-548c-5p	gene	ID	position	sub	sequence (5' ⇒ 3')	
AAAAGUAAUUG CGGUUUUUGCC	ACTR10	NM_018477	-735	13-U	AAAAGUAAUUGC UGUUUUUGCC	+
	ALDH1L2	NM_001034173	-914	12-U	AAAAGUAAUUGU GGUUUUUGCC	+
	AOX1	NM_001159	-468	20-A	AAAAGUAAUUGC GGUUUUUACC	+
	C5orf32	NM_032412	-715	13-U	AAAAGUAAUUGC UGUUUUUGCC	-
	C8B	NM_000066	-333	12-U	AAAAGUAAUUGU GGUUUUUGCC	-
	CCDC90A	NM_001031713	-913	8-G	AAAAGUAGUUGC GGUUUUUGCC	-
	CLEC12A	NM_138337	-525	12-U	AAAAGUAAUUGU GGUUUUUGCC	+
	CLEC12A	NM_201623	-525	12-U	AAAAGUAAUUGU GGUUUUUGCC	+
	LPO	NM_006151	-167	22-A	AAAAGUAAUUGC GGUUUUUGCA	-
	MS4A4A	NM_148975	-446	13-A	AAAAGUAAUUGC AGUUUUUGCC	+
	OR10W1	NM_207374	-104	13-A	AAAAGUAAUUGC AGUUUUUGCC	-
	PGM2L1	NM_173582	-789	13-U	AAAAGUAAUUGC UGUUUUUGCC	+
	PSMA1	NM_002786	-454	9-C	AAAAGUAACUGC GGUUUUUGCC	+
	PTER	NM_001001484	-887	2-G	AGAAGUAAUUGC GGUUUUUGCC	-
	PTER	NM_030664	-887	2-G	AGAAGUAAUUGC GGUUUUUGCC	-

Table 7. (Continued)

miR-548c-5p	gene	ID	position	sub	sequence (5' ⇒ 3')	
	SOX30	NM_007017	-621	12-U	AAAAGUAAUUGU GGUUUUUGCC	+
	SOX30	NM_178424	-621	12-U	AAAAGUAAUUGU GGUUUUUGCC	+

miR-548d-5p	gene	ID	position	sub	sequence (5' ⇒ 3')	
AAAAGUAAUUG UGGUUUUUGCC	ANXA3	NM_005139	-248	12-C	AAAAGUAAUUGC GGUUUUUGCC	+
	C5orf32	NM_032412	-658	22-A	AAAAGUAAUUGU GGUUUUUGCA	+
	PSTPIP2	NM_024430	-798	12-C	AAAAGUAAUUGC GGUUUUUGCC	+
	ZNF418	NM_133460	-732	12-C	AAAAGUAAUUGC GGUUUUUGCC	+

miR-566	gene	ID	position	sub	sequence (5' ⇒ 3')	
GGGCGCCUGUG AUCCCAAC	C14orf2	NM_004894	-979	11-A	GGGCGCCUGUAAU CCCAAC	+
	C20orf24	NM_018840	-980	18-G	GGGCGCCUGUGAU CCCAGC	-
	C20orf24	NM_199483	-980	18-G	GGGCGCCUGUGAU CCCAGC	-
	CT45-1	NM_001017417	-267	12-G	GGGCGCCUGUGGU CCCAAC	+
	GSTCD	NM_001031720	-444	11-A	GGGCGCCUGUAAU CCCAAC	+
	PRRG2	NM_000951	-788	11-A	GGGCGCCUGUAAU CCCAAC	+
	RASL11A	NM_206827	-104	11-A	GGGCGCCUGUAAU CCCAAC	+
	RLN3	NM_080864	-591	11-A	GGGCGCCUGUAAU CCCAAC	-
	ZNF594	NM_032530	-466	11-A	GGGCGCCUGUAAU CCCAAC	-

miR-574-5p	gene	ID	position	sub	sequence (5' ⇒ 3')	
UGAGUGUGUGU GUGUGAGUGUG U	BAHCC1	NM_001080519	-40	17-U	UGAGUGUGUGUG UGUGUGUGUGU	-
	C4orf31	NM_024574	-206	17-U	UGAGUGUGUGUG UGUGUGUGUGU	-
	CALN1	NM_031468	-899	21-G	UGAGUGUGUGUG UGUGAGUGGGU	-
	CD300A	NM_007261	-496	12-A	UGAGUGUGUGUA UGUGAGUGUGU	-
	CD300A	NM_007261	-556	10-U	UGAGUGUGUUUG UGUGAGUGUGU	-
	EMP2	NM_001424	-645	17-U	UGAGUGUGUGUG UGUGUGUGUGU	-

Table 7. (Continued)

miR-574-5p	gene	ID	position	sub	sequence (5' ⇒ 3')	
	ERC1	NM_015064	-302	3-U	UGUGUGUGUGUG UGUGAGUGUGU	-
	ERC1	NM_178037	-302	3-U	UGUGUGUGUGUG UGUGAGUGUGU	-
	ERC1	NM_178038	-302	3-U	UGUGUGUGUGUG UGUGAGUGUGU	-
	ERC1	NM_178039	-302	3-U	UGUGUGUGUGUG UGUGAGUGUGU	-
	ERC1	NM_178040	-302	3-U	UGUGUGUGUGUG UGUGAGUGUGU	-
	GDF5	NM_000557	-159	17-U	UGAGUGUGUGUG UGUGUGUGUGU	-
	GLT25D2	NM_015101	-23	17-U	UGAGUGUGUGUG UGUGUGUGUGU	+
	MAP3K13	NM_004721	-197	17-U	UGAGUGUGUGUG UGUGUGUGUGU	+
	MUC16	NM_024690	-194	17-U	UGAGUGUGUGUG UGUGUGUGUGU	+
	OPN4	NM_001030015	-154	17-U	UGAGUGUGUGUG UGUGUGUGUGU	+
	OPN4	NM_033282	-154	17-U	UGAGUGUGUGUG UGUGUGUGUGU	+
	OR1N2	NM_001004457	-423	17-U	UGAGUGUGUGUG UGUGUGUGUGU	-
	PI3	NM_002638	-801	3-U	UGUGUGUGUGUG UGUGAGUGUGU	+
	SORCS1	NM_001013031	-756	17-U	UGAGUGUGUGUG UGUGUGUGUGU	+
	SORCS1	NM_052918	-756	17-U	UGAGUGUGUGUG UGUGUGUGUGU	+
	SYT14	NM_153262	-659	17-U	UGAGUGUGUGUG UGUGUGUGUGU	-
	TRPA1	NM_007332	-273	17-U	UGAGUGUGUGUG UGUGUGUGUGU	+
	VCX	NM_013452	-143	17-U	UGAGUGUGUGUG UGUGUGUGUGU	+
	VCX3A	NM_016379	-142	3-U	UGUGUGUGUGUG UGUGAGUGUGU	+

miR-646	gene	ID	position	sub	sequence (5' ⇒ 3')	
AAGCAGCUGCCU CUGAGGC	PRELID2	NM_138492	-419	2-G	AGGCAGCUGCCU CUGAGGC	+
	PRELID2	NM_182960	-451	2-G	AGGCAGCUGCCU CUGAGGC	+
	PRELID2	NM_205846	-451	2-G	AGGCAGCUGCCU CUGAGGC	+

Table 7. (Continued)

miR-649	gene	ID	position	sub	sequence (5' ⇒ 3')	
AAACCUGUGUU GUUCAAGAGUC	ZNF211	NM_006385	-289	19-U	AAACCUGUGUUG UUCAAGUGUC	-
	ZNF211	NM_198855	-289	19-U	AAACCUGUGUUG UUCAAGUGUC	-

miR-940	gene	ID	position	sub	sequence (5' ⇒ 3')	
AAGGCAGGGCCC CCGCUCCCC	GABRA2	NM_000807	-249	15-C	AAGGCAGGGCCC CCCCUCCCC	+

Mature miRNAs and genes with complementary sequences containing 1-bp mismatches in the 1-kb upstream sequences of the TSSs. The nucleotide sequence below each miRNA designation shows the mature miRNA sequence. The gene name indicates the gene located downstream of the miRNA complementary site. ID indicates the accession ID of each transcript. Position indicates the nucleotide position relative to the 5′-end of the miRNA counted from the transcription start site (+1). "Sub" represents the mismatched position and nucleotide in miRNA. The red characters in the miRNA sequences show sequences mismatched. The +/= shows the orientation of the miRNA target sequences; this symbol indicates the relative direction with respect to downstream mRNA transcription (+/= = same/opposite).

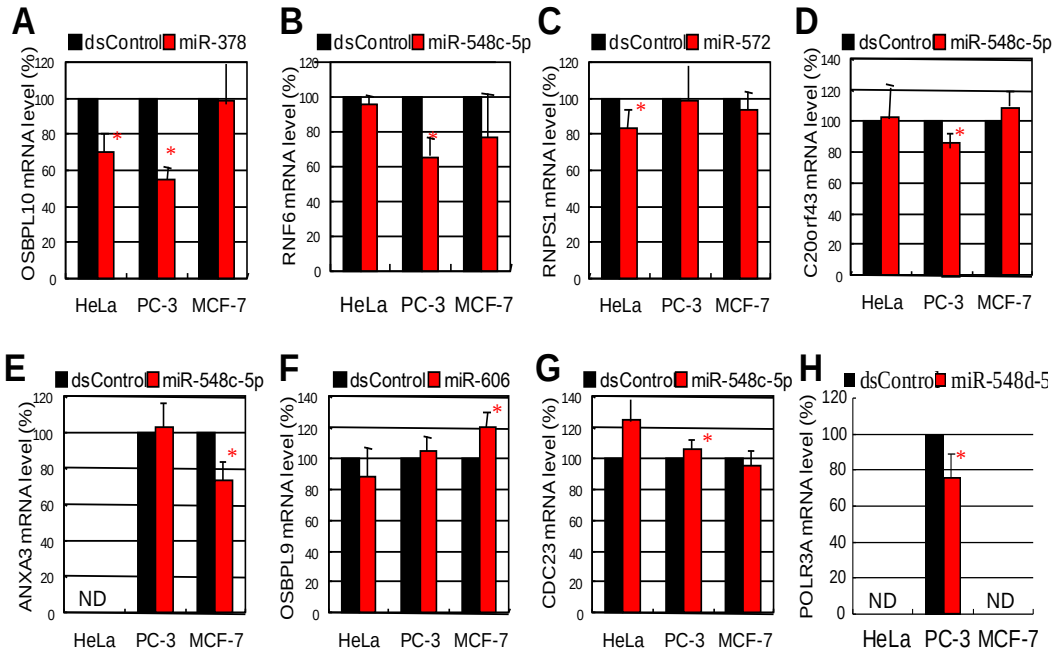

Figure 5. Changes in mRNA expression levels after transfection of miRNAs.
Downstream mRNAs were detected by real-time RT-PCR in HeLa, PC-3, and MCF-7 cells. The combinations of downstream mRNA/miRNA were as follow: (A) OSBPL10/miR-378, (B) RNF6/miR-548c-5p, (C) RNPS1/miR-572, (D) C20orf43/miR-548c-5p, (E) ANXA3/miR-548c-5p, (F) OSBPL9/miR-606, (G) CDC23/miR-548c-5p, and (H) POLR3A/miR-548d-5p. The expression level of each mRNA was normalized to the dsControl, which is presented as 100 %. The values are shown as the means ± SD from at least three independent experiments. Statistical significance was calculated using paired two-tailed Student's t-test; *P < 0.05. ND, not determined.

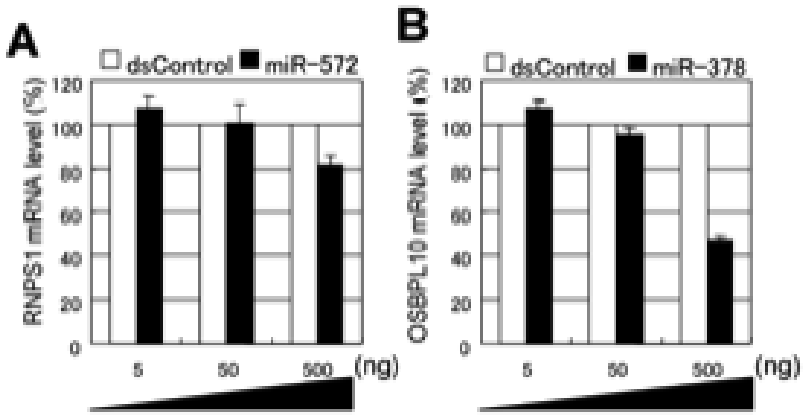

Figure 6. Dose-dependent effect of miRNA on gene expression. (A) Downstream RNPS1 mRNA was detected by real-time RT-PCR in HeLa cells transfected with miR-572 (5, 50 and 500 ng/well). (B) Downstream OSBPL10 mRNA was detected by real-time RT-PCR in PC-3 cells transfected with miR-378 (5, 50 and 500 ng/well). The expression level of each mRNA was normalized to the dsControl, which is presented as 100 %. The values are shown as the means ± SD.

Discussion

This study is the first report of global identification of miRNAs that contain completely or near-completely (i.e., 1- or 2-bp mismatched) complementary sequences to sites in the 10-kb or 1-kb upstream regions of human mRNAs. In total, 66 sites within 10-kb upstream of the coding regions were found to be completely complementary with miRNAs transcribed from the different locations (Tables 5 and 6). Although each of miR-378, miR-572, and miR-606 has only one complementary site, miR-548c-5p, miR-548d-5p and miR-574-5p appeared to be complementary with the upstream sequences of 26, 17, and 4 genes, respectively. The miR-548c-5p and miR-548d-5p miRNAs are members of a family of human miRNA genes derived from short miniature inverted-repeat transposable elements (MITEs) duplicated in many genomic regions, and are capable of forming highly stable hairpin loop recognized by the RNAi enzymatic machinery and processed to form 22 bp mature miRNA sequences [39]. Thus, the many complementary sites of miR-548 family in the upstream regions might represent evolutionary specific regulation of gene expression related to species diversification.

The mismatched positions and types of nucleotides spanning the miRNAs appeared to be highly biased based on searches for sequences containing 1- and 2-bp mismatches with mature miRNAs in the 1-kb upstream region (Figure 1 and 2). The seed sequences of miRNAs are the primary region for 3′ UTR target recognition [31-35]. In RNAi, the identities of off-targets are determined by complementarity to the seed regions of the introduced siRNA [36-38]. In this study, among human miRNAs, 10 and 44 miRNAs were revealed to have 1- and 2-bp mismatched sites, respectively, in the 1-kb upstream region of TSSs (Table 5). The search result of 1-bp mismatched target sites of miRNAs showed that 87 % of 1-bp mismatched sites and 88 % of 2-bp mismatched sites were located in the non-seed regions,

except for the target sites of miR-574-5p. The results suggest that these miRNAs are able to act on their upstream target sites with perfect complementarity in the seed region, if the mechanism of transcriptional regulation on the upstream region is similar to that of the known miRNA function. In RNAi, mismatches at the central region of the siRNA guide strand have been shown to be highly sensitive to knockdown efficiency, but the mismatches in the other site are the most tolerable [29, 30]. In both 1-bp and 2-bp mismatched site analyses, miR-548d-5p and miR-566 showed major mismatched site at positions 12 and 11 (Figs. 1 and 2), respectively, indicating that these miRNAs can act on the upstream regions through the known miRNA machinery but not RNAi machinery. In an analogous fashion, miR-548c-5p also has major mismatched site at position 12 (Figure 1B). In contract, major mismatched site of miR-548c-3p was position 7, and the miR-574-5p has mismatches at positions 3 and 17 (Figure 2). The positions 7 and 3 are involved in the seed region, suggesting that these miRNAs can not act through a known miRNA machinery.

We next evaluated the effects of the miRNAs on gene expression. Eight combinations of miRNAs that matched the upstream sites of mRNAs positioned −24 to −8654-bp relative to the TSS were used. Reductions in mRNA levels were observed in 6 of 8 combinations (Figure 5). In most cases, the reduction was observed in one of three cell lines. In only one combination, reduction was observed in two cell lines simultaneously (Figure 5A). The result indicates that the reducing effect of each miRNA is different depending on cell type. None of the miRNA sequences completely matched the mRNA coding regions. However, two complementary sites of the seed sequence of miR-548d-3p, an opposite strand of miR-548d-5p, were found in the 3′ UTR and CDS of POLR3A gene (Table 8). This suggests that the decrease in POLR3A mRNA by miR-548d-5p is a seed-dependent silencing activity. Furthermore, two seed sequence complementary sites of miR-378 were found in the CDS of OSBPL10 mRNA (Table 8). Recently, the CDS was reported to comprise miRNA and siRNA targets [38,40,41]. Thus, the decrease in OSBPL10 mRNA after transfection of miR-378 could also be a seed-dependent silencing effect that targets the OSBPL10 CDS in HeLa and PC-3 cells, although such reduction was not observed in MCF-7 cells (Figure 5A). The increases in mRNA levels were observed in two of the miRNA/mRNA combinations in only one of three cell lines (Figure 5F, G), indicating that the machinery for increasing mRNA expression is also highly cell type-specific. These results show that miRNA can both silence and activate gene expression; these effects were largely dependent on cell type and downstream gene. In this study, the indirect interaction of miRNAs through the regulation of some transcription factor(s) that control the expression of the downstream genes is hardly distinguishable from the direct interaction of miRNAs with the upstream regions. Further detailed analyses might reveal the mechanism of transcriptional regulation (e.g., RNAa, TGS) targeted the upstream region of mRNAs.

Table 8. Seed complementary sites in the mRNA coding regions

A

gene / miRNA	Region in mRNA		
	5'UTR	CDS	3'UTR
OSBPL10 / miR-378	0	2[1), 2)]	0
RNF6 / miR-548c-5p	0	0	0
CDC23 / miR-548c-5p	0	0	0

A

gene / miRNA	Region in mRNA		
	5'UTR	CDS	3'UTR
OSBPL9 / miR-606	0	0	0
RNPS1 / miR-572	0	0	0
C20orf43 / miR-548c-5p	0	0	0
ANXA3 / miR-548c-5p	0	0	0
P OLR3A / miR-548d-5p	0	0	0

1)

```
mRNA   5’-ACCAUGGAGAGGGCAGUCCAGGGC-3’
              || |   :     |||||||
miRNA  3’-GGAAGACUGAGGUUCAGGUCA-5’
```

2)

```
mRNA  5’-AGCAGGGACCCAUGGAGUCCAGGA-3’
            |       | : :|||||||
miRNA    3’-GGAAGACUGAGGUUCAGGUCA-5’
```

B

gene / miRNA	Region in mRNA		
	5'UTR	CDS	3'UTR
OSBPL10 / miR-378*	0	0	0
RNF6 / miR-548c-3p	0	0	0
CDC23 / miR-548c-3p	0	0	0
OSBPL9 / miR-606 opposite strand	0	0	0
RNPS1 / miR-572 opposite strand	0	0	0
C20orf43 / miR-548c-3p	0	0	0
ANXA3 / miR-548c-3p	0	0	0
POLR3A / miR-548d-3p	0	1[1)]	1[2)]

1)

```
mRNA                  5’-AUACGCAGAUGAAAAGGUUUUUGA-3’
                            : ||| :    ||||||||
miRNA                  3’-CGUUUUCUUUGACACCAAAAAC-5’
```

2)

```
mRNA   5’-CAAUGGGGAUGAGAUGGUUUUUUG-3’
             : : :|| :   ||||||||
miRNA   3’-CGUUUUCUUUGACACCAAAAAC-5’
```

The number of seed complementary sites of the mature miRNA guide strand (A) and opposite strand (B) in the 5′ UTR, CDS and 3′ UTR of each mRNA which expression level was changed by the corresponding miRNA. Two seed complementary sites were observed in CDS of OSBPL10 (A). The opposite strand of miR-548c-5p, miR-548c-3p, showed seed complementarity in CDS and 3′ UTR of POLR3A (B). The pairing pattern of miRNA and mRNA in the seed complementary regions are shown below each table. Seven nucleotides at nucleotide positions 2-8 from 5′ end of each miRNA were perfectly matched.

Conclusion

In the present study, we identified miRNAs and their complementary sites in upstream sequences of mRNA coding regions. Based on our results, perfectly matched complementary sites and 1- and 2-bp mismatched sites appear to be potential targets for miRNA. Using 8 combinations of miRNAs and downstream mRNAs, we found that miRNAs are able to both silence and promote the expression of downstream genes and that these changes are largely gene- and cell type-dependent.

Acknowledgments

This work was partially supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan and Core Research Project for Private University: matching fund subsidy to K.U.-T.

References

[1] Filipowicz, W., Jaskiewicz, L., Kolb, F. A. & Pillai, R. S. (2005) Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr. Opin. Struct. Biol.,* 15, 331-341.

[2] Zaratiegui, M., Irvine, D.V. & Martienssen, R. A. (2007) Noncoding RNAs and gene silencing. *Cell,* 128, 763-776.

[3] Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, J. (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science,* 305, 1434-1437.

[4] Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J.J., Hammond, S. M., Joshua-Tor, J. & Hannon, G. F. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science,* 205, 1437-1441.

[5] Borchert, G. M., Lanier, W. & Davidson, B. L. (2006) RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.,* 13, 1097-1101.

[6] Lee, R. C., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S. & Kim, V. N. (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature,* 425, 415-419.

[7] Han, J., Lee, Y., Yeom, K. H., Kim, Y. K., Jin, H. & Kim, V. N. (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.,* 18, 3016-3027.

[8] Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F. & Hannon, G. J. (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature,* 432, 231-235.

[9] Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.,* 17, 3011-3016.

[10] Morris, K. V., Chan, A. W. L., Jacobsen, S. E. & Looney, D. F. (2004) Small interfering RNA-induced transcriptional gene silencing in human cells. *Science,* 305, 1289-1292.

[11] Ting, A. H., Schuebel, K. E., Herman, J. G. & Baylin, S. B. (2005) Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation. *Nat. Genet.*, 37, 906-910.

[12] Janowski, B. A., Huffman, K. E., Schwartz, J. C., Ram, R., Nordsell, R., Shames, D. S., Minna, J. D. & Corey, D. R. (2006) Involvement of Ago1 and Ago2 in mammalian transcriptional silencing. *Nat. Struct. Mol. Biol.*, 13, 787-792.

[13] Kim, D. H., Willeneuve, L. M., Morris, K. V. & Rossi, J. J. (2006) Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat. Struct. Mol. Biol.*, 13, 793-797.

[14] Weinberg, M. S., Villneuve, L. M., Ehsani, A., Amarzguioui, M., Aagaard, L., Chen, Z. X., Riggs, A. D., Rossi, J. J. & Morris, K. V. (2006) The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. *RNA*, 12, 256-262.

[15] Han, J., Kim, D. & Morris, K. V. (2007) Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc. Natl. Acad. Sci.*, USA, 104, 12422-12427.

[16] Li, L. C., Okino, S. T., Zhao, H., Pookot, D., Place, R. F., Urakami, S., Enokida, H. & Dahiya, R. (2006) Small dsRNAs induce transcriptional activation in human cells. *Proc. Natl. Acad. Sci., USA*, 103, 17337-17342.

[17] Janowski, B. A., Younger, S. T., Hardy, D. B., Ram, R., Huffman, K. D. & Corey, D. R. (2007) Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat. Chem. Biol.*, 3, 166-173.

[18] Schwartz, J. C., Younger, S. T., Nguyen, N. B., Hardy, D. B., Monia, B. P., Corey, D. R. & Janowski, B. A. (2008) Antisense transcripts are targets for activating small RNAs. *Nat. Struct. Mol. Biol.*, 15, 842-848.

[19] Morris, K. V., Santoso, S., Turner, A. M., Pastori, C. & Hawkins, P. G. (2008) Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet.*, 4, 1-9.

[20] Weinberg, M. S., Barichievy, S., Schaffer, L., Han, J. & Morris, K. V. (2007) An RNA targeted to the HIV-1 LTR promoter modulates indiscriminate off-target gene activation. *Nucleic Acids Res.*, 35, 7303-7312.

[21] Kim, D. H., Sætrom, P., Snøve, O. Jr. & Rossi, J. J. (2008) MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc. Natl. Acad. Sci., USA*, 105, 16230-16235.

[22] Place, R. F., Li, L. C., Rookot, D., Noonan, E. J. & Dahiya, R. (2008) MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc. Natl. Acad. Sci., USA*, 105, 1608-1613.

[23] Rodriguez, A., Griffiths-Jones, S., Ashurst, J. L. & Bradley, A. (2004) Identification of mammalian microRNA host genes and transcription units. *Genome Res.*, 14, 1902-1910.

[24] Bartel, D. P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, 136, 215-233.

[25] Sayers, E. W., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L. Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D. J., Madden, T. L., Maglott, D. R., Miller, V., Mizrachi, I., Ostell, J., Pruitt, K. D., Schuler, G. D., Sequeira, E., Sherry,

S. T., Shumway, M., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusova, T. A., Wagner, L. & Yaschenko, E., Ye, J. (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.,* 37, D5-15.

[26] Kuhn, R. M., Karolchik, D., Zweig, A. S., Trumbower, H., Thomas, D. J., Thakkapallayil, A., Sugnet, C. W., Stanke, M., Smith, K. E., Siepel, A., Rosenbloom, K. R., Rhead, B., Raney, B. J., Pohl, A., Pedersen, J. S., Hsu, F., Hinrichs, A. S., Harte, R. A., Diekhans, M., Clawson, H., Bejerano, G., Barber, G. P., Baertsch, R., Haussler, D. & Kent, W. J. (2007) The UCSC genome browser database: update 2007. *Nucleic Acids Res.,* 35, D668-73.

[27] Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. (2008) miRBase tools for microRNA genomics. *Nucleic Acids Res.,* 36, D154-158.

[28] Chen, C., Ridzon, D. A., Broomer, A. J., Zhou, Z., Lee, D. H., Nguyen, J. T., Barbisin, M., Xu, N. L., Mahuvakar, V. R., Andersen, M. R., Lao, K. Q., Livak, K. L. & Guegler, K. J. (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.,* 33, e179.

[29] Du, Q., Thonberg, H., Zhang, H.-Y., Wahlestedt, C. & Liang, Z. (2005) A systematic analysis of the silencing effects of an active siRNA at all single-nucleotide mismatched target sites. *Nucleic Acids Res.,* 33, 1671-1677.

[30] Dahlgren, C., Zhang, H.-Y., Du, Q., Grahn, M., Norstedt, G., Wahlestedt, C. & Liang, Z. (2008) Analysis of siRNA specificity on targets with double-nucleotide mismatches. *Nucleic Acids Res.,* 36, e53.

[31] Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. (2003) Prediction of mammalian microRNA targets. *Cell,* 115, 787-798.

[32] Bartel, D. P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell, 116, 281-297.

[33] Lewis, B. P., Burge, C. B. & Bartel, D. P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousand of human genes are microRNA targets. *Cell,* 120, 15-20.

[34] Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., Bartel, D. P., Linsley, P. S. & Johnson, J. M. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature,* 433, 769-773.

[35] Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P. & Bartel, D. P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell,* 27, 91-105.

[36] Jackson, A. L., Burchard, J., Schelter, J., Chau, B. N., Cleary, M., Lim, L. & Linskey, P. S. (2006) Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA,* 12, 1179-1187.

[37] Birmingham, A., Anderson, E. M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., Baskerville, S., Maksimova, E., Robinson, K., Karpilow, J., Marshall, W. S. & Khvorova, A. (2006) 3′UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods,* 3, 199-204.

[38] Ui-Tei, K., Naito, Y., Nishi, K., Juni, A. & Saigo, K. (2008) Thermodynamic stability and Watson-Crick base pairing in the seed duplex are major determinants of the efficiency of the siRNA-based off-target effect. *Nucleic Acids Res.,* 36, 7100-7109.

[39] Piriyapongsa, J. & Jordan, K. (2007) Family of human microRNA genes from miniature inverted-repeat transposable elements. *PLoS ONE,* 2, e203.

[40] Duursma, A. M., Kedde, M., Schrier, M., Sage, C. & Agami, R. (2008) miR-148 targets human DNMT3b protein coding region. *RNA,* 14, 872-877.

[41] Tay, Y., Zhang, J., Thomson, A. M., Lim, B. & Rigoutsos, I. (2008) MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature,* 455, 1124-1128.