

# Genome-Wide Identification of SNPs in MicroRNA Genes and the SNP Effects on MicroRNA Target Binding and Biogenesis

Jing Gong,<sup>1,2</sup> Yin Tong,<sup>1</sup> Hong-Mei Zhang,<sup>1</sup> Kai Wang,<sup>3</sup> Tao Hu,<sup>1</sup> Ge Shan,<sup>4</sup> Jun Sun,<sup>2\*</sup> and An-Yuan Guo<sup>1\*</sup>

<sup>1</sup>Hubei Bioinformatics & Molecular Imaging Key Laboratory, Department of Systems Biology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, China; <sup>2</sup>Department of Biochemistry and Molecular Biology, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China; <sup>3</sup>Shanghai Bioladder Co., Ltd, Shanghai, China; <sup>4</sup>School of Life Science, University of Science and Technology of China, Hefei, China

Communicated by Pui-Yan Kwok

Received 3 June 2011; accepted revised manuscript 15 October 2011.

Published online 1 November 2011 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.21641

**ABSTRACT:** MicroRNAs (miRNAs) are studied as key regulators of gene expression involved in different diseases. Several single nucleotide polymorphisms (SNPs) in miRNA genes or target sites (miRNA-related SNPs) have been proved to be associated with human diseases by affecting the miRNA-mediated regulatory function. To systematically analyze miRNA-related SNPs and their effects, we performed a genome-wide scan for SNPs in human pre-miRNAs, miRNA flanking regions, target sites, and designed a pipeline to predict the effects of them on miRNA–target interaction. As a result, we identified 48 SNPs in human miRNA seed regions and thousands of SNPs in 3′ untranslated regions with the potential to either disturb or create miRNA–target interactions. Furthermore, we experimentally confirmed seven loss-of-function SNPs and one gain-of-function SNP by luciferase assay. This is the first case of experimental validation of an SNP in an miRNA creating a novel miRNA target binding. All useful data were compiled into miRNASNP, a user-friendly free online database (<http://www.bioguo.org/miRNASNP/>). These data will be a useful resource for studying miRNA function, identifying disease-associated miRNAs, and further personalized medicine.

Hum Mutat 33:254–263, 2012. © 2011 Wiley Periodicals, Inc.

**KEY WORDS:** miRNASNP; database; target loss and gain; frequency

## Introduction

MicroRNAs (miRNAs) are a family of endogenous small non-coding RNAs involved in various developmental and physiological processes by negatively regulating gene expression [Bartel 2004; Zhang et al., 2007]. To date, thousands of miRNAs in human and other species have been identified in miRBase database [Kozomara and Griffiths-Jones, 2011]. miRNA genes may reside in introns of protein coding genes or intergenic regions. They are initially transcribed in the nucleus as long primary transcripts (pri-miRNAs) and further processed by the RNase Drosha to hairpin precursor miRNAs (pre-miRNAs) [Lee et al., 2003]. Then, the pre-miRNA hairpins are exported to the cytoplasm and processed into ~22 nt miRNA duplex by RNase Dicer [Bartel, 2004]. One strand from the miRNA duplex (miR-5p/miR-3p duplex) containing the less stable 5′ end is preferentially selected and loaded onto the RNA-induced silencing complex (RISC) to produce a functional, mature miRNA (MIR) [Khvorova et al., 2003]. MIRs recognize their target mRNAs mainly by base-pairing interaction between nucleotides 2 and 8 (seed region) from its 5′ end and the complementary nucleotides on the 3′ untranslated region (3′UTR) of target mRNAs [Lai, 2002; Lewis et al., 2003]. Currently, it is estimated that an miRNA may regulate hundreds of target genes and most of human protein coding genes are regulated by miRNAs [Betel et al., 2008; Friedman et al., 2009; Krek et al., 2005].

Single nucleotide polymorphisms (SNPs) are important variations for the diversity among individuals, as well as leading to phenotypes, traits, and diseases [Shastry, 2009]. Since miRNAs are wide and key regulators of gene expression, miRNA-related SNPs including SNPs in miRNA genes and target sites may function as regulatory SNPs through modifying miRNA regulation to affect the phenotypes and disease susceptibility [Ryan et al., 2010]. Moreover, SNPs located in MIRs are likely to cause complex influence by affecting MIR maturation, functional strand selection, and target selection. To date, a number of studies have demonstrated that SNPs in target sites or miRNA genes are associated with diseases [Jazdzewski et al., 2008; Mencia et al., 2009; Ryan et al., 2010; Saunders et al., 2007; Sethupathy and Collins, 2008; Sun et al., 2009]. For example, an SNP

Additional Supporting Information may be found in the online version of this article.

\*Correspondence to: An-Yuan Guo, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, 430074, China. E-mail: guoay@mail.hust.edu.cn or Jun Sun, Department of Biochemistry and Molecular Biology, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430030, China. E-mail: sunjun99@hotmail.com

Contract grant sponsors: Huazhong University of Science and Technology (to A.Y.G.); Fundamental Research Funds for the Central Universities (2010MS045 to A.Y.G.); National Natural Science Foundation of China (31171271 to A.Y.G.); National Science and Technology Innovation Project for Undergraduates from Huazhong University of Science and Technology (2011156 to Y.T.); State Key Development Program for Basic Research of China (973, 2011CBA01103 to G.S.); National Natural Science Foundation of China (31071132 to G.S.).

in the 3'UTR of *KRAS* gene located in the binding site of miRNA let-7 weakens its inhibition and results to increase the risk of nonsmall cell lung cancer [Chin et al., 2008]. An SNP in pre-miR-146a is reported to decrease mature miRNA expression and predisposes to papillary thyroid carcinoma [Jazdzewski et al., 2008]. Sun et al. validated several SNPs in human miRNA genes that could affect the biogenesis and function of miRNAs [Sun et al., 2009]. A single mutation in pre-miR-155 creating a mismatch near the 3' end of miR-155 leads to a shift in strand selection, thereby fine-tunes their targets and results in a butterfly effect on global gene expression [Lee et al., 2011]. Another study reported that a mutation in the seed region of human miR-96 was responsible for nonsyndromic progressive hearing loss [Mencia et al., 2009].

Since 2005, several studies have systematically identified and analyzed the human polymorphisms in miRNAs and/or miRNA target sites [Bhartiya et al., 2011; Duan et al., 2009; Iwai and Naraba, 2005; Landi et al., 2008; Ryan et al., 2010; Saunders et al., 2007]. Some features of these SNPs such as SNP density, allele frequency, evolutionary conservation and effects on miRNA guide strand selection have been analyzed based on data at that time. For the convenience of biologists, several online databases and tools about SNPs in miRNA target sites or miRNAs were developed by different groups, such as dbSMR [Hariharan et al., 2009], Patrocles [Hiard et al., 2010], PolymiRTS [Bao et al., 2007], microSNiPer [Barenboim et al., 2010], and miRvar [Bhartiya et al., 2011]. However, most of them focus on SNPs in target sites and their effects; few of them mentioned the SNPs in miRNA genes and their influences on target selection and miRNA biogenesis. Moreover, the numbers of miRNAs and SNPs were increased greatly in recent 2 years. For example, 1,048 versus 706 human miRNAs are in the miRBase release 16 (September 2010) [Kozomara and Griffiths-Jones, 2011] versus release 13 (March 2009), respectively. There are more than 30 million human SNPs in current dbSNP 132 (September 2010) and only 14 million human SNPs in dbSNP 129 (April 2008). Therefore, despite of these previous efforts, a more comprehensive database for SNPs in miRNA genes based on the greatly increased data is necessary and useful. To this end, here we systematically characterized all miRNA-related SNPs, summarized their features, and analyzed their effects on target binding alteration and mature miRNA biogenesis by both prediction and experiments. All useful data about miRNA-related SNPs and target alteration information were compiled into a user-friendly database, miRNASNP, freely available at <http://www.bioguo.org/miRNASNP/>.

## Materials and Methods

### Identification of miRNA-Related SNPs and Their Context Information

miRNA data (including chromosomal location, host gene, conservation among species, and miRNA cluster etc.) of the nine studied species (human, chimpanzee, mouse, rat, dog, horse, cow, chicken, and zebrafish) were obtained from miRBase database (release 16.0). The SNP information was downloaded from the latest version of NCBI dbSNP (release 132 for human). Then, we compared the chromosomal locations of pre-miRNAs and SNPs to identify SNPs in pre-miRNAs and their adjacent upstream and downstream 1000-bp regions. SNPs in pre-miRNAs were further classified into mature miRNAs and miRNA seed regions according to their locations in miRNA. We calculated the SNP densities of each miRNA and flanking regions and then used the *t*-test statistical method to test the difference of SNP densities between miRNA regions and flanking

regions. The SNP density of each mature miRNA site was defined as the number of SNPs at the site per 1,000 miRNAs.

Data for SNP population allele frequency and iHSs were derived from HapMap (<http://www.hapmap.org/>) and Haplotter (<http://haplotter.uchicago.edu/>) [Voight et al., 2006]. Experimentally validated miRNAs targets were obtained from miR2Disease [Jiang et al., 2009] and TarBase databases [Sethupathy et al., 2006]. Together, 1,440 miRNA-target pairs with experimental evidences were included in our analysis. Minimum free energies of pre-miRNA hairpin structures were generated by RNAfold [Denman, 1993] and the pictures of pre-miRNA second structures were generated by RNAplot. All data processing scripts were written in Perl.

### miRNA Target Gain and Loss Analysis by Prediction

To predict miRNA target sites, we combined results of two popular tools, TargetScan (<http://www.TargetScan.org/>) [Friedman et al., 2009; Lewis et al., 2005] and miRanda (<http://www.microna.org/>) [Betel et al., 2008], which are regularly updated and considered with relatively good performances. In detail, we used the miRanda v3.3a with the default parameters and cutoffs (Score  $S \geq 140$  and Energy  $E \leq -7.0$ ) to predict miRNA target. For TargetScan, we used the default parameters and defined the target site conserved if it existed in one of the corresponding orthologs of mouse, rat, and dog. Sequences of 3'UTRs of human, chimp, rat, mouse, and dog were obtained from UCSC genome browser (<http://genome.ucsc.edu/>).

For the SNPs in miRNA seed regions, two different methods were used to predict the target sites for the wild-type miRNAs and SNP-miRNAs. These resulted in four groups of target gene data, which are recorded as WT (target genes of wild-type miRNAs processed by TargetScan), WM (target genes of wild-type miRNAs processed by miRanda), ST (target genes of SNP-miRNAs processed by TargetScan), and SM (target genes of SNP-miRNAs processed by miRanda). If one miRNA/target pair exists in both WT and WM, but not in either ST or SM, we called this miRNA/target pair loss. On the contrary, if one miRNA/target pair was predicted in both ST and SM, but neither in WT nor WM, we defined the SNP-miRNA gained the target gene. In addition, for each miRNA/target loss or gain pair, we obtained the sequence ( $\pm 50$  bp) of target site and used RNAhybrid [Kruger and Rehmsmeier, 2006] to calculate the minimum hybridization energy of the miRNA-target interaction. Generally, more energy change would affect the miRNA-target interaction more sharply. The binding energy changes between wild-type miRNA/target and SNP-miRNA/target were provided in our database as additional information for users making further judgments.

$$\text{Loss} = (\text{WT} \cap \text{WM}) - (\text{ST} \cup \text{SM})$$

$$\text{Gain} = (\text{ST} \cap \text{SM}) - (\text{WT} \cup \text{WM})$$

For the SNPs in mRNA 3'UTRs, the gain and loss of miRNA-target binding and energy change were analyzed by the same strategy. The difference is that the miRNA sequence did not change, while the UTR sequences were divided into wild type and corresponding mutant type.

DAVID bioinformatics online tools (<http://david.abcc.ncifcrf.gov/>) [Huang da et al., 2009] were used to identify enriched functional annotation categories for genes with conserved sites lost and gained by SNPs in miRNA seed regions. Gene Ontology (GO) terms [Barrell et al., 2009] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway [Kanehisa et al., 2004] were evaluated. Benjamini corrected  $P < 0.05$  was considered statistically significant.

**Table 1. A Summary of single nucleotide polymorphisms (SNPs) in pre-miRNAs in Nine Species**

Organism	Total pre-miRNAs	Total SNPs in dbSNP	No. of poly-miRs <sup>a</sup>	No. of SNPs in poly-miRs
Human	1048	30,442,771	440 (149, 41) <sup>b</sup>	757 (178, 50) <sup>b</sup>
Mouse	667	14,180,819	105	146
Chicken	496	3,138,475	66	86
Cow	672	2,102,505	18	16
Chimpanzee	599	1,617,731	13	15
Rat	403	2,465,386	9	9
Dog	323	3,145,822	4	4
Zebrafish	337	658,287	1	1
Horse	344	1,123,624	1	1

<sup>a</sup>poly-miRs: miRNAs with SNP in it.

<sup>b</sup>A total of 178 SNPs in 149 poly-miRs mature sequence, 50 SNPs in 41 poly-miRs seed sequence.

### miRNA Target Gain and Loss Analysis by Experiments

To test the target binding influence of SNPs in miRNA genes, nine miRNAs and three genes were selected to perform luciferase reporter assays. Among them, miR-627 and miR-379 were tested twice with different targets or different SNPs.

Dual-luciferase reporter system was constructed with 3'UTRs amplified from genomic DNA and cloned in psiCHECK<sup>TM</sup>-2 vector (Promega, Madison, WI). The primer sequences for *ATP6V0E1* were 5'-GCCTCGAGGGTCACGAGAAGAGAATGCC-3' and 5'-CAGCGGCCGAGATTTCCATAGAAAGGAGG-3' and the product length was 516 bp; the primer sequences for *BCL2* were 5'-ATTCTAGGCGATCGCTCGAGGTCAACATGCCTGCCCAACA-3' and 5'-TTATTGCGGCCAGCGGCCGCTCCATCCGTCTGCTCTTCAGAT-3' and the product length was 457 bp; the primer sequences for *SEMA3F* were 5'-CTCGAGCTCGAGATTGGTGGGTTGAA-3' and 5'-GCGGCCGCCTGTCCCACCCAGGCAGGAT-3' and the product length was 489 bp. The mimics of wild-type miRNAs and SNP-miRNAs were chemically synthesized by Shanghai Bioladder Company (Shanghai, China).

Then, for each wild-miRNA/target pair or variant-miRNA/target pair, the reporter vector with target gene 3'UTR and miRNA mimic were cotransfected into HEK-239 cells, which are commonly used in miRNA luciferase experiments. Luciferase assays

were carried out by using Luciferase Reporter Gene Assay kit (Promega Company) 48h after transfection. Renilla luciferase (target) and Firefly luciferase (internal control) activities were detected. Renilla luciferase/Firefly luciferase were calculated to compare the difference among them. All experiments were performed in triplicate.

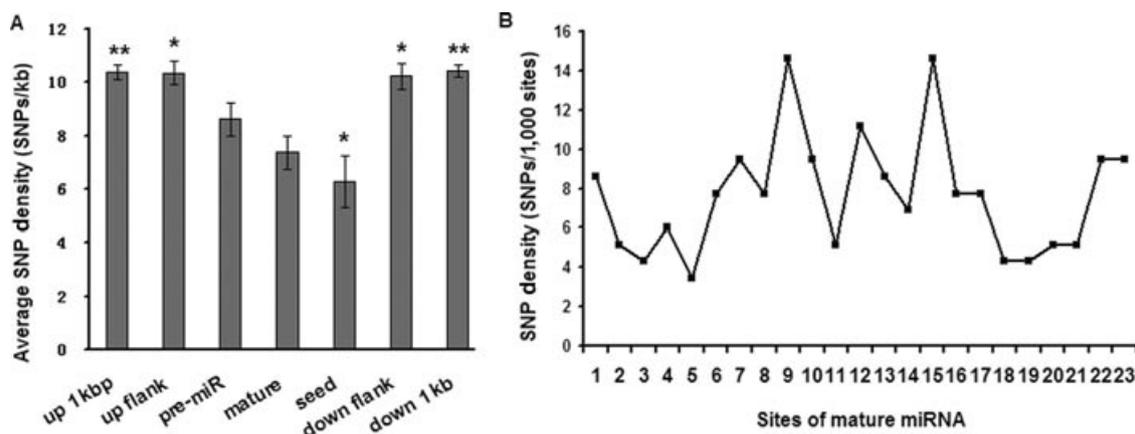
### miRNASNP Database Construction

Most of results were processed by in-house Perl scripts and then all useful information and results in this work were organized into a set of relational MySQL tables. PHP and JavaScript were used to construct the online website miRNASNP with multiple browse and search function.

## Results

### Summary and Features of miRNAs with SNPs

There are 1,048 human pre-miRNA entries in miRBase (release 16, Sept. 2010) and more than 30 million human SNPs in dbSNP (Build 132, September 2010). After mapping the SNPs onto human pre-miRNA genes, we identified 757 SNPs (including indel polymorphisms) in 440 pre-miRNAs. Among them, 178 SNPs are in 149 MIRs and 50 SNPs in 41 MIR seed regions. There are 177 miRNAs with more than one SNP, in which miR-3939, a human-specific miRNA, has the most SNPs (13) in its pre-miRNA region. Besides human, we also identified the SNPs in pre-miRNAs of other eight species, which are chimpanzee, mouse, rat, dog, cow, horse, chicken, and zebrafish (Table 1). In the following of this study, we only analyzed the human miRNAs if no specific mention because much fewer data are available in other species. To compare the SNP density among different region, we further characterized SNPs in the flanking regions of human pre-miRNAs (-1 kb ~ +1 kb). As a result, we observed that the SNP density of pre-miRNAs and flanking regions is significantly lower than that in the pre-miRNAs and flanking regions (Fig. 1A). The same trends were found in mouse and chicken that have relatively more SNPs than other six species



**Figure 1.** Single nucleotide polymorphism (SNP) density of human pre-miRNAs and flanking regions. (A) SNP density of human pre-miRNAs and flanking regions. (B) SNP density of each site in human mature miRNA regions. Up or down flank represents a window (equal to the average size of pre-miRNAs) located immediately adjacent to the pre-miRNA. *t*-Tests were used to analyze the difference of SNP density between the pre-miRNA and other region. \*  $P < 0.05$ , \*\*  $P < 0.01$ . Error bars indicate the standard error. The figure showed that the SNP density of flanks > precursor > mature > seed region.

**Table 2. The Features of miRNA Distribution in Different Categories**

miRNA	In host		Conserved in primates		Conserved in mammals		In 5-kb cluster		In 10-kb cluster		Total
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	
Total miRNAs	485	563	679	369	411	637	261	787	280	768	1048
miRNAs with SNP	208	232	258*	182	118*	322	83*	357	87*	353	440
miRNAs without SNP	277	331	421	187	293	315	178	430	193	415	608

\* $P < 0.01$  by  $\chi^2$  test comparing miRNAs with SNP or without SNP in different categories.

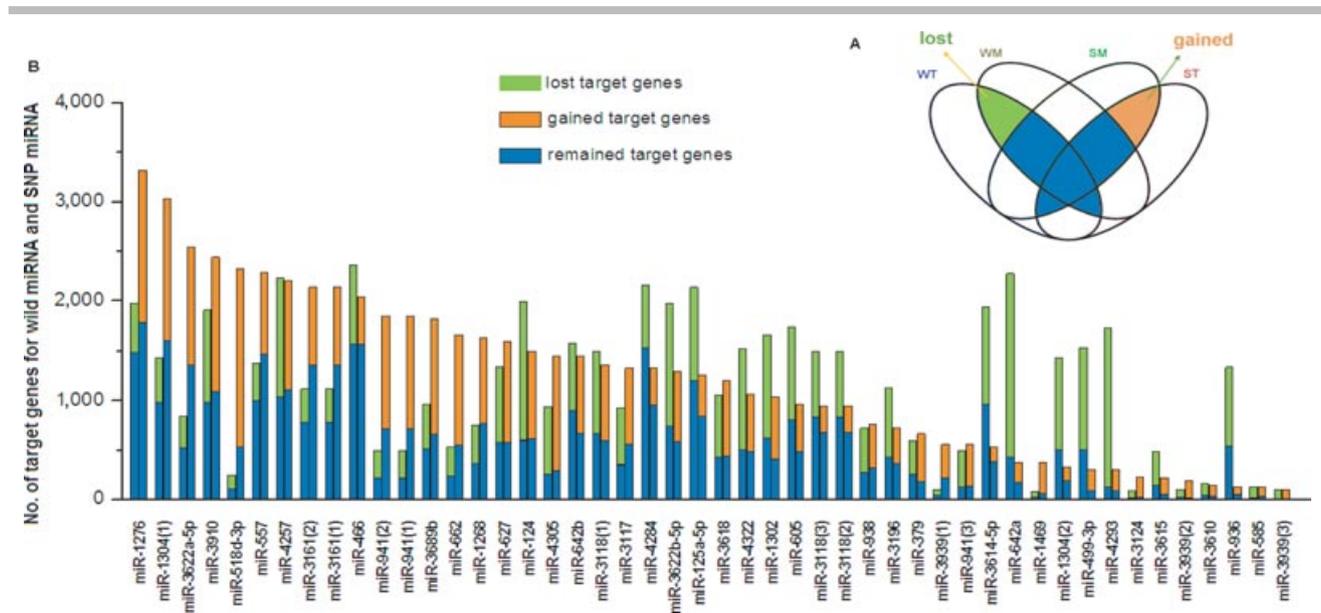
(Supp. Fig. S1). The SNP density of middle sites in human MIRs is higher than their 5' and 3' sites (Fig. 1B). According to the miRNA position on the respective genome and the conservation among species, we classified human miRNAs into different categories with SNPs or without SNPs. Then, we performed statistical test on miRNAs with and without SNPs in each miRNA category (Table 2). The results show that conserved miRNAs and miRNAs in clusters tend to have fewer SNPs ( $\chi^2$  test,  $P < 0.01$ ). miRNAs in host genes or intergenic regions have no significant difference on SNP distribution.

### miRNA Target Alteration by SNPs in miRNA Seed Regions

As miRNA target selection mainly relies on the seed region base pairing, alternative allele of an SNP in the miRNA seed region may greatly change its target spectrum, resulting in lots of targets loss and gain, therefore affect the miRNA function dramatically. As mentioned above, we have identified 50 SNPs in the seed regions of 41 human miRNAs. Because two insertion SNPs do not change the seed sequences, we just analyzed the target alteration for the other 48 SNPs (Supp. Table S1). We predicted the targets by a com-

bined strategy using both TargetScan and miRanda tools for miRNAs with reference sequences (wild-type miRNAs) and miRNAs with SNP alleles (SNP-miRNAs). By comparing the targets of wild-type miRNA with that of SNP-miRNA, we obtained the potential miRNA targets loss and gain (Fig. 2). According to our strategy, on one hand, there are 55,887 miRNA-target pairs predicted by both tools for wild-type miRNAs and 53% of them will be disturbed after changing the reference allele to another SNP allele. On the other hand, the SNP-miRNAs have 58,683 miRNA-target pairs predicted by both tools and 52% of them are novel created by the SNP allele (Fig. 2).

As shown in Figure 2, the total number of putative targets of miR-518d-3p, miR-3622a-5p, and miR-1304(1) (rs76857625) greatly increased after variation, increasing by a number of 2,075, 1,703, and 1,593, respectively. While the number of putative target genes of miR-642a, miR-4293, and miR-3614-5p decreased dramatically after SNP allele alteration, decreasing by 1,895, 1,429, and 1,406, respectively. Especially, miR-4293 lost nearly 93% of its target genes after mutation. Interestingly, miR-1304 has two SNPs in its seed region, but the alterations of target spectrum for the two SNPs were different. The number of putative targets of miR-1304 with SNP rs76857625 was greatly increased after variation, while the number



**Figure 2.** Target loss and gain analysis of miRNAs by SNPs in their seed regions. (A) The definition of loss and gain of target genes. Two tools, TargetScan and miRanda, were used to predict target genes for wild-type miRNA and SNP-miRNA. These resulted in four groups of data, which are recorded as WT (target genes of wild-type miRNAs processed by TargetScan), WM (target genes wild-type miRNAs processed by miRanda), ST (SNP-miRNA targets processed by TargetScan), and SM (SNP-miRNA targets processed by miRanda). (B) The number of target genes for wild-type miRNAs and SNP-miRNAs. The left and right columns of each pair represent the predicted targets for wild-type miRNA (WT $\cap$ WM) and SNP-miRNA (ST $\cap$ SM), respectively. The lost target genes mean (WT $\cap$ WM)–(ST $\cup$ SM), while the gained target genes mean (ST $\cap$ SM)–(WT $\cup$ WM) in (A, B). The SNPs in miRNA seed regions are shown in Supp. Table S1.

**Table 3. Experimentally Validated Human miRNA-Target Pairs and the SNPs in miRNA Genes**

miRNA	Wild mature sequence <sup>a</sup>	SNP mature sequence	SNP	In seed	Target gene	Predicted/ experimental effect
miR-379 (S1)	UGGUAG[A]CUAUGGAACGUAGG	UGGUAG[G]CUAUGGAACGUAGG	rs61991156	Yes	<i>SEMA3F</i>	Loss/loss
miR-379 (S2)	UGGUAGACUAUGGAAC[G]UAGG	UGGUAGACUAUGGAAC[A]UAGG	rs72631818	No	<i>SEMA3F</i>	Loss/loss
miR-1304	UU[U]GAGGCUACAGUGAGAUGUG	UU[C]GAGGCUACAGUGAGAUGUG	rs79759099	Yes	<i>SEMA3F</i>	Loss/no
miR-627	G[U]GAGUCUCUAAGAAAAGAGGA	G[G]GAGUCUCUAAGAAAAGAGGA	rs2620381	Yes	<i>SEMA3F</i>	Loss/loss
miR-940	AAGGCAGGGCCCC[C]GCUCCCC	AAGGCAGGGCCCC-GCUCCCC	rs35356504	No	<i>SEMA3F</i>	Loss/loss
miR-34a	UGGCAGUGUCUUA[G]CUGGUUGU	UGGCAGUGUCUUA[A]CUGGUUGU	rs35301225	No	<i>BCL2</i>	Loss/loss
miR-499-3p	AAC[A]UCACAGCAAGUCUGUGCU	AAC[G]UCACAGCAAGUCUGUGCU	rs3746444	Yes	<i>BCL2</i>	Loss/loss
miR-124	UAA[G]GCACGCGUGAAUGCC	UAA[U]GCACGCGUGAAUGCC	rs34059726	Yes	<i>ATP6V0E1</i>	Loss/loss
miR-642a	GUCC[C]UCCUCAAUGUGUCUUG	GUCC[G]UCCUCAAUGUGUCUUG	rs78902025	Yes	<i>ATP6V0E1</i>	Loss/no
miR-662	UCCCA[G]UUGUGCCCAGCAG	UCCCA[A]UUGUGCCCAGCAG	rs9745376	Yes	<i>ATP6V0E1</i>	Loss/no
miR-627	G[U]GAGUCUCUAAGAAAAGAGGA	G[G]GAGUCUCUAAGAAAAGAGGA	rs2620381	Yes	<i>ATP6V0E1</i>	Gain/gain

<sup>a</sup>The rectangle marks the SNP position.

of putative targets of miR-1304 with SNP rs79759099 was sharply decreased after variation.

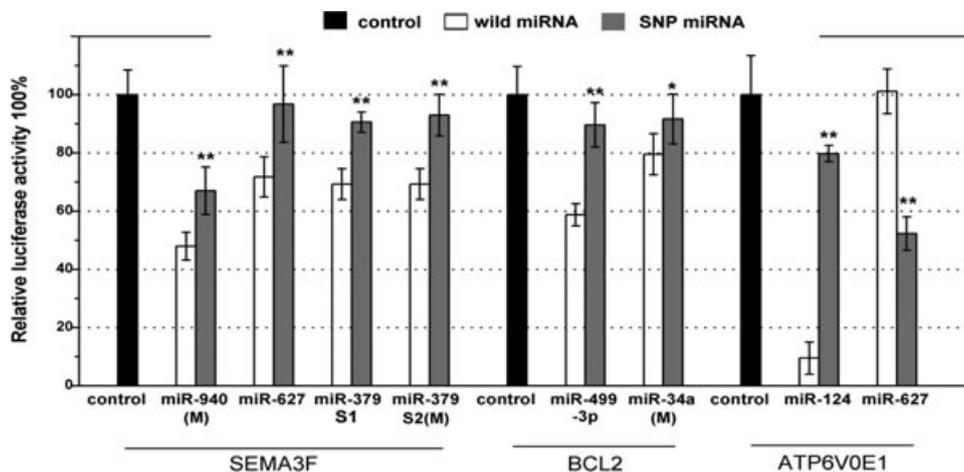
**Target Binding Alteration by Experimental Validation**

Although SNPs in miRNAs were predicted to alter miRNA target binding, to validate the effects in experiments, we selected 11 miRNA-target pairs and their corresponding SNP-miRNA target pairs to test by luciferase reporter assay. Of these 11 selected candidate SNPs, eight are in miRNA seed regions and three are in the miRNA mature region excluding the seed region (Table 3).

For the eight SNPs in miRNA seed regions, four showed all or partial loss of target binding function and one showed gain of target binding in our experiments, which were consistent with the prediction (Fig. 3). The losses of target groups were miR-627(rs2620381)/*SEMA3F*, miR-379(rs61991156)/*SEMA3F*, miR-499-3p(rs3746444)/*BCL2*, and miR-124(rs34059726)/*ATP6V0E1*, while the gain of target group was miR-627(rs2620381)/*ATP6V0E1*. Of these five SNPs showing significant different luciferase activity in our assays, SNP rs34059726 in miR-124 seed region shows the

greatest influence on the interaction between miRNA and target, with the normalized luciferase activity changing from 10% to 80%. The SNP rs2620381 in miR-627 was the first example that was confirmed to create novel target binding by an SNP in seed. MiR-627 does not bind the 3'UTR of gene *ATP6V0E1* in the wild type since the relative luciferase activity of which was the same as controls. While cotransfected the SNP-type miR-627 and reporters, it decreased half of the luciferase activity compared with control (Fig. 3).

To investigate whether SNPs in miRNA mature sequence except seed region could affect the interaction of miRNA and target genes, we also designed three SNPs in miRNA mature sequences (rs35356504 in miR-940, rs72631818 in miR-379, and rs35301225 in miR-34a) for experiments (Table 3). Among them, SNP rs35356504 in miR-940 is an indel SNP, which is the SNP type in a nucleotide deletion. As a surprising result, we found that all the three SNPs in miRNA mature sequence have effects on their target binding (Fig. 3). Although the influences of these SNPs in mature sequences were slight weaker than SNPs in seed regions, they were still statistically significant.



**Figure 3.** Dual-luciferase assay results of miRNAs with different 3' untranslated regions (3'UTRs). Luciferase activity was detected 48 hr after transfection. The luciferase activity in the absence of the oligonucleotides was set to 100% as control. Each bar represents the relative expression level normalized to control and error bars represent the standard deviation. *t*-Tests were used to analyze the difference of expression levels between the wild-miRNA and SNP-miRNA. \**P* < 0.05, \*\**P* < 0.01.

**Table 4. A list of Published SNPs in pre-miRNAs Influencing Mature miRNA Production**

miRNA-SNP allele	Wild base pairing	SNP base pairing	$\Delta\Delta G$ (kcal/mol)	Mature miR production	SNP location	Reference
hsa-miR-510-T/C	A:U	A:C	3.7	Reduced	Stem	Sun et al. [2009]
hsa-miR-890-G/C	C:G	C:C	3.4	Reduced	Stem	Sun et al. [2009]
hsa-miR-892B-T/C	A:U	A:C	4.6	Reduced	Stem	Sun et al. [2009]
hsa-miR-96-G/A	C:G	C:A	7.1	Reduced	Stem	Mencia et al. [2009]
hsa-miR-96-C/A	G:C	G:A	6.6	Reduced	Stem	Mencia et al. [2009]
hsa-miR-125a-G/U	C:G	C:U	6.1	Reduced	Stem	Duan et al. [2007]
Viral miR-K5-G/A	C:G	C:A	3.2	Reduced	Stem	Gottwein et al. [2006]
hsa-miR-502-C/G	G:C	G:G	4.1	Reduced	Stem	Sun et al. [2009]
hsa-miR-140-C/A	C:G	A:G	2.4	Reduced	Stem	Li et al. [2010]
hsa-miR-146a-G/C	U:G	U:C	2.8	Reduced	Stem	Jazdzewski et al. [2008]
hsa-miR-27-A/G	A:-	G:-	-1.21	Increased	Loop	Sun et al. [2009]
hsa-miR-510-G/A	G:U	A:U	-0.3	Increased	Stem	Sun et al. [2009]
hsa-miR-934-T/G	G:U	G:G	2.1	Reduced; MIR shift	Stem and 5p 1	Sun et al. [2009]
hsa-miR-126-A/G	A:U	G:-	NA	Reduced	Downstream pre-miRNA	Harnprasopwat et al. [2010]

“5p 1” represents the first nucleotide of 5' end of mature miRNA.

### Effects on Mature miRNA Production by SNPs in miRNA Genes

Several groups reported that SNPs in miRNA genes could affect the biogenesis process of miRNAs [Duan et al., 2007; Gottwein et al., 2006; Jazdzewski et al., 2008; Li et al., 2010; Mencia et al., 2009; Sun et al., 2009, 2010]. Are there any rules for the effects that SNPs impact on the miRNA biogenesis processing? To address this question, we made an extensive collection for the published results of SNP in pre-miRNAs (Table 4) including their energy changes and mature miRNA production. We observed that 11 SNPs, which reduced the product of mature miRNA, locate in the pre-miRNA stem regions, and change their hairpin structure from stable to unstable status, such as G:C to G:G. The energy change ( $\Delta\Delta G$ ) of the hairpin structure caused by those SNPs is often at a relatively high levels ranging from 2.1 to 7.1. Two SNPs in miRNAs elevated the mature miRNA product, one in the loop and the other in the stem region, which slightly increased the hairpin stability (G:U to A:U) (Table 4).

Based on the above observations, we roughly speculated the effects of SNPs in miRNA genes on mature miRNA production. For an SNP in the miRNA stem, if it decreases the stability of the hairpin structure, it will reduce the product of mature miRNA, otherwise increase the product. The more energy changed, the more likely the product affected. In our results, the average energy change of pre-miRNA secondary structures ( $\Delta\Delta G$ ) caused by SNPs is 2.1 kcal/mol. About 44% of energy changes are  $>2.0$  kcal/mol, which may affect the mature miRNA products significantly. However, since the inference is based on rules summarized from current uncompleted published data, there may be some exceptions and more experimental data are needed to validate.

### miRNA Target Alteration by SNPs in 3'UTR

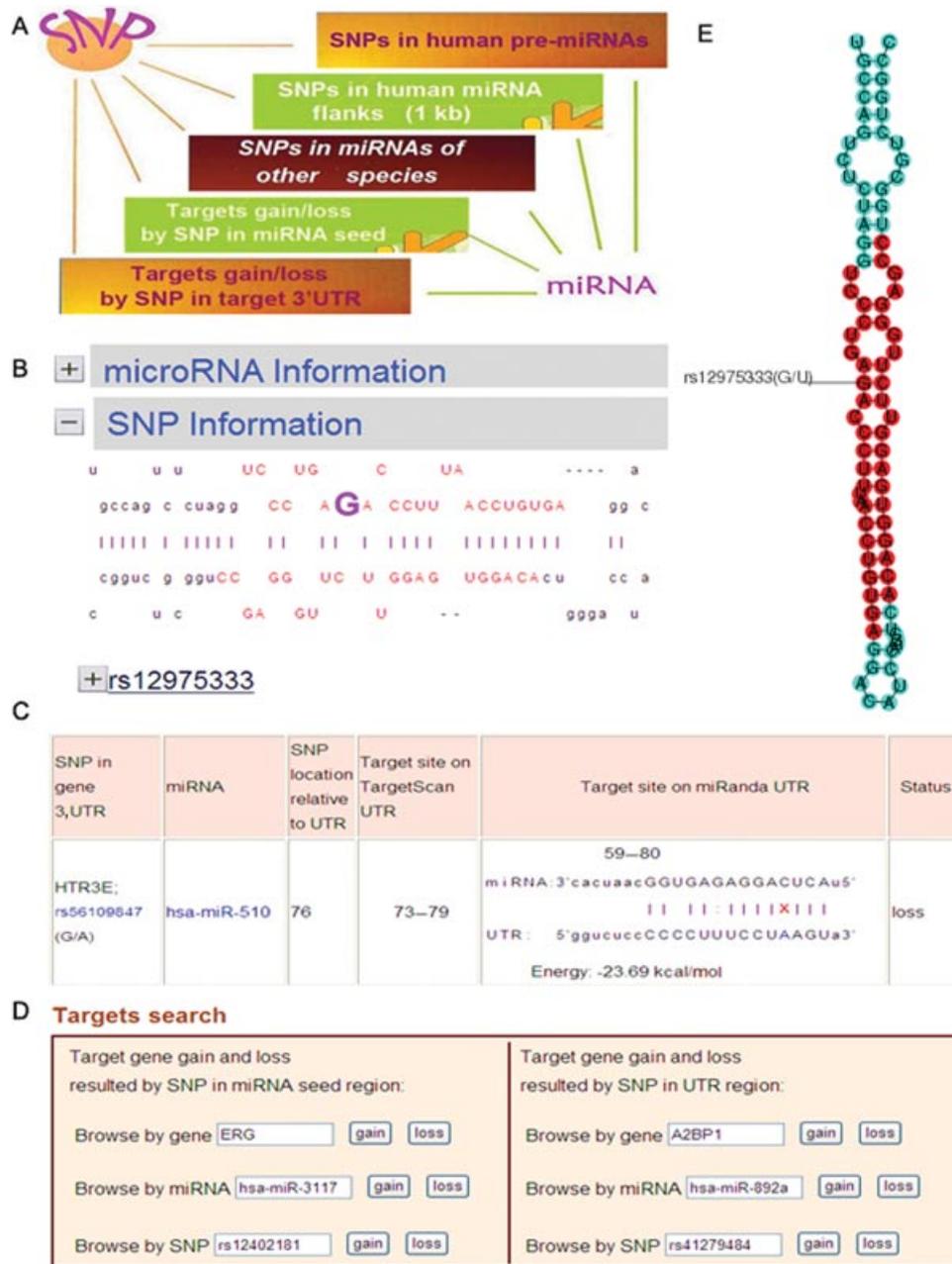
To analyze the influence of SNPs in 3'UTR on miRNA/mRNA interaction, we first mapped SNPs to 3'UTRs of all human protein coding genes and identified 225,759 SNPs in these 3'UTRs. After changing reference 3'UTR (wild-type 3'UTR) sequences to corresponding SNP-3'UTR sequences, we predicted all putative miRNA target sites for wild-type 3'UTR and SNP-3'UTR by both TargetScan and miRanda. As a result, 1,916,262 and 2,154,323 target sites (including nonconserved) were predicted for wild-type 3'UTRs and SNP-3'UTRs. By comparing these target sites, we found a total of 58,977 SNPs disturbed 90,784 original miRNA target sites, while 59,810 SNPs created 91,711 new potential miRNA target sites.

Among these SNPs, 20,779 SNPs could disturb original target sites and create new miRNA target sites at the same time. Therefore, a total of 98,008 (58,977 + 59,810–20,779) SNPs can be considered as possible functional SNPs. Here, the target site gain and loss were defined by the same strategy as SNPs in miRNA seed regions in Figure 2 and Method section. All these target site gain and loss data are shown in the miRNASNP database. In target loss dataset, we identified 31 SNPs with the potential to disrupt 30 experimentally verified miRNA-target pairs collected from the TarBase database and miR2Disease (Supp. Table S2).

To increase the credibility of our results, we further used RNAhybrid to quantitatively measure the binding energy change between miRNA with wild-type 3'UTR or SNP-3'UTR. The averages of energy changes caused by SNPs in 3'UTR were 11.5 kcal/mol and 11.7 kcal/mol in target loss and target gain dataset, respectively. About 50% energy changes in whole dataset were  $>10$  kcal/mol and some of them were even  $>30$  kcal/mol, which will greatly affect the miRNA binding. All the energy change data were shown at the online miRNASNP database.

### The frequencies of miRNA-related SNPs

Generally, functional SNPs are more importance in population genetics if they are at a high frequency or undergo positive selection. To get more information about our identified functional SNPs, we mapped them to HapMap data and iHS data from Haplotter (<http://haplotter.uchicago.edu/>) to check the population allele frequencies and test the positive selection. HapMap contains genotypes and frequency data for different population with relative high frequency. iHS data have been developed as a genomic standardized measure for recent positive selection for a given SNP in a population. After examined the HapMap data, we found only 69 of the 757 SNPs in human pre-miRNAs were sampled in HapMap (Supp. Table S3) and 40 of them have relatively high minor allele frequencies (MAF) ( $q \geq 0.10$ ), among which five in MIRs and one in seed region (rs12220909 in miR-4293). There are 11,150 SNPs in the target loss dataset and 11,190 SNPs in the target gain dataset sampled in HapMap (Supp. Tables S4 and S5), among which 7,443 and 7,339 SNPs have relatively high MAF of  $q \geq 0.10$  in at least one population, respectively. Moreover, 45 SNPs in the target loss dataset and 37 in the target gain dataset show large allele frequency differences ( $q \geq 0.8$ ) between populations (Supp. Tables S6 and S7). Since most of miRNA-related SNPs lack iHS information, we only found 100 SNPs with a relatively high iHS ( $his > 2.5$ ) (Supp. Table S8), which



**Figure 4.** The snapshot of miRNASNP database. (A) The interface of the website with five modules. (B, E) Graphics secondary structure of pre-miRNA with highlighted SNP base. (C) An example of SNP in target site making miRNA/mRNA interaction loss. (D) A part of advanced search page.

correspond to the most extreme 1% of iHS values in genome-wide outliers and represent undergoing recent positive selection pressure [Voight et al., 2006]. These SNPs with high population MAF, with high population frequency difference, or undergoing positive selection pressure will be important candidates for population phenotype research and complex trait studies.

### miRNASNP, an Online Database of miRNA-Related SNPs

To provide a useful resource of these miRNA-related SNPs and their potential target loss and gain information for all researchers, we compiled all the data into a MySQL database and developed a user-

friendly online website, miRNASNP. The miRNASNP database contains five major modules (Fig. 4A): (1) SNPs in human pre-miRNAs; (2) SNPs in human pre-miRNA flanking regions (-1 kb ~ +1kb); (3) SNPs in pre-miRNAs of other eight species; (4) targets gain/loss by SNPs in miRNA seeds; and (5) targets gain/loss by SNPs in target 3'UTRs (Fig. 4C). Besides the basic information for human miRNAs with SNPs in their precursors, we further marked the SNP in the pre-miRNA stem-loop and showed the graphics secondary structure with the SNP (Fig. 4B and E). Species module provides data of SNP-miRNAs in other eight species, which are chimpanzee, mouse, rat, dog, cow, horse, chicken, and zebrafish. The two modules of target gain and loss enable users to investigate the predicted target gain and loss due to SNPs in miRNA seed regions or in target mRNA

3'UTRs. miRNASNP provides different functions for data browsing and searching, which includes a quick search box on top-left of each page by search miRNA ID/SNP ID/gene symbol and an advanced search page for all data in the five modules in miRNASNP (Fig. 4D).

## Discussion

In recent years, SNPs in miRNA target sites have been widely studied to be associated with diseases from hereditary diseases to different cancers (review in [Sethupathy and Collins, 2008]). And SNPs in miRNA genes were reported to involve in the alteration of miRNA processing [Duan et al., 2007; Harnprasopwat et al., 2010], thyroid cancer [Jazdzewski et al., 2009], gastric cancer risk [Peng et al., 2009], ulcerative colitis [Okubo et al., 2011], squamous cell carcinoma [Liu et al., 2010], and nonsyndromic progressive hearing loss [Mencia et al., 2009]. Thus, identifying functional miRNA-related SNPs are of interest for diseases and complex trait studies. However, the effects on miRNA biogenesis and target selection of SNPs in miRNA genes have not been studied extensively. In the first step of this study, we identified the miRNA-related SNPs and summarized their features. Then, we focused on the prediction of potential effects on miRNA biogenesis and target binding by SNPs in miRNA genes through both prediction and experimental validation. Finally, we compiled all the data into the miRNASNP, a free online database. Considering the wide regulation of miRNA and widely existed SNP, our identified functional miRNA-related SNPs will be a useful resource to mine SNP-associated disease or phenotype in population.

### Promising SNPs in Human miRNA Precursors

Since miRNA functions as a top regulator involved in a wide range of regulation, SNPs in miRNA genes may affect miRNA function by influencing the miRNA biogenesis process or target interactions, thus cause serious consequences. In this study, we identified 757 SNPs in human miRNA genes, and further examined the HapMap data for these SNPs. Although lots of SNPs lack frequency information in HapMap, we still found that 69 of them were sampled in HapMap and 40 of them with relatively high MAF ( $q \geq 0.1$ ) in at least one population. In the Result section, we inferred the effects on maturation of the SNPs in pre-miRNAs by summarizing published examples. According to our speculated rules, eight of the 40 SNPs (rs11614913, rs13299349, rs13447640, rs6971711, rs11844707, rs72246410, rs4822739, and rs17797090) locate in stem regions with  $\Delta\Delta G > 2$  kcal/mol, which may decrease the MIR production. Another seven SNPs (rs2910164, rs2292832, rs10505168, rs5997893, rs12780876, rs10934682, and rs2043556) locate in stem regions with  $\Delta\Delta G < -2$  kcal/mol, which may increase the MIR production. Theoretically, these SNPs would greatly change the production of the mature miRNAs, thus may contribute to genetic difference among different population. However, a SNP in the pre-miRNA with relatively low MAF may also cause serious consequences in individuals once it occurs. For example, two SNPs (+13 G > A) and (+14 C > A) in miR-96 seed region were observed in a Spanish family with autosomal dominant progressive high-frequency hearing loss due to impaired maturation and disturbed target sites [Mencia et al., 2009].

Since the seed region of an miRNA is the most important feature for its target binding [Bartel, 2009], SNPs in miRNA seed region will influence the miRNA target binding and selection directly. Here, we identified 50 SNPs in the seed regions of 41 human miRNA genes and predicted their target gain and loss effects for these SNPs (Fig. 2

and miRNASNP website). Our predicted results indicate that SNPs in miRNA seed regions would cause nearly half targets loss and gain on average. In our dataset, five miRNAs with SNP in seed region (miR-124, miR-125-5, miR-1302, miR-379, and miR-499-3p) are conserved in mammalia (chimpanzee, mouse, rat, and dog). We extracted their conserved targets and performed KEGG pathway and Gene Ontology enrichment analyses. Results show that miRNA-mediated function will be changed greatly after SNP variants. For example, the conserved targets of wild miR-124 show significant enrichment in terms "regulation of apoptosis," "intracellular membrane-bounded organelle," and "regulation of cellular biosynthetic process," but no enrichment for the conserved targets of its variant (Benjamini corrected  $P$  value  $< 0.05$ ). Of them, miR-124 and miR-125a-5p have experimentally validated targets in TarBase [Sethupathy et al., 2006] and miR2Disease [Jiang et al., 2009], we found 135 validated targets would loss for miR-124 and one validated target would loss for miR-125a-5p once the SNP allele changes.

Notable, the SNP rs12220909 in miR-4293, which is the only one SNP located in seed region and sampled in HapMap with  $q > 0.1$ . We further analyzed the MAF of rs12220909 in HapMap populations and found the frequencies of the allele C are 0 in both Utah residents with Northern and Western European ancestry from the CEPH collection (CEU) and Yoruba populations, 0.034 in Japanese, and 0.211 in Han Chinese. Chinese has a significant higher C genotype ( $\chi^2$  test,  $P < 0.01$ ). The  $\Delta\Delta G$  affected by rs12220909 is  $-0.5$  kcal/mol, which means the SNP type is slightly more stable than wild type and it may increase the mature miRNA expression. Since it locates in the seed region, our target gene gain and loss prediction indicated that miR-4293 would loss 1,735 target genes and only gain 199 target genes after G→C substitution. Gene Ontology and KEGG analyses show that lost target genes significantly enrich in the term of "ion binding," "plasma membrane part," and "small GTPase regulator activity" (Benjamini corrected  $P < 0.05$ ), while the gained target genes by SNP variant do not show significant enrichment in any categories. Although, no studies reported the function of this miRNA currently, it is interesting to study its function and associated phenotype.

Besides seed region, other residues in mature miRNA sequence were suggested to play a modest role in target recognition [Bartel, 2009; Grimson et al., 2007]. After computational prediction for miRNA target gain and loss, we further performed experiments to validate the effects on target binding by SNPs in seed region and mature region. We selected 11 miRNA target pairs for three target genes, which are *ATP6V0E1*, *BCL2*, and *SEMA3F* (Table 3). Among these miRNA-target pairs, two (miR-34a/*BCL2* and miR-124/*ATP6V0E1*) have been validated by others [Wang et al., 2009; Wang and Wang, 2006] and we also confirmed both of them. In our experimental results, five of the eight SNPs in seed regions were proved to dysregulate their targets. Three SNPs in mature sequences other than seed regions all have slight influences on their target binding, even an indel SNP. These results support the conclusion that residues in seed region play key roles and other residues in mature sequence have modest effects on miRNA target binding [Bartel, 2009].

It is worthy of note that in this study, we proved the target gain by SNP rs2620381 in miR-627 seed region by experiments. In wild type of miR-627, it can not bind the 3'UTR of *ATP6V0E1*, while the SNP-type miR-627 gained the ability to target the *ATP6V0E1* 3'UTR and repressed its expression dramatically in our luciferase experiments. To our best knowledge, this is the first experimentally validated example for target gain affected by a SNP in an miRNA. It provides a new mechanism for miRNA dysregulation in different individuals.

Our results show that SNPs in an miRNA gene, especially in the seed region, will alter the target profile of the miRNA by losing original targets and gaining new targets. These SNPs in miRNA seed regions and their target gain and loss information will be a potential useful clue to study the miRNA function and find the SNP-associated disease or phenotype.

### Promising SNPs in 3'UTRs of Human Protein Coding Genes

In contrast to the SNPs in miRNA genes, there are more reports that studied the effect of SNPs in 3'UTRs. Loss of a potential miRNA target site may increase the protein expression, while gain of a functional miRNA target site will repress the protein expression, thus affect physiological function and clinical phenotype. Here, using our pipeline, we identified tens of thousands of SNPs locating in potential miRNA target sites and some of them show high MAF, high MAF difference between populations, or positive selection pressure during evolution. Those SNPs will be important candidates for causal variants of human disease. Currently, genome-wide association studies have uncovered many SNPs associated with traits and diseases. The NHGRI GWAS catalogue (<http://www.genome.gov/gwastudies>, accessed by 2010-12-16) described 1,227 unique SNPs associated with one or more traits ( $P < 5 \times 10^{-8}$ ) [Hindorff et al., 2009]. Among these SNPs, six are in our 3'UTR dataset and three are present in our target loss and gain dataset. They are rs1036819 associated with longevity, rs28927680 associated with triglycerides, and rs1042725 associated with height. The original papers also mentioned that these SNPs in 3'UTR may be involved in the traits by miRNA-mediated regulation but without detail miRNA information. Utilizing our database miRNASNP, users can find the detail information about miRNA and its target gain and loss. For example, when users search rs28927680 in miRNASNP, it will show the SNP locates in the potential target sites of six miRNAs (hsa-miR-1323, hsa-miR-548a-3p, hsa-miR-548e, hsa-miR-548f, hsa-miR-548o, and hsa-miR-548t) in 3'UTR of gene *BUD13*. SNP rs28927680 is reported to be associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, or triglycerides in human [Kathiresan et al., 2008], hence the SNP-associated miRNA and target site information may shed light on further experiments.

There are more than 1,000 experimentally validated miRNA-target pairs in miR2Disease and Tarbase databases. Based on these data and miRNA-related SNPs in our miRNASNP database, we identified 31 SNPs in 3'UTRs with the abilities to disturb experimental validated miRNA-target pairs. Three of them (rs5186, rs12720208, and rs56109847) have been experimental confirmed to make dysregulate their corresponding targets and associated with diseases. Sethupathy et al. demonstrated that the SNP (rs5186) in the *AGTR1* 3'UTR mediates allele-specific targeting of miR-155 to *AGTR1*, thereby modulating *AGTR1* protein levels [Sethupathy et al., 2007]. SNP rs12720208 was proved to mediate allele-specific in vitro targeting of miR-433 to the *FGF20* 3'UTR and confers risk for Parkinson disease [Wang et al., 2008]. Kapeller et al. identified rs62625044 (now merged into rs56109847) in the 3'UTR of *HTR3E*, which could mediate allele-specific miR-510 targeting. This was associated with diarrhea-predominant irritable bowel syndrome (IBS-D) in females from the United Kingdom and was also confirmed in a German cohort by replication study [Kapeller et al., 2008]. Besides the three validated SNPs, the rest will be attractive SNPs in human miRNA target sites for future studies.

### Acknowledgments

We thank Wei Liu and Hui Liu for helpful discussions.

### References

- Bao L, Zhou M, Wu L, Lu L, Goldowitz D, Williams RW, Cui Y. 2007. PolymiRTS database: linking polymorphisms in microRNA target sites with complex traits. *Nucleic Acids Res* 35:D51–D54.
- Barenboim M, Zoltick BJ, Guo Y, Weinberger DR. 2010. MicroSNiPer: a web tool for prediction of SNP effects on putative microRNA targets. *Hum Mutat* 31:1223–1232.
- Barrell D, Dimmer E, Huntley RP, Binns D, O'Donovan C, Apweiler R. 2009. The GOA database in 2009—an integrated Gene Ontology Annotation resource. *Nucleic Acids Res* 37:D396–D403.
- Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116:281–297.
- Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* 136:215–233.
- Betel D, Wilson M, Gabow A, Marks DS, Sander C. 2008. The microRNA.org resource: targets and expression. *Nucleic Acids Res* 36:D149–D153.
- Bhartiya D, Laddha SV, Mukhopadhyay A, Scaria V. 2011. miRvar: A comprehensive database for genomic variations in microRNAs. *Hum Mutat* 32:E2226–E2245.
- Chin LJ, Ratner E, Leng S, Zhai R, Nallur S, Babar I, Muller RU, Straka E, Su L, Burki EA, Crowell RE, Patel R, Kulkarni T, Homer R, Zelterman D, Kidd KK, Zhu Y, Christiani DC, Belinsky SA, Slack FJ, Weidhaas JB. 2008. A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk. *Cancer Res* 68:8535–8540.
- Denman RB. 1993. Using RNAFOLD to predict the activity of small catalytic RNAs. *Biotechniques* 15:1090–1095.
- Duan R, Pak C, Jin P. 2007. Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. *Hum Mol Genet* 16:1124–1131.
- Duan S, Mi S, Zhang W, Dolan ME. 2009. Comprehensive analysis of the impact of SNPs and CNVs on human microRNAs and their regulatory genes. *RNA Biol* 6:412–425.
- Friedman RC, Farh KK, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 19:92–105.
- Gottwein E, Cai X, Cullen BR. 2006. A novel assay for viral microRNA function identifies a single nucleotide polymorphism that affects Drosha processing. *J Virol* 80:5321–5326.
- Grimson A, Farh KK, Johnston WK, Garrett-Engle P, Lim LP, Bartel DP. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 27:91–105.
- Harihara M, Scaria V, Brahmachari SK. 2009. dbSMR: a novel resource of genome-wide SNPs affecting microRNA mediated regulation. *BMC Bioinformatics* 10:108.
- Harnprasopwat R, Ha D, Toyoshima T, Lodish H, Tojo A, Kotani A. 2010. Alteration of processing induced by a single nucleotide polymorphism in pri-miR-126. *Biochem Biophys Res Commun* 399:117–122.
- Hiard S, Charlier C, Coppieters W, Georges M, Baurain D. 2010. Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. *Nucleic Acids Res* 38:D640–D651.
- Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 106:9362–9367.
- Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4:44–57.
- Iwai N, Naraba H. 2005. Polymorphisms in human pre-miRNAs. *Biochem Biophys Res Commun* 331:1439–1444.
- Jazdzewski K, Liyanarachchi S, Swierniak M, Pachucki J, Ringel MD, Jarzab B, de la Chapelle A. 2009. Polymorphic mature microRNAs from passenger strand of pre-miR-146a contribute to thyroid cancer. *Proc Natl Acad Sci USA* 106:1502–1505.
- Jazdzewski K, Murray EL, Franssila K, Jarzab B, Schoenberg DR, de la Chapelle A. 2008. Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. *Proc Natl Acad Sci USA* 105:7269–7274.
- Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, Li M, Wang G, Liu Y. 2009. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res* 37:D98–D104.
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32:D277–D280.
- Kapeller J, Houghton LA, Monnikes H, Walstab J, Moller D, Bonisch H, Burwinkel B, Autschbach F, Funke B, Lasitschka F and others. 2008. First evidence for an association of a functional variant in the microRNA-510 target site of the serotonin receptor-type 3E gene with diarrhea predominant irritable bowel syndrome. *Hum Mol Genet* 17:2967–2977.
- Kathiresan S, Melander O, Guiducci C, Surti A, Burt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS and others. 2008. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 40:189–197.
- Khvorova A, Reynolds A, Jayasena SD. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell* 115:209–216.

- Kozomara A, Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 39:D152–157.
- Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M and others. 2005. Combinatorial microRNA target predictions. *Nat Genet* 37:495–500.
- Kruger J, Rehmsmeier M. 2006. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res* 34:W451–W454.
- Lai EC. 2002. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* 30:363–364.
- Landi D, Gemignani F, Barale R, Landi S. 2008. A catalog of polymorphisms falling in microRNA-binding regions of cancer genes. *DNA Cell Biol* 27:35–43.
- Lee HC, Yang CW, Chen CY, Au LC. 2011. Single point mutation of microRNA may cause butterfly effect on alteration of global gene expression. *Biochem Biophys Res Commun* 404:1065–1069.
- Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S and others. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425:415–419.
- Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120:15–20.
- Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. 2003. Prediction of mammalian microRNA targets. *Cell* 115:787–798.
- Li L, Meng T, Jia Z, Zhu G, Shi B. 2010. Single nucleotide polymorphism associated with nonsyndromic cleft palate influences the processing of miR-140. *Am J Med Genet A* 152A:856–862.
- Liu Z, Li G, Wei S, Niu J, El-Naggar AK, Sturgis EM, Wei Q. 2010. Genetic variants in selected pre-microRNA genes and the risk of squamous cell carcinoma of the head and neck. *Cancer* 116:4753–4760.
- Mencia A, Modamio-Hoybjor S, Redshaw N, Morin M, Mayo-Merino F, Olavarrieta L, Aguirre LA, del Castillo I, Steel KP, Dalmay T and others. 2009. Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nat Genet* 41:609–613.
- Okubo M, Tahara T, Shibata T, Yamashita H, Nakamura M, Yoshioka D, Yonemura J, Kamiya Y, Ishizuka T, Nakagawa Y and others. 2011. Association study of common genetic variants in pre-microRNAs in patients with ulcerative colitis. *J Clin Immunol* 31:69–73.
- Peng S, Kuang Z, Sheng C, Zhang Y, Xu H, Cheng Q. 2009. Association of microRNA-196a-2 gene polymorphism with gastric cancer risk in a Chinese population. *Dig Dis Sci* 55:2288–2293.
- Ryan BM, Robles AI, Harris CC. 2010. Genetic variation in microRNA networks: the implications for cancer research. *Nat Rev Cancer* 10:389–402.
- Saunders MA, Liang H, Li WH. 2007. Human polymorphism at microRNAs and microRNA target sites. *Proc Natl Acad Sci USA* 104:3300–3305.
- Sethupathy P, Borel C, Gagnebin M, Grant GR, Deutsch S, Elton TS, Hatzigeorgiou AG, Antonarakis SE. 2007. Human microRNA-155 on chromosome 21 differentially interacts with its polymorphic target in the AGTR1 3' untranslated region: a mechanism for functional single-nucleotide polymorphisms related to phenotypes. *Am J Hum Genet* 81:405–413.
- Sethupathy P, Collins FS. 2008. MicroRNA target site polymorphisms and human disease. *Trends Genet* 24:489–497.
- Sethupathy P, Corda B, Hatzigeorgiou AG. 2006. TarBase: a comprehensive database of experimentally supported animal microRNA targets. *RNA* 12:192–197.
- Shastri BS. 2009. SNPs: impact on gene function and phenotype. *Methods Mol Biol* 578:3–22.
- Sun G, Yan J, Noltner K, Feng J, Li H, Sarkis DA, Sommer SS, Rossi JJ. 2009. SNPs in human miRNA genes affect biogenesis and function. *RNA* 15:1640–1651.
- Sun Q, Gu H, Zeng Y, Xia Y, Wang Y, Jing Y, Yang L, Wang B. 2010. Hsa-mir-27a genetic variant contributes to gastric cancer susceptibility through affecting miR-27a and target gene expression. *Cancer Sci* 101:2241–2247.
- Voight BF, Kudravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol* 4:e72.
- Wang G, van der Walt JM, Mayhew G, Li YJ, Zuchner S, Scott WK, Martin ER, Vance JM. 2008. Variation in the miRNA-433 binding site of FGF20 confers risk for Parkinson disease by overexpression of alpha-synuclein. *Am J Hum Genet* 82:283–289.
- Wang X, Liu P, Zhu H, Xu Y, Ma C, Dai X, Huang L, Liu Y, Zhang L, Qin C. 2009. miR-34a, a microRNA up-regulated in a double transgenic mouse model of Alzheimer's disease, inhibits bcl2 translation. *Brain Res Bull* 80:268–273.
- Wang X, Wang X. 2006. Systematic identification of microRNA functions by combining target prediction and expression profiling. *Nucleic Acids Res* 34:1646–1652.
- Zhang B, Wang Q, Pan X. 2007. MicroRNAs and their regulatory roles in animals and plants. *J Cell Physiol* 210:279–289.