

# Biological Basis of miRNA Action when Their Targets Are Located in Human Protein Coding Region

Wanjun Gu<sup>1\*</sup>, Xiaofei Wang<sup>1</sup>, Chuanying Zhai<sup>1</sup>, Tong Zhou<sup>2,3\*</sup>, Xueying Xie<sup>1\*</sup>

**1** Research Center of Learning Sciences, Southeast University, Nanjing, Jiangsu, China, **2** Institute for Personalized Respiratory Medicine, The University of Illinois at Chicago, Chicago, Illinois, United States of America, **3** Section of Pulmonary, Critical Care, Sleep & Allergy, Department of Medicine, The University of Illinois at Chicago, Chicago, Illinois, United States of America

## Abstract

Recent analyses have revealed many functional microRNA (miRNA) targets in mammalian protein coding regions. But, the mechanisms that ensure miRNA function when their target sites are located in protein coding regions of mammalian mRNA transcripts are largely unknown. In this paper, we investigate some potential biological factors, such as target site accessibility and local translation efficiency. We computationally analyze these two factors using experimentally identified miRNA targets in human protein coding region. We find site accessibility is significantly increased in miRNA target region to facilitate miRNA binding. At the mean time, local translation efficiency is also selectively decreased near miRNA target region. GC-poor codons are preferred in the flank region of miRNA target sites to ease the access of miRNA targets. Within-genome analysis shows substantial variations of site accessibility and local translation efficiency among different miRNA targets in the genome. Further analyses suggest target gene's GC content and conservation level could explain some of the differences in site accessibility. On the other hand, target gene's functional importance and conservation level can affect local translation efficiency near miRNA target region. We hence propose both site accessibility and local translation efficiency are important in miRNA action when miRNA target sites are located in mammalian protein coding regions.

**Citation:** Gu W, Wang X, Zhai C, Zhou T, Xie X (2013) Biological Basis of miRNA Action when Their Targets Are Located in Human Protein Coding Region. PLoS ONE 8(5): e63403. doi:10.1371/journal.pone.0063403

**Editor:** Marc Tjwa, University of Frankfurt - University Hospital Frankfurt, Germany

**Received:** February 5, 2013; **Accepted:** March 30, 2013; **Published:** May 6, 2013

**Copyright:** © 2013 Gu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by grants from the National Basic Research Program of China (2012CB316501), the National High Technology Research and Development Program of China (2012AA020401), the National Natural Science Foundation of China (30900836, 61171143 and 61001175), and a Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, China. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

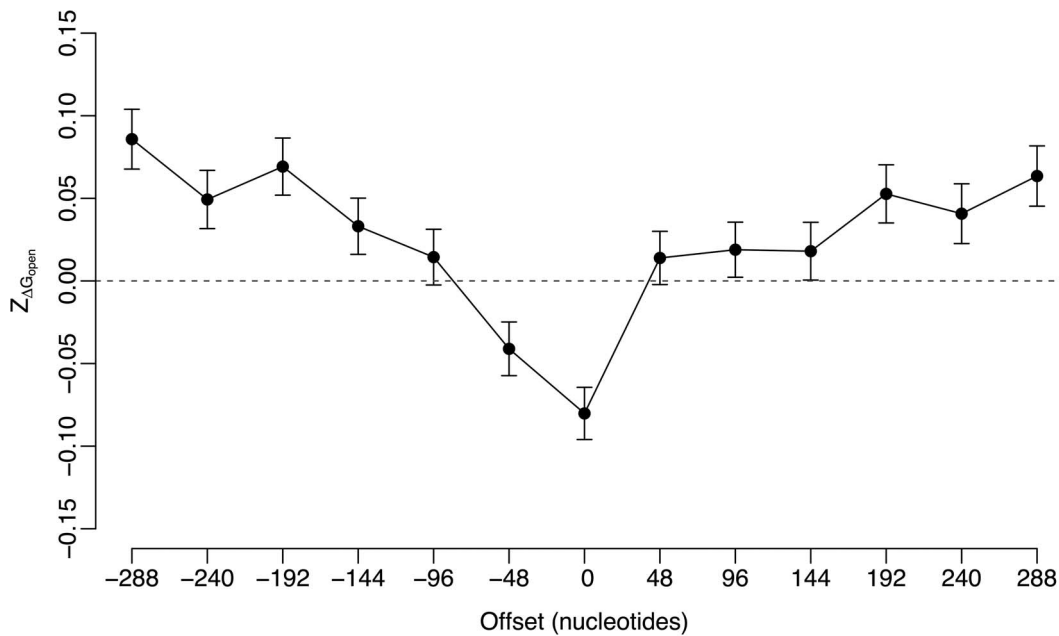
\* E-mail: wanjungu@gmail.com (WG); tongzhou@uic.edu (TZ); xxying.rcls@seu.edu.cn (XX)

## Introduction

miRNAs are a class of small non-coding RNAs that regulate gene expression in post-transcriptional stage [1]. In the process of miRNA action, miRNAs are first specifically bound to their target sites in mRNA transcripts [1]. After binding, targeted mRNA transcripts can be silenced by several mechanisms, such as gene silencing, translational repression and mRNA degradation [2]. In plants, most miRNA target sites are located in protein coding region of target gene. Unlike plants, animal miRNAs have their target sites mostly in 3' untranslated region (3'-UTR) of mRNA transcripts [1]. Two recent studies have investigated the reason why mammalian miRNA target sites are restricted to the 3'-UTR of mRNA transcripts [3,4]. They have suggested active mRNA translation may impede miRNA association with target mRNAs in mammalian genomes [3,4]. But, increasing evidences have confirmed many functional miRNA target sites are located in protein coding region of mammalian mRNA transcripts as well [5–12]. Notably, genome-wide analysis of Argonaute (AGO) HITS-CLIP data [13] and PAR-CLIP data [14] have revealed almost half AGO binding sites are located in human protein coding region. Given the prevalence of miRNA target sites occurred in mammalian protein coding sequences, it is important to investigate biological factors that may affect miRNA action

when their targets are located in mammalian protein coding regions.

Comparative genomic analyses have been successfully used in identifying important factors of miRNA action [1,15]. For example, miRNA genes are evolutionary conserved between species [1,16] and within species [17]. At the same time, nucleotides in miRNA target region are under negative selection as well [18,19]. These have implicated the importance of sequence conservation for proper miRNA function [1]. Comparing with miRNA target sites that are located in 3'-UTR of mammalian mRNA transcripts, those in mammalian protein coding region still need code amino acid information of their translated proteins. Previous studies have suggested the degeneracy of genetic code enables DNA sequences to code extra regulatory information as well as amino acid sequences [20]. Biased usage of synonymous codons have been related to many biological processes, such as DNA stability [21], nucleosome positioning [22], mRNA stability [23–25], mRNA splicing [26,27], nonsense mediated mRNA decay [28], translation initiation [29–31], translation elongation [32–36] and co-translational protein folding [37,38]. Additionally, many studies have suggested synonymous codon choices near miRNA target sites are related to miRNA function in mammalian genomes. For example, synonymous substitution rate in miRNA target region is reduced in some mammalian genomes [39–42]. Tay *et al.* [12] have found silent mutations occurred in miRNA



**Figure 1. The mean and standard error of  $Z_{\Delta G_{open}}$  of each sliding window near miRNA target region in the human genome.**  
doi:10.1371/journal.pone.0063403.g001

target sites can eliminate miRNA activity in mouse. Brest *et al.* [43] have suggested that human Crohn's disease is caused by a synonymous mutation at the binding site of miR-196 in *IRGM*. Hence, analysis of synonymous codon usage in the flank region of miRNA target sites should be able to gain some insights of miRNA action.

In this paper, we analyzed the usage of synonymous codons in the flank region of miRNA target sites that are located in human protein coding sequences. We chose human as the example of mammalian species, since many functional miRNA targets have been experimentally identified in human protein coding region [14]. We considered site accessibility and local translation efficiency as two possible biological determinants of miRNA action when their target sites are located in human protein coding region. Site accessibility is one of the most important factors that affect miRNA binding when miRNA target sites are located in 3'-UTR of mammalian mRNA transcripts [44,45]. On the other hand, local translation efficiency around miRNA target region is particularly crucial for those targets located in mammalian protein coding region [3,4]. We hypothesized that if site accessibility and local translation efficiency are important in miRNA action, they should be selectively varied at synonymous codon sites that are in the flank region of miRNA target sites. We computed site accessibility and local translation efficiency for each miRNA target in human protein coding region. To estimate the selection pressure and its significance, we permuted mRNA sequences and assessed the deviation of local translation efficiency and site accessibility from random expectation. We addressed the following several problems: 1) is site accessibility important in miRNA action when their target sites are located in protein coding region of mammalian transcripts? 2) Is local translation efficiency near miRNA target region also important for proper miRNA function? 3) If site accessibility and/or local translation efficiency are selectively varied near miRNA target region, what are the factors that can affect the selection pressure?

## Results

### Site accessibility is selectively varied in the flank region of miRNA target sites

$Z_{\Delta G_{open}}$  measures the extent to which site accessibility deviates from random expectation. A negative  $Z_{\Delta G_{open}}$  means that site accessibility is increased, and a positive  $Z_{\Delta G_{open}}$  means it is decreased. We calculated  $Z_{\Delta G_{open}}$  along mRNA sequences in sliding windows of 48 nucleotides in length. We started from the miRNA target region, which contains 21 nucleotides bound to miRNAs, 17 flank upstream nucleotides and 10 flank downstream nucleotides. We moved the sliding window both upward and downward along the mRNA sequences in a step of 48 nucleotides. We calculated  $Z_{\Delta G_{open}}$  values in 13 consecutive windows for each miRNA target. For each window, we calculated a genomic mean value of  $Z_{\Delta G_{open}}$  by averaging  $Z_{\Delta G_{open}}$  values over all miRNA targets in the human genome.

Figure 1 shows the genomic mean value of  $Z_{\Delta G_{open}}$  for all 13 consecutive windows in human. We observed a significant negative deviation of  $Z_{\Delta G_{open}}$  from zero (*t*-test,  $P < < 10^{-6}$ ) in the central window. The negative values of  $Z_{\Delta G_{open}}$  in the central window suggest selection for increased site accessibility in miRNA target region. When the sliding window moved upward or downward along the mRNA sequence,  $Z_{\Delta G_{open}}$  values increase quickly and most windows have significant positive mean  $Z_{\Delta G_{open}}$  values. The positive  $Z_{\Delta G_{open}}$  values in these windows suggest decreased site accessibility is generally preferred in mRNA segments other than the miRNA target region. In following analysis, we used  $Z_{\Delta G_{open}}$  in the central window to represent the selection signal of site accessibility in miRNA target region.

### Translation efficiency is also selectively varied near miRNA target region

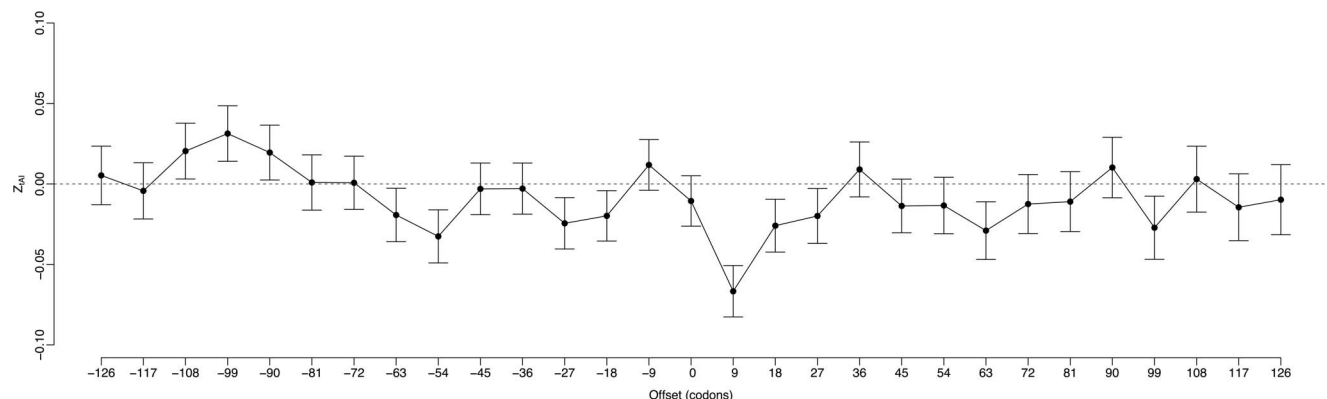
$Z_{IAI}$  measures the deviation of local translation efficiency from random expectation. A negative  $Z_{IAI}$  value means local translation efficiency is reduced, and a positive  $Z_{IAI}$  value means it is

increased. We calculated  $Z_{IAI}$  along mRNA sequences using a sliding window of nine codons (27 nucleotides) in length. We started from a window that is right downstream of miRNA target sites, and moved the window upward and downward along the mRNA sequence at a step of nine codons. We calculated  $Z_{IAI}$  in 29 consecutive windows for each miRNA target region. For each window, we calculated a genomic mean  $Z_{IAI}$  by averaging  $Z_{IAI}$  values over all miRNA target regions.

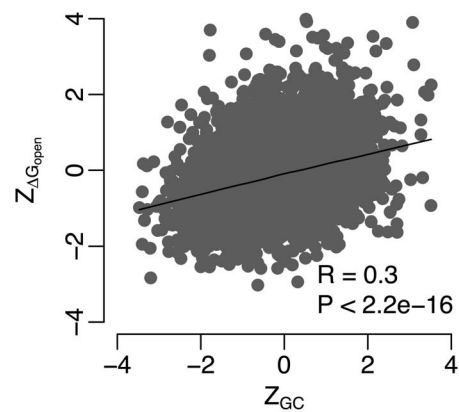
Figure 2 shows mean  $Z_{IAI}$  values in all 29 consecutive windows. We observed a significant negative deviation of  $Z_{IAI}$  value from zero ( $t$ -test,  $P < 10^{-4}$ ) in a window that is nine codons downstream from the start point. The negative  $Z_{IAI}$  values in this window suggest decreased translation efficiency is selectively preferred near miRNA target sites. We did not observe any significant deviation of  $Z_{IAI}$  values from zero in other windows. This suggests reduced translation efficiency is only selectively preferred in the flank region of miRNA target sites. We used  $Z_{IAI}$  values in the window nine codons downstream of the start point to represent selection signal of local translation efficiency for following analysis.

### GC-poor codons are preferred in miRNA target region

We observed site accessibility tended to increase in miRNA target region (Figure 1). Site accessibility is mainly determined by RNA secondary structure near miRNA target region. We calculated  $Z$ -score of local RNA secondary structure,  $Z_{\Delta G_{local}}$ , in all 13 sliding windows as we did in site accessibility analysis. We also observed decreased RNA stability in miRNA target region (Figure S1).  $Z_{\Delta G_{local}}$  in miRNA target region is correlated well to  $Z_{\Delta G_{open}}$  in that region (Pearson's product-moment correlation =  $-0.3604834$ ,  $P < 10^{-15}$ ; Figure S2). When looking into GC content in miRNA target region, we also observed a significant negative deviation of  $Z_{GC}$  values from zero ( $Z_{GC} = -0.106 \pm 0.034$ ;  $t$ -test,  $P = 1.3 \times 10^{-9}$ ). A negative  $Z_{GC}$  value in miRNA target region suggests GC-poor codons are selectively preferred in that region. When comparing  $Z_{GC}$  and  $Z_{\Delta G_{open}}$  of each miRNA target, we observed a significant positive correlation (Pearson's product-moment correlation =  $0.3$ ,  $P < 10^{-15}$ ; Figure 3). We also observed a significant negative correlation (Pearson's product-moment correlation =  $0.25$ ,  $P < 10^{-15}$ ; Figure S3) between  $Z_{GC}$  and  $Z_{\Delta G_{local}}$  in miRNA target region. This suggests GC-poor codons are selectively preferred in miRNA target region for increased site accessibility by loosening RNA structure in that region.



**Figure 2. The mean and standard error of  $Z_{IAI}$  of each sliding window near miRNA target region in the human genome.**  
doi:10.1371/journal.pone.0063403.g002



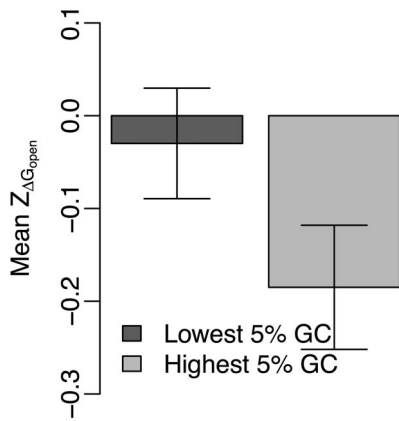
**Figure 3.  $Z_{\Delta G_{open}}$  in miRNA target region as a function of  $Z_{GC}$  in that region.** Each point represents a miRNA target in human protein coding sequences.  
doi:10.1371/journal.pone.0063403.g003

### Factors that influence the selection pressure among miRNA target regions

In previous sections, we considered the mean  $Z_{\Delta G_{open}}$  and  $Z_{IAI}$  over all miRNA targets in protein coding region of the human genome. But, we observed substantial variations of site accessibility and translation efficiency among different miRNA targets in the genome (Figures 1 and 2). We next investigated the factors that may affect site accessibility and local translation efficiency of miRNA targets.

We first considered target gene's GC content. We selected miRNA targets in genes with the highest 5% and the lowest 5% GC content, and compared the mean  $Z_{\Delta G_{open}}$  and  $Z_{IAI}$  of miRNA targets in these two groups. We observed  $Z_{\Delta G_{open}}$  values of miRNA targets in GC-rich genes were significantly smaller than those in GC-poor genes (Welch Two Sample  $t$ -test,  $P = 0.04$ ; Figure 4). This suggests miRNA targets in GC-rich genes are under higher selection pressures to increase site accessibility. But,  $Z_{IAI}$  values of miRNA targets in genes with the highest 5% GC content and those with the lowest 5% GC content are not statistically different (Welch Two Sample  $t$ -test,  $P = 0.42$ ; Figure S4).

Next we considered the conservation level of miRNA target gene. We separated miRNA targets into three groups based on gene conservation, including primate-specific targets, mammal-specific targets and vertebrate-conserved targets. We compared

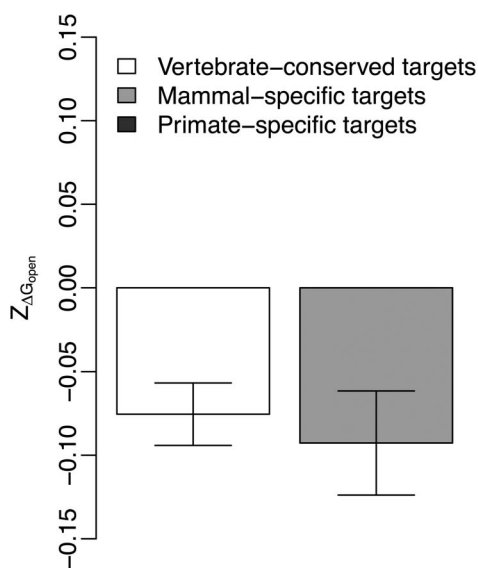


**Figure 4. Comparison of the mean  $Z_{\Delta G_{open}}$  between miRNA targets in genes with the highest 5% and the lowest 5% GC content.**

doi:10.1371/journal.pone.0063403.g004

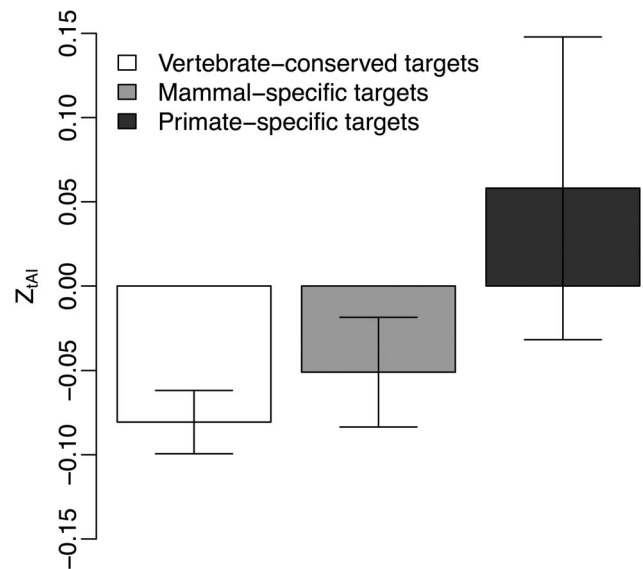
$Z_{\Delta G_{open}}$  and  $Z_{IAI}$  values of miRNA targets in these three groups. Both  $Z_{\Delta G_{open}}$  and  $Z_{IAI}$  mean of mammal-specific and vertebrate-conserved targets are smaller than those in primate-specific targets (Figures 5 and 6). While  $Z_{\Delta G_{open}}$  values of mammal-specific targets ( $t$ -test,  $P < < 10^{-4}$ ) and vertebrate-conserved targets ( $t$ -test,  $P = 0.003$ ) are significantly deviated to zero,  $Z_{IAI}$  values of mammal-specific targets ( $t$ -test,  $P < < 10^{-4}$ ) and vertebrate-conserved targets ( $t$ -test,  $P = 0.012$ ) are also significantly deviated to zero. But, neither  $Z_{\Delta G_{open}}$  ( $t$ -test,  $P = 0.938$ ) nor  $Z_{IAI}$  values ( $t$ -test,  $P = 0.519$ ) in primate-specific targets have significant deviation to zero. This suggests selection of site accessibility and translation efficiency is more pronounced for miRNA targets with longer evolutionary history.

We then considered the functional importance of target gene. We used the protein complex size to denote gene's functional importance. We compared  $Z_{\Delta G_{open}}$  and  $Z_{IAI}$  values of miRNA targets in genes with the highest 5% complex size to those with the



**Figure 5. Comparison of the mean  $Z_{\Delta G_{open}}$  between miRNA targets in genes at different conservation levels.**

doi:10.1371/journal.pone.0063403.g005

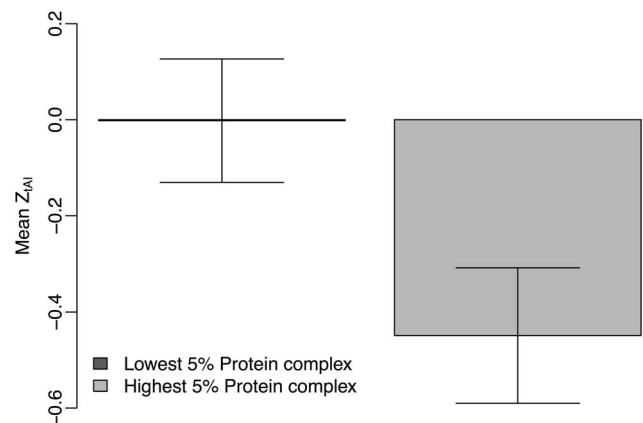


**Figure 6. Comparison of the mean  $Z_{IAI}$  between miRNA targets in genes at different conservation levels.**

doi:10.1371/journal.pone.0063403.g006

lowest 5% complex size. The mean  $Z_{IAI}$  value of miRNA targets in genes with higher functionality is smaller than that of miRNA targets in genes with less functionality (Figure 7). We did not observe any significant difference between  $Z_{\Delta G_{open}}$  values of miRNA targets in genes with different functional importance (Figure S5).

We finally considered the expression level and codon bias of target gene. We used  $ENC$  (Effective Number of Codons) to measure a gene's codon usage bias [46]. The higher a gene's codon usage bias, the lower the gene's  $ENC$ . We compared  $Z_{\Delta G_{open}}$  and  $Z_{IAI}$  values of miRNA targets in genes the highest 5% expression level to those with the lowest 5% expression level. We also compared  $Z_{\Delta G_{open}}$  and  $Z_{IAI}$  values of miRNA targets in genes with the highest 5%  $ENC$  to those with the lowest 5%  $ENC$ . But, we did not observe any significant difference in these comparisons (Figures S6, S7, S8 and S9).



**Figure 7. Comparison of the mean  $Z_{IAI}$  between miRNA targets in genes with the highest 5% and lowest 5% protein complex size.**

doi:10.1371/journal.pone.0063403.g007

## Discussion

To investigate the biological basis of miRNA action when miRNA targets are located in protein coding region of mammalian transcripts, we have performed a genome scale analysis of site accessibility and translation efficiency near miRNA target region in the human genome. We have found both site accessibility and translation efficiency is selectively varied in the flank region of miRNA target sites (Figures 1 and 2). By focusing on  $Z$ -scores relative to the random expectation, our method has excluded some biases such as amino acid content or preferred codon usage that may cause the signal. Site accessibility in miRNA target region tends to be higher in real mRNA sequence than that expected from the permuted mRNA sequences (Figure 1). Our results are comparable to the observations made in some previous studies [44,45]. They had found site accessibility was the key factor in determining miRNA activity when miRNA target sites are located in 3'-UTR of human transcripts. Another two experimental studies had found active mRNA translation could impede miRNA association with target mRNAs in mammalian genomes [3,4]. When rare codons are introduced in the upstream region of miRNA target sites, miRNA activity could be restored [3]. In our results, we have found translation efficiency is decreased in the flank region of miRNA target sites (Figure 2). This confirms the importance of local translation efficiency in miRNA action for miRNA targets occurred in mammalian protein coding region. We have not observed any significant correlation between site accessibility and translation efficiency in the flank region of miRNA target sites (Pearson's product-moment correlation = 0.02,  $P = 0.27$ ; Figure S10). Hence, site accessibility and local translation efficiency may be two independent factors in miRNA action when miRNA targets are located in mammalian protein coding regions.

In previous analysis, we had found site accessibility, but not local translation efficiency, was related to miRNA action when miRNA targets are located in protein coding region of plant genes [47,48]. This suggests site accessibility is a much general factor that affects miRNA action for different kinds of miRNA targets [44,45,47]. Comparing with site accessibility, local translation efficiency is only effective when miRNA targets are located in mammalian protein coding region [3,4,48]. Is it possible that the selection signal for reduced translation efficiency near miRNA target sites is an artifact caused by our method? To eliminate this possibility, for each miRNA target, we randomly picked a region of 21 nucleotides in length from the same mRNA sequence. We combined all these randomly selected gene regions as a dataset of randomized miRNA targets. We replicated our analysis of translation efficiency on all randomized miRNA targets. We did not observe any selection signal of reduced translation efficiency in all sliding windows (Figure S11). This implies that reduced translation efficiency near miRNA target region we have observed Figure 2 is a real signal. The reason why reduced translation efficiency is only important for miRNA targets in mammalian protein coding region is unknown. Some analysis had suggested most animal miRNAs regulated their target genes by translational repression, while plant miRNAs regulated their target genes mainly by RNA degradation [1,2]. This difference may explain the different importance of local translation efficiency in miRNA action between plants and mammals.

In our results, the region that tends to increase site accessibility is miRNA target region, which includes miRNA target sites, 17 flank upstream nucleotides and 10 downstream nucleotides (Figure 1). This region is the same as the one that we had observed in four plant genomes [29], and is comparable to those had been identified from experimental data [44,45]. We have

observed that  $Z_{\Delta G_{open}}$  value in most other windows is positive (Figure 1). This implies site accessibility in mRNA sequences is generally decreased. This is mainly caused by much tighter RNA structure in mRNA sequences (Figure S1). The latter result is comparable to the observations made by several previous studies that the overall mRNA stability is selectively increased in human [49,50].

We have observed that the region with decreased translation efficiency is located downstream of miRNA target sites (Figure 2). This region is different from the one suggested by Gu *et al.* [3]. They had employed model reporter constructs by introducing mutations into stop codons located upstream of miRNA target sites, which allowed translation to proceed through miRNA target sites [3]. They had found active translation could totally impede miRNA regulation. But, miRNA regulation can be restored when nine consecutive rare codons are introduced right upstream of miRNA target sites. In contrast, Lin *et al.* [4] had analyzed two naturally occurred miRNA target sites in viral protein coding sequences. They had suggested translation could modestly decrease miRNA regulation of those two naturally occurred miRNA targets. They had found the distribution of rare codons in the upstream region of those two miRNA targets is same as that of random sequences [4]. Our results are largely comparable to that observed in Lin *et al.* [4]. For naturally occurred miRNA targets in human protein coding sequences, we have observed reduced translation efficiency is selectively preferred near miRNA target region. But, translation efficiency in the region right upstream of miRNA target sites is not selectively varied. Instead, reduced translation efficiency has been observed in a nearby region located downstream of the miRNA target sites. A possible explanation is naturally occurred miRNA targets may use two separate regions to facilitate miRNA binding. Since the region right upstream of miRNA target sites has already been used to increase site accessibility, they may use the region downstream of the miRNA target region to slow down translational process near miRNA target sites. These two layers of selection may act together in the flank region of miRNA targets to ensure proper miRNA activity. It will be interesting to set up some experiments to validate this possibility.

Our results have suggested that the conservation level of target gene is a general factor that affects both site accessibility and translation efficiency in miRNA target region. We have found  $Z_{\Delta G_{open}}$  and  $Z_{tAI}$  values are smaller in miRNA targets located in vertebrate-conserved genes and mammal-specific genes (Figures 5 and 6). But, miRNA targets in primate-specific genes have no obvious signal of selection for site accessibility and translation efficiency (Figures 5 and 6). These are consistent with the observations made by previous works [1,17,18,51]. They had suggested purifying selection was smaller in younger miRNA targets than that in conserved miRNA targets.

We have found  $Z_{\Delta G_{open}}$  values of target regions in GC-rich genes are smaller than that in GC-poor genes (Figure 4). This finding can be explained by thermodynamic rules. Since GC pairs have three hydrogen bonds, GC-rich codons tend to fold in more stable RNA structures than AT-rich codons. Because loose RNA secondary structure is preferred in miRNA target region (Figure S1), GC-poor codons should be selectively preferred in miRNA target region to increase site accessibility, which is consistent with what we have observed in Figure 3. Assuming that selection targets the same higher site accessibility near all miRNA targets, we would expect that the increase of site accessibility is larger near miRNA targets located in GC-rich genes, simply because they start from a more-stable baseline. Whether selection actually targets the same higher site accessibility is not determined by our analysis. The

actual larger decrease of  $Z_{\Delta G_{open}}$  values for miRNA targets in GC-rich genes (Figure 4) implies this possibility.

In conclusion, we suggest site accessibility and translation efficiency have some effects on miRNA function when their targets are located in protein coding region of mammalian genes. Our results may help us better understand the process of miRNA action. It will also bring us important implications of protein coding sequence evolution in mammalian genome.

## Materials and Methods

### Data

Human PAR-CLIP data were downloaded from Supplementary Materials of Hafner *et al.* [14]. The PAR-CLIP data from Hafner *et al.* [14] contains genomic coordinates of DNA elements that were bound to human miRNAs in the experiment. Gene annotations of the human genome and mRNA sequences of all human transcripts were downloaded from Ensembl [52]. Overall, we extracted 4,344 miRNA targets that are located in human protein coding region.

To investigate the factors that may affect site accessibility and/or translation efficiency in the flank region of miRNA target sites, we considered the conservation level, the expression level and protein complex size of target gene. We downloaded a multiple alignment of 45 vertebrate genomes from UCSC genome server [53]. Using the 45-way alignment, we classified miRNA target genes into three groups (primate-specific genes, mammal-specific genes and vertebrate-conserved genes). We next downloaded expression data of human genes from Su *et al.* [54]. The expression level of each gene was quantified as the geometric mean of expression among different tissues. Finally, we downloaded protein complex data from the CORUM database [55]. For each gene, the complex size was measured as the number of proteins forming it and averaged across the number of complexes the gene participates.

### Site accessibility

Site accessibility represents the difficulty in opening a segment of mRNA sequences around miRNA target sites for their binding with RNA-induced silencing complex (RISCs) [15]. We used  $\Delta G_{open}$  to measure site accessibility for each miRNA target.  $\Delta G_{open}$  is the difference between free energy of the secondary structure of miRNA target region and free energy of the secondary structure in which miRNA target sites are unpaired [45].  $\Delta G_{open}$  has been proved to be a good indicator of site accessibility [44,45]. We used a window of 48 nucleotides, including 21 nucleotides directly bound by miRNAs, 17 flank upstream and 10 flank downstream nucleotides, to represent a miRNA target region. Previous studies have observed a significant correlation between  $\Delta G_{open}$  value of this region and miRNA activity [44,45]. We calculated  $\Delta G_{open}$  for each miRNA target using *RNAddG4* program in *PITA* package [45] with default parameter settings. At the same time, the folding energy of local RNA secondary structure,  $\Delta G_{local}$ , was also calculated for each miRNA target by *RNAddG4* [45]. In *RNAddG4* [45], *RNAfold* [56] was used to calculate the free energy of RNA secondary structures. As suggested in previous studies [45,57–60], we used a segment of mRNA sequences, rather than the full-length mRNAs, as the input sequences in calculating free energy of local RNA secondary structure. For each miRNA target region, the input mRNA segment includes 48 nucleotides in miRNA target region and additional 140 flank upstream and downstream nucleotides. This is based on the fact that the probability of base pairing when nucleotides are separated by

more than 140 nucleotides was low (data not shown) and it can substantially reduce computational complexity.

### Translation efficiency

We used tRNA adaptation index (*tAI*) to quantify translation efficiency in the flank region of mRNA target sites. *tAI* is a measure of codon adaptation to the tRNA abundance in genome [61]. *tAI* has been widely used to estimate translation efficiency in human [31,61,62]. We calculated *tAI* using codonR package [61], where tRNA copy numbers in the human genome were downloaded from the Genomic tRNA Database [63]. Since nine rare codons introduced in the upstream region of miRNA target sites are able to slow down local translational process and secure miRNA function [3], we calculated *tAI* in a window of nine consecutive codons in the flank region of miRNA target sites as a measure of local translation efficiency for each miRNA target.

### mRNA randomization

We hypothesized that if site accessibility and/or local translation efficiency is important for miRNA action when miRNA target sites are located in human protein coding sequences, synonymous codons should be selected for site accessibility and/or local translation efficiency in human. To detect the selection signal on synonymous codons, we used a randomization method to compute the statistical deviation of site accessibility and/or translation efficiency between the real mRNA sequence and permuted sequences [29,47–49]. We randomly shuffled synonymous codons among sites for each mRNA transcript, while keeping the encoded protein sequences, gene's codon usage bias and gene's GC composition the same. We did not shuffle the codons that are directly targeted by miRNAs during mRNA randomization, since those nucleotides are crucial for proper miRNA function. We generated 1,000 such permuted mRNA sequences for all mRNA transcripts with miRNA target sites in their protein-coding region.

### Z-score calculation

We used Z-score to determine the deviation of the real sequence from randomized sequences and estimate its statistical significance. For each miRNA target region, we calculated site accessibility in the real mRNA transcript ( $\Delta G_{open}$ )<sub>N</sub> and each permuted mRNA sequence ( $\Delta G_{open}$ )<sub>P</sub>. Then, we calculated Z-score of site accessibility ( $Z_{\Delta G_{open}}$ ) for each miRNA target region as formula 1.

$$Z_{\Delta G_{open}} = \frac{(\Delta G_{open})_N - \overline{(\Delta G_{open})_P}}{\sqrt{\sum_{i=1}^n \frac{((\Delta G_{open})_{P_i} - \overline{(\Delta G_{open})_P})^2}{n-1}}} \quad (1)$$

Here,  $(\Delta G_{open})_N$  is site accessibility for the naturally occurring target region under consideration.  $(\Delta G_{open})_{P_i}$  is site accessibility for the target region in  $i^{th}$  permuted sequence, and  $\overline{(\Delta G_{open})_P}$  is the mean of  $(\Delta G_{open})_{P_i}$  over all permuted sequences. The variable  $n$  represents the total number of permuted sequences, which is equal to 1,000 in our analysis.

Similarly, we calculated Z-score of local secondary structure ( $Z_{\Delta G_{local}}$ ), Z-score of local GC content ( $Z_{GC}$ ) and Z-score of local translation efficiency ( $Z_{tAI}$ ) in the flank region of each miRNA target as formulas below.

$$Z_{\Delta G_{local}} = \frac{(\Delta G_{local})_N - \overline{(\Delta G_{local})_P}}{\sqrt{\sum_{i=1}^n \frac{((\Delta G_{local})_{P_i} - \overline{(\Delta G_{local})_P})^2}{n-1}}} \quad (2)$$

$$Z_{GC} = \frac{GC_N - \overline{GC_P}}{\sqrt{\sum_{i=1}^n \frac{(GC_{P_i} - \overline{GC_P})^2}{n-1}}} \quad (3)$$

$$Z_{tAI} = \frac{tAI_N - \overline{tAI_P}}{\sqrt{\sum_{i=1}^n \frac{(tAI_{P_i} - \overline{tAI_P})^2}{n-1}}} \quad (4)$$

The definitions for  $(\Delta G_{local})_N$ ,  $GC_N$ , and  $tAI_N$  are analogous to  $(\Delta G_{open})_N$ , but refer to free energy of local RNA secondary structure, GC content and translation efficiency rather than to site accessibility. Similarly, the definitions for  $(\Delta G_{local})_{P_i}$ ,  $GC_{P_i}$ , and  $tAI_{P_i}$  are analogous to  $(\Delta G_{open})_{P_i}$ , and the definitions for  $\overline{(\Delta G_{local})_P}$ ,  $\overline{GC_P}$ , and  $\overline{tAI_P}$  are analogous to  $\overline{(\Delta G_{open})_P}$ .

## Supporting Information

**Figure S1** The mean and standard error of  $Z_{\Delta G_{local}}$  of each sliding window near miRNA target region in the human genome.

(TIFF)

**Figure S2**  $Z_{\Delta G_{open}}$  in miRNA target region as a function of  $Z_{\Delta G_{local}}$  in that region. Each point represents a miRNA target in human protein coding sequences.

(TIFF)

**Figure S3**  $Z_{\Delta G_{local}}$  in miRNA target region as a function of  $Z_{GC}$  in that region. Each point represents a miRNA target in human protein coding sequences.

(TIFF)

## References

- Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136: 215–233.
- Brodersen P, Voinnet O (2009) Revisiting the principles of microRNA target recognition and mode of action. *Nat Rev Mol Cell Biol* 10: 141–148.
- Gu S, Jin L, Zhang F, Sarnow P, Kay MA (2009) Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nat Struct Mol Biol* 16: 144–150.
- Lin HR, Ganem D (2011) Viral microRNA target allows insight into the role of translation in governing microRNA target accessibility. *Proc Natl Acad Sci U S A* 108: 5148–5153.
- Fang Z, Rajewsky N (2011) The impact of miRNA target sites in coding sequences and in 3'UTRs. *PLoS One* 6: e18067.
- Forman JJ, Collier HA (2010) The code within the code: microRNAs target coding regions. *Cell Cycle* 9: 1533–1541.
- Forman JJ, Legesse-Miller A, Collier HA (2008) A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc Natl Acad Sci U S A* 105: 14879–14884.
- Reczko M, Maragkakis M, Alexiou P, Grosse I, Hatzigeorgiou AG (2012) Functional microRNA targets in protein coding sequences. *Bioinformatics* 28: 771–776.
- Rigoutsos I (2009) New tricks for animal microRNAs: targeting of amino acid coding regions at conserved and nonconserved sites. *Cancer Res* 69: 3245–3248.
- Schnall-Levin M, Rissland OS, Johnston WK, Perrimon N, Bartel DP, et al. (2011) Unusually effective microRNA targeting within repeat-rich coding regions of mammalian mRNAs. *Genome Res* 21: 1395–1403.
- Schnall-Levin M, Zhao Y, Perrimon N, Berger B (2010) Conserved microRNA targeting in Drosophila is as widespread in coding regions as in 3'UTRs. *Proc Natl Acad Sci U S A* 107: 15751–15756.
- Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I (2008) MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* 455: 1124–1128.
- Chi SW, Zang JB, Mele A, Darnell RB (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 460: 479–486.
- Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, et al. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141: 129–141.
- Axtell MJ, Westholm JO, Lai EC (2011) Vive la difference: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol* 12: 221.
- Axtell MJ, Bowman JL (2008) Evolution of plant microRNAs and their targets. *Trends Plant Sci* 13: 343–349.
- Fahlgren N, Jogdeo S, Kasschau KD, Sullivan CM, Chapman EJ, et al. (2010) MicroRNA gene evolution in Arabidopsis lyrata and Arabidopsis thaliana. *Plant Cell* 22: 1074–1089.
- Cuperus JT, Fahlgren N, Carrington JC (2011) Evolution and functional diversification of MIRNA genes. *Plant Cell* 23: 431–442.
- Saunders MA, Liang H, Li WH (2007) Human polymorphism at microRNAs and microRNA target sites. *Proc Natl Acad Sci U S A* 104: 3300–3305.
- Itzkovitz S, Hodis E, Segal E (2010) Overlapping codes within protein-coding sequences. *Genome Res* 20: 1582–1589.
- Vinogradov AE (2003) DNA helix: the importance of being GC-rich. *Nucleic Acids Res* 31: 1838–1844.

**Figure S4** Comparison of the mean  $Z_{tAI}$  between miRNA targets in genes with the highest 5% and lowest 5% GC content.

(TIFF)

**Figure S5** Comparison of the mean  $Z_{\Delta G_{open}}$  between miRNA targets in genes with the highest 5% and lowest 5% protein complex size.

(TIFF)

**Figure S6** Comparison of the mean  $Z_{\Delta G_{open}}$  between miRNA targets in genes with the highest 5% and lowest 5% ENC.

(TIFF)

**Figure S7** Comparison of the mean  $Z_{tAI}$  between miRNA targets in genes with the highest 5% and lowest 5% ENC.

(TIFF)

**Figure S8** Comparison of the mean  $Z_{\Delta G_{open}}$  between miRNA targets in genes with the highest 5% and lowest 5% expression level.

(TIFF)

**Figure S9** Comparison of the mean  $Z_{tAI}$  between miRNA targets in genes with the highest 5% and lowest 5% expression level.

(TIFF)

**Figure S10**  $Z_{\Delta G_{open}}$  in miRNA target region as a function of  $Z_{tAI}$  in the window that is nine codons downstream of miRNA target sites. Each point represents a miRNA target in human protein coding sequences.

(TIFF)

**Figure S11** The mean and standard error of  $Z_{tAI}$  of each sliding window near randomized miRNA target region.

(TIFF)

## Author Contributions

Conceived and designed the experiments: WG XX TZ. Performed the experiments: WG XW CZ XX. Analyzed the data: WG XX TZ. Contributed reagents/materials/analysis tools: XW XX. Wrote the paper: WG XX TZ.

22. Warnecke T, Batada NN, Hurst LD (2008) The impact of the nucleosome code on protein-coding sequence evolution in yeast. *PLoS Genet* 4: e1000250.
23. Chamary JV, Hurst LD (2005) Biased codon usage near intron-exon junctions: selection on splicing enhancers, splice-site recognition or something else? *Trends Genet* 21: 256–259.
24. Chamary JV, Parmley JL, Hurst LD (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet* 7: 98–108.
25. Stoletzki N (2008) Conflicting selection pressures on synonymous codon use in yeast suggest selection on mRNA secondary structures. *BMC Evol Biol* 8: 224.
26. Parmley JL, Chamary JV, Hurst LD (2006) Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. *Mol Biol Evol* 23: 301–309.
27. Warnecke T, Hurst LD (2007) Evidence for a trade-off between translational efficiency and splicing regulation in determining synonymous codon usage in *Drosophila melanogaster*. *Mol Biol Evol* 24: 2755–2762.
28. Wilke CO (2011) Transcriptional robustness complements nonsense-mediated decay in humans. *PLoS Genet* 7: e1002296.
29. Gu W, Zhou T, Wilke CO (2010) A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput Biol* 6: e1000664.
30. Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324: 255–258.
31. Tuller T, Waldman YY, Kupiec M, Ruppin E (2010) Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci U S A* 107: 3645–3650.
32. Akashi H, Eyre-Walker A (1998) Translational selection and molecular evolution. *Curr Opin Genet Dev* 8: 688–693.
33. Duret L (2002) Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev* 12: 640–649.
34. Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 2: 13–34.
35. Morton BR, Wright SI (2007) Selective constraints on codon usage of nuclear genes from *Arabidopsis thaliana*. *Mol Biol Evol* 24: 122–129.
36. Zhou T, Weems M, Wilke CO (2009) Translationally optimal codons associate with structurally sensitive sites in proteins. *Mol Biol Evol* 26: 1571–1580.
37. Komar AA, Lesnik T, Reiss C (1999) Synonymous codon substitutions affect ribosome traffic and protein folding during *in vitro* translation. *FEBS Lett* 462: 387–391.
38. Thanaraj TA, Argos P (1996) Ribosome-mediated translational pause and protein domain organization. *Protein Sci* 5: 1594–1612.
39. Hurst LD (2006) Preliminary assessment of the impact of microRNA-mediated regulation on coding sequence evolution in mammals. *J Mol Evol* 63: 174–182.
40. Hurst LD (2011) Molecular genetics: The sound of silence. *Nature* 471: 582–583.
41. Lin MF, Kheradpour P, Washietl S, Parker BJ, Pedersen JS, et al. (2011) Locating protein-coding sequences under selection for additional, overlapping functions in 29 mammalian genomes. *Genome Res* 21: 1916–1928.
42. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, et al. (2011) A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478: 476–482.
43. Brest P, Lapaquette P, Souidi M, Lebrigand K, Cesaro A, et al. (2011) A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat Genet* 43: 242–245.
44. Hausser J, Landthaler M, Jaskiewicz L, Gaidatzis D, Zavolan M (2009) Relative contribution of sequence and structure features to the mRNA binding of Argonaute/EIF2C-miRNA complexes and the degradation of miRNA targets. *Genome Res* 19: 2009–2020.
45. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E (2007) The role of site accessibility in microRNA target recognition. *Nat Genet* 39: 1278–1284.
46. Wright F (1990) The 'effective number of codons' used in a gene. *Gene* 87: 23–29.
47. Gu W, Wang X, Zhai C, Xie X, Zhou T (2012) Selection on Synonymous Sites for Increased Accessibility around miRNA Binding Sites in Plants. *Molecular Biology and Evolution* 29: 3037–3044.
48. Gu W, Zhai C, Wang X, Xie X, Parinandi G, et al. (2012) Translation Efficiency in Upstream Region of microRNA Targets in *Arabidopsis thaliana*. *Evolutionary Bioinformatics* 8: 565.
49. Chamary J, Hurst L (2005) Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biology* 6: R75.
50. Seffens W, Digby D (1999) mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences. *Nucleic Acids Res* 27: 1578–1584.
51. Ehrenreich IM, Purugganan M (2008) MicroRNAs in plants: Possible contributions to phenotypic diversity. *Plant Signal Behav* 3: 829–830.
52. Kersey PJ, Lawson D, Birney E, Derwent PS, Haimel M, et al. (2010) Ensembl Genomes: extending Ensembl across the taxonomic space. *Nucleic Acids Res* 38: D563–569.
53. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, et al. (2011) The UCSC Genome Browser database: update 2011. *Nucleic Acids Research* 39: D876–D882.
54. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, et al. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America* 101: 6062–6067.
55. Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, et al. (2010) CORUM: the comprehensive resource of mammalian protein complexes–2009. *Nucleic Acids Res* 38: D497–501.
56. Hofacker IL (2003) Vienna RNA secondary structure server. *Nucleic Acids Res* 31: 3429–3431.
57. Lu M, Zhang Q, Deng M, Miao J, Guo Y, et al. (2008) An Analysis of Human MicroRNA and Disease Associations. *PLoS One* 3: e3420.
58. Richter AS, Schleberger C, Backofen R, Steglich C (2010) Seed-based INTARNA prediction combined with GFP-reporter system identifies mRNA targets of the small RNA Yfr1. *Bioinformatics* 26: 1–5.
59. Ying X, Cao Y, Wu J, Liu Q, Cha L, et al. (2011) sTarPicker: a method for efficient prediction of bacterial sRNA targets based on a two-step model for hybridization. *PLoS One* 6: e22705.
60. Zhao Y, Samal E, Srivastava D (2005) Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* 436: 214–220.
61. dos Reis M, Savva R, Wernisch L (2004) Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res* 32: 5036–5044.
62. Waldman YY, Tuller T, Shlomi T, Sharan R, Ruppin E (2010) Translation efficiency in humans: tissue specificity, global optimization and differences between developmental stages. *Nucleic Acids Research* 38: 2964–2974.
63. Chan PP, Lowe TM (2009) GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* 37: D93–97.