

# Genomewide Analysis of Expression of Genes Located near miRNAs in Humans and Mice

Hidenori Inaoka\*, Yutaka Fukuoka<sup>1,\*\*</sup> and Makoto Noshiro

School of Allied Health Sciences, Kitasato University,  
1-15-1, Kitasato, Minami, Sagami-hara, Kanagawa 252-0373, Japan

<sup>1</sup>Graduate School of Biomedical Science, Tokyo Medical and Dental University,  
Yushima, Bunkyo, Tokyo, 113-8510, Japan

(Received October 3, 2011; accepted January 24, 2012)

**Key words:** MicroRNA (miRNA), transcription

MicroRNAs (miRNAs) are 20- to 22-nucleotide RNAs that regulate the transcription and translation of genes. In our previous studies, we reported that a miRNA has the potential to affect the transcription of genes located near it. The number of miRNAs recorded in databases continues to increase. However, the effects of the miRNAs, including newly found miRNAs, on the transcription of genes located near them have not yet been analyzed. In this study, we investigated the effects in humans and mice. The average expression levels are increased in the 10<sup>4</sup> bp vicinity of miRNA in all the datasets investigated. Many of the newly found miRNAs are located in an intron, exon or untranslated region (UTR) of host genes. When a host gene is transcribed, the miRNA located in such a region of the host gene is also transcribed. In addition, recent reports indicate that some noncoding RNAs, including miRNAs, can upregulate the expression of genes that have the complement sequences of the noncoding RNA in their promoter regions. Therefore, if genes having complement sequences of miRNAs are located around these miRNAs, the expression levels of these genes could be increased. These phenomena may be the reasons for the increased average expression levels around miRNAs.

## 1. Introduction

MicroRNAs (miRNAs) are 20- to 22-nucleotide RNAs that regulate the transcription and translation of genes.<sup>(1)</sup> There is mounting evidence that small noncoding RNAs including miRNAs play an important role in the control of the dynamics of localized gene expression through heterochromatin formation.<sup>(2)</sup> Therefore, we have hypothesized

---

\*Corresponding author: e-mail: inaoka@kitasato-u.ac.jp

\*\*Present affiliation: Department of Electrical Engineering, Kogakuin University, Tokyo, Japan

that miRNAs may play a role in the localized expression of genes in their neighborhood. We defined a series of increasingly large chromosomal windows centered on each miRNA location. The expression levels within each window were analyzed in our previous studies. In those studies, we reported that messenger RNA expression levels obtained from genes localized within a  $10^4$  bp chromosomal distance of miRNAs were decreased in worms (103 miRNAs) and mice (228 miRNAs).<sup>(3,4)</sup> These results suggest that a miRNA has the potential to affect the transcription of genes located near it. The number of miRNAs found by experimental methods or computational estimations continues to increase. Currently, over 700 and 1,400 miRNAs are known in mouse and human, respectively. However, the effects of the miRNAs, including newly found ones, on the transcription of genes located near them have not yet been analyzed. In this study, we investigated the effects in humans and mice.

## 2. Materials and Methods

### 2.1 Materials

The positions of miRNAs in the human and mouse genomes were obtained from the microRNA database (miRBase) Release 17.<sup>(5)</sup> The positions of genes in the human and mouse genomes were obtained from the National Center for Biotechnology Information (NCBI) Build 37.2. Gene expression datasets were downloaded from NCBI Gene Expression omnibus (GEO). As human data, GEO accession numbers GSE1643, GSE3141, GSE3526, GSE4655, GSE6565, GSE8586, and GSE9520 were used.<sup>(6-12)</sup> Each of these datasets was measured using Affymetrix Human Genome U133 Plus 2.0 Array. As mouse data, GSE1049, GSE1052, GSE1301, GSE1745, GSE3234, GSE3351, and GSE26875 were used.<sup>(13-16)</sup> Each dataset was obtained using Affymetrix Mouse Expression 430B Array.

### 2.2 Methods

Gene expression values in each dataset were normalized using the average value of the whole data in the dataset. A window centered at a miRNA was set, and the average of the normalized expression levels was calculated for that window. Fifteen different window widths were used:  $10^4$  bp,  $2 \times 10^4$  bp,  $4 \times 10^4$  bp,  $10^5$  bp,  $2 \times 10^5$  bp,  $4 \times 10^5$  bp,  $10^6$  bp,  $2 \times 10^6$  bp,  $4 \times 10^6$  bp,  $10^7$  bp,  $2 \times 10^7$  bp,  $4 \times 10^7$  bp,  $10^8$  bp,  $2 \times 10^8$  bp, and  $4 \times 10^8$  bp. After this calculation, the data for all the miRNAs were merged, and the relationship between the average expression level and window width was investigated. The reason for the size of the minimum window is as follows. The average numbers of genes per  $10^4$  bp in human and mouse are 4.4 and 4.8, respectively. If the size of the minimum window is shorter than  $10^4$  bp, the number of genes in the windows will be less than 1. Therefore, in this study, the size of the minimum windows was determined to be  $10^4$  bp.

## 3. Results

The average expression level and standard error in each window in a human dataset, GSE1643, are plotted as a function of the window width (Fig. 1). Figure 2 indicates the average expression levels in all the human data. The average expression level and

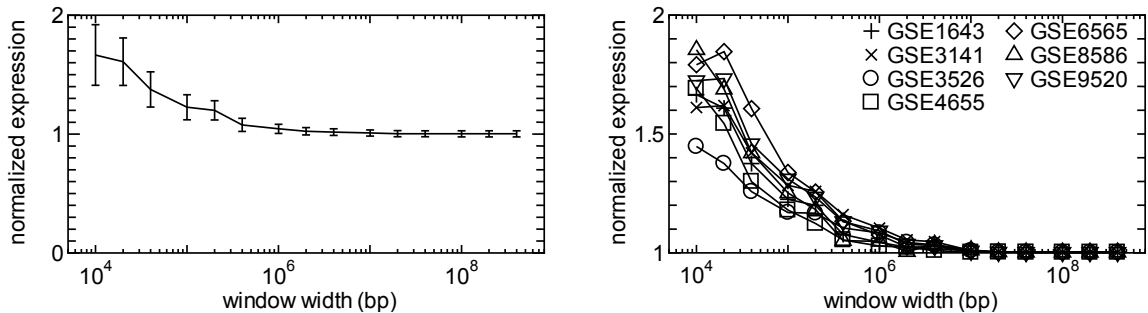


Fig. 1 (left). The average expression level of GSE1643 is plotted as a function of chromosomal distance to the miRNA in base pairs.

Fig. 2 (right). The average expression levels of all human datasets are plotted as a function of chromosomal distance to the miRNA in base pairs.

standard error in each window in a mouse dataset, GSE1049, are plotted as a function of the window width in Fig. 3. For all the mouse data, the average expression level in each window is plotted in Fig. 4. In both humans and mice, the average expression levels are increased in the  $10^4$  bp vicinity of miRNA in all the datasets. The average expression values in the  $10^4$  and  $10^6$  bp were analyzed for statistical significance using Student's t-test. A  $p$ -value  $< 0.05$  was considered statistically significant. All the datasets except GSE1301 and GSE3351 showed that the average expression level was significantly higher in  $10^4$  bp than in  $10^6$  bp.

#### 4. Discussion

The average expression levels of all the datasets investigated in this study are increased in the  $10^4$  bp vicinity of miRNA in both humans and mice. However, our previous results showed that the average expression levels are decreased in the  $10^4$  bp vicinity of miRNA in worms<sup>(3)</sup> and mice.<sup>(4)</sup> The new analysis yields results that are different those of previous analyses. In what follows, we will discuss the possible reasons for the differences.

##### 4.1 Worm analysis

The average expression levels around miRNAs were also analysed in worms using the information obtained from miRBase Release 17. Although the number of miRNAs in worms was increased from 103 to 207 in the new analysis, the result showed the same tendency as the previous one.<sup>(3)</sup> Figure 5 illustrates the average expression level as a function of the window width in each group in two worm datasets (growth and developmental,<sup>(17)</sup> germline<sup>(18)</sup>). The average expression levels are decreased in the  $10^4$

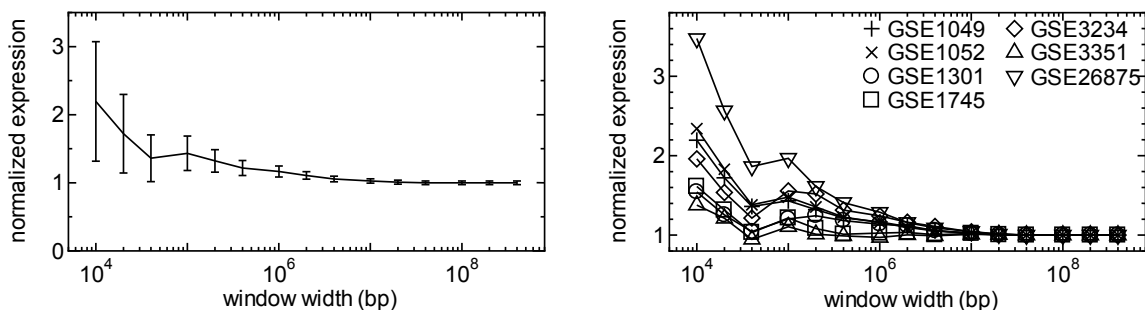


Fig. 3 (left). The average expression level of GSE1049 is plotted as a function of chromosomal distance to the miRNA in base pairs.

Fig. 4 (right). The average expression levels of all mouse datasets are plotted as a function of chromosomal distance to the miRNA in base pairs.

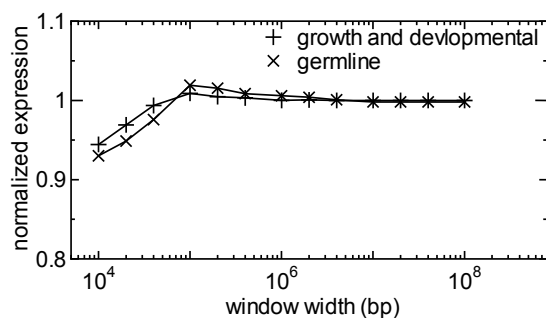


Fig. 5. The average expression level of all worm datasets are plotted as a function of chromosomal distance to the miRNA in base pairs.

bp vicinity of miRNAs. These findings indicate that the average expression levels of humans and mice tend to be different from those of worms. The gene expression levels in the former species increased in the  $10^4$  bp vicinity of miRNAs.

Many miRNAs and their targets are conserved among species. It is considered that the expression levels around the conserved miRNAs show the same tendency. Using the family gene information obtained from miRBase Release 17, the numbers of miRNAs conserved among worm, mouse, and human were counted. Only 5 out of 207 worm miRNAs were conserved in mouse and human miRNAs. These conserved miRNAs

were let-7, mir-9, mir-124, mir-1, and mir-34. These miRNAs are conserved among over 80 species. This finding indicates that most of the worm miRNAs are different from mouse and human miRNAs. It is considered that the target genes of the worm miRNAs are also different from those of mouse and human miRNAs. Therefore, the expression levels around miRNAs showed a different tendency.

#### 4.2 *Mouse analyses*

As described above, the results obtained from new analysis of mice are different from those of previous analysis. The different results could be partly explained by the different numbers of miRNAs used in the two analyses. The number of mouse miRNAs obtained from miRBase Release 17 was 715 in this study, whereas that in the previous study was 228 (miRBase Release 7.2). Therefore, the number of mouse miRNAs used in this study was over three times higher than that in the previous study. The number of miRNAs that have genes in their  $10^4$  bp vicinity was increased from 62 in Rel. 7.2 to 420 in Rel. 17. A recent change in the miRNA detection approach has considerably increased the number of miRNAs detected. In previous miRNA detection approaches, putative miRNA sequences that were conserved in other species (e.g., worm) were searched. However, in the current miRNA detection approach, the entire genome was scanned for hairpin structures, and the thermodynamic stability and structural features of the hairpins were calculated using a computational method and then putative miRNAs were determined.<sup>(19)</sup> Therefore, the types of miRNAs newly added in this study have the potential to be functionally different from those of miRNAs used in the previous study. As the numbers of different types of miRNAs were increased in the new analysis, the effects of those miRNAs on the expressions of the genes located around them may differ from those in the previous studies.

Another reason for the different results in mice could be the different information on the chromosomal positions of genes in the two analyses. The chromosomal positions of genes were obtained from NCBI Build 37.2 in this study, whereas NCBI Build 34.1 was used in the previous study. The number of genes whose chromosomal positions were determined was higher in B37.2 than in B34.1. The numbers of genes located in the  $10^4$  bp vicinity of miRNAs were 442 in B37.2 and 50 in B34.1. Therefore, the number of genes located around miRNAs was considerably increased in this study.

#### 4.3 *Effects of miRNA location*

Some miRNAs are located in an intron, exon or untranslated region (UTR) of host genes and some miRNAs are located between genes (intergenes). Using miRBase information, miRNAs were clustered into five groups: intronic, exonic, 3'-UTR, 5'-UTR, and intergenic. The number of miRNAs in each group is shown in Table 1. The average expression level in each miRNA group in a human dataset, GSE1643, is plotted as a function of the window width in Fig. 6. Figure 7 illustrates the average expression level as a function of the window width in each group in a mouse dataset, GSE1049.

In GSE1643, all the miRNA groups except the intronic group showed increasing average expression levels near miRNAs. In the intronic group, there was no relationship between the average expression level and window width. In the other human datasets,

Table 1  
Number of miRNAs in each group.

	Intronic	Exonic	3'-UTR	5'-UTR	Intergenic
Human	847	169	66	20	484
Mouse	392	140	34	8	226

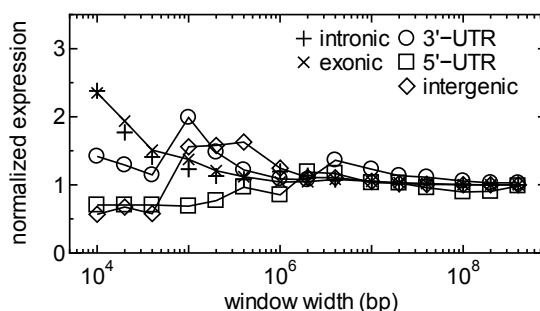
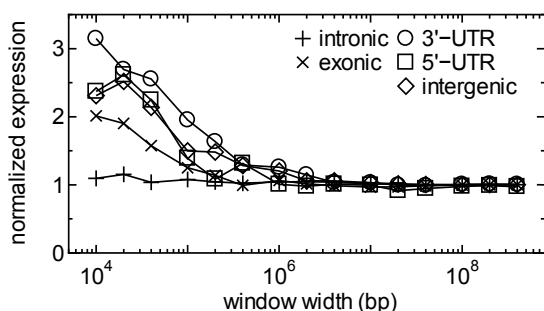


Fig. 6 (left). The average expression levels obtained from miRNA groups (intronic, exonic, 3'-UTR, 5'-UTR, and intergenic) in GSE1643 are plotted as a function of chromosomal distance to the miRNA in base pairs.

Fig. 7 (right). The average expression levels obtained from miRNA groups (intronic, exonic, 3'-UTR, 5'-UTR, and intergenic) in GSE1049 are plotted as a function of chromosomal distance to the miRNA in base pairs.

all the miRNA groups showed increasing average expression levels near miRNAs. In mouse data GSE1049, in contrast, the 5'-UTR and intergenic groups showed decreasing average expression levels near miRNAs, whereas the other groups showed increasing average expression levels near miRNAs. In the 3'-UTR group, the average expression level tends to increase in the  $10^5$  to  $10^6$  bp width and decrease in the  $10^4$  bp width. In addition, the decreasing average expression levels near miRNAs were found in the intergenic group in all the other mouse datasets. Many mouse datasets showed that the average expression levels were decreased in the  $10^4$  bp width in the 5'-UTR group and the average expression levels were increased from  $10^5$  to  $10^6$  bp in the 3'-UTR group. These findings indicate that the average expression levels near miRNAs are generally increased, whereas those of some miRNA groups are decreased in the mouse datasets.

As described above, the average expression levels near intergenic miRNAs were decreased in the  $10^4$  bp width, whereas the average expression levels near intronic/exonic miRNAs were increased in mice. Therefore, the reason for the decreasing average expression levels near miRNAs in the previous study could be partly explained on the basis of the ratio of the number of miRNAs in each group. Table 2 compares

Table 2

Differences in the number of miRNAs in each group between miRBase Release 7.2 and Release 17.

	Intronic	Exonic	3'-UTR	5'-UTR	Intergenic
Rel. 7.2	94	54	5	0	77
Rel. 17	392	140	34	8	226

the number of miRNAs in each group obtained from miRBase Release 7.2 and Release 17. As shown in Table 2, the number of intergenic groups was relatively high in Rel. 7.2. Therefore, there could be a stronger tendency of the average expression levels to decrease in the  $10^4$  bp vicinity of miRNAs in the previous analysis.

From the results described above, it is suggested that the average expression levels of the intronic and exonic groups strongly affect the overall trend of the average expression levels. The reason for the increasing average expression levels near miRNAs in the intronic and exonic groups could be explained as follows. When a host gene is transcribed, a miRNA located in an intron or exon of the host gene is also transcribed. In other words, the expression of miRNAs located in the host genes is linked to the expression of the host genes.<sup>(20,21)</sup>

Recent reports indicate that some noncoding RNAs, including miRNAs, can upregulate the expression of genes that have the complement sequences of the noncoding RNA in their promoter regions, although the exact mechanism is still unknown.<sup>(22,23)</sup> Therefore, if genes having complement sequences of miRNAs are located around these miRNAs, the expression levels of these genes could be increased because of the reason described above. This phenomenon may be one of the reasons for the increased average expression levels.

## 5. Conclusion

The results of this study indicate that the average expression levels near miRNAs obtained from mammals (humans and mice) and worms are different. These suggest that there exist some classes of miRNAs, i.e., those conserved over various species and those that evolved in higher organisms. The different tendencies in mammals and worms imply that different classes of miRNAs have different functionalities. In the future, the relationship between the chromosomal location of the miRNA and the function of the target gene will be investigated using gene functional analysis.

## Acknowledgements

This study was financially supported in part by the regular budget of Kitasato University School of Allied Health Sciences, and a scientific research grant from the Ministry of Education, Culture, Sports, Science and Technology of Japan (#20510184) to Yutaka Fukuoka.

## References

- 1 R. Liang, D. J. Bates and E. Wang: *Curr. Genomics* **10** (2009) 184.
- 2 E. Bernstein and C. D. Allis: *Genes Dev.* **19** (2005) 1635.
- 3 H. Inaoka, Y. Fukuoka and I. S. Kohane: *BMC Bioinf.* **7** (2006) 112.
- 4 H. Inaoka, Y. Fukuoka and I. S. Kohane: *Proc. Natl. Acad. Sci. USA* **104** (2007) 5020.
- 5 A. Kozomara and S. Griffiths-Jones: *Nucleic Acids Res.* **39** (2011) D152.
- 6 M. P. Gruber, C. D. Coldren, M. D. Woolum, G. P. Cosgrove, C. Zeng, A. E. Baron, M. D. Moore, C. D. Cool, G. S. Worthen, K. K. Brown and M. W. Geraci: *Am. J. Respir. Cell Mol. Biol.* **35** (2006) 65.
- 7 A. H. Bild, G. Yao, J. T. Chang, Q. Wang, A. Potti, D. Chasse, M. B. Joshi, D. Harpole, J. M. Lancaster, A. Berchuck, J. A. Olson Jr., J. R. Marks, H. K. Dressman, M. West and J. R. Nevins: *Nature* **439** (2006) 353.
- 8 R. B. Roth, P. Hevezi, J. Lee, D. Willhite, S. M. Lechner, A. C. Foster and A. Zlotnik: *Neurogenetics* **7** (2006) 67.
- 9 M. A. Keller, S. Addya, R. Vadigepalli, B. Banini, K. Delgrosso, H. Huang and S. Surrey: *Physiol. Genomics* **28** (2006) 114.
- 10 V. A. Funari, A. Day, D. Krakow, Z. A. Cohn, Z. Chen, S. F. Nelson and D. H. Cohn: *BMC Genomics* **8** (2007) 165.
- 11 J. Cohen, L. J. Van Marter, Y. Sun, E. Allred, A. Leviton and I. S. Kohane: *Genome Biol.* **8** (2007) R210.
- 12 B. L. Larson, J. Ylostalo and D. J. Prockop: *Stem Cells* **26** (2008) 193.
- 13 J. D. Porter, S. Israel, B. Gong, A. P. Merriam, J. Feuerman, S. Khanna and H. J. Kaminski: *Physiol. Genomics* **24** (2006) 264.
- 14 F. Zeng, D. A. Baldwin and R. M. Schultz: *Dev. Biol.* **272** (2004) 483.
- 15 H. Pan, M. M. O'Brien, K. Wigglesworth, J. J. Eppig and R. M. Schultz: *Dev. Biol.* **286** (2005) 493.
- 16 S. Yasuda, T. Hasegawa, T. Hosono, M. Satoh, K. Watanabe, K. Ono, S. Shimizu, T. Hayakawa, T. Yamaguchi, K. Suzuki and Y. Sato: *Biochem. J.* **437** (2011) 345.
- 17 S. K. Kim, J. Lund, M. Kiraly, K. Duke, M. Jiang, J. M. Stuart, A. Eizinger, B. N. Wylie and G. S. Davidson: *Science* **293** (2001) 2087.
- 18 V. Reinke, I. S. Gil, S. Ward and K. Kazmer: *Development* **131** (2004) 311.
- 19 I. Bentwich, A. Avniel, Y. Karov, R. Aharonov, S. Gilad, O. Barad, A. Barzilai, P. Einat, U. Einav, E. Meiri, E. Sharon, Y. Spector and Z. Bentwich: *Nat. Genet.* **37** (2005) 766.
- 20 A. Rodriguez, S. Griffiths-Jones, J. L. Ashurst and A. Bradley: *Genome Res.* **14** (2004) 1902.
- 21 S. Baskerville and D. P. Bartel: *RNA* **11** (2005) 241.
- 22 R. F. Place, L. Li, D. Pookot, E. J. Noonan and R. Dahiya: *Proc. Natl. Acad. Sci. USA* **105** (2008) 1608.
- 23 S. Majid, A. A. Dar, S. Saini, S. Yamaura, H. Hirata, Y. Tanaka, G. Deng and R. Dahiya: *Cancer* **116** (2010) 5637.