

# Database of CpG-island-associated microRNA genes

## 與 CpG 島關聯之微型核糖核酸基因資料庫

Ming-Cheng Tsai  
ScinoPharm Taiwan Ltd., Taiwan  
tsaimc0527@gmail.com

Chien-Hung Huang  
Department of Computer Science and  
Information Engineering, National  
Formosa University, Taiwan  
chhuang@sunws.nfu.edu.tw

Ka-Lok Ng\*

Department of Bioinformatics, Asia University, Taichung, Taiwan  
ppiddi@gmail.com

### Abstract

A rather high portion of mammalian miRNA genes are found to associate with CpG-islands. It is found that 74 of 462 (16.0%), 37 of 373 (9.9%), and 25 of 234 (10.7%) of the miRNA genes are located within 1000 bps of a CpG island for human, mouse, and rat respectively. These CpG-associated miRNA genes could be located near promoter regions as well. This finding suggests a potential role of these miRNAs in CpG-island methylation, and gene regulation.

Keywords: microRNA, promoter regions, CpG islands, methylation, gene regulation

### 1. Introduction

MicroRNA (miRNA) genes encode small RNA molecules involved in mRNA translation and degradation by the RNA interference (siRNA) machinery. MicroRNAs (miRNAs) are an evolutionarily conserved large class of noncoding RNAs (ncRNAs) 18–22 nucleotides long that mediate posttranscriptional silencing of genes. MiRNAs are first discovered in *Caenorhabditis elegans* with the identification of the *lin-4* and *let-7* miRNA genes, which act as posttranscriptional repressors of target

genes by antisense binding to their 3' untranslated regions (UTRs; for review, see [1]). Shortly thereafter, hundreds of other miRNAs were found in worms as well as in flies, plants, and vertebrates (for review, see [2, 3, 4]). Rapid progress has begun to unravel the genetic roles of miRNAs in development and other biological processes. For example, in *C. elegans*, *let-7* and *lin-4* miRNAs function as heterochronic genes, and mutations in either disrupt proper specification of cell fates [1, 2]. In *Drosophila*, a mutation in *miR-14* leads to a disruption in normal patterns of cell death and also defects in fat metabolism. In mammals, ~230 miRNAs have been identified from a vast array of tissues and cell types [12, 13, 14].

CpG islands are associated with genes, particularly housekeeping genes, in vertebrates. CpG islands are typically common near transcription start sites (TSS), often associated with promoter regions, and some of which are hypermethylated in cancer. It is suggested that the effect of a CpG island near the TSS could be more important than the global GC content of the region where the gene resides [16, 17].

\* corresponding author

Normally a C (cytosine) base followed immediately by a G (guanine) base (a CpG) is rare in vertebrate DNA because the Cs in such an arrangement tend to be methylated. This methylation helps distinguish the newly synthesized DNA strand from the parent strand, which aids in the final stages of DNA proofreading after duplication. However, over evolutionary time methylated Cs tends to turn into Ts because of spontaneous deamination. The result is that CpGs are relatively rare unless there is selective pressure to keep them or a region is not methylated for some reason, perhaps having to do with the regulation of gene expression. CpG islands are regions where CpGs are present at significantly higher levels than is typical for the genome as a whole [7].

According to these information, we used the UCSC Genome Browser [9, 11] to find whether mammalian (human, mouse and rat) miRNA genes are associated with CpG-islands. Afterward, through simple sequence homology searched using NCBI-BLAST, finding whether human miRNA genes are located within promoter regions (promoter data are obtained from DBTSS). We have found a rather significant proportion (~10%) of miRNA genes located within 1000 bp of a CpG-island. Although CpG-islands are frequently associated with the 5'-region of known genes, 8000 of the 28000 CpG islands in the genome are located in-between genes. A subset of miRNA genes are associated with CpG-islands and promoter regions, suggesting a potential role of these miRNAs in CpG-island methylation and gene regulation. Since siRNAs [10] may be involved in inducing methylation of target CpG-islands, it would be logical to suspect that miRNAs close to or overlapping with CpG-islands could be involved in autoregulation of the methylation of their respective CpG-islands. This is

further supported by our study in which general high CpG level are found near miRNA genes.

## 2. Materials and Methods

The mammalian miRNA precursor sequences are downloaded from the miRNA Registry [2, 5, 6] release 8.1 (<http://microrna.sanger.ac.uk/sequences/>). Next, the mammalian promoter sequences are retrieved from DBTSS [16, 17] release 5.1 (<http://dbtss.hgc.jp/>), where each promoter sequence contains a 1 kb upstream sequence and 200 bp sequence downstream from each Transcription Start Site (TSS) described. Then, the mammalian miRNA precursor sequences are aligned by NCBI-BLAST with the mammalian promoter sequences, hence, it locates precursor sequences originated from the promoter regions. Moreover, the CpG-island information associated with miRNA gene is obtained from the UCSC Genome Browser (<http://genome.ucsc.edu/cgi-bin/hgGateway>, March 2006 draft).

## 3. Results

A highly significant proportion of human miRNA genes, 74 of 462 (16.0%), are either found within or to lie within 1000 bp of a CpG-island (Table 1). This result is in consistent with the work by Holstebro and Tommerup [8]. In this study, we extended their work to a larger set of human miRNA genes, and more species' miRNA genes are taken into account. Table 1 depicts the results for the distances of human miRNA genes measure from genes and CpG islands. In Table 1, the first column denotes the distances of genes and CpG islands from human miRNA genes. The word 'within' means miRNA gene overlaps with the CpG island, and the word 'far' (near) means distances of genes or CpG

islands from human miRNA genes are longer (shorter) than 1000 bp. For instance, 14 miRNA genes are located within 1000 bp of a gene and a CpG island.

Table 1. Distances of human miRNA genes measure from genes and CpG islands.

Distances of genes and CpG islands from human miRNA genes	Number of miRNA genes
gene [far] and CpG island [within]	19
gene [far] and CpG island [near]	13
gene [near (upstream)] and CpG island [within]	8
gene [near (upstream)] and CpG island [near]	3
gene [near (downstream)] and CpG island [within]	2
gene [near (downstream)] and CpG island [near]	0
gene [within] and CpG island [within]	14
gene [within] and CpG island [near]	15
Total	74

Among the 462 miRNA genes, 239 of them are located in the intergenic regions, the rest can be grouped into exonic, intronic or mixed type of miRNA genes. On the other hand, *in vivo* data indicates that around 70% of all human genes are CpG-associated [18], this implies that one expects around 156 genes are CpG-associated. From Table 1, it is reported that 42 of the miRNA genes are associated with CpG islands as well as genes. That is

the CpG-island associated miRNA genes accounts for 27% of CpG-island associated genes. It is suggesting that these miRNAs maybe have a potential role in CpG islands methylation, involving in gene regulation [19]. Since miRNAs could be involved in inducing methylation of target CpG islands and promoter regions, it would be logical to suspect that miRNAs close to or overlapping with CpG islands and within promoter regions could be involved in autoregulation of the methylation.

Table 2 depicts the subset of human miRNA genes which are promoter-related and could be located near CpG islands or not. It is found that a rather high

portion of human promoter-related miRNA genes are CpG-island-related as well, that is 7 out of 10.

Table 2. Human miRNA genes that are located within promoter regions. MiRNA and promoter regions are located on the same strand.

chr.	pre-miRNA	within or near CpG	DBTSS promoter ID ( genegroup ; alternative )	Gene ID	Gene Name
1	hsa-mir-30c-1	No	4439 13 ; 3	4802	NFYC
20	hsa-mir-103-2	No	10812 13 ; 3	80025	PANK2
1	hsa-mir-137	Yes	2951 12 ; 2	400765	FLJ35409
			2951 11 ; 1		
1	hsa-mir-186	No	3779 12 ; 2	9406	ZNF265

6	hsa-mir-219-1	Yes	615 11 ; 1	6015	RING1
7	hsa-mir-594	Yes	15894 11 ; 1	154791	HSPC268
19	hsa-mir-637	Yes	5214 13 ; 3	1613	DAPK3
19	hsa-mir-639	Yes	5770 11 ; 1	9524	GPSN2
11	hsa-mir-611	Yes	14517 11 ; 1	746	C11orf10
17	hsa-mir-632	Yes	8401 12 ; 2	7756	ZNF207

Table 3 depicts the subset of human miRNA genes promoter region and is either CpG island-related or which are located on the opposite strand of a gene's not.

Table 3. Human miRNA genes that are located within promoter regions.

MiRNA and promoter regions are located on the opposite strand.

Chr.	pre-miRNA	within or near CpG	DBTSS promoter ID ( gene group ; alternative )	Gene ID	Gene Name
12	hsa-let-7i	Yes	16407 11 ; 1	283416	C12orf61
3	hsa-mir-191	Yes	7324 11 ; 1	25915	C3orf60
8	hsa-mir-320	Yes	12255 11 ; 1	661	POLR3D
3	hsa-mir-425	Yes	7324 11 ; 1	25915	C3orf60
16	hsa-mir-484	No	13095 11 ; 1	9665	LKAP
11	hsa-mir-611	Yes	14997 11 ; 1	2237	FEN1
3	hsa-mir-565	Yes	7670 11 ; 1	22908	SACM1L
19	hsa-mir-639	Yes	4875 11 ; 1	3337	DNAJB1

A highly significant proportion of mouse's miRNA genes, 37 of 373 (9.9%), are found to lie within 1000 bp of a CpG-island (Table 4). The total number of miRNA genes that are associated with CpG islands is 38 because one of the miRNA genes is located within 1000bp of two CpG islands. Among the 373 miRNA genes, 200 of them are located in the intergenic

regions. On the other hand, data indicates that around 60% of all mouse genes are CpG-associated [20], this implies that one expects around 104 genes are CpG-associated. From Table 4, it is reported that 17 of the miRNA genes are associated with CpG islands as well as genes. That is the CpG-island associated miRNA genes accounts for 16% of CpG-island associated genes.

Table 4. Distances of mouse miRNA genes measure from genes and CpG islands.

Distances of neighboring genes and CpG islands from mouse pre-miRNA	Number of miRNA genes
gene 【far】 & CpG island 【within】	13
gene 【far】 & CpG island 【near】	8
gene 【near (upstream)】 & CpG island 【within】	4
gene 【near (upstream)】 & CpG island 【near】	1
gene 【near (downstream)】 & CpG island 【within】	2
gene 【near (downstream)】 & CpG island 【near】	0
gene 【within】 & CpG island 【within】	6
gene 【within】 & CpG island 【near】	4
Total	38

Table 5 depicts the subset of mouse miRNA genes which are promoter-related and could be located near CpG islands or not. It is found that a rather high

portion of promoter-related miRNA genes are CpG-related as well, that is 4 out of 7 (sense) and 4 out of 4 (anti-sense).

Table 5. Mouse miRNA genes that are located within promoter regions. The strand column denotes the senses of the strand

Strand	pre-miRNA	within or near CpG	DBTSS promoter ID (genegroup;alternative)	Gene ID	Gene Name
sense	mmu-mir-199a-2	No	4691,1	18152	Npn1
	mmu-mir-707	Yes	1189,2	67605	Akt1s1
	mmu-mir-9-2	No	1941,1	320203	C130071C03Rik
	mmu-mir-208	No	11686,2	17888	Myh6
	mmu-mir-688	Yes	8307,1	67942	Atp5g2
	mmu-mir-718	Yes	6899,1	16179	Irak1
	mmu-mir-219-1	Yes	12254,1	19763	Ring1
anti-sense	mmu-mir-685	Yes	12105,1	11546	Parp2
	mmu-mir-707	Yes	201,1	233204	Tbc1d17
	mmu-mir-762	Yes	284,2	12055	Bcl7c
			284,1	12055	Bcl7c
mmu-mir-320	Yes	11838,1	67065	Polr3d	

A highly significant proportion of rat's miRNA genes, 25 of 234 (10.7%), are found to lie within 1000 bp of a CpG-island (Table 6). Among the 234 miRNA genes, 140 of them are located in the intergenic regions. It is reported rat has around 47% of all rat genes are CpG-associated [21], this implies that one expects around 56 genes are CpG-associated. From Table 6, it is reported that 5 of the miRNA genes are associated with CpG islands as well as genes. That is the CpG-island associated miRNA genes accounts for 11% of CpG-island associated genes. The promoter information for rat is not available,

therefore, we cannot provide the promoter-proximal miRNA gene result.

A web site is set up for the public to view the data, which is available at <http://140.130.34.100/960412/html/dna.php>.

Table 6. Distances of rat miRNA genes measure from genes and CpG islands.

Distances of neighboring genes and CpG islands from rat pre-miRNA	Number of miRNA genes
gene 【far】 & CpG island 【within】	14
gene 【far】 & CpG island 【near】	6
gene 【near (upstream)】 & CpG island 【within】	1
gene 【near (upstream)】 & CpG island 【near】	1
gene 【near (downstream)】 & CpG island 【within】	1
gene 【near (downstream)】 & CpG island 【near】	0
gene 【within】 & CpG island 【within】	1
gene 【within】 & CpG island 【near】	1
Total	25

## 4. Conclusions

A rather high portion of mammalian's (human, mouse, and rat) miRNA genes are located near CpG islands. It is found that 74 of 462 (16.0%), 37 of 373 (9.9%), and 25 of 234 (10.7%) of the miRNA genes are located within 1000 bps of a CpG island for human, mouse, and rat respectively. Furthermore, it is determined that a rather high portion of promoter-related miRNA genes are CpG-related as well, This finding suggests a potential role of these

miRNAs in CpG-island methylation, and gene regulation.

The possibility of other mammalian's promoter-related or CpG island-related miRNA genes are under investigation. Our preliminary study indicates that the portions of CpG island-related miRNA genes are rather high as well, which suggests that this is a rather conserve phenomena across species.

## Acknowledgments

Dr. Ka-Lok Ng and Chien-Hung Huang works are supported by the National Science Council of R.O.C. under the grant of NSC 95-2745-E-468-008-URD, and NSC 95-221-E-150-025-MY2..

## References

- [1] Ambros, V. Control of developmental timing in *Caenorhabditis elegans*. *Curr. Opin. Genet. Dev.* 10, 428–433 (2000)
- [2] Ambros, V. et al. A uniform system for microRNA annotation. *RNA* 9, 277–279 (2003)
- [3] Bartel, D.P. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297 (2004).
- [4] Carrington, J.C. and Ambros, V. Role of microRNAs in plant and animal development. *Science* 301, 336–338 (2003).
- [5] Griffiths-Jones, S. The MicroRNA Registry. *Nucleic Acids Res.* 32 (Database issue), D109–D111 (2004).
- [6] Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., Enright, A.J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34(Database Issue), D140–D144 (2006).
- [7] Gardiner-Garden, M., Frommer, M. CpG islands in vertebrate genomes. *J. Mol. Biol.* 196(2), 261–282 (1987).
- [8] Holstebro S. and Tommerup N. A subset of miRNA genes are associated with CpG-islands. European Human Genetics Conference 2005 (ESHG 2005) Prague, May 07, (2005).
- [9] Karolchik D. et al. The UCSC Genome Browser Database. *Nucleic Acids Res.* 31(1), 51–4 (2003).
- [10] Kawasaki, H., Taira, K. Transcriptional gene silencing by short interfering RNAs. *Curr. Opin. Mol. Ther.* 7(2), 125–31 (2005).
- [11] Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. The human genome browser at UCSC. *Genome Res.* 12(6), 996–1006 (2002).
- [12] Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science* 294, 853–858 (2001).
- [13] Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* 12, 735–739 (2002).
- [14] Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., and Tuschl, T. New microRNAs from mouse and human. *RNA* 9, 175–179 (2003).
- [15] Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., et al. The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425, 415–419 (2003).
- [16] Yamashita, R., Suzuki, Y., Wakaguri, H., Tsuritani, K., Nakai, K., Sugano, S. DBTSS: DataBase of Human Transcription Start Sites, progress report 2006. *Nucleic Acids Res.* 34(Database issue), D86–D89 (2006).
- [17] Yamashita R, Suzuki Y, Sugano S, Nakai K. Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity. *Gene* 9;350(2):129–36 (2005).
- [18] Beier V, Mund C, Hoheisel JD.. Monitoring methylation changes in cancer. *Adv Biochem Eng Biotechnol.*, 104, 1–11 (2007).
- [19] Mathieu O, Bender J. RNA-directed DNA methylation. *J Cell Sci.*:4881–8 (2004).

- [20] Jiang C, Han L, Su B, Li WH, Zhao Z. Features and trend of loss of promoter-associated CpG islands in the human and mouse genomes. *Mol Biol Evol.* 24(9), 1991-2000, (2007).
- [21] Xuan Z, Zhao F, Wang J, Chen G, Zhang MQ. Genome-wide promoter extraction and analysis in human, mouse, and rat. *Genome Biol.* 2005;6(8):R72 (2005)