BIOINFORMATICS DISCOVERY NOTE Vol. 24 no. 3 2008, pages 303–308
doi:10.1093/bioinformatics/btm589

Sequence analysis

Putative promoter regions of *miRNA* genes involved in evolutionarily conserved regulatory systems among vertebrates

Shuji Fujita and Hideo Iba*

Division of Host-Parasite Interaction, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku Tokyo, 108-8639, Japan

Received on October 2, 2007; revised on November 7, 2007; accepted on November 23, 2007

Advance Access publication November 30, 2007

Associate Editor: Limsoon Wong

ABSTRACT

Motivation: Just as transcription factors, *miRNA* genes modulate global patterns of gene expression during differentiation, metabolic activation, stimulus response and also carcinogenesis. However, little is currently known how the *miRNA* gene expression itself is regulated owing to lack of basic information of their gene structure. Global prediction of promoter regions of *miRNA* genes would allow us to explore the mechanisms underlying gene-regulatory mechanisms involving these miRNAs.

Results: We speculate that if specific miRNA molecules are involved in evolutionarily conserved regulatory systems in vertebrates, this would entail a high level of conservation of the promoter of *miRNA* gene as well as the miRNA molecule. By our current screening of putative promoter regions of *miRNA* genes (miPPRs) on this base, we identified 59 miPPRs that would direct production of 79 miRNAs. We present both biochemical and bioinformatical verifications of these putative promoters.

Contact: iba@ims.u-tokyo.ac.jp

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

MicroRNAs (miRNAs) are endogenous 20–24 nt RNAs known to mediate the repression of target mRNAs by suppressing translation or promoting mRNA decay in animal (Zamore and Haley, 2005). More than 500 species of miRNAs have now been identified in human (Griffiths-Jones *et al*., 2006), and also predicted to target roughly 30% of the total coding genes in human (Lewis *et al*., 2005; Xie *et al*., 2005). In spite of the strong impact of miRNAs on regulatory network comparable to transcription factors, it still remains largely unknown how human miRNA expression itself is regulated at the transcriptional level, although the vertebrate miRNA genes are thought to be generally transcribed by RNA polymerase II (pol II) to produce a pri-miRNA containing a 5′-cap structure and polyA tail (Cai *et al*., 2004; Lee *et al*., 2004). Several reports have now shown examples of vertebrate miRNA gene regulation by known transcription factors (O'Donnell *et al*., 2005; Zhao *et al*., 2005). Some overrepresented motifs have been also described in limited upstream flanking regions of miRNA embedding regions in nematodes, plants and mammals (Lee *et al*., 2007; Ohler *et al*., 2004; Zhou *et al*., 2007). However in these studies, *miRNA* gene structures and transcription units were not considered. It would be partly because identification of a transcription unit of miRNA is very difficult owing to low expression levels of certain *miRNA* genes in specific tissues (Farh *et al*., 2005), and to instability of primary transcripts of miRNAs (pri-miRNAs) (Cullen, 2004). Therefore, currently only a few pri-miRNA structures have been described biochemically (Cai *et al*., 2004; Kim and Kim, 2007; Lee *et al*., 2004; Taganov *et al*., 2006). Some miRNAs embedding regions are reported to reside in introns of certain coding genes (Kim and Kim, 2007; Rodriguez *et al*., 2004), but it is not fully understood how these miRNA are produced.

To understand regulatory networks that involve miRNAs, we are interested in important regulatory systems conserved across vertebrates. We speculated that in such miRNA molecules, a high level of conservation would be required for not only the miRNA molecule itself but for the promoter of the *miRNA* gene. We have performed a computational approach to predict promoters of such miRNAs on this basis by use of comparative genomic methods that have been successfully employed to some extent in the elucidation of functional *cis*-regulatory elements of coding genes (Blanchette and Tompa, 2002; Loots *et al*., 2002; Xie *et al*., 2005). We finally present a validity of our prediction using bioinformatical and biochemical analyses.

2 METHODS

The databases used in this study are described in online Supplementary Methods. We first searched for aligned blocks containing miRNA hairpin regions that were conserved between either human and chicken or between human and zebrafish in the Blastz-alignment at UCSC (Blanchette *et al*., 2004). We set a threshold of the original alignment scores at higher than 50 000 to filter out non-conserved miRNAs and to remove overly short or poorly conserved blocks. When a hairpin region was split into two adjacent blocks, we concatenated the blocks.

*To whom correspondence should be addressed.

We next searched for highly conserved blocks within the 100 kb upstream of the conserved blocks containing miRNA genes, with the same degree of conservation as above. Assuming that a hairpin region of miRNA is embedded in the middle of the *miRNA* gene, a limitation of 100 kb is reasonable as 95% of the coding genes in RefGene track are shorter than 200 kb. We removed the blocks whose chromosome are different from the ones containing miRNA to assure that each block is located upstream of miRNA in each species. Using FASTA (Pearson and Lipman, 1988), we also removed the upstream blocks in cases where the corresponding human sequence shares homology with the CDS or 3′UTR of human mRNA sequences (RefSeq) (≥85% sequence identity).

From the remaining upstream blocks, we further removed blocks containing other conserved miRNA hairpin regions. When the adjacent consensus regions were ≤300 bp apart, the blocks were concatenated. Consensus sequences were then extracted from the selected blocks as follows: in the position where the base is conserved either between human and chicken or between human and zebrafish and in addition between human and at least one other species among mouse, cow, lizard, chicken, frog, medaka and zebrafish, the base is regarded as a consensus base. Non-conserved bases or gaps are represented by 'n'.

We searched for core promoter elements in the consensus region by an entropy-based calculation (Frech *et al*., 1993; Schneider *et al*., 1986) with TRANSFAC weight matrices (TATA box: M00216, M00252, M00471, CAAT box: M00254 and GC box: M00255) according to the following equations;

$$Ci(i) = \frac{100}{\ln 4} \times \left\{ \sum_{b \in A,C,G,T} f(i,b) \ln f(i,b) + \ln 4 \right\} \quad (1)$$

$$S_k = \frac{100 \times \sum_{j=1}^{m} Ci(i) f(i,b)}{\sum_{j=1}^{m} Ci(i) \max\{f(i,b) | b \in A,C,G,T\}} \quad (2)$$

$$f(i,n) = \min\{f(i,b) | b \in A,C,G,T\} \quad (3)$$

where $Ci(i)$, $f(i,b)$, m and S_k denote the consensus index at position i, relative frequency of base b at position i, and length of matrix and similarity of sequence at kth slid window to consensus one, respectively.

The following cut-off values of similarity were used: $S_k > 88.6$ for TATA box, $S_k > 87.6$ for CCAAT box and $S_k > 86.2$ for GC box. Details regarding our determination of the threshold are described in the Supplementary Materials and Methods. After selecting blocks containing at least one core promoter element, we finally selected the block that was located in the closest proximity to the corresponding miRNA and defined it as an miPPR. All computations were run on 96-CPUs Sun Fire 15K systems and a 128-CPUs Linux cluster system.

3 RESULTS AND DISCUSSION

We made two critical assumptions during the screening for promoters of miRNAs operating in evolutionary conserved systems across vertebrates. First, such miRNAs would not only be derived from pre-miRNA sequences that are conserved among vertebrates, but would also have a conserved promoter sequence. Second, such a conserved promoter harbours at least one core promoter element for RNA polymerase II (TATA, CCAAT and GC box). We screened such conserved regions within a 100 kb region upstream of a series of miRNAs that are conserved among vertebrates, and defined miPPR as one which is in closest proximity to the corresponding miRNA embedding region (see Supplementary Fig. S1). 'Conservation among vertebrates' in this method indicates that conservation is detectable at least between either between human and chicken or between human and zebrafish. Details of the algorithm are described in Methods.

By this screening, 59 regions were eventually selected as the miPPRs for 79 miRNAs. These miPPRs are summarized in Table 1 together with their genomic coordinates, their length (bp) as well as the closest distance between miPPR and the corresponding miRNA hairpin region. The consensus sequences of miPPRs are available in Supplementary Material. Some of these different miRNAs share the same miPPR, suggesting that they are generated as polycistronic transcripts. Among them, such miR-1-2/-133a-1 cluster (Rao *et al*., 2006; Sempere *et al*., 2004) or miR-106a/-18b/-20b/-19b-2 clusters (Fontana *et al*., 2007), for example, has been shown to be co-regulated in several tissues.

The median distance between the human miPPR position and the corresponding miRNA was estimated to be 5.5 kb. By assuming that the hairpin region of the miRNA is located in the middle of the *miRNA* gene, this distance is less than one half of the median length of the known coding genes (11.4 kb). As shown in Figure 1A, the identified human miPPRs positions relative to the corresponding miRNAs are significantly shorter when compared with the 20 sets of background observation, where background PPRs were found from randomly selected conserved blocks as cohorts against miRNAs using the same procedure as miPPRs search. The distances between 22 miPPRs and their corresponding miRNA hairpin regions are longer than 10 kb, whereas 5 miPPRs is present very close to them (≤25 bp) (Supplementary Table S1). We believe that these miPPRs are also good candidates for the miRNA promoter and discussed the basis for it partly using computational analysis in Supplementary Note. Moreover, calculation of odds ratio of CpG dinucleotides (ρ_{CpG}) (Nakashima *et al*., 1997) shows that the human miPPRs had 1.43- and 1.64-fold higher ρ_{CpG} than that of the background cohorts described above and that of entire human genome, respectively (Fig. 1B). Considering that a significant fraction of known promoters locate in CpG island, these results support that the miPPRs are rich in functional promoter sequences.

Although only a few miRNA promoters has been reported, we found that the reported promoter regions for miR-146a (Taganov *et al*., 2006), miR-126 (Kim and Kim, 2007) and miR-10b (Ma *et al*., 2007) are consistent with the miPPR-146a (Fig. 1C), miPPR-126 (Fig. 1D) and miPPR-10b predicted here, respectively. MiPPR-146a contained the conserved NF-κB binding sites that were reported to be critical for LPS response (Supplementary Fig. S2A). In the case of miR-10b, TWIST1, a metastasis-promoting transcription factor, has been shown to induce miR-10b via binding to the most proximal E-box upstream of the miR-10b hairpin region. Importantly, we detect this E-box in human miPPR-10b (Supplementary Fig. S3).

We next performed direct biochemical analysis of miPPR-1-2 (for *miR-1-2* and *miR-133a-1*), miPPR-199a-2 (for *miR-199a-2*) and miPPR-21 (for *miR-21*) based upon unique expression patterns of these miRNAs, which would be conducive for experimental analysis (Fig. 1E–J). The expression patterns of miR-1 and miR-133a are well characterized in muscle

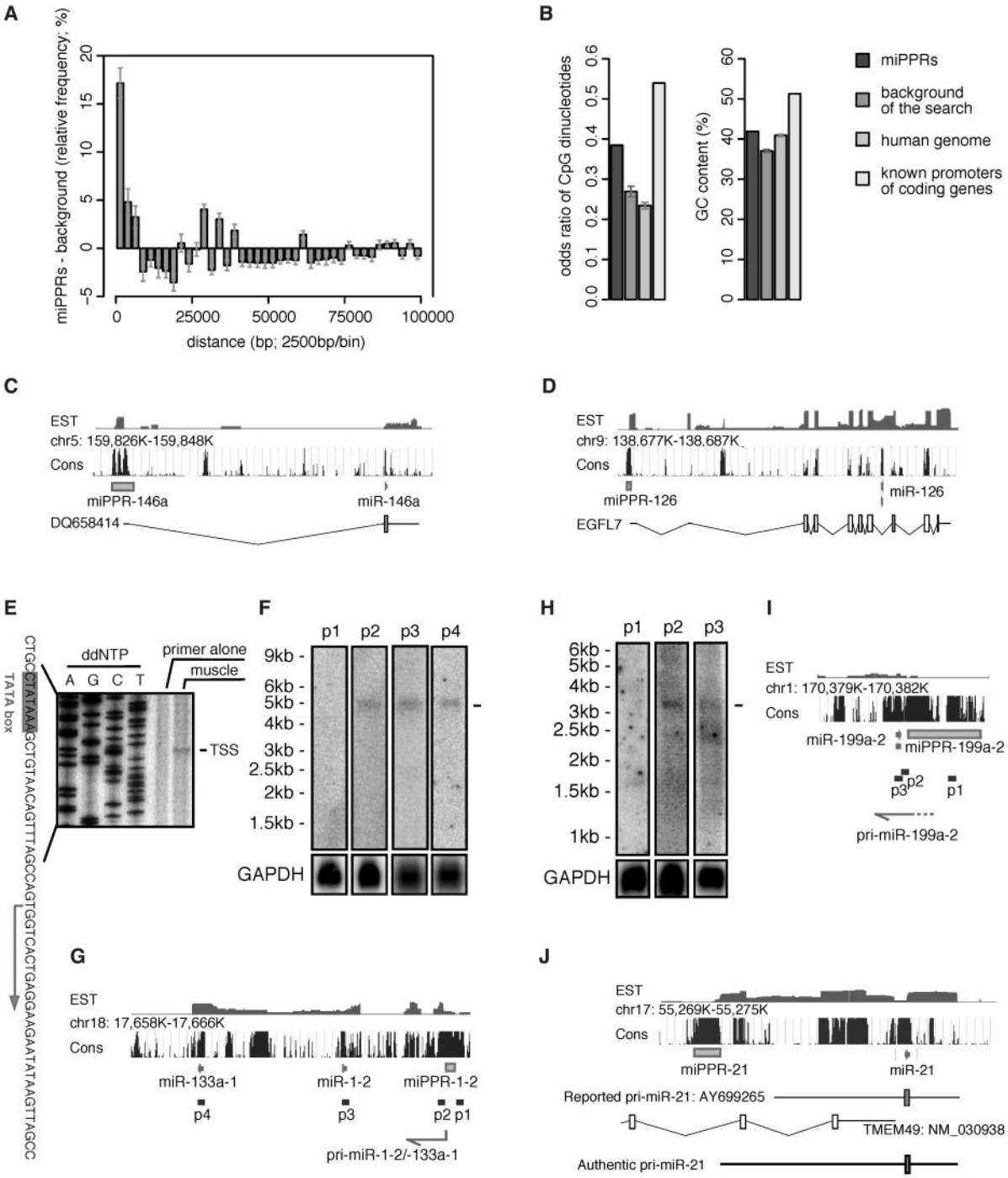

Fig. 1. The statistical (**A** and **B**) and experimental (**C–J**) analyses on human miPPRs. (A) Difference between distance distribution of miPPRs from the corresponding miRNA and that of 20 background sets [one-tailed KS-test; median (IQR) is 1.9×10^{-5} (5.5×10^{-6}–5.1×10^{-5})]. (B) The odds ratios of CpG dinucleotides (left) and the percentage of GC content (right) in the human sequences of miPPRs, the human sequences of the conserved regions of 20 background sets (the same as background sets in A), the randomly selected human genomic sequences (20 sets) and the known human promoter sequences. (C–D) Two examples of miPPRs that correspond to the *miRNA* promoters reported previously. Genomic loci of miPPR-146a (C) and miPPR-126 (D) are shown together with the EST densities of human (EST) and the degree of conservation (Cons). (E–J) Structural analysis and genomic loci of miPPR-1-2 (E–G), miPPR-199a-2 (H and I) and miPPR-21 (J), respectively. (E) Determination of transcription start site (TSS) by primer extension. In left, products of dideoxynucleotide sequencing of antisense strand using the same primers were comigrated to determine precise TSS. The sequence of template strand is shown with TATA box and the determined TSS. (F, H) Northern blot analysis. The used probes are shown in G or in I, respectively. Genomic loci of miPPR-1-2 (G), miPPR199a-2 (I) and miPPR-21 (J) are shown together with determined structures, respectively.

Table 1. Listing of the miPPRs identified in this study, each of which was named after the most proximal miRNA

miPPR	starnd	chr	coordinate[a]	length[b]	distance[c]	miRNA gene	status[d]	overlap[e]
miPPR-137	−	1	98289113	309	4798	mir-137	3	
miPPR-555	−	1	153615908	448	33048	mir-555	4	ASH1L
miPPR-9-1	−	1	154717033	162	60188	mir-9-1		C1orf61
miPPR-214	−	1	170374781	369	111	mir-214	3	
miPPR-199a-2	−	1	170380488	1349	81	mir-199a-2	1	
miPPR-181a-1	−	1	197097224	548	2319, 2490	mir-181a-1, mir-181b-1		
miPPR-29b-2	−	1	206062613	326	20122, 20706	mir-29b-2, mir-29c	4	
miPPR-205	+	1	207668788	129	3313	mir-205	3	
miPPR-194-1	−	1	218358971	419	765, 1044	mir-194-1, mir-215		
miPPR-216b	−	2	56089793	485	8359, 20095, 26078	mir-216b, mir-216a, mir-217		
miPPR-128a	+	2	136112862	165	26575	mir-128a	3′	R3HDM1
miPPR-10b	+	2	176723227	1340	50	mir-10b	2	HOXD4
miPPR-138-1	+	3	44126803	661	3905	mir-138-1		
miPPR-191	−	3	49038339	221	5193	mir-191		DALRD3
miPPR-568	−	3	115520340	718	2234	mir-568	3	
miPPR-15b	+	3	161570667	382	34403, 34560	mir-15b, mir-16-2		SMC4L1
miPPR-551b	+	3	169744643	356	7693	mir-551b		
miPPR-218-1	+	4	20136740	1175	2256	mir-218-1		SLIT2
miPPR-302b	−	4	113789309	683	147, 274, 453, 633, 763	mir-302b, mir-302c, mir-302a, mir-302d, mir-367	3	
miPPR-9-2	−	5	87998532	557	19	mir-9-2	3	
miPPR-145	+	5	148766736	387	23666	mir-145	4	
miPPR-146a	+	5	159828772	1513	16165	mir-146a	2	
miPPR-218-2	−	5	168132153	414	4315	mir-218-2		SLIT3
miPPR-30a	−	6	72170589	826	544	mir-30a		
miPPR-196b	−	7	27178988	506	3281	mir-196b	3′	HOXA10
miPPR-490	+	7	136204287	807	34167	mir-490	3′	CHRM2*
miPPR-671	+	7	150527644	449	38796	mir-671		CSGlcA-T
miPPR-153-2	−	7	157059880	241	5	mir-153-2		PTPRN2
miPPR-124-2	+	8	65451029	136	3231	mir-124-2	3′	
miPPR-875	−	8	100628786	605	10521, 10652	mir-875, mir-599		
miPPR-7-1	−	9	85774615	186	23	mir-7-1	3	HNRPK
miPPR-23b	+	9	96859033	1109	28278, 28515, 29091	mir-23b, mir-27b, mir-24-1	3′	C9orf3
miPPR-32	−	9	110887957	447	39558	mir-32	4	C9orf5
miPPR-455	+	9	115920787	420	90748	mir-455		COL27A1
miPPR-219-2	−	9	130206806	463	11992	mir-219-2		
miPPR-126	+	9	138677166	81	7709	mir-126	2	EGFL7*
miPPR-129-2	+	11	43554743	492	4777	mir-129-2	3′	
miPPR-125b-1	−	11	121477545	387	1783	mir-125b-1	3′	
miPPR-let-7a-2	−	11	121522522	44	11	let-7a-2		
miPPR-100	−	11	121534592	122	6366	mir-100	3′	
miPPR-196a-2	+	12	52665317	350	6472	mir-196a-2	3′	HOXC10
miPPR-615	+	12	52713955	110	46	mir-615		HOXC5, HOXC4
miPPR-let-7i	+	12	61246223	475	37510	let-7i	4	
miPPR-135a-2	+	12	96481471	1036	250	mir-135a-2	4	
miPPR-138-2	+	16	55373657	1503	76274	mir-138-2	4	
miPPR-328	−	16	65888841	251	95042	mir-328		
miPPR-144	−	17	24300086	1091	87324, 87502	mir-144, mir-451		
miPPR-21	−	17	55270100	462	3309	mir-21	1	TMEM49
miPPR-1-2	−	18	17665373	199	2326, 5629	mir-1-2, mir-133a-1	1	
miPPR-187	−	18	31739256	338	369	mir-187		
miPPR-499	+	20	33029109	249	12731	mir-499	4	MYH7B
miPPR-1-1	+	20	60560823	158	1135	mir-1-1	3′	C20orf166
miPPR-133a-2	+	20	60569493	245	3071	mir-133a-2	4	C20orf166
miPPR-124-3	+	20	61265799	401	14498	mir-124-3		
miPPR-99a	+	21	16830698	2645	2582, 3321	mir-99a, let-7c	3′	C21orf34
miPPR-125b-2	+	21	16884415	2456	13	mir-125b-2	3′	C21orf34
miPPR-222	−	X	45519342	822	27868, 28704	mir-222, mir-221	4	
miPPR-let-7f-2	−	X	53661209	130	60249	let-7f-2		HUWE1
miPPR-106a	−	X	133137364	274	5390, 5557, 5791, 5902	mir-106a, mir-18b, mir-20b, mir-19b-2	4	

[a]The chromosomal position of the miPPR.
[b]The length of the miPPR.
[c]The length (bp) between the 3′end of miPPR and the 5′end of the miRNA hairpin region.
[d]Current validation status of each miPPR, which was categolized as follows: (1); Confirmed by our current biochemical experiments. (2); Confirmed by previous reports. (3), (3′) and (4); Supported by the presence of cDNA clones in databases that are expressed from, or just downstream of the miPPR. Among these cDNA clones, some clones contain the corresponding miRNA (3), and others directs the transcript, the intron of which embeds the corresponding miRNA (3′).
[e]If the miRNA gene is overlapped with a coding gene, the name of the coding gene is shown.
*A promoter region of the coding gene coincides with the miPPR.

development (Rao *et al*., 2006; Zhao *et al*., 2005). Both up- and down-regulation of miR-199a has been reported in several cancer cell types, indicating its potential involvement in tumorigenesis (Volinia *et al*., 2006). MiR-21 is also highly expressed in various human cancers (Volinia *et al*., 2006).

Primer extension experiment using human muscle RNA shows a transcription 20 bp downstream of a conserved TATA box in miPPR-1-2 (Fig. 1E). By northern analysis, a single transcript of 5 kb was detected by probes for the just downstream region of miPPR-1-2 and also for the region embedding either miR-1-2 or miR-133a-1, but not by a probe for the just upstream region of miPPR-1-2 (Fig. 1F and G). These results strongly support our contention that miPPR-1-2 indeed functions as the common promoter for miR-1-2 and miR-133a-1. In addition, this showed direct evidence to support the recent suggestion that miR-1-2 and miPPR-133a-1 might be embedded in a polycistronic transcript as described above. In the consensus sequence of miPPR-1-2 (Supplementary Fig. S2B), we also identified the previously reported SRF and MyoD element (Zhao *et al*., 2005). Similar northern analysis using RNA from HeLa cells has suggested that miPPR-199a-2 also directs a transcript embedding miR-199a-2 (Figs 1H and I). By primer extension and northern analysis, we elucidated that miPPR-21 drives pri-miR-21 30 bp downstream of TATA box present in miPPR-21 (Fig. 1J). These analyses on miR-21 indicated that the authentic pri-miR-21 is transcribed just downstream of TATA box in miPPR-21 but is not drived by the previously reported region (Cai *et al*., 2004) and the details of biochemical analyses will be published elsewhere with promoter analysis including transcription factors involved in the regulation.

We categorized miPPRs with validation status as a measure of predictive accuracy (Table 1). Our direct biochemical analyses of a primary transcript of miR-1-2 and miR-133a-1 (Figs 1E–G), miR-199a-2 (Figs 1H–I) or miR-21 (Fig. 1J) were consistent with our predicted miPPRs (status 1). miPPRs-146a, miPPR-126 and miPPR-10b are consistent with previous studies of the promoter regions of the miR-146a, miR-126 and miR-10b, respectively (Figs 1C and D) (status 2).

Other than these 6 miPPR, we identified 30 miPPRs that direct transcriptional initiation according to the UCSC database (status 3, 3′ and 4) among total 59 miPPRs. Moreover, 19 out of these 30 miPPRs direct transcripts that harbour the corresponding miRNA (status 3) or transcripts that are produced by splicing of introns containing the miRNA (status 3′). All these findings support the validity of our approach. In some miPPRs with no varidation status, we found transcripts containing the miRNAs that are initialized either far upstream or downstream of the miPPR in the database. Therefore, some other promoter candidates would be major promoters, at least in the tissue from which the EST was isolated.

Since 0.03 quantile of the distribution of distances between adjacent coding genes is calculated to be 1.5 kbp, we tentatively regarded that a certain miRNA gene is overlapped with a coding gene in the genome when we find a coding gene locus within 1.5 kb flanking regions of the miRNA embedding region. As shown in Table 1, 30 conserved miRNAs are mapped to be overlapped with coding genes. Among them, miPPR-490 corresponds to the common promoter of the *miR-490* gene and the overlapping coding gene, *CHRM2*, whereas the *miR-126* and *EGFL7* genes share the common promoter, miPPR-126. These results suggest that the coding gene and miRNA could be co-regulated like these cases. However, most of these miPPRs were located distinct positions from the promoters of the overlapping gene, indicating that expression of miRNA would be regulated independent of the overlapping coding genes in these cases. For example, *miR-21* were overlapped with the *TMEM49* coding gene, however the expression of *miR-21* is regulated by its own promoter, miPPR-21 that resides in an intron of *TMEM49*, independent of the transcription of *TMEM49* (Fig. 1J).

Deciphering transcriptional regulation on *miRNA* genes is now urgent to understand regulatory networks structured with several layers including miRNA-mediated post-transcriptional repression. Biochemical promoter analysis using this information would be quite efficient and the accumulation of such biochemical data would be useful to improve the algorithm. We believe that such processes will contribute to find out conserved regulatory systems operating in vertebrates for further understanding of miRNA expression potential and misregulation in cancer or other diseases.

ACKNOWLEDGEMENTS

Computation time was provided by the Super Computer System, Human Genome Center, Institute of Medical Science, University of Tokyo. This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas and by the special coordination Funds for Promoting Science and Technology from the Ministry of Education, Culture, Sports, Science and Technology Japan.

Conflict of Interest: none declared.

REFERENCES

Blanchette,M. *et al*. (2004) Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res*., **14**, 708–715.

Blanchette,M. and Tompa,M. (2002) Discovery of regulatory elements by a computational method for phylogenetic footprinting. *Genome Res*., **12**, 739–748.

Cai,X. *et al*. (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, **10**, 1957–1966.

Cullen,B.R. (2004) Transcription and processing of human microRNA precursors. *Mol. Cell*, **16**, 861–865.

Farh,K.K. *et al*. (2005) The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science*, **310**, 1817–1821.

Fontana,L. *et al*. (2007) MicroRNAs 17-5p-20a-106a control monocytopoiesis through AML1 targeting and M-CSF receptor upregulation. *Nat. Cell Biol*., **9**, 775–787.

Frech,K. *et al*. (1993) Computer-assisted prediction, classification, and delimitation of protein binding sites in nucleic acids. *Nucleic Acids Res*., **21**, 1655–1664.

Griffiths-Jones,S. *et al*. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*., **34**, D140–D144.

Kim,Y.K. and Kim,V.N. (2007) Processing of intronic microRNAs. *EMBO J*., **26**, 775–783.

Lee,J. *et al*. (2007) Regulatory circuit of human microRNA biogenesis. *PLoS Comput. Biol*., **3**, e67.

Lee,Y. *et al*. (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*., **23**, 4051–4060.
Lewis,B.P. *et al*. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.
Loots,G.G. *et al*. (2002) rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Res*., **12**, 832–839.
Ma,L. *et al*. (2007) Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature*, **449**, 682–688.
Nakashima,H. *et al*. (1997) Differences in dinucleotide frequencies of human, yeast, and Escherichia coli genes. *DNA Res*., **4**, 185–192.
O'Donnell,K.A. *et al*. (2005) c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*, **435**, 839–843.
Ohler,U. *et al*. (2004) Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. *RNA*, **10**, 1309–1322.
Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
Rao,P.K. *et al*. (2006) Myogenic factors that regulate expression of muscle-specific microRNAs. *Proc. Natl. Acad. Sci. USA*, **103**, 8721–8726.
Rodriguez,A. *et al*. (2004) Identification of mammalian microRNA host genes and transcription units. *Genome Res*., **14**, 1902–1910.
Schneider,T.D. *et al*. (1986) Information content of binding sites on nucleotide sequences. *J. Mol. Biol*., **188**, 415–431.
Sempere,L.F. *et al*. (2004) Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol*., **5**, R13.
Taganov,K.D. *et al*. (2006) NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. *Proc. Natl Acad. Sci. USA*, **103**, 12481–12486.
Volinia,S. *et al*. (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl Acad. Sci. USA*, **103**, 2257–2261.
Xie,X. *et al*. (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature*, **434**, 338–345.
Zamore,P.D. and Haley,B. (2005) Ribo-gnome: the big world of small RNAs. *Science*, **309**, 1519–1524.
Zhao,Y. *et al*. (2005) Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature*, **436**, 214–220.
Zhou,X. *et al*. (2007) Characterization and identification of microRNA core promoters in four model species. *PLoS Comput. Biol*., **3**, e37.