2012 4th International Conference on Bioinformatics and Biomedical Technology
IPCBEE vol.29 (2012) © (2012) IACSIT Press, Singapore

An identification of microRNAs that may activate transcription through Ago2

Nilesh Gramani [1+], Apiwat Mutirangura [2] and Chatchawit Aporntewan [3]

[1] Computer Science and Information Technology

[2] Center for Excellence in Molecular Genetics of Cancer and Human Diseases, Department of Anatomy, Faculty of Medicine

[3] Department of Mathematics and Computer Science, Faculty of Science Chulalongkorn University, Bangkok, 10330, Thailand

Abstract. We performed a computer-based analysis to identify miRNAs that may activate transcription through Ago2. This cellular mechanism is so called RNA activation or RNAa. And it is believed that Ago2, a protein in Argonaute family, is required for RNA activity. We used all microarray data from a public database called, Gene Expression Omnibus (GEO). The main idea is to find the genes that were commonly 1) up-regulated in a miRNA transfection experiment and 2) down-regulated in an Ago2 knockdown experiment. Finally, we performed local sequence alignment between the transfected miRNAs and the gene promoters. Our findings show at least 6 miRNAs that could bind on promoters and may activate transcription.

Keywords: microRNA, miRNA, Ago2, Argonaute 2, RNA activation, RNAa.

1. Introduction

Ribonucleic acid (RNA) is commonly found in organic cells. It is believed that RNA serves only as an intermediate in the transcription process. Now we know that RNA is a regulatory element that mediates gene expression. The discovery of RNA interference (RNAi), awarded Nobel Prize in Physiology in 2006, indicates that RNA plays an important role in suppressing the transcription [1]. RNAi is characterized by the binding of a small RNA to a messenger RNA (mRNA). The target mRNA is either degraded or destroyed. As a result, the corresponding gene is down-regulated. An important protein for RNAi activity is Ago2 which loads a short double-strand RNA (dsRNA) and rips it into two single-strand RNAs. The passenger strand is discarded. Ago2 takes only the guide strand, and then forms RNA-Induced Silencing Complex (RISC). This complex binds to a target mRNA whose sequence is complementary to the guide strand.

The opposite mechanism that activates transcription is referred to as RNA activation (RNAa) [2,3]. RNAa is characterized by an increase of gene expression after introducing a small dsRNA which is complementary to the gene promoter. Ago2 may be required for RNAa activity because in an Ago2 knockdown experiment the gene expression level did not increase [4]. However, RNAa has been reported case by case, and the underlying process of RNAa remains largely unknown.

Although synthetic dsRNAs have been widely used in RNAi and RNAa studies, both of them are natural mechanisms. In nature, dsRNAs called microRNA (miRNA) are ubiquitous in eukaryotic cells. MiRNAs are encoded in DNA like other genes, but they are non-coding and are not translated to functional proteins. Thus, the role of miRNAs is solely a regulatory element. The regulatory functions of miRNA through RNAi have been studied extensively, but very few miRNAs that involve RNAa have been reported.

+ Corresponding author.
E-mail address: Nilesh_Gramani@hotmail.com

This paper aims to search for miRNAs that may activate transcription and increase gene expression. Our strategy employed a public gene expression database for screening the activated genes. Then, we performed local sequence alignment between miRNAs and the promoters of those genes.

2. Methods

2.1. A selection of microarray experiments from GEO

We intended to reuse a set of microarray experiments that were produced by independent principal investigators but to answer to a different research question. First, we selected microarray experiments that transfected a single or multiple miRNAs. There were plenty of them due to the study of RNAi, but we were going to use them for studying RNAa. Second, an Ago2 knockdown experiment was selected because Ago2 was required for RNAa. All microarray experiments shown in Table 1 were taken from Gene Expression Omnibus (GEO) [5].

Table 1: miRNA transfection and Ago2 knockdown experiments selected from GEO.

Series	Transfected miRNA(s)	Title
GSE6207	miR-124	miR-124 transfection time course
GSE7864	miR-34a,34b,34c	A microRNA component of the p53 tumor suppressor network
GSE11701	miR-205	Genes modulated by miR-205 in DU145 prostate cancer cells
GSE13105	miR-192,215	Coordinated regulation of cell cycle transcripts by p53-inducible microRNAs, miR-192 and -215
GSE13296	miR-155	miR-155 KO in human dendritic cells
GSE16568	miR-22	Gene expression analyses of mir-22 overexpression in ovarian clear cell cancer cell line
GSE16569	miR-30a,30d	Gene expression analyses of mir-30a knockdown in ovarian clear cell cancer cell line
GSE16571	miR-100	Gene expression analyses of mir-100 overexpression in ovarian clear cell cancer cell line
GSE16572	miR-182	Gene expression analyses of mir-182 knockdown in ovarian clear cell cancer cell line
GSE16700	miR-31	Gene expression analyses of mir-31 overexpression in ovarian serous cancer cell line
GSE16908	miR-31	miR-31 inhibits lymphatic lineage-specific differentiation in vitro and lymphatic vessel development in vivo
GSE18510	miR-193b	miR-193b represses cell proliferation and regulates cyclin D1 in melanoma: Malme-3M
GSE18545	miR-9	A MicroRNA Expression Signature For Cervical Cancer Prognosis
GSE18625	miR-145	Identification of miR-145 targets involved in colon cancer
GSE18651	miR-29a,29b,29c	miR-29 targets in human fetal lung fibroblast IMR-90 cells
GSE19777	miR-221,222	Antisense miRNA-221/222 (si221/222) and control inhibitor (GFP) treated fulvestrant-resistant breast cancer cells
GSE20293	miR-30e,30e*	miR-30e* induced gene expression alteration in glioma cells
GSE20668	miR-100	Gene expression in human umbilical cord endothelial cells following premiR-100 overexpression
GSE20679	miR-517a	mRNA expression profile modified by microRNA mir-517a (MIR517A) in human hepatocellular carcinoma cell line
GSE20745	miR-17,18a,19a, 19b,20a,92a	Members of the microRNA-17-92 cluster exhibit a cell intrinsic anti-angiogenic function in endothelial cells
GSE21577	miR-103a,103b,93, 19b,106b	Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP: miRNA inhibition data
GSE24069	kshv-miR-K12-10a	miR-K10a expression and inhibition

GSE25224	miR-124	Persistence of seed-based activity following microRNA segmentation
GSE4246	Analysis of transcripts regulated by Dicer and Argonaute proteins in human HEK-293 cells (Ago2 knockdown)	

2.2. An intersection between two microarray experiments

The principal investigators, who conducted miRNA transfection experiments, hypothesized that a transfected miRNA induced RNAi, but we hypothesized that it may induce RNAa too. If so, we should observe the target genes up regulated. Similarly to the Ago2 knockdown experiment, Ago2 is required for RNAa so we should observe the target genes down regulated. An intersection between miRNA transfection and AGO2 knockdown experiments can be made by means of a 2x2 table as shown in Table 2. We used our software called CU-DREAM for intersecting two microarrays [6]. The 2x2 table yields an odds ratio and a p-value. If OR > 1 and p-value is significant, the transfected miRNA may target at the promoters and activate the transcription of the genes in set *A*.

2.3. Local sequence alignment

We collected miRNA sequences from miRBase [7]. To find miRNA binding sites, we aligned the miRNA sequences with the promoters of the genes in set *A*. A gene promoter is defined as the 1000bp upstream from the transcription start site (TSS). The local sequence alignment was carried out by Smith-Waterman algorithm [8]. A scoring matrix is shown in Table 3. There are many scoring matrices for DNA-DNA and RNA-RNA bindings, but RNAa involves DNA-RNA binding between promoters and miRNAs. To the best of our knowledge, there are no previous definitions. Consequently, we defined our scoring matrix based on the previous reports of RNAa. A-U, T-A, C-G and G-C were strong bonds and were given 2 points. T-G and G-U were weak bonds and were given only 1 points. The penalties of opening a gap and extending a gap were set at the same value, -2.

Table 3: A scoring matrix for Smith-Waterman algorithm.

		RNA			
		A	**U**	**C**	**G**
DNA	**A**	0	2	0	0
	T	2	0	0	1
	C	0	0	0	2
	G	0	1	2	0

3. Experimental Results

The intersections between miRNA transfection and Ago2 knockdown experiments are shown in Table 4 (sorted by odds ratio). Only the top four were allowed to proceed in local sequence alignment. The best alignments between the transfected miRNAs and the promoters of the genes in set *A* are depicted in Fig. 1. Note that there are a total of 10 binding sites of GSE19777 (miR-222). Their alignment scores are the same (score = 29), only two of them are shown in Fig. 1.

Table 4: The intersections between miRNA transfection and Ago2 knockdown experiments.

GSE	**2x2 table**				**Odds ratio**	**95% confidence interval**	**Unadjusted p-value (Pearson's Chi-square)**
	$\|A\|$	$\|B\|$	$\|C\|$	$\|D\|$			
GSE19777 (1)	31	166	893	10,983	2.30	1.56 - 3.39	1.69E-05
GSE19777 (2)	26	151	898	10,998	2.11	1.38 - 3.21	3.90E-04
GSE20745	12	78	912	11,071	1.87	1.01 - 3.44	4.19E-02
GSE6207	31	214	893	10,935	1.77	1.21 - 2.60	2.94E-03
GSE24069	6	32	584	5,127	1.65	0.69 - 3.95	2.60E-01
GSE21577	50	383	869	10,650	1.60	1.18 - 2.17	2.14E-03
GSE13105	33	215	725	7,439	1.57	1.08 - 2.29	1.65E-02
GSE11701	19	128	706	6,837	1.44	0.88 - 2.34	1.43E-01

GSE18625	9	76	915	11,073	1.43	0.72 - 2.87	3.07E-01
GSE16571	70	545	790	8,544	1.39	1.07 - 1.80	1.26E-02
GSE16700	58	477	802	8,612	1.31	0.98 - 1.73	6.30E-02
GSE13105	35	276	723	7,377	1.29	0.90 - 1.85	1.59E-01
GSE13296	17	160	907	10,986	1.29	0.78 - 2.13	3.26E-01
GSE20668	56	435	789	7,918	1.29	0.97 - 1.72	8.02E-02
GSE16908	38	281	716	6,663	1.26	0.89 - 1.78	1.94E-01
GSE7864 (3)	71	566	722	7,257	1.26	0.97 - 1.63	7.81E-02
GSE7864 (2)	66	543	727	7,280	1.22	0.93 - 1.59	1.48E-01
GSE16568	29	256	831	8,833	1.20	0.81 - 1.78	3.51E-01
GSE16569	7	62	853	9,027	1.19	0.55 - 2.62	6.56E-01
GSE7864 (1)	60	520	733	7,303	1.15	0.87 - 1.52	3.25E-01
GSE7864 (7)	47	405	746	7,418	1.15	0.85 - 1.58	3.67E-01
GSE18510	64	560	781	7,793	1.14	0.87 - 1.49	3.38E-01
GSE7864 (4)	39	358	754	7,465	1.08	0.77 - 1.51	6.62E-01
GSE25224	45	433	703	7,249	1.07	0.78 - 1.47	6.68E-01
GSE7864 (8)	45	447	748	7,376	0.99	0.72 - 1.36	9.64E-01
GSE7864 (9)	33	346	760	7,477	0.94	0.65 - 1.35	7.32E-01
GSE20679	24	211	748	6,132	0.93	0.61 - 1.43	7.49E-01
GSE20293	60	640	754	7,350	0.91	0.69 - 1.20	5.21E-01
GSE18545	46	552	814	8,537	0.87	0.64 - 1.19	3.93E-01
GSE7864 (5)	32	400	761	7,423	0.78	0.54 - 1.13	1.85E-01
GSE18651	8	88	764	6,250	0.74	0.36 - 1.54	4.23E-01
GSE7864 (6)	27	407	766	7,416	0.64	0.43 - 0.95	2.74E-02
GSE16572	2	64	858	9,025	0.33	0.08 - 1.35	1.03E-01

GSE19777 (miR-221)

```
miR-221   5'-agcuacau-ugucugcuggguuuc-3' score: 33
             |||||||| :|||| ||  |||||
NM_002998 3'-tcgatgtatgcaga-ga-gcaaag-5' antisense
             ↑
             at -299bp upstream from TSS
miR-221   3'-cuuugggucgu-cuguuacaucg-5' score: 33
             ||||:|| ||| |:|: || |||
NM_022459 5'-gaaatcctgcacggcgctggagc-3' sense
             ↑
             at -448bp upstream from TSS
miR-221   5'-agcua-cauugucugcuggguuuc-3' score: 34
             |:|:| |||||||||| :::: ||
NM_152640 3'-ttggtggtaacagacgttttgcag-5' antisense
             ↑
             at -540bp upstream from TSS
```

GSE20745 (miR-17,18a,19a,19b,20a,92a)

```
miR-17    3'-gauggacgugacauucgugaaac-5' score: 34
             :|:|:| :|:| |||||: ||||
NM_006644 5'-ttgcttttattataagcggtttg-3' sense
             ↑
             at -658bp upstream from TSS
miR-20a   3'-gauggacgugauauucgugaaau-5' score: 35
             :|:|:| :|:|||||||: |||:
NM_006644 5'-ttgcttttattataagcggtttg-3' sense
             ↑
             at -658bp upstream from TSS
miR-92a   3'-ugucc-ggcccuguucacguuau-5' score: 34
             | ||| |||||:|| |||||: |
NM_031372 5'-agaggcccggggcacgtgcagaa-3' sense
             ↑
             at -169bp upstream from TSS
```

GSE19777 (miR-222)

```
miR-222   5'-agcua--caucu-ggcuacugggu-3' Score: 29
             |||||  | ||| || | ||||:|
NM_000362 3'-tcgatcagcagatcctaggaccta-5' Antisense
             ↑
             at -200bp upstream from TSS
 miR-222   5'-gcuacaucuggcuacugggu-3' Score: 29
             ||:|| ||||| ||||| ::
NM_000060 3'-cggtggagacc-atgacgtg-5' Antisense
             ↑
             at -202bp upstream from TSS
```

GSE6207 (miR-124)

```
miR-124   5'-uaaggcacgcggugaaugcc-3' score: 30
             | |||| |||:|: |||| |
NM_005342 3'-actccgggcgtcggttaccg-5' antisense
             ↑
             at -69bp upstream from TSS
miR-124   5'-uaaggcacgcgguga-augcc-3' score: 30
             || :||||  :|||| |||||
NM_005107 3'-atctcgtgaatcactgtacgg-5' antisense
             ↑
             at -648bp upstream from TSS
```

Fig. 1: The best alignments between transfected miRNAs and promoters.

4. Discussion

We have proposed a number of gene promoters that could be targets of RNAa. Moreover, we have identified the putative binding locations. There are several observations as follows.

- Unlike a synthetic RNA that is made complementary to a specific target, a perfect match between a natural miRNA and a promoter is very rare. It is expected that the binding

quality determines the effectiveness of RNAa. If so, the gene activation by a miRNA is modest compared to that of human-invented RNAs. This makes the identification of natural RNAa difficult because the difference of gene expression between cases (miRNA transfected) and controls is very small and is hardly detected with few samples.

- Although the transfected miRNAs putatively bind on the promoters of genes in set *A*, they also bind on the promoters of genes in set *D* as well. We use the 2x2 tables below to test whether the transfected miRNAs often bind on the set *A*, which is supposed to be the targets of RNAa. Unfortunately, the odds ratios are not high and the p-values are not significant. This suggests that there must be other necessary conditions for RNAa rather than RNA-DNA binding determined by sequence alignment.

GSE19777 (miR-221)	$\|A\|$	$\|D\|$
Binding score ≥ 33	3	1,267
Not bind	27	9,220

OR = 0.81, p-value = 7.27E-01

GSE20745	$\|A\|$	$\|D\|$
Binding score ≥ 34	2	1,334
Not bind	10	9,237

OR = 1.38, p-value = 6.73E-01

GSE19777 (miR-222)	$\|A\|$	$\|D\|$
Binding score ≥ 29	9	5,120
Not bind	17	5,377

OR = 0.56, p-value = 1.49E-01

GSE6207	$\|A\|$	$\|D\|$
Binding score ≥ 30	2	320
Not bind	29	10,117

OR = 2.18, p-value = 2.76E-01

We have compared the bindings between set *A* and set *D* in order to find a binding characteristic that may distinguish RNAa. The characteristics of interest may be alignment score, sense/antisense, and distance from TSS. However, we have not yet found any remarkable characteristics. Note that a large number of the bindings in set *D* are not shown here due to page limit.

5. Acknowledgements

This study is supported by the Thailand Research Fund (TRF), Research Chair Grant 2011, National Science and Technology Development Agency (NSTDA), Four Seasons Hotel Bangkok's 4th Cancer Care Charity Fun Run in association with the Thai Red Cross Society.

6. References

[1] A. Fire, C. Mello. RNA Interference. Advanced Information on The Nobel Prize in Physiology 2006.

[2] L. Li et al. Small dsRNAs induce transcriptional activation in human cells. *PNAS* 2006, **103** (46): 17337-17342.

[3] R. Place et al. MicroRNA-373 induces expression of genes with complementary promoter sequences. *PNAS* 2008, **105** (5): 1608-1603.

[4] R. Place et al. Defining features and exploring chemical modifications to manipulate RNAa activity. *Curr. Pharm. Biotechnol.* 2010, **11** (5): 518-526.

[5] T. Barrett et al. NCBI GEO: archive for functional genomics data sets – 10 years on. *Nucleic. Acids. Res.* 2011, **39** (Database issue): D1005-10.

[6] C. Aporntewan, A. Mutirangura. Connection Up- and Down-Regulation Expression Analysis of Microarrays (CU-DREAM): A physiogenomic discovery tool, *Asian Biomed.* 2011, **5** (2): 257-262.

[7] A. Kozomara, S. Griffiths-Jones. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 2011, **39** (Database issue): D152-7.

[8] T. Smith, M. Waterman. Identification of common molecular subsequences. J. Mol. Biol. 1981, **147** : 195–197.