African Journal of Biotechnology Vol. 12(10), pp. 1016-1024, 6 March, 2013
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB12.2054
ISSN 1684–5315 ©2013 Academic Journals

Full Length Research Paper

Characteristics of binding sites of intergenic, intronic and exonic miRNAs with mRNAs of oncogenes coding intronic miRNAs

Olga A. Berillo[1]*, Assel S. Issabekova[1], Mireille Régnier[2] and Anatoliy T. Ivashchenko[1]

[1]Department of Biology and Biotechnology, Al-Farabi Kazakh National University, Almaty 050038, Kazakhstan.
[2]INRIA/LIX team AMIB 1 rue Honoré d'Estienne d'Orves 91120 Palaiseau, France.

Accepted 2 October, 2012

The interaction of 784 intergenic (ig-miRNA), 686 intronic (in-miRNA) and 49 exonic miRNAs (ex-miRNA) with mRNAs of 51 oncogenes coding in-miRNAs was investigated. Out of the studied genes, 44 were targets for 94 ig-miRNAs, 29 were targets for 44 in-miRNAs and 7 were targets for 7 ex-miRNAs. The density of miRNA binding sites was higher in 5'-untranslated regions than it was in coding sequences and 3'-untranslated regions. Three types of miRNA interaction with mRNA were revealed: 5'-dominant canonical, 3'-compensatory and complementary types. In-miRNAs do not interact with mRNAs of host genes (where in-miRNA is encoded). Linkage between some mRNAs of genes encodes in-miRNAs via other in-miRNAs was revealed. These data promote the understanding of interaction mechanism of miRNA with mRNA genes participating in gastrointestinal and breast cancers.

Key words: Intergenic miRNA, intronic miRNA, exonic miRNA, mRNA, 5'-untranslated region, coding sequences, 3'-untranslated region, human, oncogene.

INTRODUCTION

MicroRNAs (miRNA) are non-protein coding RNA sequences that have length of about 22 nucleotides (nt). miRNA are evolutionary conservative in many organisms and accomplishes important regulatory function (Ibanez-Ventoso et al., 2008). The binding miRNA with mRNA leads to specific splitting, deadenylation or translation repression (Pillai et al., 2007). miRNA genes can be encoded in the intergenic or intragenic regions. Intragenic miRNAs are situated within intron or exon of their host genes (Kim and Kim, 2007). Some of the miRNAs are found on bonder of exon and intron (Rodriguez et al., 2004). The number of miRNA is considerably enlarged due to new methods of search and comparison of targets. MiRNAs regulate translation of about 60% of all human protein-coding genes (Friedman et al., 2009). mRNA can have some binding sites with one or with several miRNAs. Hence, the influence of miRNA on inhibition of mRNA translation becomes greater (Baek et al., 2008; Selbach et al., 2008; Grimson et al., 2007). Deregulation in miRNA expression is one main cause of cancer (Hamano et al., 2011), cardiovascular (Small et al., 2010) and other diseases (Jiang et al., 2010). Aberrant miRNA expression has been well characterized in different oncology diseases and their progression (Cortez et al., 2012). Expression of some miRNAs is changed in development esophageal cancer (Fang et al., 2012), gastric cancer (Lu et al., 2011), colorectal cancer (Hamfjord et al., 2012), breast cancer (Hafez et al., 2012; Hanna et al., 2012) and other cancer types.

Many programs for miRNA binding sites prediction have been developed (Sethupathy et al., 2006). A search for miRNA sites are based on revealing of complementary seed (part of 5'-end miRNA site). The programs have algorithms for identification of RNA seed with length from 6 to 8 nt (Friedman et al., 2009; Lewis et al., 2003). Site searching for only short seed leads to the prediction of considerable quantity of false sites. Some programs take the advantage of seed conservation to assess site specificity. However, there are many nonconservative

*Corresponding author. E-mail: devolia18@mail.ru. Tel/Fax: +7 (727) 377 32 02.

seed-site of miRNAs already known (Baek et al., 2008). Some programs predict not perfectly complementary seed in the presence of 3'-compensatory site (Farh et al., 2005).

Most programs allow search for sites only in 3'-untranslated regions (3'UTR) (Maragkakis et al., 2011). It has been shown that miRNA sites may also occur in open reading frame (ORF) (Baek et al., 2008). Many researchers consider that miRNAs interact with mRNAs only in 3'UTR (Delay et al., 2011; Iorio et al., 2009; Isik et al., 2010; Kruger and Rehmsmeier, 2006; Satoh and Tabunoki, 2011), however, publications which describe miRNAs binding to mRNAs in 5'-untranslated regions (5'UTR) and coding sequences (CDS) are known (Duursma et al., 2008; Elcheva et al., 2009; Kulkarni et al., 2011; Moretti et al., 2010; Tsai et al., 2009).

MiRNA binding sites, predicted by computer method, which is founded by obligatory complementarity of 5'-seed region, are checked experimentally. However, some other miRNA binding sites, based on 3'-compensatory type, are not predicted by such method, but they are identified in experiments (Qin et al., 2010; Shirdel et al., 2011).

Program RNAhybrid (http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/) relies on thermodynamics and allows the search for 5'-dominant initial, 5'-seed-dominant and 3'-compensatory types of target sites (Garcna et al., 2011). An accurate identification of miRNA:mRNA pairs should promote considerable development of diagnostic methods of various cancer diseases.

The present research aimed to reveal miRNA binding sites with various parts of mRNA (5'UTR, CDS and 3'UTR) and defined nucleotide interaction features of miRNA:mRNA complexes. This study was realized for miRNAs involved in cancer and should be extended to other genes.

MATERIALS AND METHODS

Nucleotide sequences of miRNAs and their precursors were found from miRBase (http://www.mirbase.org). Nucleotide sequences mRNAs of all genes (Homo sapiens, Genome build 37.2.) were obtained from GenBank (http://www.ncbi.nlm.nih.gov). Then, miRNAFinder 2.2 (https://sites.google.com/site/malaheenee/home) was used to find miRNA origins (intergenic, exonic or intronic). A literature review of genes coding intronic miRNAs led to 51 oncogenes (Supplementary Table 1) that encode proteins participating in gastrointestinal and breast cancer. This information was collected according to articles from US National Library of Medicine, National Institutes of Health PubMed (http://www.ncbi.nlm.nih.gov/pubmed).

RNAhybrid 2.1 was run for all pairs (miRNA:mRNA complexes) and provided positions of potential binding sites miRNAs in mRNAs, free energy of miRNA binding sites (ΔG) and scheme of their interaction.

The script E-RNAhybrid (https://sites.google.com/site/malaheenee/home) computes the ratio $\Delta G/\Delta G_m$, p-value, equalizing coefficient and type of corresponding regions (5'UTR, CDS or 3'UTR), where miRNA site disposes. To achieve this, a quantitative criterion was defined. A ratio $\Delta G/\Delta G_m$ was measured as percentage, where ΔGm equals binding energy for miRNA with perfectly complementary nucleotide sequence computed. Number of binding sites in 5'UTR, CDS and 3'UTR was calculated as the number of sites (s) divided to nucleotide length (l) of this region and multiplied by 10^3 (s/l), that is, calculation per 1000 nucleotides. Significance degree (p-value) was estimated, that relies on ΔG and its standard deviation. The binding sites had ratio $\Delta G/\Delta G_m$ that equals or more than 80%.

RESULTS

784 intergenic, 686 intronic and 49 exonic miRNAs were found by program miRNAFinder 2.2. 471 genes encode 686 intronic pre-miRNAs. In these genes are situated one or more in-miRNAs. Among them, only 51 genes encode proteins participating in the development of gastrointestinal and breast cancer according to literature reviewed of all these genes (Supplementary Table 1).

Interaction of all studied miRNAs with mRNAs of 51 oncogenes was investigated by script E-RNAhybrid. The obtained results show a linkage between several mRNA of genes encoding in-miRNAs via other in-miRNAs (these miRNAs localize in genes that are not targets). Similar results can be found on three types of miRNA. It was established that all mRNA regions (5'UTR, CDS and 3'UTR) have miRNA binding sites. The density of sites and their distribution on these mRNA regions was revealed. Complementary type of miRNA binding sites was found.

Interaction of intergenic miRNA with mRNA oncogenes coding in-miRNA

Binding sites of 784 ig-miRNAs with 51 mRNAs were studied. Shares of binding sites for ig-miRNAs were 16.7, 57.3 and 26.0% in 5'UTR, CDS, 3'UTR of mRNAs, respectively. However, the density of binding sites in 5'UTR, CDS and 3'UTR mRNA was 2.32, 0.88 and 0.80 s/l, respectively, for these genes. Ig-miRNA binding sites with the best characteristics are represented in Tables 1 to 3. 5'UTR of some mRNAs can have several miRNA binding sites despite their short length (Table 1). For example, *NR2F2* gene had four binding sites; *EPCAM*, *HNF4A* and *SLIT3* genes have three sites and *EGFL7, HDAC4, LRP1* genes have two sites in 5'UTR. Several miRNAs have some targets of studied genes. MiR-4472 had binding sites in CDS of 7 genes (*AATK*, *HDAC4*, *LRP1*, *MAP2K4*, *NOTCH1*, *PTPRJ* and *SLIT3*), where *LRP1* gene has three sites and *NOTCH1* gene has two sites (Table 2). MiR-4456, miR-4455 and miR-4711-3p had binding sites in CDS six, five and four mRNAs accordingly. These miRNAs are important as they have binding sites with many mRNAs of studied genes. Characteristics of site location in 3'UTR were the same as in 5'UTR and CDS (Table 3). *BRE, IGF1R* and *MRE11A* genes were targets for miR-4456; *BBC3* and *NOTCH1* genes were targets for miR-1587. mRNA *BBC3*

Table 1. Positions of ig-miRNA binding sites in 5'UTR mRNA.

Target gene	miRNA	Position (nt)	Target gene	miRNA	Position (nt)	Target gene	miRNA	Position (nt)
AATK	miR-4417	45	*HDAC4*	miR-1268	80	*NR2F2*	miR-4253	537
ATF2	miR-4319	63	*HNF4A*	miR-302f	71	*NR2F2*	miR-4787-5p	1152
BRE	miR-4455	39	*HNF4A*	miR-1279	1	*PRKG1*	miR-1268	142
DCC	miR-568	524	*HNF4A*	miR-4307	76	*PTK2*	miR-3676	22
DTL	miR-4309	44	*HUWE1*	miR-4266	134	*PTPRJ*	miR-3195	288
EGFL7	miR-4289	4	*LRP1*	miR-4307	188	*SDCCAG8*	miR-4309	20
EGFL7	miR-1204	301	*LRP1*	miR-4472	99	*SLIT3*	miR-4507	61
EPCAM	miR-4456	112	*MCM7*	miR-4289	453	*SLIT3*	miR-4466	253
EPCAM	miR-4492	137	*MTUS1*	miR-3195	141	*SLIT3*	miR-4481	314
EPCAM	miR-4508	140	*NR2F2*	miR-4443	556	*SPATA13*	miR-4279	12
HDAC4	miR-3195	266	*NR2F2*	miR-1538	289	-	-	-

Table 2. Positions of ig-miRNA binding sites in CDS mRNA.

Target gene	miRNA	Position (nt)	Target gene	miRNA	Position (nt)	Target gene	miRNA	Position (nt)
AATK	miR-1538	1339	*EPHB2*	miR-4316	1293	*LRP1*	miR-4266	5093
AATK	miR-3195	1352	*EPHB2*	miR-4466	1334	*LRP1*	miR-4283	12932
AATK	miR-4265	3143	*ERBB4*	miR-4443	1940	*LRP1*	miR-4455	12015
AATK	miR-4472	1684	*EVL*	miR-3180	816	*LRP1*	miR-4456	2930
AATK	miR-4492	3345	*EVL*	miR-4711-3p	802	*LRP1*	miR-4472	13015
AATK	miR-4711-3p	2715	*FOXP1*	miR-320d	2225	*LRP1*	miR-4472	11274
AKT2	miR-4492	1370	*FOXP1*	miR-1279	2111	*LRP1*	miR-4472	2781
ANTXR1	miR-3141	2010	*FOXP1*	miR-4264	2506	*MAP2K4*	miR-4472	120
ANTXR1	miR-4711-3p	1772	*FOXP1*	miR-4266	1126	*MAP7D2*	miR-4417	1395
BBC3	miR-3665	410	*FOXP1*	miR-4327	2392	*MCM7*	miR-4483	1975
BBC3	miR-4278	884	*FOXP1*	miR-4736	1146	*MRE11A*	miR-302f	852
BBC3	miR-4466	319	*GIPR*	miR-659	1441	*MRE11A*	miR-520e	693
BBC3	miR-4483	744	*HDAC4*	miR-302e	2082	*NOTCH1*	miR-516a-3p	6964
BBC3	miR-4710	796	*HDAC4*	miR-1205	976	*NOTCH1*	miR-516b*	6964
BCAS1	miR-4307	871	*HDAC4*	miR-4472	2159	*NOTCH1*	miR-4283	1245
BID	miR-596	364	*HDAC4*	miR-4481	1868	*NOTCH1*	miR-4455	6381
BIRC7	miR-4456	372	*HDAC4*	miR-4483	2802	*NOTCH1*	miR-4455	6997
BRE	miR-3201	878	*HDAC4*	miR-4746-3p	982	*NOTCH1*	miR-4472	4308
CCAR1	miR-4463	841	*HNF4A*	miR-1204	278	*NOTCH1*	miR-4472	7474
CCAR1	miR-548ak	3272	*HNF4A*	miR-4456	465	*NOTCH1*	miR-4736	75
CDH13	miR-4455	2032	*HUWE1*	miR-320c	3048	*PRKG1*	miR-599	2933
DCC	miR-1207-3p	3190	*HUWE1*	miR-320d	3050	*PTPRJ*	miR-4472	2750
DCC	miR-4318	1684	*HUWE1*	miR-1279	10486	*SLIT3*	miR-302e	3310
DCC	miR-4325	2917	*HUWE1*	miR-4264	6886	*SLIT3*	miR-302f	3310
DCC	miR-4711-3p	3297	*HUWE1*	miR-4307	3373	*SLIT3*	miR-4472	3745
DMD	miR-4282	2958	*HUWE1*	miR-4443	4959	*SLIT3*	miR-4481	4872
DMD	miR-4493	5132	*HUWE1*	miR-4792	10624	*SLIT3*	miR-4492	1201
DMD	miR-4520a-5p	1878	*IGF1R*	miR-1268	1535	*SLIT3*	miR-4508	1201
DMD	miR-4520b-5p	1878	*IGF1R*	miR-4282	528	*SLIT3*	miR-4513	499
DMD	miR-4650-5p	5135	*IGF1R*	miR-4456	462	*SPATA13*	miR-1261	1854
DNMT3A	miR-302f	3075	*LFNG*	miR-1274b	175	*SPATA13*	miR-4266	2070
DNMT3A	miR-665	1570	*LFNG*	miR-3195	178	*SPATA13*	miR-4456	2864
DNMT3A	miR-3676	1221	*LFNG*	miR-4264	1154	*TNKS*	miR-4455	3891
DNMT3A	miR-4456	2223	*LRP1*	miR-132	4503	*TNKS*	miR-4465	1769
EIF4H	miR-645	705	*LRP1*	miR-1261	7704	*TNKS*	miR-4531	647
EPHB2	miR-4253	1087	*LRP1*	miR-4253	10550	-	-	-

Table 3. Positions of ig-miRNA binding sites in 3'UTR mRNA.

Target gene	miRNA	Position (nt)	Target gene	miRNA	Position (nt)	Target gene	miRNA	Position (nt)
AATK	miR-4472	4514	*EPHB2*	miR-4455	4792	*IGF1R*	miR-4455	8927
AKT2	miR-4316	2472	*ERBB4*	miR-4279	10500	*IGF1R*	miR-769-3p	7824
AKT2	miR-4418	3951	*ERBB4*	miR-568	11012	*IGF1R*	miR-4455	10709
AKT2	miR-4264	2336	*ERBB4*	miR-3123	4973	*LFNG*	miR-466	1269
BBC3	miR-4507	1000	*ERBB4*	miR-302f	5163	*LFNG*	miR-4318	1899
BBC3	miR-4505	1000	*ERBB4*	miR-513a-5p	4123	*LRP1*	miR-4328	14372
BBC3	miR-4497	965	*FBXW7*	miR-3674	3468	*MAP2K4*	miR-1827	3050
BBC3	miR-3676	1616	*FOXP1*	miR-466	5945	*MCM7*	miR-4466	2782
BBC3	miR-1587	1000	*FOXP1*	miR-4266	3217	*MRE11A*	miR-4456	3832
BBC3	miR-4279	1636	*HDAC4*	miR-4478	8388	*MTUS1*	miR-513a-5p	4652
BID	miR-543	2428	*HDAC4*	miR-4529-5p	6541	*MTUS1*	miR-4443	6051
BRE	miR-4456	1575	*HDAC4*	miR-4482	7344	*MTUS1*	miR-3168	4674
CDH13	miR-297	3603	*HDAC4*	miR-4710	8162	*MTUS1*	miR-513b	4652
DCC	miR-302f	6493	*HDAC4*	miR-4311	7024	*NOTCH1*	miR-1587	8133
EGFL7	miR-3130-5p	1454	*HDAC4*	miR-4328	4662	*NOTCH1*	miR-1275	9070
EIF4H	miR-4309	1364	*HNF4A*	miR-3934	1350	*SPATA13*	miR-876-3p	6126
EIF4H	miR-197	1627	*IGF1R*	miR-4456	7101	-	-	-

had six miRNA binding sites in 3'UTR and three of them were located in same place (miR-4505, miR-4507 and miR-1587). Such part of mRNA is very important for its regulation via several miRNA.

Interaction of intronic miRNAs with mRNAs genes coding in-miRNA

Oncogenes (51) are host genes and target genes for in-miRNAs. Majority of these in-miRNAs are encoded in intron. Five of the studied genes (*ATF2, BID, DMD, EVL* and *FOXP1*) encode intronic pre-miRNAs in their 3'UTRs and another five genes (*BBC3, EPHB2, NR2F2, PTPRJ* and *SPATA13*) in their 5'UTRs. Some genes encode one or more in-miRNA. For example, *AATK* gene encodes pre-miR-338, pre-miR-657 and pre-miR-1250; *HDAC4* gene encodes pre-miR-2467, pre-miR-4440 and pre-miR-4441.

Binding sites of 686 in-miRNAs with 51 mRNAs were studied (Table 4). The miRNA binding sites number are significantly different for the studied mRNAs. All parts of the studied mRNAs have similar binding ability to miRNA. In-miRNAs shares were 20.9, 39.5 and 39.5% sites in 5'UTR, CDS, 3'UTR, respectively. The average binding sites densities of 5'UTR, CDS and 3'UTR computed mRNAs were 1.81, 0.61 and 0.73 s/l, respectively. These results indicate that miRNAs can bind to 5'UTR, CDS and 3'UTR. *LRP1* gene has six miRNA binding sites in all parts of its mRNA: 5'UTR (miR-4274), CDS (miR-1273f, miR-4295, miR-4296 and miR-500b) and 3'UTR (miR-4297). *HUWE1* gene has miRNA binding sites (miR-4297, miR-548an and miR-598) only in CDS and *LFNG* gene had sites (miR-1224-3p, miR-500b) only in 3'UTR. Some mRNAs can be targets to in-miRNAs that encode in different pre-mRNA genes.

The linkages between genes via in-miRNAs are presented in Table 4. Some of the studied mRNAs belong to the host and target genes. For example, *BIRC7* encode in-miR-3196 and is target for in-miR-4292, in-miR-4257 and ig-miR-4456 at the same time. *IGF1R* gene is target for miR-1268B (*CCDC40*), miR-1273F (*SCP2*), miR-3173 (*DICER1*), miR-361 (*CHM*), miR-4292 (*C9orf86*), and miR-4297 (*EBF3*), where gene origin corresponding pre-miRNA is shown in brackets. These data show a linkage between different oncogenes encoded intronic miRNAs, where each mRNA is target for some miRNAs and encodes other miRNAs.

Interaction of exonic miRNAs with mRNA genes coding in-miRNA

Binding sites of 49 ex-miRNAs with 51 mRNAs have been studied. The linkages between genes via ex-miRNAs are presented in Table 5. Despite the small number of target genes, these data have proved that ex-miRNA can bind to CDS and 3'UTR, but there is no site in 5'UTR. Only seven out of all ex-miRNAs interact with seven target mRNAs. Shares of binding sites for ex-miRNAs were 75% sites in CDS and 25% sites in 3'UTR of mRNA. *DMD* gene had two miRNA binding sites (miR-4315 and miR-1306). miR-4315 has two target genes (*DMD* and *BIRC6*). *DNMT3A* gene encodes in-miR-1301 and is a target for ex-miR-4775. These data show a linkage between host genes of exonic and intronic miRNA.

Table 4. Positions of in-miRNA binding sites in 5'UTR, CDS, 3'UTR mRNA.

Target gene	miRNA	Position (nt)	Host gene
5'UTR			
BIRC7	miR-4292	60	*C9orf86*
EPCAM	miR-4317	45	*L3MBTL4*
EPCAM	miR-4753-3p	7	*ARID4B*
GIPR	miR-26a-1*	25	*CTDSPL*
HDAC4	miR-1268b	78	*CCDC40*
HDAC4	miR-1914*	540	*UCKL1*
HDAC4	miR-4296	53	*CTBP2*
HDAC4	miR-4296	159	*CTBP2*
IGF2	miR-1273f	81	*SCP2*
LRP1	miR-4274	428	*SORCS2*
PRKG1	miR-1268b	140	*CCDC40*
PTPRJ	miR-1238	208	*ATG4D*
CDS			
AATK	miR-4651	4029	*POR*
ABCA6	miR-3941	3053	*PLEKHA1*
BBC3	miR-4655-3p	598	*MAD1L1*
BID	miR-4285	432	*SH2B2*
BIRC7	miR-4257	745	*ADAMTSL4*
BRE	miR-4259	315	*CCDC19*
DMD	miR-4753-3p	11190	*ARID4B*
DNMT3A	miR-593	674	*SND1*
ERBB4	miR-3182	3863	*CDH13*
ERBB4	miR-4797-3p	3376	*DLG1*
EVL	miR-1322	522	*PINX1*
EVL	miR-4281	1001	*SNCB*
GIPR	let-7g	140	*WDR82*
GIPR	miR-1238	182	*ATG4D*
LRP1	miR-500b	1824	*CLCN5*
LRP1	miR-1273f	3301	*SCP2*
LRP1	miR-4295	6360	*VTI1A*
NOTCH1	miR-1271	1240	*ARL10*
NOTCH1	miR-3196	688	*BIRC7*
PTPRJ	miR-342-5p	459	*EVL*
SLIT2	miR-4446-3p	4223	*SIDT1*
SLIT3	miR-500b	426	*CLCN5*
SPATA13	miR-652	482	*TMEM164*
3'UTR			
AATK	miR-648	4498	*MICAL3*
AKT2	miR-3196	1865	*BIRC7*
AKT2	miR-4281	2711	*SNCB*
ANTXR1	miR-32*	5217	*C9orf5*
BBC3	miR-3156-3p	1667	*ANKRD30B*
CDH13	miR-574-5p	3603	*FAM114A1*
EBF3	miR-500b	2252	*RPS6KA1*
EBF3	miR-1976	2260	*CLCN5*
EGFL7	miR-3130-5p	1454	*FASTKD2*
EIF4H	miR-4263	1611	*BRE*
ERBB4	miR-483-3p	10500	*IGF2*

Table 4. Continued.

ERBB4	miR-877*	10504	*ABCF1*
HDAC4	miR-1289	6515	*FSTL4*
HDAC4	miR-4326	4606	*ARFGAP1*
IGF1R	miR-361-3p	7089	*CHM*
IGF1R	miR-1273f	7624	*SCP2*
IGF1R	miR-3173-5p	5679	*DICER1*
IGF1R	miR-4292	6166	*C9orf86*
IGF2	miR-1273f	1902	*SCP2*
IGF2	miR-3972	4533	*PADI3*
IGF2	miR-4263	3055	*BRE*
IGF2	miR-4296	4307	*CTBP2*
LFNG	miR-1224-3p	2007	*VWA5B2*

Table 5. Positions of ex-miRNA binding sites in CDS, 3'UTR mRNA.

Target gene	miRNA	Position (nt)	Host gene
CDS			
BIRC6	miR-4315	11064	*PLEKHM1*
BIRC7	miR-1825	373	*POFUT1*
DMD	miR-4315	6086	*PLEKHM1*
DMD	miR-1306	2819	*DGCR8*
GIPR	miR-671-5p	1236	*CHPF2*
HUWE1	miR-3652	6967	*HSP90B1*
3'UTR			
DNMT3A	miR-4775	3420	*CCNYL1*
MTUS1	miR-1306	5360	*DGCR8*

HDAC4 mRNA has three types of miRNA binding sites

As it is known, there exist three types of binding sites: 5'-dominant canonical, 5'-dominant seed only and 3'-compensatory sites (Betel et al., 2010). Canonical site has good (or perfect) complementarity at both the 5'- and 3'-ends of the miRNA with a specific bulge in the middle. Dominant seed site has perfect seed 5'-complementarity to the miRNA but poor 3'-complementarity. Compensatory site has a mismatch (wobble) in the 5'-seed region, but compensate through excellent complementarity at the 3'-end of miRNA.

5'-dominant canonical and 3'-compensatory types of binding sites were found in the obtained data. There is no 5'-dominant seed of binding sites, because such sites have low energy of binding site and are not reliable according to our criteria. In addition, complementary type of miRNA binding sites was found after the analysis of our results. It has perfect complementarity beginning with second nucleotide and completing the last but one nucleotide of miRNA in binding site. There was no bugle in the middle (Table 6). Such binding sites have high energy of binding site and are reliable according to our criteria.

HDAC4 mRNA had three types of miRNA binding sites. For example, 5'-dominant canonical sites were in-miR-4296 and ig-miR-3195; 3'-compensatory sites were ig-miR-4311 and in-miR-1914*; complementary sites were ig-miR-4478 and in-miR-1289 (Table 6). Main contribution to energy can include all parts of miRNA site, but not only 5'seed. *HDAC4* mRNA is a target for several intergenic and intronic miRNAs. In both cases, sites may be found in any region mRNA. The sites found have high hybridization energy due to significant amount of complementary nucleotides.

DISCUSSION

Expression may be regulated by miRNAs for more than a half of human genes (Friedman et al., 2009). The number of known human miRNAs constantly increases, as well as the number of their potential target genes. The establishment of target genes depends on the efficiency of their prediction by computer methods. Indeed, experimental verification of predicted target genes depends on the accuracy of prediction of miRNA sites

Table 6. Schematic representation of types of miRNA bindings sites in mRNA HDAC4.

5'UTR,53 ΔG=-33,5 ΔG/ΔGm=84 5'-d/c	5'UTR,266 ΔG=-38,8 ΔG/ΔGm=84 5'-d/c
mRNA 5' CGAGCCCGAGCCCGCGC 3' miR-4296 3' ACUCGGACUCGGGUGUA 5'	MRNA 5' AGCCCCGGCCCGGCGCC 3' miR-3195 3' UUGGGCCCGGGCCGCGC 5'
3'UTR,7024 ΔG=-28,8 ΔG/ΔG_m=80 3'-compn.	5'UTR,540 ΔG=-46,8 ΔG/ΔG_m=81 3'-compn.
mRNA 5' CACACUCGGCUCUUCUCC 3' miR-4311 3' GUGUGAGUCGAGAGAAAG 5'	MRNA 5' UCUCCCGGUGCGGGGCCCGCGCC 3' miR-1914* 3' GGAGGGUCACGCCCUGGG-GAGG 5'
3'UTR,8388 ΔG=-34,6 ΔG/ΔG_m=87 compl.	3'UTR,6515 ΔG=-36,7 ΔG/ΔG_m=80 compl.
mRNA 5' GUUCUUAGCUCGGCCUC 3' miR-4478 3' GAGGAGUCGAGUCGGAG 5'	mRNA 5' UGAGUGCAGAUUCUUGGAUUCAC 3' miR-1289 3' UUUUACGUCUAAGGACCUGAGGU 5'

Interaction energy (ΔG) is measured in kcal/mol. The ΔG/ΔG_m value is calculated in percents. Types of binding sites: 5'-d/c – 5'-dominant canonical, 3'-comp. – 3'-compensatory, compl. – complementary site.

and their characteristics considered in corresponding programs. Besides, the essential role of prediction of binding sites and their validation have limitation depending on the quality of prediction (Grimson et al., 2007). Functional regions of mRNAs were shown to be significantly heterogeneous according to the number of binding sites and to the location density of these miRNA sites (Issabekova et al., 2011).

In our work, interaction sites of 1519 miRNAs with 51 mRNA-targets are presented. Majority of these miRNAs were revealed recently (with miRNA index number 1000 and more) and were bad-studied. Total number of miRNA binding sites was found in the studied target genes to be 439. Some miRNAs have several binding sites with the same mRNA. Average number of binding sites was six miRNAs on mRNA. miRNA sites with high ΔG/ΔG_m ratio were selected with $P < 0.0004$. In-miRNAs and ig-miRNAs have similar properties. miRNAs with length 22 nt have distinct prevalence concerning others which have bigger or smaller length. A great number of in-miRNAs and ig-miRNAs have a high GC-content (50 to 55%). Most genes under study appear to have binding sites with miRNA. Therefore, it is possible that the expression of a significant part of human genes is under a regulation control by miRNAs. Intergenic miR-4472, miR-4456, miR-4455, miR-302f and miR-3195 had thirteen, nine, seven, six and five target genes accordingly. The expression of such miRNAs can repress translation of some genes participating in cancer development.

The shares of all binding sites (18.4, 58.2 and 23.4%) in 5'UTR, CDS, 3'UTR of the studied mRNA, respectively were found. Ig-miRNA binding sites were in 3.3 times more than it is to ig-miRNA. Thus, the density of binding sites in 5'UTR was 2.6 times more than it was in CDS and 2.8 times more than it was in 3'UTR. Such data indicate that miRNAs can interact with 5'UTR and CDS, and not only with 3'UTR. Some miRNAs have a high site density in 5'UTR. A significant part of the studied miRNAs interacts with mRNA of only one gene. Effect on such miRNAs would allow selection of modification of associated target genes expression. mRNAs of some genes are targets for ig-miRNAs and in-miRNAs. Majority intronic and exonic miRNAs express together with their host genes. These miRNA provide interaction of 51 host genes of in-miRNA with other host genes via in-miRNAs and ex-miRNAs. Changes in host gene expression lead to changes of intronic and exonic miRNA expression. These processes influence the target genes translation.

Conclusion

Computational prediction of miRNA binding sites is an important stage in the investigation of biological function miRNA. These data promote understanding the features of gene regulation on post-transcription level via miRNAs. The linkage between different genes encoding in-miRNAs and being targets for in-miRNAs was revealed. These genes participate in different cellular processes and reveal linkages between such genes participating in gastrointestinal cancer.

It was found that the percentage of ig-miRNA, ex-miRNA and in-miRNA sites are approximately identical in 5'UTR,

CDS, and 3'UTR. Approximately 2/3 of these sites bind to 5'UTR or CDS regions. This suggests that sites involved in a translation regulation by miRNAs are located not only in 3'UTR, but in 5'UTR and CDS regions.

According to these data, nearly 45% of miRNAs are intronic and their synthesis directly depends on transcription of corresponding host genes. In-miRNAs had no strong binding to mRNAs of the 51 studied oncogenes. Therefore, in-miRNAs do not inhibit the expression of host genes.

Three types of schemes of interaction between mRNA and miRNA was revealed. There were 5'-dominant canonical sites, 3'-compensatory sites and complementary sites. The primary contribution to energy can include not only 5'seed, but all the parts of miRNA site. Hence, all the parts of miRNA sites can bring contribution to the total energy of binding site (ΔG/ΔGm).

High miRNA concentration, that reduce expression of gene-suppressors can be identiec to stimulation of carcinogenesis. The effect of low miRNA concentration on oncogenes can cause the development of cancer. Changes that occur in miRNA concentration can possibly be used as medicines in anti-sense therapy of breast and gastrointestinal tract cancer.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Ministry of Education and Science, Republic of Kazakhstan. We thank Vladimir Khailenko for creating programs miRNAFinder 2.2 and E-RNAhybrid.

REFERENCES

Baek D, Villen J, Shin C, Camargo F, Gygi S, Bartel D (2008). The impact of microRNAs on protein output. Nature 455:64-71.

Betel D, Koppal A, Agius P, Sander C, Betel CL (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol. pp. 11-R90.

Cortez MA, Welsh JW, Calin GA (2012). Circulating microRNAs as noninvasive biomarkers in breast cancer. Cancer Res. 195:151-161.

Delay C, Calon F, Matthews P, Hébert S (2011). Alzheimer-specific variants in the 3'UTR of amyloid precursor protein affect microRNA function. Mol. Neurodegen. pp. 6-70.

Duursma A, Kedde M, Schrier M, Sage C, Agami R (2008). miR-148 targets human DNMT3b protein coding region. RNA 14:872-877.

Elcheva I, Goswami F, Noubissi K, Spiegelman V (2009). CRD-BP protects the coding region of *βTrCP1* mRNA from miR-183-mediated degradation. Mol. Cell. 35:240-246.

Fang Y, Fang D, Hu J (2012). MicroRNA and its roles in esophageal cancer. Med. Sci. Monit. 18:RA22-30.

Farh KK-H, Grimson A, Jan C, Lewis BP, Johnston WK (2005). The widespread impact of mammalian microRNAs on mRNA repression and evolution. Sci. 310:1817-1821.

Friedman RC, Farh KK, Burge CB, Bartel DP (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 19:92-105.

Garcna D, Baek D, Shin C, Bell G, Grimson A, Bartel D (2011). Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other miRNAs. Nat. Struct. Mol. Biol. 18:1139-1146.

Grimson A, Farh K, Johnston W, Garrett-Engele P, Lim L, Bartel D (2007). MicroRNA targeting specificity in mammals:determinants beyond seed pairing. Mol. Cell. 27:91-105.

Hafez MM, Hassan ZK, Zekri AR, Gaber AA, AL Rejaie SS, Sayed-Ahmed MM (2012). microRNAs and metastasis-related gene expression in Egyptian breast cancer patients. Asian. Pac. J. Cancer Prev. 13:591-598.

Hamano R, Ishii H, Miyata H, Doki Y, Mori M (2011). Role of microRNAs in solid tumors. J. Nucleic Acids Invest. pp. 2-e2.

Hamfjord J, Stan eland AM, Hughes T, Skrede ML, Tveit KM, Ikdahl T (2012). Differential expression of miRNAs in colorectal cancer: Comparison of paired tumor tissue and adjacent normal mucosa using high-throughput sequencing. PLoS One. 7-e34150.

Hanna JA, Wimberly H, Kumar S, Slack F, Agarwal S, Rimm DL (2012). Quantitative analysis of microRNAs in tissue microarrays by *in situ* hybridization. Biotechniques 52:235-245.

Ibanez-Ventoso C, Vora M, Driscoll M (2008). Sequence relationships among *C. elegans, D. melanogaster* and human microRNAs highlight the extensive conservation of microRNAs in biology. PLoS ONE. pp. 3-e2818.

Iorio MV, Croce CM (2009). MicroRNAs in cancer:Small molecules with a huge impact. J. Clin. Oncol. 27:5848-5856.

Isik M, Korswagen H, Berezikov E (2010). Expression patterns of intronic microRNAs in *Caenorhabditis elegans.* Silence 1:1758–907X.

Issabekova AS, Berillo OA, Khailenko VA, Atambayeva SA, Regnier M, Ivachshenko AT (2011). Characteristics of intronic and intergenic human miRNAs and features of their interaction with mRNA. World Acad. Sci. Eng. Technol. 59:63-66.

Jiang Q, Hao Y, Wang G, Juan L, Zhang T, Teng M (2010). Prioritization of disease microRNAs through a human phenome-microRNAome network. BMC Syst. Biol. pp. 4-S2.

Kim YK, Kim VN (2007). Processing of intronic microRNAs. EMBO J. 26:775-783.

Kruger J, Rehmsmeier M (2006). RNAhybrid:microRNA target prediction easy, fast and flexible. Nucleic Acids Res. 34:W451-454.

Kulkarni S, Savan R, Qi Y (2011). Differential microRNA regulation of *HLA-C* expression and its association with HIV control. Nature 472:495-499.

Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003). Prediction of mammalian microRNA targets. Cell 115:787-798.

Lu L, Li Y, Li S (2011). Computational identification of potential microRNA network biomarkers for the progression stages of gastric cancer. Int. J. Data. Min. Bioinform. 5:519-531.

Maragkakis M, Vergoulis T, Alexiou P, Reczko M, Plomaritou K, Gousis M (2011). DIANA-microT web server:elucidating microRNA functions through target prediction. Nucleic. Acids Res. 37:W27-276.

Moretti F, Thermann R, Hentze M (2010). Mechanism of translational regulation by miR-2 from sites in the 5' untranslated region or the open reading frame. RNA 16:2493-2502.

Pillai RS, Bhattacharyya ST, Filipowicz W (2007). Repression of protein synthesis by miRNAs:how many mechanisms? Trends Cell Biol. 17:118-126.

Qin W, Shi Y, Zhao B, Yao C, Jin L, Ma J, Jin Y (2010). miR-24 regulates apoptosis by targeting the open reading frame (ORF) region of FAF1 in cancer cells. PLoS One. pp. 5-e9429.

Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A (2004). Identification of mammalian microRNA host genes and transcription units. Genome Res. 14:1902-1910.

Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N (2008). Widespread changes in protein synthesis induced by microRNAs. Nature 455:58-63.

Sethupathy P, Megraw M, Hatzigeorgiou AG (2006). A guide through present computational approaches for the identification of mammalian microRNA targets. Nat. Methods 3:881-886.

Shirdel EA, Xie W, Mak TW, Jurisica I (2011). NAViGaTing the micronome–using multiple microRNA prediction databases to identify signaling pathway-associated microRNAs. PLoS ONE. pp. 6-e17429.

Small EM, Frost RJ, Olson EN (2010). MicroRNAs add a new dimension to cardiovascular disease. Circulation 121:1022-1032.

Tsai NP, Lin YL, Wei LN (2009). MicroRNA mir-346 targets the 5'UTR of *RIP140* mRNA and up-regulates its protein expression. Biochem. J. 424:411–418.

APPENDIX

Supplementary Table 1. Fifty-0ne (51) Oncogenes coding intronic miRNAs.

Gene	Full name of gene	References
AATK	Apoptosis-associated tyrosine kinase	Lee et al. (2006)
ABCA6	ATP-binding cassette, sub-family A (ABC1), member 6	Hlavata et al. (2012)
ABCF1	ATP-binding cassette, sub-family F (GCN20), member 1	Hlavata et al. (2012)
AKT2	V-akt murine thymoma viral oncogene homolog 2	Li et al. (2011)
ANTXR1	Anthrax toxin receptor 1	Rmali et al. (2004)
ATF2	Activating transcription factor 2	Asting et al. (2011)
BBC3	BCL2 binding component 3	Ray et al. (2011)
BCAS1	Breast carcinoma amplified sequence 1	Correa et al. (2000)
BID	BH3 interacting domain death agonist	Ma et al. (2012)
BIRC6	Baculoviral IAP repeat-containing 6	Van et al. (2011)
BIRC7	Baculoviral IAP repeat containing 7	Oh et al. (2011)
BRE	Brain and reproductive organ-expressed	Yao et al. (2012)
CCAR1	Cell division cycle and apoptosis regulator 1	Ou et al. (2009)
CDH13	Cadherin 13, H-cadherin (heart)	Hibi et al. (2004)
DCC	Deleted in colorectal carcinoma	Derks et al. (2009)
DMD	Dystrophin	Edward et al. (2004)
DNMT3A	DNA (cytosine-5-)-methyltransferase 3 alpha	Schneider-Stock et al. (2005)
DTL	Denticleless homolog (Drosophila)	Baraniskin et al. (2012)
EBF3	Early B-cell factor 3	Kim and Kim (2007)
EGFL7	EGF-like-domain, multiple 7	Díaz et al. (2008)
EIF4H	Eukaryotic translation initiation factor 4H	Wu et al. (2011)
EPCAM	Epithelial cell adhesion molecule	Tao et al. (2012)
EPHB2	EPH receptor B2	Herath et al. (2012)
ERBB4	V-erb-a erythroblastic leukemia viral oncogene homolog 4	Frey et al. (2010)
EVL	Enah/Vasp-like	Yi et al. (2011)
FBXW7	F-box and WD repeat domain containing 7	Milne et al. (2010)
FOXP1	Forkhead box P1	Adams et al. (2009)
GIPR	Gastric inhibitory polypeptide receptor	Prabakaran et al. (2010)
HDAC4	Histone deacetylase 4	Jin et al. (2012)
LFNG	O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase	Reedijk et al. (2008).
HNF4A	Hepatocyte nuclear factor 4, alpha	Weltmeier et al. (2011)
HUWE1	HECT, UBA and WWE domain containing 1	Zhao et al. (2008)
IGF1R	Insulin-like growth factor 1 receptor	Pennarun et al. (2011)
IGF2	Insulin-like growth factor 2 (somatomedin A)	Hoyo et al. (2012)
LRP1	Low density lipoprotein receptor-related protein 1	Toquet et al. (2007)
MAP2K4	Mitogen-activated protein kinase kinase 4	Soh et al. (2001)
MAP7D2	MAP7 domain containing 2	Sjöblom et al. (2006)
MCM7	Minichromosome maintenance complex component 10	Nishihara et al. (2009)
MRE11A	MRE11 meiotic recombination 11 homolog A (*Saccharomyces cerevisiae*)	Takemura et al. (2006)
MTUS1	Microtubule associated tumor suppressor 1	Zuern et al. (2010)
NOTCH1	Notch 1p	Zhang et al. (2010)
NR2F2	Nuclear receptor subfamily 2, group F, member 2	Corneza et al. (2008)
PRKG1	Protein kinase, cGMP-dependent, type I	Savas et al. (2010)
PTK2	PTK2 protein tyrosine kinase 2	Leve et al. (2011)
PTPRJ	Protein tyrosine phosphatase, receptor type, J	Jeon et al. (2009)
SDCCAG8	Serologically defined colon cancer antigen 8	Kamio et al. (2010)
SLIT2	Slit homolog 2 (Drosophila)	Li-F et al. (2012)
SLIT3	Slit homolog 3 (Drosophila)	Zhu et al. (2011)
SPATA13	Spermatogenesis associated 13	Kawasaki et al. (2007)
TNFAIP6	Tumor necrosis factor, alpha-induced protein 6	Hamm et al. (2008)
TNKS	TRF1-interacting ankyrin-related ADP-ribose polymerase	Shebzukhov et al. (2008)