Molecules and Cells

Identification and Expression of Equine MER-Derived miRNAs

Jeong-An Gim[1,2,3], and Heui-Soo Kim[1,2,*]

[1]Department of Biological Sciences, College of Natural Sciences, [2]Genetic Engineering Institute, Pusan National University, Busan 46241, Korea, [3]The Genomics Institute, Life Sciences Department, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Korea
*Correspondence: khs307@pusan.ac.kr
http://dx.doi.org/10.14348/molcells.2017.2295
www.molcells.org

MicroRNAs (miRNAs) are single-stranded, small RNAs (21-23 nucleotides) that function in gene silencing and translational inhibition via the RNA interference mechanism. Most miRNAs originate from host genomic regions, such as intergenic regions, introns, exons, and transposable elements (TEs). Here, we focused on the palindromic structure of medium reiteration frequencies (MERs), which are similar to precursor miRNAs. Five MER consensus sequences (MER5A1, MER53, MER81, MER91C, and MER117) were matched with paralogous transcripts predicted to be precursor miRNAs in the horse genome (equCab2) and located in either intergenic regions or introns. The MER5A1, MER53, and MER91C sequences obtained from RepeatMasker were matched with the eca-miR-544b, eca-miR-1302, and eca-miR-652 precursor sequences derived from Ensembl transcript database, respectively. Each precursor form was anticipated to yield two mature forms, and we confirmed miRNA expression in six different tissues (cerebrum, cerebellum, lung, spleen, adrenal gland, and duodenum) of one thoroughbred horse. MER5A1-derived miRNAs generally showed significantly higher expression in the lung than in other tissues. MER91C-derived miRNA-5p also showed significantly higher expression in the duodenum than in other tissues (cerebellum, lung, spleen, and adrenal gland). The MER117-overlapped expressed sequence tag generated polycistronic miRNAs, which showed higher expression in the duodenum than other tissues. These data indicate that horse MER transposons encode miRNAs that are expressed in several tissues and are thought to have biological functions.

Keywords: medium reiteration frequency transposon, MicroRNA, palindromic structure, thoroughbred horse

INTRODUCTION

MicroRNAs (miRNAs) are non-coding small RNAs (21-23 nucleotides) that play a key role in inhibiting target genes by binding their 3′ UTRs, which are complementary to the miRNAs' seed regions (Bartel, 2009; Chen and Rajewsky, 2007; Shukla et al., 2011). These sequences are known to influence various biological processes in numerous animals and plants (Bartel, 2009). miRNAs undergo a series of changes from the transcript stage to maturity. First, primary miRNAs are transcribed by polymerase II in the nucleus (Lee et al., 2004). Next, they are cleaved into precursor miRNAs (70 nucleotides), with a double-stranded hairpin structure, by Drosha. Finally, precursor miRNAs are exported into the cytoplasm and converted to mature miRNAs by Dicer (Lund and Dahlberg, 2006). In a genome, miRNA-encoding genes are located in various characterized regions such as intergenic regions, exons, introns, and transposable elements (TEs) (Piriyapongsa et al., 2007). TEs contribute to miRNA sequences in various ways. For example, two adjacent TEs, as well as genomic sequences and one TE, yield one precursor miRNA. Palindrome sequences containing TEs such as miniature inverted repeat transposable elements (MITE) and medium reiteration frequency (MER) generate one precursor

Received 4 December, 2016; revised 6 March, 2017; accepted 15 March, 2017; published online 21 March, 2017

eISSN: 0219-1032

form and two mature miRNAs (Ahn et al., 2013; Yuan et al., 2010). Two mature miRNAs are derived from one precursor miRNA and present asymmetric expression or selective expression patterns (Hutvagner, 2005; Ruike et al., 2008).

MERs are non-autonomous mammalian DNA transposons (Jurka et al., 1996). A total of 240,000 copies of MERs are present in humans, accounting for 2.4% of the human genome (Lander et al., 2001). Most of these sequences were fossilized, such as long terminal repeats (LTRs) or endogenous retroviruses (ERVs) (Coffin, 2004; Smit, 1993), although some cause diseases in host organisms (Balada et al., 2010; Lamprecht et al., 2010). The genomes of primates, rodentia, and lagomorpha contain MER repeats inserted before the *Alu* family genes (Jurka et al., 1996). MITEs are a truncated form of autonomous DNA transposons (Feschotte et al., 2003; Zhang et al., 2000). Notably, MER or MITE sequences have palindromic structures, which form hairpins, similar to precursor miRNAs. Therefore, palindromic MERs or MITEs can give rise to miRNAs when they are located in transcriptionally active genomic regions. The MER53 transposon is predicted to give rise to miRNAs by forming a palindromic structure in mammalian genomes (Yuan et al., 2010). The expression of MER-derived miRNAs was experimentally confirmed in human cell lines (Ahn et al., 2013). MITE-derived miRNAs were predicted (Piriyapongsa and Jordan, 2007), and their expression confirmed in rice (Ou-Yang et al., 2013).

The reference genome sequence for horse was determined in 2007, and the horse genetic map has been available since 2009 (Wade et al., 2009). The horse genome contains numerous genetic traits linked to racing ability, such as single-nucleotide polymorphisms (SNPs), specific transcripts, and copy number variation (CNV) (Doan et al., 2012; Hill et al., 2010; Petersen et al., 2013). Based on the horse reference genome sequence, a total of 1397 horse miRNAs have been identified in the miRBase database (Kozomara and Griffiths-Jones, 2014). Subsequently, the functional study and characterization of horse miRNAs could be available (Kim et al., 2014; Zhou et al., 2009). However, horse miRNAs have not been as relatively well studied as other genetic traits, such as SNPs or CNV (Doan et al., 2012; Gim et al., 2015; Hill et al., 2010).

Many identified expressed sequence tags (ESTs) are available in databases, and ESTs can encode miRNAs and miRNA clusters (Smalheiser, 2003). The miRNA cluster transcripts are formed from ESTs, which show polycistronic expression patterns containing multiple, distinct loops separated by Drosha and Dicer. Forty-eight percent of all human mature miRNAs are derived from primary miRNAs of polycistronic transcripts (Altuvia et al., 2005). A well-known miRNA cluster, previously known as *C13orf25*, miR-17-92, can express six precursor miRNAs and is more highly expressed in B-cell lymphoma cells derived from both cell lines or patients than in normal individuals (Ota et al., 2004). In the present study, we predicted and experimentally confirmed the expression of equine MER-derived miRNAs, including a miRNA cluster in a previously identified EST. We searched consensus MER elements predicted to have palindromic structures and compared the MER sequences to horse transcripts. Next, we experimentally confirmed the expression of horse transcripts showing overlap with MER sequences. Finally, we compared the expression levels of MER-derived miRNAs in various tissues of one thoroughbred horse.

MATERIALS AND METHODS

Prediction of MER-derived miRNA sequences by bioinformatics tools

MER consensus sequences that form palindromic sequences were downloaded from the Repbase database version 21.05 of the Genetic Information Research Institute (http://www.girinst.org/repbase/) (Jurka et al., 2005). The genomic locations of MER sequences were confirmed in the horse genome (Broad/equCab2), using UCSC Genome Browser (http://genome.ucsc.edu), and their sequences were analyzed to identify regions that overlapped with transcripts, using Ensembl 84 (http://www.ensembl.org). In the UCSC table Browser, MER sequences and their genomic regions were compared using RepeatMasker 3.3 (http://www.repeatmasker.org). The sequences of horse precursor miRNAs and mature miRNAs were downloaded from miRBase v21.0 (http://www.mirbase.org), then their sequences were mapped and localized to the horse genome (equCab2).

MER consensus sequences forming palindromic sequences were predicted by RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). To identify potential MER-derived miRNA genes, we predicted the miRNA precursor forms originating from MER elements, using the miRNAfold tool (http://evryrna.ibisc.univ-evry.fr/miRNAFold/) for the horse genome. Next, mature miRNAs derived from MER precursor forms (selected for having minimum free energies) were predicted using MatureBayes program (http://mirna.imbb.forth.gr/MatureBayes.html). MatureBayes is a tool to predict mature miRNA from miRNA precursor palindromic hairpin structures. Mature forms were confirmed by RT-PCR, we used these miRNAs for expression profiling in six tissues of one horse. Strategy for transcripts analysis in MER-derived horse miRNAs were described in Supplementary Fig. 1.

RNA samples

The animal protocol and sample extraction method in this study were reviewed by the Pusan National University-Institutional Animal Care and Use Committee (PNU-IACUC) for the ethical procedures and scientific care (approval number PNU-2013-0411). Small RNA from 100 mg of different tissues from one thoroughbred horse, sacrificed for biological research, was extracted from the cerebrum, cerebellum, lung, spleen, adrenal gland, and duodenum, using Hybrid-R™ miRNA (GeneAll Biotechnology Co., Ltd., Korea) according to the manufacturer's protocol. Hybrid-R™ kit is available to simultaneously isolate large RNA and small RNA, and to separate total RNA and small RNA by using small RNA specific binding column tube. In order to exclude short RNA sequences from degreaded longer transcripts, a portion (2-3 μl) of total RNA was loaded onto an agarose gel, then only detected 28S and 18S samples were used.

Next, the turbo DNA-free™ kit (Ambion, USA) was used according to the manufacturer's protocol to remove DNA contamination from the small RNAs. A portion (2-3 μl) of isolated

small RNA was loaded onto an agarose gel, and only samples in which small-RNA (< 200 nt) was detected were used in subsequent experiments. We measured the small RNA concentrations, which ranged between 130-180 ng/μl in the samples. Each small RNA sample was quantitated to 500 ng using a NanoDrop® ND-1000 UV-vis Spectrophotometer (NanoDrop Technologies, USA). After quantifying small RNA, adenine was added to the small RNA transcripts at their 3′ ends, using poly (A) tailing kits (Ambion). Samples were incubated in E-PAP buffer at 37℃ for 30 min for the poly (A) tailing reactions. After the reactions, we performed reverse transcription reactions using Moloney murine leukemia virus reverse transcriptase with an RNase inhibitor (Promega, USA) at an annealing temperature of 42℃. To select target miRNAs in the reverse transcription reactions, we used an oligo-dT adaptor (5′-CTGTGAATGCTG CGACTACGAT-18dTs-3′) according to the methods of previous study (Fu et al., 2006). The miRNA sequence is used as forward primer, and adaptor sequences excluded poly T in the 3′ terminal is used as reverse primer (Supplementary Table S1).

Quantitative RT-PCR amplification

RT-PCR was conducted in a thermocycler (Eppendorf, Germany) and SYBR green quantitative RT-PCR was conducted using a Roter Gene Q (Qiagen, Germany). Each reaction was performed in a final volume of 20 μl containing 1 μl of cDNA sample as a template, 10 μl of Quantitect® SYBR® Green PCR MasterMix (Qiagen), 7 μl of nuclease-free water, and 1 μl each of sense primer and antisense primer (each 10 pmol). RT-PCR of genes was performed at 94℃ for 5 min, 35 cycles of 94℃ for 30 s, 55℃ for 30 s, and 72℃ for 30 s, and a final elongation at 72℃ for 7 min. Next, quantitative RT-PCR of genes was performed at 94℃ for 15 min, 55 cycles of 94℃ for 10 s, 55℃ for 15 s, and 72℃ for 15 s using the primers shown in Supplementary Table 1. Melting curves were determined by increasing the temperature from 55℃ to 99℃ over a period of 30 s. Each sample was amplified in triplicate to confirm reproducibility. U6 was used as a positive control in RT-PCR analysis and as a standard control for normalization in qRT-PCR. The relative quantitative levels of MER-derived miRNAs were normalized to U6 small-RNA expression levels. The expression values were determined by the $2^{-\Delta C}{}_t$ method, where $\Delta C_t = C_{t(miRNA)}-C_{t(U6)}$.

RESULTS

Prediction of MER-derived miRNAs

Some of consensus sequences of MER have palindromic structures, which are similar to those of precursor miRNAs. A

Table 1. Palindromic MER consensus sequences in horse genome

Repeat name	Repeat length (bp)	Structure	Number of paralogs in horse genome (equCab2)
MER5A1	160	Palindrome	196
MER53	189	Palindrome	202
MER81	114	Palindrome	161
MER91C	140	Palindrome	168
MER117	197	Palindrome	160

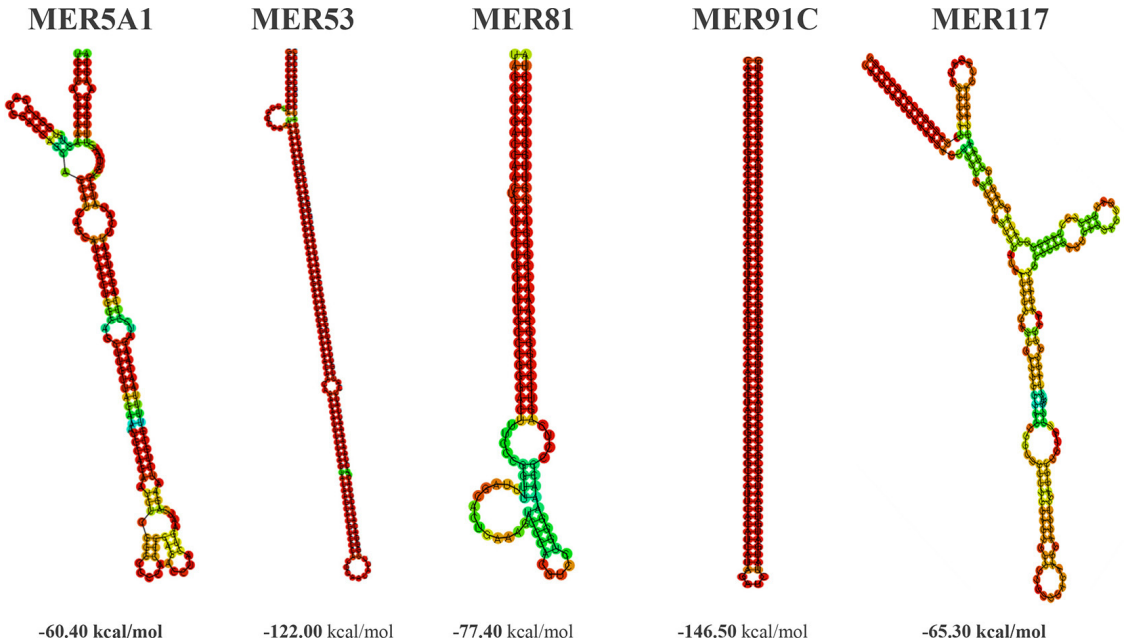

Fig. 1. MER consensus sequences matched with horse transcripts from Repbase version 21.05. MER5A1 was identified in Eutheria. MER53, MER81, MER91C, and MER117 were identified in human.

Table 2. Comparison between miRBase database sequences and palindromic MER-derived transcripts. Matched precursor sequences and E-values were the results of the miRBase BLAST seaching.

Repeat name	Transcript assession no.	Position	Strand	Transcript length (bp)	Matched precursor sequence	E-value	Genomic region (Gene or transcript accession no.)
MER5A1	ENSECAT00000029344.1	chr14:89,097,198-89,097,294	+	97	eca-miR-544b	9.00E-15	Intergenic
	ENSECAT00000029411.1	chr21:31,003,689-31,003,780	-	92	eca-miR-544b	6.00E-06	Intron (ADAMTS12)
	ENSECAT00000029517.1	chrX:10,032,472-10,032,568	+	97	eca-miR-544b	5.00E-14	Intergenic
	ENSECAT00000029551.1	chr15:74,838,110-74,838,184	+	93	eca-miR-544b	1.00E-14	Intergenic
	ENSECAT00000029718.1	chr11:53,066,395-53,066,487	-	93	eca-miR-544b	4.00E-08	Intron (MYH1, MYH2, MYH4, MYH6, MYH7, MYH7B, MYH8, MYH13)
	ENSECAT00000029813.1	chrX:19,077,964-19,078,058	-	95	eca-miR-544b	3.00E-09	Intron (POLA)
MER53	ENSECAT00000027451.1	chr16:58,884,211-58,884,360	-	150	eca-miR-1302c-5	1.00E-15	Intron (JL635408)
	ENSECAT00000027640.1	chr14:29,187,011-29,187,112	+	102	eca-miR-1302-1	6.00E-08	Intron (HTR4, GU289397, AY647163)
MER81	ENSECAT00000029217.1	chr10:36,160,777-36,160,851	+	75	Not detected		Intron (IBTK), UTR (JL626932)
	ENSECAT00000029221.1	chrX:50,972,151-50,972,233	-	83	Not detected		Intergenic
	ENSECAT00000029305.1	chr21:8,851,113-8,851,188	+	76	Not detected		Intron (JL635478, JL639598, JL624325)
	ENSECAT00000029634.1	chr2:37,821,443-37,821,510	+	68	eca-miR-8990	Not significant	Intergenic
MER91C	ENSECAT00000027964.1	chrX:86,921,316-86,921,413	+	98	eca-miR-652	2.00E-34	Intergenic
MER117	BM734541.1	chr23:7,068,821-7,073,667	-	655	Not detected		Intron (JL633484, JL641603)

total of 238 MER elements were screened from the Repbase database, and 24 MER elements were identified to have palindromic structures. In the 24 MER elements, only five MER-derived mature miRNAs were confirmed as RT-PCR (Supplementary Fig. 2). The others, 19 MER elements were not matched with horse transcripts, or horse genome (equCab2). Therefore, we selected five palindromic MER consensus sequences that matched with horse transcripts (Fig. 1). Each five actual MER sequences is provided as Supplementary Table 2. Numerous paralogs families matched the MER consensus sequences in horse genomes, based on UCSC BLAT searching (Table 1). Six transcripts overlapped with MER5A1 consensus sequences and were matched with eca-miR-544b (Hubbard et al., 2002; Jurka, 2000; Smit et al., 2004). The E-values of all transcripts were below 6.00E-06 based on miRBase BLAST searching and the transcripts contained constructed palindromic structures. Three of six miRNAs were located in the intronic region, whereas the others were located in the intergenic region. MER53 consensus sequences were matched to two transcripts related to eca-miR-1302-1 and eca-miR-1302c-5. Both miRNAs were located in the intron. Their E-values were 6.00E-08 and 1.00E-15, respectively. One transcript matched a MER91C consensus sequence in the horse transcript and express eca-miR-652 (E-value = 2.00E-34). MER117 was matched to horse EST BM734541.1 (Table 2). The BM734541.1 transcript was located in the intron of two transcripts, was predicted to contain many palindromic structures, and may generate 10 precursor miRNAs as putative polycistronic clustered miRNAs (http://www.ncbi.nlm.nih.gov/nucest/BM734541).

Validation and quantitative expression patterns of MER-derived miRNAs in a thoroughbred horse

To confirm that the mature miRNAs were derived from MER

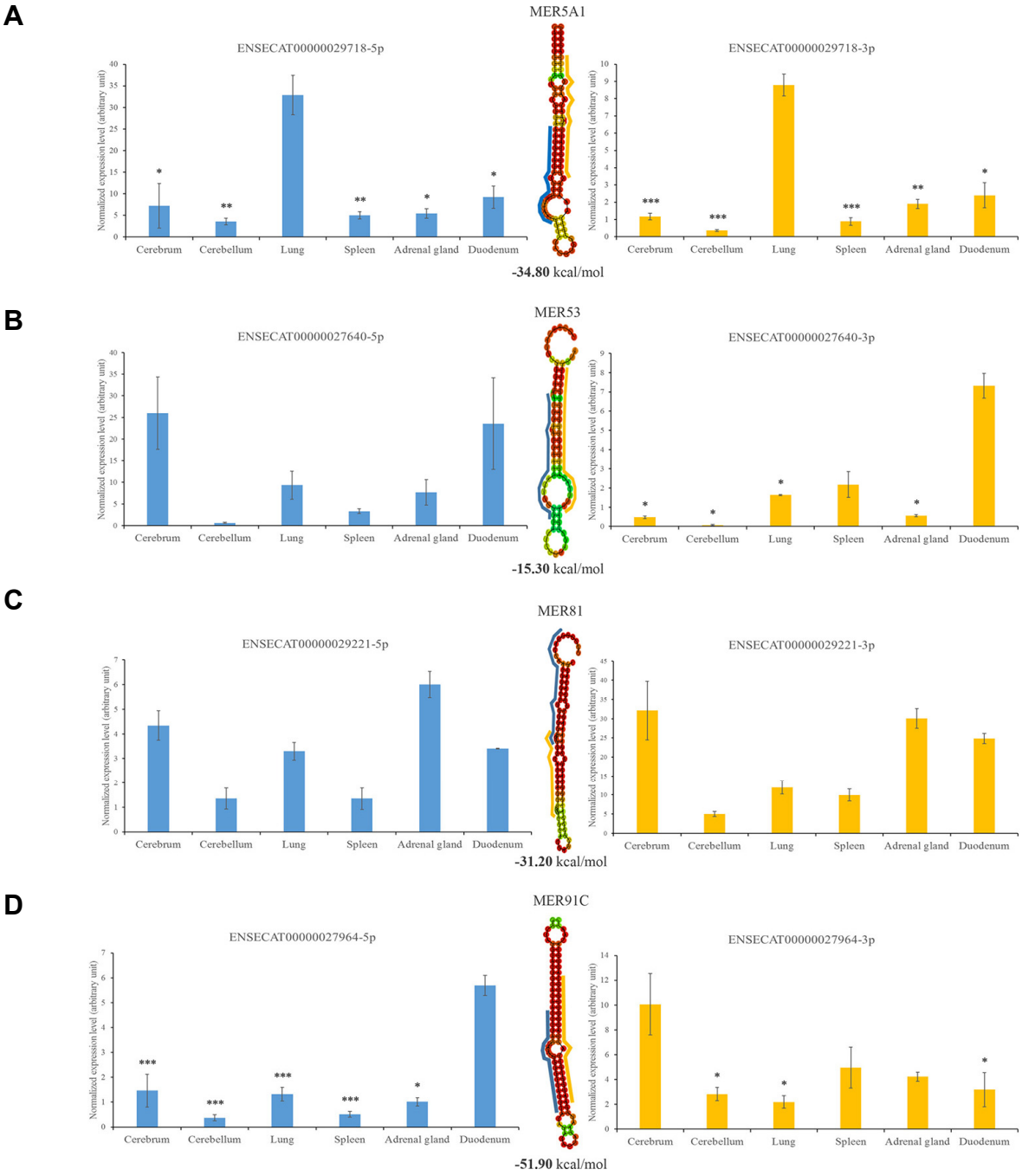

Fig. 2. Quantitative expression patterns of MER-derived transcript miRNAs. (A) MER5A1-derived transcript ENSECAT00000029718 5p (left) and 3p (right) miRNA. (B) MER53-derived transcript ENSECAT00000027640 5p (left) and 3p (right) miRNA. (C) MER81-derived transcript ENSECAT00000029221 5p (left) and 3p (right) miRNA. (D) MER91C-derived transcript ENSECAT00000027964 5p (left) and 3p (right) miRNA. Each samples was examined in triplicate (Bar: mean; Whisker: standard deviation). Paired Student's *t*-test was performed to verify statistical significance between the expression levels of two samples (*$p < 0.05$; **$p < 0.01$; ***$p < 0.005$). Two MER5A1-derived miRNAs were compared lung (A), MER91C-derived 5p miRNAs were compared duodenum (D, left), MER91C-derived 3p miRNAs were compared cerebrum (D, right), with other tissues.

transposons, all mature miRNAs were extracted from the tissues of one thoroughbred horse. Adenine and adaptor sequences were added to the 3′ end of each mature miRNA (see "Materials and Methods"). Next, RT-PCR was performed to identify the expression of 24 MER transposon-derived miRNA, and only five miRNAs were detected. Then, we per-

formed qRT-PCR of five miRNAs in the six tissues of one horse (Cerebrum, Cerebellum, Lung, Spleen, Adrenal gland, Duodenum), and MER-derived miRNA showed tissue-specific expression patterns. As shown in Fig. 2, two mature miRNAs derived from one primary miRNA showed similar expression patterns. The MER5A1-derived ENSECAT00000029718 transcript yielded two miRNAs showing lung-dominant expression patterns than other tissues (Fig. 2A). The MER53-derived ENSECAT00000027640 transcript yielded two miRNAs showing different expression patterns between the -5p and -3p forms. The -5p miRNA showed higher expression value in the cerebrum and duodenum than in other tissues, but no significant expression patterns were detected (Fig. 2B, left). The -3p miRNA showed dominant expression patterns in the duodenum, compared to cerebrum, cerebellum, lung, and adrenal gland (Fig. 2B, right). The MER81-derived ENSECAT00000029221 transcript yielded two miRNAs, that were no significant expression patterns (Fig. 2C). The MER91C-derived ENSECAT00000027964 transcript yielded two miRNAs, which showed different expression patterns. The -5p miRNA was enriched in the duodenum than other tissues (Fig. 2D, left), whereas the -3p miRNA was enriched in the cerebrum than cerebellum, lung, and duodenum (Fig. 2D, right).

Identification of MER117-derived polycistronic miRNA cluster

We confirmed that the MER117 sequence overlapped with horse EST of BM734541.1, the transcript of which was predicted to contain many palindromic structures. The BM734541.1 transcript was predicted to generate five precursor forms of polycistronic miRNAs. As shown in Fig. 3, the expression patterns of five different mature forms of the BM734541.1 were confirmed. We detected these mature forms in six different tissues, indicating that total mature forms except EST-2-derived mature form have significantly low expression patterns in the cerebellum than adrenal gland (Figs. 3A, 3C-3E). Interestingly, EST-1-derived mature form was dominantly expressed in other tissues than cerebellum, except cerebrum (Fig. 3A). Only adrenal gland had higher expression pattern than cerebellum in the EST-4-derived mature form (Fig. 3D). The location of mature miRNAs in the BM734541.1 transcript is indicated for each corresponding region in Fig. 3F.

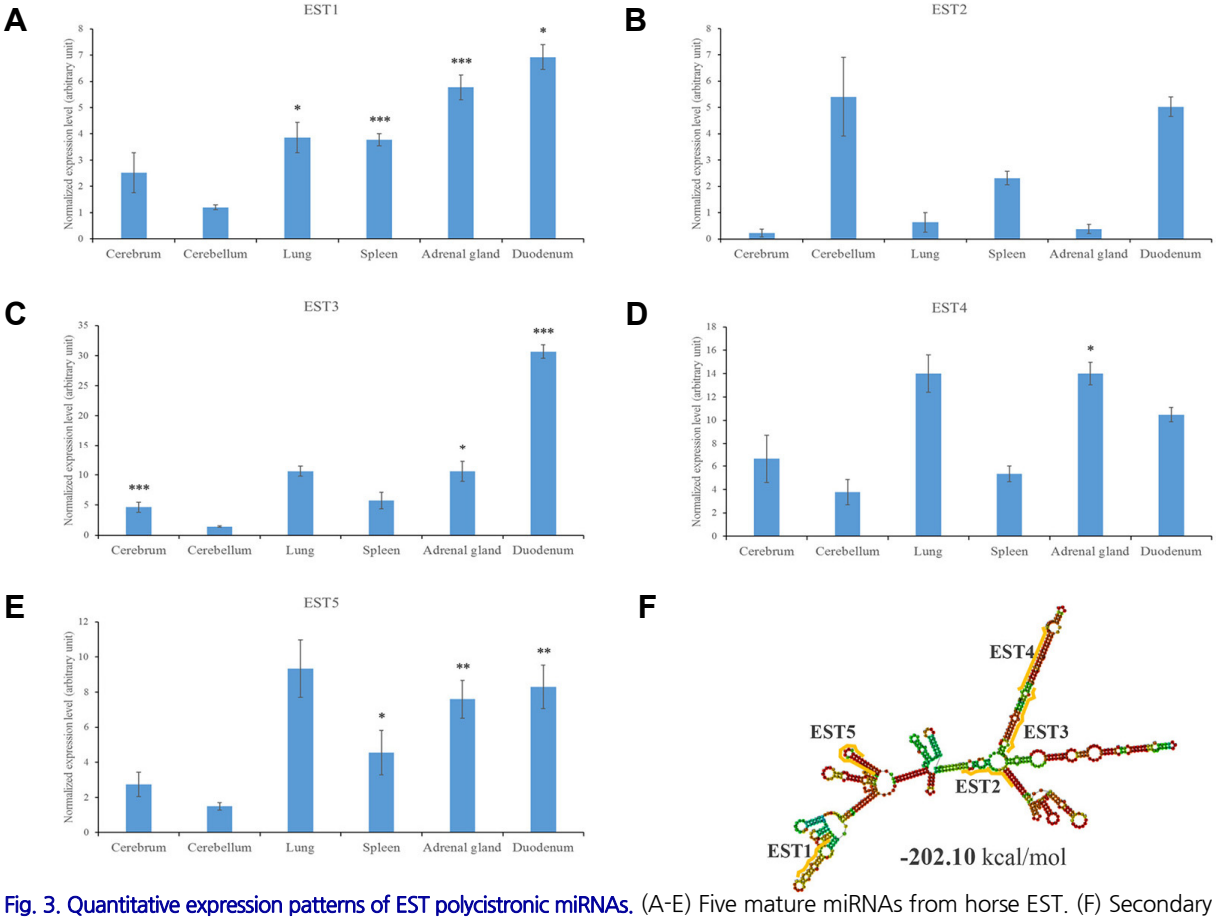

Fig. 3. Quantitative expression patterns of EST polycistronic miRNAs. (A-E) Five mature miRNAs from horse EST. (F) Secondary structure prediction of BM734541.1, indicated are each mature sequence. Each samples was examined in triplicate (Bar: mean; Whisker: standard deviation). Paired Student's *t*-test was performed to verify statistical significance between the expression levels of two samples (*$p < 0.05$). All miRNAs were compared cerebellum with other tissues.

DISCUSSION

MER-sequences are non-autonomous DNA transposons distributed throughout the host genome in a fossilized state. According to pre-RepeatMasking data, the horse genome possesses 3.61% DNA transposons containing MER-sequences (Smit et al., 2004). MER sequences can form palindromic stem-loop structures, which are similar to miRNA precursor sequences. This suggests a mode for miRNA sequence formation (Ahn et al., 2013; Gim et al., 2014). To date, miRNAs that are derived from repetitive elements have been identified in the genomes of various species (Nozawa et al., 2010; Yuan et al., 2011). In this study, we identified 13 MER-derived miRNAs and confirmed their expression in six tissues of one thoroughbred horse.

During miRNA maturation, two mature miRNAs are processed from one precursor miRNA. According to previous studies, the two mature miRNAs are expressed via asymmetric selection of each miRNA strand by the processing of precursor miRNA (Hutvagner, 2005; Ruike et al., 2008). In this study, we also observed that two miRNAs presented different expression patterns. In the MER5A1-derived miRNAs, similar expression patterns were observed for the two miRNAs from one precursor miRNA (Fig. 2A), whereas the MER53- and MER91C-derived miRNAs showed slightly different expression patterns (Figs. 2C and 2D). MER5A1-derived miRNAs showed dominant expression in the lung; thus, it may be a good target for future studies on exercise, cardiopulmonary fitness, or any lung-related process (Fig. 2A). As a part of brain, cerebrum and cerebellum are related to central nervous system (CNS). However, they expression patterns have different in MER81 and MER91C-derived miRNAs (Figs. 2C and 2D). Cerebrum is most important region of the CNS, and controls all voluntary operation in the body. Thus, these different expression patterns could be related to the cerebrum-specific roles in the CNS. Moreover, these results could provide a clue for cognitive function of the horse (Lein et al., 2007). Spleen is related to immune response, and adrenal gland produces a variety of hormones. In our data, spleen and adrenal gland have lower expressed patterns. Therefore, these two organs provide the important points in the immune and internal secretion pathways in the organisms (Nishimura and Naito, 2005). Most chemical digestion takes place in duodenum, therefore MER53-derived ENSECAT00000027640-3p miRNA, MER91C-derived ENSECAT00000027964-5p miRNA and BM734541.1-derived miRNAs could be crucial roles in digestion (Fang et al., 2006). Our six tissue expression patterns could be involved in exercise, cognition, and the physiological pathway.

MER5A1-derived miRNA precursor sequences were well-matched with eca-miR-544b (Table 2); therefore, they may be good targets for future studies on exercise and cardiopulmonary fitness. miR-544b has, to date, been identified in only three species (human, cow, and horse), and additional studies are required to determine the species- or tissue-specific roles of miR-544b. MER53 has been predicted to encode miR-1302 (Yuan et al., 2010). MER53-derived miRNAs did not show consistent expression patterns among their families. The miR-1302 subfamily derived from MER53 showed tissue-specific expression and identified in human, chimpanzee, orangutan (Kozomara and Griffiths-Jones, 2014). Therefore, it may be interesting to analyze their functions and evolutionary mechanisms in various animals in further studies. MER81-derived miRNAs have no significantly expressed patterns in all tissues to the other miRNAs. Three of four MER81-derived precursor sequences were not matched with previously identified precursor sequences. Although only one sequence matched with eca-miR-8990, its E-value was not significant. According to the latest version of miRBase (version 21.0), a total of 1397 horse miRNAs have been identified. However, when compared to the total of 4523 human miRNAs, the number of horse miRNAs is small (Kozomara and Griffiths-Jones, 2014). This means that additional miRNAs may be identified in the horse genome; our data identified some of these miRNAs. Two MER91C-derived precursor miRNAs and four mature miRNAs were identified, and their expression in human cell lines has been validated (Ahn et al., 2013). MER91C-derived precursor sequences were matched to eca-miR-652 in the horse (Table 2) and to hsa-miR-652 in humans (Ahn et al., 2013). Many studies have identified TE-derived miRNAs (Borchert et al., 2011; Piriyapongsa et al., 2007; Smalheiser and Torvik, 2005). Specifically, palindromic structures of TEs show the potential to generate miRNA precursor forms; thus, MITE- and MER-derived miRNAs are well-known (Ahn et al., 2013; Gim et al., 2014; Piriyapongsa and Jordan, 2008; Piriyapongsa et al., 2007; Yuan et al., 2010). However, few studies have examined the functions of TE-derived miRNAs. We determined the expression patterns in several tissues in a horse, and additional studies are needed to evaluate the functions of these miRNAs.

We also predicted the miRNA cluster, identified as BM734541.1, in the horse EST *SLC28A3* intron. This EST overlapped with MER117, and we predicted five precursor miRNAs. In eukaryotic protein coding, the gene expression pattern is monocistronic; therefore, one transcript encodes one amino acid chain. However, the polycistronic miRNA cluster forms multiple miRNA components, each of which has a distinct role. miRNA clusters have been identified in the human genome, and miR-17-92 is a well-known polycistronic miRNA that acts as an onco-miRNA. The *C13orf25* gene is chiefly known to show high expression in several hematopoietic malignancies (Ota et al., 2004). The miR-17-92 cluster is composed of seven mature miRNAs located in the third intron of the *C13orf25* gene in human chromosome 13 and is significantly expressed in lung cancer patients (Hayashita et al., 2005; Ota et al., 2004). To date, little is known about the function of the identified short ESTs in the intergenic regions or intron regions, and these ESTs may generate new polycistronic miRNA. In this study, we demonstrated that the EST sequence forms the miRNA cluster and confirmed their polycistronic expression of the miRNA. These results provide a basis for identifying new polycistronic miRNAs from ESTs.

As an invader of host genomes, TE sequences have undergone rapid evolution compared to other genomic sequences (Park et al., 2015). As a class II TE, DNA transposons, including MER repeats, underwent arrangement and were then conserved in the host genome (Pace and Feschotte,

2007). MER-derived miRNAs, detected in humans and horses, may be present in other mammals, including rodentia and lagomorpha. MER5A1-derived miRNAs were matched with miR-544b, which was previously identified in humans, horses, and cows. MER53-matched miR-1302 and MER91C-derived miR-652 were identified in four and fourteen species, respectively (Kozomara and Griffiths-Jones, 2014). Similarly, many cases of TE-derived miRNAs have been linked to phylogeny-specific miRNAs (Piriyapongsa et al., 2007). For instance, MITE-derived miR-548 was principally identified in primates (Liang et al., 2012).

We predict that additional miRNAs, as well as ESTs, derived from palindrome sequence TEs will be identified. In the host, these miRNAs were reported to have various roles. In a human study, the SNP in miR-1302-binding sites impaired spermatogenesis (Zhang et al., 2011), and miR-652 expression was related to heart disease. miR-548 is involved in the host antiviral response by targeting interferon λ1, and thus may be related to the immune system (Li et al., 2013). Based on these results, further studies are required to determine the functions of these miRNAs in horse and other mammalian species.

Owing to the development of next-generation sequencing technology, more precise genome sequencing data and transcript data is expected to be available. This study provides insights into novel functional transcripts. Future studies are required to understand how TE-derived miRNAs manipulate biological functions.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

ACKNOWLEDGMENTS

This work was supported by a grant from the ministry of Education, Republic of Korea (NRF-2015R1D1A1A01056650).

REFERENCES

Ahn, K., Gim, J.A., Ha, H.S., Han, K., and Kim, H.S. (2013). The novel MER transposon-derived miRNAs in human genome. Gene *512*, 422-428.

Altuvia, Y., Landgraf, P., Lithwick, G., Elefant, N., Pfeffer, S., Aravin, A., Brownstein, M.J., Tuschl, T., and Margalit, H. (2005). Clustering and conservation patterns of human microRNAs. Nucleic Acids Res. *33*, 2697-2706.

Balada, E., Vilardell-Tarres, M., and Ordi-Ros, J. (2010). Implication of human endogenous retroviruses in the development of autoimmune diseases. Int. Rev. Immunol. *29*, 351-370.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215-233.

Borchert, G.M., Holton, N.W., Williams, J.D., Hernan, W.L., Bishop, I.P., Dembosky, J.A., Elste, J.E., Gregoire, N.S., Kim, J.A., Koehler, W.W., et al. (2011). Comprehensive analysis of microRNA genomic loci identifies pervasive repetitive-element origins. Mob. Genet. Elements *1*, 8-17.

Chen, K., and Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. Nat. Rev. Genet. *8*, 93-103.

Coffin, J.M. (2004). Evolution of retroviruses: fossils in our DNA. Proc Am Philos Soc *148*, 264-280.

Doan, R., Cohen, N., Harrington, J., Veazey, K., Juras, R., Cothran, G., McCue, M.E., Skow, L., and Dindot, S.V. (2012). Identification of copy number variants in horses. Genome Res. *22*, 899-907.

Fang, R., Olds, L.C., and Sibley, E. (2006). Spatio-temporal patterns of intestine-specific transcription factor expression during postnatal mouse gut development. Gene Expr. Patterns *6*, 426-432.

Feschotte, C., Swamy, L., and Wessler, S.R. (2003). Genome-wide analysis of mariner-like transposable elements in rice reveals complex relationships with stowaway miniature inverted repeat transposable elements (MITEs). Genetics *163*, 747-758.

Fu, H.J., Zhu, J., Yang, M., Zhang, Z.Y., Tie, Y., Jiang, H., Sun, Z.X., and Zheng, X.F. (2006). A novel method to monitor the expression of microRNAs. Mol. Biotechnol. *32*, 197-204.

Gim, J.-A., Ha, H.-S., Ahn, K., Kim, D.-S., and Kim, H.-S. (2014). Genome-wide identification and classification of microRNAs derived from repetitive elements. Genomics Inform. *12*, 261-267.

Gim, J.A., Hong, C.P., Kim, D.S., Moon, J.W., Choi, Y., Eo, J., Kwon, Y.J., Lee, J.R., Jung, Y.D., Bae, J.H., et al. (2015). Genome-wide analysis of DNA methylation before-and after exercise in the thoroughbred horse with MeDIP-Seq. Mol. Cells *38*, 210-220.

Hayashita, Y., Osada, H., Tatematsu, Y., Yamada, H., Yanagisawa, K., Tomida, S., Yatabe, Y., Kawahara, K., Sekido, Y., and Takahashi, T. (2005). A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. Cancer Res. *65*, 9628-9632.

Hill, E.W., Gu, J., McGivney, B.A., and MacHugh, D.E. (2010). Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. Anim. Genet. *41 Suppl 2*, 56-63.

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., and Down, T. (2002). The Ensembl genome database project. Nucleic Acids Res. *30*, 38-41.

Hutvagner, G. (2005). Small RNA asymmetry in RNAi: function in RISC assembly and gene regulation. FEBS Lett. *579*, 5850-5857.

Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. *16*, 418-420.

Jurka, J., Kapitonov, V.V., Klonowski, P., Walichiewicz, J., and Smit, A.F. (1996). Identification of new medium reiteration frequency repeats in the genomes of Primates, Rodentia and Lagomorpha. Genetica *98*, 235-247.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. *110*, 462-467.

Kim, M.C., Lee, S.W., Ryu, D.Y., Cui, F.J., Bhak, J., and Kim, Y. (2014). Identification and characterization of microRNAs in normal equine tissues by next generation sequencing. PLoS One *9*, e93662.

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. *42*, D68-73.

Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Kochert, K., Bouhlel, M.A., Richter, J., Soler, E., et al. (2010). Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nat. Med. *16*, 571-579,

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H. and Kim, V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. EMBO J. *23*, 4051-4060.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., and Byrnes, E.J. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature *445*, 168-176.

Li, Y., Xie, J., Xu, X., Wang, J., Ao, F., Wan, Y., and Zhu, Y. (2013). MicroRNA-548 down-regulates host antiviral response via direct targeting of IFN-λ1. Protein Cell *4*, 130-141.

Liang, T., Guo, L., and Liu, C. (2012). Genome-wide analysis of mir-548 gene family reveals evolutionary and functional implications. J. Biomed. Biotechnol. *2012*, 679563.

Lund, E., and Dahlberg, J.E. (2006). Substrate selectivity of exportin 5 and Dicer in the biogenesis of microRNAs. Cold Spring Harb. Symp. Quant. Biol. *71*, 59-66.

Nishimura, M., and Naito, S. (2005). Tissue-specific mRNA expression profiles of human toll-like receptors and related genes. Biol. Pharm. Bull. *28*, 886-892.

Nozawa, M., Miura, S., and Nei, M. (2010). Origins and evolution of microRNA genes in Drosophila species. Genome Biol. Evol. *2*, 180-189.

Ota, A., Tagawa, H., Karnan, S., Tsuzuki, S., Karpas, A., Kira, S., Yoshida, Y., and Seto, M. (2004). Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. Cancer Res. *64*, 3087-3095.

Ou-Yang, F., Luo, Q.-J., Zhang, Y., Richardson, C.R., Jiang, Y., and Rock, C.D. (2013). Transposable element-associated microRNA hairpins produce 21-nt sRNAs integrated into typical microRNA pathways in rice. Funct. Integr. Genomics *13*, 207-216.

Pace, J.K., 2nd., and Feschotte, C. (2007). The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res. *17*, 422-432.

Park, S.J., Kim, Y.H., Lee, S.R., Choe, S.H., Kim, M.J., Kim, S.U., Kim, J.S., Sim, B.W., Song, B.S., Jeong, K.J., et al. (2015). Gain of a new exon by a lineage-specific alu element-integration event in the BCS1L gene during primate evolution. Mol. Cells *38*, 950-958.

Petersen, J.L., Mickelson, J.R., Rendahl, A.K., Valberg, S.J., Andersson, L.S., Axelsson, J., Bailey, E., Bannasch, D., Binns, M.M., Borges, A.S., et al. (2013). Genome-wide analysis reveals selection for important traits in domestic horse breeds. PLoS Genet. *9*, e1003211.

Piriyapongsa, J., and Jordan, I.K. (2007). A family of human microRNA genes from miniature inverted-repeat transposable elements. PLoS One *2*, e203.

Piriyapongsa, J., and Jordan, I.K. (2008). Dual coding of siRNAs and miRNAs by plant transposable elements. RNA *14*, 814-821.

Piriyapongsa, J., Marino-Ramirez, L., and Jordan, I.K. (2007). Origin and evolution of human microRNAs from transposable elements. Genetics *176*, 1323-1337.

Ruike, Y., Ichimura, A., Tsuchiya, S., Shimizu, K., Kunimoto, R., Okuno, Y., and Tsujimoto, G. (2008). Global correlation analysis for micro-RNA and mRNA expression profiles in human cell lines. J. Hum. Genet. *53*, 515-523.

Shukla, G.C., Singh, J., and Barik, S. (2011). MicroRNAs: processing, maturation, target recognition and regulatory functions. Mol. Cell Pharmacol. *3*, 83-92.

Smalheiser, N.R. (2003). EST analyses predict the existence of a population of chimeric microRNA precursor-mRNA transcripts expressed in normal human and mouse tissues. Genome Biol. *4*, 403.

Smalheiser, N.R., and Torvik, V.I. (2005). Mammalian microRNAs derived from genomic repeats. Trends Genet. *21*, 322-326.

Smit, A., Hubley, R., and Green, P. (2004). RepeatMasker Open-3.0. 2004. Seattle (WA): Institute for Systems Biology.

Smit, A.F. (1993). Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic Acids Res. *21*, 1863-1872.

Wade, C.M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T.L., Adelson, D.L., Bailey, E., Bellone, R.R., et al. (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. Science *326*, 865-867.

Yuan, Z., Sun, X., Jiang, D., Ding, Y., Lu, Z., Gong, L., Liu, H., and Xie, J. (2010). Origin and evolution of a placental-specific microRNA family in the human genome. BMC Evol. Biol. *10*, 346.

Yuan, Z., Sun, X., Liu, H., and Xie, J. (2011). MicroRNA genes derived from repetitive elements and expanded by segmental duplication events in mammalian genomes. PLoS One *6*, e17666.

Zhang, Q., Arbuckle, J., and Wessler, S.R. (2000). Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family Heartbreaker into genic regions of maize. Proc. Natl. Acad. Sci. USA *97*, 1160-1165.

Zhang, H., Liu, Y., Su, D., Yang, Y., Bai, G., Tao, D., Ma, Y., and Zhang, S. (2011). A single nucleotide polymorphism in a miR-1302 binding site in CGA increases the risk of idiopathic male infertility. Fertil. Steril. *96*, 34-39. e37.

Zhou, M., Wang, Q., Sun, J., Li, X., Xu, L., Yang, H., Shi, H., Ning, S., Chen, L., Li, Y., et al. (2009). In silico detection and characteristics of novel microRNA genes in the Equus caballus genome using an integrated ab initio and comparative genomic approach. Genomics *94*, 125-131.