

Discovering Novel MicroRNAs from Small RNA Component of the Mouse Embryo Transcriptome

Tatsuya Ando¹
andot@takara-bio.co.jp

Sachiko Okamoto²
okamotosy@takara-bio.co.jp

Masahiro Sato¹
satomy@takara-bio.co.jp

Hiroyuki Izu¹
izut@takara-bio.co.jp

Masanori Takayama¹
mitsutam@takara-bio.co.jp

Masanari Kitagawa¹
kitagawam@takara-bio.co.jp

Junichi Mineno¹
minenoj@takara-bio.co.jp

Kiyozo Asada^{1,3}
asadak@takara-bio.co.jp

Ikunoshin Kato^{2,3}
katoik@takara-bio.co.jp

¹ DNA Function Analysis Center, Takara Bio Inc., 3-4-1 Seta, Otsu, Siga 520-2193, Japan

² Center for Cell and Gene Therapy, Takara Bio Inc., 3-4-1 Seta, Otsu, Siga 520-2193, Japan

³ Biotechnology Research Laboratories, Takara Bio Inc., 3-4-1 Seta, Otsu, Siga 520-2193, Japan

Keywords: microRNA, small RNA, RNA interference, transcriptome, genome-wide analysis

1 Introduction

Small RNAs is a family of 21-27 nt non-coding RNAs that can play important regulatory roles. Recently, transcriptome analysis of small RNAs can be taken using microarray, or massively parallel signature sequencing (MPSS) [4, 5]. In small RNAs, microRNAs(miRNAs) is highly expressed and induce gene silencing through specific base-pairing with target molecules [1]. A lot of miRNAs were discovered by computational approaches which detect conserved hairpin structure [2, 3]. We applied MPSS to investigate the complexity of mammalian small RNAs and discovered new miRNA candidates in numerous expressed small RNAs.

2 Materials and Method

2.1 MPSS

MPSS sequences hundreds of thousands of molecules per reaction and provides quantitative information. Small RNA molecules are isolated by size fractionation on polyacrylamide gel. RNA adapters are ligated and reverse transcriptase generates the first strand of cDNA, which is amplified and used as the template for MPSS. We generated libraries using small RNA of mouse embryo of 9.5, 10.5 and 11.5 days. The obtained 22nt sequences or "signatures" were normalized in "transcripts per million (TPM)" for each library. Signatures more than 8TPM in any library are determined to be significantly expressed and identified genome position using mouse genome sequences (mm5) of University of California Santa Cruz (UCSC) genome database.

2.2 Structure Check of miRNA Candidates

We searched for stable stem-loop structures in neighborhoods of signatures which matched genome of which repetitive sequences are masked and don't matched rRNA, rRNA, snoRNA, snRNA. We extracted genome sequences of these signatures with 88bases on each side. A sliding window of 110bases with an increment of 1base was scanned along the extracted sequences. The sequences containing signatures were folded using the program RNAfold. We selected sequences of which free energy was minimum. We identified the selected sequences that were stem-loop structures, contained at least 16 base-pairings on signature, and didn't contain large internal loops or bulges as new precursor of miRNA candidates.

3 Results and Discussions

3.1 Summary Statistics of Small RNA s Transcription

We adapted MPSS for small RNAs in mouse embryo. We sequenced 589,738(57,632 distinct) and 731,071(55,267) and 686,314(51,286) signatures for embryo libraries of 9.5, 10.5 and 11.5 days respectively. These signatures were normalized in “transcripts per million (TPM)” and the 20,014 significantly expressed signatures were extracted. For the three libraries, 73%~75% of total signatures matched to genome (Table 1). The unmatched signatures may be derived from genomic gaps such as rRNA repeats or polymorphism of genome or centromeres or may result from sequencing errors. In signatures matching to RNA, most of them matched to rRNA, making up to 27~39% of total signatures for three libraries.

Table 1: Summary statistics of small RNA s transcription

type	subtype	distinct signature	9.5days TPM	10.5days TPM	11.5days TPM
Gene	exon	1,022	58,970(7%)	68,435(8%)	85,536(10%)
	Intron	1,989	138,307(16%)	147,065(17%)	135,239(15%)
Repeat		1,744	57,165(7%)	43,950(5%)	53,302(6%)
RNA		4,652	336,011(40%)	238,675(28%)	329,744(38%)
Genome total		9,770	614,900(73%)	653,523(75%)	643,110(73%)
No match to Genome		10,244	232,455(27%)	212,112(25%)	234,617(27%)
Total		20,014	847,355(100%)	865,635(100%)	877,727(100%)

RNA contains rRNA, tRNA, snoRNA, snRNA

3.2 Discovery of New miRNA Candidates

We selected 3,374 signatures, those matched to genome of which repetitive sequences are masked but didn't match to rRNA, tRNA, snoRNA, and snRNA. To test for a conserved RNA-folding pattern characteristic of miRNA genes, for the selected signatures, we extracted the neighbor sequences of signatures from genome. So, 390 new miRNA candidates were discovered (Table 2). About half of signatures forming stem-loop RNA are novel miRNA candidates, but the reminders are known miRNAs registered in miRBase 7.0 (Table 2). This number of expressed known miRNAs is 84% of total registered miRNAs of mouse. This fact indicated most of known miRNAs were expressed in mouse embryo. Though 40 of 195 known miRNAs have not been experimented, we could confirm the expression of these miRNAs by MPSS. 53% of novel miRNA candidates was discovered on the opposite arm of known miRNAs or conserved in other species (Table 2). Finally, 92 quite novel mature miRNA candidates were discovered (Table 2). If these miRNAs were added to registered miRNAs, the number of register would increase by 39%. These results show MPSS is a powerful tool for finding the unknown novel miRNAs.

Table 2: Expressed known miRNAs and novel miRNA candidates in mouse embryo

category	mature miRNA	precursor of miRNA
Known miRNAs	195	231
Novel miRNA candidates	195	214
Opposite arm of known miRNAs	86	91
Conserved in other species' miRNAs	24	24
Quite novel miRNAs	92	106

3.3 Searching for Other Small RNAs; Short Interfering RNAs(siRNAs)

We searched signatures which can form double strand RNA each other on the same genome position and match to the known coding genes from signatures which were not folding stem-loops. We could not find these signatures. This finding agrees with the fact that siRNAs derived from long endogenous dsRNAs were not found in mammals [1].

Our data indicates small RNAs expressed on the various regions of genome and those regulatory roles are complex. We are going to analyze small RNA data combined with transcriptome of polyA RNAs.

References

- [1] Bartel, D.P., MicroRNAs: genomics, biogenesis, mechanism, and function, *Cell*, 116:281-297, 2004.
- [2] Bentwich, I. *et al.*, Identification of hundreds of conserved and nonconserved human microRNAs, *Nat. Genet.*, 37(7):766-770, 2005.
- [3] Lim, L.P. *et al.*, The microRNAs of *Caenorhabditis elegans*, *Genes Dev.*, 17(8):991-1008, 2003.
- [4] Lu, C. *et al.*, Elucidation of the small RNA component of the transcriptome, *Science*, 309:1567-9, 2005.
- [5] Lu, J. *et al.*, MicroRNA expression profiles classify human cancers, *Nature*, 435(7043):834-8, 2005.