A Database for TSSs of Human MicroRNAs*

Malay Bhattacharyya and Sanghamitra Bandyopadhyay
Machine Intelligence Unit, Indian Statistical Institute
203 B. T. Road, Kolkata - 700108, India
E-mail: {malay_r, sanghami}@isical.ac.in.

January 25, 2012

Abstract

MicroRNAs (miRNAs) are small endogeneous non-coding RNAs of about 22nt length. These short RNAs regulate the expression of mRNAs by hybridizing with their 3′-UTRs or by translational repression. They have been shown to take crucial roles in many biological processes. Many of the current studies are focused over how mature miRNAs regulate mRNAs, even though very limited knowledge is there about their transcriptional loci. Primary miRNAs (pri-miRs) are first transcribed from the DNA, followed by the formation of precursor miRNA (pre-miR) by endonucleases activity, which finally produces mature miRNAs. Unfortunately, the identification of the loci of pri-miRs, and the associated information about transcription start sites (TSSs) and promoters is still in progress. There are reported results concerning the TSSs of about 40% of the total mature miRNAs hitherto explored in human. These information, even though limited, may be useful for further study on the regulation of miRNAs. In this paper, we provide a novel database of miRNA TSSs (miRT) that might be a valuable resource for advanced research on miRNA regulation.

1 Introduction

In the recent years, a major attention in molecular biology research is devoted over various short and long non-coding RNAs [15]. Of the various subclasses of non-coding RNAs, microRNAs (miRNAs) are the most thoroughly characterized subclass of short RNAs in the recent literature [2]. MiRNAs are short (∼22nt) single stranded RNA molecules which are endogenous and play a major role in the regulation of mRNA through cleavage by hybridizing the 3′-UTRs of their target mRNAs or by translational repression [21, 27]. The negative regulatory mechanism at the post transcriptional level ensures that miRNAs play a major

*The complete version of this paper is under consideration of publication in the *Proceedings of* 11^{th} *IEEE International Workshop on High Performance Computational Biology, 2012.*

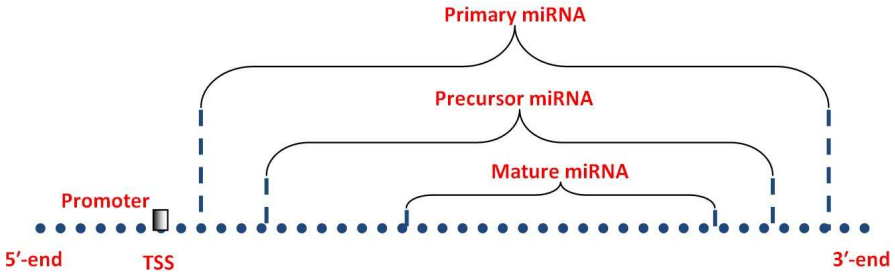

Figure 1: The locus of TSS in the upstream of miRNAs.

role in controlling diverse biological processes such as carcinogenesis, cellular proliferation and differentiation.

The biogenesis of miRNAs begins with the generation of primary miRNAs (pri-miRs) that are initially expressed as a part of mRNA hairpin which forms a longer part of the transcript [13]. The excision of the upper part of this RNA hairpin by the nuclear RNase III enzyme Drosha produces the precursor miRNA (pre-miR), a 70-110nt intermediate [7, 12, 26]. Pre-miRNAs, which form short RNA hairpins bearing a 2-nt 3′ overhang, are then bound by the nuclear export factor Exportin 5, which transports them to the cytoplasm [16, 25]. At this stage, another RNase III enzyme termed Dicer removes the terminal loop of the pre-miR to generate a ∼22-bp miRNA duplex with 2-nt 3′ overhangs [13]. One strand of this short-lived duplex degrades while the other strand remains as the mature miRNA. The strand with the less stable 5′-end (e.g., G:U pair versus G:C pair) usually survives [9, 23]. The transcriptional loci of a mature miRNA and its primary and precursor transcripts are shown in Fig. 1. The mature miRNA is finally incorporated into a multi-protein complex, termed as RNA induced silencing complex (RISC), where it functions to drive RISC to the complementary mRNA targets [8, 18, 22] to control its biological activities.

Although sufficient experimental knowledge about the biogenesis of miRNA has already been gathered, enough information about the TSSs of pri-miRs is not yet available. Based on the genomic locus, miRNAs can be broadly categorized into two types – intragenic (miRNA-coding genes located within their host protein-coding genes) and intergenic (miRNA-coding genes located in-between protein-coding genes) [7]. It is highly anticipated that most of the intergenic miRNAs (inter-miRs) are transcribed independently based on their own RNA polymerase II enzyme (Pol II) promoter, while intragenic miRNAs (intra-miRs) are transcribed along with their host genes [5]. Current research interests concentrate on the transcriptional regulation of inter-miRs, as very little is known about whether their transcription is associated with those of their neighboring genes at all. In fact, the nature of the primary transcripts of inter-miRs is still not fully explored [21]. The localization of transcription start sites (TSSs) in the upstream region of the precursor miRNAs (pre-miR) that are intergenic is thus an important object of study.

Table 1: The summary of the information content.

Source	# miRNAs	Assembly	Information Analyzed
Landgraf *et al.* [11]	204	hg17	small RNA sequencing
Saini *et al.* [21]	29	hg17	polyadenylation signals, CpG islands, EST data, TFBSs, expression ditag, CAGE tags
Ozsolak *et al.* [19]	182	hg18	nucleosome mapping, promoter chromatin signatures
Fujita *et al.* [6]	95	hg17	biochemical experiment, cDNA clones
Marson *et al.* [17]	550	hg17	ChIP-seq data
Corcoran *et al.* [5]	65	hg18	ChIP-chip data
Chien *et al.* [4]	318	hg19	CAGE tags, TSS Seq, H3K4me3 chromatin signature

Majority of the miRNA loci reside in intronic regions of the hosting transcriptional units and thus their transcriptional machinery has been studied [10, 20]. RNA polymerase II is responsible for the transcription of these miRNAs along with the heterologous mRNAs [3, 10, 14, 20]. The knowledge about pri-miR TSSs will help us to understand the transcriptional regulation of miRNA and thus its biological activity. This will also help to provide accurate information about the promoter region of miRNAs. Only a few pri-miR structures were described biochemically earlier [3, 10, 14, 24]. Some miRNAs embedding regions are reported to reside in introns of certain coding genes [10, 20], but it is not fully understood how these miRNAs are produced. Here, we report about a database that accumulates transcription start sites (TSSs) of miRNAs for the advance research. The different studies on finding TSSs are briefed in Table 1.

2 The miRT Database

The miRT database includes information about the TSSs (with a minimum support value of 1) of 588 miRNAs (closer to 40% of the known mature miRNAs) in total. Of these, the ratio of inter- and intra-miRs is almost 1:2. However, the number of reported TSSs are more as because there are multiple TSSs experimentally obtained for some of these miRNAs. The average support value, which means the number of references in which there is a report about the experimentally validated TSS of a specific miRNA, of miRT is quite high and it establishes the significance of the accumulated information. The statistical details on miRT is listed in Table 2. The entire dataset is available at: `http://www.isical.ac.in/~bioinfo_miu/miRT/miRNA_TSS_database.xls`. Even though a recent database miRGen 2.0 reports about the different transcripts of miRNAs [1], however our major focus is to accumulate the experimentally validated and updated TSS information of human miRNAs. The share of intergenic and intragenic miRNAs and their TSSs reported in the miRT database are shown in Fig. 2.

To examine the variability of the various experimental results reported in the studies covered here, we have carried out miRNA-specific analysis by converting

Table 2: The complete statistics of miRT.

# miRNAs	588
# Inter-miRs	206
# Intra-miRs	382
# TSS loci	670
Maximum support value	7
The miRNA with maximum support	hsa-let-7i
Average support value	2.27

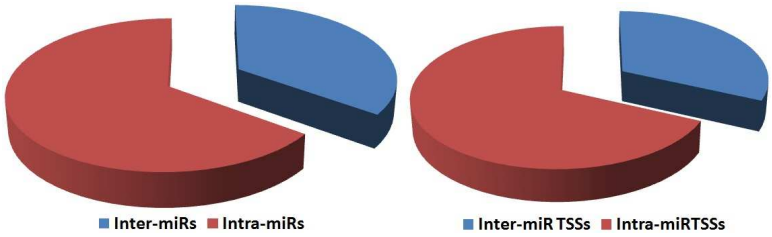

Figure 2: The share of different categories of miRNAs in miRT.

all the TSSs to a fixed gene assembly (hg19). For example, take the miRNA hsa-let-7i, for which all the seven papers report to obtain a TSS (may be a region with start TSS and end TSS or a fixed locus). We show the different TSSs experimented in these studies by converting them to a common gene built (hg19). The comparative plots of the start and end of the TSSs of hsa-let-7i are shown in Fig. 3. It includes nine loci (three from three different cell lines in [19] and one from each of the others [4, 5, 6, 11, 17, 21]) of the same miRNA that are very close with respect to the large span of the genome, even though they differ much ($\sigma \sim 17225.6$ for start TSS and $\sigma \sim 19656.4$ for end TSS).

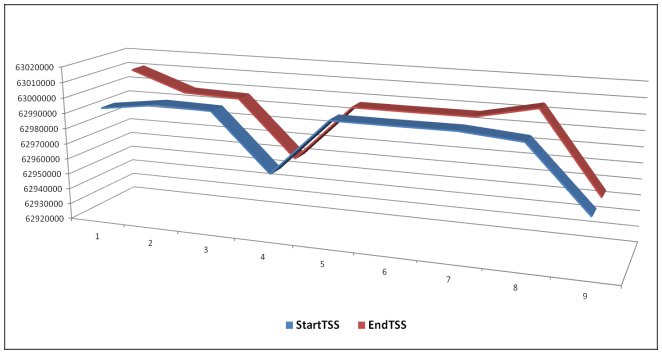

Figure 3: The variability of the TSSs experimented by various studies on the miRNA hsa-let-7i.

3 Conclusion

The current paper reports the assembly of biologically validated TSS information of a large number of human miRNAs. It accumulates the entire data in the form of a database, miRT, that is a significant resource for further analysis of transcriptional regulation of these short RNA regulators. Global prediction of promoter regions of miRNA genes, which lies in the upstream of TSSs, would allow exploring the mechanisms underlying gene-regulatory mechanisms involving these miRNAs. As highlighted by Zhou *et al.* [27], the study of conservedness may also provide robust confirmation of the loci of TSSs.

Acknowledgement

The authors would like to thank Manali Das in the Department of Bioinformatics, West Bengal University of Technology, for her support to prepare the database.

References

[1] P. Alexiou, T. Vergoulis, M. Gleditzsch, G. Prekas, T. Dalamagas, M. Megraw, I. Grosse, T. Sellis and A. G. Hatzigeorgiou. miRGen 2.0: a database of microRNA genomic information and regulation. *Nucleic Acids Research (Database Issue)*, 38:D137–D141, 2010.

[2] D. P. Bartel. MicroRNAs: target recognition and regulatory functions. *Cell*, 136(2):215–233, 2009.

[3] X. Cai, C. H. Hagedorn and B. R. Cullen. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, 10:1957–1966, 2004.

[4] C. Chien, Y. Sun, W. Chang, P. Chiang-Hsieh, T. Lee, W. Tsai, J. Horng, A. Tsou and H. Huang. Identifying transcriptional start sites of human microRNAs based on high-throughput sequencing data. *Nucleic Acids Research*, 39(21):9345–9356, 2011.

[5] D. L. Corcoran, K. V. Pandit, B. Gordon, A. Bhattacharjee, N. Kaminski, P. V. Benos. Features of mammalian microRNA promoters emerge from polymerase II chromatin immunoprecipitation data. *PLoS ONE* 4:e5279, 2009.

[6] S. Fujita and H. Iba. Putative promoter regions of miRNA genes involved in evolutionary conserved regulatory systems among vertebrates. *Bioinformatics*, 24(3):303–308, 2008.

[7] S. Griffiths-Jones *et al.* miRBase: tools for microRNA genomics. *Nucleic Acids Research (Database issue)*, 36:D154–D158, 2008.

[8] S. M. Hammond, E. Bernstein, D. Beach, G. J. Hannon. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature*, 404(6775):293–296, 2000.

[9] A. Khvorova, A. Reynolds and S. D. Jayasena. Functional siRNAs and miRNAs exhibit strand bias. *Cell*, 115(2):209–216, 2003.

[10] Y. K. Kim and V. N. Kim. Processing of intronic microRNAs. *EMBO Journal*, 26:775–783, 2007.

[11] P. Landgraf *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, 129:1401–1414, 2007.

[12] Y. Lee, K. Jeon, J. Lee, S. Kim and V. N. Kim. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO Journal* 21:4663–4670, 2002.

[13] Y. Lee, C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Rådmark, S. Kim and V. N. Kim. The nuclear RNase III drosha initiates microRNA processing. *Nature*, 425:415–419, 2003.

[14] Y. Lee, M. Kim, J. Han, K. Yeom, S. Lee, S. H. Baek and V. N. Kim. MicroRNA genes are transcribed by RNA polymerase II. *EMBO Journal*, 23:4051–4060, 2004.

[15] M. Li, C. Marin-Muller, U. Bharadwaj, K. H. Chow, Q. Yao and C. Chen. MicroRNAs: control and loss of control in human physiology and disease. *World Journal of Surgery*, 33(4):667–684, 2009.

[16] E. Lund, S. Güttinger, A. Calado, J. E. Dahlberg and U. Kutay. Nuclear export of microRNA precursors. *Science*, 303(5654):95–98, 2004.

[17] A. Marson *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell*, 134:521–533, 2008.

[18] J. Martinez, A. Patkaniowska, H. Urlaub, R. Lührmann and T. Tuschl. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell*, 110(5):563–574, 2002.

[19] F. Ozsolak, L. L. Poling, Z. Wang, H. Liu, X. S. Liu, R. G. Roeder, X. Zhang, J. S. Song and D. E. Fisher. Chromatin structure analyses identify miRNA promoters. *Genes and Development*, 22:3172–3183, 2008.

[20] A. Rodriguez, S. Griffiths-Jones, J. L. Ashurst and A. Bradley. Identification of mammalian microRNA host genes and transcription units. *Genome Research*, 14:1902–1910, 2004.

[21] H. K. Saini, S. Griffiths-Jones and A. J. Enright. Genomic analysis of human microRNA transcripts. *Proceedings of the National Academy of Sciences USA*, 104(45):17719–17724, 2007.

[22] D. S. Schwarz, G. Hutvágner, B. Haley and P. D. Zamore. Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways. *Molecular Cell*, 10(3):537–548, 2002.

[23] D. S. Schwarz, G. Hutvágner, T. Du, Z. Xu, N. Aronin and P. D. Zamore. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115(2):199–208, 2003.

[24] K. D. Taganov, M. P. Boldin, K. J. Chang and D. Baltimore. NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. *Proceedings of the National Academy of Sciences USA*, 103(33):12481–12486, 2006.

[25] R. Yi, Y. Qin, I. G. Macara and B. R. Cullen. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes and Development*, 17(24):3011–3016, 2003.

[26] Y. Zeng and B. R. Cullen. Sequence requirements for micro RNA processing and function in human cells. *RNA* 9:112–123, 2003.

[27] X. Zhou, J. Ruan, G. Wang and W. Zhang. Characterization and Identification of MicroRNA Core Promoters in Four Model Species. *PLoS Computational Biology*, 3(3), e37, 2007.