

Cap-seq Reveals Complicated miRNA Transcriptional Mechanisms in *C. elegans* and Mouse

Jiao Chen¹, Dongxiao Zhu², and Yanni Sun^{1,*}

¹Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, 48824, USA

²Department of Computer Science, Wayne State University, Detroit, MI 48202

*yannisun@msu.edu

Supplementary Materials

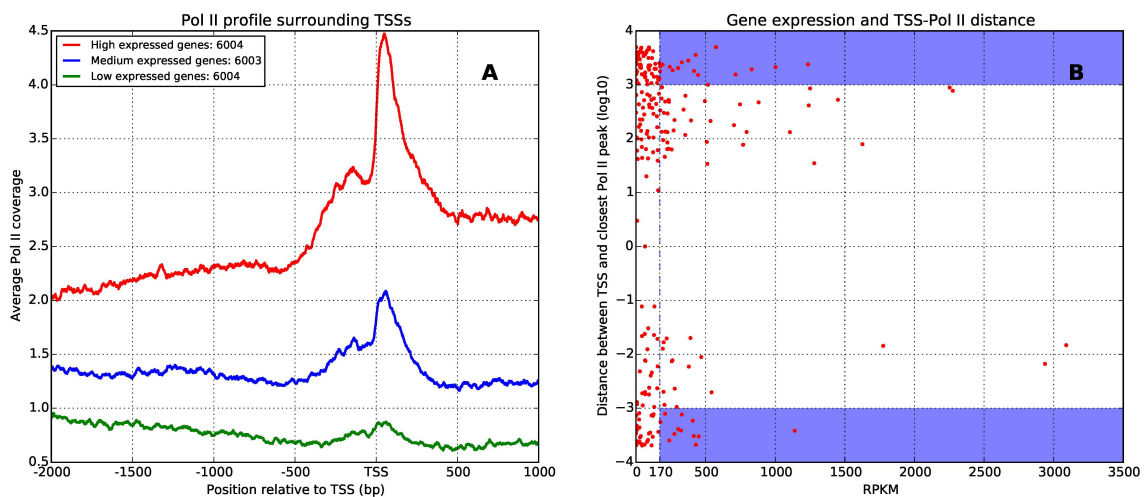


Figure S1. (A) Pol II signal profile surrounding TSSs of highly, moderately, and lowly expressed protein-coding genes. For the group of actively expressed genes, Pol II binding is highly enriched at the TSSs. The average distance from the TSSs of highly, moderately, and lowly expressed genes to the closest Pol II peaks are 168, 292, and 705 bp, respectively. (B) Each red point represents one gene. For each gene, its x-axis coordinate is the gene's RPKM (reads per kilo-base per million mapped reads) value; its y-axis is the log-transformed (\log_{10}) distance between the gene's TSS and the closest Pol II peak summit identified by MACS. The blue shadow regions show that even for many highly expressed genes (with RPKM value greater than 170, 29/85), the distance from the TSSs to the closest Pol II peak are over 1000 bp.

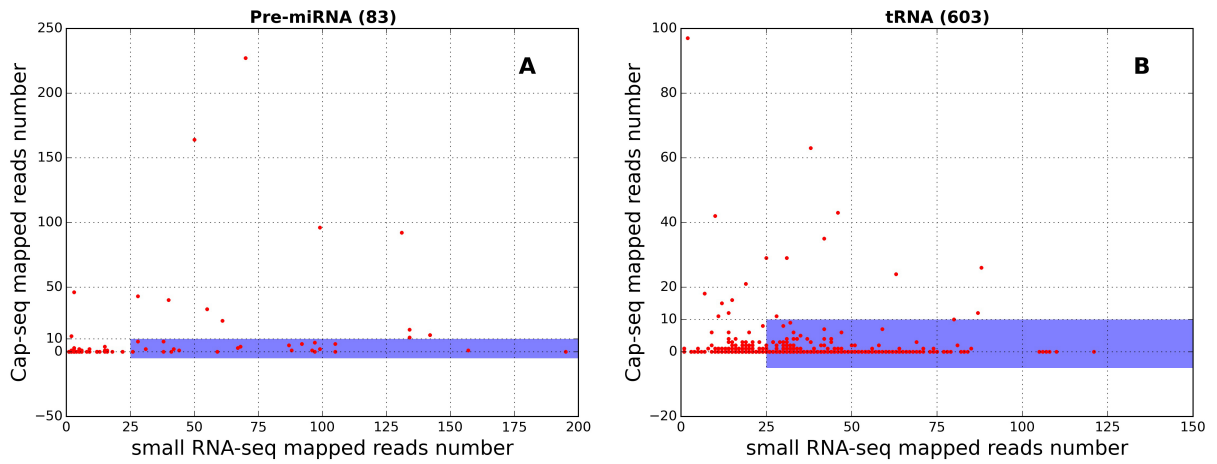


Figure S2. (A) Small RNA-seq data and Cap-seq data mapped to pre-miRNAs. Each red point represents one pre-miRNA. For each pre-miRNA, its x-axis coordinate represents the number of mapped small RNA-seq reads; its y-axis coordinate represents the number of mapped Cap-seq reads. The blue shadow region indicates that most highly expressed miRNAs (22/33) do not have many Cap-seq reads mapped (<10). (B) Small RNA-seq data and Cap-seq data mapped to tRNAs. The blue region shows that most highly expressed tRNAs (374/384) have no or less than 10 Cap-seq reads mapped.

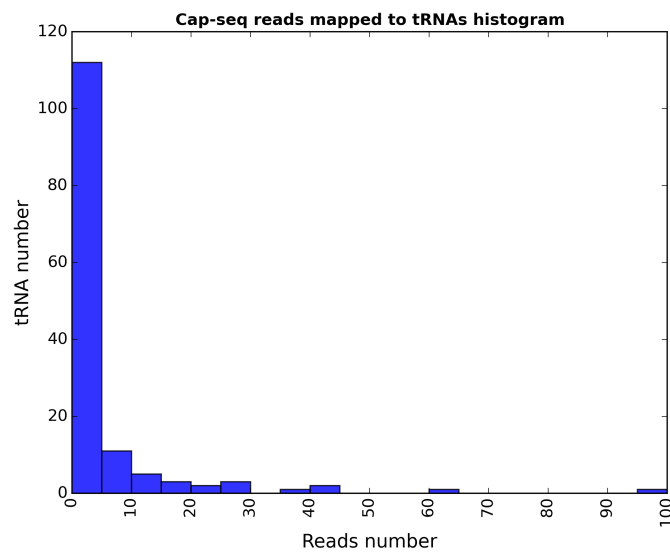


Figure S3. Histogram of Cap-seq reads mapped to tRNAs. 142 tRNAs have Cap-seq reads mapped, among which 124 have less than 10 reads.

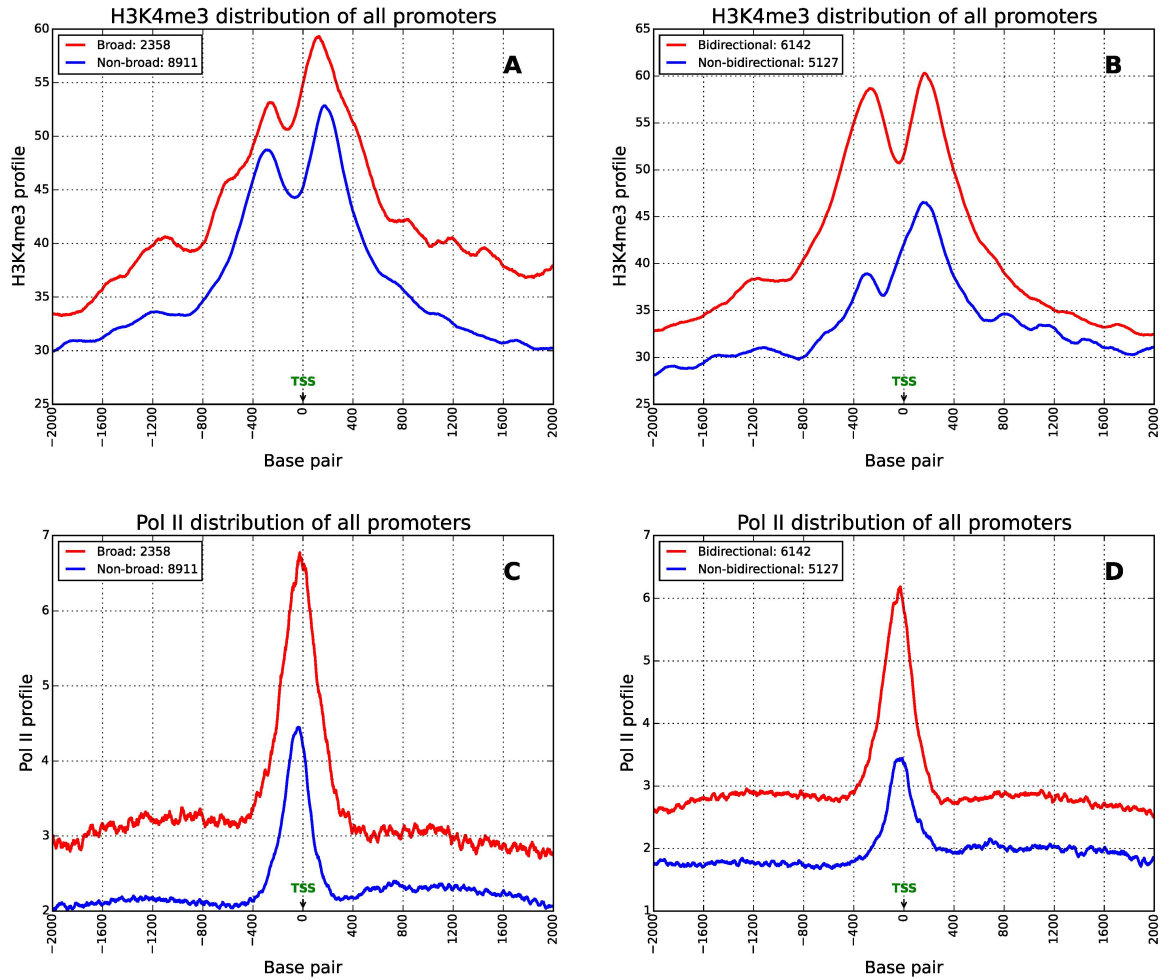


Figure S4. H3K4me3 and Pol II signal profile surrounding bidirectional and broad promoters. (A, B) Histone methylation H3K4me3 signal is highly enriched in the actively transcribed promoters, with a higher level in bidirectional and broad promoters than non-bidirectional and non-broad promoters, respectively. In bidirectional promoters, there are H3K4me3 peaks both upstream and downstream of TSS, suggesting that H3K4me3 signal is enriched for both plus and minus strand transcription. On the other hand, H3K4me3 signal has a significantly higher peak upstream of broad promoter TSSs, indicating that histone methylation elongates to promote the downstream transcriptions. (C, D) Pol II binding is highly enriched at the TSSs of actively transcribed promoters, also with a higher level in bidirectional or broad promoters than non-bidirectional or non-broad promoters, respectively. However, the distribution patterns of Pol II signal surrounding TSSs are quite consistent in different kinds promoters, suggesting that the binding of Pol II may not hold transcription directivity.

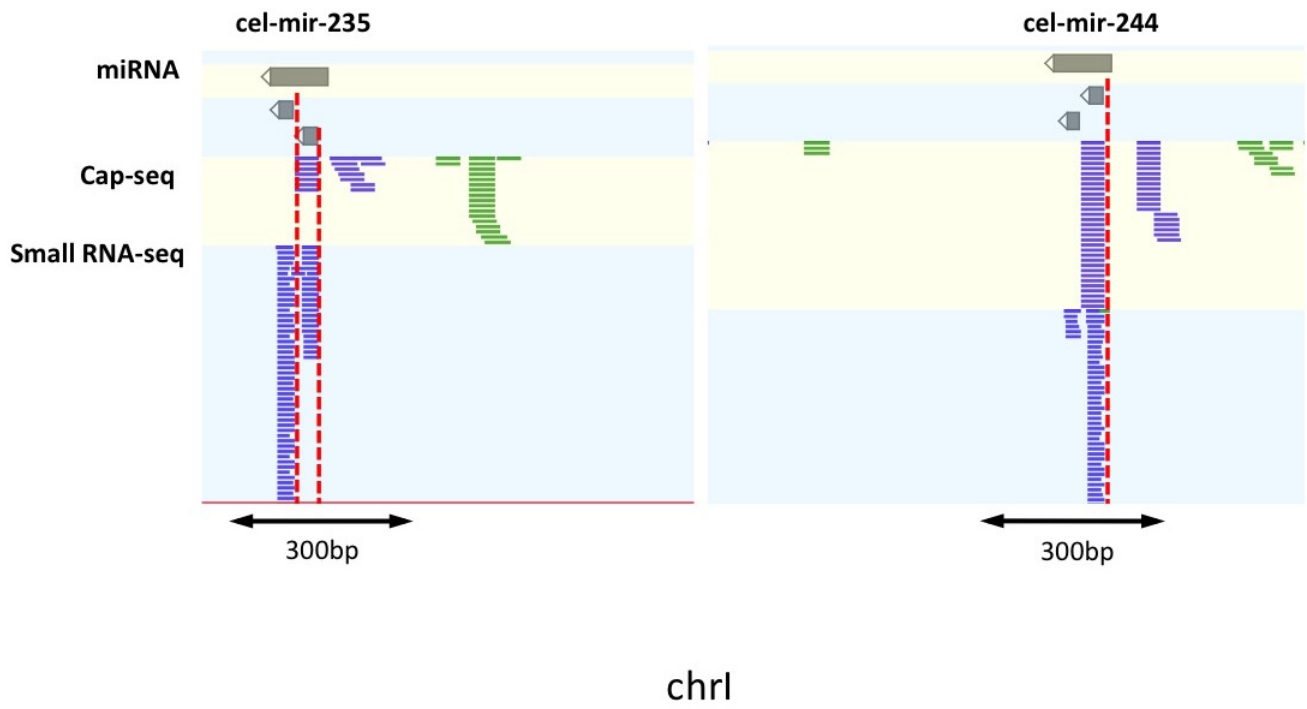


Figure S5. 5'-capped pre-miRNAs in *C. elegans*: *cel-mir-235* and *cel-mir-244*

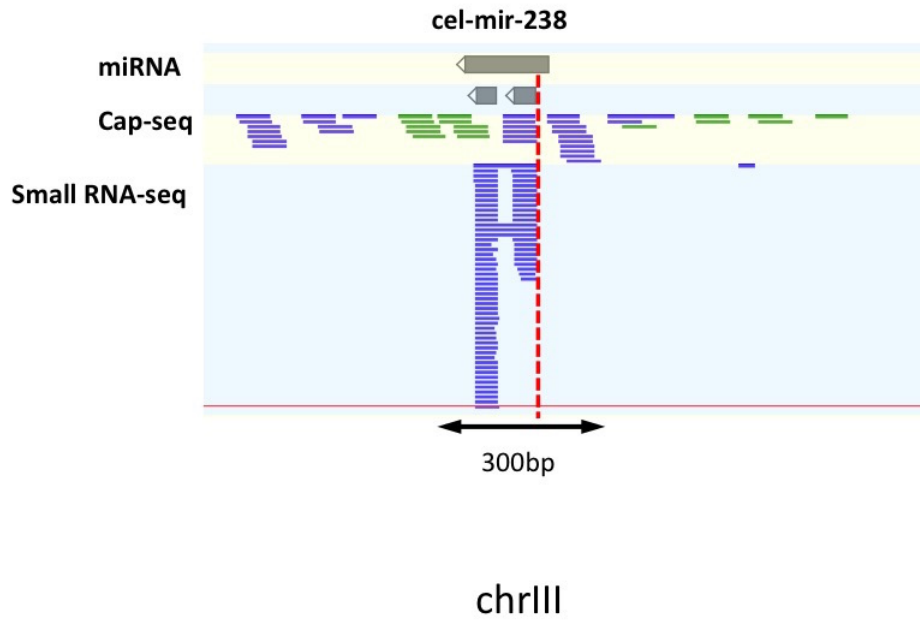


Figure S6. 5'-capped pre-miRNAs in *C. elegans*: *cel-mir-238*

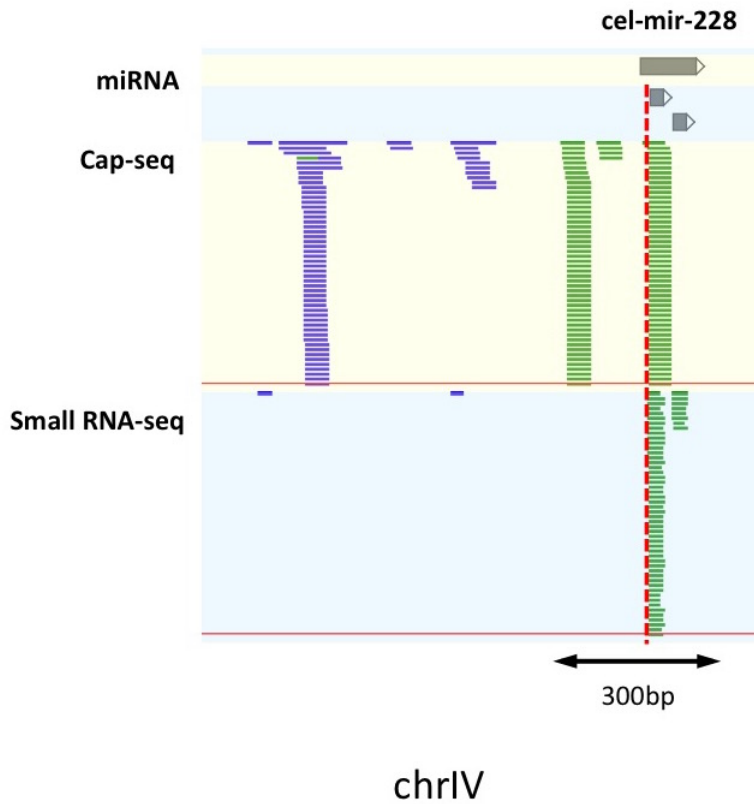


Figure S7. 5'-capped pre-miRNAs in *C. elegans*: *cel-mir-228*

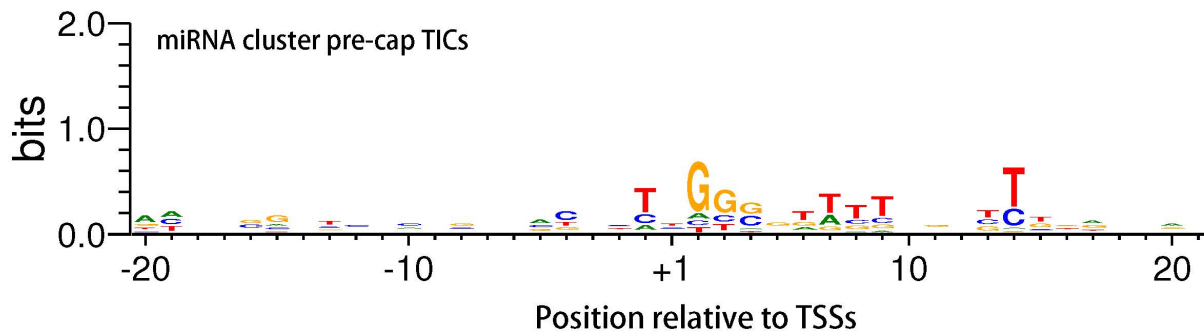


Figure S8. Weblogo analysis of nucleotides around the modes of pre-cap TICs in miRNA clusters (+1). The nucleotide height (in bits) stands for the \log_2 ratio of the observed nucleotides frequency relative to the background genomic nucleotide composition. A weak consensus sequence of “TNGG” is detected surrounding the modes of the TICs.