Liu et al. Arch Crop Sci 2018, 2(1):58-60
DOI: 10.36959/718/602 | Volume 2 | Issue 1

Archives of Crop Science

Editorial Open Access

MicroRNA Prediction Using Small RNA-Sequencing, Advances and Challenges

Kan Liu*

Department of Computer Science and Engineering, University of Nebraska-Lincoln, USA

Role of MicroRNAs in Biogenesis and Gene Expression

MicroRNAs (miRNAs) are basically a type of short (~22 nt) single strand RNA molecules predominantly shown in eukaryotes [1]. Most identified miRNAs showed evolutionary conservation throughout different species, some miRNAs even span both plants and animals and showed important and similar gene expression regulation pattern in different model species [2]. Another non-coding small RNA products called small interfering RNA (siRNA) showed similar to miRNA in biogenesis and downstream regulation. The processing of both miRNA and siRNAs are almost same mediated by Dicer and RISC, and the length are 21 to 23 nucleotides for siRNA and 18-25 nt for miRNA, generally. Unlike siRNA, which in general have perfectly complementarity to shut down gene expression. MiRNA, somehow, relies no so much perfect match to target its transcripts which makes it able to target a number of different transcripts. Another basic difference is siRNAs in general shut down gene expression at a post-transcriptional level through mRNA degradation, while miRNAs silence their target genes mainly through translational repression.

Based on the rapid breakthrough and findings in the past 20 years, many significant revelations have been reported in discovering miRNA biogenesis, target prediction and *ab initio* annotation. With the increasing development of the small RNA-sequencing (sRNA-seq) technology, thousands of plant microRNAs (miRNA) can be discovered with the high sequencing confidence, which provides an ebb of new data to analyze as well as a huge challenge for miRNA prediction and annotation. However, the miRNA annotation is still far from satisfaction. Here I review state of art miRNA discovery strategy and characterization methods that are suitable to plant genomic miRNA annotation and discuss the challenges for current novel miRNA annotation.

The High-Throughput Analysis of sRNA-seq

At present, prediction from sRNA-seq is still the most popular strategy for novel miRNA discovery and annotation. Usually, such sRNA-seq based miRNA prediction strategy contains three steps: 1) Raw data processing, 2) miRNA prediction and expression quantification and 3) Downstream analyses. For the raw data processing (using FastQC, Trimmomatic [3], Cutadapt [4], etc.), a high quality of read after trimming and filtering is necessary. Then reads were aligned to the genome using by a short-read aligner (using bwa [5], bowtie2 [6], etc.) which is the basis of miRNA prediction. To quantify the expression levels of miRNA, only uniquely aligned reads without mismatches should be used for the read counting (using HTSeq-count, featureCount, etc.), especially for those highly repetitive regions in plant genome. And for the downstream analyses, miRNA isoforms [7], differential expression, miRNA binding target prediction [8] are usually the current hot topics.

Plant Mirna Annotation is Still Beyond Satisfaction

For plant miRNAs annotation, the complex feature of their genome structure makes it quite challenging [9]. Unlike animals, there should be perfect match between the miRNA-mRNA interaction in plants, yet in animals only seed recognition will be sufficient for miRNA tar-

***Corresponding author:** Kan Liu, Department of Computer Science and Engineering, School of Biological Sciences, Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln NE 68588, USA

Accepted: November 01, 2018;

Published online: November 03, 2018

Citation: Liu K (2018) MicroRNA Prediction Using Small RNA-Sequencing, Advances and Challenges. Arch Crop Sci 2(1):58-60

geting. The plant genomes such as *Arabidopsis thaliana* and *Oryza sativa* are of high quality in both the genome assembly and gene structure annotation, are ideal and usually used as model species in novel miRNA annotation. However, other plant genome is usually more complex compared with animal genomes due to a higher proportion of repetitive sequences such as transposable elements, which burdened the precise annotation using short read sRNA-seq data. Some plant miRNAs with unknown structure and function annotation as well as low expression level however showed powerful effect on the gene expression regulation. As such, we might believe a lot of plant miRNAs are still beyond our research and annotation, especially for those low abundance plant miRNAs which are neglected in traditional method. With such difficulties, the precise identification and annotation of plant miRNAs that connect diversified biological processes for plant growth and development are still quite mysterious for us and also deserves much more analytic attention.

Current Bioinformatics Analysis Toolkit of sRNA-seq

Genome-based plant miRNA prediction methods mentioned above tend to present a higher false positive rate due to the inability of confirming hundreds and thousands of loci that showed hairpin structures. To test the sequence similarity, a lot of software used the alignment results of RNA-seq data for the prediction. Some all-in-one analysis toolkit such as UEA sRNA Workbench [10] and sRNAtoolbox [11] are designed using different miRNA prediction methods, which present a comprehensive solution for users without experience in coding and parameter tuning of those bioinformatics tools. Although using such strategy will provide a more complete list for the prediction, the hairpin and splicing is much more difficult to annotate, especially there is sequencing error and miRNA splicing regulation of your samples. Also, most miRNA annotation software is developed based on highly expressed and housing keeping miRNAs observed in several species, yet they failed to consider time and tissue specific miRNA expression [12]. To speak in detail, miRNA annotation strategies that are built exclusively on conserved information from model species might underperform when used for low expressed and tissue, development or cell type specific samples. Also, several miRNAs are derived from repetitive sequences from the same genome, which also increases the difficulty since reads that were mapped to multiple regions usually were discarded. Therefore, how to remove the noise and bias without knowing the complete miRNA datasets of given species is therefore another challenge.

sRNA-seq using next-generation sequencing technologies is becoming a mainstream for annotating miRNAs from different species in recent years with the fast development of bioinformatics software and R packages facilitated to characterize novel miRNAs by taking the advantage of high-throughput technology. However, a large number of tools will also produce unneglectable annotation error for high confidence miRNAs. Therefore, how to overcome the weakness of sRNA-seq should be combined with other genome-based methods such as RNA folding energy. The advent of more comprehensive and species-specific miRNA splicing isoform prediction is on the horizon [13]. miRNA splicing imprecision is another overlooked yet important area in miRNA prediction and annotation [14]. Different conditions using treatment such as knock-out argo factors, miRNA splicing precision can be severely affected by such treatments and this splicing imprecision is quite condition specific throughout not only plant but also animal species. In the future, to reduce the false positive prediction with the availability of public sRNA-seq data, a comprehensive evaluation of the candidate loci such as post hoc assessment is urgently needed for miRNA structural and functional analysis such as miRbase [15] and other comprehensive databases, especially for those genomes that are relatively compact to show a minimum miRNA category compared with plant that has large and complex genome.

References

1. Leung AKL (2015) The whereabouts of microRNA actions: Cytoplasm and beyond. Trends Cell Biol 25: 601-610.
2. Catalanotto C, C Cogoni, G Zardo (2016) MicroRNA in control of gene expression: An overview of nuclear functions. Int J Mol Sci 17.
3. Bolger AM, M Lohse, B Usadel (2014) Trimmomatic: A flexible trimmer for illumina sequence data. Bioinformatics 30: 2114-2120.
4. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal 17: 10-12.
5. Li H, R Durbin (2009) Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 25: 1754-1760.
6. Langmead B, SL Salzberg (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357-359.
7. Zhou H, Arcila ML, Li Z, et al. (2012) Deep annotation of mouse iso-miR and iso-moR variation. Nucleic Acids Res 40: 5864-5875.
8. Gupta R, RV Davuluri (2014) Bioinformatics approaches to the study of MicroRNAs, in non-coding RNAs and Cancer. Springer 165-245.
9. Ha M, VN Kim (2014) Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol 15: 509-524.
10. Beckers M, Mohorianu I, Stocks M, et al. (2017) Comprehensive processing of high-throughput small RNA sequencing data including quality checking, normalization, and differential expression analysis using the UEA sRNA workbench. RNA 23: 823-835.

11. Rueda A, Guillermo B, Ricardo Lebron, et al. (2015) sRNAtoolbox: An integrated collection of small rna research tools. Nucleic Acids Res 43: 467-473.

12. Bortolomeazzi M, E Gaffo, S Bortoluzzi (2017) A survey of software tools for microRNA discovery and characterization using RNA-seq. Brief Bioinform.

13. Bortoluzzi S, Bisognin A, Biasiolo M, et al. (2012) Characterisation and discovery of novel miRNAs and moRNAs in JAK2V617F mutated SET2 cells. Blood 119: e120-e130.

14. Capece V, Garcia Vizcaino JC, Vidal R, et al. (2015) Oasis: Online analysis of small RNA deep sequencing data. Bioinformatics 31: 2205-2207.

15. Kozomara A, S Griffiths-Jones (2014) miRBase: Annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42: 68-73.