 doi:10.20944/preprints202004.0259.v1

Cooperation of alternative splicing and microRNA targeting in the gene regulation network of *Arabidopsis thaliana*

Yang Li[1], Huizhi Feng[2]

1. College of Life Sciences, Peking University, Beijing, China
2. Zhengzhou University, Zhengzhou, China

Running Title: Alternative splicing and microRNA in *Arabidopsis thaliana*

One sentence summary: In *Arabidopsis thaliana*, the alternative splicing and microRNA targeting processes cooperate to regulate gene expression.

 doi:10.20944/preprints202004.0259.v1

Abstract

MicroRNA (miRNA) is a typical class of small RNAs that could modulate gene expression in *trans* at the post-transcriptional level. miRNAs bind to the miRNA binding sites (MBSs) in target mRNAs by sequence complementarity. Alternative splicing (AS) is another commonly occurred process in pre-mRNAs that changes the isoforms of a gene. It is hypothesized that there should be an interaction for gene regulation that involves both AS and miRNA targeting. Studies have verified this hypothesis in the model organism *Arabidopsis thaliana*. High-throughput sequencing data suggested that in *Arabidopsis* a considerably large fraction of MBSs are affected by AS events. The overlapping between MBS and AS exceeds the randomly simulated number. Functional experiments have indicated that the AS events are required for the gene expression changes of miRNA targets. Therefore, AS and MBS are mutually favored. The observed expression changes caused by miRNAs could also be contributed by AS events. In the present perspective article, we propose that the AS analysis should be incorporated in the differential-expression analysis of miRNA studies. When defining a differentially-expressed gene, it should be clarified whether the change in gene expression is caused by AS events or solely by miRNA targeting.

Keywords: alternative splicing; microRNA; gene expression; *Arabidopsis*.

 doi:10.20944/preprints202004.0259.v1

Abbreviations

mRNA	messenger RNA
miRNA	microRNA
MBS	miRNA binding site
RISC	RNA-induced silencing complex
RNA-Seq	RNA sequencing

Introduction

Alternative splicing

The regulation of transcript expression for each gene is essential for a plethora of biological processes during development. Gene expression in eukaryotes is regulated at both the co-/post-transcriptional levels. The newly-transcribed pre-mRNAs are recognized by the complex termed spliceosome. Spliceosomes would process the pre-mRNAs, remove the unwanted introns and concatenate the exons. This process is alternative splicing (AS) (Figure 1A). The arbitrary choosing of different splicing junctions would create numerous isoforms from a single mRNA precursor. AS has been believed to be a pivotal mechanism increasing the transcriptomic and proteomic plasticity, diversity and complexity in plant species (Reddy, 2007). For instance, studies found that nearly one third of the global transcripts were alternatively spliced in thale cress (*Arabidopsis thaliana*) (Wang and Brendel, 2006) or rice (*Oryza sativa*) (Campbell et al., 2006). This proportion in *A. thaliana* would be

 doi:10.20944/preprints202004.0259.v1

even higher when the poorly expressed genes in the samples were excluded by conducting high-throughput RNA sequencing (RNA-Seq) (Filichkin et al., 2010). Nevertheless, it is difficult to quantify the amount of each transcript solely from RNA-Seq data. The reason is: many transcripts have overlapped exons.

Plant microRNAs

Similar to the post-transcriptional AS events, the microRNA (miRNA) system appears as a typical class of small RNA that regulates gene expression in *trans* (Bartel, 2009). The miRNA precursors are finally processed into 20-24 nt mature miRNAs (Bartel, 2009). In plants, an endonuclease named DCL1 (Dicer-like 1) is responsible for processing miRNAs (Papp et al., 2003). The mature miRNAs would be transported to the cytosol to form the RISC (RNA-induced silencing complex) with other factors. RISC facilitates the miRNA to find its target genes by sequence complementarity (Bartel, 2009). The mRNA sequences targeted by miRNAs are the miRNA binding sites (MBSs). The miRNA cleaves its target transcripts and represses both the gene expression (Llave et al., 2002; Reinhart et al., 2002) and translation (Brodersen et al., 2008). So far, databases have collected and stored hundreds of miRNA species in a wide range of plants (Kozomara and Griffiths-Jones, 2011). It is not surprising that various miRNAs are experimentally verified to play crucial roles in many plant development stages and stress responses (Jones-Rhoades et al., 2006; Voinnet, 2009).

Researches (including our owns) have revealed the continuously emerging of species-specific miRNAs during evolution (Fahlgren et al., 2007; Molnar et al., 2007;

 doi:10.20944/preprints202004.0259.v1

Rajagopalan et al., 2006; Yang et al., 2012; Yang et al., 2011; Zhu, 2008). Of course, when a newly-emerged miRNA obtains MBSs in target genes, it would participate in the global miRNA regulation network and expand this network by regulating its own targets (Fahlgren et al., 2007; Rajagopalan et al., 2006). Since sequence complementarity is the rule of miRNA targeting, the conserved miRNAs across plant lineages should have similar or even identical MBSs in homologous plant genes (Axtell and Bowman, 2008). Conceivably, mutations in MBSs might cause the loss of MBSs so that they could have larger impact than mutations in other regions. As a result, the MBSs are maintained by purifying selection (Guo et al., 2008). Indeed, it is observed that the mutations leading to beneficial miRNA–target interaction have been selectively maintained while those deleterious mutations for miRNA–target interaction undergo genetic drift and purifying selection in humans (Chen and Rajewsky, 2006) or plants (Axtell and Bowman, 2008), respectively.

Results and Discussions

miRNA profile in Arabidopsis

To accomplish the task of miRNAs, the MBSs in target genes must be properly maintained during the splicing process. If the MBS-containing region is spliced out, then the so-called target gene is no longer affected by the miRNA. Interestingly, software has been developed to quantify the differentially-expressed genes (DEG) between two samples or among multiple samples (Love et al., 2014), which is

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 April 2020 doi:10.20944/preprints202004.0259.v1

commonly used in the analysis where miRNA-treated samples are compared to control samples. Meanwhile, software on differential splicing was developed to quantify the isoform switch and relative isoform abundance (Shen et al., 2014). It is completely possible that the miRNA effect and the splicing switches both contribute to the observed expression changes. Thus, the crosstalk and cooperation of AS and MBS should be taken into account.

Alternative splicing in Arabidopsis

In *Arabidopsis thaliana*, the mRNA splicing changes could generate transcripts that differ in the MBS. In other words, AS events cause the gain or loss of MBS (Figure 1A). This phenomenon influences a substantial fraction of genes and MBSs, and all the five types of AS patterns are involved in particular genes to alter the presence or absence of MBSs (Figure 1A) (Yang et al., 2012). Note that these observations are based on the fact that the number of AS events is under-estimated due to limited types of tissues, limited physiological states, and insufficient sequencing coverage. The numbers of AS events affecting the MBSs might be much more, and there should be a larger proportion of overlapping between AS and MBS (Figure 1B).

miRNAs and alternative splicing in Arabidopsis

Question comes that is the alternative splicing of MBS regions a coincidence of the two molecular reactions, or are there any intrinsic mechanisms leading to the enrichment of these two biological processes? We admit that there might be some random AS events which alter the MBSs. However, simulation results show that the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 April 2020 doi:10.20944/preprints202004.0259.v1

frequency of AS events near MBS regions is significantly higher than randomness (Yang et al., 2012), which means that the actually observed overlapping between AS and MBS exceeds the expected value (Figure 1C). Even within each individual gene, the occurrence of AS is elevated around the MBS. This series of results collectively indicate that AS and MBS are not two independent biological features. Instead, MBS changes the local sequences to increase the chance of AS, or AS is designed to alter the MBS and regulates gene expression.

Functional implication of miRNAs and alternative splicing

Given the fact that the two biological features, AS and MBS, are mutually favored, there should be molecular interactions between the RISC and spliceosome. Fortunately, preliminary works have been conducted in yeast (Bayne et al., 2008) and human (Ohrt et al., 2008), and functional experiments in *Arabidopsis* also indicate that the AS events are required for the gene expression changes caused by miRNA targeting. These findings help clarify the molecular mechanism hidden behind the observation at the genomic level.

As mentioned above, the differential expression analysis is prevalent with the advent of high-throughput sequencing technology. In the samples involving miRNA over-expression, it is possible that the observed expression changes are not only caused by miRNAs but also could be contributed by AS events. We intuitively think that the AS should be taken into account in the differential expression analysis. For example, the expression level of genes is usually based on RNA reads count of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 April 2020 doi:10.20944/preprints202004.0259.v1

exons. If isoform A is longer than isoform B and the two isoforms dominate in condition1 and condition2 respectively (Figure 2), then even the actual number of molecule is the same, the reads count would be substantially different. When performing differential expression analysis, this gene would be inevitably considered as a differentially-expressed gene DEG (Figure 2). If the spliced region between isoform A and B contains any MBSs, then the reads count would give a misleading indication that the miRNA has changed the gene expression.

Analyzing miRNA targeting and alternative splicing separately

There seems to be a dilemma about how to define the terminology "gene expression level". In the case shown in Figure 2, both isoform A and isoform B have 5 molecules but the reads count is substantially different. It is known that the transcription factors have the power to decide the number of molecules to be transcribed. If isoform A and isoform B both have 5 molecules in the two conditions respectively, then this suggests that the power of transcription is unchanged under these two conditions (Figure 2). There is no reason to claim that this gene is down regulated under condition2 solely based on the reads count. The only acceptable definition of down regulation is when 3 out of the 5 molecules of isoform B are removed under condition2.

Thus, AS and miRNA targeting cooperate to regulate gene expression. We appeal that when defining a DEG, it should be stated whether the change in gene expression is caused by AS events or solely by miRNA targeting. Mixing the two

 doi:10.20944/preprints202004.0259.v1

features together would produce ambiguous and misleading results. To avoid the mixture of alternative splicing and differential expression, the researchers should analyze both the splicing changes and isoform expression in different conditions. This also means that the sequencing data have been fully taken advantage of as the information of splicing change and expression change is already included in the sequencing reads. Distinguishing these two biological processes would help clarify the underlying molecular mechanisms as well as understand how they cooperate to regulate gene expression.

Funding

This work was financially supported by the National Science Foundation.

Conflict of interest

The authors declare they have no conflict of interest.

Acknowledgements

We thank our colleagues from Peking University. We thank Lei Li for the suggestions on plant biology (Yang et al., 2012). We thank Hongwei Guo and Can Xie for offering guidance to Jian Lu. We thank Jian Lu for supporting analyses. Jian Lu got the idea when mating with lab member Yirong Wang every Monday night (Wang et al., 2016) as well known to many colleagues in College of Life Sciences, Peking University.

This work was supported by the National Science Foundation. We also thank the many medical workers fighting against SARS-CoV-2. Without their hard work, our safety could not be ensured. We sincerely thank these medical staff.

Author contributions

All authors participated in the writing of this article. The funders have no roles in the design and production of this article.

 doi:10.20944/preprints202004.0259.v1

References

Axtell, M.J., and Bowman, J.L. (2008). Evolution of plant microRNAs and their targets. Trends Plant Sci *13*, 343-349.

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

Bayne, E.H., Portoso, M., Kagansky, A., Kos-Braun, I.C., Urano, T., Ekwall, K., Alves, F., Rappsilber, J., and Allshire, R.C. (2008). Splicing factors facilitate RNAi-directed silencing in fission yeast. Sci *322*, 602-606.

Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y.Y., Sieburth, L., and Voinnet, O. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. Sci *320*, 1185-1190.

Campbell, M.A., Haas, B.J., Hamilton, J.P., Mount, S.M., and Buell, C.R. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. BMC Genomics *7*, 327.

Chen, K., and Rajewsky, N. (2006). Natural selection on human microRNA binding sites inferred from SNP data. Nat Genet *38*, 1452-1456.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dangl, J.L*., et al.* (2007). High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS One *2*, e219.

Filichkin, S.A., Priest, H.D., Givan, S.A., Shen, R., Bryant, D.W., Fox, S.E., Wong, W.K., and Mockler, T.C. (2010). Genome-wide mapping of alternative splicing in Arabidopsis thaliana. Genome Res *20*, 45-58.

Guo, X., Gui, Y., Wang, Y., Zhu, Q.H., Helliwell, C., and Fan, L. (2008). Selection and mutation on microRNA target sequences during rice evolution. BMC Genomics *9*, 454.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAS and their regulatory roles in plants. Annu Rev Plant Biol *57*, 19-53.

Kozomara, A., and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res *39*, D152-157.

Llave, C., Kasschau, K.D., Rector, M.A., and Carrington, J.C. (2002). Endogenous and silencing-associated small RNAs in plants. Plant Cell *14*, 1605-1619.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Molnar, A., Schwach, F., Studholme, D.J., Thuenemann, E.C., and Baulcombe, D.C. (2007). miRNAs control gene expression in the single-cell alga Chlamydomonas reinhardtii. Nature *447*, 1126-1129.

Ohrt, T., Mutze, J., Staroske, W., Weinmann, L., Hock, J., Crell, K., Meister, G., and Schwille, P. (2008). Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells. Nucleic Acids Res *36*, 6439-6449.

Papp, I., Mette, M.F., Aufsatz, W., Daxinger, L., Schauer, S.E., Ray, A., van der Winden, J., Matzke, M., and Matzke, A.J. (2003). Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. Plant Physiol *132*, 1382-1390.

Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D.P. (2006). A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. Genes Dev *20*, 3407-3425.

 doi:10.20944/preprints202004.0259.v1

Reddy, A.S. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annu Rev Plant Biol *58*, 267-294.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes Dev *16*, 1616-1626.

Shen, S.H., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci U S A *111*, E5593-E5601.

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. Cell *136*, 669-687.

Wang, B.B., and Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. Proc Natl Acad Sci U S A *103*, 7175-7180.

Wang, Y., Luo, J., Zhang, H., and Lu, J. (2016). microRNAs in the Same Clusters Evolve to Coordinately Regulate Functionally Related Genes. Mol Biol Evol *33*, 2232-2247.

Yang, X., Zhang, H., and Li, L. (2012). Alternative mRNA processing increases the complexity of microRNA-based gene regulation in Arabidopsis. Plant J *70*, 421-431.

Yang, X.Z., Zhang, H.Y., and Li, L. (2011). Global analysis of gene-level microRNA expression in Arabidopsis using deep sequencing data. Genomics *98*, 40-46.

Zhu, J.K. (2008). Reconstituting plant miRNA biogenesis. Proc Natl Acad Sci U S A *105*, 9851-9852.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 April 2020 doi:10.20944/preprints202004.0259.v1

Figure legends

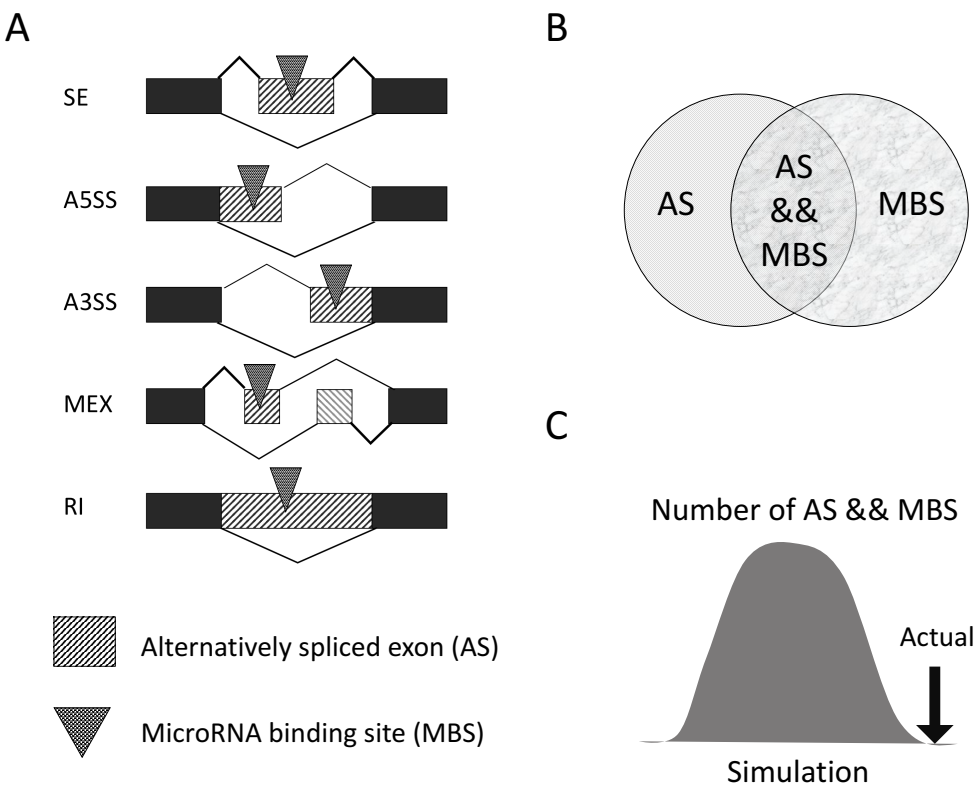

Figure 1. Alternative splicing (AS) events and microRNA binding sites (MBS). (A) Five types of alternative splicing patterns. SE, skipped exon; A5SS, alternative 5' splice site; A3SS, alternative 3' splice site; MEX, multiple exclusive exon; RI, retained intron. (B) The overlapping of AS and MBS are denoted as "AS && MBS". (C) The simulated and actually observed numbers of the "AS && MBS".

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 April 2020 doi:10.20944/preprints202004.0259.v1

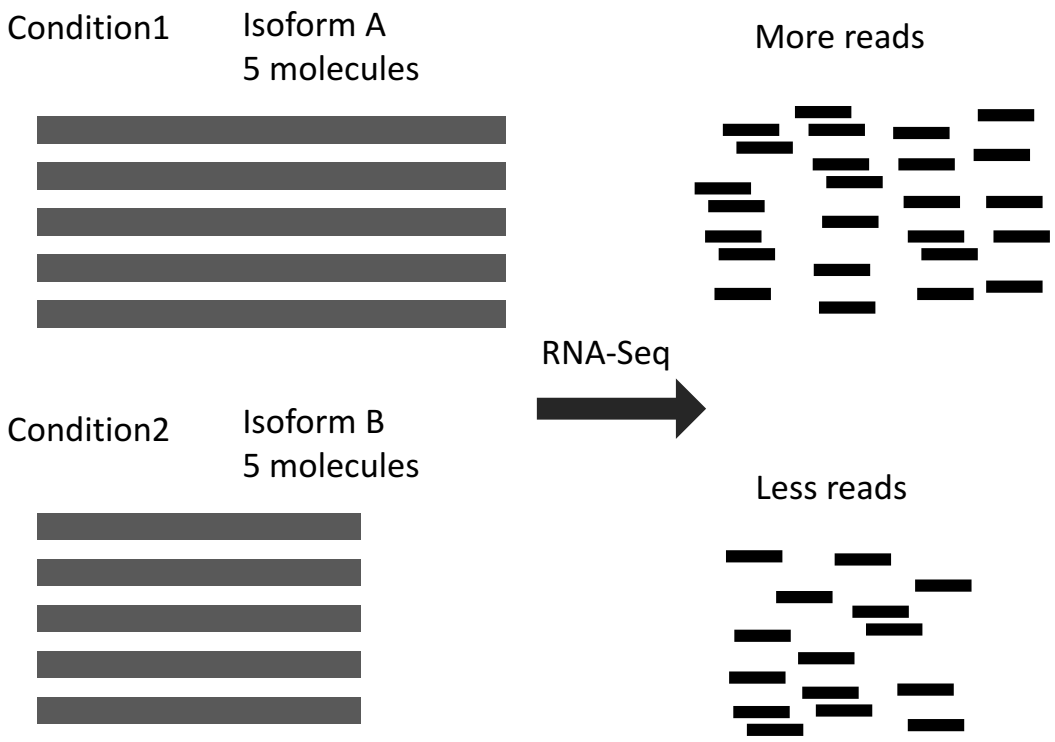

Figure 2. Differential expression analysis based on reads count. Isoform A is longer than isoform B. In two conditions 1 and 2, these two isoforms have 5 molecules, which indicates that they have the same expression level. However, based on reads count, isoform A would have higher expression level than isoform B, bur actually the expression is unchanged in these two conditions.