



University of Kentucky
UKnowledge

Theses and Dissertations--Plant and Soil
Sciences

Plant and Soil Sciences


2017

THE ROLE OF ALTERNATIVE POLYADENYLATION MEDIATED BY CPSF30 IN *ARABIDOPSIS THALIANA*

Guijie Hao

University of Kentucky, guijiehao@gmail.com

Author ORCID Identifier:

 <https://orcid.org/0000-0002-7899-0441>

Digital Object Identifier: <https://doi.org/10.13023/ETD.2017.303>

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Recommended Citation

Hao, Guijie, "THE ROLE OF ALTERNATIVE POLYADENYLATION MEDIATED BY CPSF30 IN *ARABIDOPSIS THALIANA*" (2017). *Theses and Dissertations--Plant and Soil Sciences*. 90.

https://uknowledge.uky.edu/pss_etds/90

This Doctoral Dissertation is brought to you for free and open access by the Plant and Soil Sciences at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Plant and Soil Sciences by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Guijie Hao, Student

Dr. Arthur G. Hunt, Major Professor

Dr. Mark Coyne, Director of Graduate Studies

THE ROLE OF ALTERNATIVE POLYADENYLATION MEDIATED BY CPSF30
IN *ARABIDOPSIS THALIANA*

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Agriculture, Food and Environment
at the University of Kentucky

By
Guijie Hao

Lexington, Kentucky

Director: Dr. Arthur G. Hunt, Professor of Plant Physiology

Lexington, Kentucky

2017

Copyright © Guijie Hao 2017

ABSTRACT OF DISSERTATION

THE ROLE OF ALTERNATIVE POLYADENYLATION MEDIATED BY CPSF30 IN *ARABIDOPSIS THALIANA*

Drought stress is considered one of the most devastating abiotic stress factors that limit crop productivity for modern agriculture worldwide. There is a large range of physiological and biochemical responses induced by drought stress. The responses range from physiological and biochemical to regulation at transcription and posttranscriptional levels. Post-transcription, the products encoded by eukaryotic genes must undergo a series of modifications to become a mature mRNA. Polyadenylation is an important one in terms of regulation. Polyadenylation impacts gene expression through determining the coding and regulation potential of the mRNA, especially when different mRNAs from the same gene may be polyadenylated at more than one position. This alternative polyadenylation (APA) has numerous potential effects on gene regulation and function. I have studied the impact of drought stress on APA, testing the hypothesis that drought stress may give rise to changes in the usage of poly(A) sites generating different mRNA isoforms. The results showed that usage of poly(A) sites that lie within 5'-UTRs and coding sequence (CDS) changes more than usage of sites in other regions due to drought stress.

Alternative polyadenylation is mediated by the polyadenylation complex of proteins that are conserved in eukaryotic cells. The *Arabidopsis* CPSF30 protein (AtCPSF30), which is an RNA-binding endonuclease subunit of the polyadenylation complex, plays an important role in controlling APA. Previous study showed that poly(A) site choice changes on a large scale in *oxidative stress tolerant 6 (oxt6)*, a mutant lacking AtCPSF30. Within the mutant/WT genotypes, there are three classes of poly(A) site, wild type specific, *oxt6* specific, and common (both in wild type and mutant). The wild type specific and *oxt6* specific mRNAs make up around 70% of the total of all mRNA species. I hypothesize that the stability of these various mRNA isoforms should be different, and that this is a possible way that AtCPSF30 regulates gene expression. I tested this by assessing the influence poly(A) sites can have on the mRNA isoform's stability in the wild type and *oxt6* mutant. My results show that most mRNA isoforms show similar stability profiles in the wild-type and mutant plants. However, the mRNA isoforms derived from polyadenylation within CDS are much more stable in

the mutant than the wild-type. These results implicate AtCPSF30 in the process of non-stop mRNA decay.

Messenger RNA polyadenylation occurs in the nucleus, and the subunits of the polyadenylation complex that mediate this process are expected to reside within the nucleus. However, AtCPSF30 by itself localizes not only to the nucleus, but also to the cytoplasm. AtCPSF30 protein contains three predicted CCCH-type zinc finger motifs. The first CCCH motif is the primary motif that is responsible for the bulk of its RNA-binding activity. It can bind with calmodulin, but the RNA-binding activity of AtCPSF30 is inhibited by calmodulin in a calcium-dependent manner. The third CCCH motif is associated with endonuclease activity. Previous studies demonstrated that the endonuclease activity of AtCPSF30 can be inhibited by disulfide reducing agents. These published results suggest that there are proteins that interact with AtCPSF30 and act through calmodulin binding or disulfide remodeling. To test this hypothesis, I screened for proteins that interact with AtCPSF30. For this, different approaches were performed. These screens led me to two proteins-one protein that is tyrosine-phosphorylated and whose phosphorylation state is modulated in response to ABA, which well-known ABA regulates guard cell turgor via a calcium-dependent pathway, and the other is ribosome protein L35(RPL35), which plays an important role in nuclear entry, translation activity, and endoplasmic reticulum(ER) docking. These results suggest that multiple calcium-dependent signaling mechanisms may converge on AtCPSF30, and AtCPSF30 might be directly interact with ribosome protein.

Keywords: AtCPSF30, Alternative polyadenylation, Drought stress, Protein-protein interaction, *Arabidopsis thaliana*

Guijie Hao

Student's Signature

June 23, 2017

Date

THE ROLE OF ALTERNATIVE POLYADENYLATION MEDIATED BY CPSF30
IN *ARABIDOPSIS THALIANA*

By

Guijie Hao

Arthur G. Hunt

Director of Dissertation

Mark Coyne

Director of Graduate Studies

June 23, 2017

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to my major advisor Dr. Arthur G. Hunt, for his academic guidance, financial support and friendly assistance during my study at University of Kentucky. I also would like to the Department of Plant and Soil Sciences and Graduate School for their financial support. Gratitude is also expressed to my committee members Dr. Bruce Downie, Dr. Ling Yuan, and Dr. Sharyn Perry for their helpful suggestions and encouragement. In addition, I would give a special thank you to Dr. Daniel Howe for his willingness to serve as my outside examiner. I would also like to thank Dr. Michael Goodin for his time and patience to teach me how to use the confocal microscope.

I would like to express my appreciation to the members of Hunt Lab, Dr. Manohar Chakrabarti, Dr. Laura de Lorenzo, and Carol Von Lanken for their excellent technical assistance and friendship. I would like to acknowledge Dr. Lynnette Dirk and Dr. Downie for providing the Seed Phage Library that I used to screen for CPSF30 interacting protein. I would like to express my gratitude to the Dr. Sitakanta Pattanaik and Xueyi Sui for guidance in setting up the yeast two hybrids assay in Dr. Ling Yuan's Lab. Also, I would love to express my gratitude to Dr. Jozsef Stork and Andrea Sanchez for their help.

I would express my appreciation to my family. The first appreciation is delivered to my parents. Your love and care is the priceless motivation for me to pursue and overcome each hard time. I would like to express my second appreciation to my sister Guixia Hao. Your encouragement and support allowed me to have this chance to pursue my doctor degree. Finally, I wish to have my most important thanks to my husband, Zhengshi Yang. There are not enough words to express my appreciation and love. Your devotion, sacrifice, encouragement, are the driving forces which instill in me the strong desire to complete my PhD.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	vi
LIST OF FIGURES	vii
Chapter One: Literature review	1
1.1 Nuclear polyadenylation in eukaryotes	1
1.1.1 The definition of polyadenylation	1
1.1.2 The function and significance of polyadenylation	1
1.1.3 Polyadenylation factor CPSF30 is involved in 3' end processing	2
1.2 Alternative polyadenylation in Eukaryotes	10
1.2.1 The definition and types of alternative polyadenylation	10
1.2.2 Widespread occurrence of alternative polyadenylation	11
1.2.3 Biological significance of alternative polyadenylation under stress	17
1.3 RNA stability in eukaryotes	18
1.3.1 Nuclear mRNA degradation pathway	18
1.3.2 Cytoplasm mRNA degradation pathway	19
1.3.3 RNA quality control pathways in Eukaryotes	22
1.3.4 The significance of mRNA stability	27
1.4 Summary	28
Chapter Two: Alternative polyadenylation as a potential mechanism to respond to drought stress in <i>Arabidopsis thaliana</i>	34
2.1 Introduction	34
2.2 Results	36
2.2.1 Validation of drought stress treatments with qRT-PCR	36
2.2.2 Preparation of high-throughput poly(A) tag libraries and sequencing	38
2.2.3 Drought stress incites changes in poly(A) site choice in different genomic regions	39
2.2.4 Annotation and functional analysis of poly(A) site choice in 5' UTR and CDS regions	42
2.3 Discussion	44
2.3.1 Comments regarding the validation of the drought stress treatments	44
2.3.2 Alternative polyadenylation and the molecular function of plants to drought stress	45
2.3.3 Alternative polyadenylation factors might be involved in drought stress	47
2.4 Material and Methods	50
2.4.1 Plant material and treatment	50
2.4.2 qRT-PCR condition and analysis	50
2.4.3 Poly(A) tag library preparation and sequencing	51
2.4.4 Poly(A) tag and gene expression analysis	51
2.4.5 Relative poly(A) site usage analysis	52
2.4.6 GO analysis	53

Chapter Three: AtCPSF30-mediated alternative polyadenylation controls mRNA stability	74
3.1 Introduction	74
3.2 Results	76
3.2.1 Overview of experimental approach	76
3.2.2 Validation of the cordycepin treatment	77
3.2.3 The global mRNA stabilities in the wild type and <i>oxf6</i> mutant	77
3.2.4 The stabilities of mRNA isoform defined by their dependences on AtCPSF30	78
3.2.5 Annotation and functional analysis of mRNA isoforms that AtCPSF30-dependent and -independent	80
3.3 Discussion	84
3.3.1 The stability of noncanonical mRNA isoforms derived from polyadenylation with introns	84
3.3.2 The stability of noncanonical mRNA isoforms in CDS	86
3.3.3 The stability of noncanonical mRNA isoforms in 5'UTR	88
3.3.4 A model to describe stability of nonstop transcripts	89
3.3.5 A model to describe the 5'-UTR transcript isoform stability	90
3.4 Materials and method	90
3.4.1 Plant material and treatment	Error! Bookmark not defined.
3.4.2 Poly(A) tag library preparation and sequencing	91
3.4.3 Poly(A) tag and gene expression analysis	91
3.4.4 GO analysis	92
Chapter Four: A screen for proteins that interact with AtCPSF30	106
4.1 Introduction	106
4.2 Result	107
4.2.1 Selected candidates interacting with AtCPSF30 from phage display	107
4.2.2 Yeast two hybrid assay to confirm the three candidates	109
4.2.3 Bimolecular Fluorescence Complementation to confirm corresponding interactions	110
4.3 Discussion	111
4.3.1 The approaches to identify potential interacting proteins of AtCPSF30	111
4.3.2 Possible connection between AtCPSF30 and expression protein of OBAP1A	112
4.3.3 Possible connection between AtCPSF30 and RPL35	114
4.3.4 Summary	116
4.4 Material and method	116
4.4.1 Vector Construction and site-directed mutagenesis	116
4.4.2 Protein expression and purification	117
4.4.3 Phage display sequencing	117
4.4.4 Yeast two hybrid assays	120
4.4.5 Subcellular localization and BiFC	121
Chapter Five: Summary and future perspectives	136
REFERENCES	139
VITA	162

LIST OF TABLES

Table 2.1 Selected reference genes for Arabidopsis and corresponding primer pair information.....	54
Table 2.2 PAT-seq sequence data information for different concentration of mannitol treatment	55
Table 2.3 PAT-seq sequence data information for different time of mannitol treatment.	56
Table 4.1 Primers for AtCPSF30 and its mutant	123
Table 4.2. Primers for phage libraries sequencing.....	124
Table 4.3 Primers for yeast two hybrid.....	125
Table 4.4. Primers for BiFC.....	126

LIST OF FIGURES

Figure 1.1 Mature mRNA in eukaryotes.....	31
Figure 1.2 Non-sense mediated decay pathway (cited from Lykke-Andersen et al, 2014)	32
Figure 1.3 Non-stop decay pathway (cited from Lykke-Andersen et al, 2014)	33
Figure 2.1. Overview of experimental approach to study APA under drought stress	57
Figure 2.2 qRT-PCR to validate different concentration treatment under drought stress	58
Figure 2.3 qRT-PCR to validate different time treatment under drought stress	59
Figure 2.4 Preparation for PAT libraries and sequencing data analysis workflow	60
Figure 2.5 Drought stress incites significant global gene expression changes	64
Figure 2.6 Overview of poly(A) site usage analysis.....	65
Figure 2.7 Drought induced by 300mM mannitol treatment for two weeks alters non- canonical poly(A) site usage.....	66
Figure 2.8 Drought induced by 300mM mannitol treatment for different time intervals leads to changes in the non-canonical poly(A) site usage	67
Figure 2.9 Enriched GO categories among the genes displaying APA after drought stress induced by different concentrations of mannitol treatments for two weeks	70
Figure 2.10 Enriched GO categories among the transcript isoforms terminating at poly(A) site located at 5'UTRs	71
Figure 2.11 Enriched GO categories among the transcript isoforms terminating at poly(A) site located at CDSs	73
Figure 3.1 Overview of experimental approach to study AtCPSF30-mediated APA controls mRNA stability	94
Figure 3.2 Overview of the method to calculate mRNA stability	95
Figure 3.3 Cordycepin incited significant down-regulation gene expression.....	97
Figure 3.4 The stabilities of non-canonical mRNA isoforms incites global changes.....	98
Figure 3.5. Stabilities of noncanonical mRNA isoforms from three classes of poly(A) sites changes.....	99
Figure 3.6 Enriched GO categories among the genes displaying APA after cordycepin treatment	100
Figure 3.7 The gene ontology analysis for mRNA isoform that fall into 5'UTR.....	101

Figure 3.8 The gene ontology analysis for mRNA isoforms that fall into CDS in wt-specific	102
Figure 3. The gene ontology analysis for mRNA isoforms that fall into CDS in oxt6-specific	103
Figure 3.10 A model to describe the mRNA isoform stability in the CDS region	104
Figure 3.11 A model to describe the mRNA isoform in the 5'UTR region	105
Figure 4.1 Overall graphic depiction of the phage display process (modified from Kushwaha, et al. 2014).....	127
Figure 4.2 Titering for seed phage library I.....	128
Figure 4.3 The flowchart for phage-sequencing data analysis	129
Figure 4.4 Selected Exclusive candidates interact with AtCPSF30	130
Figure 4.5 Correct or incorrect candidate as examples.....	131
Figure 4.6. Yeast two hybrid to confirm interactions with AtCPSF30.....	132
Figure 4.7 Subcellular location of At1g30460, At1g05510 and At5g06210.....	133
Figure 4.8. AtCPSF30 interacts with At5g02610 in the nucleus.....	135

Chapter One: Literature review

1.1 Nuclear polyadenylation in eukaryotes

1.1.1 The definition of polyadenylation

The expression of eukaryotic genes involves a series of modifications of the primary transcript to become a mature mRNA, including 5'-capping, exon/intron splicing and 3'-end polyadenylation. One of these steps, the addition of a poly(A) tail, is a very important step for converting the pre-mRNA to a mature mRNA (Fig 1.1). During this process, around 50-200 adenosine residues are added to the 3'-end of the processed pre-mRNA (Proudfoot et al. 2002). All eukaryote mRNA 3'ends, except for animal replication-dependent histone transcripts, are created by a two-step reaction which involves an endonucleolytic cleavage of the pre-mRNA followed by the synthesis of poly(A) tail onto the upstream cleavage product (Colgan and Manley 1997).

1.1.2 The function and significance of polyadenylation

The polyadenylation process is determined by specific signals in the transcript and the type of RNA polymerase used for transcription (Richard and Manley 2009). During this process, the polyadenylation also competes and/or cooperates with other events such as capping and splicing (Tian et al. 2007; Xing and Li 2011b). The functions of the poly(A) tail of mRNAs include protecting the mRNA from degradation, aiding in localization, promoting translational efficiency, and facilitating the transport of mRNAs from nucleus to cytoplasm (Guhaniyogi and Brewer 2001).

As mentioned above, except for animal replication-dependent histone mRNAs, almost all eukaryotic mRNAs are polyadenylated (Richard and Manley 2009; Xing and Li 2011b). Replication-dependent histone mRNAs are the only mRNAs without poly(A) tails, instead

ending in a stem-loop structure followed by a purine-rich sequence (Tian et al. 2007; Dávila López and Samuelsson 2008). Non-coding RNAs are polyadenylated, as are the precursors for microRNAs (Yoshikawa et al. 2005; Saini et al. 2007). Many protein-coding genes have more than one polyadenylation site, and polyadenylation thereby can produce more than one transcript; this process is termed alternative polyadenylation (APA) (Tian et al. 2005; Danckwardt et al. 2008). If the poly(A) tail is added to other regions of the transcript, it may give rise to the production of a truncated protein or RNA that can impact the functions of associated genes (Guhaniyogi and Brewer 2001).

1.1.3 Polyadenylation factor CPSF30 is involved in 3' end processing

mRNA 3' end processing is controlled by sequence elements in the pre-mRNA (*cis* elements) and polyadenylation factors (Tian et al. 2005). In mammals, the pre-mRNA contains three primary sequence elements and two auxiliary sequence elements. The three primary sequence elements consist of the hexamer AAUAAA (the polyadenylation signal), the cleavage site and the G/U rich downstream element. The sequence AAUAAA hexamer (or its close variant AUUAAA) is the most highly conserved motif in eukaryotes, each nucleotide substitution in this hexamer reduces or abolishes mRNA 3'-end processing (Sheets et al. 1990; Proudfoot 2011). The two auxiliary sequence elements are an upstream element and a downstream element that regulates the 3'-end processing reaction (Xing and Li 2011b).

The polyadenylation factors are a complex of 15 to 20 proteins that are required for cleavage and polyadenylation (Mandel et al. 2008). The hexamer AAUAAA is specifically recognized by the Cleavage and Polyadenylation Specificity Factor (CPSF) (Sheets et al. 1990; Proudfoot 2011), a protein complex which contains six subunits: CPSF30, CPSF73,

CPSF100, CPSF160, WD repeat domain 33 (Wdr33) and factor interacting with PAP1 (Fip1) (Takagaki et al. 1988; Zhao et al. 1999a). Other poly(A) factors are also required for the complete reaction, including the Cleavage Stimulatory Factor (CstF), Cleavage Factor I and Cleavage Factor II (CF I/II), and symplekin (Keller et al. 1991).

1.1.3.1 CPSF30 is involved in animal and yeast pre-RNA processing

Among the CPSF subunits, one of the most interesting is the CPSF30, the 30kDa cleavage and specificity polyadenylation factor (Addepalli and Hunt 2007b). Initial experiments showed that CPSF30 might be only important for polyadenylation, but it is now known that it is required for both cleavage and polyadenylation (Mandel et al. 2008; Shi et al. 2009; Chan et al. 2014).

In animals, CPSF30 is a relatively small polypeptide with a characteristic set of five CCCH-type zinc finger motifs (Cys-X8-Cys-X5-Cys-X3-His), and a C-terminal CCHC-type zinc knuckle while CCHC zinc knuckle with the consensus sequence CX2CX4HX4C (Barabino et al. 1997; Barabino et al. 2000; Addepalli and Hunt 2007b). Initially, CPSF30 was not consistently detected with active CPSF preparations, but could be immunoprecipitated with other CPSF factors. In particular, hFip1 is intimately associated with CPSF30 as well as other polyadenylation factors such as the Poly A-Polymerase (PAP) (Mandel et al. 2008). Consistent with this, CPSF30 interacts with hFip1. Other eukaryotic (animal) homologs of CPSF30 are RNA-binding proteins, and the greatest similarity between all of these involves the second of the five zinc finger motifs (Bai and Tolias 1996; Bai and Tolias 1998). Interestingly, the ortholog of CPSF30 from *Drosophila* has nucleolytic activity, raising the possibility that CPSF30 may be a processing

endonuclease in the polyadenylation reaction (Barabino et al. 1997; Bai and Tolia 1998; Kaufmann et al. 2004).

The yeast homolog of CPSF30 is Yth1, an RNA-binding protein that is part of the so-called cleavage and polyadenylation factor (CPF). Yth1 can bind RNA near the cleavage/polyadenylation site, and *in vitro* experiments showed that it is required for cleavage and polyadenylation (Barabino et al. 1997; Bai and Tolia 1998). Yth1 physically interacts with Fip, and the RNA-binding by Yth1 can interfere with this association (Bai and Tolia 1996; Zarudnaya et al. 2002). The RNA binding activity of Yth1 is found in the second zinc finger motif of this protein (Barabino et al. 2000; Tacahashi et al. 2003). In addition, the fourth zinc finger motif is involved in the interaction with Fip1, indicating a distribution of functions among different parts of Yth1 (Zhao et al. 1999a; Tacahashi et al. 2003). From sequence alignment, ZF2 is the most conserved zinc finger in CPSF30, and displays 76% identity and 96% similarity between yeast and mammals. In ZF2, point mutation of the conserved Cys residue are lethal (Tacahashi et al. 2003).

Recent studies showed that only four polypeptides (CPSF30, CPSF160, hFip1 and WDR33) are necessary and sufficient to reconstitute a CPSF subcomplex active in AAUAAA-dependent polyadenylation, whereas the CPSF73, CPSF100 and symplekin subunits are dispensable (Barabino et al. 2000). Additionally, another study showed that CPSF30 and WDR33 directly contact AAUAAA (Chan et al. 2014). The CPSF30-RNA interaction plays a pivotal role for mRNA 3' end processing and is targeted by the influenza protein NS1A to inhibit host mRNA 3' end processing (Tsu et al. 2006). In lung adenocarcinoma cell lines and tumor tissues, CPSF30 is expressed at higher levels when compared to the normal tissues (Chen et al. 2013). Remarkably, higher expression of

CPSF30 is significantly correlated with the poor survival of patients (Chen et al. 2013). Moreover, the proliferation of lung cancer cells can be suppressed by knocking-down the expression of CPSF30 with siRNA. The over-expression of CPSF30 in lung cancer activated human telomerase reverse transcriptase (hTERT), which may promote the proliferation of cancer cells (Chen et al. 2013; Chen 2014). These results indicate that changes in the activity of CPSF30 may cause large scale reprogramming of polyadenylation, and suggest that the interactions with other factor such as NS1 can change the functioning of CPSF30 in vivo (Nemeroff et al. 1998).

1.1.3.2 Characteristics of the plant CPSF30

Plants have a polyadenylation complex subunit that is evolutionarily related to CPSF30, and the core of the protein (a central set of three zinc finger motifs) is conserved in animals, yeast and plants. Unlike mammals, yeast, and *Drosophila*, the gene encoding CPSF30 in *Arabidopsis* can produce two transcripts that result from the different poly(A) site usage (Zhang et al. 2008). The smaller transcript encodes a 28kD protein, while the larger one encodes a 65kD polypeptide which contains the smaller protein and a YTH domain-containing polypeptide (Zhang et al. 2008). In higher plants, the orthologs of CPSF30 are quite diverged from the algae and yeast (Chakrabarti and Hunt 2015). Also, the monocot and dicot CPSF30 isoforms belong to distinct clades that reflect the evolution of these different groups of plant species (Chakrabarti and Hunt 2015). Additionally, the smaller and larger CPSF30 isoforms fall into similar clades, indicating that the complex plant CPSF30 gene has a common ancestor (Hunt et al. 2012) .

Biochemical properties of the Arabidopsis CPSF30

The *Arabidopsis* ortholog of CPSF30 (termed AtCPSF30) contains three predicted CCCH-type zinc finger motifs (Delaney et al. 2006). The first CCCH motif is responsible for bulk of RNA-binding activity. AtCPSF30 can bind with calmodulin, and the RNA-binding activity of AtCPSF30 is inhibited by calmodulin in a calcium-dependent manner (Delaney et al. 2006). The third motif is associated with endonuclease activity (Addepalli and Hunt 2007b). Previous studies showed that the endonuclease activity of AtCPSF30 can be inhibited by disulfide bond reducing agents (Addepalli and Hunt 2008). The reason is the existence of a disulfide bond that is required for the nuclease activity (Addepalli et al. 2010). There is an *Arabidopsis* mutant that can tolerate oxidative stress, called *oxt6*. This mutant has a T-DNA insertion in the gene that encodes AtCPSF30 (Chakrabarti and Hunt 2015). A set of genes which encode proteins containing thioredoxin- or glutaredoxin-related domains and proteins containing protein disulfide isomerase-like features are regulated by this protein (Zhang et al. 2008), consistent with the oxidative stress tolerance phenotype.

Protein-protein interactions and the localization of AtCPSF30

AtCPSF30 interacts with numerous other proteins in the polyadenylation complex (Hunt et al. 2008). One study found that AtCPSF30 can interact with itself (Delaney et al. 2006), but the function of this interaction in the polyadenylation process is not clear. Previous studies also showed that AtCPSF30 interacts with other polyadenylation factors, including CPSF100, CPSF73, CstF77 and Fip1 proteins (Delaney et al. 2006; Xu et al. 2006; Addepalli and Hunt 2007b; Chakrabarti and Hunt 2015). These studies show that AtCPSF30 is a central hub in the protein-protein interaction network of plant polyadenylation complex subunits (Hunt et al. 2008; Chakrabarti and Hunt 2015). In

transient assays, AtCPSF30 was found to localize in the cytoplasm when expressed by itself, but when co-expressed with CPSF73 or CPSF160, it localized to the nucleus. This indicates that interactions with other polyadenylation factors are necessary for nuclear localization of AtCPSF30 (Rao et al. 2009).

AtCPSF30 functions in cellular signaling and plant development

The AtCPSF30 is able to interact with calmodulin, and the calmodulin domain overlaps the first zinc finger domain of this protein (Delaney et al. 2006). Interestingly, this motif matches one of the two motifs involved in the interaction of CPSF30 with the NS1 influenza protein (NS1) in animals (Nemeroff et al. 1998). The RNA binding activity of the first zinc finger can be inhibited in a calcium-dependent manner, indicating that calmodulin might perform a regulatory role in RNA processing (Delaney et al. 2006). In plants, calcium is a secondary messenger in cellular signaling pathways that can perceive environmental and developmental conditions (Sarwat et al. 2013). Calcium binds with various calcium sensors such as calmodulin and can give rise to conformational changes, which further mediate the interactions with downstream targets (Sarwat et al. 2013). The binding of AtCPSF30 to calmodulin in a calcium-dependent manner may provide a possible link between polyadenylation and responses of environmental stress and developmental changes (Zhang et al. 2008; Liu et al. 2014).

The functional significance of calmodulin binding activity with AtCPSF30 was tested in diverse developmental processes and in response to various stimuli (Zhang et al. 2008; Liu et al. 2014). Several novel phenotypes in the *oxf6* mutant were reported (Zhang et al. 2008). These phenotypes include reduced lateral root formation, altered responses to plant growth regulators, lower fertility, and resistant to oxidative stress. The wild type growth,

development, and responses to different kinds of growth regulators and to oxidative stress can be restored by a transgene that encodes the wild type AtCPSF30 (the small protein encoded by the *OXT6* gene) (Zhang et al. 2008). This suggests that the large proteins encoded by the same gene are not necessary for the functioning of AtCPSF30 in these processes. In contrast, a mutant AtCPSF30 that cannot bind to calmodulin only partially restores wild type phenotypes to the *oxf6* mutant (Liu et al. 2014). This transgene was not able to restore wild-type lateral root development and responses to other plant hormone such as IAA, gibberellic acid and 6-BA. These results showed that AtCPSF30 has calmodulin binding-dependent or -independent roles that help to regulate different developmental and stress responses. Another study showed that AtCPSF30 plays critical role for the programmed cell death (PCD) and is required for resistance to *Pseudomonas syringe*, by regulating both basal resistance and R gene-mediated defense responses (Bruggeman et al. 2014).

AtCPSF30 mediates APA in a large set of genes

The phenotypes of the *oxf6* mutant suggest that AtCPSF30-mediated alternative polyadenylation is essential for the expression of genes related to these phenotypes. Especially, the cellular signaling system that might be involved in either the calmodulin or disulfide bond remodeling could inhibit AtCPSF30 (Delaney et al. 2006; Addepalli and Hunt 2008). This raises the possibility of alteration of poly(A) sites by cellular signaling. When genome-wide poly(A) site choice was compared between the *oxf6* mutant and wild type, more than half of all expressed genes had different poly(A) site profiles in the *oxf6* mutant compared with wild-type *Arabidopsis*. Three classes of poly(A) site were seen: wild type specific (wt-specific), *oxf6* mutant specific (*oxf6*-specific), and common (seen both in

wt-specific and *oxl6*-specific). The wt-specific and *oxl6*-specific sites were around 70% of all the sites seen in the wt and mutant. In contrast, only around 30% were common sites (Thomas et al. 2012). These sites fell into various genomic regions, such as 5'UTRs, coding regions, introns and 3'UTRs. Interestingly, the genes associated with stress responses were over-represented in the set of genes with APA sites within 5'UTRs and protein coding regions that were dependent on AtCPSF30. This suggests a role for AtCPSF30 mediated APA in the regulation of gene expression associated with stress responses (Chakrabarti and Hunt 2015).

In animals, the differential usage of poly(A) sites can give rise to shorter or longer mRNA isoforms (Ran et al. 2013; Gruber et al. 2014). In fast growing undifferentiated cells, genes subjected to alternative polyadenylation use proximal poly(A) sites and thereby produce shorter mRNA isoforms. This leads to higher levels of expression by omitting microRNA targets or recognition sites for other regulatory proteins. Usage of distal poly(A) sites can generate longer isoforms that may be subject to modes of negative regulation. A similar pattern of proximal/ distal poly(A) site usage is not seen in the case of the *oxl6* mutant (Thomas et al. 2012). However, poly(A) sites seen only in the *oxl6* mutant lacked one of the three *cis* elements (the Near Upstream Element or NUE) that together constitute a canonical plant polyadenylation signal (Loke et al. 2005). Proximal sites in animal genes have sub-optimal polyadenylation signals. Therefore, while the organization of alternative sites in these two systems might be different, it seems that the association with suboptimal polyadenylation signals with proximal poly(A) sites may be a common mechanism of alternative polyadenylation (Ran et al. 2013).

The genome-wide studies of poly(A) site choice in the *oxt6* mutant showed that AtCPSF30-mediated alternative polyadenylation has the potential to impact a large set of genes in *Arabidopsis*. This result coincides with the numerous phenotypes that are displayed in the *oxt6* mutant (Zhang et al. 2008). As mentioned above, the cellular signaling systems that cause the changes in calmodulin binding activity and/or disulfide linkages should inhibit AtCPSF30 (Delaney et al. 2006; Addepalli and Hunt 2008). This suggests that cellular signaling pathways have the possibility to change the regulation of a large set of genes at the posttranscriptional level by altering the usage of AtCPSF30-dependent sites (Chakrabarti and Hunt 2015).

1.2 Alternative polyadenylation in Eukaryotes

1.2.1 The definition and types of alternative polyadenylation

A large portion of eukaryotic genes contain more than one potential poly(A) site (as defined by the FUE, NUE, and poly(A) signal), raising the possibility of alternative polyadenylation (APA) (Di Giammartino et al. 2011; Lutz and Moreira 2011; Shi 2012). More than 50% of human genes show alternative polyadenylation was based on genome level analysis (Tian et al. 2005; Yan and Marr 2005), while more than 70% of all genes were affected by APA in *Arabidopsis* (Meyers et al. 2004; Wu et al. 2011). Four types of alternative polyadenylation have been described in higher eukaryotes. In type I alternative polyadenylation, there is more than one polyadenylation signal in the 3'-UTR in the same terminal exon. This results in more than one mRNA isoform each of which encodes an identical protein. APA in this case may alter mRNA stability or translatability, or may have other downstream effects (Di Giammartino et al. 2011; Chakrabarti and Hunt 2015). Type II alternative polyadenylation involves poly(A) sites that are present in upstream introns.

This type APA can invoke alternative splicing along with alternative polyadenylation. The truncated protein may or may not be produced, due to the stability of the mRNA, the presence of an in frame stop codon, and the translational competence of the mRNA. In addition, the products of APA in these cases may or may not trigger non-sense mediated decay to degrade these transcripts (Lutz and Moreira 2011; Shi 2012). In type III alternative polyadenylation, the poly(A) sites are present in the upstream exons. With no premature termination codon and no in-frame stop codon, these mRNA isoforms will be degraded through non-stop decay pathways (Daniel and Lynne 2012; Klauer and van Hoof 2012). The last type of alternative polyadenylation has poly(A) sites that are present in the 5'-UTR. These products might- or might- not be stable.

1.2.2 Widespread occurrence of alternative polyadenylation

In human, mouse, worm and plant, recent analysis of genome-wide data showed that alternative polyadenylation is pervasive. In many instances, alternative polyadenylation plays some important roles in control of global physiological events such as cell proliferation, differentiation, transformation and developmental programs (Sandberg et al. 2008; Ji et al. 2009; Mayr and Bartel 2009).

1.2.2.1 Widespread occurrence of alternative polyadenylation in animal

In early studies of expressed sequence tags and more recent next generation sequence analyses, alternative polyadenylation in animals was seen to be common and frequently involves 3'UTRs. At least 70% of mammalian mRNA-encoding genes produce alternative polyadenylation isoforms (Derti et al. 2012; Mainul et al. 2012). In animals, alternative polyadenylation may change the expression of mRNA and protein isoforms when poly(A) sites fall into the intronic- or coding- regions of a gene (Derti et al. 2012). However, based

on individual gene studies, most instances of alternative polyadenylation involve multiple sites in 3'UTRs (Gautheret et al. 1998). It has been demonstrated that different 3'UTR can effect mRNA stability or translatability (Mainul et al. 2012). Apparently, the longer 3'UTRs have more RNA cis regulatory elements and AU-rich content than short 3'UTRs. In addition, 3'UTRs are the most common location for microRNA target sites (Tian et al. 2005; Di Giammartino et al. 2011; Sun et al. 2012). Several studies have shown that APA affects different biological processes including immune responses, neuron activity, animal development, tumorigenesis and metastasis.

APA in immune response

In the immune response in mammals, hundreds of genes were found to switch their poly(A) sites during T cell activation (Lutz and Moreira 2011). Most of the events favored the production of shorter 3'UTR isoforms in activated T lymphocytes. This kind of event also occurred after anti-CD40 and interleukin-4 stimulation of B cells, and also upon lipopolysaccharide- and interferon- γ - stimulation of human monocytes. These studies indicate that there is a strong correlation between 3'UTR shortening and cell proliferation. The output of protein is impacted by the differential usage of 3'UTR isoforms, typically with the shorter one producing more protein through avoidance of microRNA-mediated translational repression (Ji et al. 2009), or of degradation by AU rich elements or other regulatory elements (Legendre et al. 2006).

APA in neuronal cells

In neurons, APA regulates the expression of the brain-derived neurotrophic factor (BDNF) gene, resulting in two mRNA isoforms with different 3'UTR lengths (Timmusk et al. 1993). The long and short mRNA isoforms have different localizations: the long isoforms

are found in dendrites, while the short isoforms are localized in somata. The short mRNA isoforms lead to smaller spine head diameters and more spines in the apical dendrites (Sandberg et al. 2008). In the central nervous system, a series of transcription factors are activated by environmental stimuli, such as myocyte enhancer factor 2 (MEF2) that mediates the expression of hundreds of genes (Barreau et al. 2005). For many of these neuron activity-dependent genes, extracellular stimulation can induce APA to produce truncated transcripts (An et al. 2008). The products with short transcripts increase or accelerate protein synthesis (An et al. 2008; Flavell and Greenberg 2008; Shi 2012). In addition, the regulation of transcription activity by APA seems to be widespread in neuronal responses (Flavell et al. 2008).

APA in cell differentiation and development

In contrast to 3'UTR shortening by extracellular stimulation, a global lengthening of 3'UTRs is seen during mouse early embryogenesis (Ji et al. 2009). To confirm *in silico* results, the 3'UTR lengthening process is captured in a mouse muscle differentiation where C2C12 myoblast cells differentiate into myotubes. Also, a weaker polyadenylation efficiency at proximal sites was observed in the differentiation condition from reporter assays, causing 3'UTR lengthening during myoblast differentiation. As well, 3'UTR lengthening was found during *Drosophila* development (Hilgers et al. 2011), suggesting that APA regulation is evolutionarily conserved in animals.

APA in cancer cells

Considering that alternative polyadenylation is important for cellular proliferation and differentiation, it is not surprising that alternative polyadenylation and miRNA-mediated repression are correlated in cancer cell lines (Sun et al. 2012). It has been reported that

oncogenes can be activated by the loss of miRNA target sites, such as high mobility group AT-hook 2 (*HMGA2*) (Hilgers et al. 2011; Sun et al. 2012). Also, the lengths of *Cyclin D1(CCND1)* 3'UTR in mantle cell lymphomas was shortened due to a mutation that produces a premature poly(A) signals, giving rise to an increase in mRNA stability (Wiestner et al. 2007). Another study showed that most of the 23 genes containing more than one poly(A) signal in the 3'UTR expressed shorten isoforms in cancer cell lines, caused by the utilization of the most proximal poly(A) signal (Mayr and Bartel 2009). Furthermore, when *IMP-1* 3'UTRs were fused with a luciferase reporter gene, the result showed that a shortening of UTR and the loss of miRNA target sites led to an increase in protein production (Lee and Dutta 2007). These results indicate that one of the mechanisms of oncogenic transformation is the loss of miRNA sites in the mRNAs of oncogenes. These genes can avoid the miRNA-mediated repression through 3'UTR shortening caused by the alternative polyadenylation (Mayr et al. 2007).

1.2.2.2 Widespread occurrence of alternative polyadenylation in plant

Several studies have been demonstrated that, as in animals, alternative polyadenylation is pervasive in plants. Multiple poly(A) sites were identified in genes that encode phosphoenolpyruvate carboxylase and alpha1-tubulin in maize (*Zea mays*) (Yanagisawa et al. 1988; Montoliu et al. 1990), AGAMOUS and rbohA in *Arabidopsis* (Keller et al. 1998; Cheng et al. 2003), chloroplast ascorbate peroxidase in spinach (*Spinacia oleracea*) and tobacco (*Nicotiana*) (Yoshimura et al. 2002), several soybean (*Glycine max*) genes (Xing et al. 2010), and a number of *Medicago* and rice (*Oryza sativa*) genes (Wu et al. 2014; Fu et al. 2016). Also, alternative polyadenylation plays an essential role in a range of biological processes in plants. The S locus genes are important for self-incompatibility in

Brassica. A putative protein was produced from alternative polyadenylation of the pre-mRNA of the S-locus receptor kinase gene within the third intron, while the removal this intron resulted a transcript encoding a predicted membrane-anchored protein (Tantikanjana et al. 1993; Giranton et al. 1995). In cotton (*Gossypium*), the gene that encodes lysine-ketoglutarate reductase (LKR) can produce two transcripts via alternative polyadenylation. These encode monofunctional and bifunctional LKR polypeptides, respectively. These polypeptides might be pivotal for enabling efficient flux of lysine catabolism under some specific conditions (Tang et al. 2002). The APA of LKR pre-mRNA also was conserved in *Arabidopsis* (Tang et al. 2000).

Alternative polyadenylation and flowering time control

Seasonal flowering control has a complicated but unique gene-regulation mechanism that involves alternative polyadenylation (Horniyk et al. 2010; Liu et al. 2010). Flowering time is negatively regulated by expression of the flowering locus C (*FLC*) gene. In *Arabidopsis*, flowering time control is also regulated by *FPA* and *FCA*, two RNA binding proteins that act independently to repress *FLC* expression and thereby promote flowering (Macknight et al. 2002; Horniyk et al. 2010). Both *FPA* and *FCA* have been reported to repress *FLC* expression by altering APA of non-coding antisense transcripts (Sonmez et al. 2011). On the opposite strand, one promotor located downstream of the poly(A) site of *FLC* produces anti-sense transcripts that have alternative poly(A) sites: one cluster of poly(A) sites (proximal) is situated opposite the terminal intron of *FLC* and another cluster (distal) is situated opposite the *FLC* promoter. The usage of proximal poly(A) sites is promoted by both *FPA* and *FCA* (Horniyk et al. 2010; Liu et al. 2010). Also, mutations in the CstF components, CstF64 and CstF77, showed elevation of sense *FLC* transcripts and reduction

of antisense *FLC* transcripts, indicating *FLC* antisense transcripts are sensitive to CstF activity (Liu et al. 2010). flowering locus D (*FLD*), a histone H3 Lys 4 (H3K4me2) demethylase, is also needed for effective *FLC* silencing (Liu et al. 2007; Bäurle and Dean 2008; Liu et al. 2010). In addition, FY (the homolog of the 3' processing factor WDR33; (Shi et al. 2009)) can interact with FCA to promote proximal poly(A) site selection in the FCA pre-mRNA, resulting in a truncated, non-functional *FCA* transcript (Simpson et al. 2003). As suggested by Rosonina and Manley (2010), when the proximal poly(A) sites are used in the anti-sense transcript, the transcriptionally active chromatin mark H3K4me2 in the body of the *FLC* gene was removed by the recruited *FLD* demethylase, leading to *FLC* silencing. Vice versa, when the distal poly(A) sites is used in the anti-sense transcript, the positive factors are recruited to the *FLC* promoter, giving rise to enhanced *FLC* mRNA expression (Rosonina and Manley 2010).

Alternative polyadenylation and oxidative stress response

Another well studied example in plants concerns connections between APA and oxidative stress responses. Thus, an *Arabidopsis* mutant deficient in *AtCPSF30* expression was found to be more tolerant than wild type to oxidative stress (Zhang et al. 2008). As stated above, *AtCPSF30* is inhibited by calmodulin and sulfhydryl reagents in vitro (Delaney et al. 2006; Addepalli and Hunt 2008) and thus is connected to calcium and redox signaling pathways. In addition, a genome-wide study of poly(A) site choice in the *AtCPSF30* mutant (*oxt6*) showed extensive poly(A) site choice changes between wild type and *oxt6*. Specifically, three classes of poly(A) site could be discerned: wild type specific (wt-specific) and *oxt6* mutant specific (*oxt6*-specific), and common sites (seen both in wt and *oxt6* mutant plants).

The wild wt-specific and *oxf6*-specific sites make up around 70% of all sites seen in plants (Thomas et al. 2012).

1.2.3 Biological significance of alternative polyadenylation under stress

APA was described as early as 1980, with studies identifying APA giving rise to the production of both membrane-bound and secreted IgM (Alt et al. 1980). Originally, mRNA polyadenylation events were thought of as separate and distinct, but more recent studies show that all RNA processing events are interconnected and intertwined (Tom and Robin 2002; Moore and Proudfoot 2009; Donny and Robert 2010). Interactions between polyadenylation and transcription, splicing, transcription termination, translation, export and stability are all included in these events (Proudfoot et al. 2002; Kathleen et al. 2009; Donny and Robert 2010). The RNA factors and proteins involved in many of these interconnected processes include those that mediate alternative splicing (AS) and APA (Danckwardt et al. 2007; Hall-Pogar et al. 2007; Melton et al. 2007; Newnham et al. 2010). Recently, studies using transcriptome-wide techniques showed that the transcriptomes are highly complicated and alternative processing of pre-mRNAs at post-transcriptional level significantly contributes to enlarge transcriptome diversity (1999; Wang and Brendel 2006; Shen et al. 2011; Ran et al. 2013; Gupta et al. 2014).

As mentioned above, APA is widespread in animals and plants. Recently, one study showed that APA contributes to the post-transcriptional response to stresses and affects many mRNAs and ncRNAs in mammals (Alexander et al. 2012). The length of the 3'-UTR affects the potential inclusion or exclusion of *cis*-acting RNA sequence elements under stress, such as microRNA binding sites or binding sites for regulatory RNA-binding proteins (Hollerer et al. 2014). Stress-regulated poly(A) sites were identified in intronic

and coding regions in plants (Shen et al. 2011; Fracasso et al. 2016). In addition, stress-regulated poly(A) sites occurred in intergenic regions in animals, possibly corresponding to 3' elongated mRNA isoforms (Hollerer et al. 2016). Also there is a global tendency towards increased utilization of distal poly(A) sites under stress in animals (Hollerer et al. 2014). This trend also was detected in yeast suffering DNA damage (Graber et al. 2013). These results suggest that a general stress response mechanism of poly(A) site usage is the stress-dependent inhibited usage of proximal and the increased usage of distal poly(A) sites (Hollerer et al. 2016). Interestingly, for more than 70% of genes subject to stress-induced APA, overall expression levels remain unchanged, suggesting that APA can impact the gene expression without altering overall mRNA abundance (Hollerer et al. 2016).

In plants, while there have been studies of stress-associated alternative splicing (AS) (Lopato et al. 1999; Lazar and Goodman 2000; Wang and Brendel 2006; Palusa et al. 2007; Reddy 2007; Reddy et al. 2013; Hollerer et al. 2016), stress-induced APA has not been much studied. While APA is widespread in plants, (Kalyna et al. 2003; Iida et al. 2004; Kalyna et al. 2006; Filichkin et al. 2010; Wu et al. 2011; Xing and Li 2011b), there have been no indications of stress-associated responses. Therefore, stress-induced APA in plants needs further study.

1.3 RNA stability in eukaryotes

1.3.1 Nuclear mRNA degradation pathway

Pre-mRNA synthesis by RNA pol II is coupled with 5' end capping. Also, pre-mRNAs are co-transcriptionally spliced to remove introns. Finally, pre-mRNAs are processed and polyadenylated. From the beginning of mRNA synthesis, different proteins associate with the transcript to form messenger ribonucleoproteins (mRNPs). In nuclear mRNA

processing and in the formation of an export competent mRNP, defects can cause the degradation of nuclear mRNA by surveillance pathways. Defective mRNPs may also be retained in the nucleus to allow time to complete the mRNA processing or to degrade the incorrect transcripts (Jensen et al. 2001b; Patricia et al. 2001).

In nuclear RNA retention and surveillance, an important player is the nuclear exosome that contains several subunits including the evolutionary conserved 3'-5' exonuclease Rrp6. The activity of the nuclear exosome is facilitated by the Trf4/Air2/Mtr4p Polyadenylation (TRAMP) complex that contains the unconventional poly(A) polymerase Trf4p (Lacava et al. 2005). Defective mRNAs are retained at the transcription sites and polyadenylated (Jensen et al. 2001b; Patricia et al. 2001; Thomsen et al. 2003); Rrp6 is required for this nuclear retention (Jensen et al. 2001a; Libri et al. 2002). Mutations in nuclear export factors can lead to premature transcriptional termination and polyadenylation defects, which can induce mRNA retention and degradation by the exosome (Babour et al. 2012). Unspliced pre-mRNAs can be exonucleolytically degraded in a 3' to 5' direction by the nuclear exosome, or in a 5' to 3' direction by Rat1 (Bousquet-Antonelli et al. 2000). Unspliced pre-mRNAs also can be exported to the cytoplasm where they can trigger the cytoplasmic quality control pathway (Sayani et al. 2008). In yeast and humans, the regulation of the levels of a small subset of genes involves the nuclear exosome and the RNA polyadenylation activity of TRAMP (Arigo et al. 2006; West et al. 2006; Schmid and Jensen 2008). Additionally, it has been suggested that all mRNAs are subjected to a certain rate of nuclear degradation, which is controlled by the degree of the nuclear retention of each mRNA (Kuai et al. 2005).

1.3.2 Cytoplasm mRNA degradation pathway

The degradation of mature mRNAs mainly happens in the cytoplasm. There are different degradation pathways for cytoplasmic decay of eukaryotic mRNAs (Balagopal et al. 2012; Parker 2012; Wu and Brewer 2012; Schönemann et al. 2014). Usually, mRNA decay starts at the 3'-end poly(A) tail with deadenylation. Depending on the mRNP, the poly(A) tail can be shortened by one of three different deadenylation complexes: PAN2 and 3, the Ccr4-NOT complex, or Poly(A)-Specific Ribonuclease (PARN) (Aaron and Marvin 2008). For each mRNA, the shortening rate is specific and is an important process that defines the mRNA half life (Cao and Parker 2001). Following deadenylation, the cytoplasmic exosome can exonucleolytically degrade the mRNA in a 3' to 5' direction (Anderson and Parker 1998). Usually, the shortening of the poly(A) tail is followed by the 5'-cap removal by the decapping complex Dcp2/Dcp1, and subsequent degradation in a 5'-3' direction by the exoribonuclease Xrn1 (Hsu and Stevens 1993; Muhrad et al. 1994). For some mRNAs, decay can start with internal endonucleolytic cleavage, and subsequently degradation by the exosome and Xrn1 (Bracken et al. 2011). The deadenylation of mRNAs is usually a rate-limiting step, and the length of the poly(A) tail plays a critical role in gene expression by regulating mRNA decay and translation (Decker and Parker 1993). The association of the poly(A) tail with the poly(A) binding protein Pab1 (PABP in mammals) is essential for poly(A) tail shortening as well as for mRNA translation (Mangus et al. 2003).

mRNA decapping was considered to be an irreversible step in mRNA degradation. But a recent study has described that previously cleaved RNAs can be re-capped in mammalian cells (Schoenberg and Maquat 2009; Mukherjee et al. 2012). Therefore, this raises the possibility that mRNA decay may be blocked by re-capping, even if part of the mRNA is degraded by Xrn1. Usually, decapping occurs after mRNA poly(A) tail shortening (Decker

and Parker 1993), and the poly(A) tail and poly(A)-binding protein Pab1 serve as the negative regulators of decapping (Tharun 2009). For some transcripts, deadenylation-independent decapping has been reported (Badis et al. 2004; Muhrad and Parker 2005). The cytoplasmic cap-binding complex eIF4F at the 5'-cap is replaced by the decapping complex Dcp2/Dcp1 for deadenylation-independent decapping. Therefore, the process of decapping competes with translation initiation (Li and Kiledjian 2010; Parker 2012). Decapping plays an important role for gene expression control, which is suggested by the large set of factors that modulate decapping activity (Nissan et al. 2010). To stimulate the formation of a decapping complex or by enhancing decapping activity, many factors positively affect decapping, such as Edc1, Edc2, Edc3 and the Lsm1-7 complex. Translation initiation can be directly inhibited by some of the factors that promote decapping, such as Scd6 (Rajyaguru et al. 2012) and Stm1 (Balagopal and Parker 2011). The DEAD-box helicase Dhh1, which is another decapping enhancer, seems to inhibit translation after initiation, thereby leading to the accumulation of ribosomes on the transcripts (Sweet et al. 2012). Similarly, Pat1 can directly enhance decapping by interacting with the Lsm1-7 complex and Dcp2, and thereby repress translation (Pilkington and Parker 2008; Nissan et al. 2010). The Ccr4-NOT deadenylation complex and the Dcp1-Dcp2 decapping complex associate with Pat1 in humans, thus providing a possible link for these two processes (Marnef and Standart 2010; Ozgur et al. 2010). All these results show that Pat1 might be a key player in silencing gene expression in eukaryotes by acting as a scaffold protein for the sequential binding of translational repression and decay factors onto mRNPs (Marnef and Standart 2010).

After decapping, the 5'-3' exonuclease Xrn1 acts in mRNA degradation. Xrn1 is the critical enzyme responsible for the turnover of translatable mRNAs, and is also involved in many mRNA quality control pathways including the degradation of mRNAs targeted by siRNAs and miRNAs (Jones et al. 2012). The degradation of the mRNA from the 3' end is catalyzed by the exosome. The cytoplasmic exosome is a multiprotein complex with 3'-5' exonuclease activity. The exosome consists of nine core subunits and other related subunits, some of which have nucleolytic activity, others of which regulate the catalytic activity and substrate specificity of the complex. The exosome can differ between different species (Lykke-Andersen et al. 2011).

1.3.3 RNA quality control pathways in Eukaryotes

In eukaryotes, there are at least three kinds of defects in mRNAs that affect the translating ribosome and activate co-translational quality control pathways. These are discussed in the following subsections.

Nonsense-mediated decay pathway

The first type to be discovered and best-known RNA quality control pathway is nonsense mediated mRNA decay (NMD). mRNAs that contain premature termination (stop) codons (PTCs) can lead to premature translation termination and trigger nonsense mediated mRNA decay (NMD) pathway (Figure 1.2) (Losson and Lacroute 1979; Maquat et al. 1981). Multiple mechanisms can give rise to PTCs in mRNAs. The most common one is thought to be cryptic or alternative splicing. These splicing events can lead to PTCs by either causing frame shifts within the coding region, or exposing stop codons within retained introns or alternative exons (Lareau et al. 2007; McGlincy and Smith 2008; Sayani et al. 2008). Also, transcription errors, genetic mutations, or recombination events can cause

PTCs (Chang et al. 2007; Isken and Maquat 2007; Rebbapragada and Lykke-Andersen 2009; Schweingruber et al. 2013). In addition, it seems like a subset of normal endogenous mRNA is targeted by the NMD pathway (Joshua et al. 2004; Guan et al. 2006; Johansson et al. 2007). When translation is terminated by a ribosome at a PTC, several NMD factors associate with the PTC-containing mRNA and target it for degradation by RNA decay enzymes. These NMD factors included the Upf and Smg proteins (Leeds et al. 1991; Leeds et al. 1992; Pulak and Anderson 1993). Depending on the specific mRNA and organism, NMD degradation is initiated by de-adenylation, de-capping, or endonucleolytic cleavage (Muhrad et al. 1994; Chen and Shyu 2003; Mitchell and Tollervey 2003; David and Elisa 2004). Subsequently, the degradation of mRNA body involves disassembly of the initial mRNA-protein complex, which is dependent on the ATPase activity of the central NMD factor Upf1 (Franks et al. 2010).

The specific functions of Upf/Smg proteins that distinguish premature from normal translation termination remain under investigation. Translation termination mediated by the eRF1–eRF3 termination complex at a PTC is thought to be not efficient, owing to the absence of a normal 3'-UTR (Amrani et al. 2004). The cytoplasmic poly(A)-binding protein (PABPC) that is associated with 3'-UTRs promotes efficient translation termination on normal mRNAs (Amrani et al. 2004). The translation termination event inhibits NMD when the cytoplasmic poly(A)-binding protein is in close proximity to the termination complex, whereas termination allows assembly of NMD factors when PABP is removed from the termination complex (Amrani et al. 2004; Behm-Ansmant et al. 2007; Eberle et al. 2008; Ivanov et al. 2008; Singh et al. 2008). In metazoans, NMD may also be triggered when termination occurs upstream of exon-exon junctions. These junctions are

recognized by the exon-junction complex that is deposited during pre-mRNA splicing in the nucleus and that interacts with Upf proteins in the cytoplasm (Kim et al. 2001; Le Hir et al. 2001; Lykke-Andersen and Steitz 2001; Gehring et al. 2003; Singh et al. 2007). Additional mRNP components serve to distinguish normal mRNAs from those targeted for NMD (González et al. 2000; Singh et al. 2007; Meaux et al. 2008). Moreover, the NMD pathway actively inhibits the recruitment of new ribosomes to the target mRNA (Muhlrad and Parker 1999; Isken et al. 2008). In addition, the native polypeptide that is produced from a PTC-containing mRNA is subjected to proteolysis in a way stimulated by Upf1, a core NMD factor (Kuroha et al. 2009; Verma et al. 2013).

Non-stop decay pathway

Another type of mRNA defect is one that results in the absence of a termination codon. Once the translating ribosome reaches the mRNA 3' end without encountering a termination codon, the mRNAs are subjected to the nonstop decay pathway (NSD) (Figure 1.3) (Frischmeyer et al. 2002; van Hoof et al. 2002). When polyadenylation occurs prematurely within the protein coding region, the mRNAs are likely to be non-stop mRNAs (Frischmeyer et al. 2002; van Hoof et al. 2002). Truncated mRNAs resulting from endonucleolytic cleavage events within the protein coding region are another possible source of NSD substrates (Tsuboi et al. 2012; Matsuda et al. 2014). However, the degradation of mRNAs endonucleolytically cleaved by a hammerhead ribozyme does not require the NSD machinery for rapid degradation, indicating that the general exonucleolytic mRNA decay pathway may be more predominant than the NSD pathway on such substrates (Meaux and Van Hoof 2006).

NSD has been intensely studied in *Saccharomyces cerevisiae*, where results have shown an important role for the GTPase Ski7, which is a homologue of the translation termination factor eRF3 (van Hoof et al. 2002). Ski7 also has an N-terminal domain that is important for the catalytic activity of the cytoplasmic form of the *S. cerevisiae* 3' to 5' exonuclease exosome complex (van Hoof et al. 2000; van Hoof et al. 2002). Deletion of any one of these Ski7 domains can stabilize mRNAs targeted for NSD and lead to the inactivation of exosome components (van Hoof et al. 2002; Schaeffer and van Hoof 2011). Based on these results, Ski7 might enter the empty A-site of ribosomes which are stalled at the 3' end of nonstop mRNAs and recruit the exosome to trigger mRNA degradation (van Hoof et al. 2002). Recent studies indicate a potential role for another eRF3-homologous factor Hbs1, which associates with eRF1-related factor Dom34 in *S. cerevisiae* NSD. Deletion of these factors and of 5' to 3' or 3' to 5' exonucleases can cause increased stability of NSD substrates and accumulation of degradation intermediates (Tsuboi et al. 2012). In addition, the Hbs1-Dom34 complex, along with the ribosome cycling ATPase Rli1 (that is called ABCE1 in humans), promotes disassembly of ribosomes when stalled at/near the 3' end of mRNA (Pisareva et al. 2011; Shoemaker and Green 2011). These observations suggest that the active release of the ribosome by Hbs1-Dom34 and Rli1 might be necessary for efficient degradation of an mRNA targeted for NSD (Lykke-Andersen and Bennett 2014). Besides *S. cerevisiae*, NSD has also been studied in human (*Homo sapiens*) tissue culture cells (Frischmeyer et al. 2002; Saito et al. 2013). Some studies showed that the absence of a stop codon has no effect on the decay rate of specific mRNAs in human cells (Akimitsu et al. 2007; Torres-Torronteras et al. 2011). One possible reason is that relative contributions of translation repression and mRNA decay are substrate dependent. In

addition, other studies indicate that siRNA-mediated depletion of Hbs1 or Pelota (the human orthologue of Dom34) can lead to impaired NSD in human tissue culture cells (Saito et al. 2013).

No-go decay pathway

No-go decay (NGD) is the third type of mRNA surveillance mechanism discovered in yeast. NGD can be triggered when translation elongation is inhibited, leading to endonucleolytic cleavage and rapid mRNA degradation (Meenakshi and Roy 2006). RNA structure elements (Meenakshi and Roy 2006), rare codons and polylysine (Gandhi et al. 2008), and mRNA depurination by a viral enzyme (Tsuboi et al. 2012) are reported to activate the NGD pathway. Near the ribosome stall site, NGD is initiated by endonucleolytic cleavage, which can physically impede the translational machinery from moving down the transcript (Meenakshi and Roy 2006). During this process, Hbs1 and Dom34 play important roles by binding the A site of stalled ribosomes and promoting the recycling of ribosome subunits (Kobayashi et al. 2010). One previous study showed that Hbs1 and Dom34 can stimulate an initial endonucleolytic cleavage event (Meenakshi and Roy 2006; Passos et al. 2009; Tsuboi et al. 2012), while another study demonstrated that NGD can be triggered by the stalled ribosome event in the absence of Hbs1 and Dom34 (Passos et al. 2009; Tsuboi et al. 2012). The protein complex Hbs1-Dom34 is related to eRF1 and eRF3, respectively. Therefore, Hbs1-Dom34 can directly interact with the stalled ribosome. In addition, Hbs1 is related to NSD factor Ski7, which mediates the release of stalled ribosomes at the end of transcripts without a stop codon. Perhaps Hbs1 performs the similar function in NGD as Ski7 (Chang et al. 2007; Graille et al. 2008). Besides *S. cerevisiae*, the NGD pathway was detected in *Drosophila* S2 cell (Passos et al. 2009). In

addition, the ribosome stalling has been associated with transcript decay in *Arabidopsis* (Onouchi et al. 2005), which suggests that it might be a conserved process in plants.

1.3.4 The significance of mRNA stability

The importance of mRNA stability is demonstrated by the phenotypes of mutants where the genes encoding components of systems involved in mRNA stability have been disrupted. For example, mice deficient for the tristetraprolin, an RNA-binding protein that controls the stability of mRNAs such as granulocyte/macrophage colony-stimulating factor and tumour necrosis factor, have a systemic inflammatory syndrome with autoimmunity and myeloid hyperplasia (bone marrow overgrowth) (Carballo and Blackshear 2001). In *Caenorhabditis elegans*, adults carrying mutations in dicer (*dcr-1*), a double-stranded endoribonuclease which plays an important role in the first steps of the RNA interference pathway, are sterile and have defects in developmental timing (Grishok et al. 2001; Knight and Bass 2001). *Arabidopsis* mutants with mutations in the gene that codes for DST1, that regulates the stability of RNAs such as CCL and SEN1, have defects in circadian rhythms (Gutierrez et al. 2002). Moreover, the D-subunit of photosystem I (PSI-D) is encoded by two functional genes, PsaD1 and PsaD2 in *Arabidopsis*. The single-gene mutation of the *psad1-1*, which markedly affects the accumulation of PsaD mRNA and protein, and photosynthetic electron flow (Ihnatowicz et al. 2004). In the same mutant, the levels of PSI and PSII polypeptides are decreased, and leaf color become light green and increased photosensitivity (Ihnatowicz et al. 2004). These results suggest that ribonucleases and associated factor can be specifically regulated and can target essential RNAs that play critical roles in these developmental and cellular processes.

The control of mRNA stability is intimately linked with mRNA translation. In *Drosophila* (*Drosophila melanogaster*), crucial protein gradients are often dependent on translational suppression of mRNAs encoding proteins such as NANOS and HUNCHBACK, followed by degradation (Macdonald and Smibert 1996). mRNAs being translated are known to form a ‘closed loop’ conformation, in which the 5’ cap is held in close proximity to the poly(A) tail through a series of connecting proteins, such as eIF4E (eukaryotic initiation factor 4E), eIF4G, and poly(A) binding protein (Sachs and Davis 1989; Hentze 1997; Mitchell and Tollervey 2000). This seems to provide a mechanism where mRNAs involved in translation are protected from exonucleolytic degradation. Proteins that repress translation appear to act by disrupting the RNP complex holding the 5’ and 3’ ends of the RNA together (Fátima and Matthias 2004). Subsequently, the degradation of these mRNAs might require interaction between the translation apparatus and mRNA-degradation machinery. Therefore, studying ribonucleases and ribonuclease-associated factors is not only important for understanding the control of mRNA degradation, but also for understanding mechanisms of translational repression. The links between these processes and mRNA stability need to be further study, which may also shed light on the mechanisms of degradation of specific mRNAs in eukaryotes (Newbury 2006).

1.4 Summary

In eukaryotes, the process of alternative polyadenylation plays an important role in the expression of genes. Previous studies showed that poly(A) site choices were changed in different genomic regions under stress conditions, in yeast (Yoon and Brem 2010), mammals (Sandberg et al. 2008; Mainul et al. 2012) and plants (Shen et al. 2011; Wu et al. 2011; Wu et al. 2014). Poly(A) site choice has been studied genome wide in *Arabidopsis*

subjected to drought stress. This study revealed that alternative polyadenylation might perform a mechanism to up-regulate gene expression under stress conditions in Chapter 2. A study shows that a larger number of poly (A) sites are controlled by CPSF30 in *Arabidopsis*. Three classes of poly (A) sites were noticed: wild type specific (wt-specific), *oxt6* mutant specific (*oxt6*-specific), and common (seen both in wt-specific and *oxt6*-specific) (Thomas et al. 2012). To test the mRNA isoforms from different genomic regions, the stability of different classes of mRNA isoforms was studied in Chapter 3. These mRNA isoforms had different stabilities between wild type and mutant *oxt6*. Especially the stability of mRNA isoforms that within the coding region, less stable in wt-specific while more stable in the *oxt6*-specific. What proteins might interact with AtCPSF30 to lead to the poly(A) site choice changes? Different protein-protein interaction techniques were performed to screen the candidate proteins. From localization assays, a previous study (Rao et al. 2009) and this study both found that AtCPSF30 is not only in nucleus, but also in the cytoplasm, suggesting that AtCPSF30 may interact with other proteins. The study displayed that AtCPSF30 interacts with At1g05510, OIL BODY-ASSOCIATED PROTEIN 1A that, in its tyrosine-phosphorylated state, is modulated in response to ABA in *Arabidopsis thaliana* seeds, and ribosome protein L35 (At5g02610) that belongs to Ribosomal L29 family protein.

When the aberrant RNAs are produced by a variety of events, the defective protein products are more likely to be misfolded and sometimes have dominant-negative effects (Roman et al. 1991; Ishigame et al. 2013). Thus, it is not surprising that organisms evolved a number of quality control processes to detect errors in mRNAs and subject them to rapid degradation (van Hoof and Wagner 2011). The defective mRNAs interfere with translation,

the corresponding mRNA- and protein-quality control processes all take advantage of the ribosome for aberration recognition (Shoemaker and Green 2011). The AtCPSF30 interacts with ribosome protein, and the stabilities of noncanonical mRNA isoforms are mediated by AtCPSF30. This provides a possibility that AtCPSF30 may be involved in mRNA quality control pathways.

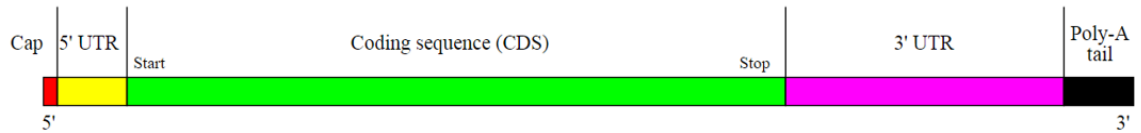


Figure 1.1 Mature mRNA in eukaryotes

A typical structure of mature mRNA in Eukaryote (Plants and Animals). It contains a CAP (A 7-methylguanosine attached from the 5' carbon by a triphosphate bridge to the 5' carbon of the terminal nucleotide of the mRNA), 5' UTR (5' end untranslated region), CDS (coding region), 3' UTR (3' untranslated region) and a poly(A) tail that interacts with the poly(A) binding proteins (PABPs).

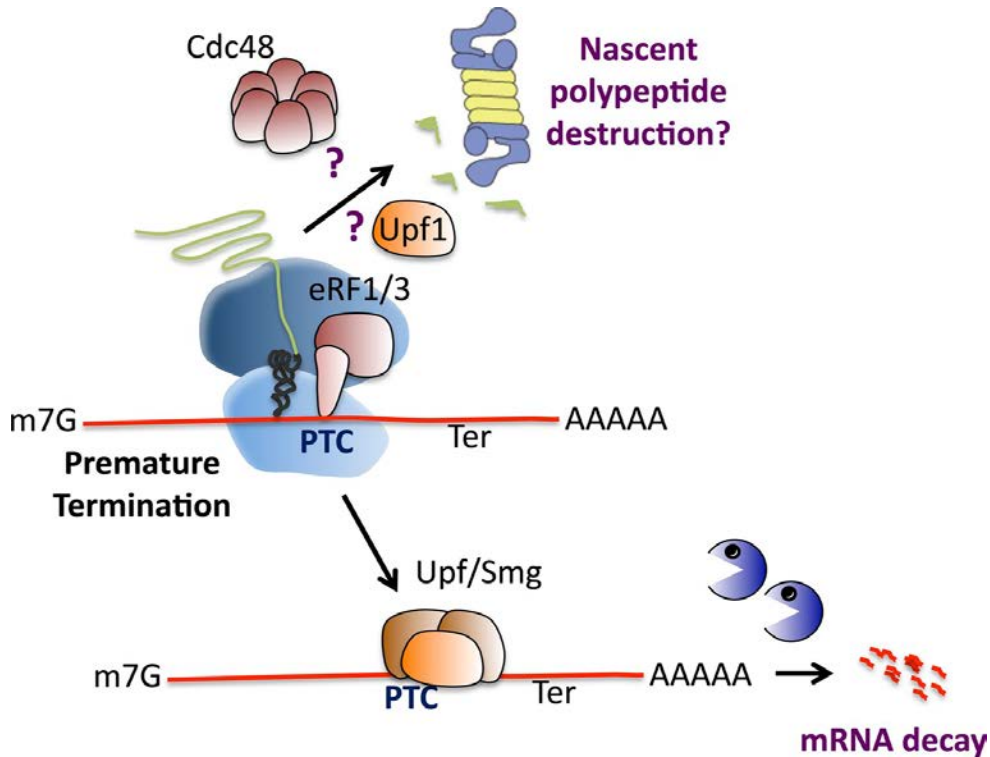


Figure 1.2 Non-sense mediated decay pathway (cited from Lykke-Andersen et al, 2014)

The mRNAs that contain a premature termination codon (PTC) are degraded when a stall in the eRF1/eRF3-dependent translation termination process is detected by Upf and Smg proteins. Here Upf3, together with Upf1 and Upf2, may signal the presence of the PTC to the 5' end of the transcript, resulting in decapping and rapid exonucleolytic digestion of the mRNA. Some evidence shows that the resulting truncated protein product can be targeted for ubiquitylation and destroyed by the proteasome in an Upf1- and Cdc48-dependent manner. m7G refers to the mRNA 7-methyl guanosine cap; AAAAA refers to the poly(A)-tail; Ter indicates a normal termination codon.

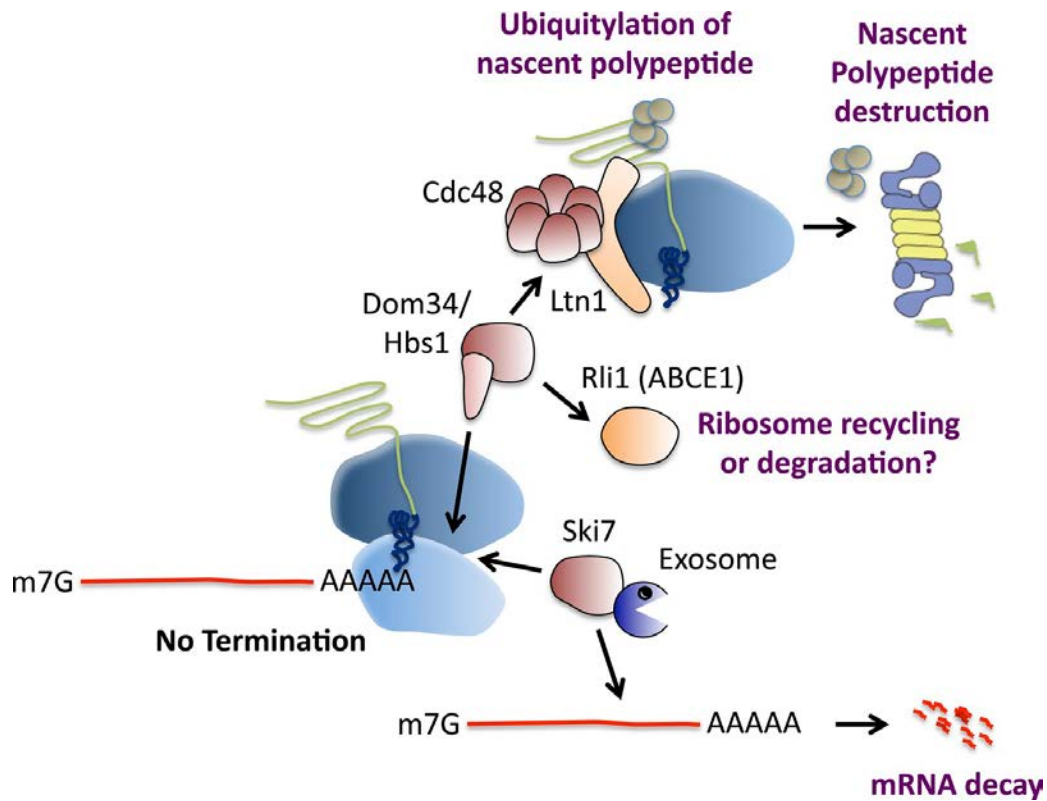


Figure 1.3 Non-stop decay pathway (cited from Lykke-Andersen et al, 2014)

The ribosome reaches the mRNA's 3' end without encountering a stop codon which is detected by the eRF3-like factor Ski7 and the mRNA targeted for decay by the exosome. Ski7 is thought to bind to the empty ribosomal A site, thereby the stalled ribosome is thought to be released by the Hbs1–Dom34–Rli1 complex, and thus recruit the exosome, resulting in rapid decay from the 3' end. The destruction of the defective protein products resulting from NSD mRNA substrates is mediated by ubiquitin pathway components Ltn1 and Cdc48.

Chapter Two: Alternative polyadenylation as a potential mechanism to respond to drought stress in *Arabidopsis thaliana*

2.1 Introduction

Plants are under several different environment stresses which all impact growth and development (Seki et al. 2003; Farooq et al. 2009a; Farooq et al. 2009b). Among these, drought is considered the most devastating abiotic stress factor limiting crop productivity in modern agriculture worldwide (Zhang 2003; Fracasso et al. 2016). Climate change and water scarcity both give rise to drought stress. More than others, water becomes a resource in demand given current and future human population and societal needs (Rosegrant and Cline 2003). Therefore, an understanding of drought stress in relation to plant growth is very important for sustainable agriculture.

The adaptation of plants to stress requires tight regulation of gene expression. Gene expression is regulated at the levels of transcription and translation and by increasing or buffering transcriptional effects by post-transcriptional mechanisms (Suzanne and Pamela 2008; Hollerer et al. 2016). Recent studies have shown the crucial role of alternative polyadenylation (APA) in different species and have connected APA to epigenetic regulation associated with various biological processes (Lutz and Moreira 2011; Xing and Li 2011b; Ma et al. 2014a). In mammals, APA can give rise to a global shortening of 3'-UTRs in proliferating cells and in cancer (Sandberg et al. 2008; Mayr and Bartel 2009; Elkou et al. 2012; Lin et al. 2012; Felice-Alessio et al. 2013). Also, the APA can result in a transcriptome-wide 3'UTR lengthening in differentiated cells and tissues (Sandberg et al. 2008; Mainul et al. 2012). Besides 3'UTRs, poly(A) sites can fall into other genomic regions of pre-mRNA. Genome-wide studies in yeast showed that the mRNA isoforms

with 3' ends within protein-coding regions were up-regulated in cells subjected to environment stresses (Yoon and Brem 2010). This was also seen in selected genes (Sparks and Dieckmann 1998). mRNA isoforms with 3' ends that lie within noncanonical regions have been described in plants (Shen et al. 2011; Wu et al. 2011; Wu et al. 2014). In *Arabidopsis* and *Medicago truncatula*, polyadenylation within the coding region seems to be evolutionarily conserved, suggestive of an important function for these mRNA isoforms (Wu et al. 2014). Genes associated with polyadenylation within coding regions and introns tend to encode proteins and enzymes associated with stress responses in *Arabidopsis* (Wu et al. 2011). These studies suggest that APA may be important for regulating stress responses in plants.

Given the connections between genes associated with stress responses and APA, we hypothesized that drought stress may cause changes in the usage of polyadenylation site to generate different mRNA isoforms. In this context, studies were conducted to evaluate differential poly(A) site usage in plants exposed to simulated drought conditions achieved by growth of plants on media containing mannitol. The results showed the levels of mRNA isoforms with 3' ends within coding sequences and 5'UTRs increase after different drought treatments. However, the mRNA isoforms with 3' ends in 5'-UTRs and in coding regions significantly increase after times as short as one hour. This suggests that APA may be among the early response mechanisms for responses to drought. The GO analysis showed that transcription, post-transcription and translation are processes overrepresented for the mRNA isoforms that fall into 5'UTR and CDS genomic region, especially the CDS region under drought stress. Taken together, these experiments indicate that the non-canonical

isoforms derived from APA within 5'UTRs and CDSs may play important roles in coping with drought stress.

2.2 Results

To study the relationship between drought stress and alternative polyadenylation, *Arabidopsis* plants were subjected to a number of different stress treatments, and poly(A) site choice subsequently analyzed. In one set of experiments, *Arabidopsis* plants were grown for two weeks in the presence of different concentrations of mannitol in MS solid medium to simulate sustained drought stress. In another set of experiments, *Arabidopsis* plants were grown under normal conditions for two weeks and then placed in liquid media containing different concentrations of mannitol. Plants so treated were harvested after different times for RNA isolation and further study (Figure 2.1).

2.2.1 Validation of drought stress treatments with qRT-PCR

To make sure that both short- and long-term drought stress was effective, potential reference genes were chosen for qRT-PCR in *Arabidopsis thaliana*. Based on previous studies and TAIR (Huang et al. 2008), one internal control Act II and eight diagnostic genes were used to test the effectiveness of the drought stress treatments. Primers were designed to assess the transcript accumulation of candidate genes (Table 2.1).

The qRT-PCR results revealed that the expression levels of the selected diagnostic genes increased when plants were grown on solid media containing increasing concentrations of mannitol (Figure 2.2). The expression levels of all the diagnostic genes increased when plants were grown on 100mM mannitol treatment, such as *MEPRIN* and *TRAF HOMOLOGY PROTEIN* (At1g58270), ABA signaling regulator *PROTEIN PHOSPHATASE2C* (*PP2C*; At3g11410) and the *SM-LIKE PROTEIN* (At5g48870).

Moreover, the expression levels of some of these genes decreased as the concentration of mannitol increased. For other genes, expression remained the same or increased with increasing mannitol concentrations. The fold change of At1g58270 is 3.4, while the previous study was 4.0 under drought (Huang et al. 2008), which is mainly for drought-responsive genes that are regulated by plant hormones. Several studies demonstrated At3g11410, which encodes the ABA regulator PP2C, is involved in drought stress (Harb et al. 2010; Joshi et al. 2016; Sinha et al. 2017). The gene At2g41190 was also used as reference to validate gene expression alterations due to treatment and increases in expression of this gene have been documented to be among the greatest under dehydration stress (Qin et al. 2008). However, the expression level change of At2g41190 in this study was only from 2.1 to 4.5 relative to the *ACTIN* housekeeping reference gene. Nevertheless, these results showed that the mannitol treatment could effectively induce the drought response.

To further explore different conditions for the drought stress study, a time and concentration series treatment was performed with plants exposed to 0mM, 100mM, 250mM, and 300mM mannitol (Figure 2.1). qRT-PCR was used to validate these treatments using the diagnostic genes listed (Table 2.1). The results showed that the expression levels for most genes increased upon exposure to 100 mM mannitol at all time points. However, as for the study in Fig 2.2, there were some difference with some of the genes. Most notable was that At1g07720 responded minimally to the treatments except to 100mM, due to this gene was mainly for the response of hormones under drought stress (Huang et al. 2008). This one instance aside, these results show that the liquid treatments were also effective at inducing drought stress.

2.2.2 Preparation of high-throughput poly(A) tag libraries and sequencing

Based on the results of qRT-PCR analysis (Figures 2.2 and 2.3), all of the drought stress treatments seemed effective. Therefore, the different concentration series (0mM, 100 mM, 250 mM, and 300 mM mannitol) and different time series (0h, 6h, 24h and 2 weeks) with 300mM mannitol were chosen for preparing poly(A) tag (PAT) libraries. The procedure for this is illustrated in Figure 2.4 and described in the Methods (Section 2.4). The total RNA was fragmented, poly(A)-enriched, and used for reverse transcription. During reverse transcription, an anchored-dT₁₈ was used to attach one sequencing adaptor and a novel strand-switching activity of the reverse transcriptase (Zhu et al. 2001) was used to attach the other adaptor. The cDNA libraries were used as templates for limited PCR amplification and size selection to obtain pools of 400-500bp. After the final quality control assessment using the Bioanalyzer, the PAT libraries were submitted for high throughput sequencing.

The sequencing data were analyzed using a series of steps. Using CLC Genomics Workbench, sequences were demultiplexed through the adaptor barcode sequence and aligned to the *Arabidopsis* genome. Each PAT library yielded from 2 million to 11 million reads after trimming. Between 2.67 million to 7.76 million reads aligned to the *Arabidopsis* genome for each sample [three PAT library (triplicate)] (Tables 2.2 and 2.3).

Aligned reads were analyzed in several different ways. First, they were used to estimate overall gene expression by counting the total number of reads that mapped to each annotated *Arabidopsis* gene. The results can be compared with previous similar published RNA-seq or microarray data to estimate the consistency of treatments. Second, the poly(A) site analysis was conducted by a dedicated computational pipeline (Bell et al. 2016). The

raw reads were trimmed to remove the adaptor sequences before mapping. Mapped reads were used to analyze poly(A) sites and poly(A) site clusters (PACs), which are groups of poly(A) sites located within 24nt of each other in the same gene (Figure 2.4). From the collection of experimentally-determined PACs, the individual PACs were analyzed for their relative usage. For this, the changes of the poly(A) site usage was calculated to generate a metric. The metrics were further analyzed to estimate the changes of poly(A) site choice induced by drought stress.

2.2.3 Drought stress incites changes in poly(A) site choice in different genomic regions

To study drought-induced global gene expression, the gene expression levels were estimated by mapping PATs to individual genes, and further analyzing the results using CLC Genomics Workbench. For these analyses, genes were assigned to be significantly different if their expression changed by at least 2-fold, and the results of a Student's t-test for the treated-control comparison returned a p-value <0.001. The results showed that the number of genes that passed these filters at higher concentration (Figure 2.5A). Also, the number of down-regulated genes were greater than the number of upregulated genes, except at 100 mM mannitol. Moreover, when looking at the up-regulated genes, 67 such genes were seen in plants treated with 250 mM that were not seen in 100 mM mannitol, and 97 genes were apparently upregulated in the 250 mM and 300 mM mannitol comparison (Figure 2.5B). For the numbers of down-regulated genes, 92 repressed genes were seen in 250 mM mannitol compared with 100mM mannitol, and 120 genes had reduced expression in the 250mM-300mM comparison (Figure 2.5C). When the results of this study were compared with a similar progressive drought study (Bechtold et al. 2016),

111 differentially expressed genes (DEGs) were seen in both studies (Figure 2.5D). Furthermore, 244 DEGs categorized in my study were also identified as drought-responsive in yet another study (Harb et al. 2010). These results confirm those shown in Figures 2.2 and 2.3, and show that the drought stress applied in my work was effective.

To determine the effect of drought stress on poly(A) site choice, individual PACs were evaluated for differential poly(A) site choice under control and drought stress condition. The number of tags that mapped to a given PAC was divided by the total number of tags that mapping to the associated gene from *Arabidopsis* genome to get values of fractional PAC usage for control and drought treatment. The \log_2 transformed value of the ratio of usage in stressed and control samples were used to evaluate changes in poly(A) site choice in response to drought stress (Figure 2.6).

A poly(A) site may fall into any genomic location, such as 5'-UTR, coding sequence, intron, 3'-UTR and intergenic region. mRNA functionalities or outcomes might be different in these different instances (de Lorenzo et al. 2017). Accordingly, the responses of these different isoforms were studied. The results for the comparison of plants grown on solid media with or without 300 mM mannitol are shown in Figure 2.7. The median change for all sites ("all" in Figure 2.7) was close to zero. Isoforms derived from polyadenylation in 3'-UTRs and introns displayed similar results, but the 25-75 percentile range for intron was much wider, suggesting there is greater variability in the usage of these sites. For isoforms derived from polyadenylation within the protein coding region ("CDS" in Figure 2.7) and the 5'-UTR, higher levels were seen as were greater site-by-site variability (Figure 2.7).

For the time series treatment, the relative usages for individual poly A sites were calculated and determined from the comparison between controls and 300mM treatments after 1h, 6h and 24h. They are shown in Figure 2.8. When comparing control plants, there is no change in usage when considering all sites together (“All” in Figure 2.8). Similarly, there is almost no change in usage of sites that fall within 3'-UTRs and introns. The usage of sites within 3'-UTRs shows less variability than sites that fall within introns. The usage of poly(A) sites that lie within 5'-UTRs and protein coding regions (“CDS”) increased after 1 and 6 hours in mannitol (Figure 2.8A and 2.8B). However, after 24 hours, only 5-UTR-localized sites showed increased usage (Figure 2.8C). These results suggest that stress-induced poly(A) site changes may play roles in early responses of plants to drought stress.

To further explore drought-associated APA, those individual sites whose differential usage changed the most significantly (P-value<0.01, determined using DEX-seq package) were identified and evaluated. The list of sites that satisfied this criterion and transcripts associated with these sites. GO analysis showed that these genes are highly enriched for those of the intracellular organelle component, non/membrane-bound organelles component, cytoplasmic part, and macromolecular complex part. Moreover, the genes associated with the protein folding process, cellular protein/macromolecular metabolic process, reproductive development process and abiotic process (Figure 2.9) are also unduly affected by APA in drought stress. Interesting, there are more transcripts that have significantly differential usage of poly(A) sites under 250 mM mannitol treatment. GO analysis showed that these genes were associated with binding activity and structural molecular activity for molecular function (Figure 2.9B).

2.2.4 Annotation and functional analysis of poly(A) site choice in 5'UTR and CDS regions

As stated in the previous section, no matter the concentration- or time-treatment, the usage of poly(A) sites that fall within 5'UTRs and CDS changes more than the usage of sites within other regions (Figures 2.7 and 2.8). To further analyze this, the annotation and function of transcripts affected by APA at sites that fall into 5'UTRs and CDS were assigned using agriGO. The annotations were verified manually and integrated using gene ontology (GO). For this, the reference is the *Arabidopsis* gene model (TAIR); the statistic model is the hypergeometric test; the multi-test adjust method is false discovery rate (FDR) with and adjusted p-value less than 0.05; and the minimum number of mapping entries is more than 9. The transcripts that fall into 5'UTR and CDS were annotated according to at least one of the three categories: cellular component, biological process or molecular function. The same parameters were used for GO analysis for all the transcripts affected by stress-induced APA in 5'UTRs and CDSs.

For one-hour treatment, 428 poly(A) sites that fall into the 5'UTR were identified; these to 407 transcripts. These 407 transcripts could be categorized into 104 functional groups. Among these groups, the terms related to biological process that are significantly enriched include stimulus (GO:0050896), abiotic stimulus (GO:0009628), water deprivation (GO:0009414), and hormone stimulus (GO:0009725) process, such as jasmonic acid (GO:0009753) and salicylic acid (GO:0009751) (Figure 2.10A). Molecular function categorizations for this group included transcription (GO:0006350), transcription regulator activity (GO:0030528), catalytic activity (GO:0003824), transporter activity (GO:0005215), transferase activity (GO:0016740), transferring phosphorus-

containing groups (GO:0016772), DNA binding (GO:0003677) and protein binding (GO:0005515) activity (Figure 2.10B).

For poly(A) sites that fall into the CDS, around 10316 differentially-utilized poly(A) sites corresponding to 5743 transcripts could be identified. These could be separated into 565 functional groups. For biological process, the set of transcripts is highly enriched for RNA processing (GO:0006396), RNA metabolic process (GO:0016070), tRNA metabolic process (GO:0006399), tRNA aminoacylation for protein translation (GO:0006418), RNA splicing (GO:0008380), cellular protein metabolic process (GO:0044267), cellular metabolic process (GO:0044237), cellular nitrogen compound metabolic process (GO:0034641), macromolecule metabolic process (GO:0043170). Also, some transcripts are enriched in response to abiotic stimulus (GO:0009628), response to oxidative stress (GO:0006979), response to inorganic substance (GO:0010035), response to osmotic stress (GO:0006970), response to cadmium ion (GO:0046686), intracellular signaling cascade (GO:0007242), and response to metal ion (GO:0010038) (Figure 2.11A). Molecular function characterizations included binding activity (GO:0005488), nucleotide binding (GO:0000166), purine nucleotide binding (GO:0017076), purine ribonucleotide binding (GO:0032555), protein binding (GO:0005515), nucleic acid binding (GO:0003676), ATP binding (GO:0005524), DNA binding (GO:0003677), zinc binding (GO:0008270). And the catalytic activity (GO:0003824), ATPase activity (GO:0016887), helicase activity (GO:0004386), translation factor activity (GO:0008135), and transcription regulator activity (GO:0030528) are involved (Figure 2.11B). These categories are dominated by biological processes and molecular functions related to transcription, post-transcriptional and translational processes under drought stress.

2.3 Discussion

2.3.1 Comments regarding the validation of the drought stress treatments

Previously, it was reported that only 27 drought-inducible genes could be seen in different drought stress studies (Bray 2004). Another study showed that 67 genes that are responsive to nine abiotic stresses (Swindell 2006). In another report, 197 genes were found to be induced by large range of diverse stress conditions in *Arabidopsis* (Ma and Bohnert 2007). Interestingly, only two genes were found in common between the last two studies. More recently, another study showed that 1067 genes were found to be stress-responsive in *Arabidopsis*, while only 111 genes were seen in other studies (Harb et al. 2010). In this study, 244 genes, more than 53%, identified as drought-inducible genes were also found in other studies of drought-stressed *Arabidopsis* (Harb et al. 2010; Bechtold et al. 2016).

The numbers of stress-responsive genes identified in drought stress experiments largely depends on the criteria used in selecting genes in a microarray or RNA-Seq experiment. Differential gene expression in earlier studies was selected based on expression ratios, not statistical thresholds. If the filter was a 5-fold change in expression, many weakly responsive genes are ignored. Therefore, the proportion of differentially expressed genes may be underestimated (Huang et al. 2008). In addition, the discrepancies from different studies may reflect the methods used for plant growth conditions, the application of stress treatment, and the purposes of the study. With these things in mind, the overlap of around 25% and 53% of stress-inducible genes in this chapter is not surprising, and supports the conclusion that the stress treatments were effective.

2.3.2 Alternative polyadenylation and the molecular function of plants to drought stress

Alternative polyadenylation (APA), a mechanism in which the same gene encodes mRNA isoforms with different 3'-ends due to the presence of multiple polyadenylation signal (PAS) elements, is pervasive in plants (Shen et al. 2011; Wu et al. 2011; Wu et al. 2014; Fu et al. 2016). In *Arabidopsis*, around 60-70% of genes have multiple poly(A) sites (Shen et al. 2011; Wu et al. 2011; Duc et al. 2013). The functions of mRNA isoforms derived from APA at promoter-proximal sites that lie within protein coding regions, introns, and 5'-UTRs have not been explained. One type of noncanonical mRNA isoform derived from polyadenylation within CDSs will, with few exceptions, lack translation termination codon. This kind of mRNAs are termed non-stop RNAs, which are less stable than canonical mRNAs in animals and yeast (Frischmeyer et al. 2002; van Hoof et al. 2002). For the intronic polyadenylation, the abundance of mRNA isoforms is associated with polysomes, suggesting these isoforms are in fact substrates for quality control processes, such as NMD or nonstop decay in plants under abiotic stresses such as hypoxia (de Lorenzo et al. 2017). In contrast to isoforms are produced by polyadenylation within intron and CDS, the isoforms derived from polyadenylation in 5'UTRs showed little difference in their stability from isoforms produced by canonical poly (A) site choice is unclear, but possibly including some sort of association with RNA quality control processes (de Lorenzo et al. 2017). These results indicate that noncanonical isoforms could be subjected to the RNA surveillance pathway.

The levels of mRNA isoforms with 3' ends that lie within 5'-UTRs and protein-coding regions increase (Figures 2.7 and 2.8) in drought-stressed plants. The mechanisms that

cause these increases have not been studied. However, other studies may shed new light on these mRNA isoforms. mRNA isoforms with 3' ends that lie within coding regions lack stop codons and should be relatively unstable. This mode of APA may be a negative regulatory mechanism. To cope with different kinds of stresses, plants can utilize transcriptional re-programming (Yamaguchi-Shinozaki and Shinozaki 2006; Chinnusamy et al. 2007; Hirayama and Shinozaki 2007; Verslues and Zhu 2007; Zhu et al. 2007). The production of non-stop mRNAs might be an additional re-programming via an analogous RNA quality control process.

The levels of 5-UTR-derived mRNA isoforms increase to the greatest extent after 1h of drought stress and is still higher than controls after 6h and 24h (Figure 2.8). The significance of this observation is unclear at this time. The stability of these mRNA isoforms is indistinguishable from isoforms with 3' ends that lie within 3'-UTRs (de Lorenzo et al. 2017), arguing against a negative regulatory function analogous to that proposed for non-stop mRNA isoforms. In yeast, previous studies showed that ribosomes and 5'UTRs may interact in ways that modulate the stability of nonaberrant mRNAs (Linz et al. 1997). This suggests that ribosomes and 5'-UTRs might interact in ways that may affect translation. However, how this relates to mRNAs with 3' ends that lie within 5'-UTRs, and thus have no protein-coding potential, is not clear.

Drought stress-responsive 5'UTR isoforms are associated with GO terms such as abiotic stimulus and water deprivation, both of which are relevant to drought stress. These characteristics were identified in many plants using microarray and RNA-seq, such as *Arabidopsis thaliana* (Harb et al. 2010), *Leymus chinensis* (one kind of sheepgrass) (Zhao et al. 2016), and *Prunus persica* (Ksouri et al. 2016). Under drought stress, ABA acts as

the central signal molecular and mediates a complex gene regulatory network (Yamaguchi-Shinozaki and Shinozaki 2006; Kim et al. 2010; Zhao et al. 2016). However, the mRNA isoforms that end within 5'UTRs are associated with different hormone stimuli, including jasmonic acid and salicylic acid, and not ABA pathways. Drought stress-responsive 5'-UTR isoforms are associated with transcription regulator activity, protein binding activity and catalytic activity. Previous studies showed that transcription factors (TFs) such as MYB, bZIP, C2H2, and NAC were expressed to a greater degree in plants under drought (Pereira et al. 2011; Thirunavukkarasu et al. 2017). These TFs include individuals that act in the ABA-dependent or ABA-independent signaling pathways (Shinozaki et al. 2003; Thirunavukkarasu et al. 2017). Thus, mRNA isoforms with 3' ends within 5'UTR may play an important role in regulating transcription under drought stress.

The associations of drought stress-responsive CDS isoforms (nonstop RNAs) are more complicated. These transcripts are significantly enriched for genes associated with RNA processing. Previous studies showed that RNA processing was critical for ability of plants to cope with drought stress (Wang et al. 2016). Specifically, the abundances of RNA processing-related proteins change when plants are subjected to drought (Wang et al. 2016). These include RNA binding proteins that can bind with RNA molecules to mediated post-transcription gene regulation (Wang et al. 2016). Drought-responsive nonstop RNAs are also associated with GO terms such as DNA binding, nucleotide binding, protein binding, ATP binding, and zinc binding.

2.3.3 Alternative polyadenylation factors might be involved in drought stress

There is a large range of physiological and biochemical responses induced by drought stress, such as osmoprotectant synthesis and changes in fatty acid composition (Krasensky

and Jonak 2012). Oxidative damage due to the accumulation of reactive oxygen species (ROS) (Smirnoff 1993) is another such outcome of drought stress. ROS results from partial reduction of atmospheric O₂, that leads to the production of four forms of cellular ROS, singlet oxygen, superoxide radicals, hydroxyl radicals and hydrogen peroxide. Singlet oxygen and hydroxyl radicals can be extremely reactive, leading to the oxidation of multiple cellular components and ultimately, if unchecked, to cell death (Foyer and Mullineaux 1994; Lerner and NetLibrary 1999; Mittler 2002; Apel and Hirt 2004). The toxic by-products of aerobic metabolism can be removed by means of antioxidants and antioxidative enzymes (Apel and Hirt 2004; Miller et al. 2010). ROS are also involved in signaling pathways that impact plant development, biotic- and abiotic-stress responses (Apel and Hirt 2004; Mittler et al. 2004). ROS can give rise to changes in gene expression, especially for the genes encoding ROS scavenging enzymes (Mittler 2002; Vranova and Villarreal 2002; Mittler et al. 2004; Gapper and Dolan 2006). In addition, ROS leads to rapid increases in intracellular calcium, resulting in a cellular signaling cascade (Price et al. 1994; Yang and Poovaiah 2002; Maize et al. 2004; Evans et al. 2005).

In plants, there is an enigmatic subunit of the polyadenylation complex, CPSF30 (Hunt et al. 2008). The *Arabidopsis* CPSF30 is impacted by calmodulin and sulfhydryl reagents in vitro (Delaney et al. 2006; Hunt et al. 2008). Several calcium binding proteins, such as calmodulin, calcium sensing receptor, calreticulin and calcium-dependent protein kinase, were changed in response to drought (Wang et al. 2016). Specifically, the RNA binding activity of CPSF30 can be inhibited by calmodulin in a calcium-dependent manner. In addition, the third zinc finger contains a disulfide bond that can be reduced by sulfhydryl reagents (Addepalli et al. 2010); reduction of this bond inactivates the endonucleolytic

activity of the protein. As stated above, generation of ROS leads to increased levels of intracellular calcium (Maibe et al. 2004; Evans et al. 2005). This could conceivably result in a calcium-calmodulin dependent inhibition of RNA binding by CPSF30, much as is seen in vitro. Also, exposure to oxidative stress may also lead to extensive disulfide bond remodeling (Barford 2004). This provides a possible mechanism by which stress, via ROS, may regulate the activity of CPSF30 (Chakrabarti and Hunt 2015). Given that CPSF30 controls the usage of many poly(A) sites in plants (Thomas et al. 2012), these considerations raise the possibility that CPSF30 may be a link between stress and APA through the combined actions of calcium and redox signaling pathways.

To tell if AtCPSF30 is involved in the response to drought stress, all the genes from 1 hour 300 mM mannitol treatment were compared with all the genes that are affected by CPSF30-mediated APA (see Chapter Three). The result show that around 50% genes overlap, and most genes possessing wt-specific poly(A) sites (7916 out of 7986) are involved in drought stress. Genes with noncanonical sites that fall within in 5'UTRs, CDS, and intron were compared in the set of those possessing wt-specific sites and those that respond to 1h of exposure to 300mM mannitol. There were 410 noncanonical transcripts are in the wt-specific set of mRNAs, while 6431 are in the set of mannitol-responsive transcripts. The overlap of these sets is 363 noncanonical transcripts (almost all noncanonical wt-specific transcripts). These results indicate that AtCPSF30 may be involved in the responses of plants to drought stress.

2.4 Material and Methods

2.4.1 Plant material and treatment

Arabidopsis thaliana wild type Col-0 was cultured vertically on solid and liquid MS media (1x Murashige and Skoog salts, MES 0.5g/L pH = 5.7, and 1% sucrose). Plants were grown under long day conditions, 16h light and 8h dark for two weeks at 23 °C. The light intensity was 30 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ from Sylvania Octron ECO 5000 K 32W light bulbs. The surface of the seeds was sterilized before plating and put at 4°C for two days. In all experiments three biological samples were prepared.

For the study involving different concentrations of mannitol, drought stress was induced on solid MS media containing different concentration of mannitol (0 mM, 100 mM, 250 mM, 300 mM). The control samples were put under the same condition as treated samples without mannitol (Pandey et al. 2013). To study stress at different times after imposition of the stress, plants were grown on MS media for two weeks under normal conditions, and then placed in liquid media containing 300mM mannitol for 1h, 6h and 24h. The control samples were grown with MS liquid media without mannitol.

2.4.2 qRT-PCR condition and analysis

For quantitative real-time PCR, total RNA was extracted using Plant Trizol (Invitrogen), and SMARTScribe™ Reverse Transcriptase (Clontech) was used for reverse transcription with 1 μg of total RNA. SsoAdvanced SYBR Green Supermix (Bio-Rad) and a CFX Connect real-time PCR system (Bio-Rad) was used for qRT-PCR in triplicate on 96-well plates. The cycling programs were as follows: initial DNA denaturation step (95°C for 30s), 40 cycles (95°C for 5s, 55°C for 30s), and melt curve of 65°C to 95°C (in 0.5°C increments, 5s per step) (Wang and Perry 2013). The reference control genes were

measured with three replicates in each PCR run, and the C_t from the average was used for relative gene expression analysis. The expression data value from candidate genes were normalized by subtracting the mean of reference gene C_t value from candidate C_t value get ΔC_t . The $\Delta\Delta C_t$ was from ΔC_t of candidate gene subtracting the average of control ΔC_t . The fold change was obtained using the expression $2^{-\Delta\Delta C_t}$.

2.4.3 Poly(A) tag library preparation and sequencing

Total RNA was isolated using Trizol reagent and purified by RNAeasy columns (Qiagen, Hilden, Germany). A Nanodrop spectrophotometer was used for quantity and quality measurement (Biotek Synergy HT Multi-Mode Microplate Reader). Very pure RNA will have an A260/A280 ratio of ~2.1. Anything higher than 1.8 is considered to be of acceptable purity. Also, The A260/A230 ratio should also be above 2.0. Each poly(A) tag library was generated from 1 μ g from total RNA using method B1 described by Ma et al. (Ma et al. 2014a). The quality of the PAT libraries was checked by on a Bioanalyzer using the Agilent High Sensitivity DNA chips (Agilent Technology). The concentration of these libraries is measured with Qubit fluorometer with Qubit® dsDNA HS assay kit (Agilent Technology). These poly(A) tag libraries were sequencing on the Illumina high-throughput sequencing platform. Three independent biological replicates were used for each sample.

2.4.4 Poly(A) tag and gene expression analysis

The poly(A) tag(PAT) libraries sequencing data were analyzed using the pipeline as described previously (Bell et al. 2016). Briefly, all the reads were demultiplexed and trimmed to remove the oligo-dT and sequencing adaptors using CLC Genomic Workbench suite of tools. The trimmed tags were mapped to the *Arabidopsis* genome (TAIR10, www.Arabidopsis.org). The pooled samples from triplicates can be exported as. bam files.

The quality can be evaluated by mapping to each sample to determine the gene expression. The sequence of all the pooled samples sequence can be used for the poly(A) analysis. The PATs that mapped to individual annotated genes can be counted using Bed Tools, and to create lists of each individual poly(A) sites (PAS) and poly(A) sites clusters (PAC). A minimum of ten individual PATs for PASs and PACs are kept for further analysis. At least ten individual PATs for sites and cluster are kept for subsequent analysis.

To determine gene expression, the PAT frequency of each gene was determined by a special file (TAIR10genes120.gff) that has only *Arabidopsis* genes. The data can be imported to CLC using the empirical analysis of DGE (differential gene expression) to compare the control and treatment. The genes with a total filter cut-off of 2-fold change and p-value <0.001 were selected as statistically significant.

To determine the PAT frequency for each PAC, the list of PACs is used as annotated master file to tell the numbers of tags in each individual sample that map to the PAC. Thus, the numbers of tags in each individual sample are compared with the number of tags in the *Arabidopsis* genome to get a proportion for poly(A) site usage analysis.

To calculate the statistically significant differences in poly(A) site usage in gene by gene analyses the DEX-seq package in R was used at a p-value <0.01 which was considered as statistically significant.

2.4.5 Relative poly(A) site usage analysis

Relative poly(A) site usage was calculated as shown in Figure 2.6. The \log_2 transformed value of PAS/gene treated sample divided PAS/gene control sample is defined as the Relative Poly(A) site Usage metric. The mRNA isoforms that fall into the CDS are used as an example for the calculation of relative mRNA stability. In the control sample, A, B,

C, D represent the numbers of tags that map to the poly(A) sites that fall into 5'UTR, intron, CDS and 3'UTR without treatment, respectively. In the treated sample, A', B', C' and D' represent the numbers of tags that map to the poly(A) sites that fall into 5'UTR, intron, CDS and 3'UTR under treatment, respectively. Therefore, the relative usage of site C is M in control sample, while N is the relative usage of site C' for the treated sample. The change in relative poly(A) site usage is N/M ; this value is \log_2 transformed before plotting results. (<http://shiny.chemgrid.org/boxplotr/>).

2.4.6 GO analysis

Gene Ontology term enrichment was performed using agriGO, which is a toolkit and database for agriculture community (Website <http://bioinfo.cau.edu.cn/agriGO/>). Parameters used were: reference is *Arabidopsis* gene model (TAIR); the statistic model is hypergeometric test; the multi-test adjust method is FDR adjust p-value; the significance is less than 0.05; and the minimum number of mapping entries is 9. In the GO figures, the X-axis is the negative log FDR adjust p-value and the Y-axis is the GO terms.

Table 2.1 Selected reference genes for *Arabidopsis* and corresponding primer pair information

Gene ID	Gene Name	Primer sequence F/R (5'-3')	Amplicon length(bp)	Amplicon T _m (°C)
AT3G18780	Actin II	GCTCTTGACCTTGCTGGACG	1134	58.5
		CACAAACGAGGGCTGGAACA		58.0
AT1G58270	Meprin and TRAF homology domain containing protein	ACAAGTACCAAATAAGCCAGGAGA	1191	55.9
		ATGTTGATGTCAACGCCAAAC		54.3
AT2G41190	Amino acid transporter	GCAGCGTCATTCAAACCATT	1611	54.3
		TTACCATACTTGCCCAACCAG		54.9
AT3G62550	Universal stress protein (USP) Gene	CCCCTCTCCCGTTFACTCT	489	56.1
		GCATCTCCTTCCAACTCG		55.3
AT4G02200	Response to water deprivation	TTGTGGGTTTGATTGTC	687	50.2
		AATCCACTGCCATTTTCAT		49.2
AT3G57520	Alkaline α Galactosidase	GATAAGCCAGGCAACCAC	2322	53.7
		GGATCAGCGAATAAGCAGT		52.3
AT5G48870	Sm-like snRNP proteins	TTGGGTGATAATGAAAGG	267	46.5
		CAGAATGGCGATGTTGTT		50.6
AT1G07720	Ketoacyl-CoA synthesis family protein	GTCATCACTCCCAAGATTC	1437	50.2
		CGATAACCGCTTTACCTCC		52.9
AT3G11410	Protein phosphatase 2C(PP2C)	TCGTAACGGTGTAGCCATTCC	1200	56.9
		ATAGTCCATCACTCGCCAAGAT		55.7

Table 2.2 PAT-seq sequence data information for different concentration of mannitol treatment

Concentration	Number of reads	Average length	Number of reads after trim	Percentage trimmed	Average length after trim	Mapped genome	Genome (%)
0R1	11,707,815	74.2	11,469,301	97.96%	58.4	4,803,600	59.97
0R2	7,873,950	73.4	7,838,672	99.55%	59.5	1,302,561	44.63
0R3	7,785,199	71.8	7,684,335	98.70%	57.2	1,657,080	40.42
100R1	7,769,335	91.6	7,697,889	99.08%	77.5	3,506,352	74.42
100R2	3,281,244	91.6	3,267,843	99.59%	77.9	583,191	48.49
100R3	3,652,714	89.5	3,621,831	99.15%	74.6	1,126,871	59.48
250R1	6,151,988	91.8	6,067,819	98.63%	75.7	3,640,708	73.16
250R2	7,168,937	89.7	7,117,643	99.28%	75.5	1,487,633	49.66
250R3	2,971,359	91.8	2,932,590	98.70%	77.1	967,624	62.81
300R1	9,518,342	73	9,392,832	98.68%	58.3	2,666,613	55.32
300R2	10,268,266	71.6	10,224,728	99.58%	57.7	1,444,568	37.21
300R3	2,731,849	91.8	2,704,432	99%	76.5	608,002	44.82

In all instances, R1, R2, R3 refers to replication 1, 2, and 3. 0 refers to the control samples, 100 refers to 100mM mannitol treatment, 250 refers to 250mM mannitol treatment, 300 refers to 300mM mannitol treatment.

Table 2.3 PAT-seq sequence data information for different time of mannitol treatment

Time*	Number of reads	Average length	Number of reads after trim	Percentage trimmed	Average length after trim	Mapped genome	Genome (%)
C1R1	2,752,919	91.9	2,729,777	99.16%	77.2	936,045	66.6
C1R2	2,635,930	92	2,570,946	97.53%	76.4	1,337,743	72.97
C1R3	2,682,875	91.9	2,640,094	98.41%	77.1	1,335,143	78.69
C6R1	3,089,237	91.9	3,064,861	99.21%	77.2	929,385	66.92
C6R2	2,122,206	91.9	2,055,926	96.88%	75.5	1,087,510	70.63
C6R3	2,785,286	91.9	2,737,572	98.29%	76.1	1,163,480	69.79
C24R1	3,533,966	92	3,511,789	99.37%	77.6	731,200	48.37
C24R2	3,089,642	91.8	2,999,843	97.09%	75.9	1,007,878	53
C24R3	5,000,558	92	4,906,102	98.11%	76.2	1,786,755	63.94
M1R1	2,756,213	91.9	2,736,602	99.29%	76.8	571,198	61.89
M1R2	1,962,515	92	1,908,908	97.27%	76.3	954,816	75.46
M1R3	2,550,958	92	2,500,046	98%	76.1	1,147,787	72
M6R1	3,038,561	92	3,017,262	99.30%	77.9	956,162	69.03
M6R2	2,065,761	91.9	2,025,557	98.05%	76.8	1,049,515	74.49
M6R3	2,960,666	92	2,933,781	99.09%	77.3	1,091,548	74.45
M24R1	5,739,327	91.9	5,641,449	98.29%	76.1	741,014	25.41
M24R2	4,370,747	91.9	4,283,686	98.01%	76.4	1,177,343	56.58
M24R3	5,802,731	92	5,731,230	98.77%	76.1	1,410,453	76.68

*Time: In all instances, R1, R2, R3 refers to replication 1, 2, and 3. C1, C6, and C24 denotes the control samples at 1, 6, and 24 hours after initiation of treatment. M1, M6 and M24 refers to the mannitol treatment samples at 1, 6, and 24 hours.

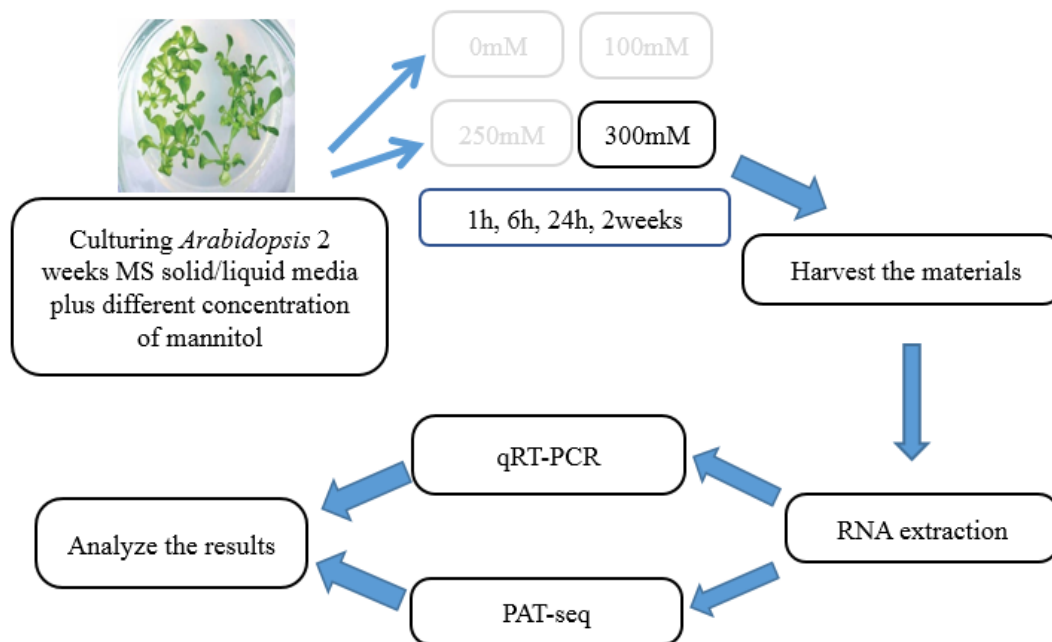


Figure 2.1. Overview of experimental approach to study APA under drought stress

This flowchart shows two experiments. First, two-week old seedlings were cultured in MS solid media under 0mM, 100mM, 250mM, and 300mM. Second, two-week old seedlings were totally emerged in different concentration treatments for 1h, 6h and 24h. After two weeks, all the materials were harvested to perform RNA extraction. All samples have three biological replicates. The drought stress was validated by qRT-PCR for concentration and time treatment, and thereafter PAT libraries for sequencing were prepared.

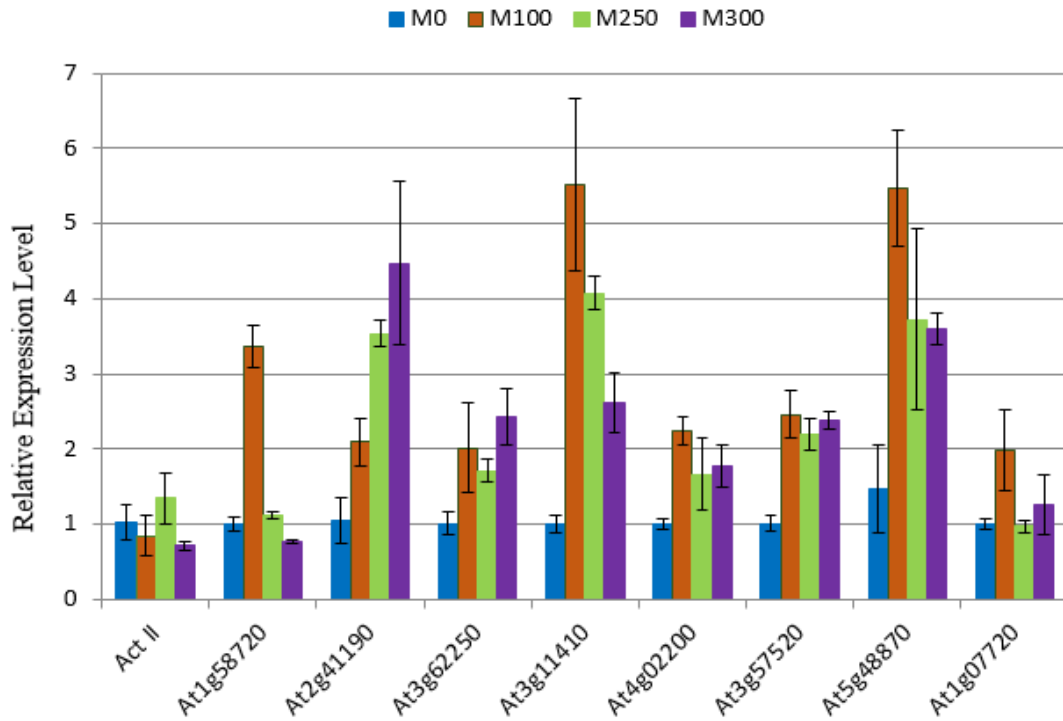


Figure 2.2 qRT-PCR to validate different concentration treatment under drought stress

qRT-PCR was performed to study the expression of eight drought-responsive genes under mannitol simulated drought conditions. Two-week old *Arabidopsis* seedlings were cultured under different concentration treatments. The RNA was extracted to make cDNA for qRT-PCR. The X-axis are the control and reference genes. The ActinII (At3g18780) was chosen as internal reference gene. The drought-response genes were chosen from previous studies under drought stress condition. For example, At3g11410, which encoded the ABA regulator PP2C, is involved drought stress. The gene fold change of At1g58270 is 3.4, while the previous study was 4.0 under drought (Harb et al. 2010; Joshi et al. 2016). Another gene At2g41190, which is response to water deprivation, was also used as reference to validate gene expression level (Qin et al. 2008). The gene At5g48870 is response to ABA and water deprivation (Cui et al. 2014). The results show that all three mannitol concentrations were effective, as each induced the expected changes in expression. The Y-axis is gene expression level from fold change.

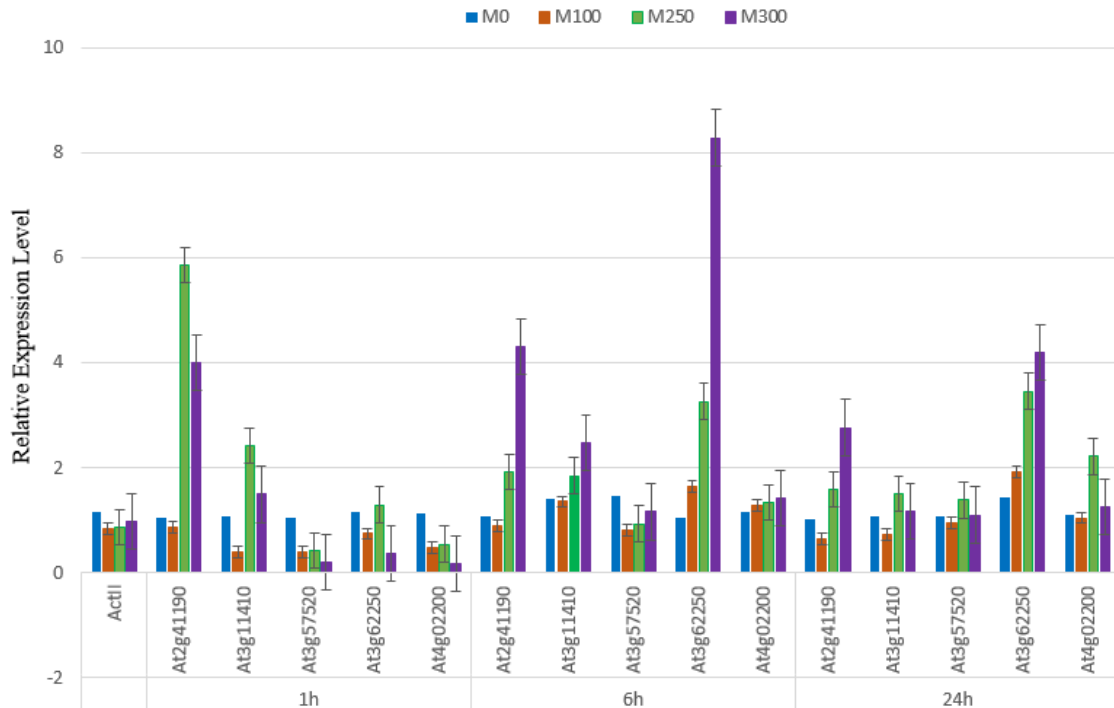


Figure 2.3 qRT-PCR to validate different time treatment under drought stress

qRT-PCR was performed to study the expression of eight drought-responsive genes under mannitol simulated drought conditions. Two-week old *Arabidopsis* seedlings were cultured under different concentration treatments. The RNA was extracted to make cDNA for qRT-PCR. The X-axis are the control and reference genes. The ActinII (At3g18780) was chosen as internal reference gene. The gene At2g41190, which is involved in response to water deprivation, was also used as reference to validate gene expression level in previous study (Qin et al. 2008). The gene At3g62250, which is involved in ubiquitin-dependent protein catabolic process, translation, was involved in ROS pathway (Rojas et al. 2012). The Y-axis is gene expression level from fold change.

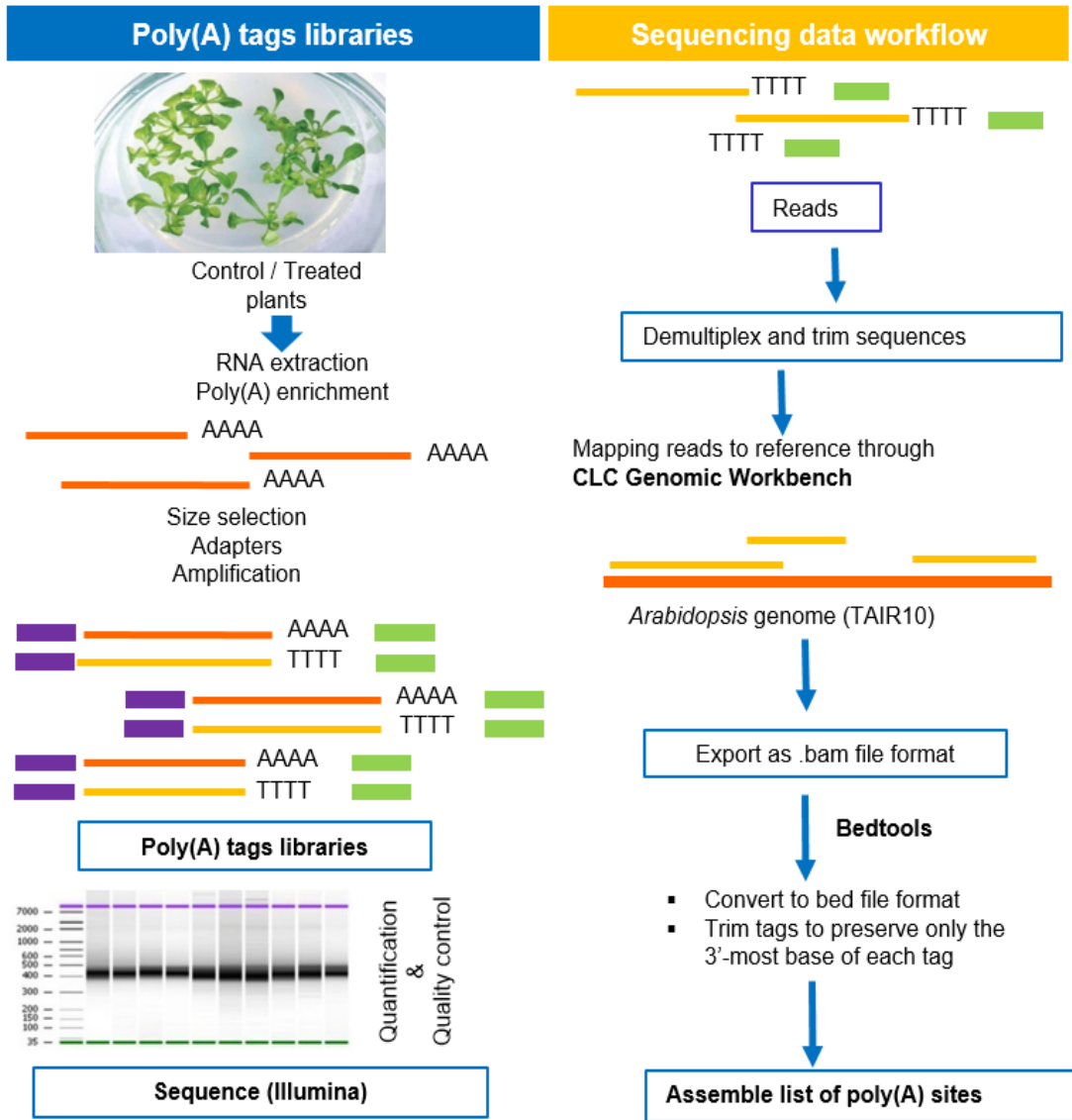
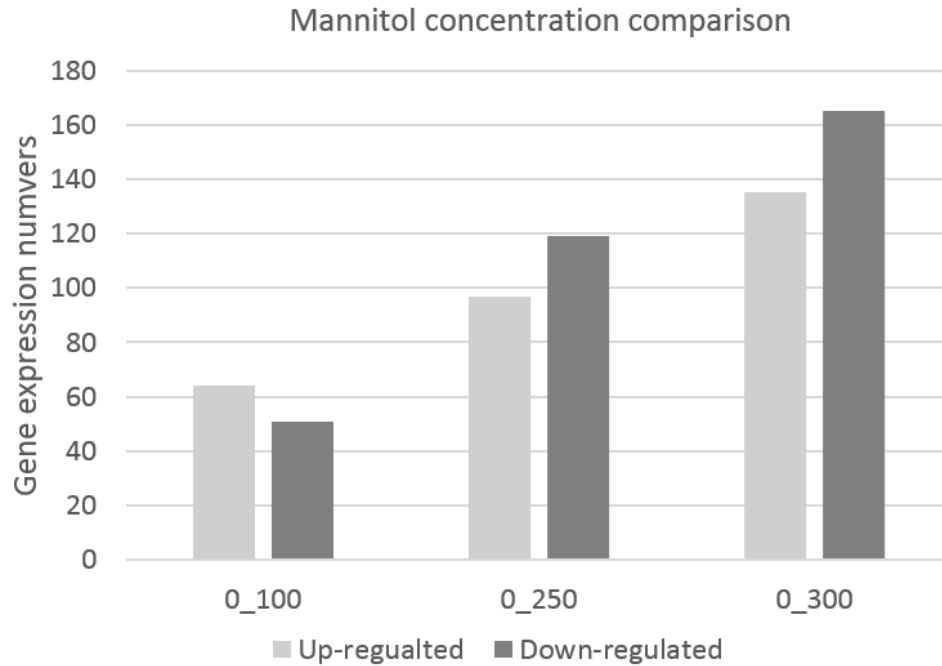


Figure 2.4 Preparation for PAT libraries and sequencing data analysis workflow

The general experimental approach carried out in this study is shown in two sections: poly(A) tag library preparation (left), the sequencing data workflow (right), and ready for next step analysis, including gene expression analysis, distribution of PATs, relative poly(A) site usage and gene ontology (GO) enrichment. PATs, poly(A) tags; PAC, poly(A) cluster.

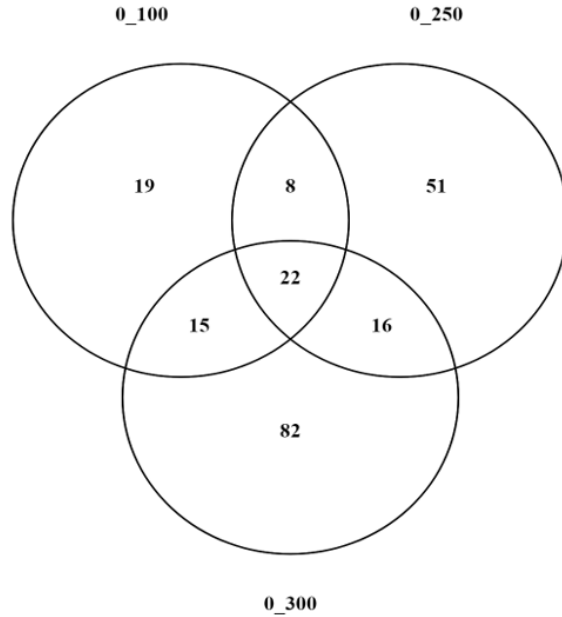
5A



5A. The tendency of up-regulated and down-regulated gene expression numbers in each comparison.

The X-axis shows different concentration treatments compared with control, and the Y-axis shows the up-, down-regulated gene expression numbers. The number of genes that passed these filters increased at higher concentration of mannitol. The number of down-regulated gene was greater than the number of upregulated genes, except at 100 mM mannitol.

5B

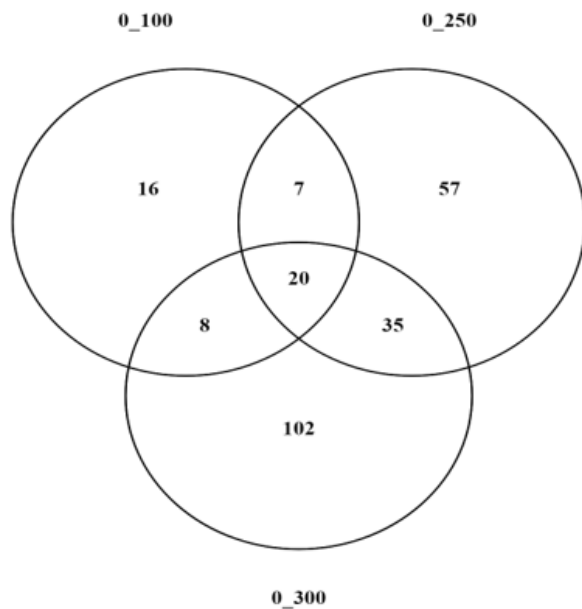


Up-regulated genes

5B. Venn diagram showing the numbers of up-regulated genes.

The numbers of up-regulated gene expression numbers in each comparison are presented using venn diagram. Moreover, when looking at the up-regulated genes, 67 new genes were seen in plants treated with 250mM that were not seen in 100 mM mannitol, and 97 new genes were apparent in the 250mM and 300Mm mannitol comparison.

5C

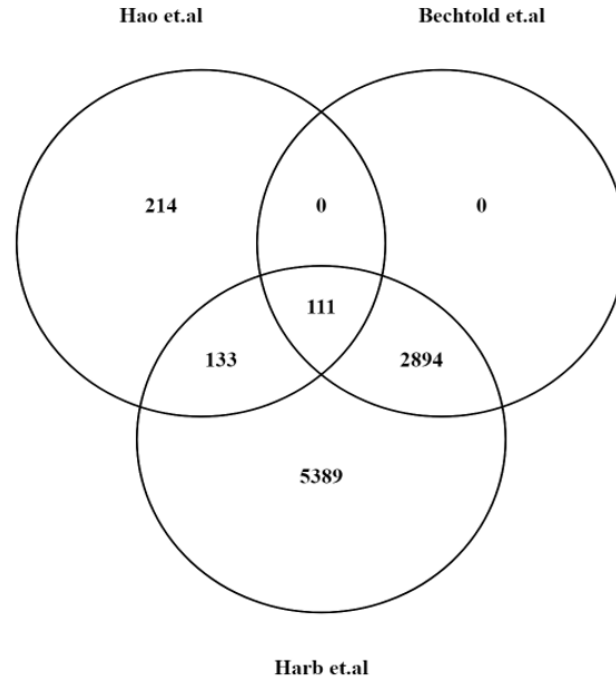


Down-regulated genes

5C. Venn diagram showing the down-regulated number of gene.

The numbers of down-regulated gene expression numbers in each comparison are presented by venn diagram. For the numbers of down-regulated genes, 92 new genes were seen in 250 mM mannitol compared with 100mM mannitol, and 120 new genes in the 250mM-300mM comparison.

5D.



5D. Venn diagram showing the total numbers of differential gene expression from this study compared with two previous studies. When the results of this study were compared with a similar progressive drought study (Bechtold et al. 2016), 111 genes were seen in both studies. 244 genes in my study were also identified as drought-responsive in yet another study (Harb et al. 2010).

Figure 2.5 Drought stress incites significant global gene expression changes

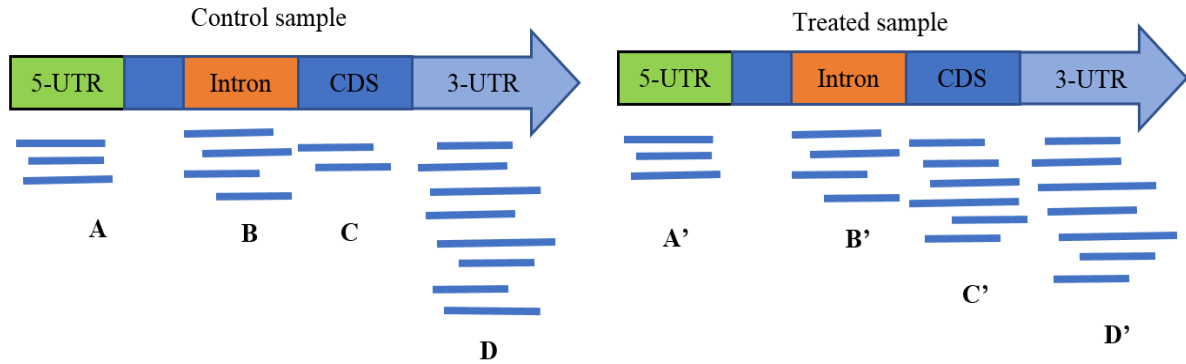
To study drought-induced global gene expression, the gene expression levels were estimated by mapping PATs to individual genes, and further analyzing the results using CLC Genomics Workbench. For these analyses, genes were assigned to be significantly different if their expression changed by at least 2-fold, and the results of a Student's t-test for the treated-control comparison returned a p-value <0.001. In these figures, 0_100 means the comparison between 100mM mannitol and control (0mM), 0_250 means the comparison between 250mM mannitol and control (0mM), 0_300 means the comparison between 300mM mannitol and control (0mM).

5A. The tendency of up-regulated and down-regulated gene expression numbers in each comparison.

5B. Venn diagram showing the numbers of up-regulated genes.

5C. Venn diagram showing the down-regulated number of gene.

5D. Venn diagram showing the total numbers of gene expression from this study compared with two previous studies.



$$M = \text{PAS} / \text{gene control plants} = C / (A+B+C+D)$$

$$N = \text{PAS} / \text{gene treated plants} = C' / (A'+B'+C'+D')$$

$$\text{Relative poly(A) usage} = \text{Log}_2 \frac{\text{PAS} / \text{gene treated plants (N)}}{\text{PAS} / \text{gene control plants (M)}}$$

If the relative poly(A) site usage > 0, the more poly(A) usage in treated plants

If the relative poly(A) site usage < 0, the less poly(A) usage in treated plants

Figure 2.6 Overview of poly(A) site usage analysis

Here using the CDS as an example to calculate relative poly(A) site usage. The left part shows the control sample has the two poly(A) sites fallen into different genomic region. The right part shows that the numbers of poly(A) sites are increased up to six under treatment. The total number of poly(A) sites could be same or different in genes. The total number of poly(A) tags are seventeen in control sample, while twenty in the treated sample. Therefore, the relative usage of site C is M in control sample, while N is the relative usage of site C' for the treated sample. The \log_2 efficiency of N/M the relative poly(A) site usage is calculated by counting tag numbers. Values greater than zero means the more poly(A) usage in treated plants, and values less than zero means less poly(A) usage in treated plants.

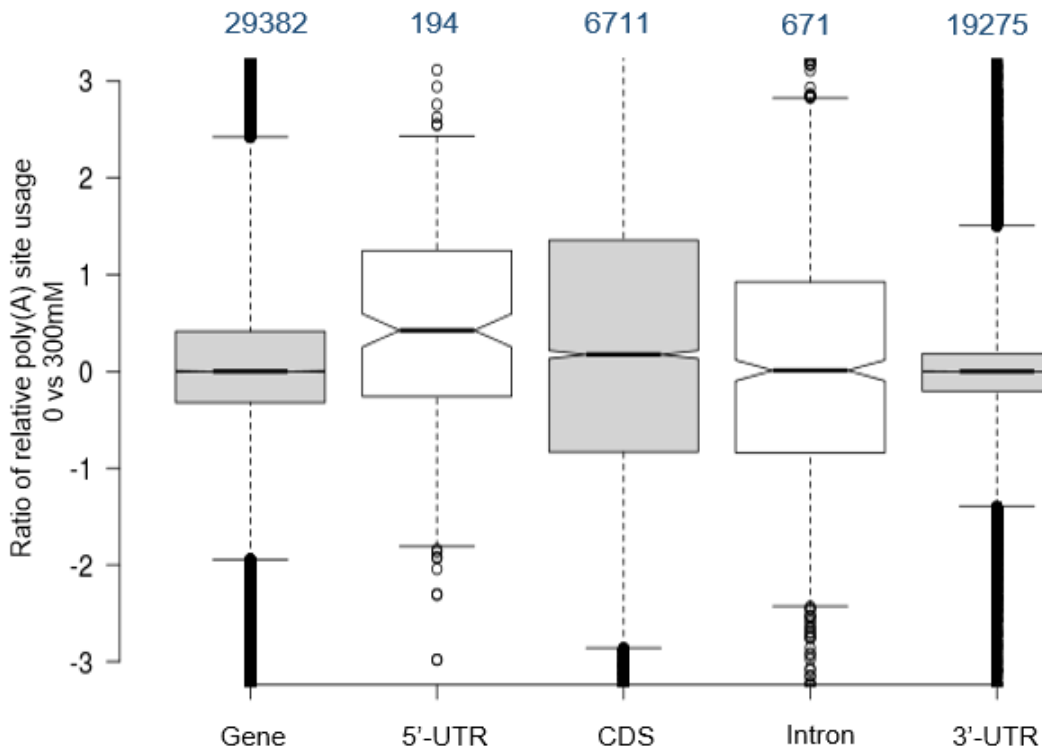


Figure 2.7 Drought induced by 300mM mannitol treatment for two weeks alters non-canonical poly(A) site usage

Boxplot showing changes in overall usage of different classes of poly(A) sites in plants grown on 300 mM mannitol for 2 weeks. The relative contributions that each PAC makes to total poly(A) site usage was evaluated on a gene by gene basis, the ratios of usage in drought and control plants were calculated and \log_2 transformed. “Gene” means the complete collection of PACs. The number above each plot shows the total number of PACs in each class. Non-canonical poly(A) sites include those situated in 5'-UTRs, CDS, and introns.

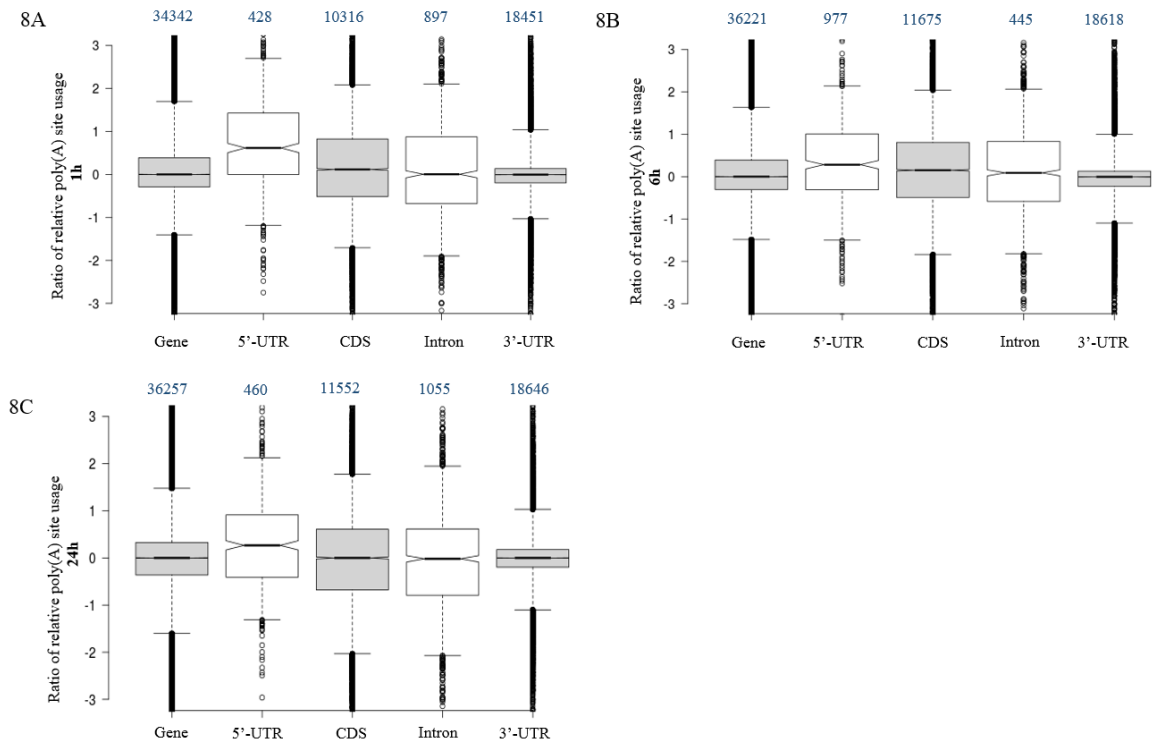
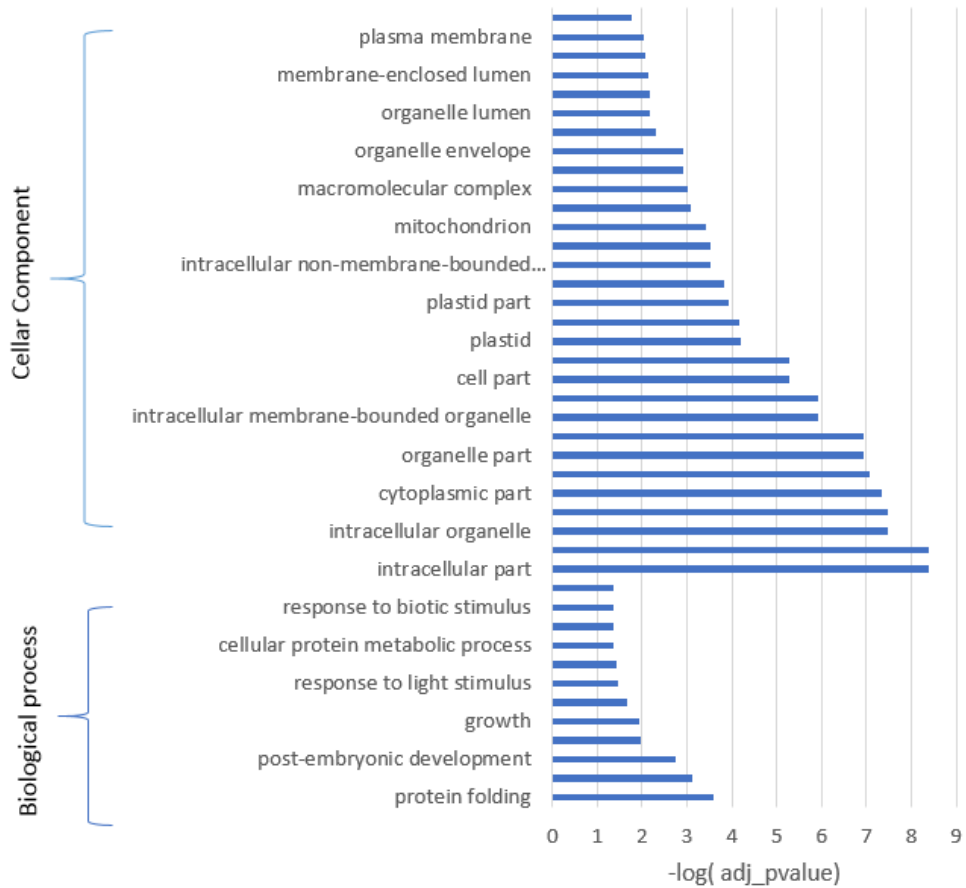


Figure 2.8 Drought induced by 300mM mannitol treatment for different time intervals leads to changes in the non-canonical poly(A) site usage

- A. Changes in poly(A) site usage in plants treated with 300 mM mannitol for 1 hr.
 - B. Changes in poly(A) site usage in plants treated with 300 mM mannitol for 6 hr.
 - C. Changes in poly(A) site usage in plants treated with 300 mM mannitol for 24 hr.
- Calculations and display are as described in Figure. 2.7.

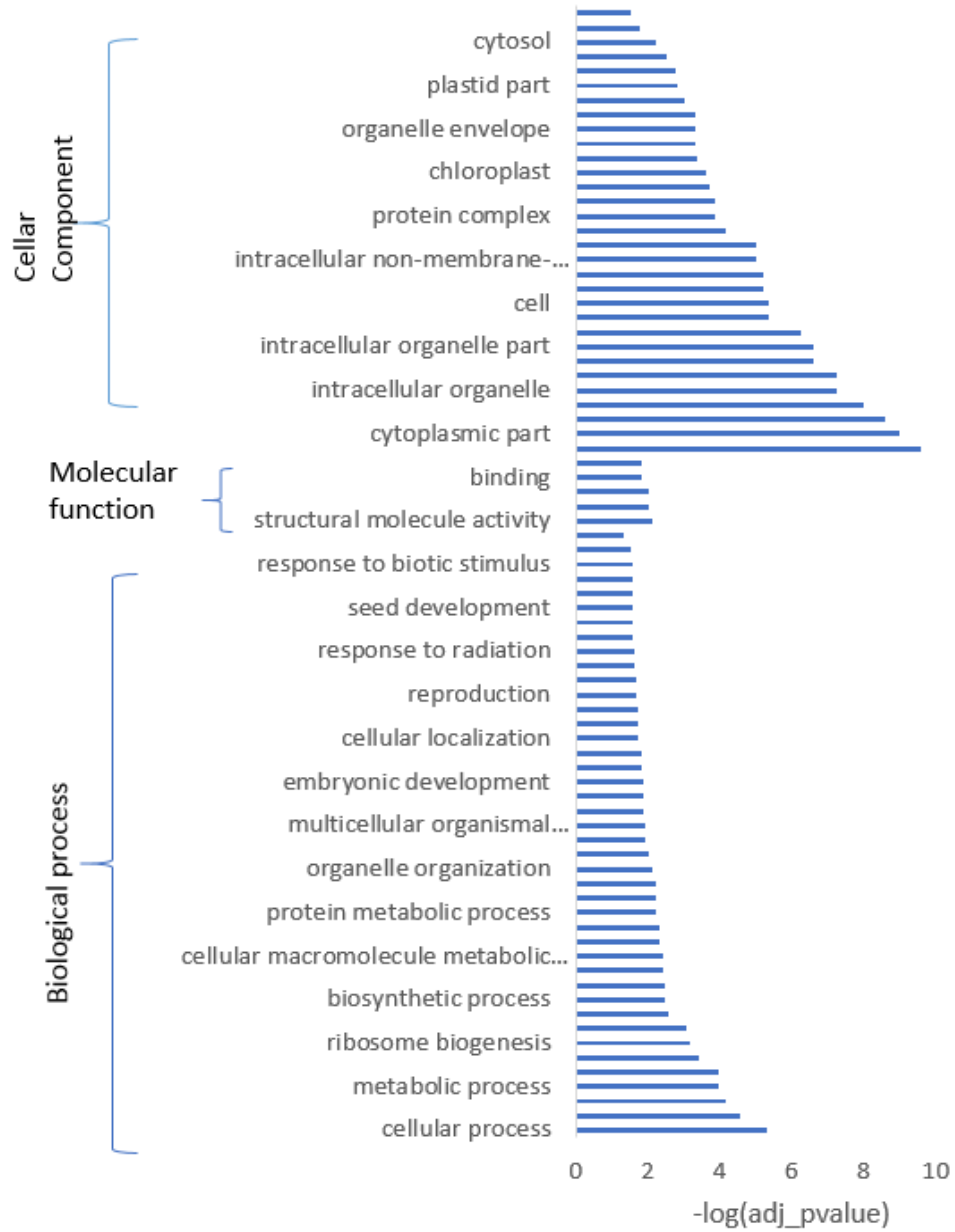
9A

0_100



9B

0_250



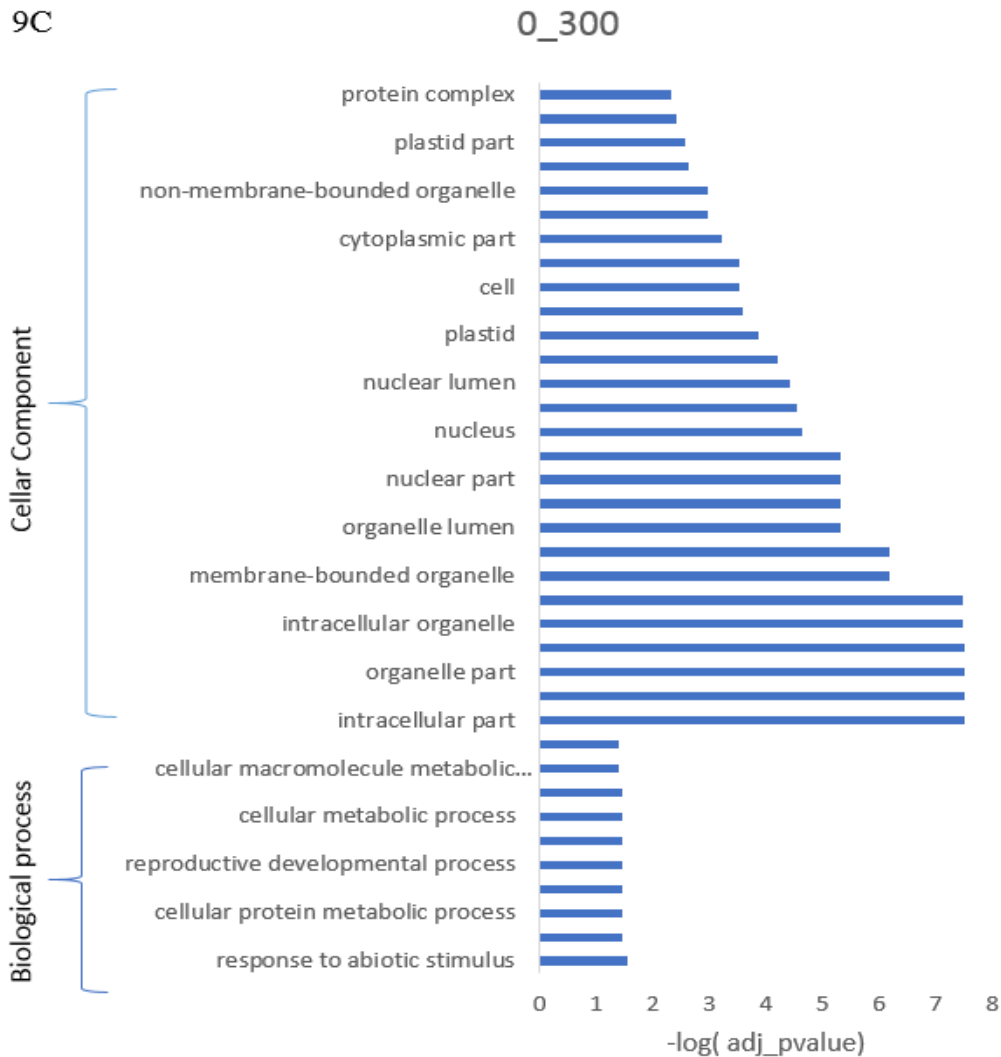


Figure 2.9 Enriched GO categories among the genes displaying APA after drought stress induced by different concentrations of mannitol treatments for two weeks

Gene Ontology term enrichment according to the agriGO, which is a toolkit and database for agriculture community (Website <http://bioinfo.cau.edu.cn/agriGO/>). Parameters are that reference is *Arabidopsis* gene model (TAIR); the statistic model is hypergeometric test; the multi-test adjust method is FDR adjust p-value; the significance is less than 0.05; and the minimum number of mapping entry is 9. The X-axis is minus log FDR adjusted p-value. The Y-axis is the GO terms.

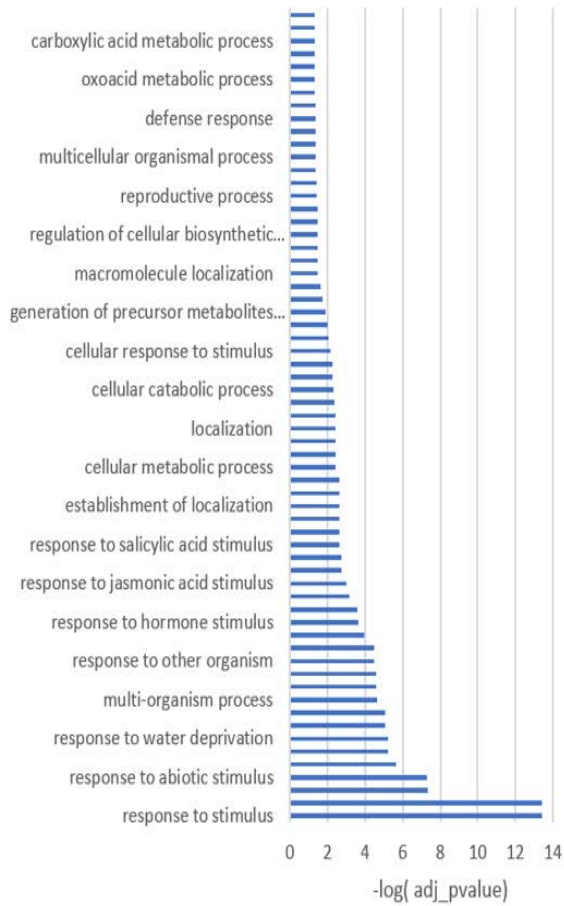
9A. The GO terms from 100mM treatment compared with control.

9B. The GO terms from 250mM treatment compared with control.

9C. The GO terms from 300mM treatment compared with control.

10A

Biological process



10B

Molecular function

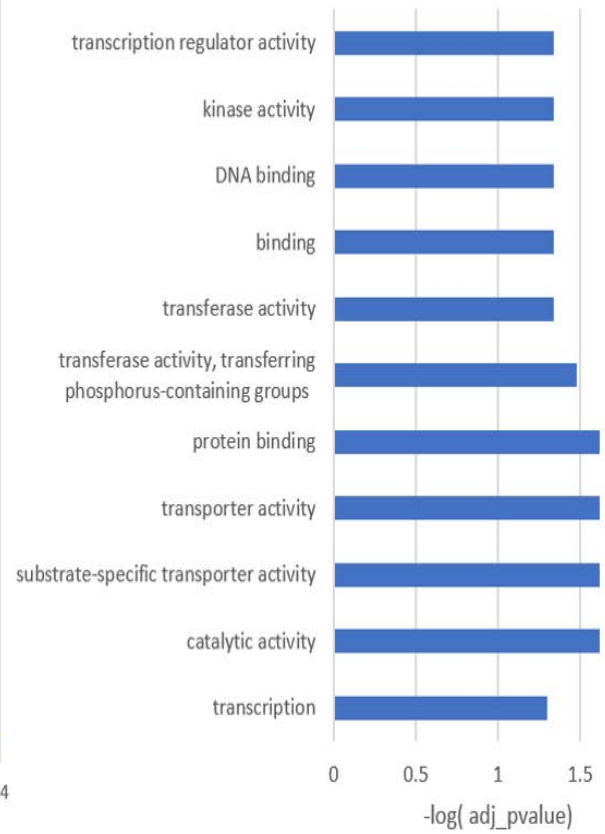


Figure 2.10 Enriched GO categories among the transcript isoforms terminating at poly(A) site located at 5'UTRs

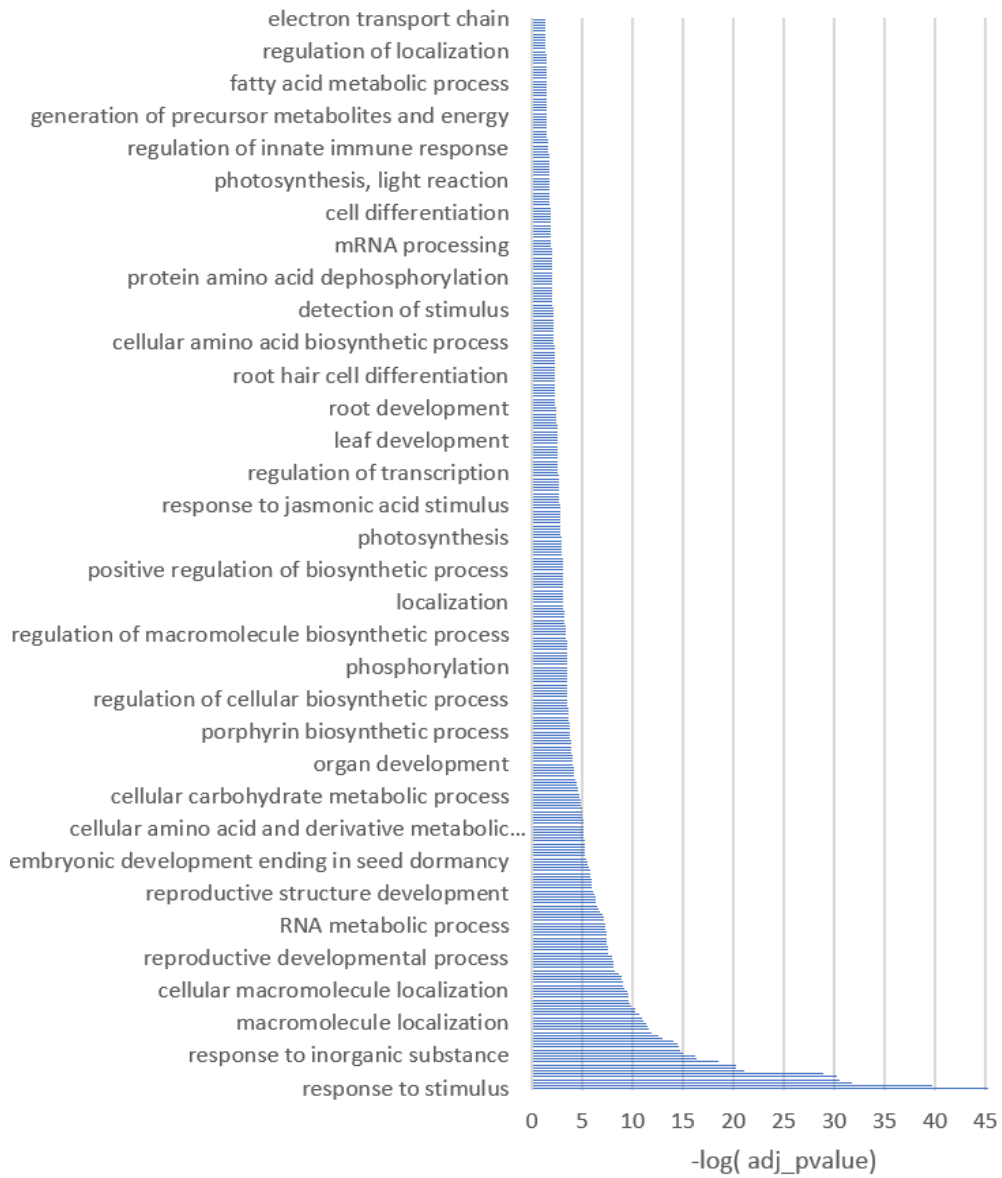
A. The biological process of 1h_300mM treatment in 5'UTR.

B. The molecular function of 1h_300mM treatment in 5'UTR.

GO analyses were conducted as described in the legend for Figure. 2.9.

11A

Biological process



11B

Molecular function

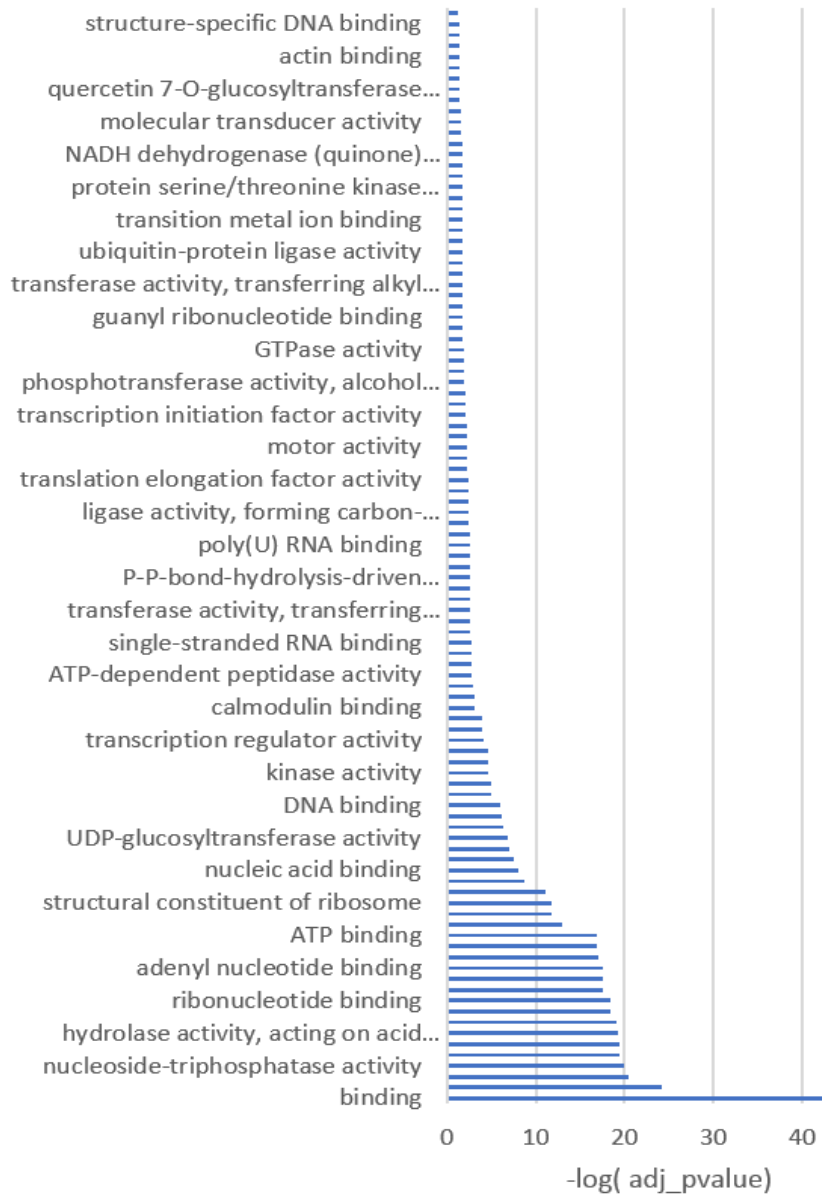


Figure 2.11 Enriched GO categories among the transcript isoforms terminating at poly(A) site located at CDSs

11A. The biological process of 1h_300mM treatment in CDS

11B. The molecular function of 1h_300mM treatment in CDS

Chapter Three: AtCPSF30-mediated alternative polyadenylation controls mRNA stability

3.1 Introduction

The process of polyadenylation is important for the function of eukaryotic mRNAs, being intimately involved in translatability (Sachs and Wahle 1993; Li and Hunt 1997) and stability (Jacobson and Peltz 1996). This is especially true for genes whose mRNAs can be polyadenylated at more than one position within the pre-mRNA (Lutz and Moreira 2011; Xing and Li 2011b), a process termed alternative polyadenylation (APA). APA in the RNA processing step may give rise to mRNAs with different functions and might ultimately change the function of the associated gene product (Hunt 2012).

In plants, there are two general classes of APA sites, canonical and noncanonical (de Lorenzo et al. 2017). Poly(A) sites that occur in the 3'-untranslated regions (3'-UTRs) of pre-mRNAs are called canonical, since they yield mRNAs that are translatable and encode full-length polypeptides (Sun et al. 2012). Noncanonical poly(A) sites reside in the upstream regions, including 5'-untranslated regions (5'-UTRs), protein coding regions and introns. Polyadenylation in the protein coding region of the pre-mRNA would produce transcripts that lack a translation termination codon (resulting in a non-stop mRNA). Non-stop mRNAs are unstable and rapidly turnover through the non-stop decay mechanism (NSD) (Frischmeyer et al. 2002; Vasudevan et al. 2002; Akimitsu et al. 2007). Polyadenylation that occurs within introns can lead to alternative protein products or impact mRNA stability through the nonsense-mediated decay mechanism (NMD) (Hogg and Goff 2010; Hwang and Maquat 2011). Interestingly, alternative polyadenylation also can happen in the 5'UTR (Wu et al. 2011; Thomas et al. 2012; Wu et al. 2014).

Alternative polyadenylation (APA) has the potential to affect gene regulation and function. In *Arabidopsis*, APA may be linked with stress responses through a polyadenylation complex subunit CPSF30 (AtCPSF30) (Hunt 2014). Zhang et al. have described an *Arabidopsis* mutant that can tolerate oxidative stress, called *oxt6*. This mutant was caused by a T-DNA insertion in the gene that encodes AtCPSF30 (Zhang et al. 2008). The AtCPSF30 is impacted by calmodulin and sulfhydryl reagents in vitro (Delaney et al. 2006). The RNA binding activity of AtCPSF30 can be inhibited by calmodulin in a calcium-dependent manner, whereas the third zinc finger contains disulfide bond that can be inhibited by DL-dithiothreitol (Addepalli et al. 2010). Consistent with its function as a polyadenylation factor subunit (Delaney et al. 2006), poly(A) site choice is altered genome-wide in the *oxt6* mutant (Thomas et al. 2012). This study revealed that there are three classes of poly A sites in *Arabidopsis*: sites seen only in the wild type (wt-specific), sites seen only in the *oxt6* mutant (*oxt6*-specific), and sites seen in both genetic backgrounds (common sites). The wt-specific and *oxt6*-specific sites comprise 70% of all sites (Thomas et al. 2012).

While AtCPSF30 controls the usage of a large number of poly(A) sites, it was not clear how AtCPSF30-mediated APA contributes to gene regulation. In this chapter, I hypothesize that alternative polyadenylation through AtCPSF30 controls mRNA stability. Two-week old *Arabidopsis* seedlings were treated with 200 μ M cordycepin for 2h to block nuclear transcription. RNA was isolated and used for PAT-seq libraries preparation and analyzed. The results showed that most mRNA isoforms show similar stability profiles in the wild-type and *oxt6* mutant plants, with the exception of mRNA isoforms with 3' ends that lie within 5'-UTRs. However, *oxt6*-specific mRNA isoforms derived from

polyadenylation within coding region (CDS isoforms) are more stable than wt-specific CDS isoforms. These results implicate AtCPSF30 in the process of non-stop mRNA decay and other aspects of RNA surveillance.

3.2 Results

3.2.1 Overview of experimental approach

To determine the relative stabilities of mRNA isoforms, the method used in similar previous studies (Johnson et al. 2000; Gutierrez et al. 2002) was used. Two-week old seedlings were treated with cordycepin for different times ranging from 15 min to 120 min to block nuclear transcription. Total RNA was isolated from these samples for PAT library preparation, sequencing, and analysis (Figure 3.1). The poly(A) site analysis was conducted using a dedicated computational pipeline (Bell et al. 2016). The rationale for this approach holds that, after treating with cordycepin (a transcription inhibitor), there is no new mRNA produced. Thus, all mRNAs will decay depending on their stabilities. Some of the mRNA isoforms will be less stable than the others, and these differences can be measured by determining the relative levels of the individual isoforms on a gene-by-gene basis. This approach is summarized in Figure 3.2. The relative poly(A) site usage is used to estimate relative mRNA stability. In Figure 3.2, the mRNA isoforms that fall into the CDS are used as an example for the calculation of relative mRNA stability. In the control sample, A, B, C, D represent the numbers of tags that map to the poly(A) sites that fall into 5'UTR, intron, CDS and 3'UTR without treatment, respectively. In the treated sample, A', B', C' and D' represent the numbers of tags that map to the poly(A) sites that fall into 5'UTR, intron, CDS and 3'UTR under treatment, respectively. Therefore, the relative usage of site C is P in the control sample, while Q is the relative usage of site C' for the

treated sample. The change in relative poly(A) site usage is Q/P; this value is \log_2 transformed before plotting results.

3.2.2 Validation of the cordycepin treatment

To validate the effectiveness of the cordycepin treatment, the relative expression of all mRNA isoforms derived from different genes was determined using PATs as described before (Ma et al. 2014b). For this, the number of PATs which were mapped to individual annotated genes were used to calculate relative gene expression levels. Subsequently, relative mRNA stability was inferred (Figure 3.3A) and the results compared with previous stability studies (Figure 3.3B). In my experiment, 1266 unstable transcripts were identified (FDR p -value <0.05 and the absolute fold change more than two) in the wild type cordycepin study. When compared with the list of genes encoding unstable transcripts noted in previous published papers, roughly 83% of the unstable transcripts in my work were also identified in Gutierrez et al. (Gutierrez et al. 2002), 77% of the unstable transcripts were identified in the Kim et al. (Kim et al. 2011), and 70% were identified in the de Lorenzo study (Figure 3.3B) (de Lorenzo et al. 2017). These results demonstrated that the cordycepin treatment was effective.

3.2.3 The global mRNA stabilities in the wild type and *oxt6* mutant

To evaluate the relative stabilities of different classes of mRNA isoforms in the wild-type and *oxt6*-mutant, the relative levels of individual mRNA isoforms were calculated from the ratio of poly(A) site usage. Since the largest number of unstable transcripts could be seen 120 min post-treatment commencement, the relative poly(A) site usage of each mRNA isoform was calculated from comparison between control (C0) and 120 min (C120) cordycepin treatment samples. The ratio of poly(A) site usage for these comparisons was

\log_2 transformed. The values less than zero reflect low stabilities, while values greater than zero indicate greater stabilities. The aggregate values for each class of mRNA isoform (with poly(A) sites that fall in 5'-UTRs, CDS, introns, and 3'-UTRs) are summarized in Figure 3.4. In the wild type, the median values for all sites and sites that fall into the 3'-UTR were similar. Sites that fell into the 5'UTR region had a slighter greater median value, but the value was not significantly different from the median value obtained from 3'UTR sites. However, mRNA isoforms with poly(A) sites that lie within the CDS and within introns had significantly lower values, -0.35 and -0.42 (Figure 3.4A). These results indicate that, in general, mRNA isoforms derived from CDS and intronic polyadenylation are somewhat less stable than other mRNA isoforms. These results corroborate those described in De Lorenzo et al. (de Lorenzo et al. 2017) .

In the *oxl6* mutant, mRNA isoforms with 3' ends within CDS and introns also had significantly lower stabilities than isoforms with 3' ends within 3'-UTRs (Figure 3.4B). Interestingly, mRNA isoforms with 3' ends within 5'UTRs also showed lower stabilities (Figure 3.4B). These results show that the relative stabilities of mRNA isoforms with 3' ends that fall within CDS and introns do not seem to be affected by the absence of AtCPSF30. However, they suggest that AtCPSF30 may be important for the isoforms with 3' ends that fall within 5'-UTRs, since these isoforms were less stable in the mutant than the wild-type.

3.2.4 The stabilities of mRNA isoform defined by their dependences on AtCPSF30

As mentioned above, poly(A) sites in *Arabidopsis* may be defined by their dependence on the presence or absence of AtCPSF30. One class is the wild type specific, which means that the poly(A) sites exclusively exist in wild type plants. Another class is the *oxl6* mutant

specific, which are poly(A) sites only seen in the *oxl6* mutant. The third class (common sites) consists of sites seen in wild type and mutant plants. To explore the stabilities of mRNA isoforms derived from usage of these three classes of poly(A) sites, the ratios of poly(A) site usage were calculated using the same method for the global mRNA stabilities, and the relative stabilities of subsets of each class displayed as shown in Figure 3.5. The stabilities of common mRNA isoforms derived from poly(A) sites within 5'UTR, coding sequence, and introns were all significantly lower than common sites ending within 3' UTRs (Figure 3.5A). Also, the stabilities of wt- and *oxl6*- specific mRNA isoforms with 3' ends within introns were lower than wt- and *oxl6*- specific mRNA isoforms with 3' ends within 3'UTRs (Figures 3.5B and 3.5C). Interestingly, naturally occurring, wt-specific isoforms derived from 5'UTR polyadenylation were relatively stable (Figure 3.5B), while mutant induced, *oxl6*-specific isoforms derived from 5'UTR polyadenylation were relatively unstable (Figure 3.5C). Moreover, naturally occurring, wt-specific isoforms derived from CDS polyadenylation were relatively unstable (Figure 3.5B), while those *oxl6*-specific isoforms, invoked due to the mutation, that are derived from CDS polyadenylation were relatively stable (Figure 3.5C). These results indicate that naturally occurring mRNA isoforms derived from 5'UTR polyadenylation are stabilized by the presence of AtCPSF30, while the stability of another type of mRNA isoform derived from 5'UTR polyadenylation and generated due to the mutation are not dependent on AtCPSF30. However, the stability of mRNA isoforms derived from CDS polyadenylation is dependent on the presence of AtCPSF30, while the stability of mRNA isoforms derived from intronic polyadenylation is not dependent on the presence of AtCPSF30. These results

reveal a very complicated role for AtCPSF30 in the stabilities of mRNA isoforms derived from 5'UTR and CDS.

To further explore the poly(A) site choice associated with mRNA stability, those individual sites whose stabilities were the most significantly (P-value <0.001, determined using DEX-seq package) lower were identified. The list of sites that fall into these parameters and transcripts associated sites are identified. For the wt-specific unstable transcripts, the GO analysis showed that they are highly enriched for response to stress, response to stimulus, response to metal ion, response to inorganic substance, response to water deprivation, response to osmotic stress, response to chemical stimulus, cellular localization, protein localization and transportation, and cellular nitrogen compound metabolic process (Figure 3.6A). Interestingly, as far as molecular function is concerned, the main affected categories were structure molecular activity and structural constituent of ribosome (Figure 3.6A). However, the biological process of GO analysis for *oxl6*-specific transcripts are similar to the wt-specific transcripts, but the affected molecular functions are different for *oxl6*-specific transcripts; for the latter, unstable transcripts were associated with binding activity and RNA binding activities (Figure 3.6B). These results suggest that the unstable transcripts are involved in stress response processes, and that the molecular functions of unstable wt-specific and *oxl6*-specific transcripts are different, suggesting different contributions of these sites to growth and development.

3.2.5 Annotation and functional analysis of mRNA isoforms that AtCPSF30-dependent and -independent

As mentioned above, the stabilities of wt-specific and *oxl6*-specific mRNAs derived from polyadenylation within 5'UTR and CDS regions are different (Figures 3.5B and 3.5C). To

further analyze this, the annotation and function of transcripts derived from polyadenylation at sites that fall into 5'UTRs and CDS were assigned using agriGO. For the GO analysis, the reference is the *Arabidopsis* gene models (TAIR); the statistical model is the hypergeometric test; the multi-test adjust method is FDR adjusted p-value less than 0.05; and the minimum number of mapping entries is greater than or equal to 5. The transcripts that fall into 5'UTRs and CDSs were annotated according to at least one of the three categories: cellular component, biological process or molecular function. The same parameters were used for GO analyses for transcripts that are dependent or independent on AtCPSF30.

For wt-specific poly(A) sites, only 41 poly(A) sites that fall into the 5'UTR were identified; those mapped to 25 transcripts. These 25 transcripts are categorized into 10 functional groups, which belong to the biological process. In this group, transcripts were enriched in cell component categories, including the cytoplasm (GO:0005737), intracellular organelle (GO:0043229) and membrane-bounded organelle (GO:0043227) (Figure 3.7A). For *oxl6*-specific poly(A) sites, 94 poly(A) sites that fall into the 5'UTR were identified; those are associated with to 94 transcripts. These 94 transcripts are categorized into 39 functional groups. For the biological process, the transcripts are mainly enriched in repose to chemical stimulus (GO:0042221), abiotic stimulus (GO:0009628), response to organic substance (GO:0010033), response to water (GO:0009415), response to endogenous stimulus (GO:0009719). Also, protein transport (GO:0015031), protein localization (GO:0008104), cellular localization (GO:0051641) and macromolecular localization (GO:0033036) are included. Molecular function categorizations for this group included transporter activity (GO:0005215), substrate-specific transporter activity (GO:0022892), substrate-specific

transmembrane transporter activity (GO:0022891), transmembrane transporter activity (GO:0022857), and oxidoreductase activity (GO:0016491) (Figure 3.7B).

For wt-specific poly(A) sites, 428 differentially-utilized poly(A) sites that fall in to CDS corresponding to 322 transcripts could be identified. These transcripts are categorized into 142 functional groups. Biological process characterization included response stimulus (GO:0050896), response to temperature stimulus (GO:0009266), response to stress (GO:0006950), response to cold (GO:0009409), response to water deprivation (GO:0009414), response to biotic stimulus (GO:0009607), response to cadmium ion (GO:0046686), response to osmotic stress (GO:0006970), response to chemical stimulus (GO:0042221), response to fungus (GO:0009620), response to inorganic substance (GO:0010035), response to metal ion (GO:0010038), response to salt stress (GO:0009651), response to organic substance (GO:0010033), carboxylic acid metabolic process (GO:0019752). Additionally, the cellular ketone metabolic process (GO:0042180), organic acid metabolic process (GO:0006082), post- transcriptional protein modification (GO:0043687), gene expression (GO:0010467), translation (GO:0006412), defense response (GO:0006952), primary metabolic process (GO:0044238), and cellular macromolecular metabolic process (GO:0044260) are all included (Figure 3.8A). For the molecular function, there is significant enrichment for binding activity, including protein binding (GO:0005515) nucleotide binding (GO:0000166), translation factor activity, nucleic acid binding (GO:0008135), ATP binding (GO:0005524), purine ribonucleotide binding (GO:0032555), ribonucleotide binding (GO:0032553) purine nucleotide binding (GO:0017076), adenylyl nucleotide binding (GO:0030554), nucleic acid binding (GO:0003676), calmodulin binding (GO:0005516). Other enriched categories included

transferase activity (GO:0016740), translation initiation factor activity (GO:0003743), transmembrane receptor activity (GO:0004888) (Figure 3.8B). For *oxl6*-specific poly(A) sites, 474 sites that fall into the CDS were identified; those are associated with 334 transcripts. These transcripts can be separated into 91 functional. In the biological process, the set of transcripts is highly enriched for genes associated with response to abiotic stimulus (GO:0009628), response to stimulus (GO:0050896) response to water (GO:0009415)/water deprivation (GO:0009414), post-embryonic development (GO:0009791), photosynthesis (GO:0015979), cation transport (GO:0006812), response to osmotic stress (GO:0006970), ion transport (GO:0006811), reproductive developmental process (GO:0003006), macromolecular localization (GO:0033036), intracellular transport (GO:0046907), embryonic development ending in seed dormancy (GO:0009793) and fruit development (GO:0010154) (Figure 3.9A). Molecular function characterizations included structural molecule activity (GO:0005198), structural constituent of ribosome (GO:0003735), hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides (GO:0016818), hydrolase activity, acting on acid anhydrides (GO:0016817), pyrophosphatase activity (GO:0016462), nucleoside-triphosphatase activity (GO:0017111), heat shock protein binding (GO:0031072) activity (Figure 3.9B).

Taken together, for the mRNAs derived from polyadenylation within the 5'UTR, there are fewer transcripts, and thus it is difficult to tell the difference of annotation between wt-specific and *oxl6*-specific. However, for the mRNA isoforms derived from polyadenylation within CDS, the annotation and function of mRNA isoform between wt-specific and *oxl6*-specific have similarities in the biological processes that are more related with stress responses, such as response stimulus, response to osmotic stress and response to water

deprivation. But differences exist in the molecular functions; the transcripts of wt-specific are mainly for binding activity, translation factor activity while the *oxl6*-specific are structural molecular activity, hydrolase activity, pyrophosphatase activity, and heat shock protein binding activity. This indicates that the molecular function of transcripts might be more dependent on AtCPSF30, while the transcripts for biological process might be partially dependent on AtCPSF30.

3.3 Discussion

A majority of eukaryotic genes have multiple polyadenylation sites (Di Giammartino et al. 2011; Lutz and Moreira 2011; Shi 2012). In previous studies, the analysis of genome-wide data in plants showed that alternative polyadenylation (APA) is pervasive in plants (Wu et al. 2011; Wu et al. 2014; Fu et al. 2016). Poly(A) sites may be classified according to their genomic location (5'-UTR, CDS, intron, or 3'-UTR). The products of the proximal poly(A) site choice give rise to noncanonical mRNA isoforms, mRNAs that do not encode full-length functional polypeptides (de Lorenzo et al. 2017). In this study, the properties of noncanonical mRNA isoforms in wild-type *Arabidopsis* and the *oxl6* mutant were studied, to better understand AtCPSF30-dependent and -independent polyadenylation.

3.3.1 The stability of noncanonical mRNA isoforms derived from polyadenylation with introns

One type of noncanonical mRNA isoforms is mRNAs with 3'ends that lie within introns. In most cases, these sites are situated between coding exons, giving rise to isoforms encoding different proteins (Mainul et al. 2012). Many human genes encode transcripts derived from combinations of alternative splicing and alternative polyadenylation. The classic example is that of the heavy chain gene of immunoglobulin (IgM); this gene has a

poly(A) site within an intron and gives rise to different mRNAs during different stages of B cell development (Edwards-Gilbert and Milcarek 1995; Takagaki et al. 1996; Bruce et al. 2003; Tian et al. 2007). Another study showed that the calcitonin/calcitonin gene-related peptide gene (*CALCA*) is regulated by the splicing factor SRp20 through the usage of an intronic poly(A) site to make mRNA variants in a tissue-specific manner (Lou et al. 1998; Tian et al. 2007). In plants, one well-studied example of alternative polyadenylation involves the regulatory factor FY, which is a core polyadenylation complex subunit involved in the network of genes that control flowering time in *Arabidopsis*. FY interacts with an RNA binding protein (FCA) leading to polyadenylation within an intron in transcripts from the FCA gene (Simpson et al. 2003). In the case of FCA, usage of the intronic poly(A) site results in the production of truncated polypeptides.

When transcripts contain premature termination codons, they are degraded through nonsense mediated decay (NMD) to prevent the accumulation of the potentially harmful proteins (Stephanie and Allan 2012). Two eukaryotic NMD evolutionary models have been proposed. One is the evolutionarily ancient form, and is based on the recognition of long 3'UTRs. The other is intron-based and involves the so-called exon junction complex. The latter mechanism has evolved in vertebrates to efficiently reduce the accumulation of abnormal products of alternative splicing (Durand and Lykke-Andersen 2011). In plants, 3'-UTR located introns can trigger nonsense mediated decay (NMD) (Kertész et al. 2006; Schwartz et al. 2006; Hori and Watanabe 2007; Wu et al. 2007), demonstrative of the existence of an analogous mechanism.

In this study, noncanonical mRNA isoforms within 3'ends that lie within introns are all unstable in both the wild type and *oxf6* mutant (Figure 3.3). Moreover, CPSF30-dependent

and independent intronic isoforms are also inherently unstable (Figure 3.4). Introns in *Arabidopsis* that contains poly(A) sites are shorter than the typical introns whose average length around 270 nt (Wu et al. 2011), but they are not long enough that polyadenylation within them may trigger NMD (Hogg and Goff 2010). Specifically, 3'UTRs shorter than 350 nt do not trigger NMD in *Arabidopsis*, while the average 3'-UTR length is 237 nt in the latest version of the TAIR10 *Arabidopsis* database (Kalyna et al. 2012). Thus, mRNA isoforms derived from intronic polyadenylation should not be substrates for NMD. The mechanism that is responsible for the lower stabilities of these mRNA isoforms is not clear. This might reflect an interplay between splicing, mRNA turnover and polyadenylation. Regardless, the results presented in this chapter show that AtCPSF30 does not control the stabilities of mRNA isoforms with 3'ends that lie within introns.

3.3.2 The stability of noncanonical mRNA isoforms in CDS

Another type of noncanonical mRNA isoform is mRNAs with 3'ends that fall within protein coding regions. These isoforms almost always lack translation termination codons and are thus non-stop RNAs. Non-stop mRNAs are rapidly degraded through the non-stop decay pathway (NSD) (Frischmeyer et al. 2002; van Hoof et al. 2002). In the non-stop pathway, ribosomes stall at the 3' ends of the mRNA, a process that results in the recruitment of NSD cofactors and the exosome complex to facilitate the degradation of the aberrant mRNA (Klauer and van Hoof 2012). This pathway was identified as a cytoplasmic pathway in both yeast and mammals (Frischmeyer et al. 2002; van Hoof et al. 2002), and is closely associated with the ribosome. During the translation of nonstop mRNAs, the ribosome proceeds through the end of the poly(A) tail, displacing the poly(A)-binding protein (PABP) and stalling at the 3' end of the mRNA. The stalled ribosome at the 3' end

of a transcript initiates non-stop decay through the binding of the C-terminus of Ski7 to the empty A site of the ribosome, thereby releasing the ribosome. Then the exosome is recruited by Ski7 and, along with the SKI complex, deadenylates and rapidly degrades the transcript (Frischmeyer et al. 2002; van Hoof et al. 2002). The 5' to 3' mRNA decay pathway can also mediate NSD without Ski7 in *S. cerevisiae* (Inada and Aiba 2005). This is because the removal of PABP is known to render transcripts susceptible to decapping and subsequent degradation by XRN1.

For the global analysis of poly(A) sites, the results showed that mRNA transcripts derived from polyadenylation in the coding regions are all unstable in the wild type and the *oxt6* mutant (Figures 3.4A and 3.4B). However, the median value of relative poly(A) site stability for these sites is -0.35 in wild type (Figure 3.4A), and -0.13 in the *oxt6* mutant (Figure 3.4B). This result suggests that the stabilities of these mRNA isoforms in the wild type is less than in the *oxt6* mutant, suggesting AtCPSF30 can affect the stabilities of these transcripts. In addition, *oxt6*-specific nonstop transcripts are more stable than wt-specific nonstop mRNAs as well as nonstop RNAs whose production is not affected by the presence or absence of AtCPSF30 (common sites; compare “CDS” samples in Figures 3.5A, 3.5B, and 3.5C). These results demonstrate that the present of AtCPSF30 can impact the stabilities of nonstop transcripts.

In Chapter 4 of this thesis, screening for proteins that interact with AtCPSF30 is described. Among the findings in this chapter is that AtCPSF30 interacts with an *Arabidopsis* 60S ribosomal protein (RPL35). In humans, one study showed that the expansion segment of RPL35 plays an important role in nuclear entry, translation activity, and endoplasmic reticulum docking (Chen et al. 2008). This raises the possibility that AtCPSF30 might

directly interact with ribosome proteins to decrease the stabilities of nonstop mRNAs. Taken together, these results indicate that the presence of AtCPSF30 might help lower the stability of transcripts derived from CDS polyadenylation.

3.3.3 The stability of noncanonical mRNA isoforms in 5'UTR

Unlike mRNA isoforms derived from polyadenylation within introns and protein-coding regions, the stabilities of mRNA isoforms with 3' ends that lie within 5'UTRs are different in the wild type and *oxl6* mutant plants (Figure 3.4 and 3.5); in the wild type, these isoforms are as stable as canonical mRNA isoforms in the 3'UTR (Figure 3.4A), while in the mutant they are less stable (Figure 3.4B). Also, common and *oxl6*-specific mRNA isoforms with 3' ends that lie within 5'UTRs seem to be relatively unstable (Figures 3.5A and 3.5C), while wt-specific mRNA isoforms with 3' ends that lie within 5'UTRs are as stable as conical 3'UTR mRNA isoforms (Figure 3.5B). These results indicate that AtCPSF30 impacts the stabilities of mRNA isoforms derived from polyadenylation in the 5'UTR.

Usually, poly(A) signals are expected at the 3'ends of genes. However, previous studies have shown that strong poly(A) signals are also found in the 5'UTRs of some genes in *Drosophila melanogaster*, and both these novel poly(A) signals and standard poly(A) signals become functionally silent when they are positioned close to transcription start sites in either *Drosophila* or human cells (Guo et al. 2011). The polymerase II (Pol II) transcription complex thus plays an important role to determine whether a putative poly(A) signal is recognized as functional. This mechanism might prevent cryptic poly(A) signals from causing premature transcription termination (Guo et al. 2011). The fact that 5'UTR site usage occurs in the *oxl6* mutant means that AtCPSF30 probably does not control the

premature termination decision. However, it may control the stabilities of these transcript isoforms.

There is a different mechanism of polyadenylation for the products of premature transcription termination so that they can be degraded by RNA exosome (Preker et al. 2008; Fox and Mosley 2016). In *Schizosaccharomyces pombe*, the RNA exosome can target the transcription machinery by terminating transcription events, which is accompanied with paused and backtracked Pol II. The backtracked Pol II can provide a free mRNA 3' end for the core exosome to lead to transcription termination with concomitant degradation of the associated transcript (Jean-François et al. 2014). Such a mechanism may explain the instability of common and *oxt6*-specific 5'-UTR transcript isoforms. Also, the results presented in this chapter raise the possibility that CPSF30 might interact with the Pol II complex to enhance the stabilities of the transcripts derived from polyadenylation within the 5'-UTR.

3.3.4 A model to describe stability of nonstop transcripts

As mentioned above, the non-stop RNAs can be degraded by the non-stop pathway (Frischmeyer et al. 2002; van Hoof et al. 2002). In this pathway, ribosomes stalling at the 3' end can recruit NSD cofactors and the exosome complex to facilitate the aberrant mRNA degradation (Klauer and van Hoof 2012). There are two possible ways for mRNA degradation, Ski7-dependent and -independent. AtCPSF30 might be involved in these two possible ways to mediate the stabilities of these transcript isoforms. A model thereby is proposed in Figure 3.10.

In this model, AtCPSF30 might act with Ski7 to recruit the exosome to degrade the mRNA isoforms. During the translation of nonstop mRNAs, the ribosome proceeds through the

end of the poly(A) tail, displacing the poly(A)-binding protein (PABP) and stalling at the 3' end of the mRNA. The stalled ribosome at the 3' end of a transcript initiates non-stop decay through the binding of the C-terminus of Ski7 to the empty A site of the ribosome, thereby releasing the ribosome. Then the exosome is recruited by Ski7 and AtCPSF30, along with the SKI complex, deadenylates and rapidly degrades the transcripts. Therefore, the mRNA isoforms derived from polyadenylation in the CDS are unstable in the wild type and wt-specific. Without AtCPSF30, Ski7 may be unable to recruit the exosome, so the mRNA isoforms derived from polyadenylation in the CDS would be more stable.

3.3.5 A model to describe the 5'-UTR transcript isoform stability

The properties of mRNA isoforms derived from polyadenylation in the 5'UTR region are more complicated, since the wild type and wt-specific 5'UTR isoforms are as stable as the canonical mRNA isoforms. Therefore, two models are proposed.

In the wild type, AtCPSF30 can form a polyadenylation complex with CPSF160, CPSF100, CPSF73 and Wdr33 to perform recognition and cleavage (Chan et al. 2014; Schönemann et al. 2014), followed by the poly(A) tail addition (Figure 3.11A). However, a complex that contains only CPSF160, CPSF100, CPSF73 and Wdr33 also may perform the polyadenylation function to generate the mRNA isoforms (Figure 3.11B), since polyadenylation occurs in the *oxl6* mutant. However, the fact is that *oxl6*-specific transcripts do not have an AAUAAA-like motif (Figure 3.11C), may be the presence or absence of this motif in the 5'-UTR is involved in stabilization, perhaps through interactions with FY or other proteins.

3.4 Materials and method

3.4.1 Plant material and treatment

Arabidopsis thaliana wild type Col-0 and mutant *oxl6* plants were cultured vertically on solid MS media (1x Murashige and Skoog salts, MES 0.5g/L pH = 5.7, and 1% (w/v) sucrose). The surface of the seeds was sterilized before plating and put at 4°C for two days. Plants were then grown under long day conditions (16h light and 8h dark) for two weeks at 23 °C. In all experiments, three biological replicates were prepared.

For the cordycepin treatment, whole seedlings were transferred to a Petri dish containing incubation buffer (15 mM sucrose, 1 mM KCl/ 1 mM Pipes/ 1 mM sodium citrate, pH 6.5) and incubated for 30 min at room temperature before addition of cordycepin (200µM). Vacuum was applied for 30 seconds, and plants then incubated for the indicated periods of time. Controls were treated identically, but cordycepin was not added. Plants were harvested at different time intervals (15 min, 30 min, 60 min, and 120 min) and were immediately frozen in liquid nitrogen.

3.4.2 Poly(A) tag library preparation and sequencing

Total RNA was isolated using Trizol reagent and purified using RNAeasy columns (Qiagen, Hilden, Germany). A Nanodrop spectrophotometer was used for total RNA quantitation. Poly(A) tag libraries were generated from 1µg of total RNA using method B1 from Ma et al. (Ma et al. 2014a). These libraries were sequenced on the Illumina high-throughput sequencing platform. Three independent biological replicates were prepared for each treatment or condition.

3.4.3 Poly(A) tag and gene expression analysis

The sequencing data were analyzed using the pipeline as described previously (Bell et al. 2016). The raw data were demultiplexed and trimmed to remove the oligo-dT and sequencing adaptors using CLC Genomic Workbench. The trimmed tags were mapped to

the *Arabidopsis* genome (TAIR10) to get the bam file. The bam files were converted to bed files with the BEDtools suite (Quinlan and Hall 2010; Quinlan 2014) to obtain lists of individual poly(A) sites (PAS) and poly(A) site clusters (PACs) (see Fig 3.4 and Bell et al.) (Bell et al. 2016). A minimum of ten individual PATs for PASs and PACs are kept for subsequent analysis.

To determine gene expression, the PAT frequency of each gene was determined by a special file (TAIR10genes120.gff) that has only *Arabidopsis* genes. The data were then imported to CLC using the empirical analysis of DGE to compare the control and treatment. The genes with a total filter cut-off of 2-fold change and p-value <0.001 were selected as statistically significant.

To determine the PAT frequency for each PAC, the list of PACs is used as annotated master file to tell the numbers of tags in each individual sample that map to the PAC. Thus, the numbers of tags in each individual sample are compared with the number of tags in the *Arabidopsis* genome to get a proportion for poly(A) site usage analysis.

To calculate the statistically significant differences in poly(A) site usage in gene by gene analysis using DEX-seq package in R, and a *p-value*<0.01 is considered as statistically significant.

3.4.4 GO analysis

Gene Ontology term enrichment was performed using agriGO, which is a toolkit and database for the agriculture community (Website<http://bioinfo.cau.edu.cn/agriGO/>). Parameters used were: reference is *Arabidopsis* gene model (TAIR); the statistic model is hypergeometric test; the multi-test adjust method is FDR adjust p-value; the significance is less than 0.05; and the minimum number of mapping entries is 5. In the GO figures, the

X-axis is the negative log FDR adjust p-value and the Y-axis is the GO terms.

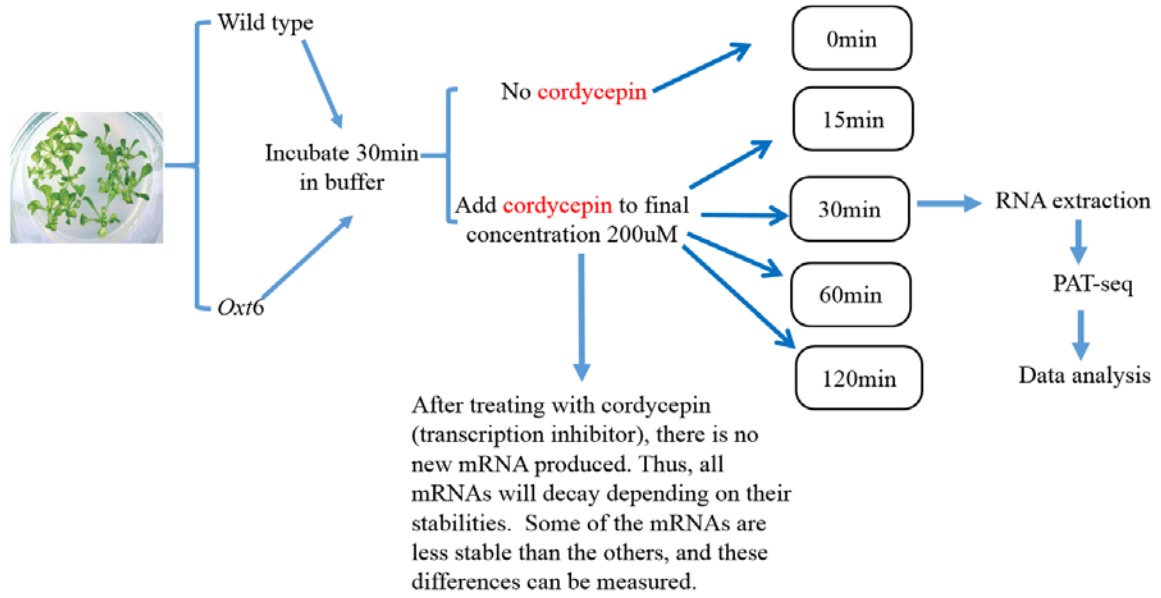
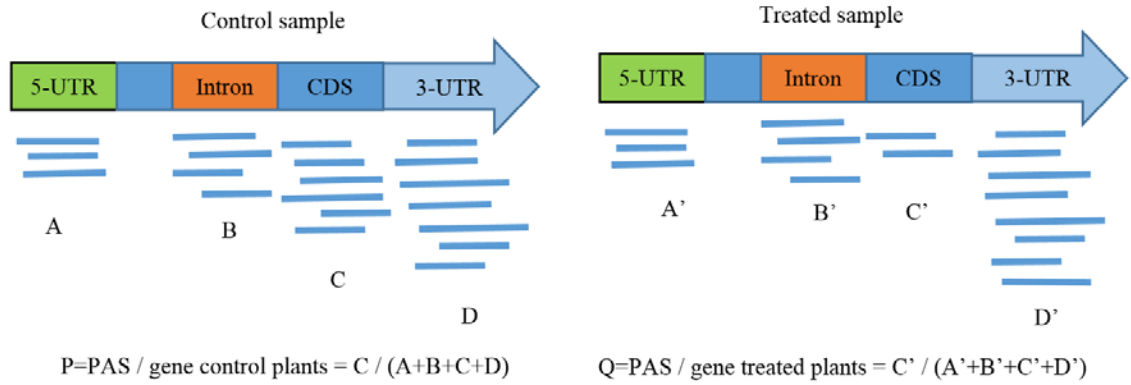


Figure 3.1 Overview of experimental approach to study AtCPSF30-mediated APA controls mRNA stability

Two-week old seedlings were incubated in buffer for 30min, and then cordycepin added to 200uM final concentration so no new mRNA is produced. Thus, all mRNAs will decay depending on their stabilities. Some of the mRNAs are less stable than the others, and these differences can be measured. The seedlings after different times were harvested for RNA extraction, and thereafter PATs libraries were prepared and sequenced. Next, the PATs libraries data was analyzed to evaluate mRNA stability.



$$\text{Relative poly(A) usage} = \text{Log}_2 \frac{\text{PAS / gene treated plants (Q)}}{\text{PAS / gene control plants (P)}}$$

If all mRNA isoforms have the same stabilities, relative usage will be 0

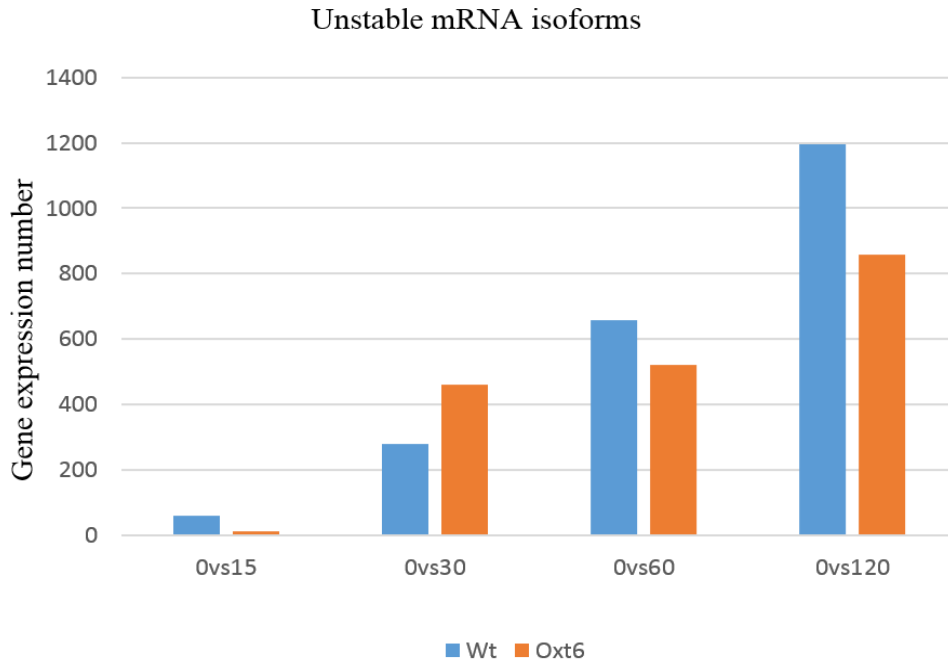
If one isoform is less stable than others, relative poly(A) site usage will be <0

If one isoform is more stable than others, relative poly(A) site usage will be > 0

Figure 3.2 Overview of the method to calculate mRNA stability

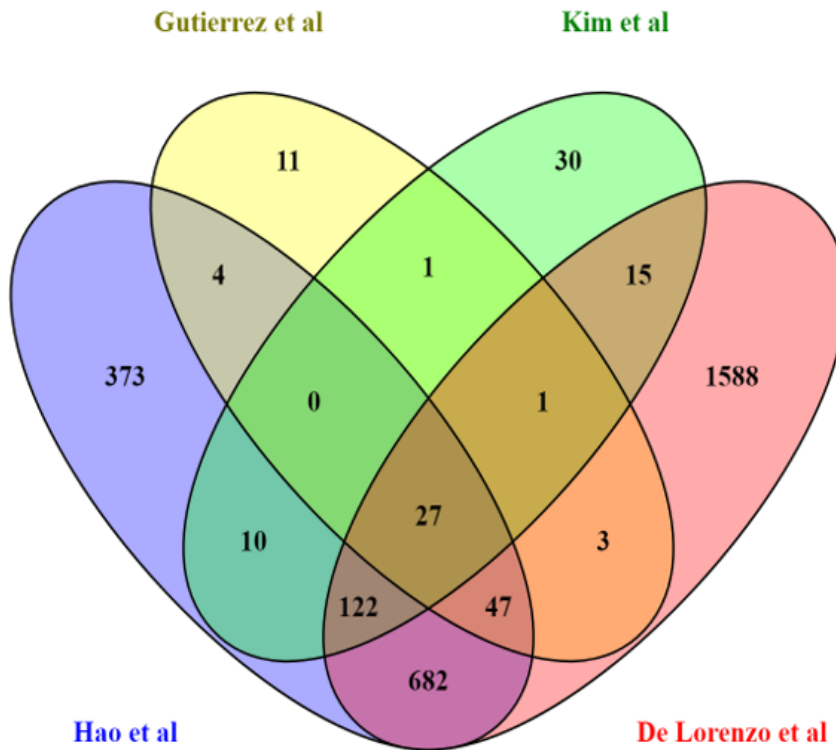
In the control sample, A, B, C, D shows the poly(A) sites that fall into 5'UTR, intron, CDS and 3'UTR without treatment, respectively. While in the treated sample, A', B', C' and D' means the proportion of poly(A) site choice that fall into 5'UTR, intron, CDS and 3'UTR under treatment, respectively. Here the poly(A) site choice in CDS region as an example, the proportion of poly(A) site choice that fall into CDS is M in control sample, whole N is the percentage of poly(A) site choice that fall into CDS for treated sample. The relative poly(A) site usage is \log_2 transformed for P/Q. The relative mRNA stability is determined by the value of relative poly(A) site usage compared with zero.

3A



3A. The number of unstable mRNA isoforms are increased after longer time cordycepin treatment. The X-axis shows the comparisons between different time treatments and control. 0vs15 means the comparison between 15mins and control (0min), 0vs30 means the comparison between 30mins and control (0min), 0vs60 means the comparison between 60mins and control (0min), 0vs120 means the comparison between 120mins and control (0min). The Y-axis shows the gene expression number of unstable mRNA isoforms.

3B



3B. Unstable transcripts in this study to compare with previous studies

The list of genes encoding unstable transcripts noted in this study are compared with previous published papers, roughly 83% of the unstable transcripts in my work were also identified in Gutierrez et al. (Gutierrez et al. 2002), 77% of the unstable transcripts were identified in Kim et al. study (Kim et al. 2011), and 70% were identified in (de Lorenzo et al. 2017).

Figure 3.3 Cordycepin incited significant down-regulation gene expression

To validate the effectiveness of the cordycepin treatment, the relative expression of all mRNA isoforms derived from different genes was determined using PATs as described before (Ma et al. 2014b). For this, the number of PATs which were mapped to individual annotated genes were used to calculate relative gene expression levels. 1266 unstable transcripts were identified (FDR p-value < 0.05 and the absolute fold change more than two) in the wild type cordycepin study.

3A. The number of unstable mRNA isoforms are increased after longer time cordycepin treatment.

3B. Unstable transcripts in this study compared with previous studies

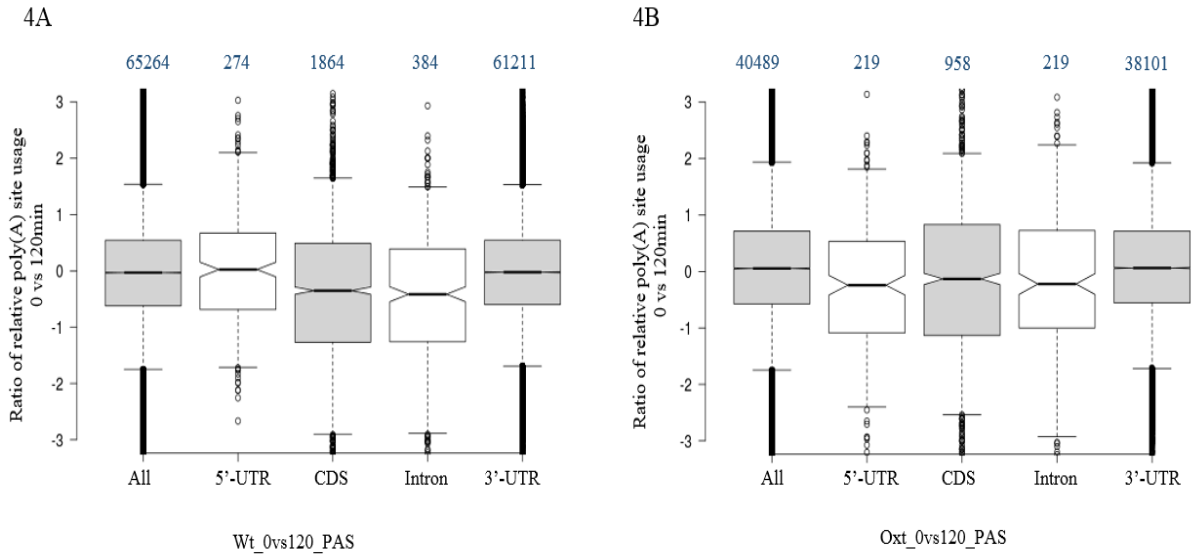


Figure 3.4 The stabilities of non-canonical mRNA isoforms incites global changes

Boxplot showing changes in different classes of poly(A) sites usage after 120min of treatment with cordycepin. The relative contribution that each PAS makes to total poly(A) usage was determined as in Figure 3.2. Total number of PASs in each class show above the boxplot.

- A. The global mRNA stability in wild type
- B. The global mRNA stability in *oxt6* mutant

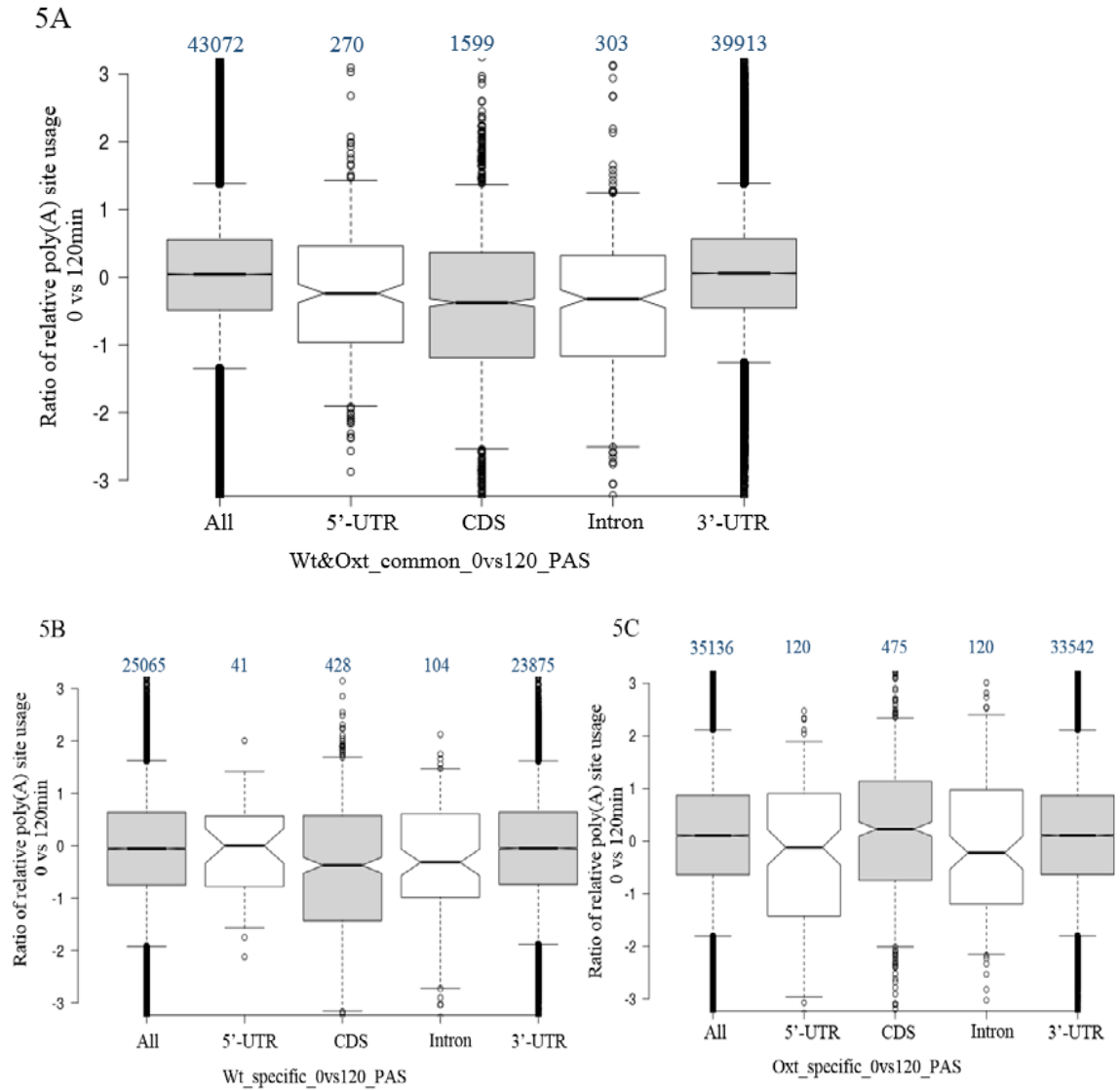


Figure 3.5. Stabilities of noncanonical mRNA isoforms from three classes of poly(A) sites changes

Boxplot showing changes in different classes of poly(A) sites usage after 120 min of treatment with cordycepin. The relative contribution that each PAS makes to total poly(A) usage was determined as in Figure 3.2. Total number of PASs in each class are shown above the boxplot.

- A. The stabilities of common mRNA isoforms.
- B. The stabilities of wt-specific mRNA isoforms.
- C. The stabilities of *oxt6*-specific mRNA isoforms.

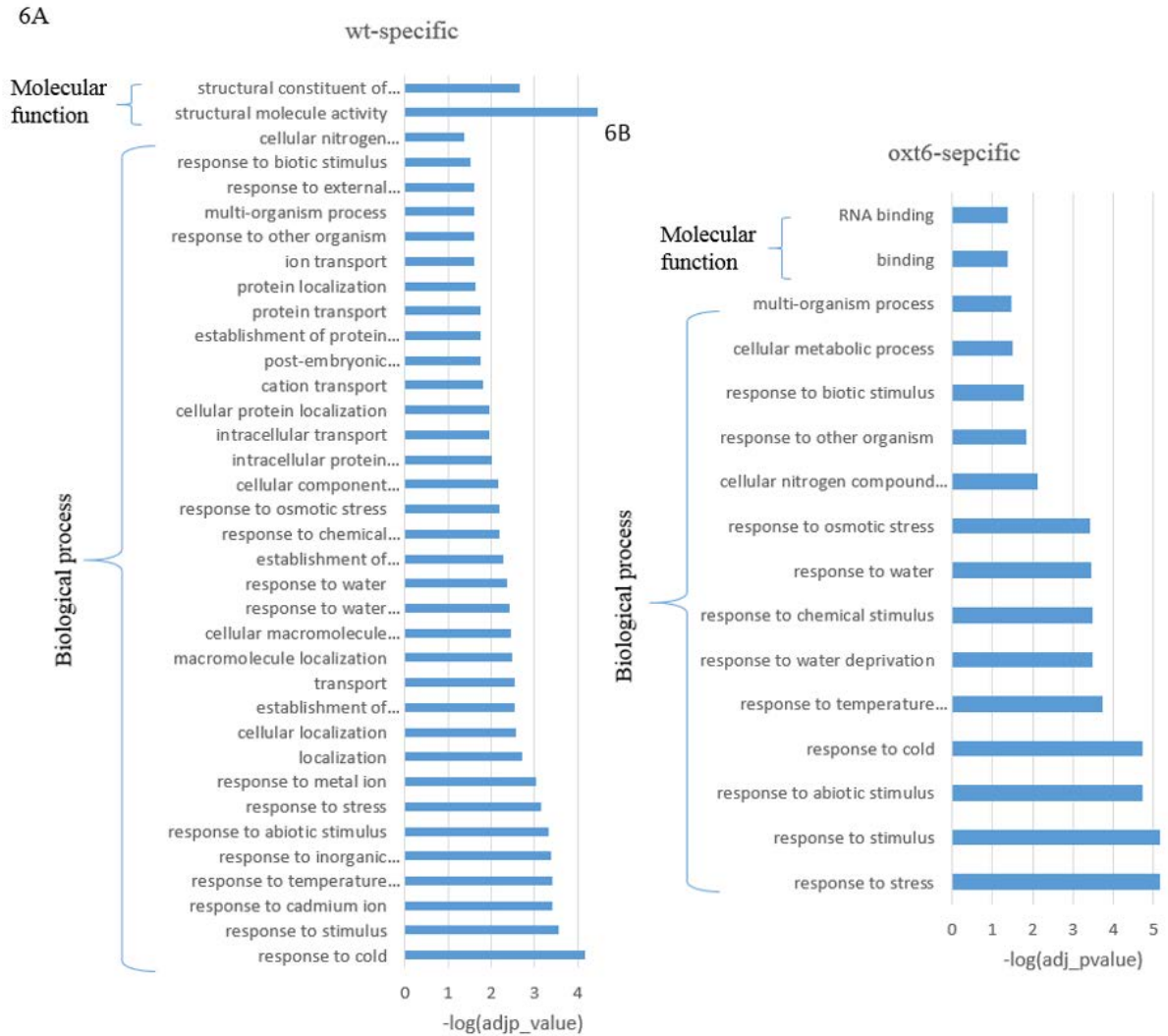


Figure 3.6 Enriched GO categories among the genes displaying APA after cordycepin treatment

Gene Ontology term enrichment according to the agriGo, which is a toolkit and database for agriculture community (Website<http://bioinfo.cau.edu.cn/agriGO/>). Parameters are that reference is *Arabidopsis* genemodel (TAIR); the statistic model is hypergeometric test; the multi-test adjust method is FDR adjust p-value; the significance is less than 0.05; and the minimum number of mapping entry is 9. The X-axis is minus log FDR adjusted p-value. The Y-axis is the GO terms.

- A. The GO analysis for genes encoding unstable wt-specific sites.
- B. The GO analysis for genes encoding unstable *oxt6*-specific sites.

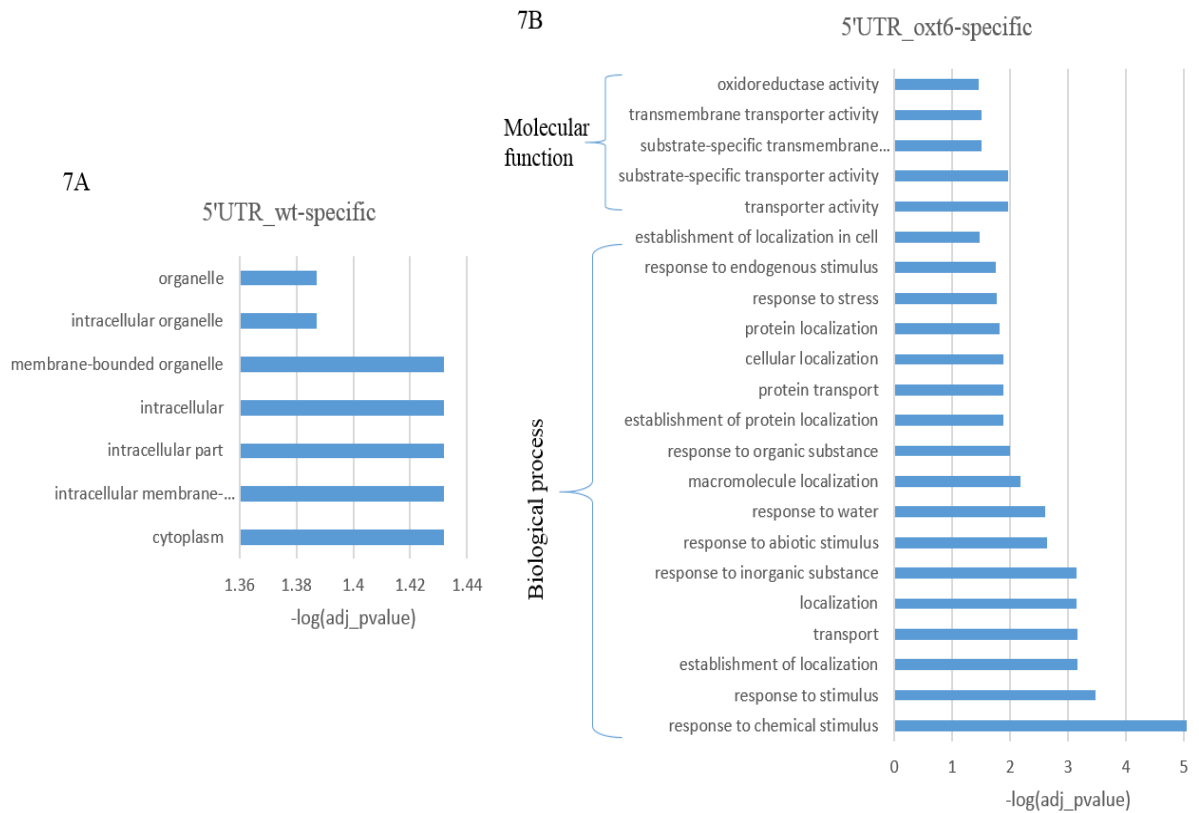


Figure 3.7 The gene ontology analysis for mRNA isoform that fall into 5'UTR

Gene Ontology term enrichment according to the agriGO, which is a toolkit and database for agriculture community (Website <http://bioinfo.cau.edu.cn/agriGO/>). Parameters are that reference is *Arabidopsis* genemodel (TAIR); the statistic model is hypergeometric test; the multi-test adjust method is FDR adjust p-value; the significance is less than 0.05; and the minimum number of mapping entry is 9. The X-axis is minus log FDR adjusted p-value. The Y-axis is the GO terms.

- A. GO analysis for unstable wt-specific mRNA isoforms that fall into 5'UTRs.
- B. GO analysis for unstable *oxl6*-specific mRNA isoforms that fall into 5'UTRs.

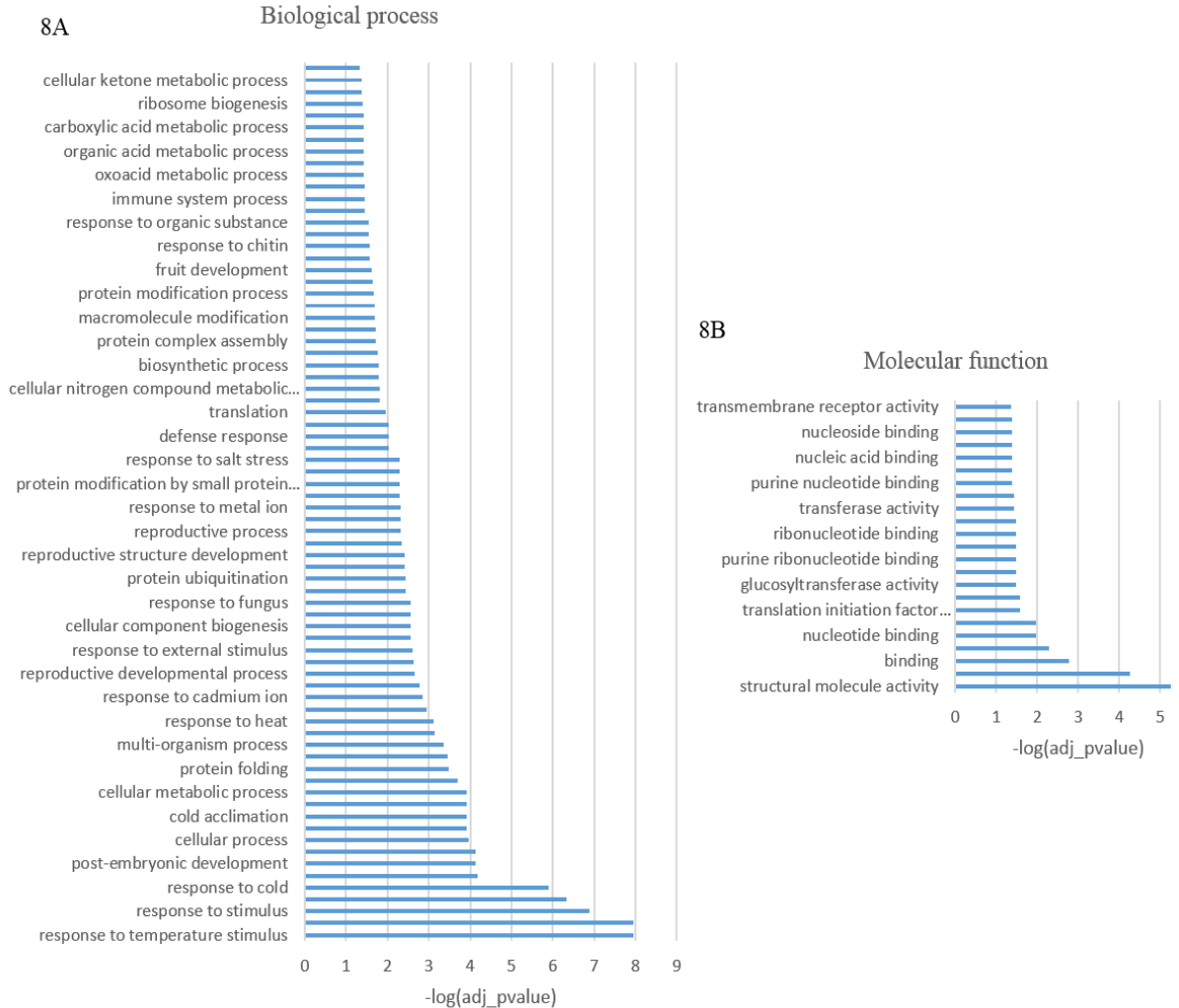


Figure 3.8 The gene ontology analysis for mRNA isoforms that fall into CDS in wt-specific

Gene Ontology term enrichment according to the agriGO, which is a toolkit and database for agriculture community (Website <http://bioinfo.cau.edu.cn/agriGO/>). Parameters are that reference is *Arabidopsis* genemodel (TAIR); the statistic model is hypergeometric test; the multi-test adjust method is FDR adjust p-value; the significance is less than 0.05; and the minimum number of mapping entry is 9. The X-axis is minus log FDR adjusted p-value. The Y-axis is the GO terms.

- A. Biological process for unstable wt-specific mRNA isoforms that fall into CDS.
- B. Molecular function for unstable wt-specific mRNA isoforms that fall into CDS.

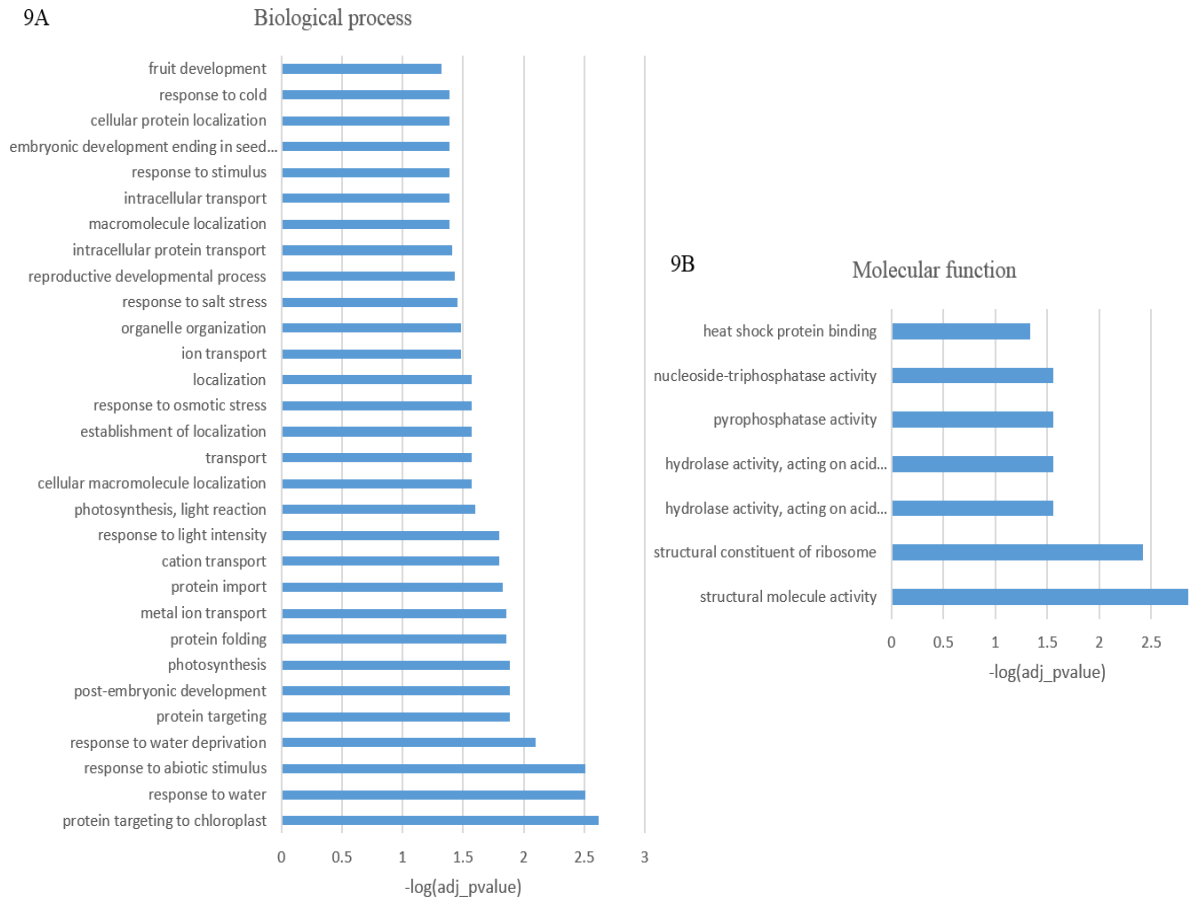


Figure 3. The gene ontology analysis for mRNA isoforms that fall into CDS in *oxt6*-specific

Gene Ontology term enrichment according to the agriGO, which is a toolkit and database for agriculture community (Website <http://bioinfo.cau.edu.cn/agriGO/>). Parameters are that reference is *Arabidopsis* genemodel (TAIR); the statistic model is hypergeometric test; the multi-test adjust method is FDR adjust p-value; the significance is less than 0.05; and the minimum number of mapping entry is 9. The X-axis is minus log FDR adjusted p-value. The Y-axis is the GO terms.

A. Biological process for unstable *oxt6*-specific mRNA isoforms that fall into CDS.

B. Molecular function for unstable *oxt6*-specific mRNA isoforms that fall into CDS.

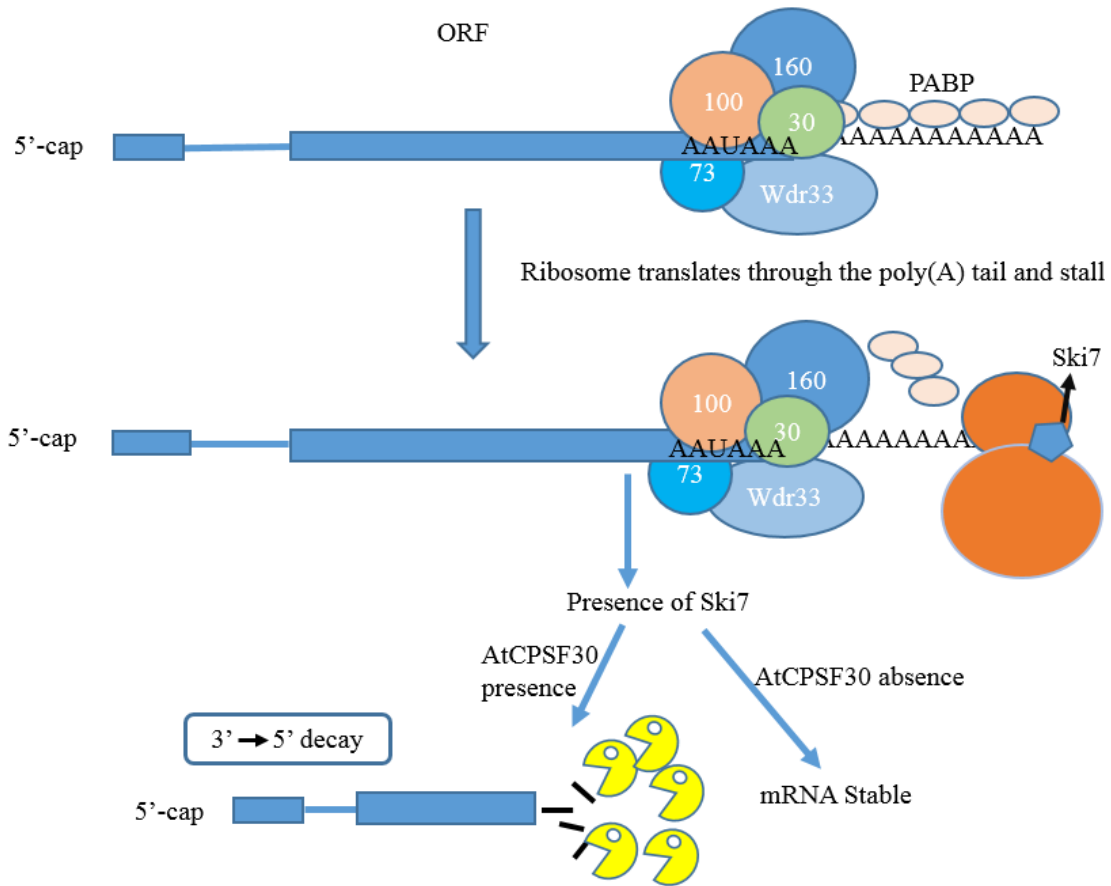


Figure 3.10 A model to describe the mRNA isoform stability in the CDS region

In this model, AtCPSF30 might act with Ski7 to recruit the exosome to degrade the mRNA isoforms. When the nonstop mRNAs are translated, the ribosome proceeds through the end of the poly(A) tail, displacing the poly(A)-binding protein (PABP) and stalling at the 3' end of the mRNA. The stalled ribosome at the 3' end of a transcript initiates non-stop decay through the binding of the C-terminus of Ski7 to the empty A site of the ribosome, thus releasing the ribosome. Then the exosome is recruited by Ski7 and AtCPSF30, along with the SKI complex, deadenylates and rapidly degrades the transcripts. Therefore, wt-specific mRNA isoforms derived from polyadenylation in the CDS are unstable. Without AtCPSF30, Ski7 may be unable to recruit the exosome, so the mRNA isoforms derived from polyadenylation in the CDS would be more stable in the *ox16* mutant.

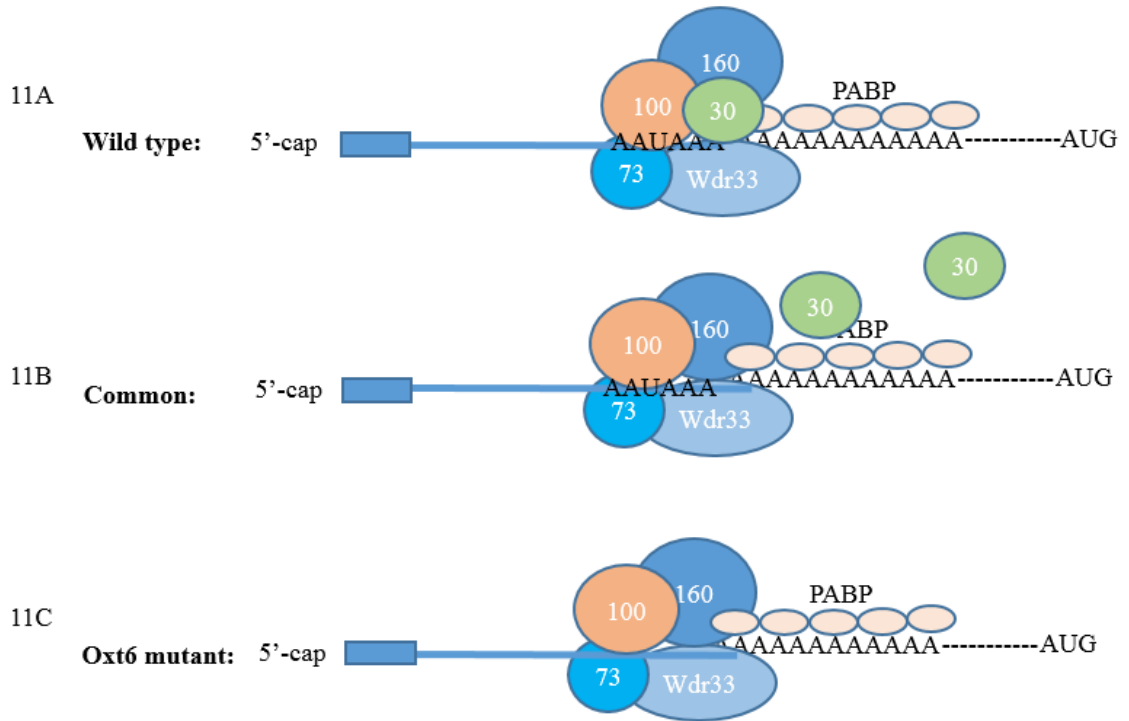


Figure 3.11 A model to describe the mRNA isoform stability in the 5'UTR region

- A. The possible model for AtCPSF30 acting with the polyadenylation complex as a unit.
- B. The possible model for AtCPSF30 acting without interactions with the polyadenylation complex.
- C. The possible model mRNA stability in the absence of AtCPSF30.

Chapter Four: A screen for proteins that interact with AtCPSF30

4.1 Introduction

mRNA 3' end processing is not only an important step in eukaryotic gene expression but also a critical mechanism for gene regulation (Colgan and Manley 1997; Zhao et al. 1999b; Millevoi and Vagner 2010; Chan et al. 2011). Polyadenylation affects gene expression through determining the coding and regulatory capacity of the mRNA, especially for genes whose transcripts may be polyadenylated at more than one position (Lutz and Moreira 2011; Xing and Li 2011a).

Polyadenylation is mediated by a complex that includes several subunits that are evolutionarily conserved for the most part (Belostotsky and Rose 2005; Hunt et al. 2008). Among these subunits, the 30 kD subunit is enigmatic. In *Arabidopsis*, CPSF30 (AtCPSF30) plays an important role in regulating APA. This protein contains three predicted CCCH-type zinc finger motifs. The first CCCH motif is the primary motif that is responsible for the bulk of RNA-binding activity. It can also bind calmodulin, and the RNA-binding activity of AtCPSF30 is inhibited by calmodulin in a calcium-dependent manner. The function of the second zinc finger is not precisely known (Delaney et al. 2006; Addepalli and Hunt 2007b; Chakrabarti and Hunt 2015). The third motif is associated with endonuclease activity. Previous studies demonstrated that the endonuclease activity of AtCPSF30 could be inhibited by disulfide reducing agents (Addepalli and Hunt 2008).

Messenger RNA polyadenylation occurs in the nucleus, and the subunits of the polyadenylation complex that mediated this process are expected to reside within the nucleus. However, AtCPSF30 by itself localizes to the nucleus and also to the cytoplasm

(Rao et al. 2009). Another study showed that CPSF30 has interactions with other subunits of the mRNA polyadenylation complex (Hunt et al. 2008). There is a model proposed whereby CPSF30 might act as an interface between cellular signaling and alternative polyadenylation (Hunt 2014). In this model, either calmodulin or redox status would cause an inhibition of CPSF30 as a signaling cues. The effects of calmodulin are obvious, but the disulfide remodeling might play an important role for the redox status. Based on these studies, we know that CPSF30 is not only in the nucleus, but also the cytoplasm. However, what proteins might interact with AtCPSF30 in the cytoplasm are still unknown. In this chapter, screen was conducted in an effort to identify such proteins. The results show AtCPSF30 interacts with two proteins, one of which is tyrosine-phosphorylated and whose phosphorylation state is modulated in response to ABA, and the other is ribosomal protein RPL35.

4.2 Result

4.2.1 Selected candidates interacting with AtCPSF30 from phage display

To identify novel proteins that can interaction with AtCPSF30, a phage display approach was chosen, since this is one of the most effective techniques to canvass a wide diversity of proteins and peptides (Bazan et al. 2012). The recombinant phage was used as a scaffold to present various protein portions encoded by a directionally cloned cDNA library to immobilized bait molecules (Kushwaha et al. 2014). For the *Arabidopsis* seed phage display cDNA libraries, the total RNA from mature, dehydrated (0h) or 12h-, 24h-, or 36h-germinated (on water at 25 °C with constant light) *Arabidopsis* seeds was the source of polyA⁺ mRNA (Chen et al. 2010). The cDNAs were ligated into T7Select10-3b vector arms, the libraries packaged into bacteriophage T7 in vitro, amplified and stored in aliquots

at -80C until use. Aliquots were used for serial dilutions to infect the *E. coli* strain BLT5403 to determine the titer of the primary recombinants (Chen et al. 2010). In this study, the seed library I, the titer of which was 1.6×10^7 pfu/ml, was used for initial affinity screening. After amplification, the titers of library I were 8.15×10^9 pfu/ml (Chen et al. 2010).

Based on these results, the MBP-tag(MBP), MBP-tag CPSF30(C30) and the third zinc finger mutation of CPSF30 with MBP tag(mC30) were used to perform biopanning to capture phage that interact with each bait as described in the Methods and illustrated in Figure 4.1. Here the plaque forming units (PFU) were used to determine and calculate titer. At the fourth round of phage display, the plaque forming units went down compare with the third round (Figure 4.2) suggesting further affinity selection would be unproductive after the fourth round. Therefore, the phage from the fourth round of screening were used to preparing phage display libraries for sequencing.

The Miseq data representing the affinity-purified phage was analyzed using CLC Genomics Workbench. Reads were demultiplexed, trimmed, and mapped to a set of reference sequences (Figure 4.3). The mapping references used consisted of the set of *Arabidopsis* protein coding sequences. Using the mapping results, gene lists were obtained. The gene lists for each of the three replicates for the different “baits” (MBP-CPSF30, MBP, etc.) were pooled for further analysis. The results showed that there are 245 proteins that exclusively interact with C30 (Figure 4.4A). To narrow down the gene list, the gene lists from individual C30 replicates were compared; this yielded a set of 107 genes present in two or more replicates. These have the greatest possibility of representing bone fide interacting partners with C30. These lists of 245 and 107 were further compared to get a set of 15 high-confidence candidates. These 15 candidates were screened first to make sure

they were in-frame (Figure 4.4B). Finally, the *Arabidopsis* sequences that were in-frame with the phage coat protein coding region yielded three candidates, At1g05510 (OIL BODY-ASSOCIATED PROTEIN 1A; OBAP1A), At1g09590 (CYTOSOLIC LARGE RIBOSOMAL SUBUNIT 21E; RPL21E) and At5g02610 (RIBOSOMAL PROTEIN L35; RPL35). These were chosen for further study.

4.2.2 Yeast two hybrid assay to confirm the three candidates

To confirm these three candidates, yeast two hybrid assays were performed. Because AtCPSF30 self-activates when cloned into the DNA-binding domain of the two-hybrid system, it was cloned into the activation domain vector. The three candidates were cloned into the binding domain vector. The full length AtCPSF30 cloning into pAD, and the empty BD plasmid, were used as negative controls. The AtMYC2 in pAD and AtGBF1 in pBD were used as positively controls (Maurya et al. 2015). Transformants were identified by selecting on synthetic dropout (SD) medium without Leu and Trp (-LT) (Figure 4.6A). These were then grown on SD media without Leu, Trp and His (-HLT) to test for protein-protein interactions. The results showed that only AtCPSF30 and OBAP1A interact with each other (Figure 4.6B). Interestingly, no transformants were obtained with the RPL35-AtCPSF30 combination, suggesting that this combination might be toxic to yeast cells.

To further confirm these results, transformants were grown on media containing x-gal, taking advantage of the presence of the *LacZ* reporter gene in the yeast host strain. The results confirmed the interaction between AtCPSF30 and OBAP1A (Figure 4.6C).

4.2.3 Bimolecular Fluorescence Complementation to confirm corresponding interactions

These above results showed that OBAP1A interacts with AtCPSF30 in yeast cells. To further confirm interaction, Bimolecular Fluorescence Complementation assays (BiFC) were performed. In addition, this assay was used to test for interactions between AtCPSF30 and RPL35 in plant cells; this is an alternative to the two-hybrid assay that failed due to the apparent toxicity of the AtCPSF30- RPL35 combination in yeast cells.

Before the BiFC assay, localization experiments were performed. In this assay, AtCPSF30, OBAP1A and RPL35 were fused to the GFP coding region present in the pSITE2N and pSITE2C vectors, differing with respect to where the GFP fusion protein is situated relative to the experimental protein (i.e. N- or C-terminus, respectively), and expressed transiently in *Nicotiana benthamiana* (Chakrabarty et al. 2007). The results showed that expression of the fusion proteins containing AtCPSF30 and OBAP1A could be observed with the pSITE2C vector, while RPL35 protein expression could be obtained with two vectors. However, for the latter protein, the pSITE2N vector yielded higher expression levels. The results of the localization experiments showed that AtCPSF30 is located in both the nucleus and cytoplasm (Figure 4.7A), as reported earlier (Rao et al., 2009). The proteins encoded by OBAP1A and RPL35 were located mainly in the nucleus (Figure 4.7B-C).

To test the functionality of the binary BiFC vectors in plant, the wild type AtCPSF30, OBAP1A At1g05510 and RPL35 cDNA were cloned into pSITEII-nEYFP and pSITEII-cEYFP, respectively. Here the empty pSITEII-nEYFP and pSITEII-cEYFP act as negative controls, while the *CoRSV_ORF2* was used as a positive control (Ramalho et al. 2014). The constructs were delivered into leaf cells of CFP-H2B tobacco (*Nicotiana benthamiana*)

by *Agrobacterium* infiltration (Goodin et al. 2008). The result shows that AtCPSF30 and RPL35 interact in the nucleus (Figure 4.8).

4.3 Discussion

4.3.1 The approaches to identify potential interacting proteins of AtCPSF30

CPSF30, one of the subunits of the eukaryotic polyadenylation complex, is a small protein that consists of a characteristic array of conserved CCCH-type zinc finger proteins, binds RNA, and interacts with several other polyadenylation factor subunits. One previous study reported that AtCPSF30 interacts with AtCPSF160, AtCPSF100, AtCFIS2, both FIPS orthologs, AtPCPS1, AtPCFS5, and AtCLPS3, which may as a central hub in the protein-protein interaction network of plant polyadenylation complex subunits (Hunt et al. 2008). This study mainly used yeast two yeast hybrid to test protein-protein interactions amongst *Arabidopsis* polyadenylation complex subunits. Another study showed that AtCPSF30 interacts with other polyadenylation factor subunits, AtCPSF160 or AtCPSF73(I) in the nucleus when co-expressed (Rao et al. 2009). These two experiments studied specific targets that reside in the poly(A) complex. However, these proteins are expected to reside within the nucleus, whereas AtCPSF30 may also reside in the cytoplasm (Rao, et. al, 2009). Screening a phage display library has the potential to identify additional, unknown interacting partners.

The advantage of this approach is that the cDNA phage display library is a global one for seeds that may include novel partners that interact with the bait protein (AtCPSF30). Thus, large libraries of proteins can be screened by the in vitro selection. However, the number of genes returned is too large to confirm all of them. To narrow down the gene number and increase the possibility of identifying bonafide candidates, the overlap among three

replicate wells was used. In the hypothesis, we expected to find proteins that might play a role in redox regulation, but found none. Possibly our strategy to narrow down the number of genes screened out possible redox-related proteins or the conditions leading to interaction must actually be oxidative.

4.3.2 Possible connection between AtCPSF30 and expression protein of OBAP1A

AtCPSF30 contains three predicted CCCH-type zinc finger motifs. The first CCCH motif is the primary motif responsible for the bulk of CPSF30's RNA-binding activity. It can bind with calmodulin, but the RNA-binding activity of AtCPSF30 is inhibited by calmodulin in a calcium-dependent manner. The third motif is associated with endonuclease activity. Previous studies demonstrated that the endonuclease activity of AtCPSF30 can be inhibited by disulfide reducing agents. Here we expected proteins that interact with AtCPSF30 and act through calmodulin binding or disulfide remodeling.

One of the proteins identified in this study as interacting with AtCPSF30 is encoded by At1g05510 (OBAP1A). This gene is expressed in *Arabidopsis* seeds and during germination (Job et al. 2005). OBAP1A, encoded by At1g05510, is tyrosine-phosphorylated and this phosphorylation state is modulated in response to ABA (Ghelis et al. 2008). Protein tyrosine phosphorylation plays important roles in cell growth and differentiation in animals. Also, protein tyrosine phosphorylation has been detected in *Arabidopsis thaliana*. OBAP1A, coded by At1g055010, was identified in a mass spectrometry-based proteomics study of responses of *Arabidopsis* to abscisic acid (ABA) (Ghelis et al. 2008). Specifically, the overall abundance of this protein decreases, but tyrosine phosphorylation increases in ABA-treated seeds.

Several studies on tyrosine phosphorylation used phenylarsine oxide (PAO), a specific protein tyrosine phosphatase (PTP) inhibitor (Quettier et al. 2006; Ghelis et al. 2008). In the cytosolic and detergent soluble fractions as well as on surface of *Setaria cervi*, the PTP activity present was significantly inhibited by PAO (Singh et al. 2016). Also, PAO can prevent stomatal closure in *Commelina communis*. Stomatal closure can be induced by ABA, high external Ca^{2+} , darkness, and hydrogen peroxide (Quettier et al. 2006). Moreover, with the increased ROS level, the activity of different antioxidant enzymes like thioredoxin reductase, glutathione reductase and glutathione transferases was decreased in PAO-treated plants (Singh et al. 2016). Along with previous studies showing that CPSF30 might be involved in ROS regulation (Zhang et al. 2008), this provides a possible reason for the interaction of CPSF30 with OBAP1A.

Protein tyrosine phosphorylation activity is essential for stomatal closure induced by ABA, external calcium, darkness and H_2O_2 (Enid 2002). It is well known that ABA regulates guard cell turgor via a calcium-dependent pathway (Julian et al. 2001; Luan 2002). H_2O_2 serves as the downstream messenger for ABA and activates calcium channels. The phosphorylation of tyrosine lies downstream from the calcium response (Enid 2002). These facts suggest that multiple calcium-dependent signaling mechanisms may converge on AtCPSF30, and that one may involve OBAP1A.

Additionally, a BLASTP search of the plant genome database shows that the protein encoded by OBAP1A has a strong similarity to the embryo-specific protein Ose731 from rice. Ose731 has a lipoprotein domain that is also found in oleosins. Oleosins stabilize oil bodies in seeds during desiccation and facilitate the hydrolysis of these reserves by lipases during germination (Keddie et al. 1992). Also, the levels of oleosins decrease in *Brassica*

napus and rice after ABA treatment (Holbrook et al. 1991; Konishi et al. 2005). Oleosins have been shown to be phosphorylated in *B. napus* (Holbrook et al. 1991). Therefore, tyrosine phosphorylation through an interaction between AtCPSF30 and OBAP1A might regulate oil body biogenesis or oleosin targeting to modulate the synthesis or hydrolysis of lipid reserves.

From the localization experiment, the protein encoded by OBAP1A is located both in the cytoplasm and nucleus. Previous studies showed that it is involved in oleosome biogenesis and the ABA response (Job et al. 2005; Ghelis et al. 2008), but no study detected it in the nucleus. However, one study showed that the protein encoded OBAP1A might be involved in the protein/transcript expression patterns by protein profiling in *Arabidopsis thaliana* seed development (Hajdich et al. 2010).

4.3.3 Possible connection between AtCPSF30 and RPL35

Another protein identified in this study as interacting with AtCPSF30 is encoded by At5g06210. This protein belongs to ribosomal L29 family and is termed *Arabidopsis* 60S ribosomal protein L35 (RPL35). RPL35 is an important component of the 60S ribosomal subunit, and is required for 60S r-subunit accumulation (Zhong and Arndt 1993).

In eukaryotes, the few studies that have focused RPL35 have shown it to have a role in translation and endoplasmic reticulum docking (Chen et al. 2008; Babiano and de la Cruz 2010). Eukaryotic RPL35 shares a remarkable sequence identity with archaeal and eubacterial RPL29, although RPL35 has a specific C-terminal extension (Babiano and de la Cruz 2010). The location of RPL35 is close by RPL25 and RPL26, thus flanks the nascent peptide exit tunnel in a similar way to their archaeal and eubacterial RPL29, RPL23 and RPL24 counterparts, respectively (Klein et al. 2004). RPL35 with RPL25 together may

have a function as a general docking site for nascent polypeptide chain-associated factors, like N-terminal processing and modification enzymes, chaperones, the signal recognition particle (SRP) and the Sec61/translocon complex (Becker et al. 2009; Günter et al. 2009). Another study showed that RPL35 is indeed required for the normal accumulation of 60S r-subunit because the defective production of RPL35 causes defects in pre-rRNA processing and in intra-nuclear transport and nuclear export of pre-60S r-particles (Babiano and de la Cruz 2010). This study also showed that RPL35 assembles in the nucle(ol)us and interacts with 27S pre-rRNA stably, indicating that RPL35 might be added early in the ribosomal assembly pathway (Babiano and de la Cruz 2010). In our study, the localization and BiFC experiment shows that the protein encoded by At5g06210 is located in the nucleolus, suggesting that AtCPSF30 might have a function in ribosomal biogenesis.

In zebrafish (*Danio rerio*), the protein RPL35 is an haploinsufficient tumor suppressor operating through an as yet unknown mechanism, similar to many other r-protein gene products (Amsterdam et al. 2004). In humans, RPL35 carries a 54-aa eukaryotic-specific expansion segment (ES) at the C-terminal end. Within the ES, the first 25 amino acid residues play an important role for the nuclear import of the protein, while the last 29 residues are one of the contact sites for ribosomal docking to endoplasmic reticulum (Chen et al. 2008). In addition, *RPL35* expression is significantly up-regulated during bovine lactation (Bionaz and Looor 2011). Another study showed that RPL35 could positively control translational elongation by interacting with eEF2 and by locating to the ER during CSN2 secretion, indicating that RPL35 is an important regulatory factor involved in the Met-mediated regulation of CSN2 translational elongation and secretion (Jiang et al. 2015). In yeast, RPL35 assembly is essential for the efficient cleavage of the internal transcribed

spacer 2 at site C2 (Babiano and de la Cruz 2010). One possible role for the AtCPSF30-RPL35 interaction might be to recruit AtCPSF30, which has endonuclease activity (Addepalli and Hunt 2007a), to the pre-rRNA for processing of the precursor.

4.3.4 Summary

In summary, a large number of AtCPSF30-interacting protein candidates were identified by screening an mRNA phage display library. Further analyses confirmed interactions of two of these, encoded by OBAP1A and RPL35. Based on other studies of these proteins, these studies show that AtCPSF30 might have different roles in the plant, and other proteins that interact with this protein might affect its functioning in polyadenylation in the nucleus and in the cytoplasm.

4.4 Material and method

4.4.1 Vector construction and site-directed mutagenesis

Total RNA was extracted for wild type *Arabidopsis*, and the cDNA encoding AtCPSF30 was isolated by RT-PCR as described by Addepalli (Addepalli et al. 2004). The DNA polymerase, *Taq* DNA polymerase was used to amplify the products with A overhangs. The A residue overhang product was cloned into pGEM T vector and the inserts sequenced. The *AtCPSF30* coding region was excised from a confirmed pGEM clone as an NdeI-EcoRI fragment and cloned into similarly-digested pMAL-c5x. The recombinant plasmid pMAL-c5x-AtCPSF30 was obtained after ligation and transformation. This plasmid was transformed into Rosetta cells for protein expression.

The mZF1, mZF3 and mZF13 mutants were generated from pMAL-c5x-AtCPSF30 plasmids as the templates using the Quick-Change site-directed mutagenesis kit

(Stratagene) and the oligonucleotides shown in Table 1. The mutant clones were confirmed by Sanger sequencing.

4.4.2 Protein expression and purification

To produce MBP fusion proteins, 10 mL overnight culture of transformed Rosetta cells grown and used to inoculate 200 mL of LB media. The culture was grown at 37°C for 3 to 4hr, until the OD600 reached 0.5. The fusion protein gene expression was induced by adding isopropylthio- β -galactoside to a final concentration of 0.3mM. After adding the inducer, the culture was grown for an additional 2 h. Cells were harvested and resuspended in 5ml of lysis buffer (GLB; 20 mM Tris-HCl (pH 8.0), 200 mM NaCl and 1 mM EDTA). The resuspended cells were sonicated using three bursts for 30s each on ice and debris was removed by centrifugation at 4°C (8,000,000 \times g). The centrifugation was repeated to completely clarify the final solution.

This supernatant added to Amylose Resin High Flow that had been washed twice with GLB and incubated in a 500 μ L Eppendorf tube for 15min. The resin was collected by centrifugation (8,000,000 x g) and washed sequentially with 1 ml GLB+ 2M NaCl and then 1 mL of GLB. MBP fusion proteins were eluted with MBP elution buffer. The eluted protein was stored at -80°C.

Protein preparations were analyzed by SDS-PAGE and staining with Coomassie Brilliant Blue, and confirmed by Western blot. Quantities were estimated by comparison with known quantities of bovine serum albumin.

4.4.3 Phage display sequencing

Screening the libraries by biopanning

The seed phage display cDNA library I (Library titer: 6.1×10^7 pfu/mL) was obtained from Dr. Bruce Downie, and this experiment was performed in the Downie lab. To make the seed phage display cDNA libraries, hot borate extracted total RNA from mature, dehydrated (0) or 12h-, 24h, or 36h- germinated (on water at 25 °C with constant light) *Arabidopsis* seeds was the source of polyA+ mRNA (Chen et al. 2010). Phage synthesis consisted of acquiring mRNA following two rounds of selection over Oligotex resin (QIAGEN Inc., Valencia, CA, USA). One μ g of polyA+ mRNA from each stage of germination was combined for cDNA synthesis. A kit (OrientExpress random primer cDNA synthesis kit; Novagen, San Diego, CA, USA) was used for first strand synthesis (randomly primed) to synthesize first- and second-strand cDNAs. One μ g random primers (Hind III RP) were used to generate the Seed Library I (SLI) and 0.1 μ g to generate the second- (SLII) phage display library. At the end, T4 DNA polymerase blunted the ends, linkers were ligated to the cDNAs which were digested with Eco RI/Hind III, and cleavage products were size fractionated according to the instructions of the kit manufacturer (T7Select10-3 Cloning Kit, Novagen). The cDNAs were directionally ligated into T7Select10-3b vector arms, and the libraries packaged into bacteriophage T7 in vitro. An aliquot was used for serial dilutions to infect the *E. coli* strain BLT5403 to determine the titer of the primary recombinants (Chen et al. 2010). For each library, the primary libraries were then amplified using plate lysates, the phage extraction buffer (20 mM Tris, pH 8.0, 100 mM NaCl, 6 mM MgSO₄) from the plates was combined, treated with chloroform, centrifuged (3,000 X g, 5 min) and the supernatant recovered. After the titer of the two amplified libraries was determined, the libraries were mixed with 0.1 volume 80% glycerol, and 1 mL aliquots were prepared and stored at -80 °C (Chen et al. 2010).

One hundred μL of $10\mu\text{g/ml}$ protein solutions were placed in each microtiter plate well (Special protein binding plates) and the plates covered with plastic wrap. For this, three proteins-MBP, MPB-CPSF30, MPB-mCPSF30, were used. The microtiter plate was placed at 4°C and incubated overnight. The microtiter plate was taken out and washed in TBS, blocked using $200\mu\text{L}$ milk blocking reagent 5% (W/V) with shaking for 1h at room temperature, and washed with TBST buffer. After the last wash, $100\mu\text{L}$ phage library was added to wells and plates were shaken at room temperature for one hour. The supernatant was removed and the wells washed with TBST buffer. After the last wash, $200\mu\text{L}$ aliquots of *Escherichia coli* (BLT5403 cells $\text{OD}_{600}=0.5$ to 0.6) in LB were added to each well, and the plate was sealed with adhesive film, incubated for 20min at 37°C to allow infection. After 20min, a portion of the cells was plated at low density to determine the titer. The remainder was added to a 250mL Erlenmeyer flask contained 50mL BLT5403. Cells were shaken at 37°C for 3h or until visible lysis occurred. Amplified phage was centrifuged with final concentration 0.5M NaCl at $8000\times g$ 10min at 4°C , and the supernatant moved to a clean, sterile 50mL falcon tube. A few drops of chloroform were added and the tubes and stored at 4°C for the next biopanning round.

PCR and Sequence Analysis of Phage Recovered during Biopanning

After four rounds of biopanning, a $1\mu\text{L}$ aliquot of the final suspension was used for phage display sequencing preparation by PCR. Two steps were taken for the primers design of phage display. First, part of the Illumina first read adapter with a barcode and some random nucleotides was included to allow sequencing and subsequent analysis. Next, the PE-PCR1 and PE-PCR2 primers used for amplifying the cleaned reactions from the first step. These

primers give products fully compatible with the HiSeq and MiSeq platforms. The primers were designed with Illumina adaptor (Table 2).

The first PCR program was: initial denaturation of 30 seconds at 98°C, denaturation of 15 seconds at 98°C, annealing at 60°C for 15 seconds, extension at 72°C for 1 minute for 15 cycles for the first step. The second PCR program was the same as the first PCR, but only 12 cycles. The libraries were quantified by qRT-PCR and pooled together, and each sample contained at least 25ng of amplified cDNA library. The libraries were sequenced on a Miseq instrument.

The sequencing data was analyzed with CLC genomic workbench. These samples were demultiplexed, trimmed and mapping to the TAIR10 coding sequence to get the gene list. The number of genes were narrowed down by comparing the results of three replicates. Sequence analysis was conducted to assure that the respective protein coding regions were in frame with the phage coat protein gene present in the phage display vector.

4.4.4 Yeast two hybrid assays

The full-length *AtCPSF30* was cloned into pAD-GAL4-2.1, and the protein-coding regions for *At1g05510*, *At1g09590* and *At5g02610* were cloned into pBD-GAL4 Cam (Stratagene), respectively (oligonucleotide primers are given in Table (4.3)). Different combinations of bait and target plasmids were transformed into yeast strain AH109 (Clontech), and the corresponding transformants were selected on synthetic dropout (SD) medium without Leu and Trp. Transformed colonies were then streaked on SD-His-Leu-Trp and SD-Ade-His-Leu-Trp medium to check protein-protein interaction. To further confirm the positive clones in yeast system, experiments were performed in AH109 yeast strain to detect the transcription of *LacZ* reporter gene.

4.4.5 Subcellular localization and BiFC

Gateway cloning was performed for full-length *AtCPSF30*, *At1g05510* and *At5g02610* cDNAs using pSITEII-2NA, pSITEII-2CA, pSITE-nYEFP-C1, pSITE-cYEFP-C1, pSITE-nYEFP-N1 pSITE-cYEFP-N1, respectively (oligonucleotide primers are in Table 3). The vector was provided by Dr. Michael Goodin, Department of Plant Pathology, University of Kentucky, USA. All the constructs and control vectors were transformed into *Agrobacterium* strain LBA3108 using the freeze thaw method (An, Ebert et al. 1988). *Agrobacterium* was freshly grown on LB agar plates with respective antibiotics for 2 days at 28 °C. For agroinfiltrations, 1ml of an overnight culture was put into 25ml LB with appropriate antibiotics and grown at 28°C overnight. The suspension was harvested and suspended in MES buffer (10mM MES and 10mM MgCl₂, pH=5.9); the cell density was set to OD₆₀₀ of 0.6 to 1.0. Acetosyringone was added to all samples at 150µM concentrations and incubated at room temperature for 2-3 hours. For co-infiltrations with different constructs, equal volumes of different *Agrobacterium* suspensions were mixed before infiltrations.

Three-week old *Nicotiana benthamiana* that express CFP (cyan fluorescent protein) fused to histone 2B (H2B-CFP) (Kathleen et al. 2009) plants were used for infiltrations. For this, leaves were infiltrated with the *Agrobacterium* suspensions using a 1ml syringe without a needle. Fully expanded and healthy lower leaves were selected for infiltration experiments. The abaxial surface of the leaf was injected with the suspension buffer while, at the same time applying gentle pressure with a finger on the adaxial surface of the leaf. The buffer with bacterial suspension was infiltrated in the intravenous regions. At least two leaves were infiltrated with the same suspension buffer for a single plant. The infiltrated

plants were incubated under 16hr/8hr light and dark photoperiods at 25°C for a 48-72h period.

The infiltrated leaves were cut into small sections near the infiltration site. The abaxial side facing the cover slip was placed on a glass slide in a drop of water and mounted gently with a cover slip. The water-mounted sections of leaf tissue were chosen to examine by confocal microscopy. Images were observed using an Olympus FV1000 laser scanning confocal microscopy. The microscope is equipped with lasers for excitation wavelengths ranging from 488 to 514 nm. Micrographs for dual-color imaging were obtained sequentially, as described in Goodin et al. (2007a). Images were visualized under 20x and 40x objective lens and image processing was performed by using Olympus Fluoview software (Olympus). The images are typically acquired are a pixel resolution of 1024x1024 at a scan rate of 10 μ s/pixel.

Table 4.1 Primers for AtCPSF30 and its mutant

Primer Name	Sequence (5'-3')	Used for
AtCPSF30-F1	ATGGAGGATGCTGATGGACTTAG	Cloning into pGEM
AtCPSF30-R1	TTACAGAACCCAATTA AAAACCTTAGGACT	Cloning into pGEM
AtCPSF30-F2	CATATGGAGGATGCTGATGGACTT	Cloning into pMAL-C5
AtCPSF30-R2	GAATTCTTACAGAACCCAATTA AAAAC	Cloning into pMAL-C5
mZF1-F	AGGTGACGCCAGTACTTTTCTCTATCAGTTCGATA	Site-directed mutagenesis
mZF1-R	TTCATACACAGACCTCGAAGCCAGTGTCTACAAAC	Site-directed mutagenesis
mZF2-F	AGCAGGATAGTACTTATAAAATATACCAATGAAGAT	Site-directed mutagenesis
mZF2-R	CTCGACATTCACCGTATAACCGGAAAAAACGGCAG	Site-directed mutagenesis
mZF3-F	TGGTCCTGATAGTACTTACAGGTATGCAAAGCTTC	Site-directed mutagenesis
mZF3-R	TTGGGACAAAACCCAGCTTGACATATTGCATTC	Site-directed mutagenesis

Table 4.2. Primers for phage libraries sequencing

Primer Name	Sequence (5'-3')
F1-T7a	ACACTCTTTCCTACACGACGCTCTTCCGATCTtcCCCggagctgctgtattccagtcagg
F1-T7b	ACACTCTTTCCTACACGACGCTCTTCCGATCTtcCGGggagctgctgtattccagtcagg
F1-T7c	ACACTCTTTCCTACACGACGCTCTTCCGATCTtcAACggagctgctgtattccagtcagg
F1-T7d	ACACTCTTTCCTACACGACGCTCTTCCGATCTtcAGCggagctgctgtattccagtcagg
F1-T7e	ACACTCTTTCCTACACGACGCTCTTCCGATCTtcAGAggagctgctgtattccagtcagg
F1-T7f	ACACTCTTTCCTACACGACGCTCTTCCGATCTtcACGggagctgctgtattccagtcagg
F1-T7g	ACACTCTTTCCTACACGACGCTCTTCCGATCTtcCCGggagctgctgtattccagtcagg
F1-T7h	ACACTCTTTCCTACACGACGCTCTTCCGATCTtcCAAggagctgctgtattccagtcagg
F1-T7i	ACACTCTTTCCTACACGACGCTCTTCCGATCTtcCAGggagctgctgtattccagtcagg
F1-T7j	ACACTCTTTCCTACACGACGCTCTTCCGATCTtcCCAaggagctgctgtattccagtcagg
R1-T7	CGGTCTCGGCATTCTTGCTGAACCGCTCTTCCGATCTaaccctcaagaccgtttagagg
PE-PCR1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCTACACGACGCTCTTCCGATCT
PE-PCR2	CAAGCAGAAGACGGCATAACGAGATCGGTCTCGGCATTCTTGCTGAACCGCTCTTCCGATCT

Table 4.3 Primers for yeast two hybrid

Primer Name	Sequence (5'-3')	Used for
At1g30460-F	GGATCCATGGAGGATGCTGATGGACT	Cloning into PBD vector
At1g30460-R	CGCCGG AATTCTTACAGAACCCAATTA AAAACCTT	Cloning into PBD vector
At1g05510-F	CCGGAATTCATGGAGAAGGCAGTTCATTTA	Cloning into PAD vector
At1g05510-R	GTCCGATTAACGAAGACTCTTGGAACGG	Cloning into PAD vector
At1g09590-F	GAATTCATGCCGGCTGGACATGGAGT	Cloning into PAD vector
At1g09590-R	GTCGACCTAATAGCCTCCTTTGAGATCGTTGAC	Cloning into PAD vector
At5g02610-F	GCGTCGACATGGCGAGAATTAAGGTTC A	Cloning into PAD vector
At5g02610-R	CTGCAGCTACACTTTGATAGCGTACTTCCTCA	Cloning into PAD vector

Table 4.4. Primers for BiFC

Primer Name	Sequence (5'-3')
At1g30460-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGGATGCTGATGGACTTAG
At1g30460-R	GGGGACCACTTTGTACAAGAAAGCTGGGTCCAGAACCCAATTA AAAACCTTA
At1g05510-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATG GAGAAGGCAGTT CATTATCCA
At1g05510-R	GGGGACCACTTTGTACAAGAAAGCTGGGTCAACGAAGACTCTTGGAACGG
At5g02610-F	GGGGACAAGTTTGTACAAA AAAGCAGGCTTCATGGCGAGAATTAAGGTTTCATGAGC
At5g02610-R	GGGGACCACTTTGTACAAGAAAGCTGGGTCCACTTTGATAGCGTACTTCC

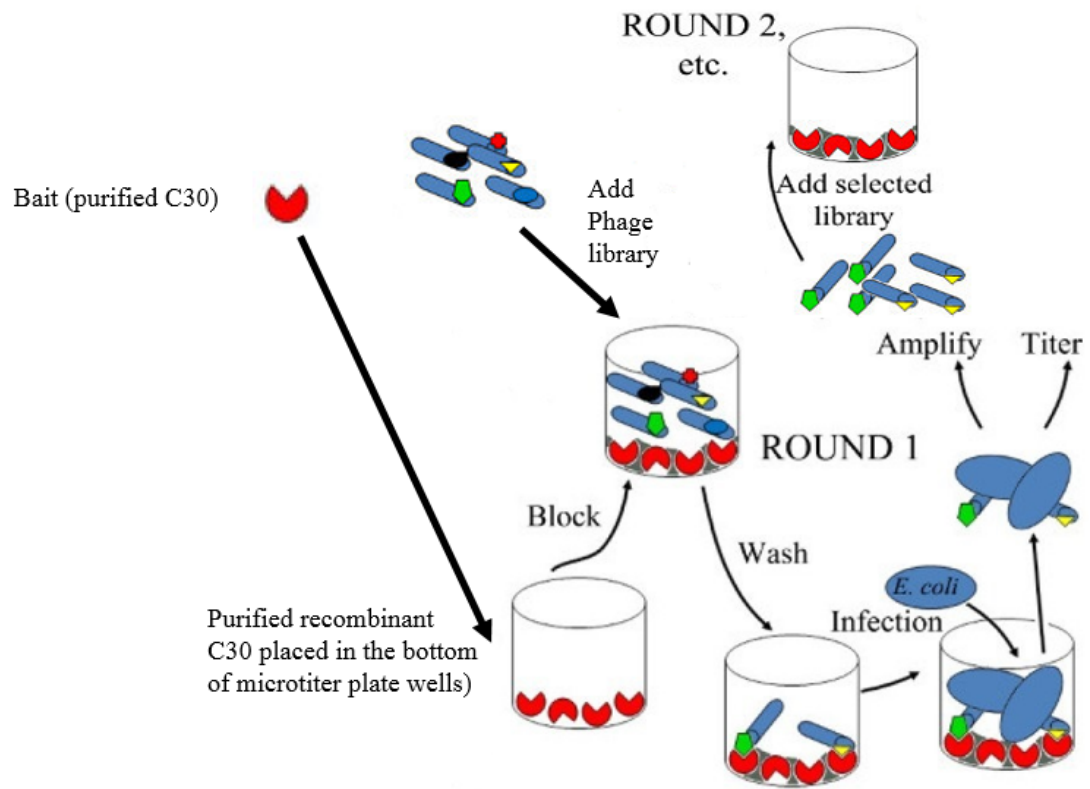


Figure 4.1 Overall graphic depiction of the phage display process (modified from Kushwaha, *et al.* 2014)

A bait was added at the microtiter plates for phage display screening using the *E. coli* (BLT5403). After *E. coli* was infected, one part was used for titering, while the rest for amplification. Amplified phage is used as sub-libraries for the next round of panning.

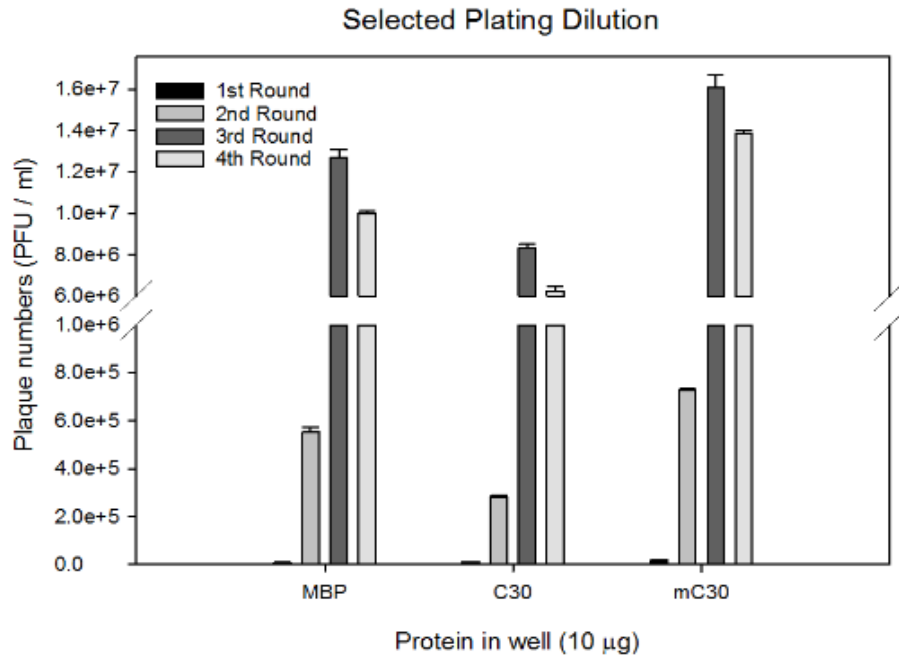


Figure 4.2 Titering for seed phage library I

The X-axis shows different kinds of protein as bait for phage display, including MBP, MBP-CPSF30(C30) and MBP-mCPSF30 (mC30). The Y-axis shows the PFU (Plaque Form Unit), which is used to calculate titer. This figure showed the titer increases dramatically for the all the baits until the third round. In the fourth round, titer from bait wells decrease, and further affinity selection is unproductive.

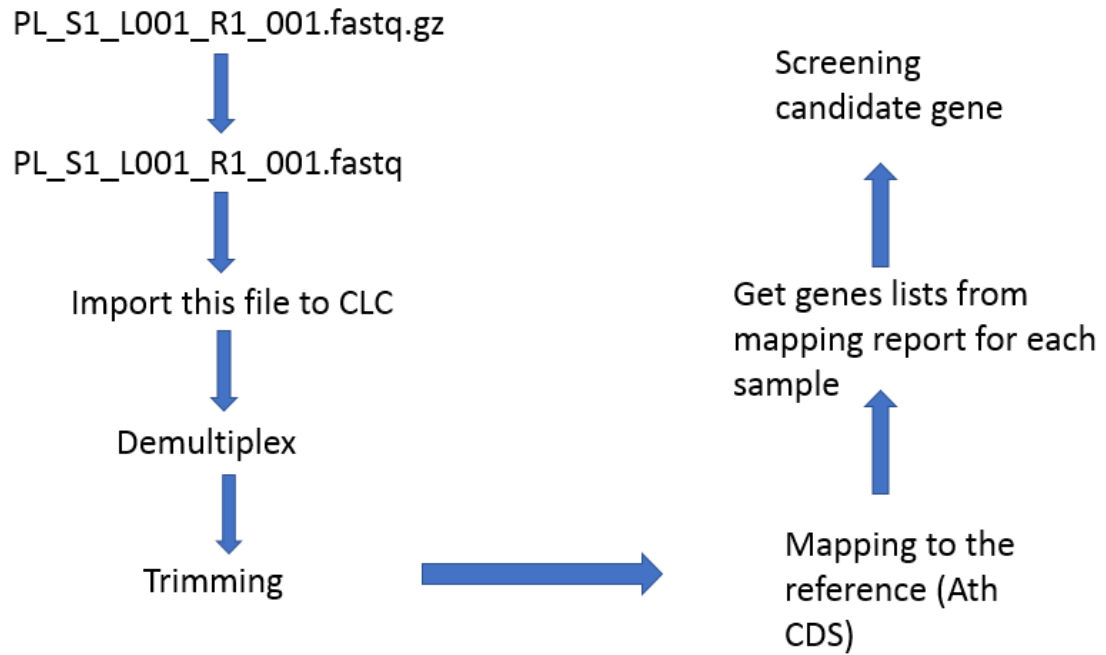


Figure 4.3 The flowchart for phage-sequencing data analysis

This figure shows the method for analysis of PEPSeq (Pair-end Phage Display sequencing).

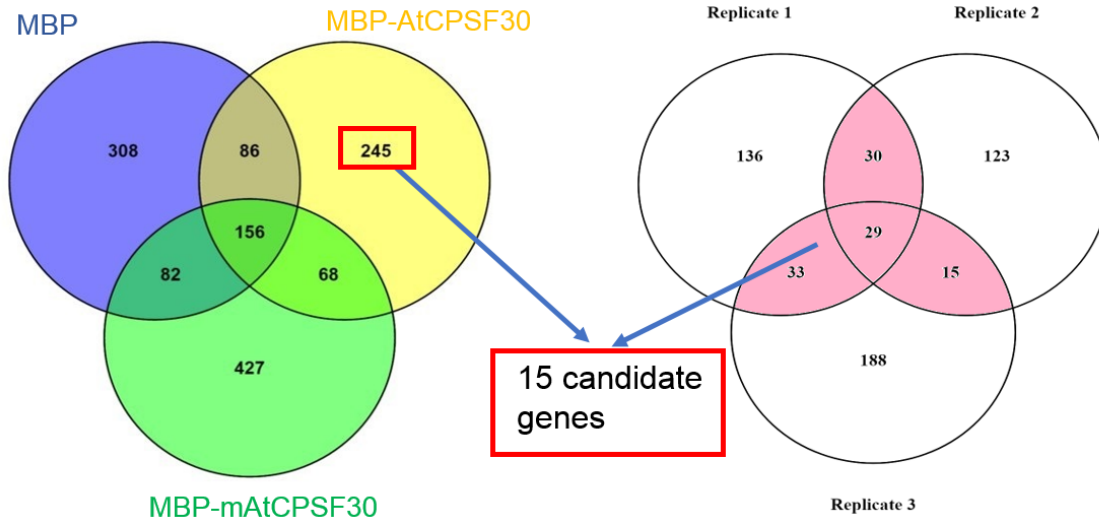


Figure 4.4 Selected Exclusive candidates interact with AtCPSF30

The gene numbers of MBP, C30 (MBP-ACPSF30), mC30 (mutant of AtCPSF30) are from the Miseq data representing the affinity-purified phage was analyzed using CLC Genomics Workbench. Reads were demultiplexed, trimmed, and mapped to a set of reference sequences (Figure 4.3). The mapping references used consisted of the set of *Arabidopsis* protein coding sequences. Using the mapping results, gene lists were obtained. These gene lists are compared by Venn diagram.

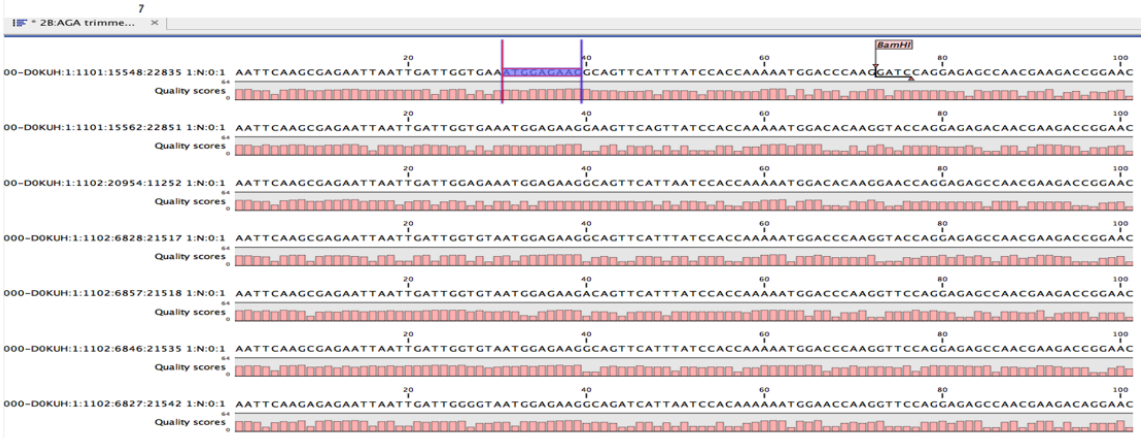
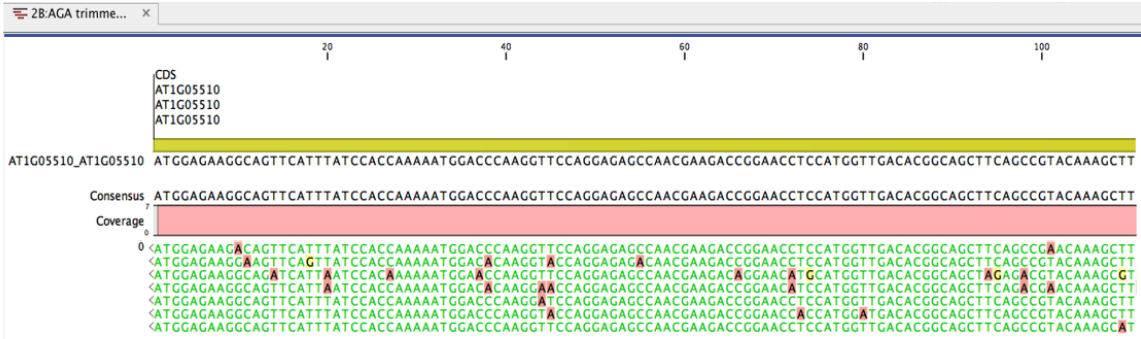
4A. To get the exclusive genes interacting with C30 (AtCPSF30)

This figure shows that there are 245 proteins that exclusively interact with C30.

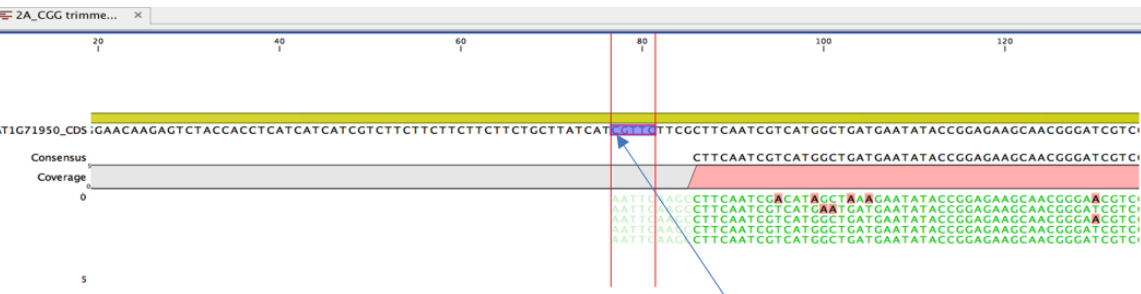
4B. To narrow down the number of genes that interact with AtCPSF30

To narrow down the gene list, the gene lists from individual C30 replicates (MBP-AtCPSF30) were compared; this yielded a set of 107 genes present in two or more replicates (pink), which have the greatest possibility of interacting with C30. These lists of 245 and 29 were further compared to get a set of 15 high-confidence candidates.

5A



5B



77

Figure 4.5 Correct or incorrect candidate as examples

A. This figure is to explain how to make sure candidate gene is in the open reading frame when it maps to a reference gene. The upper figure shows the At1g05510 sequences from Miseq are mapping with the CDS of this gene (yellow bar). After calculating the position of ATG for the phage coat protein, this gene is found to be in the proper open reading frame.

B. At1g71950 is an example that the candidate gene is not on the open reading frame.

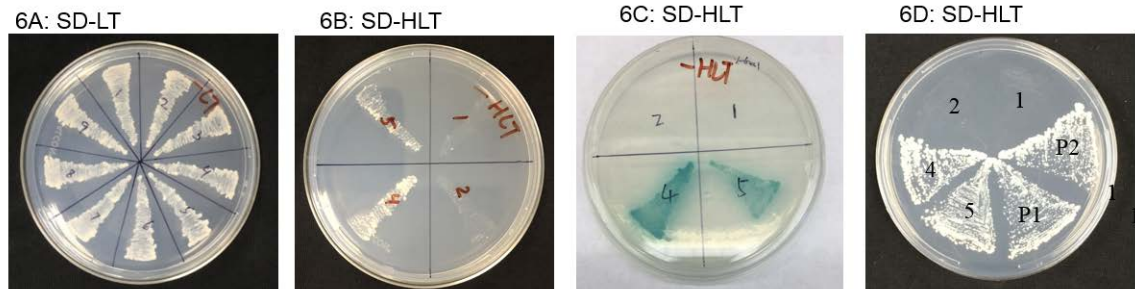


Figure 4.6. Yeast two hybrid to confirm interactions with AtCPSF30

To confirm these three candidates (At1g05510, At1g09590 and At5g02610) interact with AtCPSF30, yeast two hybrid assays were performed. AtCPSF30 self-activates when cloned into the activation domain vector. The three candidates were cloned into the binding domain vector. The full length AtCPSF30 was cloned into pAD, and the empty BD plasmid, were used as a negative control. The atMYC2 in pAD and atGBF1 in pBD were used as a positive control (Maurya et al. 2015). Transformants were identified by selecting on synthetic dropout (SD) medium without Leu and Trp (-LT) (Figure 4.6A). These were then grown on SD media without Leu, Trp and His (-HLT) (4.6B,4.6C, 4.6D) to test for protein-protein interactions. Here number 1-3 are the control, the combination of At1g30460 (AtCPSF30) in pAD and empty pBD; number 4-5 is the combination of At1g30460 and At1g05510; number 6-9 is the combination of At1g30460 and At1g09590. P1 and P2 mean positive controls.

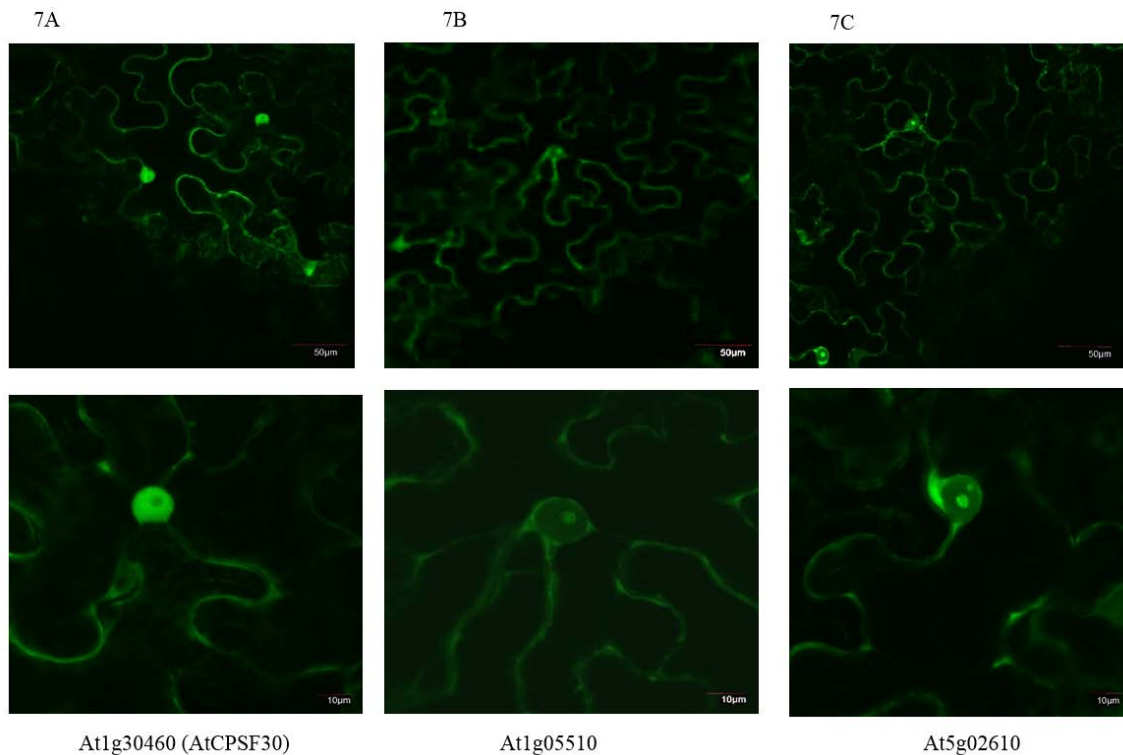


Figure 4.7 Subcellular location of At1g30460, At1g05510 and At5g02610

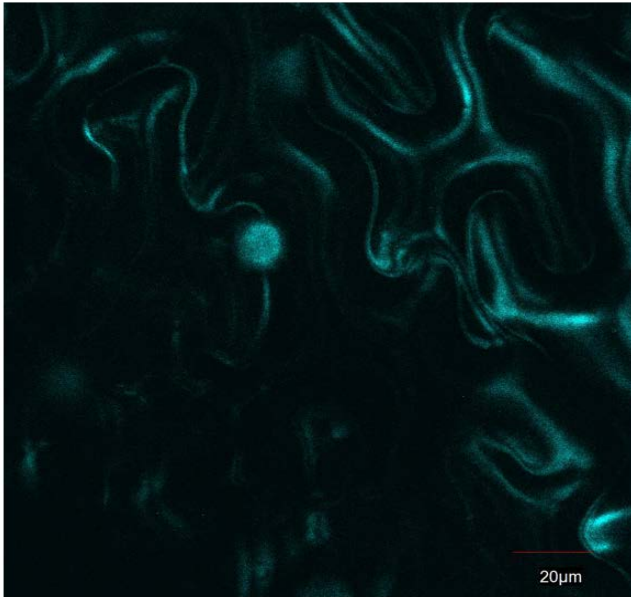
AtCPSF30 and At1505510 were cloned pSITE2-CA and the At5g02610 was cloned to Psite2-NA using Gateway cloning. After the recombination constructs confirmed by Sanger sequence, the Agrobacterium transformation was performed for three weeks old wild type *N.benthamiana* tobacco plants using a syringe. Plants were incubated under light at 18-28°C for 48 hours, followed by confocal microscopy was performed using Olympus FV1000 laser scanning. The upper panel pictures were taking by 20x objective lens and the lower panel by 40x.

7A. Subcellular distribution of AtCPSF30.

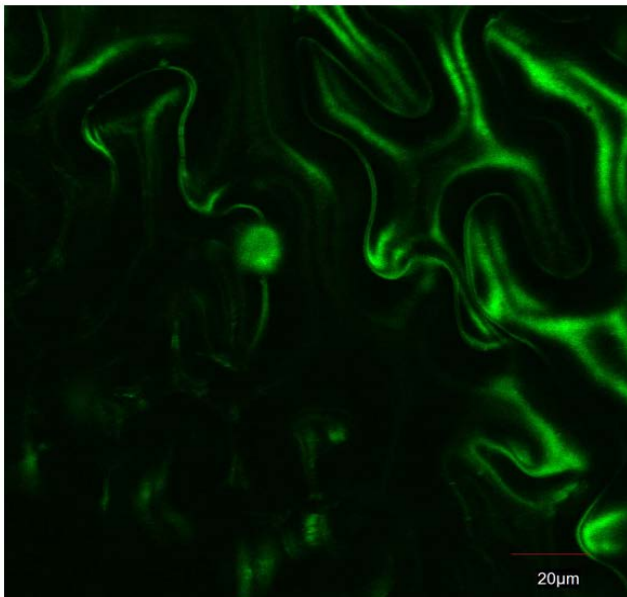
7B. Subcellular distribution of the protein encoded by At1g05510.

7C. Subcellular distribution of the protein encoded by At5g02610.

8A CFP-H2B



8B BiFC



8C

Overlay

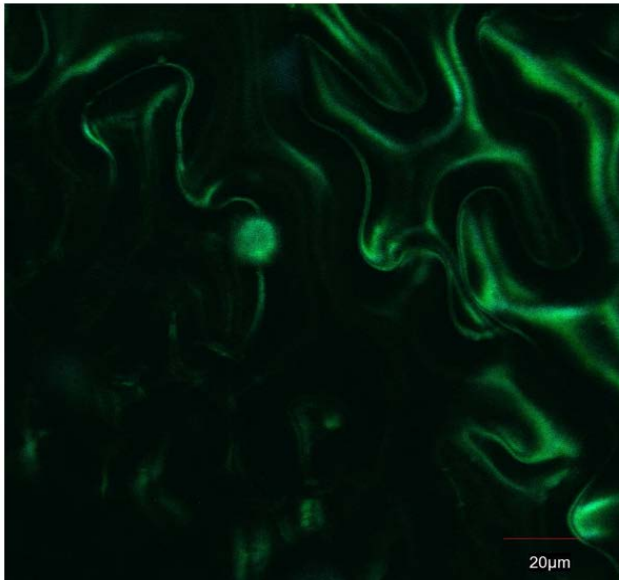


Figure 4.8. AtCPSF30 interacts with At5g02610 in the nucleus

AtCPSF30 was cloned into pSITE-cEYFP and At5g02610 cloned into pSITE-nEYFP using Gateway cloning. After the recombination constructs confirmed by Sanger sequence, the *Agrobacterium* transformation was performed on three weeks old wild type *N.benthamiana* tobacco plants using syringe. Plants were incubated under light at 18-28°C for 48 hours, followed by confocal microscopy was performed using Olympus FV1000 laser scanning. The pictures were taking by 20x and the lower panel by 40x objective lens.

Chapter Five: Summary and future perspectives

The process of mRNA regulation is well known to perform a fundamental role in determining the products of gene expression, and alternative polyadenylation is a major factor impacting the dynamics of gene regulation. Alternative polyadenylation requires several protein factors to define the correct cleavage site by recognition of sequence elements within the pre-mRNA. These protein factors are CSPF (cleavage and polyadenylation specificity factor), CstF (cleavage stimulation factor), CF (cleavage factor) along with poly (A) polymerase. It has been reported that alternative polyadenylation is a widespread phenomenon across the transcriptomes of higher eukaryotes, and this process is regulated by developmental and environment cues.

Since environment stress has been reported to lead to alternative polyadenylation, I hypothesized that drought stress should give rise to the poly (A) site changes. This is particularly the case since drought stress causes increases in reactive oxygen species, a characteristic that affects the functioning of CPSF30. My results showed that poly (A) site choice within 5'UTR and CDS change in drought-stressed plants. Interestingly, poly (A) sites within 5'UTR region show significantly increased usage after only one hour. Subsequently, the relative poly (A) site usage within 5'-UTRs decreases after six and twenty-four hours treatment, but again increases after two weeks. Why does relative poly (A) site within 5'UTRs change so rapidly? This might be caused by altering the efficiency of the polyadenylation complex or by changes in the stabilities of different mRNA isoforms, or the combination of these two mechanisms. Because this polyadenylation complex contains many proteins, it is difficult which of these might be impacted by stress. One of these may be CPSF30, since this protein is controlled by calmodulin and by reactive

oxygen species, two signaling systems associated with drought. Another possible reason is the stability of the mRNA isoforms that derived from polyadenylation within 5'UTRs and CDS. In particular, mRNA isoforms with 3' ends that lie within CDS regions lack stop codons and should be unstable. It is conceivable that stress may inhibit RNA quality control processes, leading to the result obtained in this thesis.

It is known that CPSF30 controls a large number of poly (A) sites in *Arabidopsis*. Between the *oxt6* mutant and wild type, there are three classes of poly(A) site, wild type specific, *oxt6* specific, and common (seen both in wild type and mutant). Around 70% of all sites are wt-specific or *oxt6*-specific sites. The stabilities of these mRNA may be different, which may provide one mechanism by which AtCPSF30 regulates gene expression. My results showed that the mRNA isoforms show similar stability profiles in the wild type and mutant plants except those with 3' ends within CDS. These latter mRNA isoforms are much more stable in the mutant than the wild type. This result suggests that AtCPSF30 may be involved in the process of non-stop mRNA decay.

The combined results of the drought stress and stability assays indicates that the efficiency of the polyadenylation complex might be the main reason for significant increases in poly (A) site usage within 5'UTRs. Also, other results suggest that CPSF30 could be involved in drought stress-induced APA. Therefore, CPSF30 might be a target for changing the efficiency of the polyadenylation complex in response to stress.

What proteins can interact with CPSF30 to cause the changes of poly (A) site choice and the mRNA stability? The protein-protein assays showed that CPSF30 interacts with OBAP1A, whose tyrosine phosphorylation state is modulated in response to ABA. ABA regulates guard cell turgor via a calcium-dependent pathway, thus bringing an additional

calcium signaling mechanism to CPSF30. CPSF30 also interacts with ribosome protein L35(RPL35), which plays an important role in nuclear entry, translation activity, and endoplasmic reticulum(ER) docking. These results indicate that multiple calcium-dependent signaling mechanisms may converge on CPSF30, and CPSF30 might directly interact the ribosome. Since CPSF30 interacts with OBAP1A and RPL35, it would be interesting to study OBAP1A and RPL35 mutants, and to characterize the double mutant of CPSF30 with OBAP1A and RPL35.

REFERENCES

- Aaron CG, Marvin W. 2008. Multifunctional deadenylase complexes diversify mRNA control. *Nature Reviews Molecular Cell Biology* **9**: 337.
- Addepalli B, Hunt AG. 2007a. A novel endonuclease activity associated with the Arabidopsis ortholog of the 30-kDa subunit of cleavage and polyadenylation specificity factor. *Nucleic acids research* **35**: 4453-4463.
- Addepalli B, Hunt AG. 2007b. A novel endonuclease activity associated with the Arabidopsis ortholog of the 30-kDa subunit of cleavage and polyadenylation specificity factor. *Nucleic acids research* **35**: 4453-4463.
- Addepalli B, Hunt AG. 2008. Redox and heavy metal effects on the biochemical activities of an Arabidopsis polyadenylation factor subunit. *Archives of Biochemistry and Biophysics* **473**: 88-95.
- Addepalli B, Limbach PA, Hunt AG. 2010. A disulfide linkage in a CCCH zinc finger motif of an Arabidopsis CPSF30 ortholog. *FEBS Letters* **584**: 4408-4412.
- Addepalli B, Meeks LR, Forbes KP, Hunt AG. 2004. Novel alternative splicing of mRNAs encoding poly(A) polymerases in Arabidopsis. *BBA-Gene Structure and Expression* **1679**: 117-128.
- Akimitsu N, Tanaka J, Pelletier J. 2007. Translation of nonSTOP mRNA is repressed post - initiation in mammalian cells. *EMBO Journal* **26**: 2327-2338.
- Alexander S, Céline D, Christian C, Vasiliki Z, Csaba H, Fatih O, Patrice MM, Geoffrey JB, Gordon GS. 2012. Direct sequencing of *Arabidopsis thaliana* RNA reveals patterns of cleavage and polyadenylation. *Nature Structural & Molecular Biology* **19**: 845.
- Alt FW, Bothwell ALM, Knapp M, Siden E, Mather E, Koshland M, Baltimore D. 1980. Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. *Cell* **20**: 293-301.
- Amrani N, Ganesan R, Kervestin S, Mangus DA, Ghosh S, Jacobson A. 2004. A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* **432**: 112.
- Amsterdam A, Sadler KC, Lai K, Farrington S, Bronson RT, Lees JA, Hopkins N. 2004. Many Ribosomal Protein Genes Are Cancer Genes in Zebrafish (Tumor Suppressor Genes in Zebrafish). *PLoS Biology* **2**: e139.
- An JJ, Gharami K, Liao G-Y, Woo NH, Lau AG, Vanevski F, Torre ER, Jones KR, Feng Y, Lu B et al. 2008. Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. *Cell* **134**: 175-187.
- Anderson JSJ, Parker R. 1998. The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO Journal* **17**: 1497-1506.
- Apel K, Hirt H. 2004. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annual review of plant biology* **55**: 373.

- Arigo JT, Eyler DE, Carroll KL, Corden JL. 2006. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Molecular Cell* **23**: 841-851.
- Babiano R, de la Cruz J. 2010. Ribosomal protein L35 is required for 27SB pre-rRNA processing in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **38**: 5177-5192.
- Babour A, Dargemont C, Stutz F. 2012. Ubiquitin and assembly of export competent mRNP. *BBA - Gene Regulatory Mechanisms* **1819**: 521-530.
- Badis G, Saveanu C, Fromont-Racine M, Jacquier A. 2004. Targeted mRNA degradation by deadenylation-independent decapping. *Molecular Cell* **15**: 5-15.
- Bai C, Tolia P. 1998. *Drosophila* clipper/CPSF 30K is a post-transcriptionally regulated nuclear protein that binds RNA containing GC clusters. *Nucleic Acids Research* **26**: 1597-1604.
- Bai C, Tolia PP. 1996. Cleavage of RNA hairpins mediated by a developmentally regulated CCH zinc finger protein. *Molecular and Cellular Biology* **16**: 6661.
- Balagopal V, Fluch L, Nissan T. 2012. Ways and means of eukaryotic mRNA decay. Vol 1819, pp. 593-603, New York.
- Balagopal V, Parker R. 2011. Stm1 modulates translation after 80S formation in *Saccharomyces cerevisiae*. *RNA (New York, NY)* **17**: 835.
- Barabino S, Huebner W, Jenny A, Minvielle-Sebastia L, Keller W. 1997. The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes & Development* **11**: 1703-1716.
- Barabino SML, Ohnacker M, Keller W. 2000. Distinct roles of two Yth1p domains in 3' - end cleavage and polyadenylation of yeast pre-mRNAs. *EMBO Journal* **19**: 3778-3787.
- Barford D. 2004. The role of cysteine residues as redox- sensitive regulatory switches. *Current Opinion in Structural Biology* **14**: 679-686.
- Barreau C, Paillard L, Osborne HB. 2005. AU-rich elements and associated factors: are there unifying principles? *Nucleic acids research* **33**: 7138.
- Bäurle I, Dean C. 2008. Differential Interactions of the Autonomous Pathway RRM Proteins and Chromatin Regulators in the Silencing of Arabidopsis Targets (RRM- Proteins and Chromatin). *PLoS ONE* **3**: e2733.
- Bazan J, Całkosiński I, Gamian A. 2012. Phage display--a powerful technique for immunotherapy: 2. Vaccine delivery. *Human vaccines & immunotherapeutics* **8**: 1829.
- Bechtold U, Penfold CA, Jenkins DJ, Legaie R, Moore JD, Lawson T, Matthews JSA, Violet-Chabrand SRM, Baxter L, Subramaniam S et al. 2016. Time-series transcriptomics reveals that AGAMOUS-LIKE22 affects primary metabolism and developmental processes in drought-stressed Arabidopsis *The Plant cell* **28**: 345.
- Becker T, Bhushan S, Jarasch A, Armache J-P, Funes S, Jossinet F, Gumbart J, Mielke T, Berninghausen O, Schulten K et al. 2009. Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science (New York, NY)* **326**: 1369.

- Behm-Ansmant I, Gatfield D, Rehwinkel J, Hilgers V, Izaurralde E. 2007. A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense - mediated mRNA decay. *EMBO Journal* **26**: 1591-1601.
- Bell SA, Shen C, Brown A, Hunt AG. 2016. Experimental Genome-Wide Determination of RNA Polyadenylation in *Chlamydomonas reinhardtii*. *PLoS One* **11**: e0146107.
- Belostotsky DA, Rose AB. 2005. Plant gene expression in the age of systems biology: integrating transcriptional and post-transcriptional events. *Trends in Plant Science* **10**: 347-353.
- Bionaz M, Loor JJ. 2011. Gene Networks Driving Bovine Mammary Protein Synthesis During the Lactation Cycle. *Bioinformatics and Biology Insights* **5**: 83-98.
- Bousquet-Antonelli C, Presutti C, Tollervey D. 2000. Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell* **102**: 765-775.
- Bracken CP, Szubert JM, Mercer TR, Dinger ME, Thomson DW, Mattick JS, Michael MZ, Goodall GJ. 2011. Global analysis of the mammalian RNA degradome reveals widespread miRNA- dependent and miRNA-independent endonucleolytic cleavage. *Nucleic acids research* **39**: 5658.
- Bray E. 2004. Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana*. *Journal of Experimental Botany* **55**: 2331-2341.
- Bruce SR, Dingle RWC, Peterson ML. 2003. B-cell and plasma-cell splicing differences: a potential role in regulated immunoglobulin RNA processing. *RNA (New York, NY)* **9**: 1264.
- Bruggeman Q, Garmier M, de Bont L, Soubigou-Taconnat L, Mazubert C, Benhamed M, Raynaud C, Bergounioux C, Delarue M. 2014. The Polyadenylation Factor Subunit CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR30: A Key Factor of Programmed Cell Death and a Regulator of Immunity in *Arabidopsis*. *Plant physiology* **165**: 732.
- Cao D, Parker R. 2001. Computational modeling of eukaryotic mRNA turnover. *RNA (New York, NY)* **7**: 1192.
- Carballo E, Blackshear PJ. 2001. Roles of tumor necrosis factor-alpha receptor subtypes in the pathogenesis of the tristetraprolin- deficiency syndrome. *Blood* **98**: 2389.
- Chakrabarti M, Hunt AG. 2015. CPSF30 at the Interface of Alternative Polyadenylation and Cellular Signaling in Plants. *Biomolecules* **5**: 1151.
- Chakrabarty R, Banerjee R, Chung S-M, Farman M, Citovsky V, Hogenhout SA, Tzfira T, Goodin M. 2007. PSITE vectors for stable integration or transient expression of autofluorescent protein fusions in plants: probing *Nicotiana benthamiana*-virus interactions. *Molecular plant-microbe interactions : MPMI* **20**: 740.
- Chan S, Choi E-A, Shi Y. 2011. Pre-mRNA 3'-end processing complex assembly and function. *Wiley interdisciplinary reviews RNA* **2**: 321.
- Chan SL, Huppertz I, Yao C, Weng L, Moresco JJ, Yates JR, Ule J, Manley JL, Shi Y. 2014. CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes & development* **28**: 2370.

- Chang Y-F, Imam JS, Wilkinson MF. 2007. The nonsense-mediated decay RNA surveillance pathway. *76*: 51-74.
- Chen C-YA, Shyu A-B. 2003. Rapid deadenylation triggered by a nonsense codon precedes decay of the RNA body in a mammalian cytoplasmic nonsense-mediated decay pathway. *Molecular and Cellular Biology* **23**: 4805.
- Chen I-J, Wang IA, Tai L-R, Lin A. 2008. The role of expansion segment of human ribosomal protein L35 in nuclear entry, translation activity, and endoplasmic reticulum docking. *Biochemistry and Cell Biology* **86**: 271-277.
- Chen T, Nayak N, Majee SM, Lowenson J, Schäfermeyer KR, Eliopoulos AC, Lloyd TD, Dinkins R, Perry SE, Forsthoefel NR et al. 2010. Substrates of the *Arabidopsis thaliana* protein isoaspartyl methyltransferase 1 identified using phage display and biopanning. *The Journal of biological chemistry* **285**: 37281.
- Chen W, Guo W, Li M, Shi D, Tian Y, Li Z, Wang J, Fu L, Xiao X, Liu QQ et al. 2013. Upregulation of cleavage and polyadenylation specific factor 4 in lung adenocarcinoma and its critical role for cancer cell survival and proliferation. *PLoS one* **8**: e82728.
- Chen WB, Qin, L.J., Wang, S.S., Li, M., Shi, D.B., Tian, Y., Wang, J.S., Fu, L.Y., Li, Z.L., Guo, W., et al. 2014. CPSF4 activates telomerase reverse transcriptase and predicts poor prognosis in human lung adenocarcinomas. *Molecular Oncology* **8**: 704-716.
- Cheng Y, Kato N, Wang W, Li J, Chen X. 2003. Two RNA binding proteins, HEN4 and HUA1, act in the processing of AGAMOUS pre-mRNA in *Arabidopsis thaliana*. *Developmental Cell* **4**: 53-66.
- Chinnusamy V, Zhu J, Zhu J-K. 2007. Cold stress regulation of gene expression in plants. *Trends in Plant Science* **12**: 444-451.
- Colgan D, Manley J. 1997. Mechanism and regulation of mRNA polyadenylation. Vol 11 (ed. D Colgan), pp. 2755-2766.
- Cui P, Zhang S, Ding F, Ali S, Xiong L. 2014. Dynamic regulation of genome-wide pre-mRNA splicing and stress tolerance by the Sm-like protein LSM5 in *Arabidopsis*. *Genome Biology* **15**: R1-R1.
- Danckwardt S, Hentze MW, Kulozik AE. 2008. 3' end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO Journal* **27**: 482-498.
- Danckwardt S, Kaufmann I, Gentzel M, Foerstner KU, Gantzer AS, Gehring NH, Neu-Yilik G, Bork P, Keller W, Wilm M et al. 2007. Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals. *EMBO Journal* **26**: 2658-2669.
- Daniel RS, Lynne EM. 2012. Regulation of cytoplasmic mRNA decay. *Nature Reviews Genetics* **13**: 448.
- David G, Elisa I. 2004. Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in *Drosophila*. *Nature* **429**: 575.
- Dávila López M, Samuelsson T. 2008. Early evolution of histone mRNA 3' end processing. *RNA (New York, NY)* **14**: 1.
- de Lorenzo L, Sorenson R, Bailey-Serres J, Hunt AG. 2017. Noncanonical alternative polyadenylation contributes to gene regulation in response to hypoxia. *The Plant Cell* **in press**.

- Decker CJ, Parker R. 1993. A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes & development* **7**: 1632.
- Delaney K, Xu R, Zhang J. 2006. Calmodulin Interacts with and Regulates the RNA-Binding Activity of an Arabidopsis Polyadenylation Factor Subunit1[OA]. *Plant Physiology* **140**: 1507-1521.
- Derti A, Garrett-Engle P, Macisaac K, Stevens R, Sriram S, Chen R, Rohl C, Johnson J, Babak T. 2012. A quantitative atlas of polyadenylation in five mammals. *Genome Research* **22**: 1173-1183.
- Di Giammartino Dafne C, Nishida K, Manley James I. 2011. Mechanisms and Consequences of Alternative Polyadenylation. *Molecular Cell* **43**: 853-866.
- Donny DL, Robert BD. 2010. RNA processing and its regulation: global insights into biological networks. *Nature Reviews Genetics* **11**: 75.
- Duc C, Sherstnev A, Cole C, Barton GJ, Simpson GG. 2013. Transcription termination and chimeric RNA formation controlled by *Arabidopsis thaliana* FPA **9**: e1003867.
- Durand S, Lykke-Andersen J. 2011. SnapShot: nonsense-mediated mRNA decay. *Cell* **145**: 324-324.e322.
- Eberle AB, Stalder L, Mathys H, Orozco RZ, Mühlemann O. 2008. Posttranscriptional gene regulation by spatial rearrangement of the 3' untranslated region *PLoS Biology* **6**: e92.
- Edwalds-Gilbert G, Milcarek C. 1995. Regulation of poly(A) site use during mouse B-cell development involves a change in the binding of a general polyadenylation factor in a B-cell stage-specific manner. *Molecular and Cellular Biology* **15**: 6420.
- Elkon R, Drost J, van Haaften G, Jenal M, Schrier M, Vrieling JAO, Agami R. 2012. E2F mediates enhanced alternative polyadenylation in proliferation. *Genome biology* **13**: 2895.
- Enid ACM. 2002. Evidence for a Role for Protein Tyrosine Phosphatase in the Control of Ion Release from the Guard Cell Vacuole in Stomatal Closure. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 11963-11968.
- Evans NH, McAinsh MR, Hetherington AM, Knight MR. 2005. ROS perception in *Arabidopsis thaliana* : the ozone - induced calcium response. *Plant Journal* **41**: 615-626.
- Farooq M, Wahid A, Kobayashi N, Fujita D, Basra SMA. 2009a. Plant drought stress: effects, mechanisms and management. *Agronomy for sustainable development* **29**: 185-212.
- Farooq M, Wahid A, Lee D-J. 2009b. Exogenously applied polyamines increase drought tolerance of rice by improving leaf water status, photosynthesis and membrane properties. *Acta Physiol Plant* **31**: 937-945.
- Fátima G, Matthias WH. 2004. Molecular mechanisms of translational control. *Nature Reviews Molecular Cell Biology* **5**: 827.

- Felice-Alessio B, Carolina E, Pedro GF, Belen M, Claudia B-D, Roderic G, Juan V, Raúl M. 2013. CPEB1 coordinates alternative 3' -UTR formation with translational regulation. *Nature* **495**: 121.
- Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, Fox SE, Wong W-K, Mockler TC. 2010. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome research* **20**: 45.
- Flavell SW, Greenberg ME. 2008. Signaling mechanisms linking neuronal activity to gene expression and plasticity of the nervous system. **31**: 563-590.
- Flavell SW, Kim T-K, Gray JM, Harmin DA, Hemberg M, Hong EJ, Markenscoff-Papadimitriou E, Bear DM, Greenberg ME. 2008. Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron* **60**: 1022-1038.
- Fox MJ, Mosley AL. 2016. Rrp6: Integrated roles in nuclear RNA metabolism and transcription termination. *Wiley Interdiscip Rev RNA* **7**: 91-104.
- Foyer CH, Mullineaux PM. 1994. *Causes of photooxidative stress and amelioration of defense systems in plants*. Boca Raton, FL : CRC Press, Boca Raton, FL.
- Fracasso A, Trindade LM, Amaducci S. 2016. Drought stress tolerance strategies revealed by RNA-Seq in two sorghum genotypes with contrasting WUE. *BMC Plant Biology* **16**.
- Franks TM, Singh G, Lykke-Andersen J. 2010. Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense- mediated mRNA decay. *Cell* **143**: 938-950.
- Frischmeyer PA, van Hoof A, O'Donnell K, Guerrerio AL, Parker R, Dietz HC. 2002. An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* **295**: 2258-2261.
- Fu H, Yang D, Su W, Ma L, Shen Y, Ji G, Ye X, Wu X, Li QQ. 2016. Genome-wide dynamics of alternative polyadenylation in rice. *Genome research* **26**: 1753.
- Gandhi R, Manzoor M, Hudak KA. 2008. Depurination of Brome mosaic virus RNA3 in vivo results in translation- dependent accelerated degradation of the viral RNA. *The Journal of biological chemistry* **283**: 32218.
- Gapper C, Dolan L. 2006. Control of plant development by reactive oxygen species. *Plant Physiology* **141**: 341-345.
- Gautheret D, Poirot O, Lopez F, Audic S, Claverie JM. 1998. Alternate polyadenylation in human mRNAs: A large- scale analysis by EST clustering. *Genome Research* **8**: 524-530.
- Gehring NH, Neu-Yilik G, Schell T, Hentze MW, Kulozik AE. 2003. Y14 and hUpf3b form an NMD-activating complex. *Mol Cell* **11**: 939-949.
- Ghelis T, Bolbach G, Clodic G, Habricot Y, Miginiac E, Sotta B, Jeannette E. 2008. Protein Tyrosine Kinases and Protein Tyrosine Phosphatases Are Involved in Abscisic Acid-Dependent Processes in Arabidopsis Seeds and Suspension Cells1. *Plant Physiology* **148**: 1668-1680.
- Giranton JL, Ariza MJ, Dumas C, Cock JM, Gaude T. 1995. The S locus receptor kinase gene encodes a soluble glycoprotein corresponding to the SRK extracellular domain in Brassica oleracea. *Plant Journal* **8**: 827-834.

- González CI, Ruiz-Echevarría MJ, Vasudevan S, Henry MF, Peltz SW. 2000. The Yeast hnRNP-like Protein Hrp1/ Nab4 marks a transcript for nonsense-mediated mRNA decay. *Molecular Cell* **5**: 489-499.
- Goodin M, Chakrabarty R, Yelton S. 2008. Membrane and protein dynamics in virus-infected plant cells. *Methods in molecular biology (Clifton, NJ)* **451**: 377.
- Graber JH, Nazeer FI, Yeh P-C, Kuehner JN, Borikar S, Hoskinson D, Moore CL. 2013. DNA damage induces targeted, genome-wide variation of poly(A) sites in budding yeast. *Genome research* **23**: 1690.
- Graille M, Chaillet M, van Tilbeurgh H. 2008. Structure of yeast Dom34: a protein related to translation termination factor Erf1 and involved in No-Go decay. *Journal of biological chemistry* **283**: 7145-7154.
- Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, Mello CC. 2001. Genes and Mechanisms Related to RNA Interference Regulate Expression of the Small Temporal RNAs that Control *C. elegans* Developmental Timing. *Cell* **106**: 23-34.
- Gruber AR, Martin G, Keller W, Zavolan M. 2014. Means to an end: mechanisms of alternative polyadenylation of messenger RNA precursors. *Wiley Interdiscip Rev RNA* **5**: 183-196.
- Guan Q, Zheng W, Tang S, Liu X, Zinkel RA, Tsui K-W, Yandell BS, Culbertson MR. 2006. Impact of Nonsense-Mediated mRNA Decay on the Global Expression Profile of Budding Yeast (Global Impact of NMD in Yeast). *PLoS Genetics* **2**: e203.
- Guhaniyogi J, Brewer G. 2001. Regulation of mRNA stability in mammalian cells. *Gene* **265**: 11-23.
- Günter K, Daniel B, Nenad B, Bernd B. 2009. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nature Structural & Molecular Biology* **16**: 589.
- Guo J, Garrett M, Micklem G, Brogna S. 2011. Poly(A) signals located near the 5' end of genes are silenced by a general mechanism that prevents premature 3'-end processing. *Molecular and Cellular Biology* **31**: 639.
- Gupta I, Clauder - Münster S, Klaus B, Järvelin AI, Aiyar RS, Benes V, Wilkening S, Huber W, Pelechano V, Steinmetz LM. 2014. Alternative polyadenylation diversifies post - transcriptional regulation by selective RNA - protein interactions. *Molecular Systems Biology* **10**: n/a-n/a.
- Gutierrez R, Ewing R, Cherry J, Green P. 2002. Identification of unstable transcripts in Arabidopsis by cDNA microarray analysis: Rapid decay is associated with. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 11513-11518.
- Hajdуч M, Hearne LB, Miernyk JA, Casteel JE, Joshi T, Agrawal GK, Song Z, Zhou M, Xu D, Thelen JJ. 2010. Systems analysis of seed filling in Arabidopsis: using general linear modeling to assess concordance of transcript and protein expression. *Plant physiology* **152**: 2078.
- Hall-Pogar T, Liang S, Hague LK, Lutz CS. 2007. Specific trans-acting proteins interact with auxiliary RNA polyadenylation elements in the COX-2 3'-UTR. *RNA (New York, NY)* **13**: 1103.

- Harb A, Krishnan A, Ambavaram M, Pereira A. 2010. Molecular and Physiological Analysis of Drought Stress in Arabidopsis Reveals Early Responses Leading to Acclimation in Plant Growth1[C][W][OA]. *Plant Physiology* **154**: 1254-1271.
- Hentze M. 1997. eIF4G: A multipurpose ribosome adapter? , Vol 275 (ed. M Hentze), pp. 500-501.
- Hilgers V, Perry MW, Hendrix D, Stark A, Levine M, Haley B. 2011. Neural-specific elongation of 3' UTRs during Drosophila development. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 15864.
- Hirayama T, Shinozaki K. 2007. Perception and transduction of abscisic acid signals: keys to the function of the versatile plant hormone ABA. *Trends in plant science* **12**: 343-351.
- Hogg JR, Goff SP. 2010. Upf1 senses 3' UTR length to potentiate mRNA decay. *Cell* **143**: 379-389.
- Holbrook LA, Van Rooijen GJH, Wilen RW, Moloney MM. 1991. Oilbody proteins in microspore-derived embryos of Brassica napus. *Plant physiology*: 1051-1058.
- Hollerer I, Curk T, Haase B, Benes V, Hauer C, Neu-Yilik G, Bhuvanagiri M, Hentze MW, Kulozik AE. 2016. The differential expression of alternatively polyadenylated transcripts is a common stress-induced response mechanism that modulates mammalian mRNA expression in a quantitative and qualitative fashion. *RNA (New York, NY)* **22**: 1441.
- Hollerer I, Grund K, Hentze MW, Kulozik AE. 2014. mRNA 3' end processing: A tale of the tail reaches the clinic. Vol 6, pp. 16-26.
- Hori K, Watanabe Y. 2007. Context Analysis of Termination Codons in mRNA that are Recognized by Plant NMD. *Plant and Cell Physiology* **48**: 1072-1078.
- Hornyik C, Terzi LC, Simpson GG. 2010. The spen family protein FPA controls alternative cleavage and polyadenylation of RNA. *Developmental Cell* **18**: 203-213.
- Hsu CL, Stevens A. 1993. Yeast cells lacking 5'-->3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. *Molecular and Cellular Biology* **13**: 4826.
- Huang D, Wu W, Abrams SR, Cutler AJ. 2008. Trelationship of drought-related gene expression in *Arabidopsis thaliana* to hormonal and environmental factors. *Journal of experimental botany* **59**: 2991-3007.
- Hunt AG. 2012. RNA regulatory elements and polyadenylation in plants. In *Front Plant Sci*, Vol 2.
- Hunt AG. 2014. The Arabidopsis polyadenylation factor subunit CPSF30 as conceptual link between mRNA polyadenylation and cellular signaling. *Current opinion in plant biology* **21**: 128.
- Hunt AG, Xing D, Li QQ. 2012. Plant polyadenylation factors: conservation and variety in the polyadenylation complex in plants. *BMC genomics* **13**: 641.
- Hunt AG, Xu R, Addepalli B, Rao S, Forbes KP, Meeks LR, Xing D, Mo M, Zhao H, Bandyopadhyay A et al. 2008. Arabidopsis mRNA polyadenylation machinery: comprehensive analysis of protein- protein interactions and gene expression profiling. *BMC Genomics* **9**: 220-220.

- Hwang J, Maquat LE. 2011. Nonsense-mediated mRNA decay(NMD) in animal embryogenesis: to die or not to die, that is the question. *Current Opinion in Genetics & Development* doi:10.1016/j.gde.2011.03.008.
- Ihnatowicz A, Pesaresi P, Varotto C, Richly E, Schneider A, Jahns P, Salamini F, Leister D. 2004. Mutants for photosystem I subunit D of *Arabidopsis thaliana*: effects on photosynthesis, photosystem I stability and expression of nuclear genes for chloroplast functions. *Plant Journal* **37**: 839-852.
- Iida K, Seki M, Sakurai T, Satou M, Akiyama K, Toyoda T, Konagaya A, Shinozaki K. 2004. Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences. *Nucleic acids research* **32**: 5096.
- Inada T, Aiba H. 2005. Translation of aberrant mRNAs lacking a termination codon or with a shortened 3' - UTR is repressed after initiation in yeast. *EMBO Journal* **24**: 1584-1595.
- Ishigame H, Mosaheb MM, Sanjabi S, Flavell RA. 2013. Truncated form of transforming growth factor- β receptor II, but not its absence, induces memory CD8⁺ T cell expansion and lymphoproliferative disorder in mice. Vol 63, pp. 273-273.
- Isken O, Kim YK, Hosoda N, Mayeur GL, Hershey JWB, Maquat LE. 2008. Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. *Cell* **133**: 314-327.
- Isken O, Maquat LE. 2007. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes & development* **21**: 1833.
- Ivanov PV, Gehring NH, Kunz JB, Hentze MW, Kulozik AE. 2008. Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *EMBO Journal* **27**: 736-747.
- Jacobson A, Peltz SW. 1996. Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. *Annual Review of Biochemistry* **65**: 693-739.
- Jean-François L, Marc L, Samuel M, Sophie A, Jürg B, François B. 2014. The RNA exosome promotes transcription termination of backtracked RNA polymerase II. *Nature Structural & Molecular Biology* doi:10.1038/nsmb.2893.
- Jensen TH, Boulay J, Rosbash M, Libri D. 2001a. The DECD box putative ATPase Sub2p is an early mRNA export factor. *Current Biology* **11**: 1711-1715.
- Jensen TH, Patricio K, McCarthy T, Rosbash M. 2001b. A block to mRNA nuclear export in *S. cerevisiae* leads to hyperadenylation of transcripts that accumulate at the site of transcription. *Molecular Cell* **7**: 887-898.
- Ji Z, Lee JY, Pan Z, Jiang B, Tian B. 2009. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 7028.
- Jiang N, Hu L, Liu C, Gao X, Zheng S. 2015. 60S ribosomal protein L35 regulates β -casein translational elongation and secretion in bovine mammary epithelial cells. *Archives of Biochemistry and Biophysics* **583**: 130-139.

- Job C, Rajjou L, Lovigny Y, Belghazi M, Job D. 2005. Patterns of Protein Oxidation in Arabidopsis Seeds and during Germination1[w]. *Plant Physiology* **138**: 790-802.
- Johansson MJO, He F, Spatrack P, Li C, Jacobson A. 2007. Association of yeast Upf1p with direct substrates of the NMD pathway. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 20872.
- Johnson M, Perez-Amador M, Lidder P, Green P. 2000. Mutants of Arabidopsis defective in a sequence-specific mRNA degradation pathway. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 13991-13996.
- Jones CI, Zabolotskaya MV, Newbury SF. 2012. The 5'→3' exoribonuclease XRN1/Pacman and its functions in cellular processes and development. *Wiley interdisciplinary reviews RNA* **3**: 455.
- Joshi R, Wani SH, Singh B, Bohra A, Dar ZA, Lone AA, Pareek A, Singla-Pareek SL. 2016. Transcription Factors and Plants Response to Drought Stress: Current Understanding and Future Directions. *Frontiers in Plant Science* **7**.
- Joshua TM, Neda AS, Jennifer LM, Francisco M-M, Harry CD. 2004. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nature Genetics* **36**: 1073.
- Julian IS, June MK, Gethyn JA. 2001. Guard cell abscisic acid signalling and engineering drought hardiness in plants. *Nature* **410**: 327.
- Kalya M, Lopato S, Barta A. 2003. Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development. *Molecular biology of the cell* **14**: 3565.
- Kalya M, Lopato S, Voronin V, Barta A. 2006. Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins. *Nucleic acids research* **34**: 4395.
- Kalya M, Simpson CG, Syed NH, Lewandowska D, Marquez Y, Kusenda B, Marshall J, Fuller J, Cardle L, McNicol J et al. 2012. Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic acids research* **40**: 2454.
- Kathleen M, Serkan AA, Alys C, Lisa B, Christine M. 2009. Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. *Nature Immunology* **10**: 1102.
- Kaufmann I, Martin G, Friedlein A, Langen H, Keller W. 2004. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO Journal* **23**: 616-626.
- Keddie JSDoBaOR, Huebner G, Slocombe SP, Jarvis RP, Cummins I, Edwards E, Shaw CH, Murphy DJ. 1992. Cloning and characterisation of an oleosin gene from Brassica napus. *Plant Molecular Biology*: 443-453.
- Keller T, Damude HG, Werner D, Doerner P, Dixon RA, Lamb C. 1998. A plant homolog of the neutrophil NADPH oxidase gp91phox subunit gene encodes a plasma membrane protein with Ca²⁺ binding motifs. *The Plant cell* **10**: 255.

- Keller W, Bienroth S, Lang K, Christofori G. 1991. Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. *Embo Journal* **10**: 4241-4249.
- Kertész S, Kerényi Z, Mérai Z, Bartos I, Pálffy T, Barta E, Silhavy D. 2006. Both introns and long 3'-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants. *Nucleic acids research* **34**: 6147.
- Kim E-D, Chen ZJ, Hazen SP. 2011. Unstable transcripts in Arabidopsis allotetraploids are associated with nonadditive gene expression in response to abiotic and biotic stresses. *PLoS ONE* **6**.
- Kim T-H, Böhmer M, Hu H, Nishimura N, Schroeder JI. 2010. Guard cell signal transduction network: advances in understanding abscisic acid, CO₂, and Ca²⁺ signaling. *Annual review of plant biology* **61**: 561-591.
- Kim V, Kataoka N, Dreyfuss G. 2001. Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent Exon-Exon junction complex. *Science* **293**: 1832-1836.
- Klauer AA, van Hoof A. 2012. Degradation of mRNAs that lack a stop codon: a decade of nonstop progress. *Wiley interdisciplinary reviews RNA* **3**: 649.
- Klein DJ, Moore PB, Steitz TA. 2004. The Roles of Ribosomal Proteins in the Structure Assembly, and Evolution of the Large Ribosomal Subunit. *Journal of Molecular Biology* **340**: 141-177.
- Knight S, Bass B. 2001. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science (Washington)* **293**: 2269-2271.
- Kobayashi K, Kikuno I, Kuroha K, Saito K, Ito K, Ishitani R, Inada T, Nureki O. 2010. Structural basis for mRNA surveillance by archaeal Pelota and GTP-bound EF1[alpha] complex. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 17575.
- Konishi H, Kitano H, Komatsu S. 2005. Identification of rice root proteins regulated by gibberellin using proteome analysis. *Plant, Cell & Environment* **28**: 328-339.
- Krasensky J, Jonak C. 2012. Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *Journal of Experimental Botany* **63**: 1593-1608.
- Ksouri N, Jiménez S, Wells CE, Contreras-Moreira B, Gogorcena Y. 2016. Transcriptional responses in root and leaf of *Prunus persica* under drought stress using RNA sequencing. *Frontiers in plant science* **7**: 1715.
- Kuai L, Das B, Sherman F. 2005. A nuclear degradation pathway controls the abundance of normal mRNAs in *Saccharomyces cerevisiae*. In *nuclear degradation pathway controls the abundance of normal mRNAs in Saccharomyces cerevisiae*, Vol 102, pp. 13962-13967.
- Kuroha K, Tatematsu T, Inada T. 2009. Upf1 stimulates degradation of the product derived from aberrant messenger RNA containing a specific nonsense mutation by the proteasome. *EMBO reports* **10**: 1265-1271.
- Kushwaha R, Schäfermeyer KR, Downie AB. 2014. A protocol for phage display and affinity selection using recombinant protein baits. *Journal of visualized experiments : JoVE* doi:10.3791/50685: e50685.

- Lacava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D. 2005. RNA Degradation by the Exosome Is Promoted by a Nuclear Polyadenylation Complex. *Cell* **121**: 713-724.
- Lareau LF, Brooks AN, Soergel DAW, Meng Q, Brenner SE. 2007. The coupling of alternative splicing and nonsense-mediated mRNA decay.
- Lazar G, Goodman H. 2000. The Arabidopsis splicing factor SR1 is regulated by alternative splicing. *Plant Mol Biol* **42**: 571-581.
- Le Hir H, Gatfield D, Izaurralde E, Moore MJ. 2001. The exon - exon junction complex provides a binding platform for factors involved in mRNA export and nonsense - mediated mRNA decay. *EMBO Journal* **20**: 4987-4997.
- Lee Y, Dutta A. 2007. The tumor suppressor microRNA let- 7 represses the HMGA2 oncogene. *Genes & Development* **21**: 1025-1030.
- Leeds P, Peltz SW, Jacobson A, Culbertson MR. 1991. The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes & development* **5**: 2303.
- Leeds P, Wood JM, Lee BS, Culbertson MR. 1992. Gene products that promote mRNA turnover in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **12**: 2165.
- Legendre M, Ritchie W, Lopez F, Gautheret D. 2006. Differential repression of alternative transcripts: a screen for miRNA targets. *PLoS Computational Biology* **2**: e43.
- Lerner HR, NetLibrary I. 1999. *Plant responses to environmental stresses : from phytohormones to genome reorganization*. New York : M. Dekker, New York.
- Li Q, Hunt AG. 1997. The polyadenylation of RNA in plants. *Plant Physiology* **115**: 321-325.
- Li Y, Kiledjian M. 2010. Regulation of mRNA decapping. *Wiley Interdisciplinary Reviews - RNA* **1**: 253-265.
- Libri D, Dower K, Boulay J, Thomsen R, Rosbash M, Jensen TH. 2002. Interactions between mRNA Export Commitment, 3'- End Quality Control, and Nuclear Degradation. *Molecular and Cellular Biology* **22**: 8254.
- Lin Y, Li Z, Ozsolak F, Kim SW, Arango-Argoty G, Liu TT, Tenenbaum SA, Bailey T, Monaghan AP, Milos PM et al. 2012. An in- depth map of polyadenylation sites in cancer. *Nucleic acids research* **40**: 8460.
- Linz B, Koloteva N, Vasilescu S, McCarthy JEG. 1997. Disruption of ribosomal scanning on the 5'-untranslated region, and not restriction of translational initiation per se, modulates the stability of nonaberrant mRNAs in the yeast *Saccharomyces cerevisiae*. *The Journal of biological chemistry*: 9131-9140.
- Liu F, Marquardt S, Lister C, Swiezewski S, Dean C. 2010. Targeted 3' processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. *Science* **327**: 94-97.
- Liu F, Quesada V, Crevillén P, Bäurle I, Swiezewski S, Dean C. 2007. The Arabidopsis RNA-Binding Protein FCA Requires a Lysine- Specific Demethylase 1 Homolog to Downregulate FLC. *Molecular Cell* **28**: 398-407.

- Liu M, Xu R, Merrill C, Hong L, Von Lanken C, Hunt AG, Li QQ, He Y. 2014. Integration of Developmental and Environmental Signals via a Polyadenylation Factor in Arabidopsis. *PLoS ONE* **9**.
- Loke J, Stahlberg E, Strenski D, Haas B. 2005. Compilation of mRNA Polyadenylation Signals in Arabidopsis Revealed a New Signal Element and Potential Secondary Structures1[w]. *Plant Physiology* **138**: 1457-1468.
- Lopato S, Kalyna M, Dorner S, Kobayashi R, Krainer AR, Barta A. 1999. atSRp30, one of two SF2/ ASF- like proteins from *Arabidopsis thaliana*, regulates splicing of specific plant genes. *Genes & development* **13**: 987.
- Losson R, Lacroute F. 1979. Interference of nonsense mutations with eukaryotic messenger RNA stability [Saccharomyces cerevisiae, yeast]. In *Interference of nonsense mutations with eukaryotic messenger RNA stability [Saccharomyces cerevisiae, yeast]*, Vol 76, pp. 5134-5137.
- Lou H, Neugebauer KM, Gagel RF, Berget SM. 1998. Regulation of alternative polyadenylation by U1 snRNPs and SRp20. *Molecular and Cellular Biology* **18**: 4977.
- Luan S. 2002. Signalling drought in guard cells. *Plant, Cell & Environment* **25**: 229-237.
- Lutz CS, Moreira A. 2011. Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. *Wiley interdisciplinary reviews RNA* **2**: 23.
- Lykke-Andersen J, Bennett EJ. 2014. Protecting the proteome: Eukaryotic cotranslational quality control pathways. *The Journal of cell biology* **204**: 467.
- Lykke-Andersen J, Steitz J. 2001. Communication of the position of Exon-Exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* **293**: 1836-1839.
- Lykke-Andersen S, Tomecki R, Jensen TH, Dziembowski A. 2011. The eukaryotic RNA exosome: same scaffold but variable catalytic subunits. *RNA biology* **8**: 61.
- Ma L, Guo C, Li QQ. 2014a. Role of alternative polyadenylation in epigenetic silencing and antisilencing. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 9.
- Ma L, Pati PK, Liu M, Li QQ, Hunt AG. 2014b. High throughput characterizations of poly(A) site choice in plants. *Methods* **67**: 74-83.
- Ma S, Bohnert HJ. 2007. Integration of *Arabidopsis thaliana* stress-related transcript profiles, promoter structures, and cell-specific expression. *Genome Biology* **8**: R49-R49.
- Macdonald PM, Smibert CA. 1996. Translational regulation of maternal mRNAs. *Current Opinion in Genetics & Development* **6**: 403-407.
- Macknight R, Duroux M, Laurie R, Dijkwel P, Simpson G, Dean C. 2002. Functional Significance of the Alternative Transcript Processing of the Arabidopsis Floral Promoter FCA. *The Plant Cell* **14**: 877-888.
- Maike CR, David L, Fatma O, Sarah LU, Lindsay P, Haruko O, Heather K, Scott CP, Claire SG, Heribert H et al. 2004. OXI1 kinase is necessary for oxidative burst-mediated signalling in Arabidopsis. *Nature* **427**: 858.

- Mainul H, Zhe J, Dinghai Z, Wenting L, Wencheng L, Bei Y, Ji Yeon P, Ghassan Y, Bin T. 2012. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nature Methods* **10**: 133.
- Mandel C, Bai Y, Tong L. 2008. Protein factors in pre-mRNA 3' - end processing. *Cell Mol Life Sci* **65**: 1099-1122.
- Mangus DA, Evans MC, Jacobson A. 2003. Poly(A)-binding proteins: multifunctional scaffolds for the post- transcriptional control of gene expression. *Genome Biology* **4**: 223-223.
- Maquat LE, Kinniburgh AJ, Rachmilewitz EA, Ross J. 1981. Unstable beta-globin mRNA in mRNA- deficient beta o thalassemia. *Cell* **27**: 543.
- Marnef A, Standart N. 2010. Pat1 proteins: a life in translation, translation repression and mRNA decay. *Biochemical Society transactions* **38**: 1602.
- Matsuda R, Ikeuchi K, Nomura S, Inada T. 2014. Protein quality control systems associated with no - go and nonstop mRNA surveillance in yeast. *Genes to Cells* **19**: 1-12.
- Maurya JP, Sethi V, Gangappa SN, Gupta N, Chattopadhyay S. 2015. Interaction of MYC2 and GBF1 results in functional antagonism in blue light - mediated Arabidopsis seedling development. *Plant Journal* **83**: 439-450.
- Mayr C, Bartel DP. 2009. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**: 673-684.
- Mayr C, Hemann MT, Bartel DP. 2007. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science (New York, NY)* **315**: 1576.
- McGlinchy NJ, Smith CWJ. 2008. Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends in Biochemical Sciences* **33**: 385-393.
- Meaux S, Van Hoof A. 2006. Yeast transcripts cleaved by an internal ribozyme provide new insight into the role of the cap and poly(A) tail in translation and mRNA decay. *RNA (New York, NY)* **12**: 1323.
- Meaux S, van Hoof A, Baker KE. 2008. Nonsense-mediated mRNA decay in yeast does not require PAB1 or a poly(A) tail. *Molecular Cell* **29**: 134-140.
- Meenakshi KD, Roy P. 2006. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* **440**: 561.
- Melton AA, Jackson J, Wang J, Lynch KW. 2007. Combinatorial control of signal-induced exon repression by hnRNP L and PSF. *Molecular and cellular biology* **27**: 6972.
- Meyers BC, Vu TH, Tej SS, Ghazal H, Matvienko M, Agrawal V, Ning JC, Haudenschild CD. 2004. Analysis of the transcriptional complexity of *Arabidopsis thaliana* by massively parallel signature sequencing. *Nature biotechnology* **22**: 1006-1011.
- Miller G, Suzuki N, Ciftci - Yilmaz S, Mittler R. 2010. Reactive oxygen species homeostasis and signalling during drought and salinity stresses. *Plant, Cell & Environment* **33**: 453-467.
- Millevoi S, Vagner S. 2010. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic acids research* **38**: 2757.

- Mitchell P, Tollervey D. 2000. mRNA stability in eukaryotes. *Current Opinion in Genetics & Development* **10**: 193-198.
- Mitchell P, Tollervey D. 2003. An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'→5' degradation. *Molecular Cell* **11**: 1405-1413.
- Mittler R. 2002. Oxidative stress, antioxidants and stress tolerance. *Trends in Plant Science* **7**: 405-410.
- Mittler R, Vanderauwera S, Gollery M, Van Breusegem F. 2004. Reactive oxygen gene network of plants. *Trends in Plant Science* **9**: 490-498.
- Montoliu L, Rigau J, Puigdomènech P. 1990. Multiple polyadenylation sites are active in the α 1-tubulin gene from *Zea mays*. *FEBS Letters* **277**: 29-32.
- Moore MJ, Proudfoot NJ. 2009. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**: 688-700.
- Muhlrad D, Decker CJ, Parker R. 1994. Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'→3' digestion of the transcript. *Genes & development* **8**: 855.
- Muhlrad D, Parker R. 1999. Recognition of yeast mRNAs as "nonsense containing" leads to both inhibition of mRNA translation and mRNA degradation: implications for the control of mRNA decapping. *Molecular biology of the cell* **10**: 3971.
- Muhlrad D, Parker R. 2005. The yeast EDC1 mRNA undergoes deadenylation - independent decapping stimulated by Not2p, Not4p, and Not5p. *EMBO Journal* **24**: 1033-1045.
- Mukherjee C, Patil Deepak p, Kennedy Brian a, Bakthavachalu B, Bundschuh R, Schoenberg Daniel r. 2012. Identification of Cytoplasmic Capping Targets Reveals a Role for Cap Homeostasis in Translation and mRNA Stability. *Cell Reports* **2**: 674-684.
- Nemeroff ME, Barabino SML, Li Y, Keller W, Krug RM. 1998. Influenza Virus NS1 Protein Interacts with the Cellular 30 kDa Subunit of CPSF and Inhibits 3' End Formation of Cellular Pre- mRNAs. *Molecular Cell* **1**: 991-1000.
- Newbury SF. 2006. Control of mRNA stability in eukaryotes. *Biochemical Society transactions* **34**: 30.
- Newnham CM, Hall-Pogar T, Liang S, Wu J, Tian B, Hu J, Lutz CS. 2010. Alternative polyadenylation of MeCP2: Influence of cis-acting elements and trans-acting factors. *RNA biology* **7**: 361.
- Nissan T, Rajyaguru P, She M, Song H, Parker R. 2010. Decapping activators in *Saccharomyces cerevisiae* act by multiple mechanisms. *Molecular Cell* **39**: 773-783.
- Onouchi H, Nagami Y, Haraguchi Y, Nakamoto M, Nishimura Y, Sakurai R, Nagao N, Kawasaki D, Kadokura Y, Naito S. 2005. Nascent peptide-mediated translation elongation arrest coupled with mRNA degradation in the CGS1 gene of *Arabidopsis*. *Genes & development* **19**: 1799.
- Ozgur S, Chekulaeva M, Stoecklin G. 2010. Human Pat1b connects deadenylation with mRNA decapping and controls the assembly of processing bodies. *Molecular and Cellular Biology* **30**: 4308.

- Palusa SG, Ali GS, Reddy ASN. 2007. Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses. *Plant journal for cell and molecular biology* **49**: 1091-1107.
- Pandey N, Ranjan A, Pant P, Tripathi RK, Ateek F, Pandey HP, Patre UV, Sawant SV. 2013. CAMTA1 regulates drought responses in *Arabidopsis thaliana*. *BMC Genomics* **14**: 216-216.
- Parker R. 2012. RNA degradation in *Saccharomyces cerevisiae*. *Genetics* **191**: 671.
- Passos DO, Doma MK, Shoemaker CJ, Muhlrud D, Green R, Weissman J, Hollien J, Parker R. 2009. Analysis of Dom34 and its function in no-go decay. *Molecular biology of the cell* **20**: 3025.
- Patricia H, Terri M, Michael R, Roy P, Torben Heick J. 2001. Quality control of mRNA 3' -end processing is linked to the nuclear exosome. *Nature* **413**: 538.
- Pereira SS, Guimaraes FC, Carvalho JF, Stolf-Moreira R, Oliveira MC, Rolla AA, Farias JR, Neumaier N, Nepomuceno AL. 2011. Transcription factors expressed in soybean roots under drought stress. *Genet Mol Res* **10**: 3689-3701.
- Pilkington GR, Parker R. 2008. Pat1 contains distinct functional domains that promote P-body assembly and activation of decapping. *Molecular and Cellular Biology* **28**: 1298.
- Pisareva VP, Skabkin MA, Hellen CUT, Pestova TV, Pisarev AV. 2011. Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. *EMBO Journal* **30**: 1804-1817.
- Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH. 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**: 1851-1854.
- Price AH, Taylor A, Ripley SJ, Griffiths A, Trewavas AJ, Knight MR. 1994. Oxidative signals in tobacco increase cytosolic calcium. *The Plant Cell* **6**: 1301-1310.
- Proudfoot N. 2011. Ending the message: poly(A) signals then and now. In *Genes Dev*, Vol 25, pp. 1770-1782.
- Proudfoot NJ, Furger A, Dye MJ. 2002. Integrating mRNA Processing with Transcription. *Cell* **108**: 501-512.
- Pulak R, Anderson P. 1993. mRNA surveillance by the *Caenorhabditis elegans* smg genes. *Genes & development* **7**: 1885.
- Qin F, Sakuma Y, Tran L-S, Maruyama K, Kidokoro S, Fujita Y, Fujita M, Umezawa T, Sawano Y, Miyazono K-I et al. 2008. Arabidopsis DREB2A- Interacting Proteins Function as RING E3 Ligases and Negatively Regulate Plant Drought Stress- Responsive Gene Expression(W). *Plant Cell* **20**: 1693-1707.
- Quettier AL, Bertrand C, Habricot Y, Miginiac E, Agnes C, Jeannette E, Maldiney R. 2006. The phs1-3 mutation in a putative dual-specificity protein tyrosine phosphatase gene provokes hypersensitive responses to abscisic acid in *Arabidopsis thaliana*. *Plant Journal* **47**: 711-719.
- Quinlan AR. 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**: 11 12 11-34.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

- Rajyaguru P, She M, Parker R. 2012. Scd6 Targets eIF4G to Repress Translation: RGG Motif Proteins as a Class of eIF4G- Binding Proteins. *Molecular Cell* **45**: 244-254.
- Ramalho TO, Figueira AR, Sotero AJ, Wang R, Geraldino Duarte PS, Farman M, Goodin MM. 2014. Characterization of Coffee ringspot virus-Lavras: A model for an emerging threat to coffee production and quality. *Virology* **464-465**: 385-396.
- Ran E, Alejandro PU, Reuven A. 2013. Alternative cleavage and polyadenylation: extent, regulation and function. *Nature Reviews Genetics* **14**: 496.
- Rao S, Dinkins RD, Hunt AG. 2009. Distinctive interactions of the Arabidopsis homolog of the 30 kD subunit of the cleavage and polyadenylation specificity factor (AtCPSF30) with other polyadenylation factor subunits. *BMC Cell Biology* **10**: 51-51.
- Rebbapragada I, Lykke-Andersen J. 2009. Execution of nonsense-mediated mRNA decay: what defines a substrate? *Current Opinion in Cell Biology* **21**: 394-402.
- Reddy ASN. 2007. Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annual review of plant biology* **58**: 267.
- Reddy ASN, Marquez Y, Kalyna M, Barta A. 2013. Complexity of the alternative splicing landscape in plants. *The Plant Cell* **25**: 3657-3683.
- Richard P, Manley JL. 2009. Transcription termination by nuclear RNA polymerases. *Genes & development* **23**: 1247.
- Rojas C, Senthil-Kumar M, Wang K, Ryu C-M, Kaundal A, Mysore K. 2012. Glycolate Oxidase Modulates Reactive Oxygen Species-Mediated Signal Transduction during Nonhost Resistance in *Nicotiana benthamiana* and *Arabidopsis*(W). *Plant Cell* **24**: 336-352.
- Roman C, Cohn L, Calame K. 1991. A Dominant Negative Form of Transcription Activator mTFE3 Created by Differential Splicing. *Science* **254**: 94-97.
- Rosegrant MW, Cline SA. 2003. Global food security: challenges and policies. *Science* **302**: 1917-1919.
- Rosonina E, Manley JL. 2010. Alternative Polyadenylation Blooms. *Developmental Cell* **18**: 172-174.
- Sachs A, Wahle E. 1993. Poly(A) tail metabolism and function in eucaryotes. In *J Biol Chem*, Vol 268, pp. 22955-22958.
- Sachs AB, Davis RW. 1989. The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit- dependent translation initiation. *Cell* **58**: 857-867.
- Saini HK, Griffiths-Jones S, Enright AJ. 2007. Genomic analysis of human microRNA transcripts. In *Genomic analysis of human microRNA transcripts*, Vol 104, pp. 17719-17724. National Academy of Sciences.
- Saito S, Hosoda N, Hoshino S-I. 2013. The Hbs1-Dom34 protein complex functions in non-stop mRNA decay in mammalian cells. *The Journal of biological chemistry* **288**: 17832.
- Sandberg R, Sarma A, Sharp P, Burge C. 2008. Proliferating cells express mRNAs with shortened 3[variant prime] untranslated regions and fewer microRNA target sites. *Science* **320**: 1643-1647.

- Sarwat M, Ahmad P, Nabi G, Hu X. 2013. Ca(2+) signals: the versatile decoders of environmental cues. *Critical Reviews in Biotechnology*, 2013, Vol33(1), p97-109 **33**: 97-109.
- Sayani S, Janis M, Lee CY, Toesca I, Chanfreau GF. 2008. Widespread impact of nonsense-mediated mRNA decay on the yeast intronome. *Molecular Cell* **31**: 360-370.
- Schaeffer D, van Hoof A. 2011. Different nuclease requirements for exosome-mediated degradation of normal and nonstop mRNAs. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 2366.
- Schmid M, Jensen T. 2008. Quality control of mRNP in the nucleus. *Chromosoma* **117**: 419-429.
- Schoenberg DR, Maquat LE. 2009. Re-capping the message. *Trends in Biochemical Sciences* **34**: 435-442.
- Schönemann L, Kühn U, Martin G, Schäfer P, Gruber AR, Keller W, Zavolan M, Wahle E. 2014. Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes & development* **28**: 2381.
- Schwartz A, Komarova T, Skulachev M, Zvereva A, Dorokhov Y, Atabekov J. 2006. Stability of plant mRNAs depends on the length of the 3' -untranslated region. *Biochemistry (Moscow)* **71**: 1377-1384.
- Schweingruber C, Rufener SC, Zünd D, Yamashita A, Mühlemann O. 2013. Nonsense-mediated mRNA decay—Mechanisms of substrate mRNA recognition and degradation in mammalian cells. *BBA - Gene Regulatory Mechanisms* **1829**: 612-623.
- Seki M, Kamei A, Yamaguchi-Shinozaki K, Shinozaki K. 2003. Molecular responses to drought, salinity and frost: common and different paths for plant protection. *Current Opinion in Biotechnology* **14**: 194-199.
- Sheets MD, Ogg SC, Wickens MP. 1990. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Research* **18**: 5799-5805.
- Shen Y, Venu RC, Nobuta K, Wu X, Notibala V, Demirci C, Meyers BC, Wang G-L, Ji G, Li QQ. 2011. Transcriptome dynamics through alternative polyadenylation in developmental and environmental responses in plants revealed by deep sequencing. *Genome research* **21**: 1478.
- Shi Y. 2012. Alternative polyadenylation: new insights from global analyses. *RNA (New York, NY)* **18**: 2105.
- Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, Frank J, Manley JL. 2009. Molecular architecture of the human pre-mRNA 3' processing complex. *Molecular Cell* **33**: 365-376.
- Shinozaki K, Yamaguchi-Shinozaki K, Seki M. 2003. Regulatory network of gene expression in the drought and cold stress responses. *Current Opinion in Plant Biology* **6**: 410-417.
- Shoemaker C, Green R. 2011. Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *Proceedings of the National Academy of Sciences, USA* **108**: E1392-E1398.

- Simpson GG, Dijkwel PP, Quesada V, Henderson I, Dean C. 2003. FY is an RNA 3' end-processing factor that interacts with FCA to control the Arabidopsis floral transition. *Cell* **113**: 777.
- Singh G, Jakob S, Kleedehn MG, Lykke-Andersen J. 2007. Communication with the exon-junction complex and activation of nonsense-mediated decay by human Upf proteins occur in the cytoplasm. *Molecular Cell* **27**: 780-792.
- Singh G, Rebbapragada I, Lykke-Andersen J. 2008. A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biology* **6**: e111.
- Singh N, Wadhawan M, Tiwari S, Kumar R, Rathaur S. 2016. Inhibition of *Setaria cervi* protein tyrosine phosphatases by Phenylarsine oxide: A proteomic and biochemical study. *Acta Tropica* **159**: 20-28.
- Sinha R, Gupta A, Senthil-Kumar M. 2017. Concurrent Drought Stress and Vascular Pathogen Infection Induce Common and Distinct Transcriptomic Responses in Chickpea. *Frontiers in Plant Science* **8**.
- Smirnoff N. 1993. The role of active oxygen in the response of plants to water deficit and desiccation. *New Phytologist* **125**: 27-58.
- Sonmez C, Bäurle I, Magusin A, Dreos R, Laubinger S, Weigel D, Dean C. 2011. RNA 3' processing functions of Arabidopsis FCA and FPA limit intergenic transcription. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 8508.
- Sparks KA, Dieckmann CL. 1998. Regulation of poly(A) site choice of several yeast mRNAs. *Nucleic acids research* **26**: 4676-4687.
- Stephanie K, Allan J. 2012. NMD: a multifaceted response to premature translational termination. *Nature Reviews Molecular Cell Biology* **13**: 700.
- Sun Y, Fu Y, Li Y, Xu A. 2012. Genome-wide alternative polyadenylation in animals: insights from high-throughput technologies. *Journal of Molecular Cell Biology* **4**: 352-361.
- Suzanne K, Pamela AS. 2008. Coupling and coordination in gene expression processes: a systems biology view. *Nature Reviews Genetics* **9**: 38.
- Sweet T, Kovalak C, Collier J. 2012. The DEAD-box protein Dhh1 promotes decapping by slowing ribosome movement. *PLoS biology* **10**: e1001342.
- Swindell W. 2006. The association among gene expression responses to nine abiotic stress treatments in *Arabidopsis thaliana*. *Genetics* **174**: 1811-1824.
- Takahashi Y, Helmling S, Moore CL. 2003. Functional dissection of the zinc finger and flanking domains of the Yth1 cleavage/ polyadenylation factor. *Nucleic acids research* **31**: 1744.
- Takagaki Y, Ryner LC, Manley JL. 1988. Separation and characterization of a poly(A) polymerase and a cleavage/ specificity factor required for pre-mRNA polyadenylation. *Cell* **52**: 731-742.
- Takagaki Y, Seipelt RL, Peterson ML, Manley JL. 1996. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain Pre-mRNA during B cell differentiation. *Cell* **87**: 941-952.
- Tang G, Zhu X, Gakiere B, Levanony H, Galili G. 2002. The bifunctional LKR/ SDH locus of plants also encodes a highly active monofunctional lysine-

- ketoglutarate reductase using a polyadenylation signal located within an intron. *Plant Physiology* **130**: 147-154.
- Tang W, Gunn TM, McLaughlin DF, Barsh GS, Schlossman SF, Duke-Cohan JS. 2000. Secreted and membrane attractin result from alternative splicing of the human ATRN gene. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 6025-6030.
- Tantikanjana T, Nasrallah ME, Stein JC, Chen CH, Nasrallah JB. 1993. An alternative transcript of the S locus glycoprotein gene in a class II pollen-recessive self-incompatibility haplotype of *Brassica oleracea* encodes a membrane-anchored protein. *The Plant cell*: 657-666.
- Tharun S. 2009. Lsm1-7-Pat1 complex: a link between 3' and 5'-ends in mRNA decay? *RNA biology* **6**: 228.
- Thirunavukkarasu N, Sharma R, Singh N, Shiriga K, Mohan S, Mittal S, Mittal S, Mallikarjuna MG, Rao AR, Dash PK et al. 2017. Genomewide Expression and Functional Interactions of Genes under Drought Stress in Maize. *International Journal of Genomics* **2017**.
- Thomas PE, Wu X, Liu M, Gaffney B, Ji G, Li QQ, Hunt AG. 2012. Genome-wide control of polyadenylation site choice by CPSF30 in *Arabidopsis*. *Plant Cell* **24**: 4376-4388.
- Thomsen R, Libri D, Boulay J, Rosbash M, Jensen TH. 2003. Localization of nuclear retained mRNAs in *Saccharomyces cerevisiae*. *RNA (New York, NY)* **9**: 1049.
- Tian B, Hu J, Zhang H, Lutz CS. 2005. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic acids research* **33**: 201.
- Tian B, Pan Z, Lee JY. 2007. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome research* **17**: 156.
- Timmusk T, Palm K, Metsis M, Reintam T, Paalme V, Saarma M, Persson H. 1993. Multiple promoters direct tissue-specific expression of the rat BDNF gene. *Neuron* **10**: 475-489.
- Tom M, Robin R. 2002. An extensive network of coupling among gene expression machines. *Nature* **416**: 499.
- Torres-Torronteras J, Rodriguez - Palmero A, Pinós T, Accarino A, Andreu AL, Pintos - Morell G, Martí R. 2011. A novel nonstop mutation in TYMP does not induce nonstop mRNA decay in a MNGIE patient with severe neuropathy. *Human Mutation* **32**: E2061-E2068.
- Tsuboi T, Kuroha K, Kudo K, Makino S, Inoue E, Kashima I, Inada T. 2012. Dom34:hbs1 plays a general role in quality-control systems by dissociation of a stalled ribosome at the 3' end of aberrant mRNA. *Molecular Cell* **46**: 518-529.
- Twu KY, Noah DL, Rao P, Kuo R-L, Krug RM. 2006. The CPSF30 Binding Site on the NS1A Protein of Influenza A Virus Is a Potential Antiviral Target. *The Journal of Virology* **80**: 3957.
- van Hoof A, Frischmeyer PA, Dietz HC, Parker R. 2002. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science (New York, NY)* **295**: 2262.

- van Hoof A, Staples RR, Baker RE, Parker R. 2000. Function of the Ski4p (Csl4p) and Ski7p Proteins in 3'-to-5' Degradation of mRNA. *Molecular and Cellular Biology* **20**: 8230.
- van Hoof A, Wagner EJ. 2011. A brief survey of mRNA surveillance. *Trends in Biochemical Sciences* **36**: 585-592.
- Vasudevan S, Peltz SW, Wilusz CJ. 2002. Non-stop decay--a new mRNA surveillance pathway. *Bioessays* **24**: 785-788.
- Verma R, Oania RS, Kolawa NJ, Deshaies RJ, Dikic I. 2013. Cdc48/ p97 promotes degradation of aberrant nascent polypeptides bound to the ribosome. *eLife* **2**.
- Verslues PE, Zhu J-K. 2007. New developments in abscisic acid perception and metabolism. *Current Opinion in Plant Biology* **10**: 447-452.
- Vranova E, Villarreal R. 2002. Comprehensive analysis of gene expression in *Nicotiana tabacum* leaves acclimated to oxidative stress. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 10870-10875.
- Wang BB, Brendel V. 2006. Genomewide comparative analysis of alternative splicing in plants. In *Genomewide comparative analysis of alternative splicing in plants*, Vol 103, pp. 7175-7180.
- Wang F, Perry SE. 2013. Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development. *Plant physiology* **161**: 1251.
- Wang X, Cai X, Xu C, Wang Q, Dai S. 2016. Drought-Responsive Mechanisms in Plant Leaves Revealed by Proteomics. Vol 17, p. 1706. MDPI AG, Basel.
- West S, Gromak N, Norbury CJ, Proudfoot NJ. 2006. Adenylation and exosome-mediated degradation of cotranscriptionally cleaved pre-messenger RNA in human cells. *Molecular cell* **21**: 437.
- Wiestner A, Tehrani M, Chiorazzi M, Wright G, Gibellini F, Nakayama K, Liu H, Rosenwald A, Muller-Hermelink HK, Ott G et al. 2007. Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival. *Blood* **109**: 4599.
- Wu J, Kang JH, Hettenhausen C, Baldwin IT. 2007. Nonsense - mediated mRNA decay (NMD) silences the accumulation of aberrant trypsin proteinase inhibitor mRNA in *Nicotiana attenuata*. *Plant Journal* **51**: 693-706.
- Wu X, Brewer G. 2012. The regulation of mRNA stability in mammalian cells: 2.0. *Gene* **500**: 10-21.
- Wu X, Gaffney B, Hunt AG, Li QQ. 2014. Genome-wide determination of poly(A) sites in *Medicago truncatula*: evolutionary conservation of alternative poly(A) site choice. *BMC genomics* **15**: 615.
- Wu X, Liu M, Downie B, Liang C, Ji G, Li QQ, Hunt AG. 2011. Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 12533.
- Xing A, Moon BP, Mills KM, Falco SC, Li Z. 2010. Revealing frequent alternative polyadenylation and widespread low-level transcription read-through of

- novel plant transcription terminators. *Plant Biotechnology Journal* **8**: 772-782.
- Xing D, Li Q. 2011a. Alternative polyadenylation and gene expression regulation in plants. In *Wiley Interdiscip Rev-RNA*, Vol 2, pp. 445-458.
- Xing D, Li QQ. 2011b. Alternative polyadenylation and gene expression regulation in plants. *Wiley interdisciplinary reviews RNA* **2**: 445.
- Xu R, Zhao H, Dinkins R, Cheng X, Carberry G, Li Q. 2006. The 73 kD Subunit of the cleavage and polyadenylation specificity factor (CPSF) complex affects reproductive development in Arabidopsis. *Plant Mol Biol* **61**: 799-815.
- Yamaguchi-Shinozaki K, Shinozaki K. 2006. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annual review of plant biology* **57**: 781-803.
- Yan J, Marr TG. 2005. Computational analysis of 3'-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat. *Genome research* **15**: 369.
- Yanagisawa S, Izui K, Yamaguchi Y, Shigesada K, Katsuki H. 1988. Further analysis of cDNA clones for maize phosphoenolpyruvate carboxylase involved in C4 photosynthesis Nucleotide sequence of entire open reading frame and evidence for polyadenylation of mRNA at multiple sites in vivo. *FEBS Letters* **229**: 107-110.
- Yang T, Poovaiah BW. 2002. Hydrogen peroxide homeostasis: activation of plant catalase by calcium/ calmodulin. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 4097-4102.
- Yoon OK, Brem RB. 2010. Noncanonical transcript forms in yeast and their regulation during environmental stress. *RNA (New York, NY)* **16**: 1256.
- Yoshikawa M, Peragine A, Park MY, Poethig RS. 2005. A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes & development* **19**: 2164.
- Yoshimura K, Yabuta Y, Ishikawa T, Shigeoka S. 2002. Identification of a cis element for tissue-specific alternative splicing of chloroplast ascorbate peroxidase pre-mRNA in higher plants. *The Journal of biological chemistry* **277**: 40623.
- Zarudnaya MI, Kolomiets IM, Hovorun DM. 2002. What nuclease cleaves pre - mRNA in the process of polyadenylation? *IUBMB Life* **54**: 27-31.
- Zhang H. 2003. *Improving water productivity through deficit irrigation: examples from Syria, the North China Plain and Oregon, USA*. CABI Publishing: Wallingford, UK, UK.
- Zhang J, Addepalli B, Yun K-Y, Hunt AG, Xu R, Rao S, Li QQ, Falcone DL. 2008. A Polyadenylation Factor Subunit Implicated in Regulating Oxidative Signaling in *Arabidopsis thaliana* (Signaling in RNA Processing). *PLoS ONE* **3**: e2410.
- Zhao J, Hyman L, Moore C. 1999a. Formation of mRNA 3'-ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiology and Molecular Biology Reviews* **63**: 405.
- Zhao J, Hyman L, Moore C. 1999b. Formation of mRNA 3' Ends in Eukaryotes: Mechanism, Regulation, and Interrelationships with Other Steps in mRNA Synthesis. *Microbiology and Molecular Biology Reviews* **63**: 405.

- Zhao P, Liu P, Yuan G, Jia J, Li X, Qi D, Chen S, Ma T, Liu G, Cheng L. 2016. New Insights on Drought Stress Response by Global Investigation of Gene Expression Changes in Sheepgrass (*Leymus chinensis*). *Frontiers in plant science* **7**: 954.
- Zhong T, Arndt KT. 1993. The yeast SIS1 protein, a DnaJ homolog, is required for the initiation of translation. *Cell* **73**: 1175-1186.
- Zhu J, Dong C-H, Zhu J-K. 2007. Interplay between cold-responsive gene regulation, metabolism and RNA processing during plant cold acclimation. *Current Opinion in Plant Biology* **10**: 290-295.
- Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. 2001. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *BioTechniques* **30**: 892-897.

VITA

Name: Guijie Hao

Education:

2009-2012: Shenyang Agriculture University

Master Degree in Plant Science

2007-2009: Liaocheng University

Bachelor Degree in Bioscience

2004-2007: Binzhou University

Associate Degree in Education of Biology

Research paper:

Guijie Hao, Laura De Lorenzo, Manohar Chakrabarti, Arthur G. Hunt. Alternative Polyadenylation as a Potential Mechanism to Response Drought Stress in *Arabidopsis thaliana*. (Manuscript in preparation)

Guijie Hao, Laura De Lorenzo, Manohar Chakrabarti, Arthur G. Hunt. AtCPSF30-mediated Alternative Polyadenylation Controls mRNA Stability *Arabidopsis thaliana*. (Manuscript in preparation)

Guijie Hao, Xueyi Sui, Allan Bruce Downie, Arthur G. Hunt. A screen for proteins that interact with AtCPSF30. (Manuscript in preparation)