



Dehydration stress extends mRNA 3' untranslated regions with noncoding RNA functions in *Arabidopsis*

Hai-Xi Sun, Yan Li, Qi-Wen Niu, et al.

Genome Res. published online May 18, 2017

Access the most recent version at doi:[10.1101/gr.218669.116](https://doi.org/10.1101/gr.218669.116)

P<P Published online May 18, 2017 in advance of the print journal.

Accepted Manuscript Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Comprehensive immune receptor profiling.
Discover the **DriverMap™ AIR Assay** difference.

LEARN
MORE



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Published by Cold Spring Harbor Laboratory Press

Dehydration stress extends mRNA 3' untranslated regions with noncoding RNA functions in *Arabidopsis*

Hai-Xi Sun,^{1,3} Yan Li,^{1,3} Qi-Wen Niu,^{1,2} and Nam-Hai Chua^{1,4}

¹ Laboratory of Plant Molecular Biology, Rockefeller University, 1230 York Avenue, New York, New York 10065, USA

² Present address: Wilmar (Shanghai) Biotechnology Research & Development Co., Ltd, No. 118 Gaodong Road, Pudong New District, Shanghai 200137, China

³ These authors contributed equally to this work

⁴ Corresponding author: Nam-Hai Chua

Andrew W. Mellon Professor

Head, Lab of Plant Molecular Biology

Rockefeller University

1230 York Avenue

New York, NY 10065

chua@mail.rockefeller.edu

Tel: 212-327-8126

Running title:

Dehydration stress extends 3' UTRs in *Arabidopsis*

Keywords:

3' UTR extensions, dehydration stress, alternative cleavage and polyadenylation, noncoding RNA, *Arabidopsis*

Abstract:

The 3' untranslated regions (3' UTRs) of mRNAs play important roles in the regulation of mRNA localization, translation and stability. Alternative cleavage and polyadenylation (APA) generates mRNAs with different 3' UTRs, but the involvement of this process in stress response has not yet been clarified. Here, we report that a subset of stress-related genes exhibits 3' UTR extensions of their mRNAs during dehydration stress. These extended 3' UTRs have characteristics of long noncoding RNAs and likely do not interact with miRNAs. Functional studies using T-DNA insertion mutants reveal that they can act as antisense transcripts to repress expression levels of sense genes from the opposite strand, or can activate the transcription or lead to read-through transcription of their downstream genes. Further analysis suggests that transcripts with 3' UTR extensions have weaker poly(A) signals than those without 3' UTR extensions. Finally, we show that their biogenesis is partially dependent on a *trans*-acting factor FPA. Taken together, we report that dehydration stress could induce transcript 3' UTR extensions and elucidate a novel function for these stress-induced 3' UTR extensions as long noncoding RNAs in the regulation of their neighboring genes.

INTRODUCTION

The 3' untranslated regions (3' UTRs) of mRNAs play important roles in regulating mRNA localization, translation and stability (Elkon et al. 2013; Tian and Manley 2013). Regulatory elements within 3' UTRs are essential for their interactions with *trans*-acting factors, such as microRNAs (miRNAs) and RNA-binding proteins (RBPs) (Kim et al. 2007; Fabian et al. 2010). Alternative cleavage and polyadenylation (APA) generates mRNAs with different 3' UTR sequences and it is widespread in plants and animals. About 70% of mRNAs contain more than one poly(A) sites in human (Derti et al. 2012), mouse (Hoque et al. 2013) and *Arabidopsis* (Wu et al. 2011). The 3' UTR length is dynamically regulated under different biological conditions, e.g., mRNAs tend to have shorter 3' UTRs in proliferating T cells (Sandberg et al. 2008; Gruber et al. 2014), induced pluripotent stem cells (Ji and Tian 2009) and cancer cells (Mayr and Bartel 2009). By contrast, in nervous system (Hilgers et al. 2011; Hilgers et al. 2012) and during embryonic development (Ji et al. 2009), mRNAs tend to have longer 3' UTRs.

APA is regulated by *cis*-acting elements within mRNAs, which determine the position and cleavage efficiency of the poly(A) sites (Di Giammartino et al. 2011; Elkon et al. 2013). The major *cis*-acting element is the poly(A) signal located at 15-30 nt upstream of the cleavage sites. The canonical form is AAUAAA. APA is also regulated by several *trans*-acting factors that are shown to regulate poly(A) site selection (Licatalosi et al. 2008; Hilgers et al. 2012; Jenal et al. 2012; Thomas et al. 2012; Bava et al. 2013; Duc et al. 2013). For example, the 3' UTR extensions in *Drosophila* nervous system are mediated by ELAV (Hilgers et al. 2012). In *elav* mutants the 3' UTR extensions are diminished whereas ectopic expression of ELAV in non-neural embryonic tissues induces 3' UTR extensions. Further study reveals that ELAV directly binds to the proximal poly(A) signals and inhibits their usage.

Previous studies have identified some *Arabidopsis* *trans*-acting factors involved in transcript maturation (Horniyk et al. 2010b; Thomas et al. 2012; Zhang et al. 2015). Among them, FPA is one of the best characterized polyadenylation regulatory factors which has been studied since more than 10 years ago. FPA was first found as a positive regulator of flowering and the protein carries an RNA recognition motif (RRM) (Schomburg et al. 2001; Simpson et al. 2004; Quesada et al. 2005). Further studies revealed that FPA could function as a *trans*-acting factor to promote the proximal poly(A) site selection of *FLC* antisense transcript. In the absence of FPA, *FLC* antisense transcript with distal poly(A) site is induced, which activates *FLC* expression to repress flowering (Horniyk et al. 2010a; Horniyk et al. 2010b; Liu et al. 2010). In addition to *FLC* antisense transcript, FPA also regulates poly(A) site choice of many other transcripts (Sonmez et al. 2011; Sonmez and Dean 2012; Duc et al. 2013; Lyons et al. 2013). Furthermore, FPA is also involved in RNA-mediated chromatin silencing of several endogenous transposons and retroelements by interacting with chromatin regulators (Baurle et al. 2007; Baurle and Dean 2008). And recently, the crystal structure of FPA SPOC (Spen paralogue and orthologue C-terminal) domain has been reported (Zhang et al. 2016).

Although APA induced alteration of 3' UTR sequences plays important roles in the regulation of mRNA functions, the involvement of this process in stress response has not yet been clarified. To study the effects of stress on APA, here we performed strand-specific RNA-seq (ssRNA-seq) on *Arabidopsis* seedlings grown in normal, ABA-treated and dehydration-stressed conditions. We developed a computational pipeline to identify and compare 3' ends in stressed conditions with those in normal condition in order to identify stress-induced APA events.

99

100 **RESULTS**

101

102 **Dehydration stress induces 3' UTR extensions**

103 We collected samples at two time points (1 h and 3 h) with three biological replicates for
 104 each time point (Supplemental Fig S1A; Supplemental Table S1). Global correlation
 105 analysis revealed that these replicates were highly reproducible and replicates of each
 106 condition showed unique expression profiles (Supplemental Fig S1B). Generally, about
 107 71% of TAIR10 annotated genes (23,706 of 33,602) were detected, including 22,526
 108 protein coding genes, 61 microRNA (miRNA) genes and 1,119 other noncoding genes. In
 109 addition, we also detected 1,081 long intergenic noncoding RNA (lincRNA) genes
 110 (Supplemental Fig S1C; Supplemental Table S2). Among these genes we identified 468
 111 protein coding genes that were up-regulated in all stressed samples (fold change ≥ 2 , P
 112 value < 0.05 ; Supplemental Fig S2A). Gene Ontology (GO) enrichment analysis revealed
 113 that they were significantly enriched in ABA response and dehydration stress associated
 114 biological processes (Supplemental Fig S2B) demonstrating that our transcriptome data
 115 sets exhibited unique features of plants in response to ABA and dehydration stress.

116

117 In the IGV genome browser we observed some mRNAs had 3' UTR extensions in
 118 dehydration stress (Fig. 1A; Supplemental Fig S3A). To determine whether this
 119 phenomenon was widespread, we developed a computational pipeline to identify
 120 stress-induced 5'/3' UTR extensions followed by a series of quality controls
 121 (Supplemental Fig S3B). Consistent with our observations, a large number of mRNAs
 122 exhibited 3' UTR extensions specifically in dehydration stress (Fig. 1B). The number of 3'
 123 UTR extensions was 5 times greater than that of 5' UTR extensions (1,552 vs. 342, P
 124 value = 3.22×10^{-191} , Fisher's exact test). When the ambiguous extensions were
 125 excluded (Supplemental Fig S3C), it was 8 times greater (1,324 vs. 165, P value = $1.34 \times$

10⁻²³¹, Fisher's exact test; Supplemental Fig S3D). Moreover, the number of 3' UTR extensions induced by dehydration stress was significantly greater than that induced by ABA treatment (1,530 vs. 96, *P* value → 0, Fisher's exact test).

The 3' UTR extensions were further validated by RT-PCR (Fig. 1C). While transcripts with canonical 3' UTRs (P1 + P2) were detected in all conditions the extended 3' UTRs (P3 + P4) were only detected in dehydration stress. Moreover, using primer sets flanking the proximal cleavage sites (P5 + P6) as well as northern blot analysis we also confirmed that the extended 3' UTRs were only detected in dehydration stress and more important, they were indeed 3' UTR extensions rather than independent transcripts (Fig. 1D,E; Supplemental Fig S3E).

Using the same criteria we identified only 55 and 77 of 5' and 3' UTR shortenings (Supplemental Fig S3F) indicating that transcripts tend to have longer 3' UTRs under dehydration stress. This is in agreement with previous studies in animals demonstrating that arrested cells tend to express mRNAs with longer 3' UTRs (Sandberg et al. 2008; Ji et al. 2009; Ji and Tian 2009; Elkon et al. 2012), because dehydration stress inhibits plant growth (Claeys and Inze 2013) including cell proliferation. Taken together, these results demonstrated that dehydration stress could induce 3' UTR extensions.

Expression levels of extended 3' UTRs are time-dependent

A total of 631 3' UTR extensions were induced in both 1 h and 3 h (Fig. 2A) and 80% of them had higher expression levels at 3 h (Fig. 2B). In addition, there were also 778 3' UTR extensions that were only induced at 3 h. As the majority of them had increased expression levels over time, we focused on 3' UTR extensions induced in 3 h (*n* = 1,409; Supplemental Table S3) in further analysis. We found the majority (75%) of the extended 3' UTRs accounted for less than 10% of the total transcript abundance indicating that in

dehydration stress the transcripts with canonical 3' UTRs were still the dominant isoform (Supplemental Fig S4). GO enrichment analysis of these 1,409 genes revealed that their functions were significantly enriched in biological processes associated with stress responses, including response to water stress (Fig. 2C).

Extended 3' UTRs have characteristics of long noncoding RNAs

Recently, a large group of long noncoding RNAs, including lincRNAs and long noncoding natural antisense transcripts (lncNATs), have been characterized in *Arabidopsis* (Liu et al. 2012; Zhu et al. 2013; Di et al. 2014; Wang et al. 2014). We found that the extended 3' UTRs also had characteristics of long noncoding RNAs. Length distribution revealed that the majority of 3' UTR extensions were 200 to 800 nt in length, which were similar to lincRNAs and shorter than protein coding mRNAs and primary miRNAs (pri-miRNAs) (Fig. 3A). In addition, coding potential assessment revealed that they were distinct from protein coding mRNAs but similar to lincRNAs and pri-miRNAs (Fig. 3B,C). Moreover, their expression levels were 10 fold lower than protein coding mRNAs and a little higher than lincRNAs and pri-miRNAs (Fig. 3D), which was also a feature of noncoding RNAs (Cabili et al. 2011; Liu et al. 2012; Pauli et al. 2012; Sun et al. 2013; Di et al. 2014). Taken together, these results demonstrated that the extended 3' UTRs had characteristics of long noncoding RNAs, raising the possibility that they might function as noncoding RNAs.

Extended 3' UTRs likely do not interact with miRNAs

Previous studies in animals found that proliferating cells tend to express mRNAs with shortened 3' UTRs and fewer miRNA target sites leading to increased protein production by these mRNAs (Sandberg et al. 2008). Therefore, we asked whether the extended 3' UTRs have more miRNA target sites and accordingly we predicted miRNA target sites (Wu et al. 2012) in sequences of 1,409 extended 3' UTRs. However, we found that only 4% of them (63 of 1,409) had gained miRNA target sites (Supplemental Table S3;

Supplemental Data S1). In addition, only 6% of them (4 of 63) were downregulated in dehydration stress (Supplemental Fig S5A) suggesting that they were not targeted by miRNAs.

Target mimicry refers to a negative regulatory mechanism of miRNA functions mediated by noncoding RNAs (Franco-Zorrilla et al. 2007). Recently a genome-wide study identified 36 putative endogenous miRNA target mimics (eTMs) in *Arabidopsis* (Wu et al. 2013). As the identified 3' UTR extensions had features of noncoding RNAs, we asked whether they had gained target mimic sites. However, none of these eTMs were found in extended 3' UTR sequences suggesting that they did not function as miRNA target mimics.

Repression of expression levels of sense genes by extended 3' UTRs

Previous study using human and mouse cells uncovered an anti-correlation between 3' UTR lengths and gene expression levels (Ji et al. 2011). However, in our data only a small group (3%) of transcripts with 3' UTR extensions were downregulated in dehydration stress (Supplemental Fig S5B). By contrast, the majority (74%) of them were not differentially expressed suggesting that 3' UTR extensions had little effect on their own total expression levels (short isoforms + long isoforms). In addition, as the identified 3' UTR extensions were extended from the 3' end of canonical 3' UTRs they should not change the coding sequences (CDS) of their own transcripts. These observations suggested that the identified 3' UTR extensions were more likely used to regulate other genes.

Previous studies in *Arabidopsis* have demonstrated that the expression of natural antisense transcripts (NATs) could result in the repression of sense gene expression (Borsani et al. 2005; Katiyar-Agarwal et al. 2006; Katiyar-Agarwal et al. 2007; Faghihi and

Wahlestedt 2009; Ron et al. 2010). Such overlapping pairs were also observed when transcripts (referred as “sense transcripts” hereafter) overlapped by 3' UTR extensions transcribed from the opposite strand (referred as “antisense transcripts” hereafter). We found 43% of 3' UTR extensions (605 of 1,409) overlapped in sequences with sense transcripts (Supplemental Table S3), resulting in the formation or elongation of overlapping regions (Fig. 4A,B). About 47% of the canonical 3' UTRs did not overlap with sense transcripts, and for the overlapped pairs the majority of them overlapped at 3' UTR of sense transcripts. By contrast, for the extended 3' UTRs, the number of pairs with overlapping sequences at CDS (44%) and promoter (41%) was elevated, leading to the increased coverage (~10 fold in average coverage, P value = 9.93×10^{-167} , t test) of sense transcripts.

To investigate whether the expression levels of sense transcripts were repressed by 3' UTR extensions, we compared their expression levels in dehydration stress to those under ABA treatment. Among 41 ABA treatment-induced sense transcripts (ABA_3h/NC $\geq$ 2), 46% (19 of 41) were repressed by dehydration stress (Dehy_3h/ABA_3h $\leq$ 0.5; Supplemental Fig S6A,B; Supplemental Table S4). Meanwhile, among 41 dehydration stress-repressed sense transcripts (Dehy_3h/NC $\leq$ 0.5), 59% (24 of 41) were not repressed by ABA treatment (ABA_3h/NC $>$ 0.5; Supplemental Fig S6A,C; Supplemental Table S4). Moreover, as the 3' UTR extensions increased in their expression levels over time (Fig. 2), the sense transcripts were predicted to decrease in their expression levels (relative to ABA treatment) over time. As expected, we found that 74% of them (32 of 43) had increased expression levels of 3' UTR extensions and decreased expression levels of sense transcripts over time (Fig. 4C, left; Supplemental Table S4), suggesting that the repression of sense transcripts by 3' UTR extensions was dose-dependent. However, using convergently transcribed transcripts one of which (at least) was up-regulated by dehydration stress and neither of which were subject to 3' UTR extensions, we did not

find such correlation (Fig. 4C, right): only 33% of them had increased expression levels of antisense genes and decreased expression levels of sense genes over time, which was significantly different from that of 3' UTR extension-sense gene pairs (33% vs. 74%, P value = 4.07×10^{-6} , Fisher's exact test).

One pair of such regulation was further functionally validated using T-DNA insertion mutants. *SP1-LIKE 2 (SPL2)* mRNA displayed 3' UTR extensions under dehydration stress and the extended 3' UTR elongated the overlapping regions with the sense *NFYA5* mRNA (*Nuclear Factor Y, Subunit A5*) transcribed from the opposite strand (Fig. 4D). *NYFA5* was reported as a transcription factor to promote drought resistance in *Arabidopsis* (Li et al. 2008). The overlapping region elongated about 9 fold in length (from 154 nt to 1,423 nt), which covered the entire 3' UTR and CDS of the *NFYA5* mRNA under dehydration stress. A *SPL2* T-DNA insertion knockout line (WiscDsLoxHs077_05E) was employed to investigate whether the extended *SPL2* 3' UTR could repress *NFYA5* transcription (Fig. 4D). As expected, in *spl2* mutant, the induction of extended *SPL2* 3' UTR was eliminated in dehydration stress and, as a result, the *NFYA5* gene was derepressed (Fig. 4E). Taken together, these results showed that 3' UTR extensions could function as long antisense noncoding RNAs repressing the expression levels of sense genes in dehydration stress.

Activation or read-through transcription of downstream genes by 3' UTR extensions

In addition to 605 3' UTR extensions that overlapped with sense mRNAs, we also identified 429 extensions that extended into the promoter regions (defined as sequences 1 kb upstream of transcriptional start site) of their downstream genes from the same strand, or lead to read-through transcription (Fig. 5A; Supplemental Table S3). Compared with canonical 3' UTRs, the length of intergenic sequences was significantly shortened by

the extended 3' UTRs (~3 fold in average length, P value = 1.56×10^{-125} , t test).

We found for the downstream genes that were lowly expressed in untreated control plants (FPKM < 2; $n = 119$), about 75% of them (89 of 119) were up-regulated in dehydration stress (Dehy_3h/NC ≥ 2). Interestingly, the majority (91%) of them (81 of 89) were also up-regulated compared to ABA treatment (Dehy_3h/ABA_3h ≥ 2 ; Supplemental Fig S7A; Supplemental Table S5) suggesting that they would have higher expression levels if their promoter sequences were overlapped by upstream 3' UTR extensions. Similarly, as 3' UTR extensions had increased expression levels over time (Fig. 2), we found that 86% of the downstream genes (70 of 81) also showed elevated expression levels in dehydration stress for 3 h compared to 1 h (Fig. 5B; Supplemental Table S5), suggesting that the activation of downstream genes by upstream 3' UTR extensions was dose-dependent.

Such proposed regulation was functionally validated using T-DNA insertion mutants. *NITRATE EXCRETION TRANSPORTER1* (*NAXT1*) mRNA displayed read-through transcription with the downstream gene *AT3G45660* under dehydration stress (Fig. 5C; Supplemental Fig S7B). Real-time PCR analysis also validated that the extended *NAXT1* 3' UTR was induced by dehydration stress and *AT3G45660* was up-regulated (Fig. 5D). A *NAXT1* T-DNA insertion knockout line (SALK_091226) was used to see whether the up-regulation of *AT3G45660* was due to read-through transcription (Fig. 5C). Consistent with our expectations, when the read-through transcription was depleted in *naxt1-1* mutant, the downstream *AT3G45660* gene was deactivated under dehydration stress (Fig. 5D).

Another example is *FARNESOIC ACID CARBOXYL-O-METHYLTRANSFERASE* (*FAMT*) whose 3' UTR extended into the promoter region of the downstream gene *AT3G44870*

(Supplemental Fig S7C). Real-time PCR assay also verified the induction of *FAMT* extended 3' UTR and *AT3G44870* transcription in dehydration stress (Supplemental Fig S7D). We identified a *FAMT* T-DNA insertion line (SALK_119380) to assess whether the activation of *AT3G44870* expression is dependent on *FAMT* extended 3' UTR (Supplemental Fig S7C). We found that this T-DNA insertion led to activation of *FAMT* gene, yielding even higher expression level of the extended 3' UTR in dehydration stress (Supplemental Fig S7D). Consistent with this, increased expression level of the downstream *AT3G44870* gene was observed under dehydration stress in the T-DNA insertion line compared with wild type (WT) (Supplemental Fig S7D). Taken together, these results provide evidence that the 3' UTR extensions could lead to read-through transcription or activation of downstream genes in dehydration stress.

mRNAs with 3' UTR extensions tend to have weaker poly(A) signals

The overall expression levels of transcripts with 3' UTR extensions were higher than those without 3' UTR extensions (~ 3.2 fold in median FPKM, P value = 1.85×10^{-123} , t test; Supplemental Fig S8A), raising the possibility that the biogenesis of 3' UTR extensions were due to their high expression levels. However, when we analyzed their expression levels in ABA treatment, we found that 83% of them (1,178 of 1,409) had the same or higher expression levels in ABA treatment, but only 3% of them (40 of 1,178) extended their 3' UTRs (Supplemental Fig S8B). Similarly, for the 778 transcripts that only had 3' UTR extensions in dehydration stress for 3 h (Fig. 2A), 81% of them (632 of 778) had the same or higher expression levels in dehydration stress for 1 h (Supplemental Fig S8C). Taken together, these results showed that 3' UTR extensions were not induced by high expression levels.

Previous study in *Arabidopsis* using direct RNA sequencing (DRS) revealed that the formation of mRNA 3' ends was heterogeneous and the sequences in the vicinity of the

most preferred cleavage sites were characterized by an AAUAAA-like motif and an A-rich peak at the -20 locus, as well as an U-rich peak at the -7 locus (Sherstnev et al. 2012). An absence of these sequence signatures was observed in the non-preferred cleavage sites. One explanation for the biogenesis of 3' UTR extensions is that the poly(A) signals of transcripts with 3' UTR extensions are weaker than those of transcripts without 3' UTR extensions. Our analysis showed that the proportion of upstream sequences with AAUAAA-like motifs was significantly lower in mRNAs with 3' UTR extensions (Fig. 6A; Supplemental Fig S9). In addition, the AAUAAA-like motifs were not enriched at the -20 locus of transcripts with 3' UTR extensions (Fig. 6B), and consistent with this the absence of A-rich peak at the -20 locus and of U-rich peak at the -7 locus was observed (Fig. 6C-E). However, we did not find any obvious difference between the proximal and the distal cleavage site of transcripts with 3' UTR extensions (Fig. 6A-D; Supplemental Fig S9). Taken together, these computational analyses suggest that transcripts with 3' UTR extensions have weaker poly(A) signals than those without 3' UTR extensions.

FPA partially regulates biogenesis of dehydration stress-induced 3' UTR extensions

In addition to *cis*-acting poly(A) signals on transcripts, previous studies also identified several *trans*-acting factors that are involved in the control of poly(A) site choice (Licatalosi et al. 2008; Hilgers et al. 2012; Jenal et al. 2012; Thomas et al. 2012; Bava et al. 2013; Duc et al. 2013). One of them is FPA, which was first identified as a positive regulator of flowering in *Arabidopsis* (Schomburg et al. 2001; Hornyik et al. 2010b). The FPA protein has three RRM domains at the N terminus, and one SPOC domain (protein-protein interaction domain) at the C terminus (Fig. 7A). Recently, a genome-wide study using DRS revealed that FPA could regulate poly(A) site choice of many transcripts in *Arabidopsis*. In *fpa* knockout mutants (*fpa-7*), hundreds of intergenic read-through transcripts were induced, resulting from defective 3' end formation of their upstream

transcripts (Duc et al. 2013). By reanalysis of the DRS data sets we identified transcripts with enhanced distal polyadenylation in *fpa-7* mutant compared to WT (Supplemental Fig S10A; Supplemental Table S6). We found that 76% of them (320 of 420) were also induced in dehydration stress for 3 h whereas only 8% of them (32 of 420) were induced in ABA treatment for 3 h (Fig. 7B). These results raised the possibility that the dehydration stress-induced 3' UTR extensions were dependent on FPA.

To test this hypothesis, we determined the induction of 3' UTR extensions by dehydration stress in WT plants and *fpa-7* mutants. If their biogenesis were completely dependent on FPA, the induction of 3' UTR extensions by dehydration stress should be abolished in *fpa-7* mutants; by contrast, if their biogenesis were through FPA-independent pathway, this induction should be the same for WT plants and *fpa-7* mutants. To examine these possibilities, we selected five genes including *PHYTOCHROME-INTERACTING FACTOR 5 (PIF5)* and *ETHYLENE RESPONSIVE ELEMENT BINDING FACTOR 4 (ERF4)*, whose transcript 3' UTR extensions were previously shown to be regulated by FPA (Duc et al. 2013; Lyons et al. 2013). Real-time PCR analysis confirmed that all of these genes exhibited 3' UTR extensions of their transcripts in WT plants under dehydration stress and in *fpa-7* mutants grown in normal condition (Supplemental Fig S10B,C). There was no obvious difference in total transcript abundance of these genes between WT plants and *fpa-7* mutants (Fig. 7C). However, the induction of dehydration stress-induced 3' UTR extensions was significantly impaired in *fpa-7* mutants (Fig. 7D), showing that FPA regulates the biogenesis of dehydration stress-induced 3' UTR extensions. In addition, as this induction was not completely eliminated in *fpa-7* mutants, we conclude that there should be also a FPA-independent pathway that participates in the regulation of their biogenesis.

Next we asked whether *FPA* expression was decreased in dehydration stress. We found its total mRNA abundance was not decreased (Supplemental Fig S10D) raising the

possibility that its function was impaired at protein level. Previous study revealed that alternative polyadenylation of *FPA* mRNA was regulated by the FPA protein itself: *FPA* mRNA contains a proximal poly(A) site in its first intron, and the use of this poly(A) site resulted in a shorten transcript encoding a truncated FPA protein with only one RRM domain (Fig. 7A); the full-length FPA protein promotes the usage of proximal poly(A) site forming a negative autoregulation loop (Hornyik et al. 2010b). A similar phenomenon might be happening here: we found the *FPA* first intron was retained specifically in dehydration stress (Fig. 7E,F), which introduced a premature termination codon in its CDS leading to the translation of a similar truncated FPA protein containing only one of the four domains (Fig. 7A). Therefore, it is possible that the truncated FPA proteins competed with full-length FPA proteins to interfere with FPA functions in dehydration stress thus giving rise to 3' UTR extensions. Although due to technical issues such as the lack of FPA antibody we were not able to fully validate this competition, our results still provide preliminary evidence that the biogenesis of dehydration stress-induced 3' UTR extensions is partially dependent on the *trans*-acting factor, FPA.

DISCUSSION

APA generates transcripts with different 3' UTR sequences; however, the role of this process in stress response has not been investigated. Here, we report an unanticipated global induction of 3' UTR extensions in *Arabidopsis* under dehydration stress, linking 3' UTR extensions to plant stress response (Fig. 1). Their expression levels are time-dependent (Fig. 2) and they have characteristics of long noncoding RNAs (Fig. 3). Unlike transcripts 3' UTR extensions in animals (Sandberg et al. 2008; Ji and Tian 2009; Jenal et al. 2012), the dehydration stress-induced transcript 3' UTR extensions appear not to possess miRNA binding sites (Supplemental Fig S5). One reason for this difference might be the distinct regulatory mechanisms between animal miRNAs and plant miRNAs: animal miRNAs usually have multiple and imperfectly complementary target sites in 3'

UTRs of mRNAs; by contrast, plant miRNAs usually target CDS of mRNAs through single and highly complementary target site (Voinnet 2009). Here, we found that plant miRNAs did not target extended 3' UTRs, providing additional evidence to the point that plant miRNAs rarely target 3' UTRs.

About 43% of 3' UTR extensions lead to the formation or elongation of overlapping sequences between sense and antisense transcripts (Fig. 4), a mechanism that has been demonstrated to repress sense gene transcription (Faghihi and Wahlestedt 2009). Using T-DNA insertion mutants we found that extended 3' UTRs can function as antisense transcripts to repress expression of sense genes. This result is not consistent with previous studies showing no general anti-correlation of convergently transcribed gene pairs in *Arabidopsis* (Wu et al. 2011; Sherstnev et al. 2012). One reason for this divergence might be that their analysis cannot distinguish whether the expression changes of sense genes is due to an increase in levels of antisense transcripts or other factors such as changes in the transcription activity of sense genes. By contrast, in our analysis we have approached the resolution of this issue on two levels: first, to reduce the effects caused by other factors, we assessed their expression variance between dehydration stress and ABA treatment, because expressions of dehydration stress-responsive genes are partially regulated by ABA-dependent pathway (Golldack et al. 2014; Yoshida et al. 2014) and in our study ABA stimulus did not induce 3' UTR extensions; second, as 3' UTR extensions increased in their expression levels over time (Fig. 2), the predicted inhibitory effect on sense gene expression should be stronger in 3 h than in 1 h. Therefore, in our analysis we were able to identify 43 anti-correlated pairs and 74% of them were dose-dependent (Fig. 4C).

In addition, 429 of the other transcript 3' UTR extensions elongated into promoter regions of their downstream genes, or lead to read-through transcription. Among this group a

positive regulation was observed for 81 downstream genes that were lowly expressed in untreated controls (Fig. 5). One putative mechanism to explain this activation is chromatin remodeling: the 3' UTR extensions may help to create open, relaxed chromatin structures for the downstream genes, which is required for transcriptional activation (Bell et al. 2011). Genome-wide mapping of DNase I hypersensitive (DH) sites has been used to identify open chromatin structures in human (Boyle et al. 2008), yeast (Hesselberth et al. 2009) and *Arabidopsis* (Zhang et al. 2012). Analysis of the presence of DH sites (Zhang et al. 2012) within gene promoters showed that the proportion of promoters with DH sites was significantly decreased in these downstream genes compared to that in other genes, in both non-stressed leaf sample (43% vs. 82%, P value = 3.66×10^{-15} , Fisher's exact test) and flower sample (54% vs. 85%, P value = 1.45×10^{-10} , Fisher's exact test). This is consistent with our hypothesis that these downstream genes have closed chromatin structures in untreated controls, and 3' UTR extensions might help them to establish open chromatin structures in dehydration stress, like long noncoding RNAs in the regulation of their neighbor genes (Orom et al. 2010; Wang et al. 2011).

APA is regulated by *cis*-acting poly(A) signals within transcripts and *trans*-acting factors that regulate poly(A) site selections. We propose that weak poly(A) signals, reduced FPA function, together with a FPA-independent pathway contribute to the biogenesis of dehydration stress-induced 3' UTR extensions (Fig. 8). Taken together, our results identify and uncover novel functions and biogenesis of stress-induced 3' UTR extensions, which expand our knowledge of transcriptome complexity in *Arabidopsis* under stress conditions.

METHODS

Accession numbers

The DRS data sets (Duc et al. 2013) had been deposited at the European Nucleotide Archive (ENA) under accession number ERP003245. Sequence data in this article can be found in TAIR databases under the following accession numbers: ACT2 (AT3G18780), COR15A (AT2G42540), COR15B (AT2G42530), DREB1A (AT4G25480), ERF4 (AT3G15210), FAMT (AT3G44860), FPA (AT2G43410), NAXT1 (AT3G45650), NFYA5 (AT1G54160), P5CS1 (AT2G39800), PIF5 (AT3G59060), RD29A (AT5G52310), RD29B (AT5G52300), RGP2 (AT5G15650), SCL13 (AT4G17230), SPL2 (AT1G54150), SRX (AT1G31170), UBL5 (AT5G42300).

Plant materials, growth conditions and treatments

The *Arabidopsis thaliana* ecotype Columbia (Col-0) was used as wild type in this study. All T-DNA insertion lines are in Col-0 background. Detailed methods are described in Supplemental Data S2.

ssRNA-seq data analysis

Strand-specific RNA-seq data were analyzed using TopHat/Cufflinks pipeline (Trapnell et al. 2013). Detailed parameters are described in Supplemental Data S2.

Identification of 3' UTR extensions

A computational pipeline was developed to identify stress-induced 5'/3' UTR extensions (Supplemental Fig S3B). Details are described in Supplemental Data S2.

Coding potential assessment of 3' UTR extensions

Coding Potential Calculator (CPC) (Kong et al. 2007) was used to estimate the coding capacities. The ORFs were predicted using GenScan with *Arabidopsis* specific parameters (Burge and Karlin 1997).

miRNA target site prediction

Mature sequences of 337 *Arabidopsis* miRNAs were obtained from miRBase (release 20) (Kozomara and Griffiths-Jones 2014). miRNA target sites were predicted using psRobot (Wu et al. 2012) with parameters “-ts 2.5 -fp 2 -tp 17 -gl 17 -p 8 -gn 1”.

Identification of sense mRNAs overlapped in sequences with extended 3' UTRs

The sense-antisense pairs were identified using a previously reported method (Wang et al. 2014). Details are described in Supplemental Data S2.

Poly(A) signal analysis

The poly(A) signal was analyzed using a previously reported method (Sherstnev et al. 2012). The AAUAAA-like motif refers to the canonical AAUAAA signal and 18 of single-nucleotide mutations. Details are described in Supplemental Data S2.

GO enrichment analysis

GO enrichment analysis was performed using agriGO (Du et al. 2010) with TAIR10 annotation as the background. The smaller the *P* value is, the more the GO term is significantly enriched. The top enriched GO terms are those with the lowest *P* values.

Reanalysis of DRS data sets

DRS reads were analyzed using HeliSphere software (version 1.1.498.63) with previously described parameters (Duc et al. 2013). Details are described in Supplemental Data S2.

RT-PCR and RNA gel blot analysis of gene expression

Detailed methods of reverse transcription, semiquantitative PCR, real-time PCR and RNA gel blot analysis are described in Supplemental Data S2. Primers used in all RT-PCR experiments are listed in Supplemental Table S7.

DATA ACCESS

A total of 15 ssRNA-seq data sets generated in this work have been submitted to the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo>) under accession number GSE74864.

ACKNOWLEDGMENTS

We thank Connie Zhao and Bin Zhang (Genomics Resource Center, The Rockefeller University) for technical support; Li-Fang Huang for genotyping and RNA extraction; Prof. Scott Michaels (Departments of Biology/Molecular and Cellular Biochemistry, Indiana University) for providing *fpa-7* mutants; Jun Liu and Huan Wang for suggestions about the bioinformatics analysis; Shulin Deng and other lab members for helpful discussion. This work was supported by Kleinwanzlebener Saatzucht (KWS) SAAT AG, Germany.

AUTHOR CONTRIBUTIONS

H.-X.S., Y.L., and N.-H.C. conceived this project. H.-X.S. collected data and performed bioinformatics and statistical analysis. H.-X.S., Y.L., and Q.-W.N. performed experiments. H.-X.S., Y.L., and N.-H.C. wrote the manuscript.

FIGURE LEGENDS

Figure 1. Dehydration stress induces 3' UTR extensions.

(A) Examples of dehydration stress-induced 3' UTR extensions. Blue and pink colors represent read abundance of forward strand and reverse strand, respectively. Extended

3' UTRs are depicted by red rectangles. **(B)** Number of transcripts with 5'/3' UTR extensions in ABA treatment and dehydration stress. **(C)** Schematic diagram of RT-PCR primers used to validate 3' UTR extensions. **(D)** Validation of dehydration stress-induced 3' UTR extensions (A) using RT-PCR. Primer set P1+P2 indicates transcripts with canonical and extended 3' UTRs. Primer sets P3+P4 and P5+P6 indicate extended 3' UTRs. *ACT2* mRNA served as a positive control. **(E)** Validation of *ERF4* 3' UTR extension using RNA gel blot analysis. ³²P labeled riboprobe targeting *ERF4* 3' UTR region were used to detect both the canonical and extended *ERF4* transcripts. Methylene blue-stained ribosomal RNA (*rRNA*) is shown as loading control. The percentage (%) of extended *ERF4* in total *ERF4* transcripts are presented as mean $\pm$ SD of two biological replicates.

Figure 2. Increased expression levels of extended 3' UTRs over time.

(A) Venn diagram of 3' UTR extensions induced in dehydration stress for 1 h and 3 h. **(B)** Scatter plot showing expression levels (FPKMs) of 631 3' UTR extensions induced by dehydration stress for both 1 h (x-axis) and 3 h (y-axis). **(C)** GO enrichment analysis of 1,409 transcripts with 3' UTR extensions in dehydration stress. Top 15 (with the lowest *P* values) enriched GO terms of biological process category are shown.

Figure 3. Extended 3' UTRs have characteristics of long noncoding RNAs.

(A) Length distribution of 3' UTR extensions, lincRNAs, protein coding transcripts and pri-miRNAs. **(B)** Coding potential assessment of 3' UTR extensions, lincRNAs, protein coding transcripts and pri-miRNAs. Transcripts with higher coding potential have higher CPC scores. **(C)** Cumulative distribution of the longest ORF length across 3' UTR extensions, lincRNAs, protein coding transcripts and pri-miRNAs. **(D)** Cumulative distribution of expression levels (FPKMs) across 3' UTR extensions, lincRNAs, protein coding transcripts and pri-miRNAs. The fold changes of median expression levels for

protein coding transcripts relative to 3' UTR extensions (bottom), pri-miRNAs (middle) and lincRNAs (top) are shown.

Figure 4. Repression of expression levels of sense genes by extended 3' UTRs.

(A) Composition of overlapping regions between sense and antisense transcripts. A total of 605 pairs between sense transcripts and antisense transcripts (canonical 3' UTRs and extended 3' UTRs) were used for the plot. **(B)** Coverage of sense transcripts overlapped by canonical 3' UTRs and extended 3' UTRs. Red line indicates the average of coverage. The coverage is defined as the length of overlapping region divided by the length of sense gene, ranging from 0% to 100%. **(C)** Fold change of expression levels (FPKM) of 3' UTR extensions (left) or transcripts without 3' UTR extensions (right) versus fold change of expression levels (relative to ABA treatment) of sense transcripts in Dehy_3h compared with Dehy_1h. Log₂-transformed fold changes were used for this plot. **(D)** The IGV genome browser view showing *SPL2* and *NFYA5*. Blue and pink colors represent read abundance of forward strand and reverse strand, respectively. Extended 3' UTR is depicted by a red rectangle. T-DNA insertion site is indicated by a black inverted triangle. **(E)** Real-time PCR analysis of expression levels of *SPL2* (top), *SPL2* extended 3' UTR (middle) and *NFYA5* (bottom) in WT and *sp2* mutant. Transcript levels were normalized to *ACT2* expression. The expression level of each gene in WT plants grown in normal condition (NC) was set to 1. Data are shown as mean $\pm$ s.d. (n = 3) of one representative result of three independent biological replicates. Asterisks indicate statistically significant differences (* P < 0.05, ** P < 0.01; *** P < 0.001; *t* test). N.S., not significant.

Figure 5. Activation or read-through transcription of downstream genes by 3' UTR extensions.

(A) Composition of upstream-downstream gene pairs of canonical 3' UTRs and extended 3' UTRs. A total of 429 gene pairs were used for the plot. **(B)** Fold change of expression

levels (FPKM) of 3' UTR extensions (x-axis) versus fold change of expression levels of sense genes (y-axis) in Dehy_3h compared with Dehy_1h. Log₂-transformed fold changes were used for this plot. **(C)** The IGV genome browser view showing *NAXT1* and *AT3G45660*. Blue color represents read abundance of forward strand. Extended 3' UTRs is depicted by a red rectangle. T-DNA insertion site is indicated by a black inverted triangle. **(D)** Real-time PCR analysis of expression levels of *NAXT1* (top), *NAXT1* extended 3' UTR (middle) and *AT3G45660* (bottom) in WT and *naxt1-1* mutant. Transcript levels were normalized to *ACT2* expression. The expression level of each gene in WT plants grown in normal condition (NC) was set to 1. Data are shown as mean $\pm$ s.d. (n = 3) of one representative result of three independent biological replicates. Asterisks indicate statistically significant differences (* P < 0.05, ** P < 0.01; *** P < 0.001; *t* test). N.S., not significant.

Figure 6. mRNAs with 3' UTR extensions tend to have weaker poly(A) signals.

(A) Composition of proximal cleavage sites of 1,409 transcripts with 3' UTR extensions, distal cleavage sites of 1,409 transcripts with 3' UTR extensions and cleavage sites of expressed transcripts without 3' UTR extensions. **(B)** Frequency of occurrence and distribution of AAUAAA-like motifs across 50 nt upstream sequences of cleavage sites. The -20 locus is indicated by red line. **(C-E)** Nucleotide composition of 100 nt up- and downstream flanking sequences of: **(C)** proximal cleavage sites of 1,409 transcripts with 3' UTR extensions; **(D)** distal cleavage sites of 1,409 transcripts with 3' UTR extensions; and **(E)** cleavage sites of expressed transcripts without 3' UTR extensions. The -20 locus, -7 locus and -1 locus are indicated by red lines. N.S., not significant.

Figure 7. FPA partially regulates biogenesis of 3' UTR extensions.

(A) Schematic diagram showing domains of the full-length FPA protein (top), intron retention-produced truncated FPA protein (middle) and proximal

polyadenylation-produced FPA protein (bottom). **(B)** Proportion of ABA treatment- and dehydration stress-induced 3' UTR extensions among 3' UTR extensions induced in *fpa-7* mutant. **(C)** Real-time PCR analysis of total expression levels of *PIF5*, *ERF4*, *UBL5*, *SRX* and *AT2G31040* in WT and *fpa-7* mutant. The expression level of each gene in WT plants grown in normal condition (NC) was set to 1. **(D)** Real-time PCR analysis showing induction of 3' UTR extensions by dehydration stress of *PIF5*, *ERF4*, *UBL5*, *SRX* and *AT2G31040* in WT and *fpa-7* mutant. The expression levels of each gene in WT plants and *fpa-7* mutants grown in normal condition (NC) were set to 1. **(E)** The IGV genome browser view showing *FPA*. Pink color represents read abundance of reverse strand. Retained intron is depicted by a red rectangle. The position of introduced premature termination codon is indicated by a red inverted triangle. **(F)** Real-time PCR analysis of expression levels of intron-retained *FPA*. The expression level in normal condition (NC) was set to 1. Transcript levels were normalized to *ACT2* expression. Data are shown as mean $\pm$ s.d. ($n = 3$) of one representative result of three independent biological replicates. Asterisks indicate statistically significant differences compared with WT (* $P < 0.05$, ** $P < 0.01$; *** $P < 0.001$; t test).

Figure 8. A proposed model for the biogenesis and function of dehydration stress-induced 3' UTR extensions.

Biogenesis of 3' UTR extensions depends on *cis*-acting poly(A) signals, *trans*-acting regulator FPA and a FPA-independent pathway. In normal growth condition, transcripts are polyadenylated at the proximal poly(A) site regardless of the strengths of poly(A) signals, whereas in dehydration-stressed condition, reduced FPA function causes 3' UTR extensions of transcripts only if they have weaker poly(A) signals, which might be mediated by the competition of truncated FPA proteins. In addition, a FPA-independent pathway also contributes to the biogenesis of 3' UTR extensions. The extended 3' UTRs have characteristics of long noncoding RNAs and can function as repressors of their

sense genes or activators of their downstream genes.

REFERENCES

- Baurle I, Dean C. 2008. Differential interactions of the autonomous pathway RRM proteins and chromatin regulators in the silencing of Arabidopsis targets. *PLoS one* **3**(7): e2733.
- Baurle I, Smith L, Baulcombe DC, Dean C. 2007. Widespread role for the flowering-time regulators FCA and FPA in RNA-mediated chromatin silencing. *Science* **318**(5847): 109-112.
- Bava FA, Eliscovich C, Ferreira PG, Minana B, Ben-Dov C, Guigo R, Valcarcel J, Mendez R. 2013. CPEB1 coordinates alternative 3'-UTR formation with translational regulation. *Nature* **495**(7439): 121-125.
- Bell O, Tiwari VK, Thoma NH, Schubeler D. 2011. Determinants and dynamics of genome accessibility. *Nat Rev Genet* **12**(8): 554-564.
- Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK. 2005. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* **123**(7): 1279-1291.
- Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**(2): 311-322.
- Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *Journal of molecular biology* **268**(1): 78-94.
- Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25**(18): 1915-1927.
- Claeys H, Inze D. 2013. The agony of choice: how plants balance growth and survival under water-limiting conditions. *Plant physiology* **162**(4): 1768-1779.
- Derti A, Garrett-Engle P, Macisaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, Babak T. 2012. A quantitative atlas of polyadenylation in five mammals. *Genome Res* **22**(6): 1173-1183.
- Di C, Yuan J, Wu Y, Li J, Lin H, Hu L, Zhang T, Qi Y, Gerstein MB, Guo Y et al. 2014. Characterization of stress-responsive lncRNAs in Arabidopsis thaliana by integrating expression, epigenetic and structural features. *Plant J* **80**(5): 848-861.
- Di Giammartino DC, Nishida K, Manley JL. 2011. Mechanisms and consequences of alternative polyadenylation. *Mol Cell* **43**(6): 853-866.
- Du Z, Zhou X, Ling Y, Zhang Z, Su Z. 2010. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic acids research* **38**(Web Server issue): W64-70.
- Duc C, Sherstnev A, Cole C, Barton GJ, Simpson GG. 2013. Transcription termination and chimeric RNA formation controlled by Arabidopsis thaliana FPA. *PLoS Genet* **9**(10): e1003867.
- Elkon R, Drost J, van Haaften G, Jenal M, Schrier M, Oude Vrielink JA, Agami R. 2012. E2F mediates enhanced alternative polyadenylation in proliferation. *Genome biology* **13**(7): R59.
- Elkon R, Ugalde AP, Agami R. 2013. Alternative cleavage and polyadenylation: extent, regulation and

- function. *Nat Rev Genet* **14**(7): 496-506.
- Fabian MR, Sonenberg N, Filipowicz W. 2010. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* **79**: 351-379.
- Faghihi MA, Wahlestedt C. 2009. Regulatory roles of natural antisense transcripts. *Nature reviews Molecular cell biology* **10**(9): 637-643.
- Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, Garcia JA, Paz-Ares J. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature genetics* **39**(8): 1033-1037.
- Golldack D, Li C, Mohan H, Probst N. 2014. Tolerance to drought and salt stress in plants: Unraveling the signaling networks. *Frontiers in plant science* **5**: 151.
- Gruber AR, Martin G, Muller P, Schmidt A, Gruber AJ, Gumieny R, Mittal N, Jayachandran R, Pieters J, Keller W et al. 2014. Global 3' UTR shortening has a limited effect on protein abundance in proliferating T cells. *Nat Commun* **5**: 5465.
- Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS et al. 2009. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat Methods* **6**(4): 283-289.
- Hilgers V, Lemke SB, Levine M. 2012. ELAV mediates 3' UTR extension in the Drosophila nervous system. *Genes Dev* **26**(20): 2259-2264.
- 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**(38): 15864-15869.
- Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, Tian B. 2013. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat Methods* **10**(2): 133-139.
- Horniyk C, Duc C, Rataj K, Terzi LC, Simpson GG. 2010a. Alternative polyadenylation of antisense RNAs and flowering time control. *Biochemical Society transactions* **38**(4): 1077-1081.
- Horniyk C, Terzi LC, Simpson GG. 2010b. The spen family protein FPA controls alternative cleavage and polyadenylation of RNA. *Developmental cell* **18**(2): 203-213.
- Jenal M, Elkon R, Loayza-Puch F, van Haaften G, Kuhn U, Menzies FM, Oude Vrielink JA, Bos AJ, Drost J, Rooijers K et al. 2012. The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* **149**(3): 538-553.
- 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**(17): 7028-7033.
- Ji Z, Luo W, Li W, Hoque M, Pan Z, Zhao Y, Tian B. 2011. Transcriptional activity regulates alternative cleavage and polyadenylation. *Mol Syst Biol* **7**: 534.
- Ji Z, Tian B. 2009. Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PloS one* **4**(12): e8419.
- Katiyar-Agarwal S, Gao S, Vivian-Smith A, Jin H. 2007. A novel class of bacteria-induced small RNAs in Arabidopsis. *Genes Dev* **21**(23): 3123-3134.
- Katiyar-Agarwal S, Morgan R, Dahlbeck D, Borsani O, Villegas A, Jr., Zhu JK, Staskawicz BJ, Jin H. 2006. A pathogen-inducible endogenous siRNA in plant immunity. *Proceedings of the National Academy*

of *Sciences of the United States of America* **103**(47): 18002-18007.

Kim YK, Furic L, Parisien M, Major F, DesGroseillers L, Maquat LE. 2007. Staufen1 regulates diverse classes of mammalian transcripts. *The EMBO journal* **26**(11): 2670-2681.

Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G. 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic acids research* **35**(Web Server issue): W345-349.

Kozomara A, Griffiths-Jones S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research* **42**(Database issue): D68-73.

Li WX, Oono Y, Zhu J, He XJ, Wu JM, Iida K, Lu XY, Cui X, Jin H, Zhu JK. 2008. The Arabidopsis NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. *The Plant cell* **20**(8): 2238-2251.

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X et al. 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**(7221): 464-469.

Liu F, Marquardt S, Lister C, Swiezewski S, Dean C. 2010. Targeted 3' processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. *Science* **327**(5961): 94-97.

Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua NH. 2012. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. *The Plant cell* **24**(11): 4333-4345.

Lyons R, Iwase A, Gansewig T, Sherstnev A, Duc C, Barton GJ, Hanada K, Higuchi-Takeuchi M, Matsui M, Sugimoto K et al. 2013. The RNA-binding protein FPA regulates flg22-triggered defense responses and transcription factor activity by alternative polyadenylation. *Scientific reports* **3**: 2866.

Mayr C, Bartel DP. 2009. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**(4): 673-684.

Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q et al. 2010. Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**(1): 46-58.

Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, Fan L, Sandelin A, Rinn JL, Regev A et al. 2012. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res* **22**(3): 577-591.

Quesada V, Dean C, Simpson GG. 2005. Regulated RNA processing in the control of Arabidopsis flowering. *The International journal of developmental biology* **49**(5-6): 773-780.

Ron M, Alandete Saez M, Eshed Williams L, Fletcher JC, McCormick S. 2010. Proper regulation of a sperm-specific cis-nat-siRNA is essential for double fertilization in Arabidopsis. *Genes Dev* **24**(10): 1010-1021.

Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. 2008. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **320**(5883): 1643-1647.

Schomburg FM, Patton DA, Meinke DW, Amasino RM. 2001. FPA, a gene involved in floral induction in Arabidopsis, encodes a protein containing RNA-recognition motifs. *The Plant cell* **13**(6): 1427-1436.

- Sherstnev A, Duc C, Cole C, Zacharaki V, Hornyik C, Oszolak F, Milos PM, Barton GJ, Simpson GG. 2012. Direct sequencing of *Arabidopsis thaliana* RNA reveals patterns of cleavage and polyadenylation. *Nat Struct Mol Biol* **19**(8): 845-852.
- Simpson GG, Quesada V, Henderson IR, Dijkwel PP, Macknight R, Dean C. 2004. RNA processing and *Arabidopsis* flowering time control. *Biochemical Society transactions* **32**(Pt 4): 565-566.
- Sonmez C, Baurle 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**(20): 8508-8513.
- Sonmez C, Dean C. 2012. Transcription beyond borders has downstream consequences. *RNA biology* **9**(2): 143-147.
- Sun J, Lin Y, Wu J. 2013. Long non-coding RNA expression profiling of mouse testis during postnatal development. *PloS one* **8**(10): e75750.
- 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*. *The Plant cell* **24**(11): 4376-4388.
- Tian B, Manley JL. 2013. Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem Sci* **38**(6): 312-320.
- Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology* **31**(1): 46-53.
- Voinnet O. 2009. Origin, biogenesis, and activity of plant microRNAs. *Cell* **136**(4): 669-687.
- Wang H, Chung PJ, Liu J, Jang IC, Kean MJ, Xu J, Chua NH. 2014. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Res* **24**(3): 444-453.
- Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA et al. 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**(7341): 120-124.
- Wu HJ, Ma YK, Chen T, Wang M, Wang XJ. 2012. PsRobot: a web-based plant small RNA meta-analysis toolbox. *Nucleic acids research* **40**(Web Server issue): W22-28.
- Wu HJ, Wang ZM, Wang M, Wang XJ. 2013. Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. *Plant physiology* **161**(4): 1875-1884.
- 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**(30): 12533-12538.
- Yoshida T, Mogami J, Yamaguchi-Shinozaki K. 2014. ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. *Current opinion in plant biology* **21**: 133-139.
- Zhang W, Zhang T, Wu Y, Jiang J. 2012. Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in *Arabidopsis*. *The Plant cell* **24**(7): 2719-2731.
- Zhang Y, Gu L, Hou Y, Wang L, Deng X, Hang R, Chen D, Zhang X, Zhang Y, Liu C et al. 2015. Integrative genome-wide analysis reveals HLP1, a novel RNA-binding protein, regulates plant flowering by targeting alternative polyadenylation. *Cell research* **25**(7): 864-876.
- Zhang Y, Rataj K, Simpson GG, Tong L. 2016. Crystal Structure of the SPOC Domain of the *Arabidopsis*

803 Flowering Regulator FPA. *PLoS one* **11**(8): e0160694.

804 Zhu QH, Stephen S, Taylor J, Helliwell CA, Wang MB. 2013. Long noncoding RNAs responsive to *Fusarium*

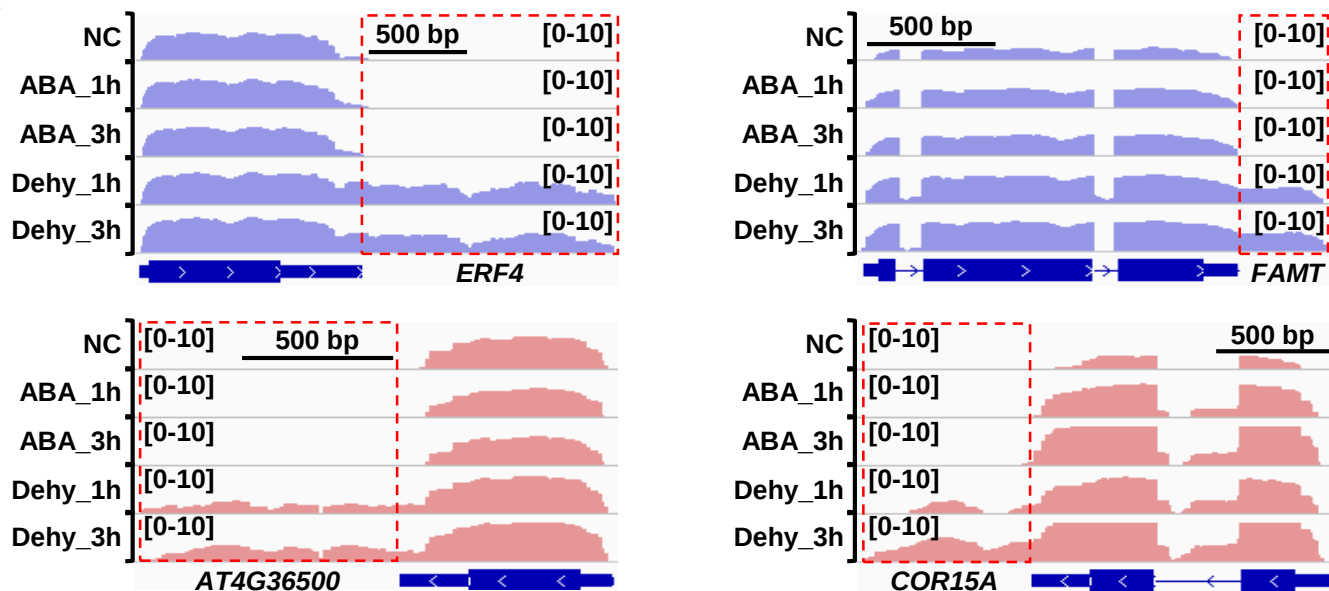
805 *oxysporum* infection in *Arabidopsis thaliana*. *New Phytol.*

806

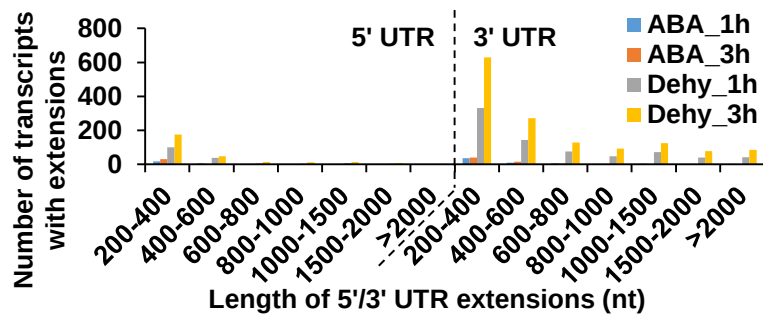
807

Figure 1

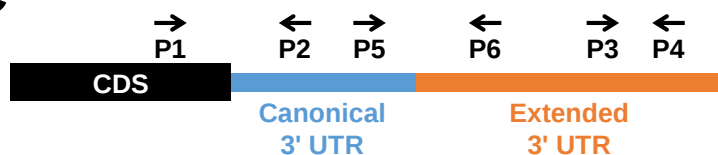
A



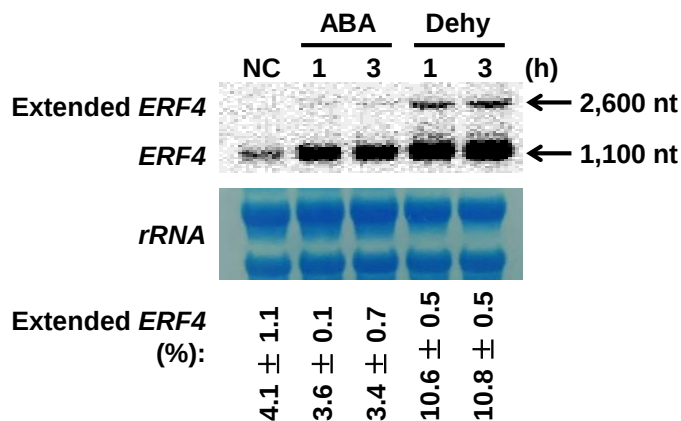
B



C



E



D

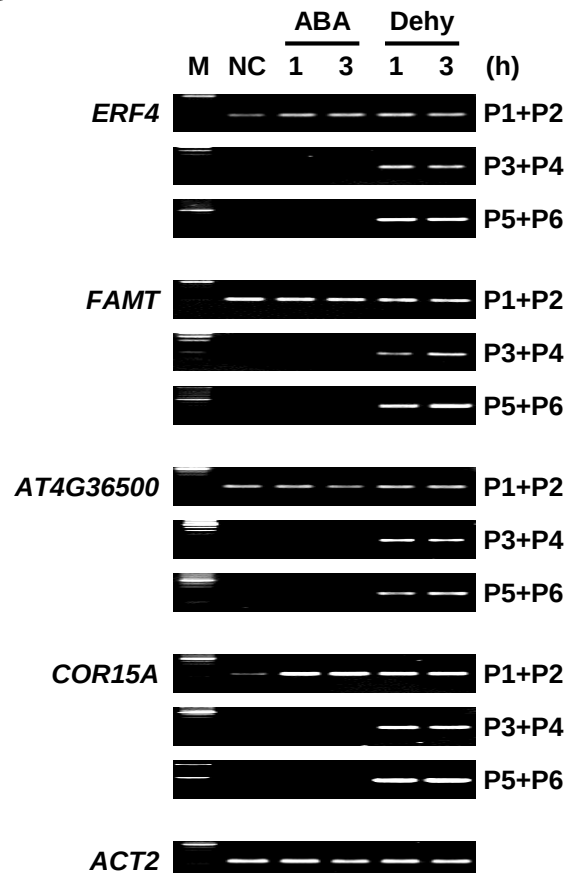


Figure 2

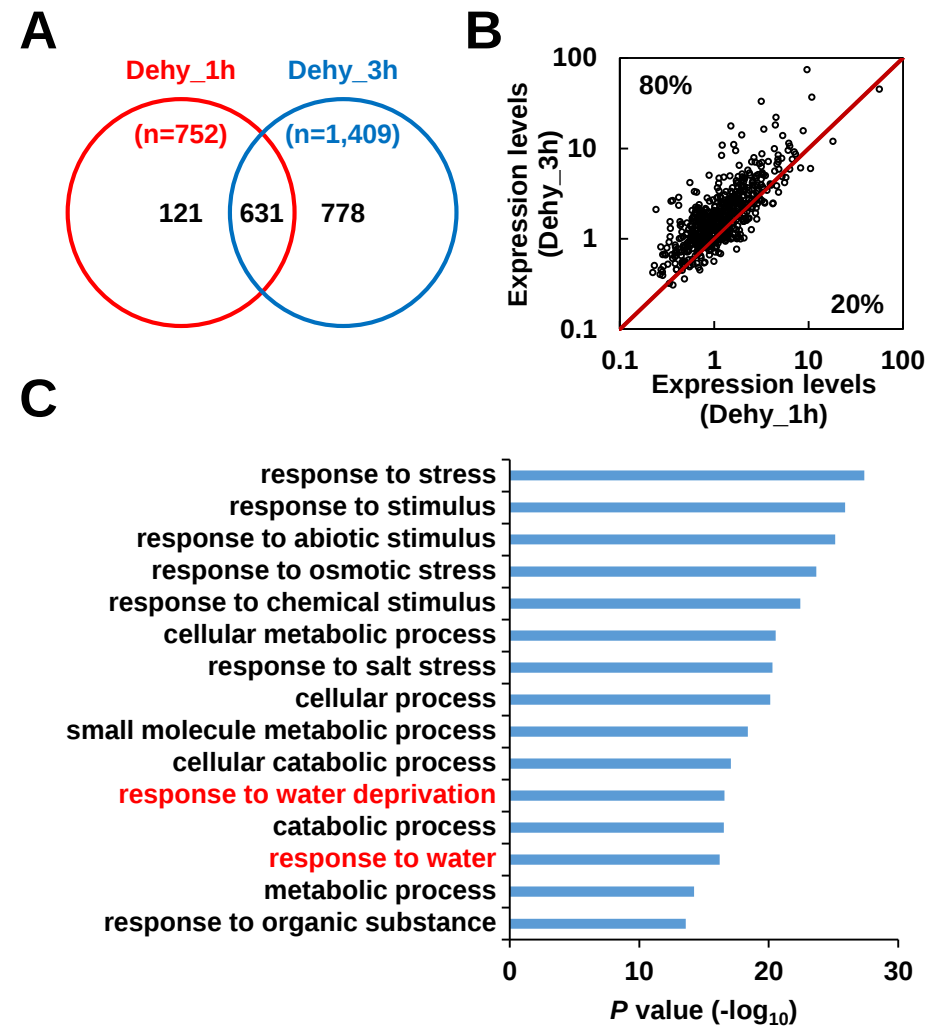


Figure 3

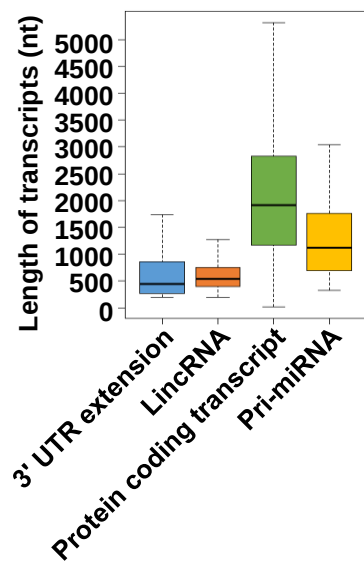
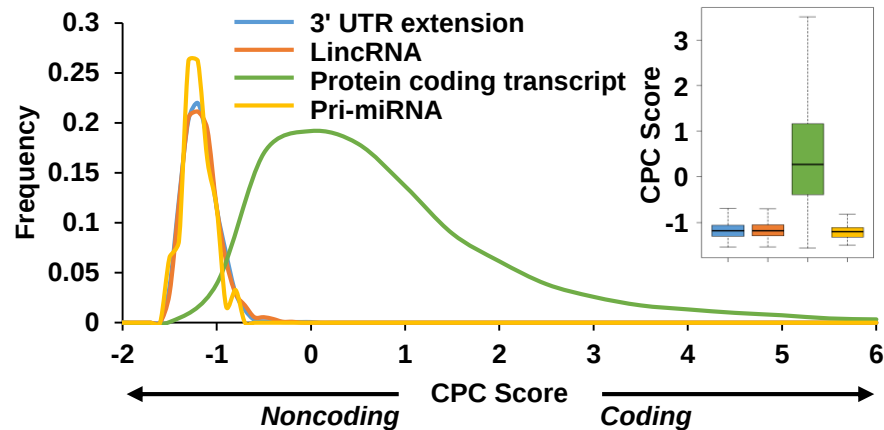
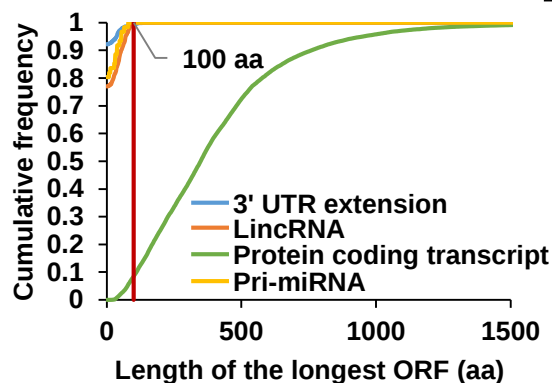
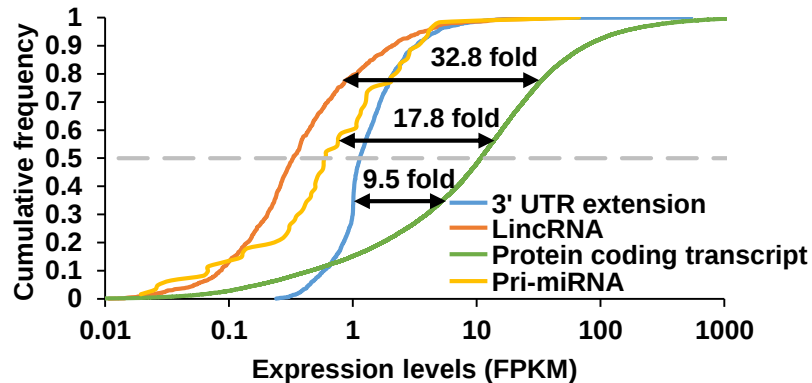
A**B****C****D**

Figure 4

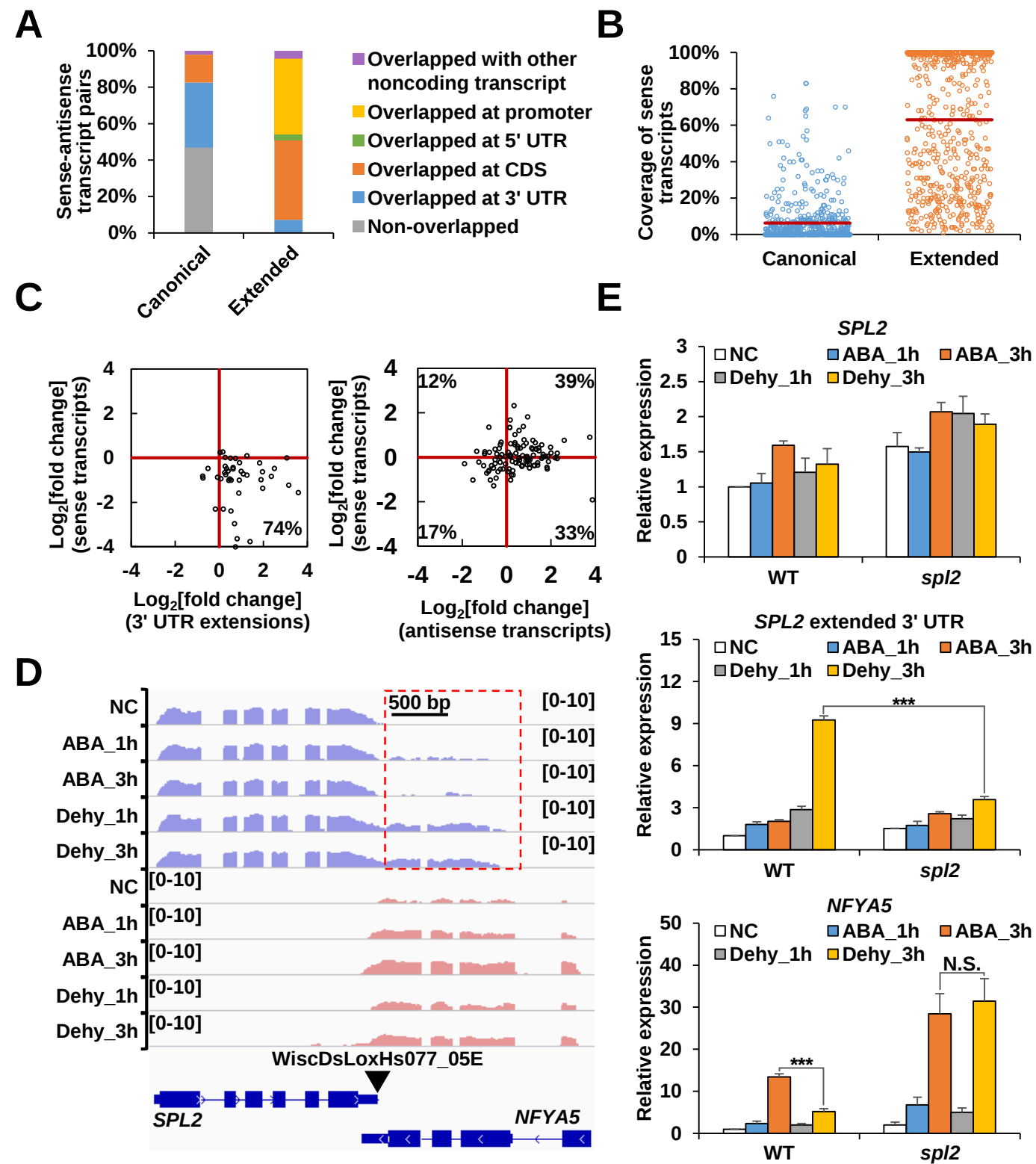
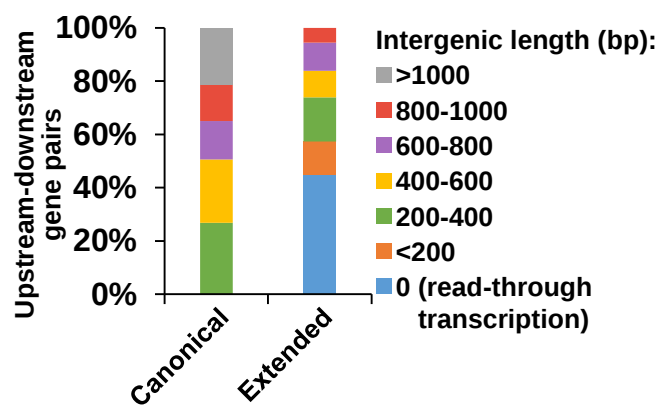
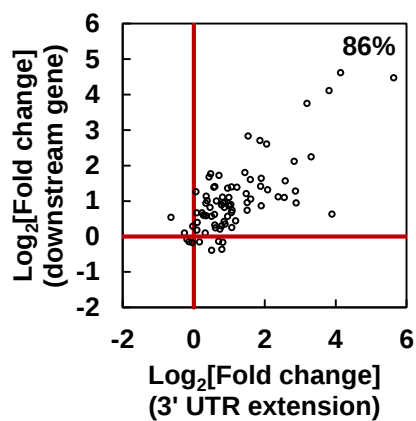


Figure 5

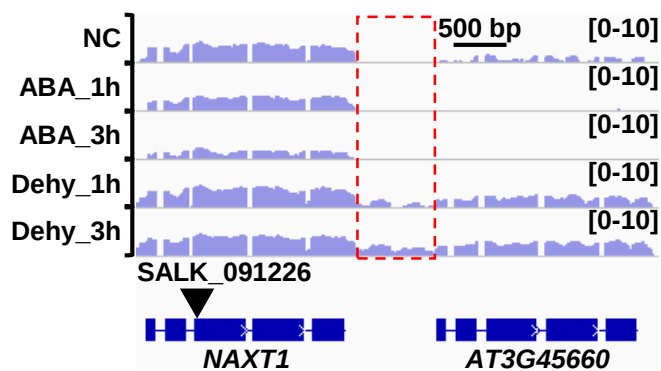
A



B



C



D

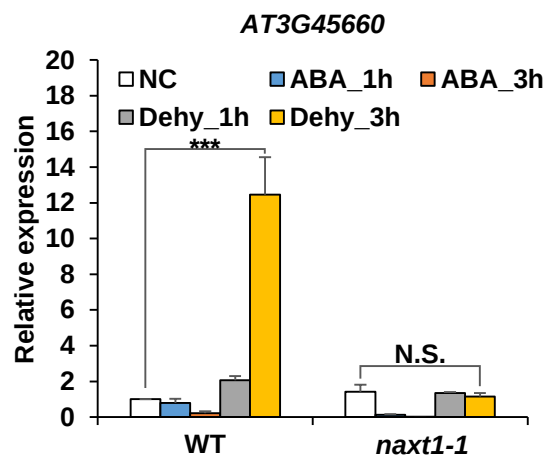
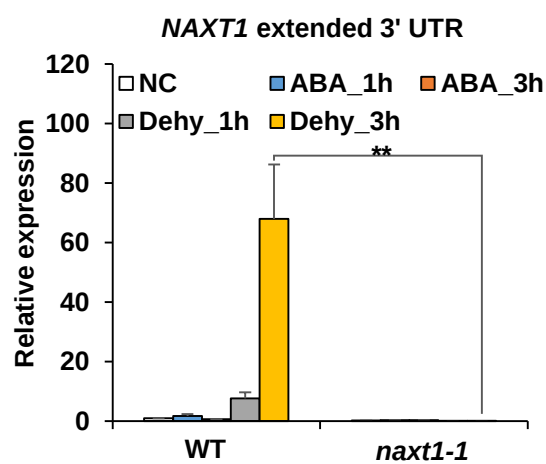
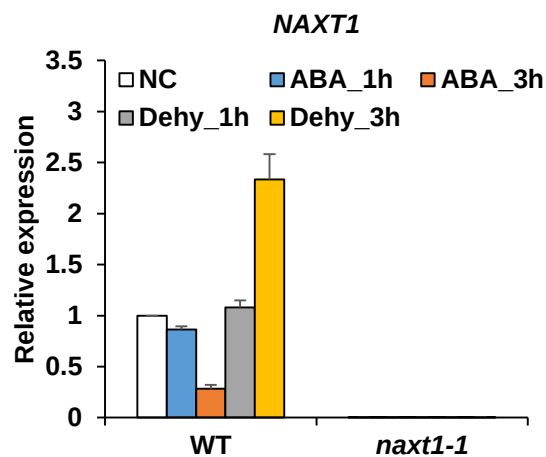
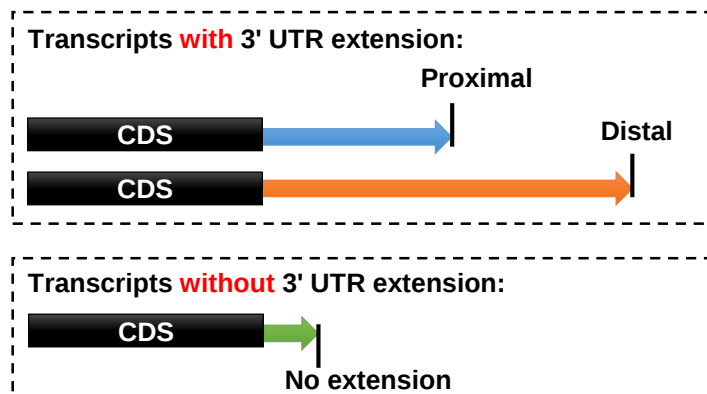
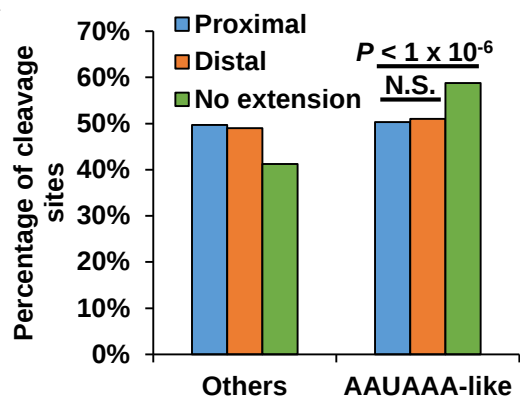
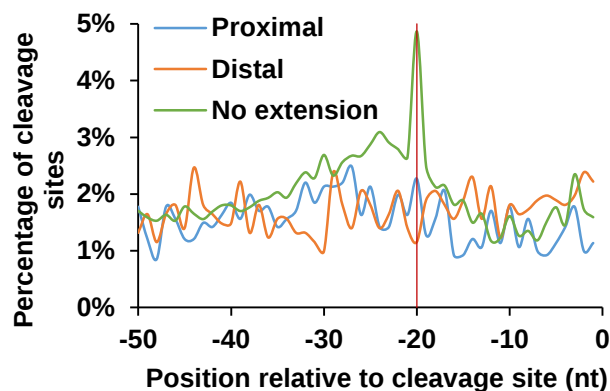


Figure 6

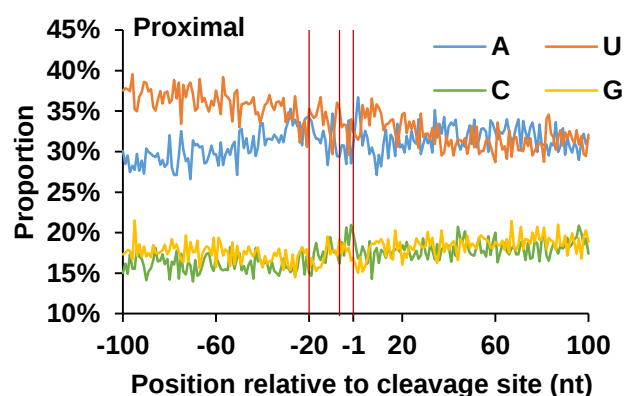
A



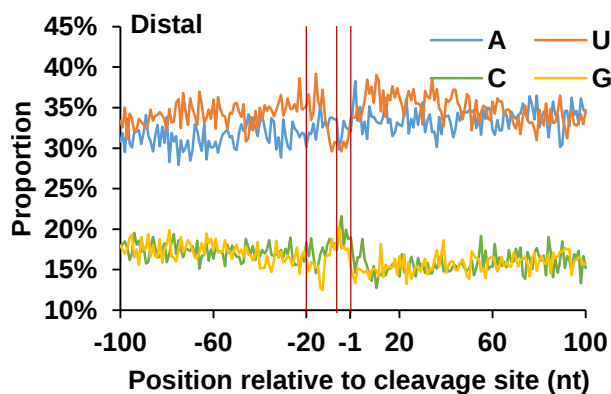
B



C



D



E

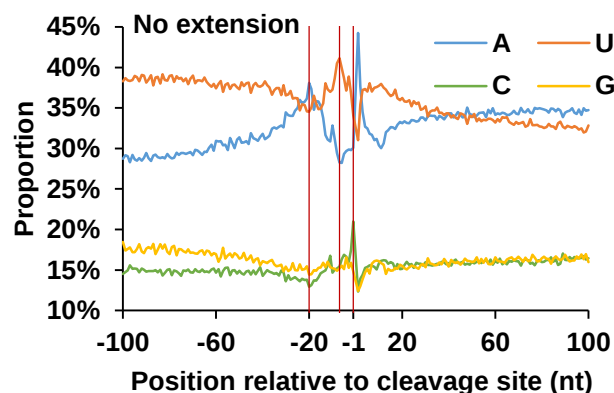
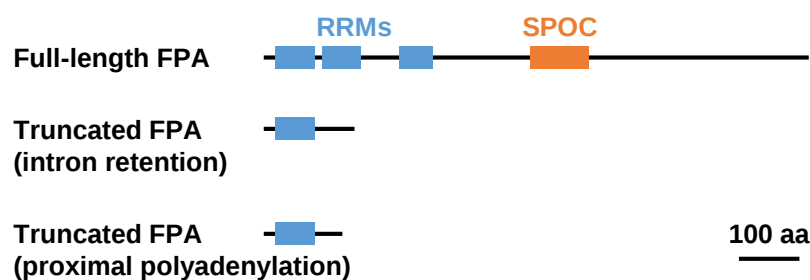
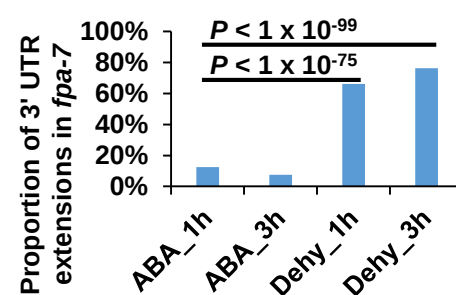


Figure 7

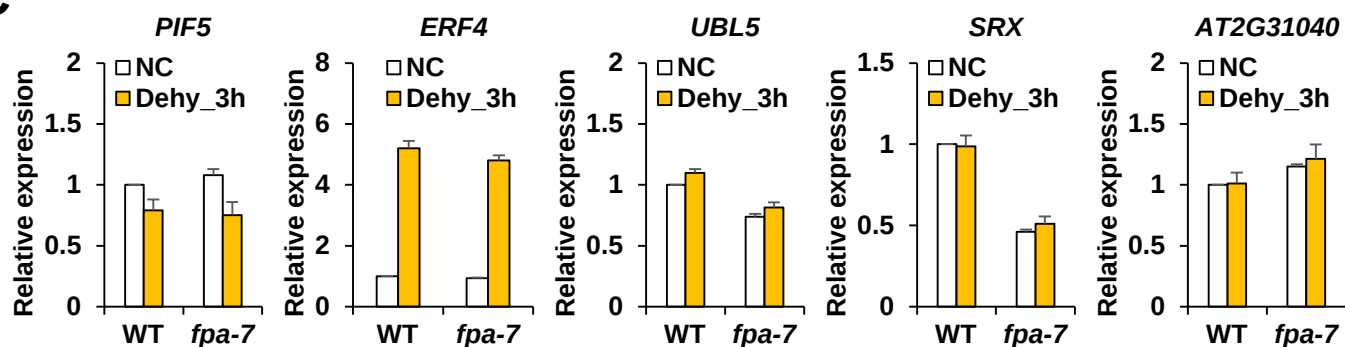
A



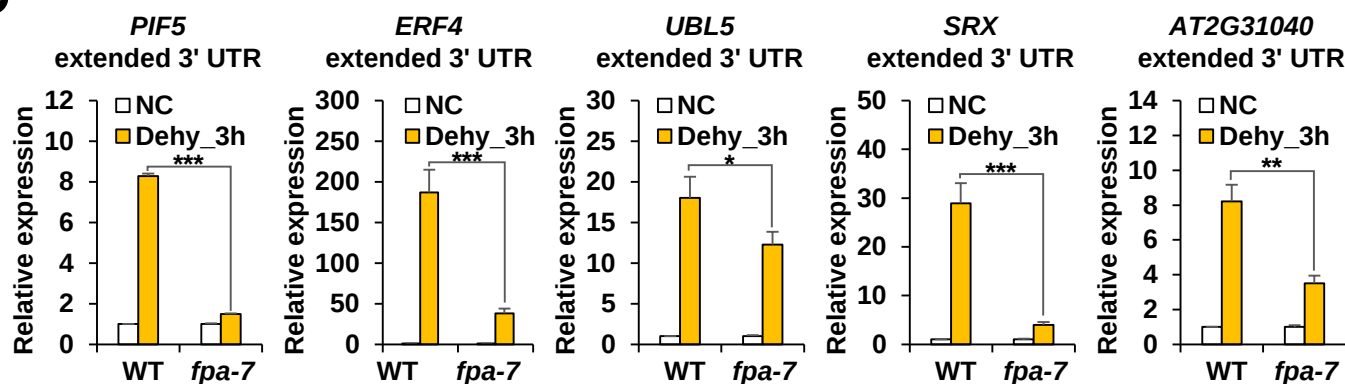
B



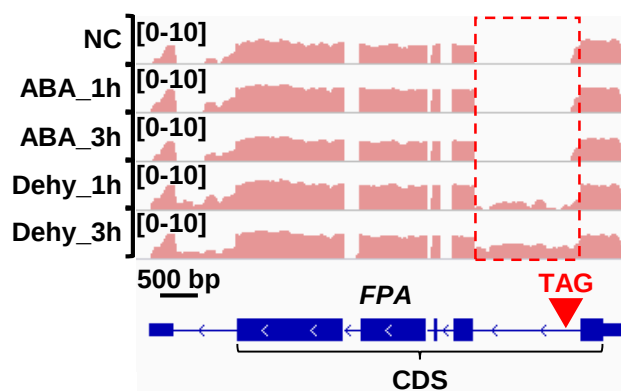
C



D



E



F

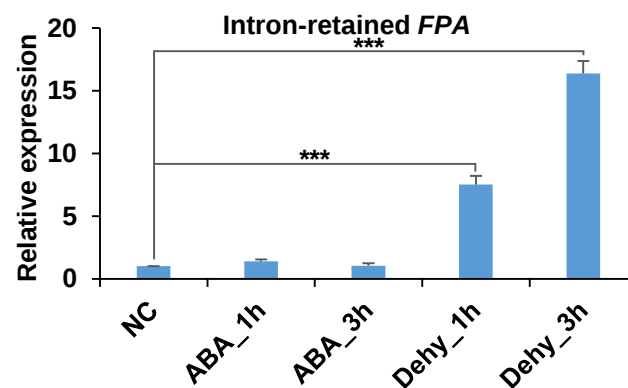


Figure 8

