Zhibo Yu, Juncheng Lin, **Qingshun Q. Li**. 2019. Transcriptome analyses of FY mutants reveal its role in mRNA alternative polyadenylation. ***Plant Cell***, 31(10):2332-2352. DOI: 10.1105/tpc.18.00545

LARGE-SCALE BIOLOGY ARTICLE

Transcriptome Analyses of FY Mutants Reveal its Role in mRNA Alternative Polyadenylation

Zhibo Yu[1], Juncheng Lin [1,2], Qingshun Quinn Li[1,2] *

[1]Key Laboratory of the Ministry of Education for Coastal and Wetland Ecosystems, College of the Environment and Ecology, Xiamen University, Xiamen, Fujian, China 361102
[2]Graduate College of Biomedical Sciences, Western University of Health Sciences, Pomona, CA 91766, USA

* Corresponding author: Q. Quinn Li, College of the Environment and Ecology, Xiamen University, Xiamen, Fujian, China 361102; Phone: 86-592-218-9836; email: liqq@xmu.edu.cn

Short title： Role of FY in polyadenylation.

One-sentence summary: FY is directly involved in poly(A) signal recognition and affects genome-wide poly(A) site usage through alternative polyadenylation.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Q. Quinn Li (liqq@xmu.edu.cn).

ABSTRACT

A crucial step for mRNA polyadenylation is poly(A) signal recognition by trans-acting factors. The mammalian cleavage and polyadenylation specificity factor (CPSF) complex components CPSF30 and WDR33 recognize the canonical AAUAAA signal for efficient polyadenylation. In *Arabidopsis thaliana*, the flowering time regulator FY is the homologue of WDR33. However, its role in mRNA polyadenylation is poorly understood.

Using poly(A) tag sequencing, we found that over 50% of alternative polyadenylation (APA) events are altered in *fy* single mutants or double mutants with *Atcpsf30*, but mutation of the FY WD40-repeat has a stronger effect than deletion of the plant-unique PPLPP domain. *fy* mutations disrupt AAUAAA or AAUAAA-like poly(A) signal recognition. Notably, A-rich signal usage is suppressed in the WD40-repeat mutation, but promoted in Pro-Pro-Leu-Pro-Pro (PPLPP)-domain deficiency. However, *fy* mutations do not aggravate the alteration of signal usage in the *Atcpsf30* null mutant. Furthermore, the WD40-repeat mutation shows a preference for 3'UTR shortening, but the PPLPP-domain deficiency shows a preference for lengthening. Interestingly, the WD40-repeat mutant exhibits shortened primary roots and late flowering with alteration of APA of related genes. Importantly, the long transcripts of two APA genes affected in *fy* are related to abiotic stress responses. These results reveal a conserved and specific role of FY in mRNA polyadenylation.

INTRODUCTION

Polyadenylation of eukaryotic mRNA is an essential posttranscriptional process achieved by poly(A) signal recognition, cleavage, and the addition of a poly(A) tail (Colgan and Manley, 1997; Elkon R, 2013). Polyadenylation affects mRNA stability, nuclear export, and translation initiation (Tian and Manley, 2017). At least 50% of genes in humans, animals, algae and plants have more than one poly(A) cleavage site (Wu et al., 2011; Derti et al., 2012; Smibert et al., 2012; Ulitsky et al., 2012; Zhao et al., 2014; Fu et al., 2016). This common phenomenon is designated alternative polyadenylation (APA), which increases the complexity and diversity of transcriptomes and proteomes. In humans, APA affects immunity regulation, cancer formation, and cell reprogramming (Mayr and Bartel, 2009; Fu et al., 2011; Lin et al., 2012; Carpenter et al., 2014). In plants, APA functions in disease resistance, flowering time control, symbiosis, development, and reproduction (Bruggeman and Delarue, 2014; Liu et al., 2014; Zhang et al., 2015; Cyrek et al., 2016;

Pan et al., 2016; Lin et al., 2017; Zhou et al., 2019; Riester et al., 2019; Zeng et al., 2019).

During polyadenylation, poly(A) signals anchor the position of a cleavage site that eventually becomes the poly(A) site. In mammalian cells, four parts of the poly(A) signals are located around the pre-mRNA cleavage site. The predominant AAUAAA hexamer is located between 10 and 30 nucleotides (nt) upstream of the cleavage site. More than 50% of transcripts in humans preferentially use the AAUAAA poly(A) signal (Neve et al., 2016). The sequence elements at the cleavage site and its downstream element (DSE) are mainly composed of the dinucleotide CA and U/GU-rich sequences, respectively. Some genes have an upstream element (USE, before the AAUAAA) with UGUA-containing hexamers or other similar repeats (Shi and Manley, 2015).

In plants, poly(A) signals consist of three major elements (Loke et al., 2005). Near upstream elements (NUE) predominantly consist of AAUAAA. However, this canonical signal is less conserved in plants and embedded in only about 10% of transcripts (Loke et al., 2005). The far upstream element (FUE) is a U-rich signal that is similar to USEs in humans. A cleavage element (CE) resides on both sides of the cleavage site and includes two U-rich regions, which is different from the elements at the cleavage site in human mRNAs. However, plant signals lack a DSE (Loke et al., 2005). These similarities and differences indicate that the underlying mechanism of polyadenylation between mammals and plants is conserved but exhibits variability.

Poly(A) signals are recognized by trans-acting factors (Clerici et al., 2017; Clerici et al., 2018; Sun et al., 2018). In mammals, biochemical studies have shown that pre-mRNA 3’ end processing requires four multi-unit protein complexes, cleavage and polyadenylation specificity factor (CPSF), cleavage stimulatory factor (CstF), cleavage factor I (CFIm) and cleavage factor II (CFIIm), in addition to the single subunit poly(A) polymerase (PAP) (Takagaki et al., 1989; Colgan and Manley, 1997). Among these, CPSF, (assembled from CPSF160, WDR33, CPSF100, CPSF73, Fip1, and CPSF30), serves as the central complex for the recognition of the predominant AAUAAA signal and pre-mRNA cleavage (Shi et

al., 2009; Michalski and Steiniger, 2015). CPSF30 and WDR33 directly bind to the AAUAAA signal (Chan et al., 2014; Schönemann et al., 2014). Recent studies have demonstrated that the CPSF160-WDR33-CPSF30 ternary complex has a high affinity for the AAUAAA signal, and CPSF160 functions as an essential scaffold that organizes CPSF30 and WDR33 to bind AAUAAA (Clerici et al., 2017; Sun et al., 2018). In addition, the A1 and A2 bases of AAUAAA are recognized specifically by the zinc finger 2 (ZF2) of CPSF30 as well as A4 and A5 bases are recognized specifically by the ZF3 of CPSF30. WDR33 interacts with RNA at least in part via its N-terminus, and the WD40-repeat of WDR33 contacts with the U3-A6 bases, indicating that the highly conserved WD40-repeat plays an important role in AAUAAA signal recognition (Schönemann et al., 2014; Sun et al., 2018).

Genetic and phylogenetic studies have revealed that polyadenylation trans-acting factors are evolutionarily conserved among eukaryotes (Hunt et al., 2008; Hunt et al., 2012). Notably, genetic evidence has shown that the plant CPSF30 is involved in NUE signal choice in *Arabidopsis thaliana*, where the knockout of *AtCPSF30* leads to a shift from A-rich poly(A) signals to U-rich poly(A) signals (Thomas et al., 2012). However, the role of the WDR33 homolog, FY, in recognizing plant NUE signals remains unclear. FY is known as a flowering time regulator (Simpson et al., 2003), and a subunit of the CPSF complex with a special C-terminus next to seven conserved WD40-repeats (Henderson et al., 2005). The C-terminus harbors two plant unique Pro-Pro-Leu-Pro-Pro (PPLPP) domains, which can interact with the Trp-Trp (WW) domain of the nuclear RNA-binding protein FCA and control plant flowering time (Simpson et al., 2003; Henderson et al., 2005). FCA/FY interaction suppresses FCA protein abundance by promoting the polyadenylation of *FCA* within intron 3 to generate a nonfunctional *FCA-β* transcript. The FCA/FY interaction is also important for properly positioning the polyadenylation site of the floral inhibitor gene *FLOWERING LOCUS C* (*FLC*), and controls flowering time (Henderson et al., 2005; Feng et al., 2011). In addition, FY influences seed dormancy by regulating the APA of *DELAY*

OF GERMINATION1 (Cyrek et al., 2016).

In this report, we used a poly(A) tag sequencing (PAT-seq) approach to study the poly(A) profile in a set of defective *fy* mutants and *fy oxt6* double mutants with the an *AtCPSF30* knockout mutant *oxidative stress tolerant6* (*oxt6*). We demonstrated that FY is indeed involved in AAUAAA signal recognition. Interestingly, we found that the WD40-repeat mutation of FY suppresses, whereas PPLPP-domain deficiency promotes, the A-rich signal recognition. Furthermore, the mutated WD40-repeat of FY shows a preference for using the proximal poly(A) site in 3'UTRs. However, the PPLPP-domain deficiency of FY results in a preference for using the distal poly(A) site in the 3'UTR. Importantly, we provided *in vivo* evidence that the long transcripts of *ARK2* and a zinc ion binding protein affected by FY and AtCPSF30 play roles in salt and oxidative stress responses. Overall, our results reveal the role of FY in genome-wide mRNA polyadenylation.

RESULTS

The *fy* mutants and poly(A) profiling

FY has seven WD40-repeats in the N terminus and two PPLPP-domains in the C terminus, and previous studies have reported a set of *fy* mutants with significant phenotypic outcomes (Henderson et al., 2005). Among these, *fy-1* (G-A mutation at a splice-acceptor site results in premature termination), *fy-2*, and *fy-5* are defective on the PPLPP-domain (Simpson et al., 2003; Henderson et al., 2005). Figure 1 shows a side-by-side comparison of various *fy* alleles using reverse transcription quantitative PCR (RT-qPCR). Both *fy1* and *fy2* have low transcript abundance and encode proteins that lack two PPLPP-domains, while the *fy5* mutant lacks the last PPLPP-domain (Figure 1). For uncertain reasons, the transcript levels of *fy5* were repeatedly quantified as a lower expression of *FY* than the wild type, which is inconsistent with what was reported before (Feng et al., 2011). The *fy3* allele introduces a glycine to serine (G141S) change at the first WD40-repeat. The expression level of *FY* was not affected in *fy3*, which is consistent with a previous report (Henderson

et al., 2005). A T-DNA insertion on the promoter of *FY* resulted in its overexpression, and this mutant was designated *fy6* in this study to distinguish other FY WD40-repeat or PPLPP-domain mutants (Figure 1). These mutants were crossed with *oxidative stress tolerant6* (*oxt6*, an *AtCPSF30* knockout mutant) to generate double mutants.

Poly(A) tag sequencing (PAT-seq) is an efficient method for genome-wide profiling of poly(A) site usage, mature transcripts abundance, and functional gene expression as described in our previous publications (Fu et al., 2016; Lin et al., 2017; Hong et al., 2018). Thus, PAT-seq was employed to uncover the poly(A) site usage and transcriptomic profiling of mutants described above (Supplemental Figure 1). A summary of the raw reads, mapped PATs, and PACs (poly(A) site clusters) for each library are listed in Table 1. A total of 48,457 PACs were identified as shown in Supplemental Data Set 1. Principal component analysis showed replicates are repeated well (Supplemental Figure 2). Overall, these PACs mapped to 19,601 genes. Of these, 10,351 genes contain more than one PAC (defined as APA genes), reflecting about 53% of APA genes (Supplemental Figure 3A), and 97.5% of the 19,601 genes are protein-coding (Supplemental Figure 3B).

WD40-repeat mutation of FY has a stronger effect on poly(A) site usage than PPLPP-domain deficiency

To assess the impact of different mutants on genome-wide poly(A) site usage, the fraction of each PAC within one gene was calculated to show the relative abundance of each isoform, which can be represented by "Poly(A) Usage" (PAU). Hierarchical cluster analysis based on PAU was used to distinguish distances among different samples. The results showed that *fy3* and *fy6* clustered together, as did Col-0 (WT), *fy2* and *fy5* were grouped (Figure 2A). This reflected that WD40-repeat mutation and the overexpression of *FY* may have a similar impact on the poly(A) profile. The profile of PPLPP-domain mutants (*fy2* and *fy5*) were much closer to Col-0, indicating that they have less impact on global PAU than *fy3* and *fy6*. Double mutants of *fy2*, *fy3*, and *fy6* with *oxt6* were grouped

in a cluster with the *oxt6* single mutant, and *fy2 oxt6* was distinguished from *fy3 oxt6* and *fy6 oxt6.* Again, this indicates that the PPLPP-domain deficiency may be different from the WD40-repeat mutation in terms of affecting PAU. Mutants of *fy1* and *fca-1* were grouped with a different ecotype, L*er*-0 (Figure 2A). However, *fy1* was further away from L*er*-0 than *fca-1*, suggesting that FY has a greater impact on polyadenylation than FCA in Arabidopsis.

The PAU values were plotted by Cumulative Distribution Function (CDF) at the genomic level of individual mutants (Figure 2B and C). The PAU profiles of all mutants are significantly different from their wild types (K–S test, *P*-value<0.001), reflecting the important role of FY in polyadenylation. In general, CDF curves merge together between 0.5-0.6 on the right y-axis, indicating that more than 50% of the PAU in mutants differed from that of the WT (Figure 2B and C). Notably, the start site of CDF and the median (the point where the curves are folded) curves differed among samples, indicating a different sample unique poly(A) site usage and different PAU distribution profiles among those samples. The *oxt6* mutant was reported to have a significantly different genome-wide poly(A) profile (Thomas et al., 2012). Accordingly, FY also coordinates genome-wide poly(A) site usage (Figure 2B and C).

WD40-repeat mutation and overexpression of FY shift PAU profiles more than PPLPP-domain deficiency (Figure 2C). The PPLPP-domains located in the C-terminus of FY are specifically found in plants and not in human WDR33. These domains interact with FCA and function in flowering time control (Henderson et al., 2005). We found that the poly(A) profile of *fca-1* is significantly different from WT (L*er*-0) (Figure 2B), suggesting that FCA also affects poly(A) site choice. Moreover, the poly(A) profiles of *FY* and *CPSF30* double mutants were also different from both WT and single mutants with various alterations (Figure 2C). The CDF curve of *fy3 oxt6* is closer to Col-0 than their single mutants, whereas curves of *fy2 oxt6* and *fy6 oxt6* were much further away from Col-0 compared to their single mutants (Figure 2C).

196 Collectively, these lines of genetic evidence suggest that FY functions in the
197 determination of poly(A) site usage, as does AtCPSF30. The WD40-repeat mutation of FY
198 has a more significant influence than PPLPP-domain deficiency on poly(A) site choices.
199 In addition, the overexpression of FY has the strongest impact and would be different with
200 other single mutants for the PAU regulation. However, the interaction of plant unique
201 PPLPP-domains on FY and FCA results in a different and more complicated mechanism
202 of polyadenylation than that in mammals.
203

204 **_FY_ mutations affect poly(A) signal usage of near upstream element (NUE)**

205 The cis-elements surrounding poly(A) sites, FUE, NUE, and CE are important for
206 plant polyadenylation (Loke et al., 2005). Previous studies revealed that mutations of one
207 CPSF complex component (AtCPSF30 or AtCPSF100) resulted in an abnormal single
208 nucleotide profile in NUE or FUE (Thomas et al., 2012; Lin et al., 2017). Thus, to elucidate
209 the role of FY in poly(A) signal usage, poly(A) sites were grouped into three sets according
210 to Thomas et al (Thomas et al., 2012): those seen only in the wild type (WT-unique PACs),
211 those seen only in the mutants (mutant-unique PACs), and those seen in both samples
212 (common PACs) as shown in Supplemental Data Set 2.

213 In order to identify canonical poly(A) signals, we focused on NUE regions between
214 10 and 35 bases upstream of poly(A) sites. As shown in Figure 3, an A-rich peak and low
215 U content centered around 20 nucleotides upstream from the poly(A) site is shown in
216 common PACs. A dramatic decrease in A usage is found in *oxt6* unique PACs, as well as a
217 dramatic increase in U usage (Figure 3A). The profile of WT unique PACs is almost the
218 same as common PACs (Figure 3A). These results are consistent with previous findings
219 (Thomas et al., 2012). Generally, profiles of both WT unique and mutant unique PACs of
220 *fy* mutants are different from that of their common PACs (Figure 3B-F). The profiles of FY
221 cryptic (mutant unique) PACs are consistent with that of AtCPSF30 cryptic PACs, and have
222 lower frequency of A usage than their common PACs. However, the profiles of FY

223 authentic (WT unique) PACs have lower A usage than common PACs, which is different
224 from that of AtCPSF30 authentic PACs. These indicate that FY function is not fully
225 redundant with AtCPSF30 in poly(A) signal choice, but it is associated with A-rich NUE
226 poly(A) signal usage. Moreover, in *fy3* and *fy6*, cryptic PACs appear to have a lower A
227 usage than that in *fy1*, *fy2*, and *fy5*, suggesting a different influence between the WD40-
228 repeat mutation and the PPLPP-domain deficiency of FY. Moreover, the overexpression of
229 both WD40-repeat and PPLPP-domain gives a similar phenotype to the WD40-repeat
230 mutation, suggesting that the WD40-repeat may play a dominant role in poly(A) signal
231 recognition. By coincidence, nucleotide composition profiles of cryptic PACs in *fca-1* is
232 similar to that of PPLPP-domain mutants, suggesting that FCA may interact with PPLPP-
233 domain of FY to regulate APA (Figure 3J). Since *oxt6* is a *AtCPSF30* null mutant, whereas
234 *fy* mutants are hypomorphic or overexpression plants, the single nucleotide profiles of
235 double mutants with *oxt6* appear to be similar to that of *oxt6* (Figure 3G-I).
236

237 **Poly(A) site usage pattern coordinated by FY is associated with canonical NUE poly(A)**
238 **signals**

239 As described above, FY widely affects poly(A) site usage depending on its domains
240 (Figure 2). Thus, to reduce the fuzziness of such a large amount data, weighted gene co-
241 expression network analysis (WGCNA) was employed to cluster transcripts with similar
242 poly(A) site usage patterns between samples. Totally, 31,184 PAU from APA gene
243 transcripts were clustered into 19 modules (M) plus a M20, which PAU pattern were not
244 correlated well (Figure 4). Different modules had a different correlation with the samples.
245 A higher correlation value (red) indicates that this module is positively associated with the
246 sample, while a lower correlation value (blue) indicates that this module is negatively
247 associated with the sample.

248 Such module clustering distinguishes each other by NUE poly(A) signal usage of
249 transcripts, which shows the fractions of AAUAAA and 1-nt variant of AAUAAA are

different (Figure 5A). This analysis identified M13 as having a higher AAUAAA usage (~15%) than others. However, M13 is negatively correlated with *oxt6*, indicating that AAUAAA usage in *oxt6* is lower than Col-0. M13 has a higher correlation with PPLPP-domain mutants (*fy1*, *fy2* and *fy5*) than WT. However, M13 has a weaker correlation with *fy3* mutant and *fy6* line than Col-0. These results indicate that the PPLPP-domain deficiency and WD40-repeat mutation of FY oppositely affect polyadenylation. The co-expression network of M13 with 63 transcripts was visualized (Figure 5B). Two hub transcripts from AT2G34420 and AT1G63770 using AACAAA and AAUAAA signals, respectively, were identified. AT2G34420 (LHB1B2) is a Chlorophyll A-B binding protein, which is related to growth and seed dormancy (Li et al., 2007). AT1G63770 is a putative aminopeptidase, which is involved in indole-3-acetic acid (IAA) content, root development and seed germination (Job et al., 2005; Sorin et al., 2006). Previous studies have shown that FY and AtCPSF30 play important roles in plant growth and seed germination (Jiang et al., 2012; Chakrabarti and Hunt, 2015). Thus, genes in this module may be mainly involved in plant growth and development, and modulated by FY and AtCPSF30 through poly(A) signal choices.

Another module, M3, is positive correlated with *fy3*, *fy6*, and *oxt6*, but negatively correlated with *fy2* and *fy5* (both are PPLPP-domain mutants). This module contains the lowest AAUAAA usage, suggesting that non-canonical NUE poly(A) signal usages are overrepresented in the WD40-repeat defective FY mutant (*fy3*), overexpressed FY (*fy6* line), and *AtCPSF30* knockout mutant (*oxt6*). Moreover, two hub transcripts from AT5G46420 and AT5G63530 were identified from M3 (Figure 5C). They used 1-nt variant signal, AACAAA and UAUAAA, respectively. AT5G46420 and AT5G63530 encode 16S rRNA processing protein and farnesylated protein (FP3), respectively. The microarray result revealed that they are significantly reduced in CaLCuV infected leaves, indicating that they play a key role in the defense response (Ascencio-Ibanez et al., 2008). Previous research showed that AtCPSF30 is required for *Pseudomonas syringae* bacterial resistance

(Bruggeman and Delarue, 2014). Therefore, both FY and AtCPSF30 are involved in defense response by modulating poly(A) signal usage of related genes.

Taken together, FY and AtCPSF30 are both associated with AAUAAA poly(A) signal usage. However, FY function does not fully overlap with AtCPSF30. Interestingly, the PPLPP-domain deficiency of FY may act oppositely with its WD40-repeat mutation on canonical poly(A) signal usage.

FY widely affects alternative polyadenylation and APA gene expression

To further explore the role of FY in alternative polyadenylation, the expression of each PAC was analyzed using the DESeq2 package. Different alleles result in a large variation of differentially expressed (DE) PAC APA gene and DE gene numbers (Figure 6A and B). These DE PAC APA genes were significantly enriched in plenty of biological processes, such as Cellular Process, Response to Stimulus, and Developmental Process (Figure 6C). Transcripts from >2000 APA genes were significantly differentially expressed in *fy3* (*padj*<0.05, Figure 6A). Moreover, total expression (all transcripts of one gene) of >3000 genes differed significantly in *fy3* (Figure 6B). These results indicate that the WD40-repeat mutation in FY widely impacts poly(A) site usage and gene expression. However, poly(A) site usage and gene expression were only significantly altered in several hundreds of genes in PPLPP-domain-related mutants, *fy2* and *fy5* (Figure 6A and B). This suggests that the PPLPP-domain deficiency has less impact on both poly(A) site usage and gene expression, which are consistent with the findings shown in Figure 2. Surprisingly, thousands of genes were affected in *fy1* and *fca-1*, which may be affected by the different genetic backgrounds. Two double mutants enhance the DE number (*fy2 oxt6*, *fy3 oxt6*). However, *fy6 oxt6* contains fewer DE PAC and DE genes than single mutants only, indicating that overexpression of FY could partially rescue the expression variation induced by knocking out *AtCPSF30* (Figure 6A and B). Detailed information of DE PAC is provided in Supplemental Data Set 3. Moreover, we found a large proportion of overlap (mostly >50%)

between DE PAC APA genes and DE genes of each mutant (Figure 6D). These results suggest that DE genes in each mutant may be mainly contributed by APA of genes.

WD40-repeat mutation and PPLPP-domain deficiency in FY antagonistically affect NUE poly(A) signal usage in 3'UTRs

Single nucleotide profiles were different among genomic regions, and 3'UTR polyadenylation was the most frequent event (Thomas et al., 2012; Lin et al., 2017). Thus, 3'UTR PACs were extracted to study the NUE poly(A) signal. Generally, A usage of FY authentic PACs (WT-unique) were less abundant than that in *oxt6* (Figure 7A-F). Moreover, in WT-*fy3* and WT-*fy6* comparisons, WT unique PACs have a higher A usage than WT unique PACs identified from WT-*fy1*, WT-*fy2*, and WT-*fy5* pairwise comparisons. This trend was reversed in mutant unique PACs. For example, the profiles of WT unique and *fy1* unique were opposite (Figure 7B), and this trend inverted in the WT-*fy3* and WT-*fy6* comparisons (Figure 7D and F). These results suggest that the WD40-repeat mutation and PPLPP-domain deficiency of FY may differently affect NUE poly(A) signal choice of 3'UTR polyadenylation. Again, the WD40-repeat in FY may act an important role in poly(A) signal usage.

Furthermore, the frequency of the canonical poly(A) signal, AAUAAA, and its 1-nt variants were calculated (Figure 7H and I). By referring to WD40-repeat mutant (*fy3*), WT unique PACs have a higher AAUAAA frequency than *fy3* unique PACs (Figure 7H). Conversely, by referring to PPLPP-domain mutants, WT unique PACs have a lower AAUAAA frequency than PPLPP-domain mutants' unique PACs. Since both WD40-repeat and PPLPP-domain were overexpressed in *fy6*, the frequency of AAUAAA in the WT unique PACs was not as high as that in *fy3*, but still higher than *fy6* unique PACs. These results indicate that WD40-repeat mutation and PPLPP-domain deficiency of FY antagonistically affect AAUAAA usage in 3'UTR polyadenylation. The same trend (but weaker) was found in 1-nt variants of AAUAAA (Figure 7I). Furthermore, FCA was

reported to form a complex with FY and participated in polyadenylation (Simpson et al., 2003). The nucleotide composition and AAUAAA frequency of *fca-1* appeared similar to PPLPP-domain deficiency mutants, rather than the WD40-repeat mutant or *fy6* (Figure 7G and H). This confirmed that FCA engages in polyadenylation through interaction with the PPLPP-domain of FY.

From the above analysis, different domains mutation of FY inversely affect AAUAAA usage in 3'UTR polyadenylation. In addition, it is clear that FY extensively impacts poly(A) site choices. Therefore, we speculated that mutations in FY would disrupt the distribution of 3'UTR PATs and PACs at the genome level. To test this hypothesis, the genomic distributions of DE-PACs (|fold change|≥2) and their PATs were determined. We found that the fraction of PAT reads and PACs in 3'UTR were significantly increased in *fy3* mutant and *fy6* line (Figure 8), indicating that the WD40-repeat mutation in FY increases poly(A) site usage in 3'UTR. However, compared with that in *fy3* and *fy6*, the distribution of PAT reads and PACs in 3'UTR was opposite to that of *fy1* and *fy5*, i.e., the PPLPP-domain deficiency in FY decreases the poly(A) site usage in the 3'UTR. Interestingly, *oxt6* and its double mutants decrease the distribution of PAT reads and PACs in the 3'UTR (Figure 8). These results show that the WD40-repeat mutation and PPLPP-domain deficiency in FY influences the expression of full-length transcripts differently.

Importantly, we found that more genes used longer 3'UTR in *fy1*, *fy2* and *fy5* mutants (Figure 9A), indicating that the PPLPP-domain deficiency in FY results in a preference for using distal poly(A) sites rather than proximal sites in 3'UTR. By contrast, more genes containing a shorter 3'UTR were observed in the *fy3*, *fy6* and *oxt6* (Figure 9A). This tendency also occurred in the double mutants. The result implies that the WD40-repeat mutation in FY and the AtCPF30 knockouts show a preference for using proximal poly(A) sites in 3'UTRs. Again, the WD40-repeat mutation and PPLPP-domain deficiency in FY functional antagonistically affect poly(A) site usage in 3'UTR. In 3'UTR significantly lengthened genes, the average 3'UTR length was increased by 25 nt in *fy3* (Figure 9B); in

3’UTR significantly shortened genes, the length was decreased by 22 nt in *fy3* (Figure 9C). The variation of average 3’UTR length was above 20 nt in other mutants.

WD40-repeat mutation of FY affects the APA of genes and contributes to phenotypic outputs

Phenotyping showed that primary root length varied between WT and mutants (including double mutants) (Figure 10A and B). Among *fy* single mutants, primary root length had the most significant difference between *fy3* and Col-0. A similar trend was observed between *oxt6*, *fy oxt6* double mutants, and WT (Figure 10A and B). Remarkably, the reduction of primary root length in *fy3 oxt6* was emphasized, potentially contributed by the combination effect of *fy3* and *oxt6.*

SAHH, a gene encoding a S-adenosylhomocysteine hydrolase, was reported to control primary root length (Wu et al., 2009). We found that the poly(A) profiles among 3’UTR of *SAHH* were different among WT, *fy3*, *oxt6*, and *fy3 oxt6*, indicating that APA happened in mutants (Figure 10C). Normally, the distal poly(A) site (PA2) is preferred rather than the proximal site (PA1). By contrast, the preference usage was switched to PA1 in *fy3 oxt6*, as well as in *oxt6*. Gene expression (total expression) of *SAHH* was quantified by RT-qPCR, and showed that it was all significantly decreased in *fy3*, *oxt6*, and *fy3 oxt6* by comparing to WT (Figure 10D, left). The PA2 abundance was also validated and appeared consistent with PAT-seq (Figure 10D, right). In *fy3*, the distal site (PA2) was unchanged, whereas, gene expression of *SAHH* significantly decreased, suggesting that PA1 usage decreased. Thus, in the case of *SAHH,* FY enhances PA1 usage, whereas AtCPSF30 promotes PA2 usage. To evaluate the impact of RNA turnover on gene expression, RNA stability assay was conducted. The results showed that the *SAHH* mRNA is stable in each mutant, even though it is a little bit less stable in *oxt6* (Figure 10E). Thus, the variation of *SAHH* expression in *fy3*, *oxt6*, and *fy3 oxt6* may be mainly contributed by APA.

Furthermore, we checked through another APA gene, *ATHB13*, which encodes a

385 homeodomain leucine zipper class I (HD-Zip I) protein that regulates primary root
386 development (Silva et al., 2016). The poly(A) profiles of *ATHB13* were shifted to the
387 proximal site in *fy3*, *oxt6*, and *fy3 oxt6* (Figure 10F). PA2 of *ATHB13* was mildly inhibited
388 in *fy3* and strongly inhibited in *oxt6* and *fy3 oxt6* (Figure 10F and G). This indicates that
389 PA2 of *ATHB13* may be affected by the combination of intact FY and AtCPSF30. However,
390 total gene expression of *ATHB13* was not changed in *fy3*, *fy3 oxt6*, and *oxt6*, which reflects
391 the increase of PA1 that was compensated by the decrease of PA2. The RNA stability assay
392 showed that *ATHB13.PA2* was less stable in *oxt6*, reflecting that the APA of *ATHB13* results
393 in different isoform stability (Figure 10H).

394 Both *fy3* and *oxt6* exhibit significant late flowering, and *fy3 oxt6* double mutants
395 flower much later than the others (Figure 11A). Thus, FY and AtCPSF30 synergistically
396 affect flowering time. Indeed the full-length transcript and total expression of
397 *FLOWERING LOCUS C* (*FLC*), which encodes an inhibitor of flowering, was
398 overrepresented in *fy3* and *oxt6*, and dramatically overrepresented in *fy3 oxt6* (Figure 11 B
399 and D). Moreover, proximal poly(A) site usage within intron 3 and total expression levels
400 of *FCA* decreased in *fy3*, *oxt6*, and *fy3 oxt6* (Figure 11 C and E). These are consistent with
401 previous findings (Simpson et al., 2003), but now include the role of FY/CPSF30 in the
402 APA of *FCA* and *FLC*.

403 Mutation of *FY* and *AtCPSF30* altered the APA of a transcriptional regulator
404 (AT3G47610) and *AKR2* (AT4G35450) (Supplemental Figure 4 and 5). Moreover, we
405 found that the two mutants of these two genes carry T-DNA insertions between their APA
406 sites, which may result in the loss of their full-length transcripts. Phenotypic studies
407 showed that these two mutants have higher seed germination rates and green cotyledon
408 rates than WT under salt stress (Figure 12), suggesting that these two mutants were less
409 sensitive to salt stress. Moreover, the primary root length of the two mutants were longer
410 than WT under oxidative stress induced by treatment with methyl viologen (MV),
411 especially for the SALK_205297 mutant (Figure 12). Importantly, 3'RACE confirmed that

412 the two mutants were long transcript deletions (Figure 12F), revealing that the above
413 phenotypes indeed are related to the function of the long transcript. Collectively, we
414 demonstrated *in vivo* that APA mediated by FY/CPSF30 can function in plant stress
415 responses.
416

417 # DISCUSSION

418 **Role of FY-mediated alternative polyadenylation**

419 FY is the Arabidopsis homolog of the polyadenylation factor Pfs2p in yeast and
420 WDR33 in mammals (Simpson et al., 2003; Chan et al., 2014). However, the role of FY in
421 poly(A) site choice at the genome level remained unclear. Our results herein demonstrate
422 that FY is definitely involved in poly(A) site usage. Furthermore, the WD40-repeat
423 mutation of FY has more influence on genome-wide poly(A) site usage than the PPLPP-
424 domain deficiency (Figure 2). Interestingly, the WD40-repeat mutation in FY acts in an
425 opposite manner as the PPLPP-domain deficiency in 3'UTR APA, especially in the
426 preference of single nucleotide usage and 3'UTR length. Since the PPLPP-domain of FY
427 is not found in WDR33 or Pfs2p (Henderson et al., 2005), this antagonistic effect of WD40-
428 repeat and PPLPP-domain is plant unique. Moreover, FY was differentially expressed
429 among tissues of Arabidopsis (Henderson et al., 2005), implying that FY may be involved
430 in the differentiation of APA among tissues. Previous studies demonstrated that shortening
431 3'UTR in mammalian cells resulted in the exception of miRNA targeting, leading to an
432 increase of protein production (Sandberg et al., 2008; Bartel, 2009; Mayr and Bartel, 2009).
433 However, we found that there is no obvious correlation between the 3'UTR length
434 switching and the gene expression in *fy*, *oxt6*, and their double mutants (data not shown).
435 Moreover, our previous work also did not find an obvious negative correlation between
436 3'UTR length variation and gene expressions in rice different tissue (Zhou et al., 2019). It
437 was reported that plant miRNAs predominately targeted to CDS (Carthew and Sontheimer,
438 2009). However, plant 3'UTR were targeted by phasiRNA, which is produced by miRNA

targeted genes (Ma et al., 2018). Therefore, the relationship between 3'UTR length and microRNA regulation in plants is complicated and remains to be further investigated.

Role of FY in the recognition of plant poly(A) signals

In this study, by using the PAT-seq approach, we provided genetic evidence that the mutation of WD40-repeat or PPLPP-domain deficiency in FY disrupt AAUAAA signal usage (Figures 3, 5 and 7). Importantly, WD40-repeat mutation and PPLPP-domain deficiency might also function antagonistically in polyadenylation (Figures 4-9). In addition, overexpression of both WD40-repeat and PPLPP-domain (*fy6* line) possess similar change with the WD40-repeat mutation (*fy3* mutant). These results indicate that the WD40-repeat in FY may play an important role in PAS recognition in the NUE region. By modeling in SWISS-MODEL (http://swissmodel.expasy.org), a protein 3D structure of FY was found to be similar to WDR33 with 54.61% protein sequence identity and significant QMEAN Z-scores (-3.82). It was clear that the WD40-repeat of WDR33 can directly bind to AAUAAA signal (Schönemann et al., 2014; Clerici et al., 2017; Sun et al., 2018). Thus, WD40-repeat mutation of FY may directly affect AAUAAA signal recognition. However, further genetic and biochemical experiments should be carried out to reveal the mechanism by which FY functions in plant polyadenylation.

Previous studies have shown that the PPLPP-domain (plant unique) of FY binds to the Trp-Trp (WW) domain of FCA *in vitro* (Henderson et al., 2005). The FCA/FY interaction is well characterized *in vitro* and can be reproduced by using FCA/FY counterparts from other plant species (Lu et al., 2006). However, FY forms a stable complex with AtCPSF100 and AtCPSF160 *in vivo* but not with FCA (Manzano et al., 2009). Thus, the FCA/FY interaction *in vivo* may be regulated or transient. Furthermore, FY/AtCPSF160 containing fractions and those containing FCA did not appear to overlap, suggesting that FY/AtCPSF and FCA/FY are two separate complexes. Importantly, it was found that FCA/FY interaction leads to an altered interactions in the FY/AtCPSF

complexes (Manzano et al., 2009). Therefore, we speculate that FCA may compete with other CPSF factors to recruit FY in or out of the CPSF complex to affect poly(A) signal recognition and polyadenylation.

Sine *oxt6* is an AtCPSF30 null mutant and *fy* mutants are hypomorphic, *fy2 oxt6*, *fy3 oxt6*, and *oxt6* have similar patterns for recognizing poly(A) signals (Figures 3 and 5). We also found fewer unique PACs in *fy6 oxt6* mutant compared with other double mutants (Figure 3). This result makes one speculate that FY overexpression may partially complement AtCPSF30 function in the recognition of poly(A) signals. Nevertheless, individual contributions of FY and AtCPSF30 to CPSF RNA-binding specificity in plants remains to be determined by additional biochemical experiments. We also found that 2%-4% AAUAAA signal still is used in pre-mRNA transcripts of *fy oxt6* double mutants, suggesting that other polyadenylation factors may participate in the recognition of the AAUAAA signal in the absence of AtCPSF30 and FY. Previous studies showed that human Fip1 and CstF64 appeared to be able to crosslink with the AAUAAA signal (Martin et al., 2012), and Fip1 is in close association with the CPSF complex (Schönemann et al., 2014; Clerici et al., 2017; Sun et al., 2018). CstF64 and Fip1 are homologs of Arabidopsis CstF64 (AtCstF64), AtFIPS3, and AtFIPS5, respectively, and the three proteins can also interact with each other. In addition, AtFIPS can directly interact with AtCPSF30 (Hunt et al., 2008). The C-terminal of AtFIPS5 contains an RNA binding domain. AtFIPS5 may be the FUE recognition factor for polyadenylation in plants, suggesting that one or more of its interacting protein partners may be involved in the recognition of the NUE, FUE and/or cleavage site (Forbes et al., 2006). AtCPSF100 resides at the center of the CPSF protein–protein interaction network (Hunt et al., 2008). However, recent studies have revealed that AtCPSF100 does not participate in NUE poly(A) signal selection but does affect the poly(A) signal recognition of the FUE (Lin et al., 2017). Hence, determining the full machinery of poly(A) signal recognition in plants requires additional research.

493 **The biological functions of FY**

494 APA regulation of gene expression participates in a subset of biological processes,
495 including development, disease resistance and abiotic stress tolerance in plants (Xu et al.,
496 2006; Zhang et al., 2008; Xing and Li, 2011; Bruggeman and Delarue, 2014; Ma et al.,
497 2014). The results presented in this article demonstrate that FY comprehensively affects
498 APA and gene expression, and that these DE-PAC APA genes are involved in many
499 biological processes, including 'cellular process', 'developmental process', and
500 'reproductive process' as determined by the GO analysis. Indeed, our findings are
501 consistent with previously known biological processes in which FY is involved, like
502 flowering time regulation. They are also consistent with previous reports where PPLPP-
503 domain deficient mutants (*fy1*, *fy2*, and *fy5*) led to alterations in the poly(A) site usage of
504 *FCA* and increased expression of *FLC*, especially in the *fy2* mutant (Supplemental Figure
505 6 and 7) (Henderson et al., 2005; Feng et al., 2011). In addition, the glycine (G141) residue
506 substitution occurred in the first WD40-repeat (*fy3*) also demonstrates that intact WD40-
507 repeats are required for mediating *FLC* expression (Supplemental Figure 7), as previously
508 reported (Henderson et al., 2005). The amino acid substitution in *fy3* is predicted to affect
509 a structural residue of the B-β-strand in the first propeller blade (Smith et al., 1999).
510 Therefore, the G141S substitution may have a specific effect on FY-WD40-repeat
511 interactions, resulting in an increase of *FLC* expression.

512 We also found that the WD40-repeat mutation (*fy3*) affects primary root growth. This
513 phenomenon could be related to the confirmed APA events of a couple of genes relate with
514 root growth, *SAHH* and *ATHB13*. The RT-qPCR results showed that the mRNA level of
515 *SAHH* was decreased in *fy3* (Figure 10). Previous reports have shown that *sahh* null
516 mutants showed decreased primary root length (Wu et al., 2009). Therefore,
517 downregulation of the *SAHH* gene may lead to shortened primary roots in *fy3*. Knockout
518 mutants, *athb13*, showed increased primary root length, suggesting that this transcription
519 factor is a negative regulator of early root growth (Silva et al., 2016). Interestingly, the

expression of distal poly(A) transcript from *ATHB13* was decreased in *fy3*, but gene expression of *ATHB13* did not change, reflecting a switched usage of the poly(A) site. Moreover, *SAHH* and *ATHB13* transcripts in WD40-repeat mutant are stable, further suggesting that FY regulates gene expression by mediating poly(A) site usage rather than directly modulating RNA stability. However, AtCPSF30 alters the stabilities of *ATHB13* mRNA isoforms. Previous research showed that AtCPSF30 could localize in the cytoplasm by itself or co-localized with CPSF100 and is present in P-bodies (Rao et al., 2009), which are foci for mRNA surveillance and mRNA decay (Eulalio et al., 2007). Therefore, our results provided further evidence that AtCPSF30 plays a role in mRNA degradation.

METHODS

Plant materials, growth conditions and phenotype assays

The *Arabidopsis thaliana fy1*, *fy2*, and *fy3* mutants were provided by Dr. Caroline Dean (John Innes Centre, UK). SALK_005697 (designated as *fy5*), SALK_048649 (designated as *fy6* line), SALK_146237 (T-DNA insertion mutant of AT3G47610) and SALK_205297 (T-DNA insertion mutant of AT4G35450) were ordered from the Arabidopsis Biological Resource Center (ABRC, http://www.arabidopsis.org). The *fca-1* mutant (ABRC stock: NS52) carries a point mutation at exon 13 introducing a premature termination codon, which contains both RRMs but lacks the WW domain (Macknight et al., 1997). A description of *fy* mutants and the PPLPP-domain and WD40-repeats are shown in Figure 1. The double mutants *fy2 oxt6, fy3 oxt6* and *fy6 oxt6* were generated by crossing *fy2*, *fy3* or *fy6* with *oxt6*, respectively. The *fy1* and *fca-1* are in the ecotype Landsberg *erecta* (L*er*-0) genetic background. Other mutants are in the Col-0 background. L*er*-0 and Col-0 were referred to as wild type (WT) in this study. Arabidopsis plants were grown under long-day conditions (16 h of illumination at 120 μmol m^{-2} s^{-1} of white light and 8 h dark cycle) at a constant temperature of 22°C. Seeds for the following phenotypic analyses were collected at the same time.

For root length analyses, seeds were surface sterilized for 3 min and then washed five times with sterilized distilled water, and then placed in the dark for 3 days at 4°C for synchronization, after which they were grown on 0.8% agar plates containing ½-strength Murashige and Skoog (MS) medium or ½ MS medium supplemented with 50 nM methyl viologen (MV) and 1% sucrose for 11 days. At least six seeds of mutants and their WT were sown on the same plates side-by-side. Three biological replicates were performed and each replicate contained 3 plates. The root length was measured by ImageJ software. One-way ANOVA was applied to analyze statistically significant differences between the wild type and mutants. A P-value<0.05 threshold was considered as statistical significance.

For flowering time tests, seeds were synchronized and then planted in soil. Each 6 × 6 cm pot contained one plant. Each experiment comprised of 18 pools and three independent experiments were completed. Plants were grown in a controlled environment under long-day photoperiods in a growth chamber. Flowering time was measured by counting the number of rosette leaves at flowering as previously described (Macknight et al., 2002). For seed germination assays, the sterilized seeds were placed in the dark for 3 days at 4°C for synchronization, after which they were grown on ½ MS medium (0.8% agar, 1% sucrose) or ½ MS medium supplemented with 125 mM NaCl. Three biological replicates were performed and each replicate contained 40 seeds for each line on the same plate. Germination (emergence of radicles) and post-germination growth (green cotyledon appearance) were scored at the indicated time points.

For PAT-seq, seeds were synchronized and planted in soil for 14 days. At least 10 seedling shoots were collected for one replicate. Three biological replicates from different shoots and independent pools were accomplished for PAT-seq.

PAT-seq library preparation and sequencing

For PAT-seq libraries construction, samples of mutants and wild type were prepared from three independent biological replicates. Total RNAs were isolated using the TRIzol reagent (Invitrogen), and their DNA was removed by using DNase I (Takara) following a

column-based RNA purification. PAT-seq libraries were prepared from two μg of total RNA as described (de Lorenzo et al., 2017) with modifications. Briefly, RNA was fragmented in 5 × first strand buffer (Invitrogen) at 94°C for 4 min. RNA fragments with poly(A) tails were enrichened via oligo$(dT)_{25}$ magnetic beads (New England Biolabs). Reverse-transcription was performed using barcoded oligo$(dT)_{18}$ primers with SMARTSCRIBE enzyme (Clontech) for 2 h and then 5' adaptor for template switching added. The last nucleotide of the 5' adaptor was modified by locked nucleic acid modification (LNA). The generated cDNA was purified by AMPURE XP beads (Beckman), following by 18 PCR cycles with Phire II (Thermo Fisher Scientific) to produce PAT-seq libraries. The library was run on a 2% agarose gel, and 300-500 bp library fragments were purified. Libraries were qualified and quantified by Agilent Bioanalyzer 2100, Qubit 2.0 and qPCR. Finally, libraries were sequenced on the Illumina HiSeq 2500 platform at the facility located in the College of the Environment and Ecology, Xiamen University.

Poly(A) Tag (PAT) and poly(A) site cluster (PAC) generation

The sequencing data were processed using previously described methods (Wu et al., 2011; Fu et al., 2016). Briefly, low-quality raw data were filtered out using FASTX-Toolkit (Version 0.0.14, parameters "-q 10 -p 50 -v -Q 33"), and barcodes and poly(T) stretches of raw reads were trimmed. The remaining reads were mapped to the Arabidopsis reference genome (TAIR10, www.arabidopsis.org) by Bowtie2 software (Version 2.1.0, parameters "-L 25 -N 1 -i S,1,1.15 --no-unal"). Potential internal priming reads were filtered out (Loke et al., 2005). As poly(A) site microheterogeneity is pervasive in plants, the mapped poly(A) tags (PATs) within 24 nucleotides (nt) were grouped into one poly(A) cluster (PAC) which represents a cleavage site (known as a poly(A) site) (Wu et al., 2011). To facilitate the assignments of PACs to annotated genome, genes with annotated 3'UTRs were extended for 120 nt, and genes without annotated 3' UTRs were extended by 338 nt (Wu et al., 2011). To avoid uncertainty from low read counts, total reads of a PAC among all samples with less than 20 were discarded.

Poly(A) usage analysis

Filtered PACs were used for calculating "Poly(A) Usage" (PAU). PAU represents the ratio of reads in one PAC relative to total reads of the gene (Ha et al., 2018). Average PAU among three biological replicates were used for calculating cumulative distribution function (CDF) and plotted by "mountainplot" package in R (Monti, 1995). By "mountainplot", CDF was folded at 50% frequency to show the median of genome-wide PAU profile. Kolmogorov-smirnov test (K–S test) was applied to judge the significant difference between two CDF (Haslinger et al., 2010). A P-value<0.05 threshold was considered as statistical significance.

Poly(A) signal analyses

The sequences 300 nt upstream and 100 nt downstream of unique and common poly(A) sites were extracted for single nucleotide profile analysis, as previously reported (Loke et al., 2005). In order to identify poly(A) signals, we focused on NUE regions between 10 and 35 bases upstream of poly(A) sites. The canonical AAUAAA signal and its 1-nt variants were analyzed across all unique and common poly(A) sites as described (Loke et al., 2005). Sample-unique PACs: PACs only expressed in a mutant (the sum of all PATs in three biological repeats was greater than three), but not in the WT (PAT of each biological repeat was equal to zero); or PACs only expressed in the WT, but not in the mutant. Sample-common PACs: these PACs expressed simultaneously in both the WT and mutant. 1-nt variants of AAUAAA contained eighteen hexamers (UAUAAA, CAUAAA, GAUAAA, AUUAAA, ACUAAA, AGUAAA, AAAAAA, AACAAA, AAGAAA, AAUUAA, AAUCAA, AAUGAA, AAUAUA, AAUACA, AAUAGA, AAUAAU, AAUAAC, AAUAAG).

Transcript co-expression analysis

The weighted gene co-expression network analysis (WGCNA) R package (Langfelder and Horvath, 2008; Zhan et al., 2015) was used to assess PAU profiles of APA genes across different mutants and WT. The average PAU values of 31,184 transcripts from three

biological replicates were used for WGCNA. To calculate the adjacency matrix, we first calculated the Pearson correlation coefficients between every two transcripts across different mutants and WT. A soft threshold value of 7 was used to transform the adjacency matrix, which was then transformed into a topological overlap (TO) matrix by the TOM similarity algorithm. Transcripts were hierarchically clustered based on TO similarity. The Dynamic Tree Cut algorithm was used to detect clusters; the mergeCutHeight was 0.4. The modules were defined as branches from the tree cutting, and the minModuleSize was 30. These transcripts were clustered into 19 modules. The networks of M3 and M13 were filtered at adjacency thresholds of 0.1 and 0.3, respectively, and visualized in Cytoscape 3.6.0 software (Shannon et al., 2003). Intramodular connectivity was also calculated. Transcripts with high intramodular connectivity were considered as intramodular hub transcripts. The hub transcripts were obtained with more than 40 connectivity degree and were shown in yellow in network maps.

Differentially expressed PACs (DE-PAC) and DE gene analysis

DESeq2 package (version 1.14.1) was used to normalize read counts and process differential expression PACs (Anders and Huber, 2010). DE-PACs were calculated to uncover the poly(A) profile shift and to estimate the variance of expression levels for a set of genomic regions (5'UTR, 3'UTR, introns, CDS, and intergenic regions) based on read number within each feature. All poly(A) tags of the genes were summed for representing gene expression levels. Similarly, differentially expressed genes were calculated by DESeq2 package. An adjusted *P*-value (*padj*) was corrected using Benjamini–Hochberg method. A *padj*<0.05 threshold was considered statistically significant. DE-PAC APA-associated gene ontology enrichment was performed using agriGO with TAIR10 annotation as the background (Du et al., 2010). FDR-corrected *P*-values < 0.05 were selected as statistically significant.

Identification of 3' UTR length

For 3'UTR APA analysis, the average weighted length of each 3'UTR of a gene was

calculated as described (Fu et al., 2016). The 3'UTR length of each PAC is the distance from each PAC location to the stop codon. For each gene, 3'UTR average weighted length was defined as the sum of 3'UTR length of each PAC multiplied by its expression level (average of three biological repeats normalized PATs) and then divided by the total expression level. A cut-off *P*-value of 0.05 was adopted for both significantly longer and shorter 3'UTR. The box plot was used to show the length distribution.

RT-qPCR analysis of poly(A) sites

About two μg of high quality total RNA free of DNA contamination were reverse transcribed with oligo$(dT)_{18}$ primer by SMARTScribe Reverse Transcriptase (Clontech). RT-qPCR assays were performed using the CFX96TM Real-Time PCR Detection System (Bio-Rad, Inc., Hercules, CA, USA) with SYBR Premix Ex TaqII fluorescent dye (Roche). The relative expression values were determined by using *UBQ10* as a housekeeping gene (Wang and Auwerx, 2017). Three biological replicates were performed for all experiments. Moreover, each replicate comprised of three technical repetitions. One-way ANOVA was used to analyze statistically significant differences between the wild type and mutants. A *P*-value<0.05 threshold was considered as statistical significance. All primers used herein are listed in Supplemental Table 1.

mRNA stability assay

RNA stability assay was performed by using cordycepin to inhibit transcription (de Lorenzo et al., 2017). Briefly, 2 week-old seedlings were harvested, the soil attached on root surface was gently washed away and whole plants were then transferred to a flask containing incubation buffer (15 mM sucrose, 1 mM KCl, 1 mM PIPES, and 1 mM sodium citrate, pH6.5). Cordycepin (Sigma) was dissolved in 50% EtOH. After 30 min of incubation (time 0), cordycepin solution was added to a final concentration of 200 mM. Seedlings were collected after 120 min and frozen in liquid nitrogen. Triplicate biological replicates were conducted with a pooling of ~10 plants for each replicate. RNA extraction and RT-qPCR analysis were performed as described above. *EIF4A* was used as a reference

gene (Fedak et al., 2016).

3' rapid amplification of cDNA end (3'RACE)

3'RACE was performed using SMART RACE cDNA Amplification Protocol (Clontech) according to the manufacturer's instructions. One μg DNA-free total RNA was used to reverse to cDNA with oligo$(dT)_{30}$ 3'RACE CDS primer A. The first PCR was amplified using Universal Primer A Mix (UPM) (UPM-long and UPM-short mix) and gene special primer (GSP1) with Phusion High-Fidelity DNA Polymerase (Life Technology). The second PCR was amplified using Nested Universal Primer A (NUP) and GSP2. Multiple PCR products were purified and sequenced. All primers used herein are listed in Supplemental Table 1. Sequencing results were mapped to target gene by DNAMAN, and single nucleotide peaks were visualized by SeqMan.

Statistical Analysis

P-values were calculated with one-way ANOVA. See Supplemental Data Set 4 for detailed statistical results.

Accession numbers

All PAT-seq raw data for this study are available at NCBI website under accession number SRP145554.

Supplemental Data

Supplemental Figure 1. The general experimental process.

Supplemental Figure 2. Principal Component Analysis (PCA) of repeatability of three biological replications.

Supplemental Figure 3. APA gene number and gene type analysis.

Supplemental Figure 4. Sequencing coverage of AT3G47610 gene among WT, *fy* mutants and double mutants.

Supplemental Figure 5. Sequencing coverage of *AKR2* gene among WT, *fy* mutants and double mutants.

Supplemental Figure 6. Sequencing coverage of *FCA* among WT, *fy* mutants and double mutants.

Supplemental Figure 7. Sequencing coverage of *FLC* among WT, *fy* mutants and double mutants.

Supplemental Table 1. RT-qPCR and 3'RACE primers used in this study.

Supplemental Data Set 1. List of poly(A) site clusters (PACs).

Supplemental Data Set 2. List of unique PACs and common PACs.

Supplemental Data Set 3. List of DE-PAC between mutant and wild type.

Supplemental Data Set 4. Statistical test results for one-way ANOVA.

ACKNOWLEDGMENTS

We thank Caroline Dean for providing *fy1*, *fy2*, and *fy3* mutant seeds. The authors are grateful to Haidong Qu, Xiuxiu Wang, Xiaoxuan Zhou, and Wenjia Lu for technical assistance, to American Journal Experts and Taylor Li for language editing. This work was supported in part by a grant from the National Key R&D Project of China (2016YFE0108800), a grant from the U.S. NSF (IOS-154173), both to QQL, and grants (2017M620274, 2018T110649) from China Postdoctoral Science Foundation to JL.

AUTHOR CONTRIBUTIONS

QQL and ZY designed the research. ZY performed experiments. ZY, QQL and JL contributed to data analysis, wrote and revised the manuscript.

Conflict of interest

All authors declare that there is no conflict of interest.

REFERENCES

735 **Anders, S., and Huber, W.** (2010). Differential expression analysis for sequence count data. Genome
736 Biol **11,** R106.
737 **Ascencio-Ibanez, J.T., Sozzani, R., Lee, T.J., Chu, T.M., Wolfinger, R.D., Cella, R., and Hanley-**
738 **Bowdoin, L.** (2008). Global analysis of *Arabidopsis* gene expression uncovers a complex array
739 of changes impacting pathogen response and cell cycle during geminivirus infection. Plant
740 Physiol **148,** 436-454.
741 **Bartel, D.P.** (2009). MicroRNAs: target recognition and regulatory functions. Cell **136,** 215-233.
742 **Bruggeman, Q., and Delarue, M.** (2014). The Polyadenylation Factor Subunit CLEAVAGE AND
743 POLYADENYLATION SPECIFICITY FACTOR30: A Key Factor of Programmed Cell Death
744 and a Regulator of Immunity in *Arabidopsis*. Plant Physiol **165,** 732-746.
745 **Carpenter, S., Ricci, E.P., Mercier, B.C., Moore, M.J., and Fitzgerald, K.A.** (2014). Post-
746 transcriptional regulation of gene expression in innate immunity. Nat Rev Immunol **14,** 361-
747 376.
748 **Carthew, R.W., and Sontheimer, E.J.** (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell
749 **136,** 642-655.
750 **Chakrabarti, M., and Hunt, A.G.** (2015). CPSF30 at the Interface of Alternative Polyadenylation and
751 Cellular Signaling in Plants. Biomolecules **5,** 1151-1168.
752 **Chan, S.L., Huppertz, I., Yao, C., Weng, L., Moresco, J.J., Rd, Y.J., Ule, J., Manley, J.L., and Shi,**
753 **Y.** (2014). CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing.
754 Genes Dev **28,** 2370-2380.
755 **Clerici, M., Faini, M., Aebersold, R., and Jinek, M.** (2017). Structural insights into the assembly and
756 polyA signal recognition mechanism of the human CPSF complex. Elife **6,** e33111.
757 **Clerici, M., Faini, M., Muckenfuss, L.M., Aebersold, R., and Jinek, M.** (2018). Structural basis of
758 AAUAAA polyadenylation signal recognition by the human CPSF complex. Nat Struct Mol
759 Biol **25,** 135-138.
760 **Colgan, D.F., and Manley, J.L.** (1997). Mechanism and regulation of mRNA polyadenylation. Genes
761 Dev **11,** 2755-2766.
762 **Cyrek, M., Fedak, H., Ciesielski, A., Guo, Y., A, Ś., Brzeźniak, L., Krzyczmonik, K., Pietras, Z.,**
763 **Liu, F., and Kaczanowski, S.** (2016). Seed dormancy in *Arabidopsis thaliana* is controlled by
764 alternative polyadenylation of DOG1. Plant Physiol **170,** 947–955.
765 **de Lorenzo, L., Sorenson, R., Bailey-Serres, J., and Hunt, A.G.** (2017). Noncanonical Alternative
766 Polyadenylation Contributes to Gene Regulation in Response to Hypoxia. Plant Cell **29,** 1262-
767 1277.
768 **Derti, A., Garrettengele, P., Macisaac, K.D., Stevens, R.C., Sriram, S., Chen, R., Rohl, C.A.,**
769 **Johnson, J.M., and Babak, T.** (2012). A quantitative atlas of polyadenylation in five mammals.
770 Genome Res **22,** 1173-1183.
771 **Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z.** (2010). agriGO: a GO analysis toolkit for the
772 agricultural community. Nucleic Acids Res **38,** W64-W70.
773 **Elkon R, U.A., Agami R.** (2013). Alternative cleavage and polyadenylation: extent, regulation and
774 function. Nat Rev Genet **14,** 496-506.
775 **Eulalio, A., Behm-Ansmant, I., and Izaurralde, E.** (2007). P bodies: at the crossroads of post-

transcriptional pathways. Nat Rev Mol Cell Biol **8,** 9-22.

Fedak, H., Palusinska, M., Krzyczmonik, K., Brzezniak, L., Yatusevich, R., Pietras, Z., Kaczanowski, S., and Swiezewski, S. (2016). Control of seed dormancy in *Arabidopsis* by a cis-acting noncoding antisense transcript. Proc Natl Acad Sci U S A **113,** E7846-E7855.

Feng, W., Jacob, Y., Veley, K.M., Ding, L., Yu, X., Choe, G., and Michaels, S.D. (2011). Hypomorphic Alleles Reveal FCA-Independent Roles for FY in the Regulation of FLOWERING LOCUS C. Plant Physiol **155,** 1425-1434.

Forbes, K.P., Addepalli, B., and Hunt, A.G. (2006). An *Arabidopsis* Fip1 homolog interacts with RNA and provides conceptual links with a number of other polyadenylation factor subunits. J Biol Chem **281,** 176-186.

Fu, H., Yang, D., Su, W., Ma, L., Shen, Y., Ji, G., Ye, X., Wu, X., and Li, Q.Q. (2016). Genome-wide dynamics of alternative polyadenylation in rice. Genome Res **26,** 1753–1760.

Fu, Y., Sun, Y., Li, Y., Li, J., Rao, X., Chen, C., and Xu, A. (2011). Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. Genome Res **21,** 741-747.

Ha, K.C.H., Blencowe, B.J., and Morris, Q. (2018). QAPA: a new method for the systematic analysis of alternative polyadenylation from RNA-seq data. Genome Biol **19,** 45.

Haslinger, R., Pipa, G., and Brown, E. (2010). Discrete time rescaling theorem: determining goodness of fit for discrete time statistical models of neural spiking. Neural Comput **22,** 2477-2506.

Henderson, I.R., Liu, F., Drea, S., Simpson, G.G., and Dean, C. (2005). An allelic series reveals essential roles for FY in plant development in addition to flowering-time control. Development **132,** 3597-3607.

Hong, L., Ye, C., Lin, J., Fu, H., Wu, X., and Li, Q.Q. (2018). Alternative polyadenylation is involved in auxin-based plant growth and development. Plant J **93,** 246-258.

Hunt, A.G., Xing, D., and Li, Q.Q. (2012). Plant polyadenylation factors: conservation and variety in the polyadenylation complex in plants. BMC Genomics **13,** 641.

Hunt, A.G., Xu, R., Addepalli, B., Rao, S., Forbes, K.P., Meeks, L.R., Xing, D., Mo, M., Zhao, H., and Bandyopadhyay, A. (2008). *Arabidopsis* mRNA polyadenylation machinery: comprehensive analysis of protein-protein interactions and gene expression profiling. BMC Genomics **9,** 220.

Jiang, S., Kumar, S., Eu, Y.J., Jami, S.K., Stasolla, C., and Hill, R.D. (2012). The *Arabidopsis* mutant, *fy-1*, has an ABA-insensitive germination phenotype. J Exp Bot **63,** 2693-2703.

Job, C., Rajjou, L., Lovigny, Y., Belghazi, M., and Job, D. (2005). Patterns of protein oxidation in *Arabidopsis* seeds and during germination. Plant Physiol **138,** 790-802.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics **9,** 559.

Li, F., Asami, T., Wu, X., Tsang, E.W., and Cutler, A.J. (2007). A putative hydroxysteroid dehydrogenase involved in regulating plant growth and development. Plant Physiol **145,** 87-97.

Lin, J., Xu, R., Wu, X., Shen, Y., and Li, Q.Q. (2017). Role of cleavage and polyadenylation specificity factor 100: anchoring poly(A) sites and modulating transcription termination. Plant J **91,** 829-839.

Lin, Y., Li, Z., Ozsolak, F., Kim, S.W., Arangoargoty, G., Liu, T.T., Tenenbaum, S.A., Bailey, T., Monaghan, A.P., and Milos, P.M. (2012). An in-depth map of polyadenylation sites in cancer. Nucleic Acids Res **40,** 8460-8471.

Liu, M., Xu, R., Carrie, M., Hong, L., Carol, V.L., Hunt, A.G., and Li, Q.Q. (2014). Integration of Developmental and Environmental Signals via a Polyadenylation Factor in *Arabidopsis*. PLos One **9,** e115779.

Loke, J.C., Stahlberg, E.A., Strenski, D.G., Haas, B.J., Wood, P.C., and Li, Q.Q. (2005). Compilation of mRNA Polyadenylation Signals in *Arabidopsis* Revealed a New Signal Element and Potential Secondary Structures. Plant Physiol **138,** 1457-1468.

Lu, Q., Xu, Z.K., and Song, R.T. (2006). OsFY, a homolog of AtFY, encodes a protein that can interact with OsFCA-gamma in rice (*Oryza sativa* L.). Acta Biochim Biophys Sin (Shanghai) **38,** 492-499.

Ma, L., Guo, C., and Li, Q.Q. (2014). Role of alternative polyadenylation in epigenetic silencing and antisilencing. Proc Natl Acad Sci U S A **111,** 9-10.

Ma, W., Chen, C., Liu, Y., Zeng, M., Meyers, B.C., Li, J., and Xia, R. (2018). Coupling of microRNA-directed phased small interfering RNA generation from long noncoding genes with alternative splicing and alternative polyadenylation in small RNA-mediated gene silencing. New Phytol **217,** 1535-1550.

Macknight, R., Duroux, M., Laurie, R., Dijkwel, P., Simpson, G., and Dean, C. (2002). Functional significance of the alternative transcript processing of the *Arabidopsis* floral promoter FCA. Plant Cell **14,** 877-888.

Macknight, R., Bancroft, I., Page, T., Lister, C., Schmidt, R., Love, K., Westphal, L., Murphy, G., Sherson, S., Cobbett, C., and Dean, C. (1997). FCA, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. Cell **89,** 737-745.

Manzano, D., Marquardt, S., Jones, A.M.E., Bäurle, I., Liu, F., and Dean, C. (2009). Altered interactions within FY/AtCPSF complexes required for *Arabidopsis* FCA-mediated chromatin silencing. Proc Natl Acad Sci U S A **106,** 8772-8777.

Martin, G., Gruber, A.R., Keller, W., and Zavolan, M. (2012). Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. Cell Rep **1,** 753-763.

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

Michalski, D., and Steiniger, M. (2015). In vivo characterization of the *Drosophila* mRNA 3' end processing core cleavage complex. RNA **21,** 1404-1418.

Monti, K.L. (1995). Folded Empirical Distribution Function Curves-Mountain Plots. The American Statistician **49,** 342-345.

Neve, J., Burger, K., Li, W., Hoque, M., Patel, R., Tian, B., Gullerova, M., and Furger, A. (2016). Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. Genome Res **26,** 24-35.

Pan, H., Oztas, O., Zhang, X., Wu, X., Stonoha, C., Wang, E., Wang, B., and Wang, D. (2016). A symbiotic SNARE protein generated by alternative termination of transcription. Nat Plants **2,**

15197-15201.

Rao, S., Dinkins, R.D., and Hunt, A.G. (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 Biol. **10,** 51.

Riester, L., Koster-Hofmann, S., Doll, J., Berendzen, K.W., and Zentgraf, U. (2019). Impact of Alternatively Polyadenylated Isoforms of ETHYLENE RESPONSE FACTOR4 with Activator and Repressor Function on Senescence in *Arabidopsis thaliana* L. Genes **10**.

Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A., and Burge, C.B. (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science **320,** 1643-1647.

Schönemann, L., Kühn, U., Martin, G., Schäfer, P., Gruber, A.R., Keller, W., Zavolan, M., and Wahle, E. (2014). Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. Genes Dev **28,** 2381-2393.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res **13,** 2498-2504.

Shi, Y., and Manley, J.L. (2015). The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. Genes Dev **29,** 889-897.

Shi, Y., Giammartino, D.C.D., Taylor, D., Sarkeshik, A., Rice, W.J., Iii, J.R.Y., Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3′ processing complex. Mol Cell **33,** 365-376.

Silva, A.T., Ribone, P.A., Chan, R.L., Ligterink, W., and Hilhorst, H.W. (2016). A Predictive Coexpression Network Identifies Novel Genes Controlling the Seed-to-Seedling Phase Transition in *Arabidopsis thaliana*. Plant Physiol **170,** 2218-2231.

Simpson, G.G., Dijkwel, P.P., Quesada, V., Henderson, I., and Dean, C. (2003). FY Is an RNA 3′ End-Processing Factor that Interacts with FCA to Control the *Arabidopsis* Floral Transition. Cell **113,** 777-787.

Smibert, P., Miura, P., Westholm, J.O., Shenker, S., May, G., Duff, M.O., Zhang, D., Eads, B.D., Carlson, J., and Brown, J.B. (2012). Global Patterns of Tissue-Specific Alternative Polyadenylation in *Drosophila*. Cell Rep **1,** 277-289.

Smith, T.F., Gaitatzes, C., Saxena, K., and Neer, E.J. (1999). The WD repeat: a common architecture for diverse functions. Trends Biochem Sci **24,** 181-185.

Sorin, C., Negroni, L., Balliau, T., Corti, H., Jacquemot, M.P., Davanture, M., Sandberg, G., Zivy, M., and Bellini, C. (2006). Proteomic analysis of different mutant genotypes of *Arabidopsis* led to the identification of 11 proteins correlating with adventitious root development. Plant Physiol **140,** 349-364.

Sun, Y., Zhang, Y., Hamilton, K., Manley, J.L., Shi, Y., Walz, T., and Tong, L. (2018). Molecular basis for the recognition of the human AAUAAA polyadenylation signal. Proc Natl Acad Sci U S A **115,** E1419-E1428.

Takagaki, Y., Ryner, L.C., and Manley, J.L. (1989). Four factors are required for 3'-end cleavage of pre-mRNAs. Genes Dev **3,** 1711-1724.

Thomas, P.E., Wu, X., Liu, M., Gaffney, B., Ji, G., Li, Q.Q., and Hunt, A.G. (2012). Genome-wide control of polyadenylation site choice by CPSF30 in *Arabidopsis*. Plant Cell **24,** 4376-4388.

Tian, B., and Manley, J.L. (2017). Alternative polyadenylation of mRNA precursors. Nat Rev Mol Cell Biol **18,** 18-30.

Ulitsky, I., Shkumatava, A., Jan, C.H., Subtelny, A.O., Koppstein, D., Bell, G.W., Sive, H., and Bartel, D.P. (2012). Extensive alternative polyadenylation during zebrafish development. Genome Res **22,** 2054-2066.

Wang, X., and Auwerx, J. (2017). Systems Phytohormone Responses to Mitochondrial Proteotoxic Stress. Mol Cell **68,** 540-551.

Wu, X., Liu, M., Downie, B., Liang, C., Ji, G., Li, Q.Q., and Hunt, A.G. (2011). Genome-wide landscape of polyadenylation in *Arabidopsis* provides evidence for extensive alternative polyadenylation. Proc Natl Acad Sci U S A **108,** 12533-12538.

Wu, X.Z., Li, F.L., Kolenovsky, A., Caplan, A., Cui, Y.H., Cutler, A., Tsang, E.W.T., Peterson, R.L., and Shelp, B.J. (2009). A mutant deficient in S-adenosylhomocysteine hydrolase in *Arabidopsis* shows defects in root-hair development. Botany **87,** 571-584.

Xing, D., and Li, Q.Q. (2011). Alternative polyadenylation and gene expression regulation in plants. Wiley Interdiscip Rev RNA **2,** 445-458.

Xu, R., Zhao, H., Dinkins, R.D., Cheng, X., Carberry, G., and Li, Q.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.

Zeng, W., Dai, X., Sun, J., Hou, Y., Ma, X., Cao, X., Zhao, Y., and Cheng, Y. (2019). Modulation of Auxin Signaling and Development by Polyadenylation Machinery. Plant Physiol **179,** 686-699.

Zhan, J., Thakare, D., Ma, C., Lloyd, A., Nixon, N.M., Arakaki, A.M., Burnett, W.J., Logan, K.O., Wang, D., and Wang, X. (2015). RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. Plant Cell **27,** 513-531.

Zhang, J., Addepalli, B., Yun, K.Y., Hunt, A.G., Xu, R., Rao, S., Li, Q.Q., and Falcone, D.L. (2008). A polyadenylation factor subunit implicated in regulating oxidative signaling in *Arabidopsis thaliana*. PLos One **3,** e2410.

Zhang, Y., Gu, L., Hou, Y., Wang, L., Deng, X., Hang, R., Chen, D., Zhang, X., Zhang, Y., Liu, C., and Cao, X. (2015). Integrative genome-wide analysis reveals HLP1, a novel RNA-binding protein, regulates plant flowering by targeting alternative polyadenylation. Cell Res **25,** 864-876.

Zhao, Z., Wu, X., Kumar, P.K.R., Dong, M., Ji, G., Li, Q.Q., and Liang, C. (2014). Bioinformatics Analysis of Alternative Polyadenylation in Green Alga Chlamydomonas reinhardtii Using Transcriptome Sequences from Three Different Sequencing Platforms. G3 (Bethesda) **4,** 871-883.

Zhou, Q., Fu, H., Yang, D., Ye, C., Zhu, S., Lin, J., Ye, W., Ji, G., Ye, X., Wu, X., and Li, Q.Q. (2019). Differential alternative polyadenylation contributes to the developmental divergence between two rice subspecies, japonica and indica. Plant J **98,** 260-276.

940
941

FIGURE LEGENDS

Figure 1. Schematic representation of *FY* and its mutants, and the transcriptional level of *FY*. (A) The top and bottom represent *FY* gene and protein. The top of the gene shows the position of *fy* mutations. The black boxes and lines represent exons and introns, respectively. The FY protein contains seven WD domains and two PPLPP-domains. The purple, blue and red bars show the RT-qPCR amplicons located in 7×WD (P1), the first PPLPP-domain (P2) and the second PPLPP-domain (P3) regions. **(B)** Schema of FY in mutants. PPLPP-domain deficiency in *fy1*, *fy2* and *fy5*. The first WD domain amino acid was changed in *fy3* and indicated by ⋆. *FY* is overexpressed in *fy6* line. **(C)** and **(D)** RT-qPCR quantification of *FY* transcription level in the mutants and wild types. The wild type of *fy1* is L*er,* and wild type for the rest is Col-0. RT-qPCR quantification of *FY* expression levels in *fy1*, *fy2*, *fy3*, *fy5*, *fy6* and WT were done in the P1 and P2 regions, and only *fy5* was performed in the P3 region. Error bars represent standard deviation from three biological replicates and asterisks are indicative of statistically significant differences between wild type and mutant using one-way ANOVA (⋆ indicates *P*-value<0.05. ⋆⋆ indicates *P*-value<0.01).

Figure 2. The analysis of hierarchical clustering and Cumulative Distribution Function (CDF) based on Poly(A) Usage (PAU). PAU values were calculated as the ratio of its expression to the sum of the expression of all isoforms for each APA gene, and based on the average of three biological replicates. **(A)** Hierarchical cluster analysis of PAU. **(B and C)** The curves of CDF. The x-axis is the log values of the ratio of poly(A) site in all isoforms of a single gene. The curve of CDF was based on a mountain plot to examine PAU distribution, and the mountain plot is formed by reflecting the two halves, folded at y = 50%. The Kolmogorov-Smirnov (K–S) test is used to detect the differences in both location and shape of the empirical cumulative distribution functions between the mutant

and its wild type. Compared to WT, the *P*-value of *fy3*, *fy5*, *fy6*, *oxt6*, *fy2 oxt6, fy3 oxt6*, *fy6 oxt6* and *fca-1* was less than 2.2e-16. *fy1* and *fy2* were 1.16e-06 and 1.88e-10, respectively. **B and C** were separated because of two different ecotype backgrounds. The numbers in the figure insert represented the maximum distance in paired comparison between two CDF curves. ** indicates *P*-value<0.01.

Figure 3. Single nucleotide profiles of NUE of the mutants. (A-J) Nucleotide profiles of unique PACs in different mutants, WT and common PACs. "*oxt6* unique" - sites seen only in the *oxt6* mutant relative to WT, "WT unique" - sites seen only in WT relative to *oxt6*, "Common" - poly(A) sites seen in both the WT and *oxt6*. All other mutants are shown in the same way. "n" represents transcript number. The y-axis indicates the fraction of nucleotide composition at x-axis locations, e.g., -10 indicates 10 nucleotide up-stream of the poly(A) site.

Figure 4. A heat map of module-sample associations. The left panel shows the 20 modules (M) and the number in parentheses represent transcript number. Each row corresponds to a module, and each column corresponds to a mutant line or wild type. The color scale on right shows module-mutant correlations from 1 (red) to –1 (blue). The color of each cell at the row-column intersection represents the correlation coefficient (upper values) between the modules and samples. Red color indicates a high degree of positive correlation, and blue color indicates a high degree of negative correlation, between each module and the mutant or wild type. Each cell also contains the corresponding *P*-value (lower values).

Figure 5. Co-expression network analysis of specific modules. (A) AAUAAA signal and its 1-nt variant signal of the NUE region (between 10 and 35 bases upstream of poly(A) sites) are analyzed in each module. **(B)** and **(C)** The network of the 63 and 91 highly

connected transcripts in M13 and M3, respectively. The networks were visualized using Cytoscape 3.6.0 software and the protein name or gene ID (no protein name) were shown in Figure. Candidate hub genes in the module are shown in yellow.

Figure 6. Analysis of differentially expressed (DE) PAC APA genes and DE genes. (A) and (B) A bar graph showing the number of DE-PAC APA genes and DE genes, which were analyzed by DESeq2 package. A *padj*<0.05 threshold was considered statistically significant. **(C)** GO enrichment analysis of DE-PAC APA genes. All significant GO terms of biological process at the second level were shown. FDR: false discovery rate; solid line: FDR=0.01. dashed line: FDR=0.05. **(D)** Venn diagram of DE-PAC APA genes and DE genes. The blue circle represents DE-PAC APA gene; the orange circle represents DE gene. The number in the circles show DE-PAC APA gene or DE gene count. The percentage of DE gene belong to DE-PAC APA gene is 58% in *fy1*, 53% in *fy2*, 59% in *fy3*, 62% in *fy5*, 60% in *fy6*, 57% in *oxt6*, 58% in *fy2 oxt6*, 59% in *fy3 oxt6*, and 59% in *fy6 oxt6*.

Figure 7. Single nucleotide profiles of NUE located in the 3'UTR. Nucleotide profiles **(A-G),** AAUAAA signal usage **(H)** and 1-nt variants of AAUAAA signal usage **(I)** of different mutants unique PACs. WT unique PACs and common PACs. "*oxt6* unique" - sites seen only in the *oxt6* mutant relative to WT; "WT unique" - sites seen only in WT relative to *oxt6;* "Common" - poly(A) sites seen in both the WT and *oxt6*. All other mutants are shown in the same way. "n" represents transcript number. The y-axis indicates the nucleotide composition at x-axis locations. (**H)** and (**I)**, "WT-*oxt6*" indicates the control for *oxt6* mutant. Others are labeled the same way.

Figure 8. Distributions of PATs and PACs selected from DE-PACs (|fold change|>=2) in 3'UTR. (A) PATs distribution. **(B)** PACs distribution. Error bars represent standard deviation from three biological replicates, with 10 plants in each repeat, and asterisks are

indicative of statistically significant differences using one-way ANOVA (⋆ indicates *P*-value<0.05. ⋆⋆ indicates *P*-value<0.01).

Figure 9. 3'UTR APA analysis. (A) Comparison of 3'UTR significantly lengthen or shorten genes. For each gene, 3'UTR average weighted length was defined as the sum of 3'UTR length (the distance from each PAC location to the stop codon) of each PAC multiplied by its expression level (average value of three biological repeats normalized PATs) and then divided by the total expression level. A cut-off *P*-value of 0.05 was adopted for both significantly longer and shorter 3'UTR between mutant and wild type. **(B) and (C)** The box plot was used to show the 3'UTR average weighted length distribution of significantly lengthen or shorten in *fy3*, respectively.

Figure 10. Analysis of primary root phenotype and related gene APA. (A) The phenotype of root. Each line contained two seedlings. **(B)** The root length was measured by ImageJ software. Box plots showing change in primary root length. **(C) and (F)** The sequencing coverage of primary root related gene *SAHH* and *ATHB13* were visualized by Integrative Genomics Viewer (IGV) software. **(D) and (G)** Distal transcript and total gene expression of primary root related gene *SAHH* and *ATHB13* were verified by RT-qPCR. Error bars represent standard deviation from three biological replicates. **(E) and (H)** RNA stability assay. RT-qPCR analysis of distal transcript and total gene expression of primary root related gene *SAHH* and *ATHB13* in control and after 120 min of cordycepin conditions. Error bars represent standard deviation from three biological replicates (pooling ~10 plants per condition), and asterisks are indicative of statistically significant differences using one-way ANOVA (⋆ indicates *P*-value<0.05. ⋆⋆ indicates *P*-value<0.01). C0, control conditions; C120, mRNA after 120 min of cordycepin treatment.

Figure 11. Analysis of flowering phenotype and related gene APA. (A) Flowering time was measured by counting the number of rosette leaves at flowering under long-day photoperiods in the incubator. Each pool contained one plant. Each experiment comprised 18 pools and three independent experiments were completed. **(B) and (C)** The sequencing coverage of *FLC* and *FCA* were visualized by Integrative Genomics Viewer (IGV) software. **(D) and (E)** Distal transcript and total gene expression of *FLC* and *FCA* were verified by RT-qPCR. Error bars represent standard deviation from three biological replicates and asterisks are indicative of statistically significant differences using one-way ANOVA (⋆ indicates *P*-value<0.05. ⋆⋆ indicates *P*-value<0.01).

Figure 12. Abiotic stress responses of the SALK_146237 and SALK_205297 and 3'RACE analysis. (A) Photographs of seedlings grown on ½ MS medium or ½ MS medium containing 125 mM NaCl at day 14 after the end of stratification and grown on ½ MS or containing 50 nM MV at day 12. **(B-C)** Seed germination rates of the indicated genotypes grown on ½ MS medium or ½ MS medium containing 125 mM NaCl were quantified every day from the 2nd day to the 14th day after sowing. Three independent experiments were conducted. 40 seeds per genotype were measured in each replicate. Values are mean ± SD of three replications. **(D)** Cotyledon-greening percentages of the 14th day were recorded. **(E)** The root length was measured by ImageJ software. Error bars represent standard deviation from three biological replicates and asterisks are indicative of statistically significant differences using one-way ANOVA (⋆ indicates *P*-value<0.05. ⋆⋆ indicates *P*-value<0.01). **(F)** Illustration of gene constructs and 3'RACE experiment results of AT3G47610 and *AKR2*. The red and blue region represent PA1 and PA2 region, respectively.

Table 1. Summary of PAT mapping and poly(A) site.

Sample	raw read#	PAT#	PAC#	Sample	raw read#	PAT#	PAC#
L*er*-rep1	20,492,374	13,402,960	37,477	*fy6*-rep1	6,729,599	4,230,734	33,644
L*er*-rep2	22,410,997	14,642,333	42,909	*fy6*-rep2	7,172,240	5,095,320	27,691
L*er*-rep3	23,726,124	15,800,072	43,454	*fy6*-rep3	15,213,057	7,498,700	37,900
fy1-rep1	20,289,695	13,194,738	43,544	*fy2/oxt6*-rep1	12,195,428	7,781,310	37,748
fy1-rep2	21,784,455	13,256,921	40,955	*fy2/oxt6*-rep2	10,590,120	6,098,745	35,439
fy1-rep3	21,389,641	14,460,622	43,889	*fy2/oxt6*-rep3	13,756,239	7,442,659	38,363
Col-rep1	37,153,184	14,091,055	44,772	*fy3/oxt6*-rep1	17,320,310	8,939,522	40,325
Col-rep2	31,094,544	10,583,248	42,190	*fy3/oxt6*-rep2	14,761,712	7,113,164	38,526
Col-rep3	31,520,633	10,754,403	43,224	*fy3/oxt6*-rep3	11,275,339	5,462,609	38,468
fy2-rep1	17,334,713	8,520,124	42,288	*fy6/oxt6*-rep1	2,742,850	1,506,088	26,686
fy2-rep2	22,023,128	9,919,939	42,351	*fy6/oxt6*-rep2	4,060,464	1,511,615	26,496
fy2-rep3	20,640,430	9,105,296	41,932	*fy6/oxt6*-rep3	7,009,595	3,028,461	32,240
fy3-rep1	14,116,814	6,948,370	35,481	*oxt6*-rep1	15,606,616	7,585,236	37,009
fy3-rep2	9,543,788	5,634,440	36,012	*oxt6*-rep2	18,124,350	9,253,847	40,657
fy3-rep3	10,646,487	6,721,523	37,016	*oxt6*-rep3	16,747,050	8,538,587	39,665
fy5-rep1	14,382,474	5,856,273	40,390	*fca-1*-rep1	9,854,259	7,320,494	45,185
fy5-rep2	15,826,806	5,901,180	39,267	*fca-1*-rep2	11,637,455	9,283,112	45,303
fy5-rep3	18,091,769	6,900,259	41,313	*fca-1*-rep3	12,299,062	10,249,445	45,719

Note: PAT#, Numbers of individual tags after curation (to remove low-quality reads, invalid poly(T) reads and unmapped tags). PAC#, Numbers of PACs obtained after grouping poly(A) sites that lie within 24 nt of adjacent sites. “rep” represents “repeat”.

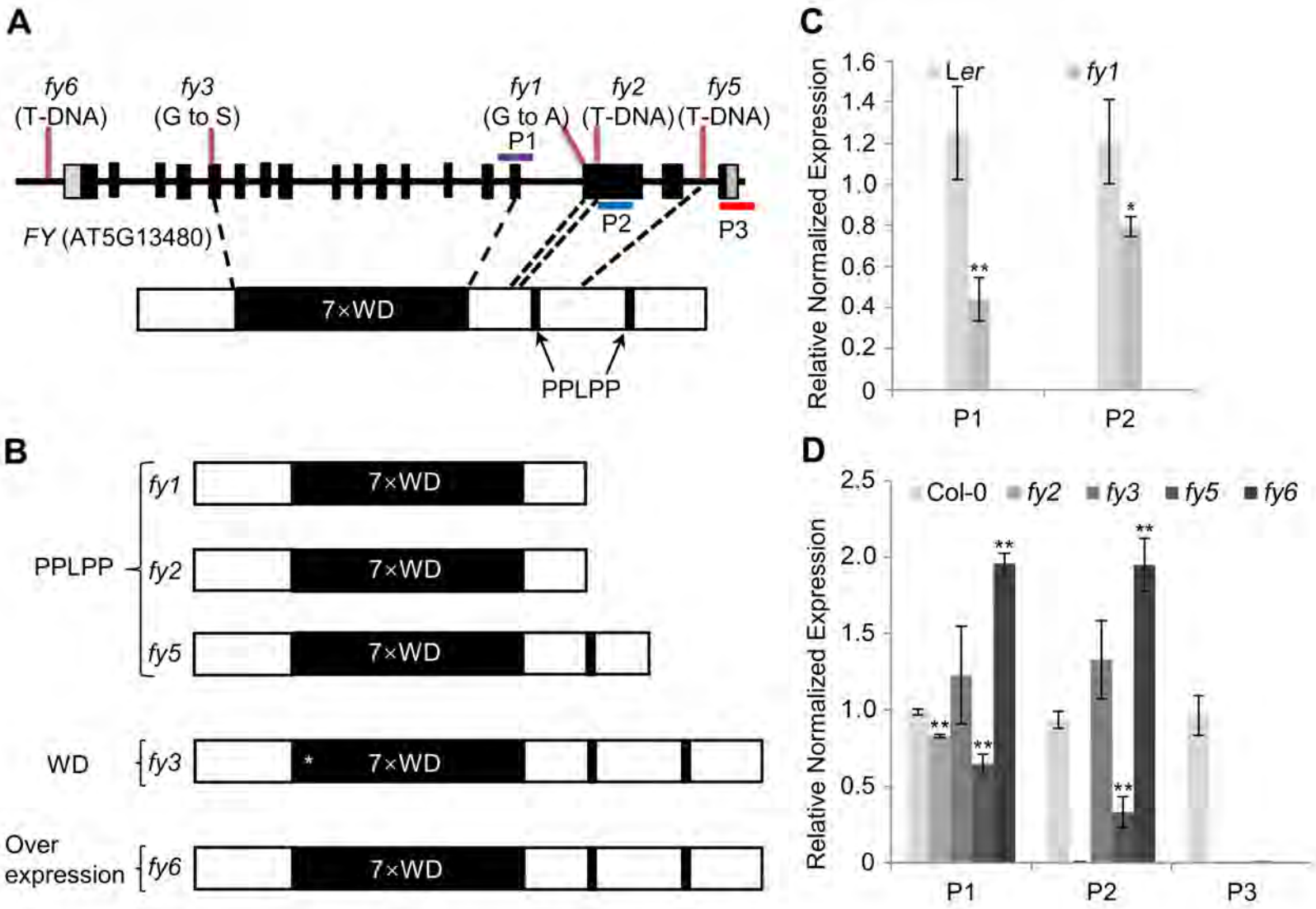

Figure 1. Schematic representation of *FY* and its mutants, and the transcriptional level of *FY*. **(A)** The top and bottom represent *FY* gene and protein. The top of the gene shows the position of *fy* mutations. The black boxes and lines represent exons and introns, respectively. The FY protein contains seven WD domains and two PPLPP domains. The purple, blue and red bars show the qRT-PCR amplicons located in 7×WD (P1), the first PPLPP (P2) and the second PPLPP (P3) regions. **(B)** Schema of FY in mutants. PPLPP domain deficiency in *fy1*, *fy2* and *fy5*. The first WD domain amino acid was changed in *fy3* and indicated by *. *FY* is overexpressed in *fy6*. **(C)** and **(D)** qRT-PCR quantification of *FY* transcription level in the mutants and wild types. The wild type of *fy1* is L*er,* and wild type for the rest is Col-0. qRT-PCR quantification of *FY* expression levels in *fy1*, *fy2*, *fy3*, *fy5*, *fy6* and WT were done in the P1 and P2 regions, and only *fy5* was performed in the P3 region. Error bars represent standard deviation from three biological replicates and asterisks are indicative of statistically significant differences between wild type and mutant using one-way ANOVA (* indicates *P*-value<0.05. ** indicates *P*-value<0.01).

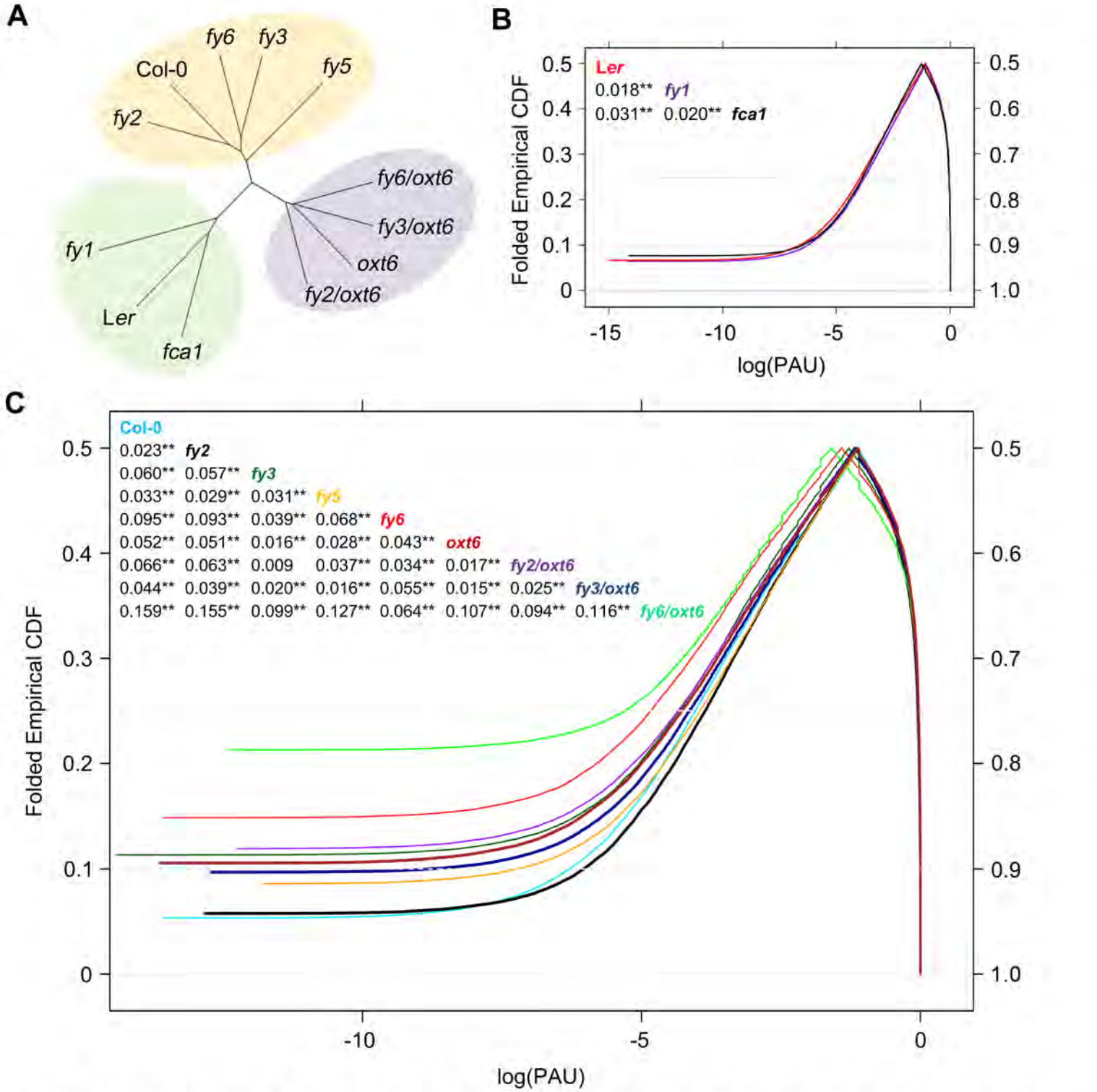

Figure 2. The analysis of hierarchical clustering and Cumulative Distribution Frequency (CDF) based on Poly(A) Usage (PAU). PAU values were calculated as the ratio of its expression to the sum of the expression of all isoforms for each APA gene, and based on the average of three biological replicates. **(A)** Hierarchical cluster analysis of PAU. **(B and C)** The curves of CDF. The x-axis is the log values of the ratio of poly(A) site in all isoforms of a single gene. The curve of CDF was based on a mountain plot to examine PAU distribution, and the mountain plot is formed by reflecting the two halves, folded at y = 50%. The Kolmogorov-smirnov (K–S) test is used to detect the differences in both location and shape of the empirical cumulative distribution functions between the mutant and its wild type. Compared to WT, the *p*-value of *fy3*, *fy5*, *fy6*, *oxt6*, *fy2/oxt6*, *fy3/oxt6*, *fy6/oxt6* and *fca1* was less than 2.2e-16. *fy1* and *fy2* were 1.16e-06 and 1.88e-10, respectively. B and C were separated because of two different ecotype backgrounds. The numbers in the figure insert represented the maximum distance in paired comparison between two CDF curves. ** indicates *P*-value<0.01.

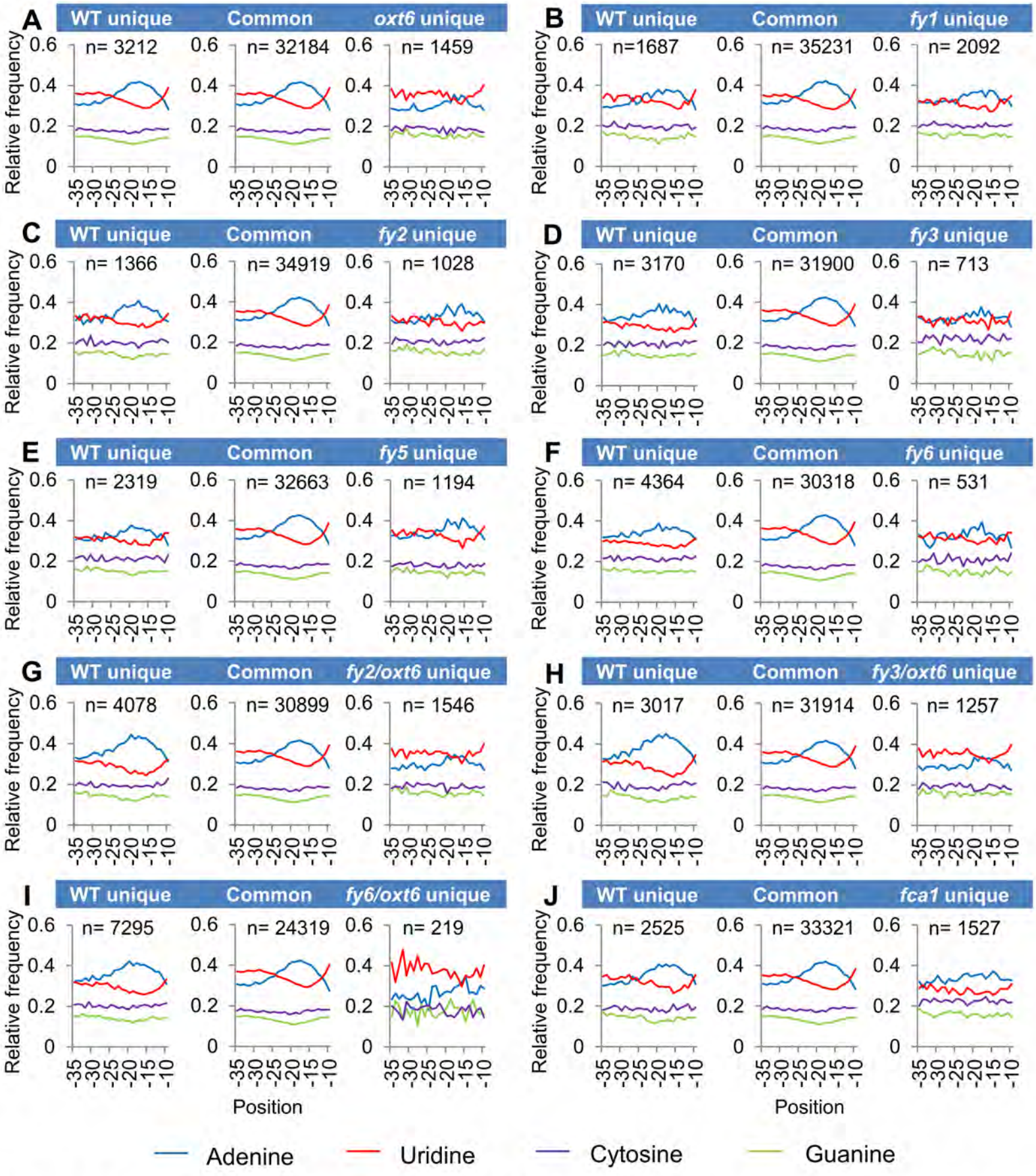

Figure 3. Single nucleotide profiles of NUE of the mutants. (A-J) Nucleotide profiles of unique PACs in different mutants, WT and common PACs. "*oxt6* unique" - sites seen only in the *oxt6* mutant relative to WT, "WT unique" - sites seen only in WT relative to *oxt6*, "Common" - poly(A) sites seen in both the WT and *oxt6*. All other mutants are shown in the same way. "n" represents transcript number. The y-axis indicates the fraction of nucleotide composition at x-axis locations, e.g., -10 indicates 10 nucleotide up-stream of the poly(A) site.

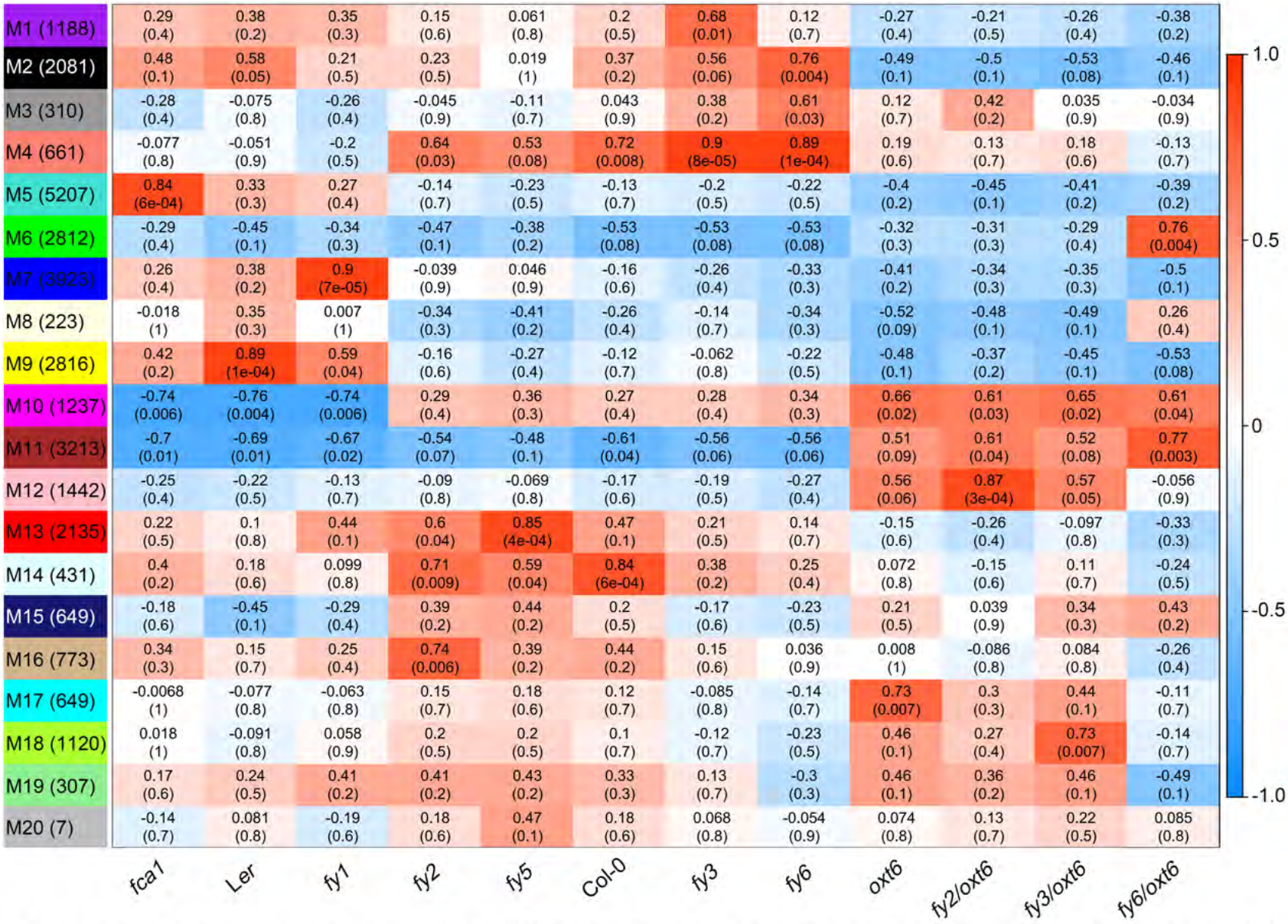

Figure 4. A heat map of module-sample associations. The left panel shows the 20 modules (M) and the number in parentheses represent transcript number. Each row corresponds to a module, and each column corresponds to a mutant line or wild-type. The color scale on right shows module-mutant correlations from 1 (red) to –1 (blue). The color of each cell at the row-column intersection represents the correlation coefficient (upper values) between the modules and samples. Red color indicates a high degree of positive correlation, and blue color indicates a high degree of negative correlation, between each module and the mutant or wild-type. Each cell also contains the corresponding *P*-value (lower values).

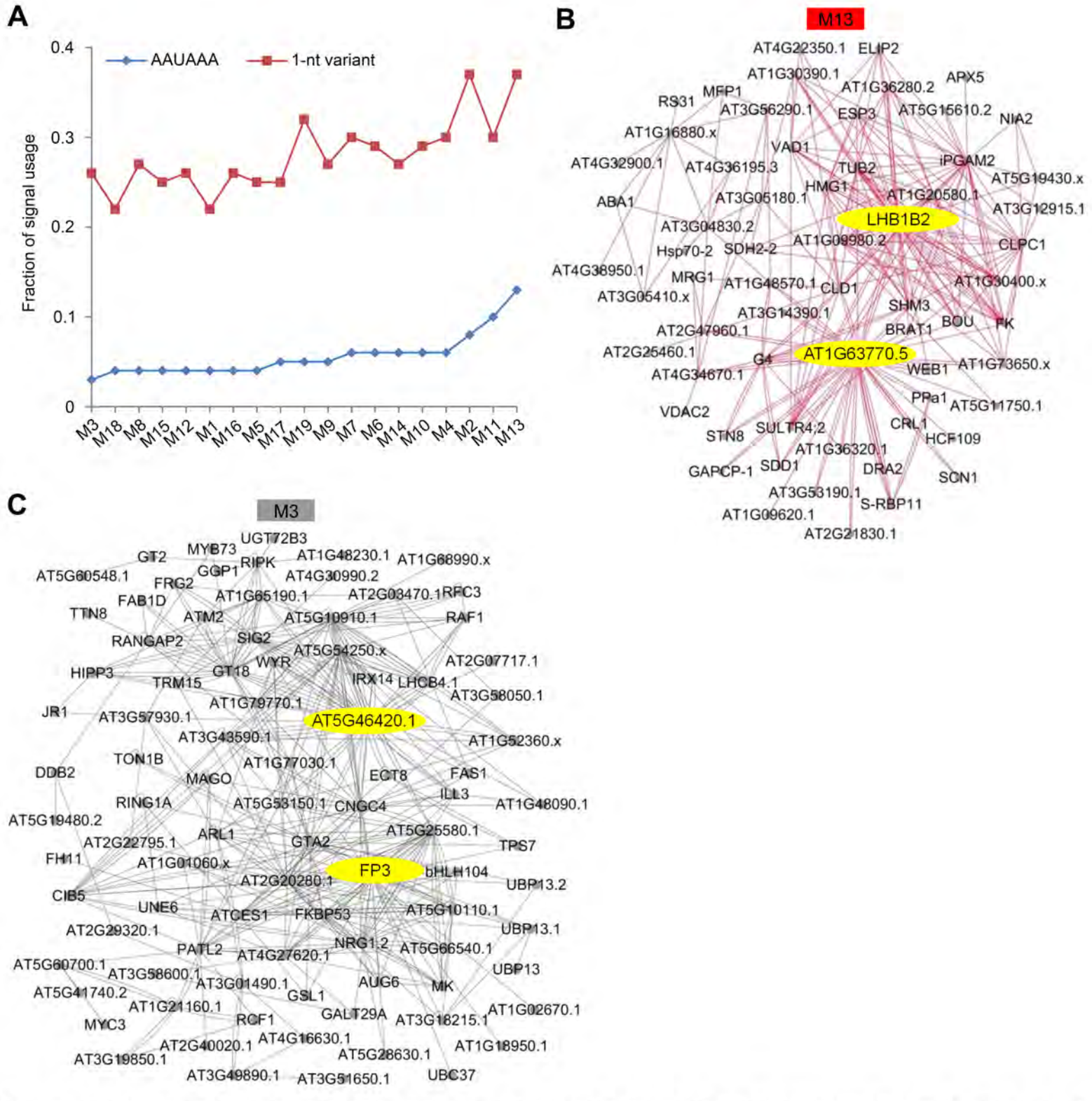

Figure 5. Co-expression network analysis of specific modules. (A) AAUAAA signal and its 1-nt variant signal of the NUE region (between 10 and 35 bases upstream of poly(A) sites) are analyzed in each module. **(B) and (C)** The network of the 63 and 91 highly connected transcripts in M13 and M3, respectively. The networks were visualized using Cytoscape 3.6.0 software and the protein name or gene ID (no protein name) were shown in Figure. Candidate hub genes in the module are shown in yellow.

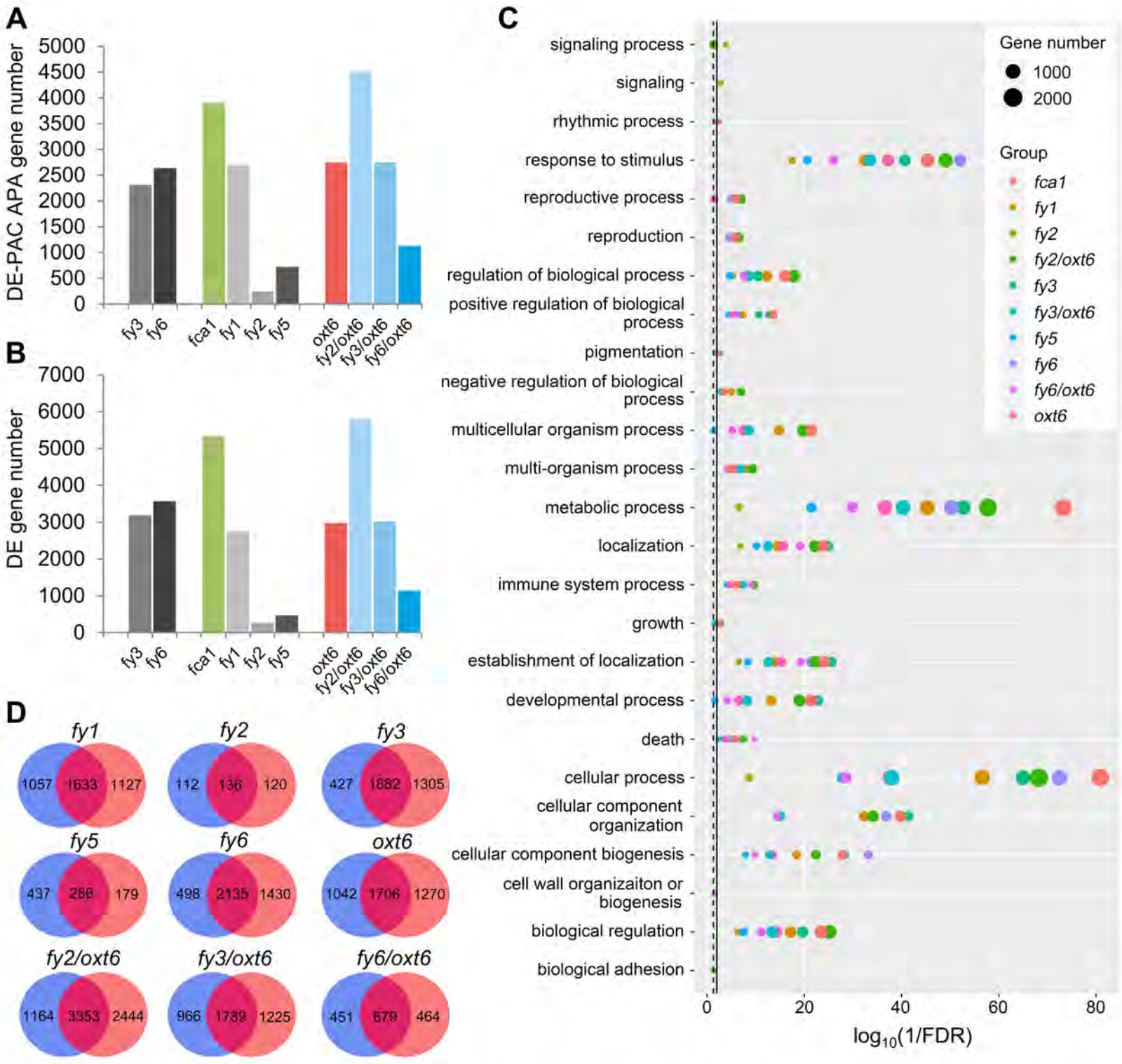

Figure 6. Analysis of differentially expressed (DE) PAC APA genes and DE genes. (A) and (B) A bar graph showing the number of DE-PAC APA genes and DE genes, which were analyzed by DESeq2 package. A *padj*<0.05 threshold was considered statistically significant. **(C)** GO enrichment analysis of DE-PAC APA genes. All significant GO terms of biological process at the second level were shown. FDR: false discovery rate; solid line: FDR=0.01. dashed line: FDR=0.05. **(D)** Venn diagram of DE-PAC APA genes and DE genes. The blue circle represents DE-PAC APA gene; the orange circle represents DE gene. The number in the circles show DE-PAC APA gene or DE gene count. The percentage of DE gene belong to DE-PAC APA gene is 58% in *fy1*, 53% in *fy2*, 62% in *fy5*, 59% in *fy3*, 60% in *fy6*, 57% in *oxt6*, 58% in *fy2/oxt6*, 59% in *fy3/oxt6*, 59% in *fy6/oxt6* and 58% in *fca1*.

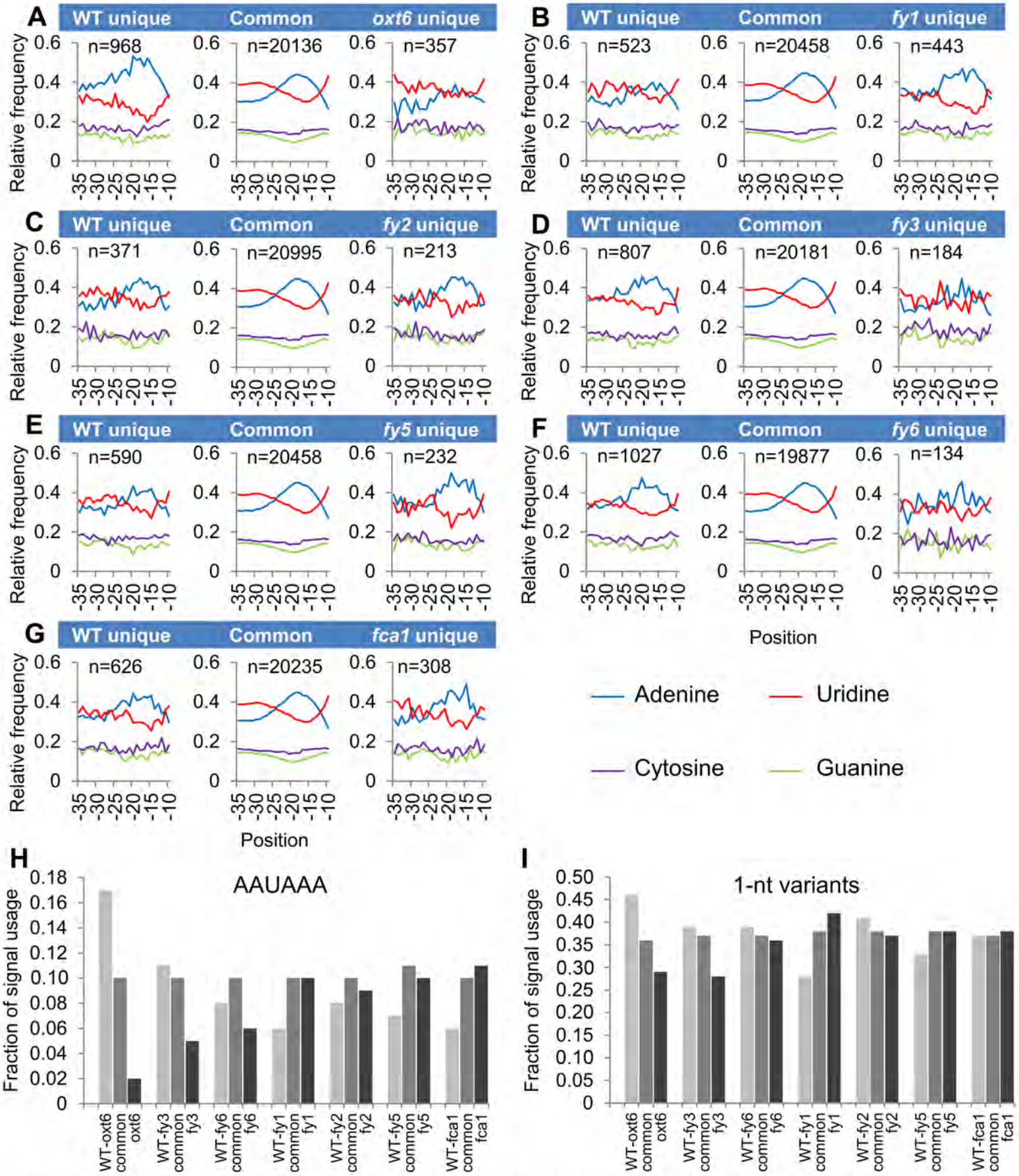

Figure 7. Single nucleotide profiles of NUE located in the 3'UTR. Nucleotide profiles **(A-G),** AAUAAA signal usage **(H)** and 1-nt variants of AAUAAA signal usage **(I)** of different mutants unique PACs. WT unique PACs and common PACs. "*oxt6* unique" - sites seen only in the *oxt6* mutant relative to WT; "WT unique" - sites seen only in WT relative to *oxt6;* "Common" - poly(A) sites seen in both the WT and *oxt6*. All other mutants are shown in the same way. "n" represent transcript number. The y-axis indicates the nucleotide composition at x-axis locations. (**H**) and (**I**), "WT-*oxt6*" indicates the control for *oxt6* mutant. Others are labelled the same way.

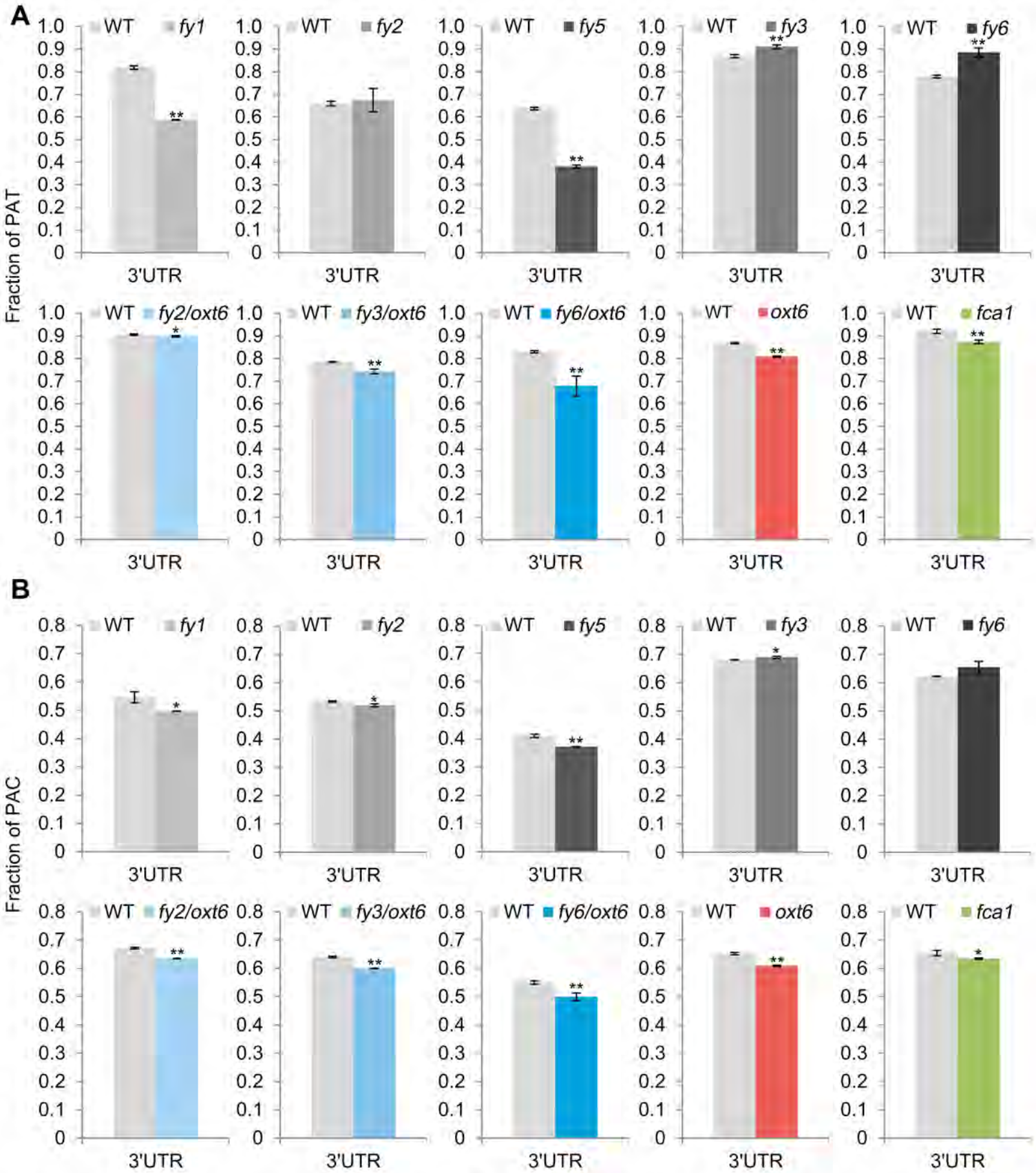

Figure 8. Distributions of PATs and PACs selected from DE-PACs (|fold change|>=2) in 3'UTR. (A) PATs distribution. **(B)** PACs distribution. Error bars represent standard deviation from three biological replicates, with 10 plants in each repeat, and asterisks are indicative of statistically significant differences using one-way ANOVA (* indicates *P*-value<0.05. ** indicates *P*-value<0.01).

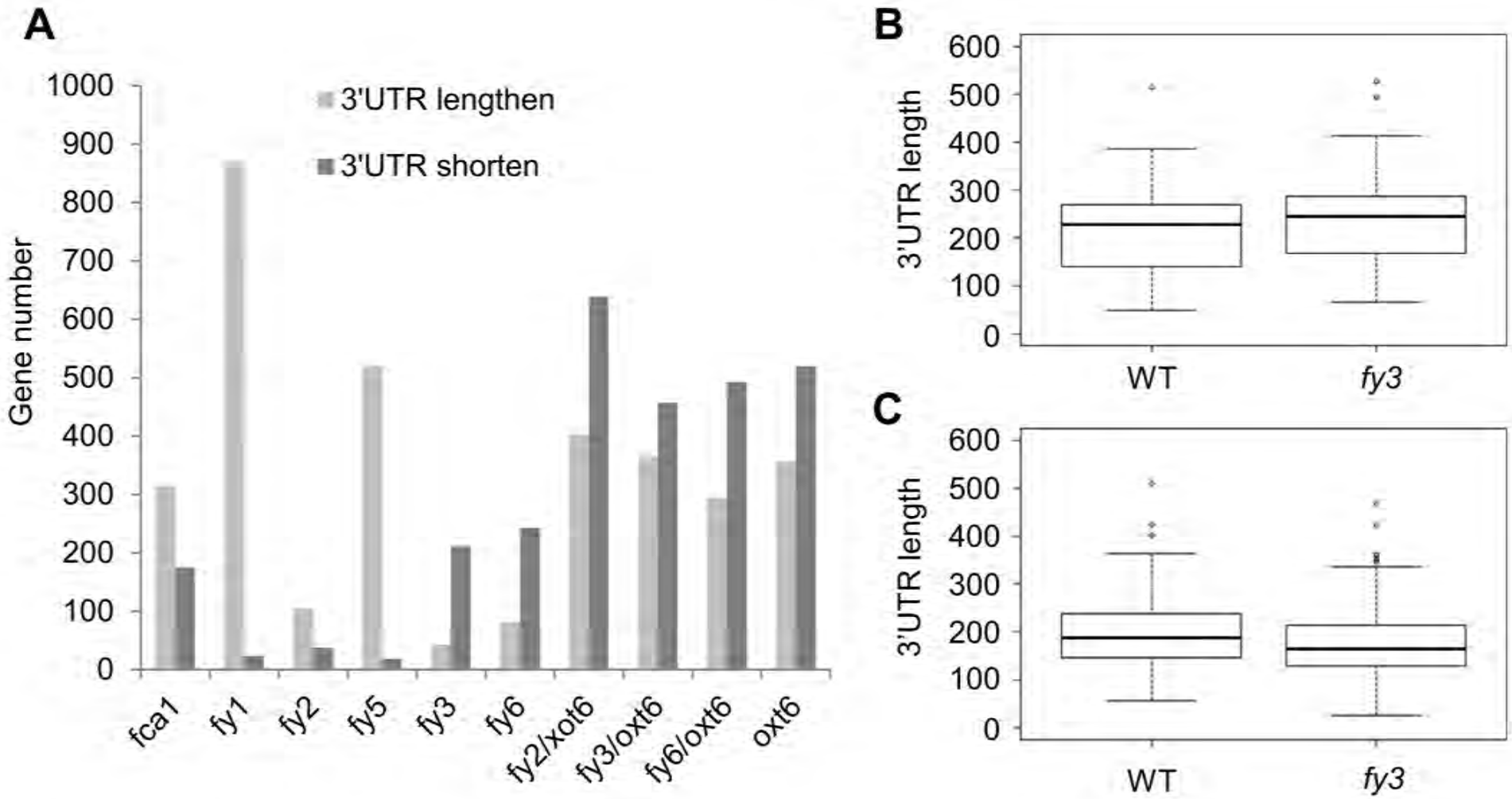

Figure 9. 3'UTR APA analysis. (A) Comparison of 3'UTR significantly lengthen or shorten genes. For each gene, 3'UTR average weighted length was defined as the sum of 3'UTR length (the distance from each PAC location to the stop codon) of each PAC multiplied by its expression level (average value of three biological repeats normalized PATs) and then divided by the total expression level. A cut-off *P*-value of 0.05 was adopted for both significantly longer and shorter 3'UTR between mutant and wild type. **(B) and (C)** The box plot was used to show the 3'UTR average weighted length distribution of significantly lengthen or shorten in *fy3*, respectively.

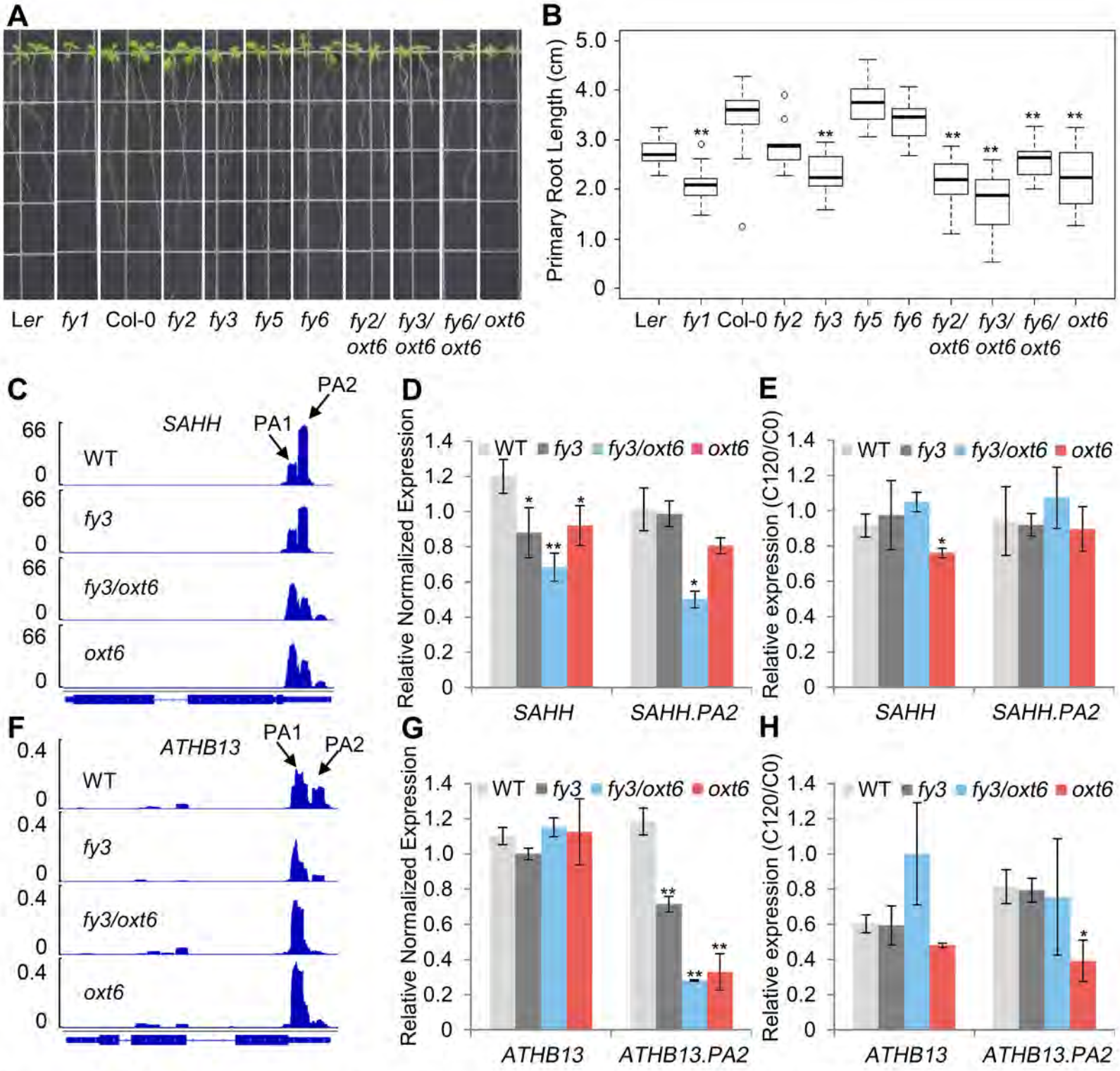

Figure 10. Analysis of primary root phenotype and related gene APA. (A) The phenotype of root. Each lines contained two seedlings. **(B)** The root length was measured by ImageJ software. Box plots showing change in primary root length. **(C) and (F)** The sequencing coverage of primary root related gene *SAHH* and *ATHB13* were visualized by Integrative Genomics Viewer (IGV) software. **(D) and (G)** Distal transcript and total gene expression of primary root related gene *SAHH* and *ATHB13* were verified by qRT-PCR. Error bars represent standard deviation from three biological replicates. **(E) and (H)** RNA stability assay. Quantitative RT-PCR analysis of distal transcript and total gene expression of primary root related gene *SAHH* and *ATHB13* in control and after 120 min of cordycepin conditions. Error bars represent standard deviation from three biological replicates (pooling ~10 plants per condition), and asterisks are indicative of statistically significant differences using one-way ANOVA (* indicates *P*-value<0.05. ** indicates *P*-value<0.01). C0, control conditions; C120, mRNA after 120 min of cordycepin treatment.

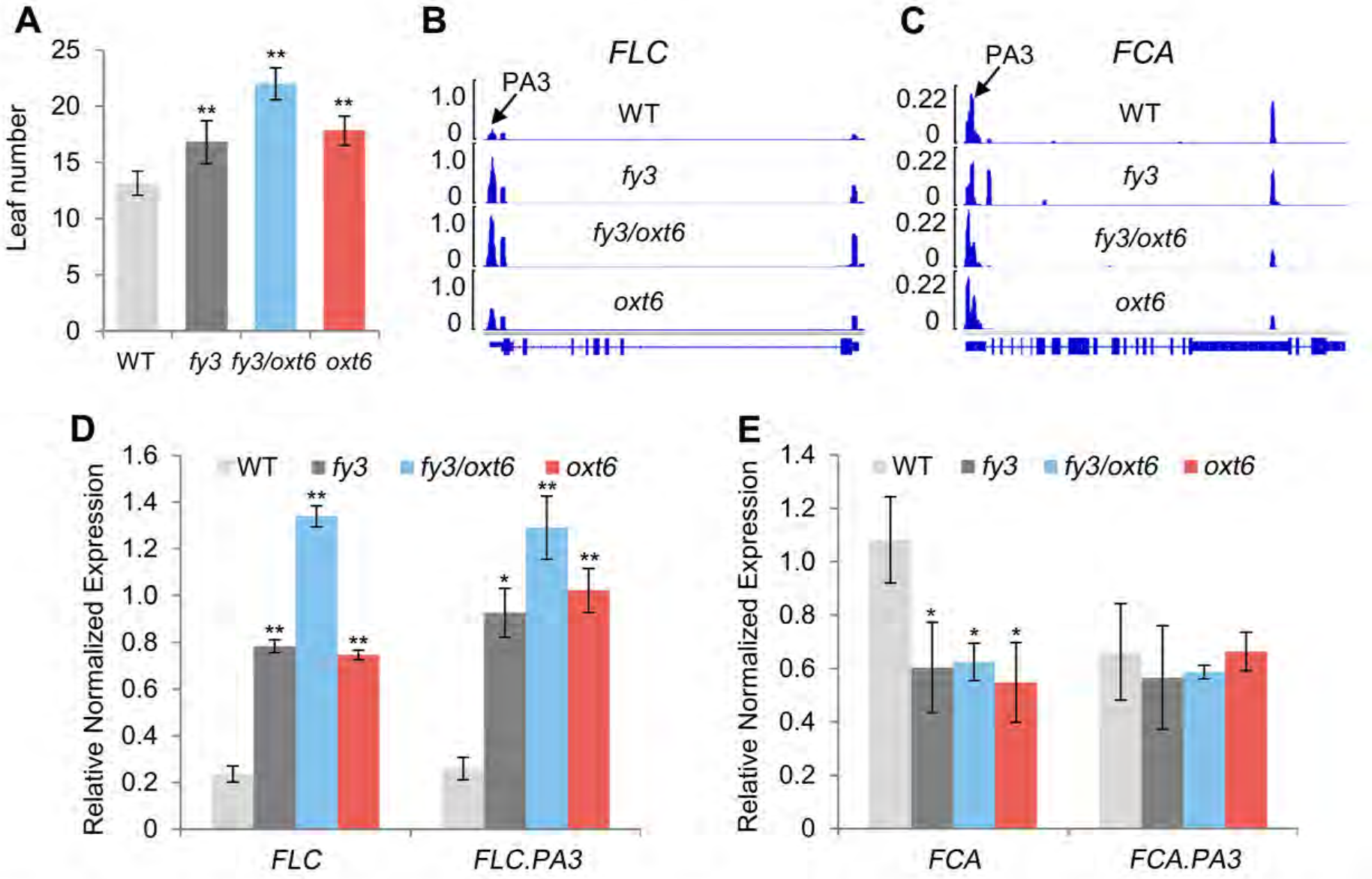

Figure 11. Analysis of flowering phenotype and related gene APA. (A) Flowering time was measured by counting the number of rosette leaves at flowering under long-day photoperiods in the incubator. Each pool contained one plant. Each experiment comprised 18 pools and three independent experiments were completed. **(B) and (C)** The sequencing coverage of FLC and FCA were visualized by Integrative Genomics Viewer (IGV) software. **(D) and (E)** Distal transcript and total gene expression of FLC and FCA were verified by qRT-PCR. Error bars represent standard deviation from three biological replicates and asterisks are indicative of statistically significant differences using one-way ANOVA (* indicates *P*-value<0.05. ** indicates *P*-value<0.01).

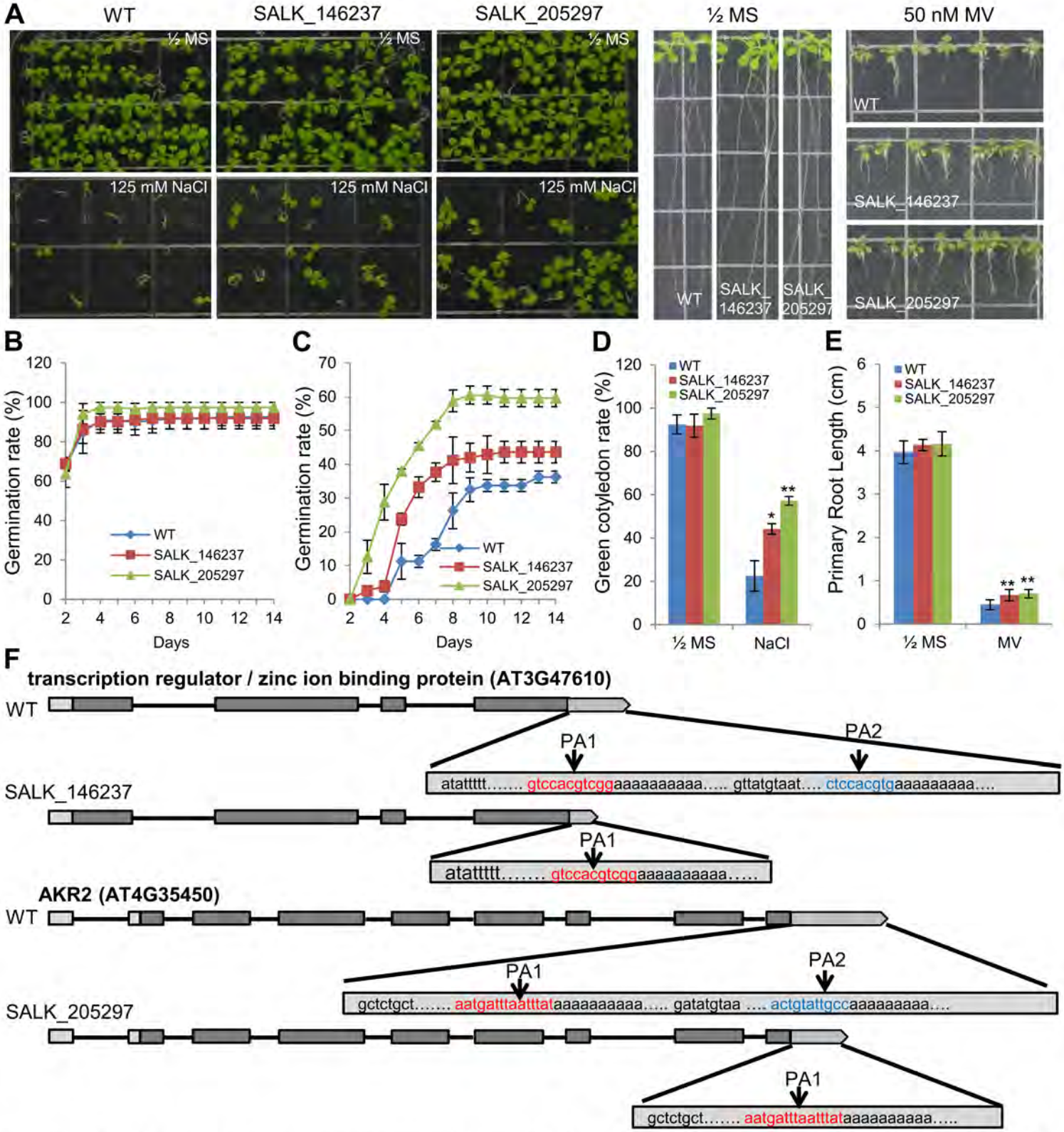

Figure 12. Abiotic stress responses of the SALK_146237 and SALK_205297 and 3'RACE analysis. (A) Photographs of seedlings grown on ½ MS medium or ½ MS medium containing 125 mM NaCl at day 14 after the end of stratification and grown on ½ MS or containing 50 nM MV at day 12. **(B-C)** Seed germination rates of the indicated genotypes grown on ½ MS medium or ½ MS medium containing 125 mM NaCl were quantified every day from the 2nd day to the 14th day after sowing. Three independent experiments were conducted. 40 seeds per genotype were measured in each replicate. Values are mean ± SD of three replications. **(D)** Cotyledon-greening percentages of the 14th day were recorded. **(E)** The root length was measured by ImageJ software. Error bars represent standard deviation from three biological replicates and asterisks are indicative of statistically significant differences using one-way ANOVA (* indicates *P*-value<0.05. ** indicates *P*-value<0.01). **(F)** Illustration of gene constructs and 3'RACE experiment results of AT3G47610 and *AKR2*. The red and blue region represent PA1 and PA2 region, respectively.