XAP5 CIRCADIAN TIMEKEEPER regulates RNA splicing and the circadian clock via genetically separable pathways

Authors:

Hongtao Zhang (张弘韬)[1], Roderick W. Kumimoto[1], Shajahan Anver[1,2], Stacey L. Harmer[1*]

Author affiliations:

[1]Department of Plant Biology, College of Biological Sciences, University of California, Davis, CA 95616, USA

[2]Department of Genetics, Evolution and Environment, University College London, London, UK

[*]**Corresponding author:**

slharmer@ucdavis.edu

Running title:

Separable roles for *XCT* in splicing and the clock

One sentence summary:

XCT plays genetically separable roles in controlling the fidelity of 3' splice site selection during pre-mRNA splicing and the pace of circadian clock in Arabidopsis

Abstract

The circadian oscillator allows organisms to synchronize their cellular and physiological activities with diurnal environmental changes. In plants, the circadian clock is primarily composed of multiple transcriptional-translational feedback loops. Regulators of post-transcriptional events, such as pre-mRNA splicing factors, are also involved in controlling the pace of the clock. However, in most cases the underlying mechanisms remain unclear. We have previously identified *XAP5 CIRCADIAN TIMEKEEPER* (*XCT*) as an *Arabidopsis thaliana* circadian clock regulator with uncharacterized molecular functions. Here, we report that XCT physically interacts with components of the spliceosome, including members of the Nineteen Complex (NTC). PacBio Iso-Seq data show that *xct* mutants have transcriptome-wide pre-mRNA splicing defects, predominantly aberrant 3' splice site selection. Expression of a genomic copy of *XCT* fully rescues those splicing defects, demonstrating that functional *XCT* is important for splicing. Dawn-expressed genes are significantly enriched among those aberrantly spliced in *xct* mutants, suggesting that the splicing activity of *XCT* may be circadian regulated. Furthermore, we show that loss of function mutations in *PRP19A* or *PRP19B*, two homologous core NTC components, suppress the short circadian period phenotype of *xct-2*. Interestingly, we do not see rescue of the splicing defects of core clock genes in *prp19 xct* mutants. Similarly, we find that transgenic expression of the fission yeast ortholog of *XCT* in Arabidopsis shortens the circadian period without affecting splicing. Taken together, our results suggest that *XCT* may regulate splicing and the clock function through genetically separable pathways.

Introduction

Most eukaryotes have evolved an endogenous timekeeper known as the circadian clock, which allows them to anticipate the daily fluctuating environmental conditions caused by the earth's rotation (Harmer, 2009). Although the central oscillators of circadian clocks in diverse eukaryotes lack conserved individual components, they share similar general architectures (Nohales and Kay, 2016). In plants, the approximately 24-h periodicity of the clock is maintained by a complex gene regulatory network consisting primarily of repressors and activators of transcription. Those regulators, often referred to as core circadian clock genes, regulate each other's expression and the expression of thousands of output genes (Creux and Harmer, 2019).

Additionally, post-transcriptional and post-translational mechanisms such as alternative splicing of precursor messenger RNAs (pre-mRNAs) provide critical regulation of clock function. It has been suggested that changes in splicing of core circadian clock genes are important for Arabidopsis in response to environmental stresses (James et al., 2012; Kwon et al., 2014). For example, cold temperatures significantly suppress production of a splice variant of *CIRCADIAN CLOCK-ASSOCIATED1* (*CCA1*) in which intron four is retained. This incompletely spliced isoform encodes a truncated protein which competitively inhibits the function of fully spliced *CCA1* (Seo et al., 2012). However, in most cases the effects of alternative splicing of pre-mRNAs on circadian clock function remain unclear.

The mechanisms underlying pre-mRNA splicing are increasingly well understood (Wilkinson et al., 2020). There are two catalytic transesterification steps, which sequentially remove first the 5' and then the 3' splice sites (5'SS and 3'SS) of introns from their adjacent exons (Shi, 2017). This process is carried out by five small nuclear ribonucleoproteins (snRNPs) and hundreds of non-snRNP splicing factor proteins which assemble on a pre-mRNA to form the spliceosome complex (Wilkinson et al., 2020). One of these spliceosome complex components is the Nineteen Complex (NTC, also known as PRP19 complex), named after PRECURSOR RNA PROCESSING 19 (PRP19), a U-box E3 ubiquitin ligase that forms the core of the NTC (Hatakeyama et al., 2001; Koncz et al., 2012). The NTC is associated with the spliceosome during the two transesterification steps and helps facilitate conformational rearrangements and promote splicing fidelity (Fig. 1; Hogg et al., 2010). The NTC is highly conserved across eukaryotes. In Arabidopsis, multiple orthologs of yeast NTC proteins including two PRP19 homologs (PRP19A/MAC3A and PRP19B/MAC3B) have been shown to physically interact with the spliceosome (Monaghan et al., 2009; Deng et al., 2016). Plants mutant for *prp19a prp19b* or homologs of other NTC-associated proteins such as *pleiotropic regulatory locus1* (*prl1*) and *snw/ski-interacting protein* (*skip*) have genome-wide intron retention defects (Jia et al., 2017; Wang et al., 2012; Li et al., 2019). These data indicate that NTC components act as evolutionarily conserved splicing factors in Arabidopsis.

Mutation of splicing factors can lead to disruption of circadian clock function (Shakhmantsir and Sehgal, 2019). For example, loss-of-function alleles of NTC components, including *PRP19*, *PRL1* and *SKIP*, cause lengthening of circadian period (Wang et al., 2012; Feke et al., 2019; Li et al., 2019). Aberrantly spliced mRNA variants of core circadian clock genes have been detected in these splicing factor mutants (Sanchez et al., 2010; Wang et al., 2012; Jones et al., 2012; Schlaen et al., 2015; Marshall et al., 2016; Li et al., 2019; Feke et al., 2019), suggesting that changes in the pace of the clock might be caused by aberrant splicing of core clock genes. In some cases, epistasis analysis suggests that this may be true (Marshall et al., 2016; Schlaen et al., 2015). However, in other cases, genetic analysis has either not been performed or has revealed additive interactions between the splicing factor and clock gene mutants. Besides, the levels of mis-spliced mRNA variants of clock genes are usually only a small fraction of total transcripts (Jones et al., 2012; Perez-Santángelo et al., 2014; Feke et al., 2019; Sanchez et al., 2010; Wang et al., 2012). Thus, it is unclear whether splicing factors affect clock function solely by controlling the splicing of core clock genes.

RNA splicing factors often function in multiple biological pathways: plants deficient for components of the NTC have defects in immunity, microRNA biogenesis, DNA damage response and transcriptional elongation (Koncz et al., 2012; Monaghan et al., 2009; Zhang et al., 2013; Jia et al., 2017). Some splicing factors are known to carry out roles in nuclear processes biochemically separable from their roles in splicing. In addition to its structural role in the spliceosome, PRP19 also senses DNA damage and promotes DNA repair via its ubiquitin ligase activity (Maréchal et al., 2014). Additionally, SKIP acts in two distinct complexes to regulate splicing and the transcription of genes involved in flowering time control (Li et al., 2019). Nonetheless, whether splicing factors might control circadian clock function via splicing-independent activities has not been investigated.

We previously identified *XAP5 CIRCADIAN TIMEKEEPER* (*XCT*) as a novel regulator of the Arabidopsis circadian clock (Martin-Tryon and Harmer, 2008). Like NTC components, XCT is also well-conserved across eukaryotes. Homologs of XCT (also known as XAP5 proteins) share a highly conserved C-terminal protein domain and are nuclear-localized (Martin-Tryon and Harmer, 2008; Anver et al., 2014; Li et al., 2018; Lee et al., 2020). Previously, we found that

transgenic expression of Arabidopsis *XCT* fully rescued the slow-growth phenotype of fission yeast mutant for *xap5* (Anver et al., 2014). Taken together, these data suggest that XCT homologs might share similar molecular and cellular functions across kingdoms.

It has been reported that FAM50A, one of the two XCT orthologs in humans, physically associates with the spliceosomal C complex and its mutants have defects in RNA splicing (Bessonov et al., 2008; Lee et al., 2020). Similarly, a recent study in Arabidopsis also reported the association of XCT with the spliceosome (Liu et al., 2022), implying that XAP5 proteins may be evolutionarily conserved splicing factors. However, evidence suggests that XAP5 proteins may also participate in fundamental biological processes other than splicing. Fission yeast and *Chlamydomonas* XAP5 proteins associate with chromatin and directly regulate transcription (Anver et al., 2014; Li et al., 2018). In addition to its role in the clock, Arabidopsis *XCT* has been implicated in diverse processes including small RNA production, immune signaling, and DNA damage responses (Fang et al., 2015; Xu et al., 2017; Kumimoto et al., 2021). Notably, NTC components have also been reported to function in all these pathways (Jia et al., 2017; Maréchal et al., 2014; Chanarat and Sträßer, 2013). Although pleiotropic defects are seen in *xct* mutants, interconnections between these phenotypes and the molecular function of *XCT* have not been extensively studied.

In the current work, we report that XCT physically interacts with NTC and other spliceosome-associated proteins in Arabidopsis. We use long-read RNA sequencing to reveal that *XCT* controls the fidelity of 3' splice site selection for hundreds of pre-mRNA splicing events, probably by rejecting downstream suboptimal 3' splice sites. Intriguingly, circadian-regulated genes that are also aberrantly spliced in *xct* mutants are significantly enriched for peak expression at subjective dawn. This implies that the splicing-related activity of *XCT* may be circadian regulated. We also demonstrate that *PRP19A* and *PRP19B* display epistatic genetic interactions with *XCT* in control of the circadian clock period length but not of the splicing of core clock genes. Finally, our data show that expression of *SpXAP5*, the fission yeast ortholog of *XCT*, shortens the circadian period without impacting splicing of clock genes. Collectively, our results suggest that XCT works close to NTC and may regulate pre-mRNA splicing and the circadian clock function via distinct pathways.

Results

XCT physically interacts with the Nineteen Complex and other spliceosome-associated proteins

To uncover the molecular functions of XCT, we carried out mass spectrometry (MS) experiments to identify its possible interactors. We expressed epitope-tagged XCT protein under the control of the native *XCT* promoter in *xct-2* mutant background (*xct-2 XCT*). This transgene largely rescued the morphological and circadian clock defects of *xct-2* (Supplemental Fig. 1). Next, we affinity purified tagged XCT from plant extracts and analyzed co-purifying proteins via MS. In total, 26 proteins were significantly enriched in the XCT immunoprecipitation compared to control immunoprecipitations (Welch's two sample *t*-test, $P < 0.05$; Supplemental Dataset 1). Among those XCT-associated proteins, 15 (57.7%) were annotated as mRNA splicing-related, including 11 core NTC or NTC-associated proteins (Table 1), consistent with the MS and co-immunoprecipitation data from a recent report (Liu et al., 2022). Notably, studies in human and yeast have revealed that NTC is physically associated with the spliceosomal complex throughout the two catalytic transesterification steps (Fig. 1; Hogg et al., 2010). Taken together, those results imply that *XCT* may act close to NTC and function in the catalytic steps during pre-mRNA splicing. Interestingly, we also observed 8 (30.8%) chloroplast proteins enriched in the XCT immunoprecipitation (Supplemental Dataset 1), which may be related to the delayed leaf greening phenotypes observed in *xct-2* mutant plants (Martin-Tryon and Harmer, 2008).

XCT is required for the fidelity of 3' splice site selection during pre-mRNA splicing

To investigate a possible role for *XCT* in RNA splicing, we performed PacBio Isoform Sequencing (Iso-Seq) on wild-type Col-0, reduction-of-function allele *xct-1*, null allele *xct-2*, and the complemented line *xct-2 XCT*. Additionally, we also sequenced *prl1-2*, which contains a T-DNA insertion mutation *in PRL1*, a NTC member that has been demonstrated to control genome-wide splicing efficiency (Jia et al., 2017). Since both transcript levels and RNA splicing of a large proportion of the Arabidopsis transcriptome are circadian regulated (Romanowski et al., 2020; Yang et al., 2020), the time of day at which plants are harvested has significant effects on

gene expression and splicing analysis (Hsu and Harmer, 2012). Therefore, to minimize any differences in subjective time of day between wild-type plants, the long-circadian-period *prl1-2*, and the short-period *xct* mutants (Supplemental Fig. 2), we grew plants in constant environmental conditions for ten days and then pooled seedlings harvested at 2-hour intervals across a twenty-four hour period (Fig. 2A). For each genotype, we sequenced three biological replicates and acquired an average of 784,832 full-length transcript reads per genotype per replicate (Supplemental Dataset 2). With no alignment needed, those reads were directly mapped to the TAIR 10 Arabidopsis reference genome and then subjected to differential splicing analysis using the JunctionSeq R package (Hartley and Mullikin, 2016).

Previous studies have demonstrated that intron retention is the most prevalent type of alternative splicing event in Arabidopsis (Wang and Brendel, 2006; Filichkin et al., 2010). Indeed, plants mutant for multiple NTC components have been reported to have global intron retention defects (Jia et al., 2017; Meng et al., 2022). In our analysis of *prl1-2*, we found 5730 out of 44,496 detected splicing events were significantly differentially enriched from Col-0 (Supplemental Dataset 3). Notably, 99% of these events were decreases in known junctions (Supplemental Fig. 3), consistent with previous studies suggesting that loss of *PRL1* function mainly causes intron retention without affecting splice site selection (Jia et al., 2017).

We next examined transcript composition in *xct* mutants. In the null allele *xct-2*, we detected 75,073 splicing events, of which 875 were significantly down-regulated and 532 were up-regulated compared with Col-0 (FDR<0.05, Fig. 2B-C). Meanwhile, 209 and 337 splicing events were significantly decreased or increased, respectively, in the partial loss-of-function mutant *xct-1*. Comparing mis-regulated splicing between the two *xct* mutant alleles, 196 (93.8%) of the down-regulated and 329 (97.6%) of the up-regulated events were shared. Thus, the two *xct* alleles have similar splicing phenotypes but the severity of the defects depends on the degree of XCT function lost. Next, we analyzed the splicing events in the complemented line *xct-2 XCT*. Only 16 out of 84,965 events were significantly differentially enriched compared with Col-0, indicating that the splicing defects of *xct-2* were almost completely rescued by the restored *XCT* expression. Taken together, these results demonstrate that *XCT* is important for transcriptome-wide pre-mRNA splicing.

To further characterize the major types of splicing defects caused by loss of *XCT* function, we analyzed the splice sites for the differentially spliced junctions in *xct* mutants. Specifically, we categorized *xct*-induced mis-splicing events into four classes based on whether the 5' and 3' splice sites used were previously annotated (known) or not (novel). As expected, most down-regulated splicing events displayed decreases in usage of junctions with known 5' and 3' splice sites (Fig. 2D). Interestingly, 307 (90.8%) in *xct-1* and 462 (86.5%) in *xct-2* of up-regulated splicing events involved increased usage of junctions with a known 5' splice site but a novel 3' splice site (Fig. 2E). The decreases in abundance of junctions with known splice sites hence reflects both intron retention events and novel 3' splice sites usage. Therefore, our results demonstrate that in addition to controlling the efficient removal of introns, *XCT* is also responsible for the fidelity of 3' splice site selection.

To investigate how *XCT* contributes to the fidelity of 3' splice site selection, we compared the significantly up- or down-regulated 3' splice sites in *xct* mutants with all detected 3' splice sites in Col-0. Intriguingly, most of the up-regulated novel 3' splice sites in *xct-1* and *xct-2* were located less than 50 nucleotides downstream of the wild-type 3' splice sites (Fig. 3B; Supplemental Fig. 4A). Studies in human and yeast have demonstrated that the sequences preceding a 3' splice site, including the polypyrimidine tract (PPT) and the pyrimidine at the -3 position (Fig. 3A), are important for the strength of the 3' splice site (Horowitz, 2012). Therefore, we further examined the 23-mer sequences flanking the 3' splice sites (from -20 to +3 position). We found that the frequency of the canonical AG sequence at the 3' splice sites was not altered in either *xct-1* or *xct-2* (Fig. 3C; Supplemental Fig. 4B-C). However, the average predicted strength of the 875 down-regulated 3' splice sites in *xct-2* was significantly weaker than the average of total detected 3' splice sites in Col-0 (Fig. 3D). Specifically, the percentage of pyrimidine residues at the -3 position was significantly reduced in down-regulated 3' splice sites compared to that found in all detected 3' splice sites (Fisher's exact test, $P < 0.001$, Fig. 3E), suggesting that functional *XCT* is important for the removal of 3' splice sites with a suboptimal sequence at the -3 position. Moreover, the 532 up-regulated 3' splice sites had an even lower 3' splice site score than the down-regulated sites (Fig. 3D). The frequency of pyrimidines both at the -3 position and throughout the PPT region was significantly lower in the

up-regulated junctions (Fig. 3E-F), showing that *xct-2* is less able to discriminate between 3' splice sites during splicing. Taken together, our Iso-Seq data demonstrates that *XCT* controls the accuracy of 3' splice site selection, possibly by helping to recognize weaker 3' splice sites and reject sub-optimal downstream sites.

Genes expressed at subjective dawn are enriched among those aberrantly spliced in *xct* mutants

To explore the possible link between the circadian clock and splicing phenotypes in *xct* mutants, we looked for enrichment of circadian clock regulation within genes aberrantly spliced in *xct.* Among the 5602 previously reported clock-regulated genes (Romanowski et al., 2020) that are detected in our Iso-Seq data, we identified 182 and 402 genes that are mis-spliced in *xct-1* and *xct-2*, respectively (Fig. 4A; Supplemental Dataset 4). However, there is no significant over-representation in either mutant (one-tailed Fisher's exact test, $P = 0.87$ for *xct-1* and 0.99 for *xct-2*), suggesting that *XCT* does not preferentially affect the splicing of clock-regulated genes. Previous studies have identified transcripts that are differentially spliced at various time of a day (Romanowski et al., 2020; Yang et al., 2020). This suggest that the splicing activity may be circadian clock regulated. Therefore, to test whether the circadian regulator *XCT* preferentially affects pre-mRNA splicing at certain time of a day, we examined the distribution of estimated peaked expression times of genes aberrantly spliced in *xct* mutants. We grouped all 5602 clock-regulated genes into twelve 2-hour intervals based on their estimated peaked expression time (Fig. 4B). Genes that are aberrantly spliced in *xct-2* are significantly enriched for peaked expression between ZT22 and ZT2 compared with the other intervals of the day (Fig. 4C-D, Fisher's exact test, $P < 1e\text{-}4$). Similarly, clock-regulated genes that are aberrantly spliced in *xct-1* also showed a dawn-enriched expression pattern (Supplemental Fig. 5). Collectively, those results indicate that *XCT* activity within the spliceosome may be regulated by the circadian clock.

Alternative splicing of core clock genes is important for Arabidopsis responses to environmental stresses such as shifts in ambient temperature (James et al., 2012; Kwon et al., 2014). Recent studies showed that aberrant splicing of core clock genes might fully or partially contribute to the

altered circadian clock period phenotype in some splicing factor mutants (Sanchez et al., 2010; Wang et al., 2012; Jones et al., 2012; Schlaen et al., 2015; Marshall et al., 2016; Li et al., 2019; Feke et al., 2019). Similarly, in the *xct-1* and *xct-2* mutants, we detected shared aberrant splicing events in five core circadian clock genes, including *LHY*, *LNK2*, *PRR7*, *TOC1* and *TIC* (Fig. 4A; Fig. 5A-D), one or more of which might potentially cause the short clock period phenotype. Four of these genes are transcriptionally regulated by the circadian clock, with peak *LHY* and *LNK2* expression at dawn and peak *PRR7* and *TOC1* expression late in the afternoon (Creux and Harmer, 2019). Thus, to further investigate whether *XCT* preferentially affects splicing of transcripts produced at dawn, we monitored the abundance of the splice variants of these four circadian-clock-regulated core clock genes together with one non-clock-regulated gene via quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) over a 24-hour light/dark period (LD) followed by 48 hours in constant light (LL).

As we expected, total transcript levels of clock-regulated genes displayed rhythmic abundance with anticipated peak phases of expression (Supplemental Fig. 6). Notably, both the peak and average abundance of the aberrantly spliced isoforms were significantly different in *xct-2* compared with Col-0 (Fig. 5F-J). This serves as validation of the pooling strategy we used to generate our Iso-Seq libraries, where the daily abundance of transcripts was averaged (Fig. 2A; Fig. 5A-E). However, instead of preferentially accumulating at dawn, abundance of all splice variants in *xct-2* was synchronized with total transcript levels (Fig. 5F-I; Supplemental Fig. 6). Likewise, the mis-spliced isoform of the non-clock-regulated gene *SPPA* did not show a significant morning peak (Fig. 5J). Meanwhile, the genomic *XCT* complemented line almost completely rescued the splicing defects across the whole experimental period. Taken together, those results suggest that although morning-expressed genes are preferentially enriched among *XCT* targets, *XCT* is important for the accuracy of pre-mRNA splicing throughout the day under both LD and free-running LL conditions.

Reduction of *PRP19* function rescues the circadian clock but not the splicing defects in *xct-2*

Previous studies revealed that NTC components participate in regulation of circadian clock function in Arabidopsis (Feke et al., 2019; Li et al., 2019). Here we examined circadian period in mutants for several NTC components. We found that mutation in either of the two Arabidopsis *PRP19* homologs, *PRP19A* (*MAC3A*) or *PRP19B* (*MAC3B*), caused minor lengthening of circadian period. Moreover, *prp19a-1 prp19b-1* double mutants had a significantly slower circadian clock than Col-0 (Supplemental Fig. 7A), consistent with previous reports (Feke et al., 2019). Similarly, loss of function of other NTC members, including *PRL1*, *CDC5* and *SKIP*, also lengthened the clock period by 1 to 3 hours (Supplemental Fig. 2B; Supplemental Fig. 7B-C), similar to previous reports for *prl1-9* and *skip-1* mutants (Li et al., 2019).

Since *PRP19* co-purifies with *XCT* (Table 1) and they both control the pace of the circadian clock (Supplemental Fig. 7A), we hypothesized they might function in the same pathway to regulate the clock. To test this hypothesis, we introduced *prp19a-1* and *prp19b-1* mutant alleles, which express greatly decreased and near-null levels of *PRP19A* and *PRP19B*, respectively (Supplemental Fig. 8), into the *xct-2* background and assayed their circadian clock phenotypes. Interestingly, we found that the circadian periods of both *prp19a-1 xct-2* and *prp19b-1 xct-2* double mutants were not significantly different from Col-0 (Fig. 6A; Supplemental Fig. 9), indicating that *prp19a-1* and *prp19b-1* both fully suppress the short period phenotype of *xct-2*. Thus, our data suggest that functional *PRP19A* and *PRP19B* are both necessary for *XCT* in regulation of circadian period.

To further investigate the functional association between *XCT* and *PRP19*, we next asked whether *prp19* mutants suppress other developmental and physiological defects observed in *xct-2*. In contrast to the circadian period, the reduced rosette size in *xct-2* was not significantly restored in either *prp19a-1 xct-2* or *prp19b-1 xct-2* (Supplemental Fig. 10), suggesting that the regulation of *XCT* on rosette development and the circadian clock period are genetically separable. This is consistent with our observation that *xct-1* and *xct-2* have similar short circadian period phenotypes but only *xct-2* is morphologically different from Col-0 (Supplemental Fig. 1).

Next, to determine whether suppression of the short clock period phenotype in *xct-2 prp19* mutants is due to reversal of the *xct-2* splicing defects, we conducted qRT-PCR experiments to examine the abundance of *xct*-induced aberrantly-spliced mRNA isoforms of core clock genes in *prp19a-1 xct-2* and *prp19b-1 xct-2*. Surprisingly, none of the aberrantly-spliced isoforms tested showed significantly different abundance in *prp19a-1 xct-2* and *prp19b-1 xct-2* relative to *xct-2* (Fig. 6B-F). We also checked the expression levels of fully spliced or total mRNA isoforms of the five core clock genes that were mis-spliced in *xct-2*. In all cases, addition of *prp19a-1* or *prp19b-1* to *xct-2* did not alter the abundance of functionally spliced isoforms of the clock genes (Supplemental Fig. 11). Thus, similar to the developmental defects, the splicing defects of core clock genes are genetically separable from the circadian clock phenotype of *xct-2*.

Fission yeast *XCT* ortholog *SpXAP5* accelerates the Arabidopsis circadian clock without affecting splicing

XAP5 proteins are highly conserved across eukaryotes (Martin-Tryon and Harmer, 2008). Previous studies showed that transgenic expression of Arabidopsis *XCT* in the fission yeast *Schizosaccharomyces pombe xap5* mutant can fully rescue its slow-growing phenotype (Anver et al., 2014), indicating that there are shared molecular functions between these two homologous genes. Therefore, to further investigate the molecular function of *XCT* that regulates the clock in Arabidopsis, we examined whether a *SpXAP5* transgene could complement the short circadian period phenotype of Arabidopsis *xct* mutants. To our surprise, we found that expression of *SpXAP5* under the control of the native Arabidopsis *XCT* promoter not only failed to rescue the short circadian period phenotype in *xct-1* mutant, but actually further accelerated the clock in both Col-0 and *xct-1* backgrounds (Fig. 7A; Supplemental Fig. 12), suggesting a gain-of-function effect of *SpXAP5* on controlling the pace of the Arabidopsis circadian clock.

To determine whether *SpXAP5* accelerates the Arabidopsis clock by exacerbating the splicing defects of *xct* mutants, we performed PacBio Iso-Seq to examine changes in the transcriptome-wide splicing pattern caused by expressing the *SpXAP5* transgene in the *xct-1* background. Overall, only 3 out of the 90,114 detected splicing events were significantly, albeit only slightly, enriched in *xct-1 SpXAP5* compared to *xct-1* (FDR<0.05, Fig. 7B). Furthermore, we conducted

qRT-PCR experiment to determine whether *SpXAP5* alters the splicing patterns of the core clock genes targeted by *XCT*. Consistent with the Iso-Seq results (Fig. 5A-E), ratios of aberrantly spliced isoforms to fully spliced isoforms in *xct-1* and *xct-2* were significantly different from Col-0, whereas those ratios in *SpXAP5* and *xct-1 SpXAP5* showed no significant difference compared to Col-0 and *xct-1*, respectively (Fig. 7C-F). Taken together, these results indicate that the *SpXAP5* transgene shortens the Arabidopsis circadian clock period length in a gain-of-function manner, but not by affecting splicing of the same set of core clock genes targeted by *XCT*.

Discussion

The accurate removal of introns from pre-mRNAs is an essential step of gene expression in all eukaryotes. In this work, we report that *XCT* is a global regulator of pre-mRNA splicing in Arabidopsis. Multiple lines of evidence suggest that orthologs of XCT share conserved molecular and cellular functions across kingdoms. Previous mass spectrometry data showed that FAM50A and FAM50B, two human homologs of XCT, physically associate with affinity-purified spliceosomal C complex (Bessonov et al., 2008; Bessonov et al., 2010). Loss of *FAM50A* function induces transcriptome-wide pre-mRNA splicing defects in both human and zebrafish (Lee et al., 2020). Similarly, a recent study in Arabidopsis also found that XCT associates with spliceosomal proteins and regulates splicing (Liu et al., 2022). In this paper, we show that XCT physically co-purifies with PRP19 and other NTC-associated proteins (Table 1). Furthermore, our PacBio Iso-Seq data indicate that the efficiency and fidelity of pre-mRNA splicing are negatively impacted in both the partial loss-of-function mutant *xct-1* and the null mutant *xct-2* but are largely restored in *xct-2 XCT* (Fig. 2). Therefore, XCT orthologs likely play evolutionarily conserved roles in pre-mRNA splicing. This is supported by our prior finding that Arabidopsis *XCT* rescues the growth defects in yeast *xap5* mutant (Anver et al., 2014). However, we now report that expression of fission yeast *SpXAP5* fails to rescue splicing defects in *xct-1* mutant (Fig. 7B-F), indicating that any conserved biochemical functions between yeast SpXAP5 and XCT are not sufficient for appropriate splicing regulation in Arabidopsis. It would be interesting to apply structural biology approaches to further investigate what conserved functional domains among XCT homologs contribute to their splicing activities.

Studies have shown that intron retention is the most prevalent type of alternative splicing event in Arabidopsis (Wang and Brendel, 2006; Filichkin et al., 2010). Correspondingly, mutations in many Arabidopsis splicing factors mainly cause intron retention (Supplemental Fig. 3; Schlaen et al., 2015; Li et al., 2019). Here we identify *XCT* as an unusual splicing regulator that specifically controls the fidelity of 3' splice site selection (Fig. 2D-E; Fig. 3). Biochemical studies of splicing in yeast revealed that several DEAH-box ATPases are important for the selection of 3' splice sites (Horowitz, 2012). For example, PRP22, a DEAH-box ATPases that co-purifies with XCT (Table 1; Liu et al., 2022), represses usage of aberrant 3' splice sites and promotes spliceosome scanning for downstream alternative 3' splice sites in yeast (Mayas et al., 2006; Semlow et al., 2016). Studies in human showed that hPRP22, FAM50A and FAM50B, are all abundant in the spliceosomal C complex but nearly absent in B and B* complexes (Bessonov et al., 2008; Bessonov et al., 2010; Zhan et al., 2022). This implies that PRP22 and XCT may have evolutionarily conserved interactions during the second transesterification reaction of splicing, when 5' and 3' splice sites are joined (Fig. 1). Intriguingly, we detected decreased fidelity of 3' splice site selection in both *xct-1* and *xct-2* mutants (Fig. 3; Supplementary Fig. 4), which is also observed in yeast *prp22* mutants (Semlow et al., 2016). Therefore, our results imply that XCT may work with PRP22 to control the fidelity of 3' splice site selection. Future research on the biochemical functions of XCT using *in vitro* and *in vivo* systems could help reveal how it controls splicing fidelity.

Previous microarray and RNAseq studies demonstrated that transcription and splicing of a large proportion of Arabidopsis genes are under circadian clock regulation (Covington and Harmer, 2007; Hsu and Harmer, 2012; Romanowski et al., 2020; Yang et al., 2020). Consequently, timing of sample collection is an important consideration in gene expression and splicing analysis, especially when comparing genotypes with different circadian periods. Here we employed a more efficient sampling method by harvesting plants at twelve evenly distributed time points of a day and pooling them before analysis. Our pooling strategy enabled us to detect transcripts of genes with the distribution of peak phases of expression mirroring that of all clock-regulated genes (Fig. 4B; Romanowski et al., 2020). Overall, this strategy allowed us to identify more clock-regulated differentially spliced transcripts in *xct-2* mutants than in a recent study in which plants were collected at a single time point (Liu et al., 2022). For example, we report novel

differential splicing events in afternoon-peaked genes including *PRR7* and *TOC1* (Fig. 4A; Supplemental Fig. 6C-D). In fact, our time-course qPCR data suggest that abundance of the aberrantly spliced isoforms of these clock-regulated genes fluctuates by over 99% depending on time of sample collection (Fig. 5F-I). Thus, our results demonstrate the advantages of pooling samples when conducting transcriptome-wide splicing analysis of clock-regulated genes.

Using this strategy, we were able to investigate whether *XCT* preferentially affects splicing at certain time of the day. Indeed, we found that dawn-expressed genes are significantly more likely to be mis-spliced in *xct* mutants than those with other peak phases (Fig. 4C-D; Supplemental Fig. 5). This suggests that *XCT* activity in or its association with the spliceosome may be clock-regulated. An alternative explanation for the overrepresented mis-splicing of dawn-expressed genes could be that *cis*-regulatory elements of those genes might preferentially recruit XCT to control their splicing. This possibility is supported by previous studies showing that XCT orthologs in fission yeast and *Chlamydomonas* are chromatin-associated and in the latter case recruit RNA polymerase II (Pol II) to promoter regions (Anver et al., 2014; Li et al., 2018).

Alternative splicing regulates various biological processes, including the function of the circadian clock (Hsu and Harmer, 2014; Nolte and Staiger, 2015). In Arabidopsis, such regulation is supported by identification of splicing factor mutants that alter circadian clock period length (Shakhmantsir and Sehgal, 2019). Some splicing factors, such as *GEMIN2* and *SICKLE*, interact epistatically with one or more alternatively spliced clock genes in the control of period length (Schlaen et al., 2015; Marshall et al., 2016), suggesting changes in the pace of the clock are due to altered splicing of these clock genes. Whereas in other cases, the mis-regulated circadian period can only be partially attributed to changes in levels of splicing variants of clock genes (Sanchez et al., 2010; Wang et al., 2012). Additionally, in most splicing factor mutants, only small fractions of total transcripts are aberrantly processed (Jones et al., 2012; Perez-Santángelo et al., 2014; Feke et al., 2019; Li et al., 2019). Therefore, whether these small changes in levels of aberrantly spliced isoforms could lead to significantly decreased levels of functional proteins and thereby cause the observed circadian phenotypes remains unclear.

Here, we demonstrate that in *xct* mutants the circadian clock phenotypes are genetically separable from altered levels of aberrantly spliced mRNA variants of core clock gene. We found that loss of *XCT* function accelerates the clock and causes aberrant splicing of five core clock genes (Supplemental Fig. 1; Fig. 4A). Surprisingly, the clock phenotype but not the splicing defects of the five clock genes in *xct-2* is suppressed by reduction of *PRP19A* or *PRP19B* function (Fig. 6). Similarly, the clock in *xct-1* runs slightly faster than *xct-2* but the splicing defects, including the aberrant splicing of *TOC1* and *TIC*, are more severe in the latter (Fig. 2; Fig. 7D, F). Furthermore, we show that expression of *SpXAP5* in Col-0 or *xct-1* shortens the circadian period without impacting splicing of *XCT*-targeted clock genes (Fig. 7). Collectively, these data demonstrate that the short period phenotypes and altered levels of clock gene splicing variants are genetically separable in *xct* mutants.

How *XCT* regulates the pace of the circadian clock is still an outstanding question. Although we argue that changes in levels of aberrantly spliced clock mRNAs are not responsible for the accelerated clock in *xct* mutants, one possibility is that alterations in the overall kinetics of pre-mRNA splicing may cause circadian period phenotypes. Indeed, pharmacological perturbations of global transcription and translation efficiency can both lengthen the circadian period (de Melo et al., 2021; Uehara et al., 2022). Those results suggest that changes in the kinetics of RNA processing might cause the period phenotypes observed in plants mutated for some splicing factors.

Alternatively, *XCT* may control the circadian clock function independent of its role in pre-mRNA splicing. In Chlamydomonas, the XCT homolog XAP5 co-immunoprecipitates with RNA Pol II (Li et al., 2018), suggesting that XCT orthologs may possess transcriptional regulatory activities. In this paper, we show that NTC components physically and genetically interact with XCT to regulate circadian clock period (Table 1; Fig. 6A). Besides splicing, another well-characterized role of the NTC in gene expression is regulation of transcriptional elongation (Chanarat and Sträßer, 2013). Studies in both yeast and Arabidopsis have revealed that multiple NTC members, including PRP19, PRL1 and CDC5, physically associate with RNA Pol II and participate in transcription elongation (Chanarat et al., 2011; Zhang et al., 2013; Zhang et al., 2014). In addition, the NTC component SKIP interacts with Polymerase-Associated Factor 1

complex to regulate transcription in a splicing-independent manner (Li et al., 2019). Interestingly, a recent study showed that inhibition of transcriptional elongation by decreasing phosphorylation of the RNA Pol II C-terminal domain lengthens circadian period in Arabidopsis (Uehara et al., 2022). It is therefore possible that *XCT* affects circadian period by altering transcriptional elongation.

Yet another possibility is that *XCT* affects a cellular process independently of its role in gene expression. For example, PRP19 is known to facilitate DNA repair by its E3 ubiquitin ligase activity independent of its involvement in the spliceosomal complex (Maréchal et al., 2014; Idrissou and Maréchal, 2022). Indeed, there is increasing evidence that many RNA binding proteins directly participate in DNA double-strand break responses (Klaric et al., 2021). A recent study showed that the NTC-associated protein MOS4-ASSOCIATED COMPLEX SUBUNIT 5A (MAC5A) physically interacts with the 26S proteasome and regulates its activities in response to DNA damage (Meng et al., 2022). Intriguingly, we recently reported that loss of *XCT* function also disturbs the DNA damage response pathway (Kumimoto et al., 2021). Thus, it is possible that *XCT* works with *PRP19* in this pathway separately from its role in RNA processing. Future research is required to understand the nature of the relationship of between XCT and the NTC in the control of circadian clock function and other essential biological processes.

Materials and Methods

Plant materials and growth conditions

All *Arabidopsis thaliana* plants used in this study are Columbia-0 (Col-0) ecotype and contain a *CCR2*::*LUC* reporter for circadian clock assays. The *xct-1*, *xct-2*, *xct-2* p*XCT*::g*XCT-YFP-HA*, *prp19a-1*, *prp19b-1*, *prl1-2*, *cdc5-1* and *skip-1* genotypes have been previously described (Martin-Tryon and Harmer, 2008; Monaghan et al., 2009; Zhang et al., 2014; Wang et al., 2012). All the double mutants in this study were produced by crossing. The fission yeast transgene was generated by cloning *SpXAP5* cDNA sequence driven by an Arabidopsis *XCT* promoter into the pCR8/GW/TOPO gateway entry vector (Invitrogen) and then transferring it into the promoter-less pEarleyGate destination vector pEG301. The construct was transformed into *xct-1* mutant plants using *Agrobacterium*-mediated floral dip method (Clough and Bent, 1998) and introduced

493 into the Col-0 background by crossing. Unless otherwise specified, seeds were surface sterilized
494 with chlorine gas for 3 hours (50 ml 100% bleach + 3ml 1M HCl) and then plated on 1x
495 Murashige and Skoog (MS) growth media containing 0.7% agar, pH 5.7. After 3d stratification
496 in dark at 4˚C, plates were transferred to 12-h light (cool white fluorescent bulbs, 55 µmol m^{-2} s^{-1}) / 12-h dark cycles at 22˚C for a variable number of days depending on the experiment.
497
498
499 **Immunoprecipitations and mass spectrometry**
500
501 Plants were grown in 12-h light / 12-h dark cycles and harvested on day 10 at ZT3 or ZT17.
502 Approximately 7.5g of vegetative tissue was flash frozen in liquid nitrogen and ground into fine
503 powder with a mortar and pestle. Ground tissue was resuspended in nuclei enrichment buffer (50
504 mM Tris pH7.5, 400 mM sucrose, 2.5 mM $MgCl_2$, 1 mM DTT, 1 mM PMSF, cOmplete Protease
505 Inhibitor cocktail (Roche)), the nuclear pellet was collected, resuspended in IP buffer (100mM
506 Tris pH7.5, 1mM EDTA, 75mM NaCl, 10% glycerol, 0.3% Triton X-100, 0.05% SDS, 10 mM
507 MG-132, 1 mM PMSF, , cOmplete Protease Inhibitor cocktail (Roche)) and pelleted again.
508 Nuclei were resuspended in IP buffer, disrupted via sonication, and adjusted to 1 mg/mL in IP
509 buffer. 1 mg of extract was incubated with µMACS MicroBeads conjugated to an anti-HA
510 monoclonal antibody (Miltenyi Biotec) and beads were then captured on µMACS M-colums.
511 Beads were washed 3x with ice cold IP buffer and 1x with ice cold TE buffer (10 mM Tris-HCl
512 pH 8.0, 0.1 mM EDTA). Proteins were eluted off the beads with elution buffer (Miltenyi Biotec).
513 Proteins were resolved using SDS polyacrylamide gel electrophoresis and gels were sent to the
514 Rutgers Biological Mass Spectrometry Facility, Robert Wood Johnson Medical School. Proteins
515 were eluted and subjected to mass spectrometry as previously described (Wei et al., 2020).
516
517 **PacBio Iso-Seq and bioinformatic analysis**
518
519 Arabidopsis seeds were surface sterilized in 70% ethanol prepared in 0.1% (v/v) Triton X-100
520 (Sigma) for 5 minutes and then in 100% Ethanol for 20 minutes. After air drying, seeds were
521 plated on 1x MS growth media containing 0.7% agar. After cold stratification for 3 days, plates
522 were transferred to constant light (55 µmol m^{-2} s^{-1}) and temperature (22˚C). Approximately 60
523 mg of whole seedlings of each genotype were collected every 2h over 24 hours on day 11 and

pooled as indicated in Fig. 2A. Three biological replicates were collected for each genotype. Samples were flash-frozen in liquid nitrogen, and ground into fine powder using a beadbeater. Total RNA was extracted using an RNeasy Plant Mini Kit (Qiagen) followed by purification using the RNA Clean & Concentrator Kit (Zymo Research). Purified RNAs were quantified using a Nanodrop (ThermoFisher) and quality control was performed by an Agilent Bioanalyzer 2100. All samples had a 260/230 nm ratio higher than 1.9, 260/280nm ratio between 2 and 2.15, and RIN score higher than 8. PacBio Sequel II library preparation and RNA-sequencing (Iso-Seq) was performed by UC Davis DNA Technologies & Expression Analysis Core (https://dnatech.genomecenter.ucdavis.edu/pacbio-library-prep-sequencing/).

Raw reads generated by the PacBio Sequel II sequencer were imported into PacBio SMRT Link v 8.0 for Circular Consensus Sequence (CCS) calling and demultiplexing. Next, poly(A) tails and concatemers were removed using the refine command from isoseq3 package via Bioconda (v 3.10) with the option '--require-polya'. Then the fastq files containing full-length non-concatemer reads were mapped to Arabidopsis TAIR 10 genome assembly using minimap2 (v 2.17) with the parameter '-ax splice -t 30 -uf --secondary=no -C5'. For differential splicing analysis, counts of exonic regions and known/novel splice junctions were generated using QoRTs software package (v 1.3.6) (Hartley and Mullikin, 2015) with the parameter '--stranded --singleEnded --stranded_fr_secondstrand --keepMultiMapped'. To adapt to the long-read data, the 'maxPhredScore' was set to 93 and the 'maxReadLength' was adjusted manually for each library. To increase the power of detecting novel splice junctions, the '--minCount' threshold was set to 5. The raw count data was then loaded to JunctionSeq R package (v 1.5.4) (Hartley and Mullikin, 2016) to determine differential usage of exons or splice junctions relative to the overall expression of the corresponding gene with a false discovery rate (FDR) of 0.05. Genes with at least one differentially spliced exon/junction were compared to the total circadian-clock-regulated genes (Romanowski et al., 2020) as described in Fig. 4.

RNA extraction and RT-PCR

For time-course qRT-PCR analysis, Arabidopsis seedlings were entrained in 12-h light / 12-h dark cycles for 9 days before transferred to constant light and temperature (55 µmol m^{-2} s^{-1},

22˚C). Starting from day 9, approximately 15-20 whole seedlings of each genotype were harvested every 4h over the 72-h period. For single time point gene expression and splicing analysis, 10-d-old seedlings grown in 12-h light / 12-h dark cycles at 22˚C were harvested at estimated peak time of expression for each studied gene. Collected tissue was immediately frozen in liquid nitrogen and ground into fine powder using a beadbeater. Total RNA was isolated using TRIzol reagent (Invitrogen) and quantified with a Nanodrop (ThermoFisher). 200 ng of total RNA was used for cDNA synthesis with an oligo(dT)18 primer and SuperScriptIII (Invitrogen) reverse transcriptase. Diluted cDNAs were used as templates for qRT-PCR and semi-qRT-PCR reactions. The qRT-PCR was carried out as previously described (Martin-Tryon et al., 2007) using a Bio-Rad CFX96 thermocycler. Primers amplifying the aberrantly spliced or total transcripts were tested by standard curve and melt curve assays. Relative expression (ΔΔCq) values were normalized to the geometric mean of *PROTEIN PHOSPHATASE 2a* (*PP2A*) and *ISOPENTENYL DIPHOSPHATE ISOMERASE 2* (*IPP2*) expression levels. For semi-qRT-PCR, splice-junction specific primer pairs were designed to amplify regions flanking the aberrantly spliced introns of interest. The size and abundance of the resultant PCR products were then analyzed and quantified by LabChip GX bioanalyzer (PerkinElmer). The expression analyses represent two to three biological replicates. All primers used in this study are described in Supplemental Dataset 5.

Circadian period analysis

After growing on MS plates under 12-h light / 12-h dark cycles for 6 days, Arabidopsis seedlings were sprayed with 3 mM D-luciferin (Biosynth) prepared in 0.01% (v/v) Triton X-100 (Sigma) and then transferred to a growth chamber with a constant 22˚C and constant light provided by red and blue LED SnapLites (Quantum Devices, 35 µmol m^{-2} s^{-1} each) for imaging. Luciferase activity was detected using a cooled CCD camera (DU434-BV [Andor Technology]). Bioluminescence signals from the images were quantified using MetaMorph 7.7.1.0 software (Molecular Devices). Subsequent circadian period estimation and rhythmicity analysis were performed using Biological Rhythms Analysis Software System 3.0 (BRASS) by fitting the bioluminescence data to a cosine wave through Fourier Fast Transform-Non-Linear Least Squares (Plautz et al., 1997).

Rosette size measurement

Seeds on MS plates were germinated in a growth chamber with 12-h light / 12-h dark cycles at a constant 22˚C. 15-day-old seedlings were transferred to soil and grown in 16-h light (cool-white fluorescent bulbs, 75 µmol m^{-2} s^{-1}) / 8-h dark long day condition at a constant 22˚C. Rosette sizes of 39-day-old plants were determined by the greatest distance between rosette leaves using a digital caliper.

Accession Numbers

All the *Arabidopsis thaliana* and *Schizosaccharomyces pombe* genes studied in this paper can be found under the following accession numbers:

XCT, AT2G21150
PRP19A, AT1G04510
PRP19B, AT2G33340
PRL1, AT4G15900
SKIP, AT1G77180
CDC5, AT1G09770
LHY, AT1G01060
LNK2, AT3G54500
TIC, AT3G22380
PRR7, AT5G02810
TOC1, AT5G61380
SPPA, AT1G73990
IPP2, AT3G02780
PP2A, AT1G69960
SpXAP5, NP_587947

Data Availability

All our MS data were deposited and available at Center for Computational Mass Spectrometry (CCMS, Dataset: MSV000090830). Our PacBio Iso-Seq data were deposited and available in NCBI Gene Expression Omnibus (GEO, Accession number: GSE220902).

Fundings

This work was supported by awards from the National Institutes of Health (R01 GM069418 to SLH) and the United States Department of Agriculture-National Institute of Food and Agriculture (CA-D-PLB-2259-H to SLH). HZ was supported by a fellowship from China Scholarship Council (CSC #201806010204 to HZ).

Acknowledgments

We thank the Arabidopsis Biological Resource Center for providing seeds. We thank Haiyan Zheng and the Biological Mass Spectrometry Facility of Robert Wood Johnson Medical School and Rutgers for mass spectrometry analysis. They are supported by NIH Shared Instrumentation Grant S10OD01640. We thank the DNA Technologies and Expression Analysis Core at the UC Davis Genome Center for performing PacBio Iso-Seq experiment. They are supported by NIH Shared Instrumentation Grant 1S10OD010786-01. We thank Julin Maloof for statistical advice, as well as members of the Harmer, Maloof, and Shabek labs for helpful discussions.

Author Contributions

HZ, RWK and SLH designed the research; HZ, RWK, and SA performed the research; HZ, RWK, and SLH analyzed data; HZ and SLH wrote the paper.

Figure Legends

Figure 1. Simplified overview of pre-mRNA splicing reactions. A schematic diagram highlighting the two catalytic transesterification steps and the association between the NTC and the spliceosomal complex during pre-mRNA splicing. Gray boxes and solid black lines represent

exons and introns, respectively. U1-U6 small nuclear ribonucleoproteins (snRNPs) are indicated by yellow circles. The NTC is indicated by a green oval.

Table 1. XCT physically associates with splicing-related proteins, especially the NTC components, in Arabidopsis. Mass Spectrometry (MS) data showing that spliceosome-associated proteins are significantly more enriched in XCT-YFP-HA-IP than control IP (fold enrichment > 3; *P*-value < 0.05, Welch two sample t.test). The full list of proteins co-purified with XCT are described in Supplemental Dataset 1. Only proteins detected in each biological replicate and with 20 or more total peptides counts are shown.

Figure 2. Transcriptome-wide analysis reveals *XCT* as a global pre-mRNA splicing regulator. A, Experimental design and sampling method for the PacBio Iso-Seq experiment. White and grey boxes represent subjective day and subjective night, respectively. Arabidopsis seedlings were grown at constant light and temperature for 10 days before being harvested and pooled. B and C, Numbers of differentially enriched splicing events represented by significantly differentially down- (B) or up-regulated (C) splice junctions in *xct-1*, *xct-2* and *xct-2* complemented with p*XCT*::g*XCT-YFP-HA* compared with Col-0 (false discovery rate < 0.05). D and E, Frequency of different classes of 5' and 3' splice sites among down- (D) or up-regulated (E) splice junctions in *xct-1* and *xct-2*. Known or novel splice sites were classified by comparing to TAIR10 genome annotation.

Figure 3. *XCT* is required for the fidelity of 3' splice site selection during pre-mRNA splicing. A, A schematic diagram showing the structure of a typical U2-type splice junction. Y: pyrimidine. B, Distribution of the distance between each pair of novel 3' splice site and its corresponding canonical 3' splice sites in *xct-2*. C, Pictograms showing the frequency of nucleotides in the 23-mers sequences flanking the 3' splice sites in all expressed, down-regulated and up-regulated splice junctions in *xct-2*. D, Maximum Entropy Model scores showing the strength of 3' splice sites of all expressed, down-regulated and up-regulated junctions in *xct-2*. E, Percentage of pyrimidines at the -3 position (i.e. the nucleotide preceding the AG at 3' splice site) in *xct-2*. F, Counts of pyrimidines in the Y10 polypyrimidine tract upstream of the 3' splice sites in *xct-2*. PPT, polypyrimidine tract. The lines in the boxplot represent the 75% quartile,

median and 25% quartile of the data, respectively. Statistical significance in (D) and (F) was determined using linear regression model with junction class as a fixed effect and is shown in lower case letters (Tukey's multiple comparison test, $P < 0.05$). Statistical significance in (E) was determined by Fisher's exact test: **: $P < 0.01$; ***: $P < 0.001$.

Figure 4. Dawn-phased genes are significantly enriched among all the circadian-clock-regulated genes that are aberrantly spliced in *xct-2*. A, Venn diagram showing the overlaps of aberrantly spliced genes in *xct-1, xct-2*, and total circadian-clock-regulated genes (Romanowski et al., 2020). The core circadian clock genes that are mis-spliced in both *xct* mutants are indicated in red (non-clock-regulated) and blue (clock-regulated) fonts. Only genes considered as 'detected' in all three RNA-Seq datasets are included. B and C, Phases of estimated peak expression of all circadian clock-regulated genes (Romanowski et al., 2020) that are detected (B) or significantly aberrantly spliced (C) in *xct-2*. The white and gray backgrounds represent the subjective day and subjective night, respectively. D, Distribution of *P*-values from Fisher's extact tests calculating whether the ratio of the number of aberrantly spliced genes in (B) to total clock-regulated genes in (A) is significantly higher than expected by chance in each 2-hour interval.

Figure 5. Time-course qRT-PCR experiments validate the role of *XCT* in regulating pre-mRNA splicing of clock-regulated and non-clock-regulated genes. A - E, Sashimi plots showing PacBio Iso-Seq reads mapped to *LHY* (A), *LNK2* (B), *PRR7* (C), *TOC1* (D) and *SPPA* (E) in Col-0 (teal), *xct-2* (orange) and *xct-2 XCT* (purple). The red rectangles highlight the aberrantly spliced exon-exon junctions that are examined by qRT-PCR in (F) - (J). F - J, Normalized expression of the aberrantly spliced isoforms of *LHY* (F), *LNK2* (G), *PRR7* (H), *TOC1* (I) and *SPPA* (J) in Col-0, *xct-2* and *xct-2 XCT*. Samples were collected every four hours over a 72-h window. Expression levels were examined by qRT-PCR using splice-junction-specific primers and normalized to *PP2*A and *IPP2*. Data points represent mean ± se from three independent biological replicates. For each isoform in each biological replicate, the normalized expression levels were relative to the highest expression levels of their total transcripts in Col-0 across all time points. Teal lines, wild type Col-0; orange lines, *xct-2* muants; purple lines, *xct-2 XCT*. Black background, dark period; white and gray background, light period during subjective day and night, respectively.

Figure 6. Loss of *PRP19* function suppresses the short circadian clock period phenotype but not the splicing defects of core clock genes in *xct-2*. A, Circadian periods of Col-0, *xct-2*, *prp19a-1*, *prp19b-1*, *prp19a-1 prp19b-1*, *prp19a-1 xct-2* and *prp19b-1 xct-2* plants. B - F, Normalized expression of the aberrantly spliced isoforms of *LHY, LNK2, TIC, PRR7* and *TOC1* in Col-0, *xct-2*, *prp19a-1*, *prp19b-1*, *prp19a-1 prp19b-1*, *prp19a-1 xct-2* and *prp19b-1 xct-2*. Samples were collected at the estimated peak expression time for each gene. Expression levels were examined by qRT-PCR using splice-junction-specific primers and normalized to *PP2*A and *IPP2.* Data points represent mean ± se from two independent biological replicates. Statistical significance was determined using linear regression model with genotype as a fixed effect and is shown in lower case letters (Tukey's multiple comparison test, $P < 0.05$).

Figure 7. Expression of *S.pombe XAP5* in Arabidopsis causes a short circadian clock period phenotype without inducing pre-mRNA splicing defects. A, Circadian period of Col-0, *SpXAP5*, *xct-1* and *xct-1 SpXAP5* plants. B, A scatter plot showing differential splicing events in *xct-1 SpXAP5* compared with *xct-1* indicated by normalized PacBio Iso-Seq read counts of exons and intron elements. Differentially spliced elements that are statistically significant (FDR < 0.05) and insignificant (FDR > 0.05) are indicated with red and grey dots, respectively. The dashed line represents y=x. C - F, Relative abundance of aberrantly-spliced to fully-spliced isoforms of *LNK2*, *TIC*, *PRR7* and *TOC1* in Col-0, *SpXAP5, xct-1, xct-1 SpXAP5* and *xct-2*. Differentially spliced isoforms were amplified by semi-qRT-PCR using primers flanking the examined regions, followed by separation and quantification using a Lab Chip GX bioanalyzer. Statistical significance in (A) and (C) - (F) was determined using linear regression models with genotype as a fixed effect and is shown in lower case letters (Tukey's multiple comparison test, $P < 0.05$).

Supplemental Materials

Supplemental Figure 1. *XCT* regulates plant development and the pace of the circadian clock in *Arabidopsis thaliana.* A, Morphology of 52-day-old Col-0, *xct-1*, *xct-2* mutant plants and *xct-2* complemented with p*XCT*::g*XCT-YFP-HA.* Bar = 10 mm. B, Circadian periods of genotypes described in (A). The lines in the boxplot represent the 75% quartile, median and 25%

quartile of the data, respectively. Statistical significance was determined using linear regression model with genotype as a fixed effect and is shown in lower case letters (Tukey's multiple comparison test, $P < 0.05$).

Supplemental Figure 2. *XCT* and *PRL1* control the pace of the circadian clock in an opposite manner. A, Morphology of 21-day-old Col-0, *xct-1*, *xct-2* and *prl1-2* mutant Arabidopsis seedlings. Bar = 1 mm. B, Circadian clock periods of plant genotypes described in (A). The lines in the boxplot represent the 75% quartile, median and 25% quartile of the data, respectively. Statistical significance was determined using linear regression model with genotype as a fixed effect and is shown in lower case letters (Tukey's multiple comparison test, $P < 0.05$).

Supplemental Figure 3. *prl1-2* mutant mainly induces intron retention defects in Arabidopsis. Frequency of different classes of 5' and 3' splice sites of down- (teal) or up-regulated (orange) junctions in *prl1-2* compared to wild type. Known or novel splice sites were classified based on TAIR10 genome annotation.

Supplemental Figure 4. *xct-1* mutant displays reduced fidelity of 3' splice site selection during pre-mRNA splicing. A, Distribution of the distances between each mis-selected novel 3' splice site and its corresponding canonical 3' splice site in *xct-1*. B - C, Pictograms showing the frequency of nucleotides in the 23-mers sequences flanking the 3' splice sites in splice junctions that are down-regulated or up-regulated in *xct-1*.

Supplemental Figure 5. Dawn-phased genes are significantly enriched among genes mis-spliced in *xct-1*. A - B, Phases of peak expression of all the circadian-clock-regulated genes (Romanowski et al., 2020) that are detected (A) or aberrantly spliced (B) in *xct-1*. The white and gray backgrounds represent the subjective day and subjective night, respectively. C, Distribution of *P*-values from Fisher's exact test calculating whether the ratio of the number of mis-spliced genes in (B) to all clock-regulated genes in (A) is significantly higher than expected by chance for each 2-hour interval.

Supplemental Figure 6. Loss of *XCT* function does not affect overall abundance of transcripts of core-circadian-clock genes that are aberrantly spliced in *xct* mutants. A - F, Normalized expression of the total transcript levels of *LHY* (A), *LNK2* (B), *PRR7* (C) and *TOC1* (D) in Col-0, *xct-2* and *xct-2 XCT*. Samples were collected every four hours over a 72-h window. Expression levels were examined by qRT-PCR using primers that detect abundance of all transcripts and normalized to *PP2*A and *IPP2*. Data points represent mean ± se from three independent biological replicates. For each isoform in each biological replicate, the normalized expression levels were relative to the highest expression level in Col-0 across all time points. Teal lines, wild type Col-0; orange lines, *xct-2* muants; purple lines, *xct-2* complemented with p*XCT*::g*XCT-YFP-HA*. Black background, dark period; white and gray background, light period during subjective day and night, respectively.

Supplemental Figure 7. Mutation of Arabidopsis NTC members lengthens circadian period. A - C, Circadian periods of Col-0 and NTC mutant plants, including (A) *prp19a-1*, *prp19b-1*, *prp19a-1 prp19b-1*, (B) *cdc5-1*, and (C) *skip-1*. The lines in the boxplot represent the 75% quartile, median and 25% quartile of the data, respectively. Statistical significance was determined using linear regression model with genotype as a fixed effect and is shown in lower case letters (Tukey's multiple comparison test, $P < 0.05$).

Supplemental Figure 8. *PRP19A* (*MAC3A*) and *PRP19B* (*MAC3B*) transcript levels are highly reduced in *prp19a-1* and *prp19b-1* mutant backgrounds, respectively. A - B, Normalized expression of (A) *PRP19A* and (B) *PRP19B* in Col-0, *xct-2*, *prp19a-1*, *prp19b-1*, *prp19a-1 prp19b-1*, *prp19a-1 xct-2* and *prp19b-1 xct-2*. Expression levels were examined by qRT-PCR and normalized to *PP2*A and *IPP2*. Data points represent mean ± se from two independent biological replicates. Samples were collected at ZT0. Statistical significance was determined using linear regression model with genotype as a fixed effect and is shown in lower case letters (Tukey's multiple comparison test, $P < 0.05$).

Supplemental Figure 9. Loss of *PRP19* function suppresses the *xct-2* short circadian period phenotype. An independent replicate of the experiment in Figure 6A showing the circadian

periods of Col-0, *xct-2*, *prp19a-1*, *prp19b-1*, *prp19a-1 prp19b-1*, *prp19a-1 xct-2* and *prp19b-1 xct-2* plants.

Supplemental Figure 10. *PRP19* and *XCT* additively regulate Arabidopsis rosette size. A, Morphology of 39-day-old Col-0, *xct-2*, *prp19a-1*, *prp19b-1*, *prp19a-1 prp19b-1*, *prp19b-1 xct-2* and *prp19a-1 xct-2* plants. Bar = 10 mm. B, Rosette diameter of 36-day-old plants grown under long-day conditions at 22˚C (n = 5-11). Data are shown as mean ± se. Statistical significance was determined using linear regression model with genotype as a fixed effect and is shown in lower case letters (Tukey's multiple comparison test, $P < 0.05$).

Supplemental Figure 11. Levels of total transcripts of *XCT*-targeted core clock genes are not altered in *prp19a-1 xct-2* or *prp19b-1 xct-2* compared to *xct-2.* A - F, Normalized expression of total transcript levels (A, C, E, F) or specific normally spliced isoforms (B, D) of *LHY, LNK2, TIC, PRR7* and *TOC1* in Col-0, *xct-2*, *prp19a-1*, *prp19b-1*, *prp19a-1 prp19b-1*, *prp19a-1 xct-2* and *prp19b-1 xct-2*. Expression levels were examined by qRT-PCR using primers that detect abundance of all transcripts or normally spliced isoforms and normalized to *PP2*A and *IPP2.* Data points represent mean ± se from two independent biological replicates. Samples were collected at the estimated peak expression time for each gene. Statistical significance was determined using linear regression model with genotype as a fixed effect and is shown in lower case letters (Tukey's multiple comparison test, $P < 0.05$).

Supplemental Figure 12. Arabidopsis plants expressing *SpXAP5* don't mimic the developmental defects caused by *xct-2* mutant. Morphology of 50–day-old Col-0, *SpXAP5, xct-1*, *xct-1 SpXAP5* and *xct-2*. Bar = 10 mm.

Supplemental Dataset 1. A list of proteins co-purified with XCT-IP detected by MS.

Supplemental Dataset 2. Summary of full-length reads obtained from each PacBio Iso-Seq library.

Supplemental Dataset 3. Lists of significant differential splicing events identified by JunctionSeq.

Supplemental Dataset 4. Lists of total and aberrantly spliced circadian-clock-regulated genes described in this study.

Supplemental Dataset 5. A list of primers used in this study.

References

Anver S, Roguev A, Zofall M, Krogan NJ, Grewal SIS, Harmer SL (2014) Yeast X-chromosome-associated protein 5 (Xap5) functions with H2A.Z to suppress aberrant transcripts. EMBO Rep **15**: 894–902

Bessonov S, Anokhina M, Krasauskas A, Golas MM, Sander B, Will CL, Urlaub H, Stark H, Lührmann R (2010) Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. RNA **16**: 2384–2403

Bessonov S, Anokhina M, Will CL, Urlaub H, Lührmann R (2008) Isolation of an active step I spliceosome and composition of its RNP core. Nature **452**: 846–850

Chanarat S, Seizl M, Strässer K (2011) The Prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. Genes Dev **25**: 1147–1158

Chanarat S, Sträßer K (2013) Splicing and beyond: the many faces of the Prp19 complex. Biochim Biophys Acta **1833**: 2126–2134

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J **16**: 735–743

Covington MF, Harmer SL (2007) The circadian clock regulates auxin signaling and responses in Arabidopsis. PLoS Biol **5**: e222

Creux N, Harmer S (2019) Circadian Rhythms in Plants. Cold Spring Harb Perspect Biol **11**: a034611

Deng X, Lu T, Wang L, Gu L, Sun J, Kong X, Liu C, Cao X (2016) Recruitment of the NineTeen Complex to the activated spliceosome requires AtPRMT5. Proc Natl Acad Sci U S A **113**: 5447–5452

Fang X, Shi Y, Lu X, Chen Z, Qi Y (2015) CMA33/XCT Regulates Small RNA Production through Modulating the Transcription of Dicer-Like Genes in Arabidopsis. Mol Plant **8**: 1227–1236

Feke A, Liu W, Hong J, Li M-W, Lee C-M, Zhou EK, Gendron JM (2019) Decoys provide a scalable platform for the identification of plant E3 ubiquitin ligases that regulate circadian function. Elife **8**: e44558

Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, Fox SE, Wong W-K, Mockler TC (2010) Genome-wide mapping of alternative splicing in Arabidopsis thaliana. Genome Res **20**: 45–58

Harmer SL (2009) The circadian system in higher plants. Annu Rev Plant Biol **60**: 357–377

Hartley SW, Mullikin JC (2016) Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. Nucleic Acids Res **44**: e127

Hartley SW, Mullikin JC (2015) QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. BMC Bioinformatics **16**: 224

Hatakeyama S, Yada M, Matsumoto M, Ishida N, Nakayama KI (2001) U box proteins as a new family of ubiquitin-protein ligases. J Biol Chem **276**: 33111–33120

Hogg R, McGrail JC, O'Keefe RT (2010) The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. Biochem Soc Trans **38**: 1110–1115

Horowitz DS (2012) The mechanism of the second step of pre-mRNA splicing. Wiley Interdiscip Rev RNA **3**: 331–350

Hsu PY, Harmer SL (2012) Circadian phase has profound effects on differential expression analysis. PLoS One **7**: e49853

Hsu PY, Harmer SL (2014) Wheels within wheels: the plant circadian system. Trends Plant Sci **19**: 240–249

Idrissou M, Maréchal A (2022) The PRP19 Ubiquitin Ligase, Standing at the Cross-Roads of mRNA Processing and Genome Stability. Cancers (Basel) **14**: 878

James AB, Syed NH, Bordage S, Marshall J, Nimmo GA, Jenkins GI, Herzyk P, Brown JWS, Nimmo HG (2012) Alternative splicing mediates responses of the Arabidopsis circadian clock to temperature changes. Plant Cell **24**: 961–981

Jia T, Zhang B, You C, Zhang Y, Zeng L, Li S, Johnson KCM, Yu B, Li X, Chen X (2017) The Arabidopsis MOS4-Associated Complex Promotes MicroRNA Biogenesis and Precursor Messenger RNA Splicing. Plant Cell **29**: 2626–2643

Jones MA, Williams BA, McNicol J, Simpson CG, Brown JWS, Harmer SL (2012) Mutation of Arabidopsis spliceosomal timekeeper locus1 causes circadian clock defects. Plant Cell **24**: 4066–4082

Klaric JA, Wüst S, Panier S (2021) New Faces of old Friends: Emerging new Roles of RNA-Binding Proteins in the DNA Double-Strand Break Response. Front Mol Biosci **8**: 668821

Koncz C, Dejong F, Villacorta N, Szakonyi D, Koncz Z (2012) The spliceosome-activating complex: molecular mechanisms underlying the function of a pleiotropic regulator. Front Plant Sci **3**: 9

Kumimoto RW, Ellison CT, Toruño TY, Bak A, Zhang H, Casteel CL, Coaker G, Harmer SL (2021) XAP5 CIRCADIAN TIMEKEEPER Affects Both DNA Damage Responses and Immune Signaling in Arabidopsis. Front Plant Sci **12**: 707923

Kwon Y-J, Park M-J, Kim S-G, Baldwin IT, Park C-M (2014) Alternative splicing and nonsense-mediated decay of circadian clock genes under environmental stress conditions in Arabidopsis. BMC Plant Biol **14**: 136

Lee Y-R, Khan K, Armfield-Uhas K, Srikanth S, Thompson NA, Pardo M, Yu L, Norris JW, Peng Y, Gripp KW, et al (2020) Mutations in FAM50A suggest that Armfield XLID syndrome is a spliceosomopathy. Nat Commun **11**: 3698

Li L, Tian G, Peng H, Meng D, Wang L, Hu X, Tian C, He M, Zhou J, Chen L, et al (2018) New class of transcription factors controls flagellar assembly by recruiting RNA polymerase II in Chlamydomonas. Proc Natl Acad Sci U S A **115**: 4435–4440

Li Y, Yang J, Shang X, Lv W, Xia C, Wang C, Feng J, Cao Y, He H, Li L, et al (2019) SKIP regulates environmental fitness and floral transition by forming two distinct complexes in Arabidopsis. New Phytol **224**: 321–335

Liu L, Li X, Yuan L, Zhang G, Gao H, Xu X, Zhao H (2022) XAP5 CIRCADIAN TIMEKEEPER specifically modulates 3' splice site recognition and is important for circadian clock regulation partly by alternative splicing of LHY and TIC. Plant Physiol Biochem **172**: 151–157

Maréchal A, Li J-M, Ji XY, Wu C-S, Yazinski SA, Nguyen HD, Liu S, Jiménez AE, Jin J, Zou L (2014) PRP19 transforms into a sensor of RPA-ssDNA after DNA damage and drives ATR activation via a ubiquitin-mediated circuitry. Mol Cell **53**: 235–246

Marshall CM, Tartaglio V, Duarte M, Harmon FG (2016) The Arabidopsis sickle Mutant Exhibits Altered Circadian Clock Responses to Cool Temperatures and Temperature-Dependent Alternative Splicing. Plant Cell **28**: 2560–2575

Martin-Tryon EL, Harmer SL (2008) XAP5 CIRCADIAN TIMEKEEPER coordinates light signals for proper timing of photomorphogenesis and the circadian clock in Arabidopsis. Plant Cell **20**: 1244–1259

Martin-Tryon EL, Kreps JA, Harmer SL (2007) GIGANTEA acts in blue light signaling and has biochemically separable roles in circadian clock and flowering time regulation. Plant Physiol **143**: 473–486

Mayas RM, Maita H, Staley JP (2006) Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat Struct Mol Biol **13**: 482–490

de Melo JRF, Gutsch A, Caluwé TD, Leloup J-C, Gonze D, Hermans C, Webb AAR, Verbruggen N (2021) Magnesium maintains the length of the circadian period in Arabidopsis. Plant Physiol **185**: 519–532

Meng X, Wang Q, Hao R, Li X, Li M, Hu R, Du H, Hu Z, Yu B, Li S (2022) RNA-binding protein MAC5A interacts with the 26S proteasome to regulate DNA damage response in Arabidopsis. Plant Physiol kiac510

Monaghan J, Xu F, Gao M, Zhao Q, Palma K, Long C, Chen S, Zhang Y, Li X (2009) Two Prp19-like U-box proteins in the MOS4-associated complex play redundant roles in plant innate immunity. PLoS Pathog **5**: e1000526

Nohales MA, Kay SA (2016) Molecular mechanisms at the core of the plant circadian oscillator. Nat Struct Mol Biol **23**: 1061–1069

Nolte C, Staiger D (2015) RNA around the clock - regulation at the RNA level in biological timing. Front Plant Sci **6**: 311

Perez-Santángelo S, Mancini E, Francey LJ, Schlaen RG, Chernomoretz A, Hogenesch JB, Yanovsky MJ (2014) Role for LSM genes in the regulation of circadian rhythms. Proc Natl Acad Sci U S A **111**: 15166–15171

Plautz JD, Straume M, Stanewsky R, Jamison CF, Brandes C, Dowse HB, Hall JC, Kay SA (1997) Quantitative analysis of Drosophila period gene transcription in living animals. J Biol Rhythms **12**: 204–217

Romanowski A, Schlaen RG, Perez-Santangelo S, Mancini E, Yanovsky MJ (2020) Global transcriptome analysis reveals circadian control of splicing events in Arabidopsis thaliana. Plant J **103**: 889–902

Sanchez SE, Petrillo E, Beckwith EJ, Zhang X, Rugnone ML, Hernando CE, Cuevas JC, Godoy Herz MA, Depetris-Chauvin A, Simpson CG, et al (2010) A methyl transferase links the circadian clock to the regulation of alternative splicing. Nature **468**: 112–116

Schlaen RG, Mancini E, Sanchez SE, Perez-Santángelo S, Rugnone ML, Simpson CG, Brown JWS, Zhang X, Chernomoretz A, Yanovsky MJ (2015) The spliceosome assembly factor GEMIN2 attenuates the effects of temperature on alternative splicing and circadian rhythms. Proc Natl Acad Sci U S A **112**: 9382–9387

Semlow DR, Blanco MR, Walter NG, Staley JP (2016) Spliceosomal DEAH-Box ATPases Remodel Pre-mRNA to Activate Alternative Splice Sites. Cell **164**: 985–998

Seo PJ, Park M-J, Lim M-H, Kim S-G, Lee M, Baldwin IT, Park C-M (2012) A self-regulatory circuit of CIRCADIAN CLOCK-ASSOCIATED1 underlies the circadian clock regulation of temperature responses in Arabidopsis. Plant Cell **24**: 2427–2442

Shakhmantsir I, Sehgal A (2019) Splicing the Clock to Maintain and Entrain Circadian Rhythms. J Biol Rhythms **34**: 584–595

Shi Y (2017) Mechanistic insights into precursor messenger RNA splicing by the spliceosome. Nat Rev Mol Cell Biol **18**: 655–670

Uehara TN, Nonoyama T, Taki K, Kuwata K, Sato A, Fujimoto KJ, Hirota T, Matsuo H, Maeda AE, Ono A, et al (2022) Phosphorylation of RNA Polymerase II by CDKC;2 Maintains the Arabidopsis Circadian Clock Period. Plant Cell Physiol **63**: 450–462

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

Wang X, Wu F, Xie Q, Wang H, Wang Y, Yue Y, Gahura O, Ma S, Liu L, Cao Y, et al (2012) SKIP is a component of the spliceosome linking alternative splicing and the circadian clock in Arabidopsis. Plant Cell **24**: 3278–3295

Wei Y, Yee PP, Liu Z, Zhang L, Guo H, Zheng H, Anderson B, Gulley M, Li W (2020) NEDD4L-mediated Merlin ubiquitination facilitates Hippo pathway activation. EMBO Rep **21**: e50642

Wilkinson ME, Charenton C, Nagai K (2020) RNA Splicing by the Spliceosome. Annu Rev Biochem **89**: 359–388

Xu Y-J, Lei Y, Li R, Zhang L-L, Zhao Z-X, Zhao J-H, Fan J, Li Y, Yang H, Shang J, et al (2017) XAP5 CIRCADIAN TIMEKEEPER Positively Regulates RESISTANCE TO POWDERY MILDEW8.1-Mediated Immunity in Arabidopsis. Front Plant Sci **8**: 2044

Yang Y, Li Y, Sancar A, Oztas O (2020) The circadian clock shapes the Arabidopsis transcriptome by regulating alternative splicing and alternative polyadenylation. J Biol Chem **295**: 7608–7619

Zhan X, Lu Y, Zhang X, Yan C, Shi Y (2022) Mechanism of exon ligation by human spliceosome. Mol Cell **82**: 2769-2778.e4

Zhang S, Liu Y, Yu B (2014) PRL1, an RNA-binding protein, positively regulates the accumulation of miRNAs and siRNAs in Arabidopsis. PLoS Genet **10**: e1004841

Zhang S, Xie M, Ren G, Yu B (2013) CDC5, a DNA binding protein, positively regulates posttranscriptional processing and/or transcription of primary microRNA transcripts. Proc Natl Acad Sci U S A **110**: 17588–17593

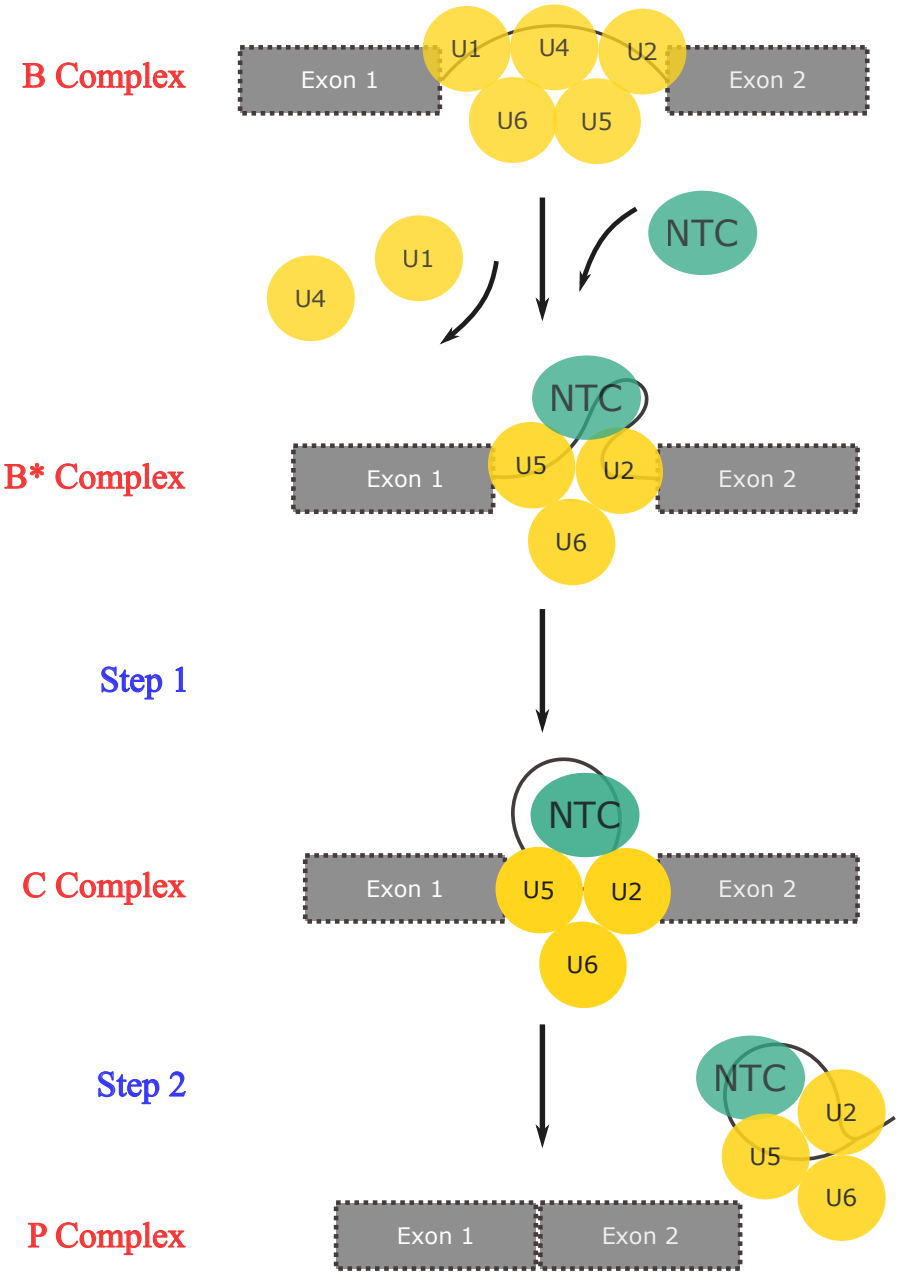

Figure 1. Simplified overview of pre-mRNA splicing reactions. A schematic diagram highlighting the two catalytic transesterification steps and the association between the NTC and the spliceosomal complex during pre-mRNA splicing. Gray boxes and solid black lines represent exons and introns, respectively. U1-U6 small nuclear ribonucleoproteins (snRNPs) are indicated by yellow circles. The NTC is indicated by a green oval.

Table 1. XCT physically associates with splicing-related proteins, especially the NTC components, in Arabidopsis. Mass Spectrometry (MS) data showing that spliceosome-associated proteins are significantly more enriched in XCT-YFP-HA-IP than control IP (fold enrichment > 3; *P*-value < 0.05, Welch two sample t.test). The full list of proteins co-purified with XCT are described in Supplemental Dataset 1. Only proteins detected in each biological replicate and with 20 or more total peptides counts are shown.

AGI	*A. thaliana* gene name	*P*-value	fold enrichment	category	*S. cerevisiae* homolog	*H. sapiens* homolog
AT2G21150	XCT	6.21E-04	Inf			FAM50A
AT1G04510	PRP19A/MAC3A/PUB59	1.52E-02	4.2	NTC	Prp19p	PRPF19; PSO4; SNEV
AT2G33340	PRP19B/MAC3B/PUB60	2.17E-03	4.8	NTC		
AT4G15900	PRL1/MAC2	1.24E-02	7.7	NTC	Prp46p	PLRG1
AT1G77180	SKIP/MAC6	1.36E-03	4.9	NTC	Prp45p	SNW1
AT1G09770	CDC5/MAC1	4.06E-03	6.7	NTC	Cef1p	CDC5L
AT1G07360	MAC5A	2.15E-03	8.8	NTC	Ecm2p	RBM22
AT5G41770	CRNK1/MAC10	5.78E-03	7.7	NTC	Clf1p/Syf3	CRNKL1/SYF3
AT3G18790	ISY1/MAC8	1.26E-03	6.1	NTC	Isy1p	ISY1
AT2G38770	EMB2765/MAC7	1.36E-02	4.4	NTC	Cwf11p	AQR
AT5G28740		8.86E-03	4.6	NTC	Syf1p	XAB2
AT3G18165	MOS4	4.09E-03	4.5	NTC	Snt309p	BCAS2/SPF27
AT5G51280		1.65E-02	4.3	other splicing	Dbp2p	DDX41
AT2G47640		4.93E-03	4.3	other splicing	Smd2p	SNRPD2
AT3G26560		8.15E-03	24	other splicing	Prp22p	DHX8
AT1G09760	U2A'	3.51E-02	3.3	U2 complex	Lea1p	SNRPA1

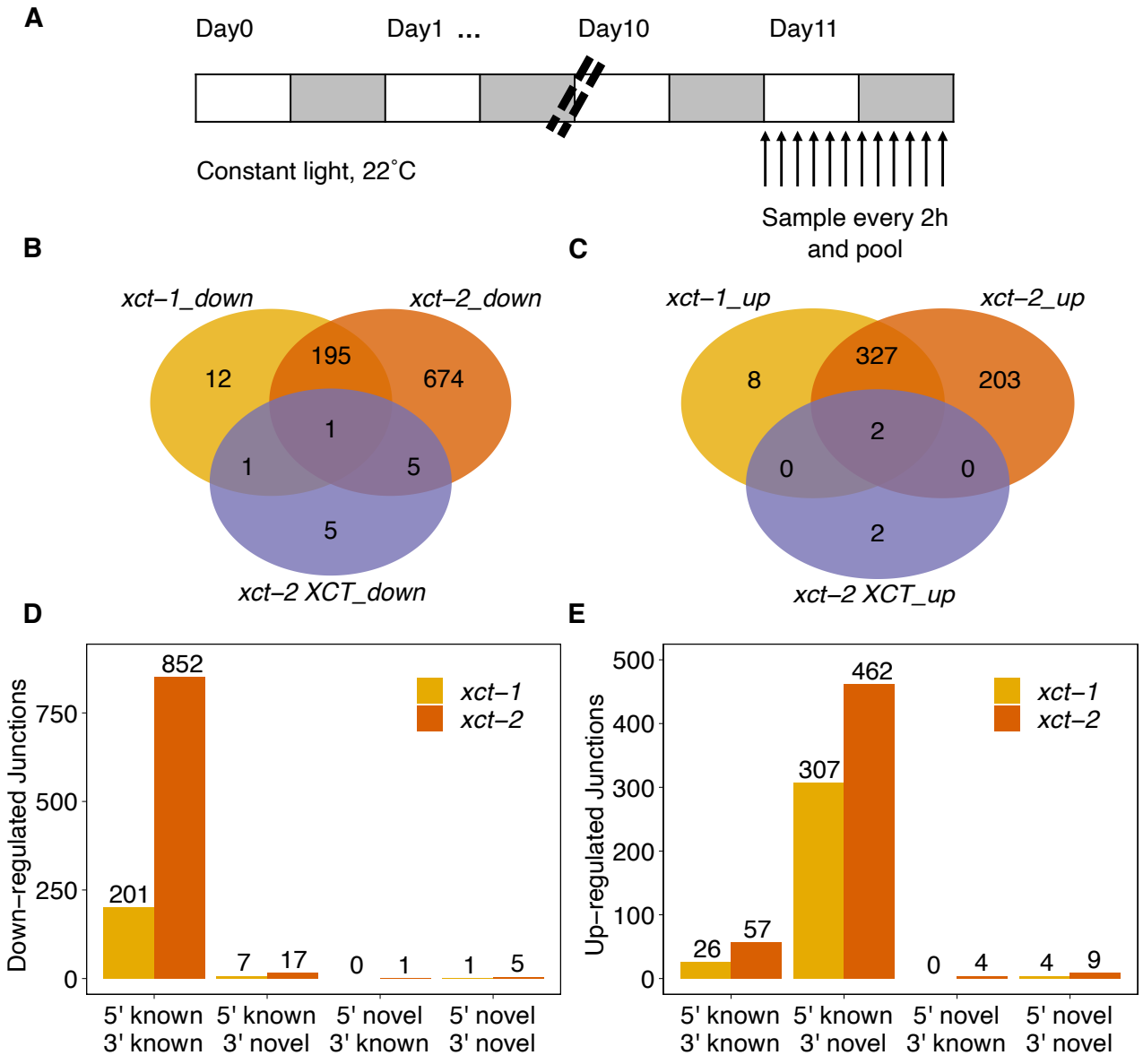

Figure 2. Transcriptome-wide analysis reveals *XCT* as a global pre-mRNA splicing regulator. A, Experimental design and sampling method for the PacBio Iso-Seq experiment. White and grey boxes represent subjective day and subjective night, respectively. Arabidopsis seedlings were grown at constant light and temperature for 10 days before being harvested and pooled. B and C, Numbers of differentially enriched splicing events represented by significantly differentially down- (B) or up-regulated (C) splice junctions in *xct-1*, *xct-2* and *xct-2* complemented with p*XCT*::g*XCT-YFP-HA* compared with Col-0 (false discovery rate < 0.05). D and E, Frequency of different classes of 5' and 3' splice sites among down- (D) or up-regulated (E) splice junctions in *xct-1* and *xct-2*. Known or novel splice sites were classified by comparing to TAIR10 genome annotation.

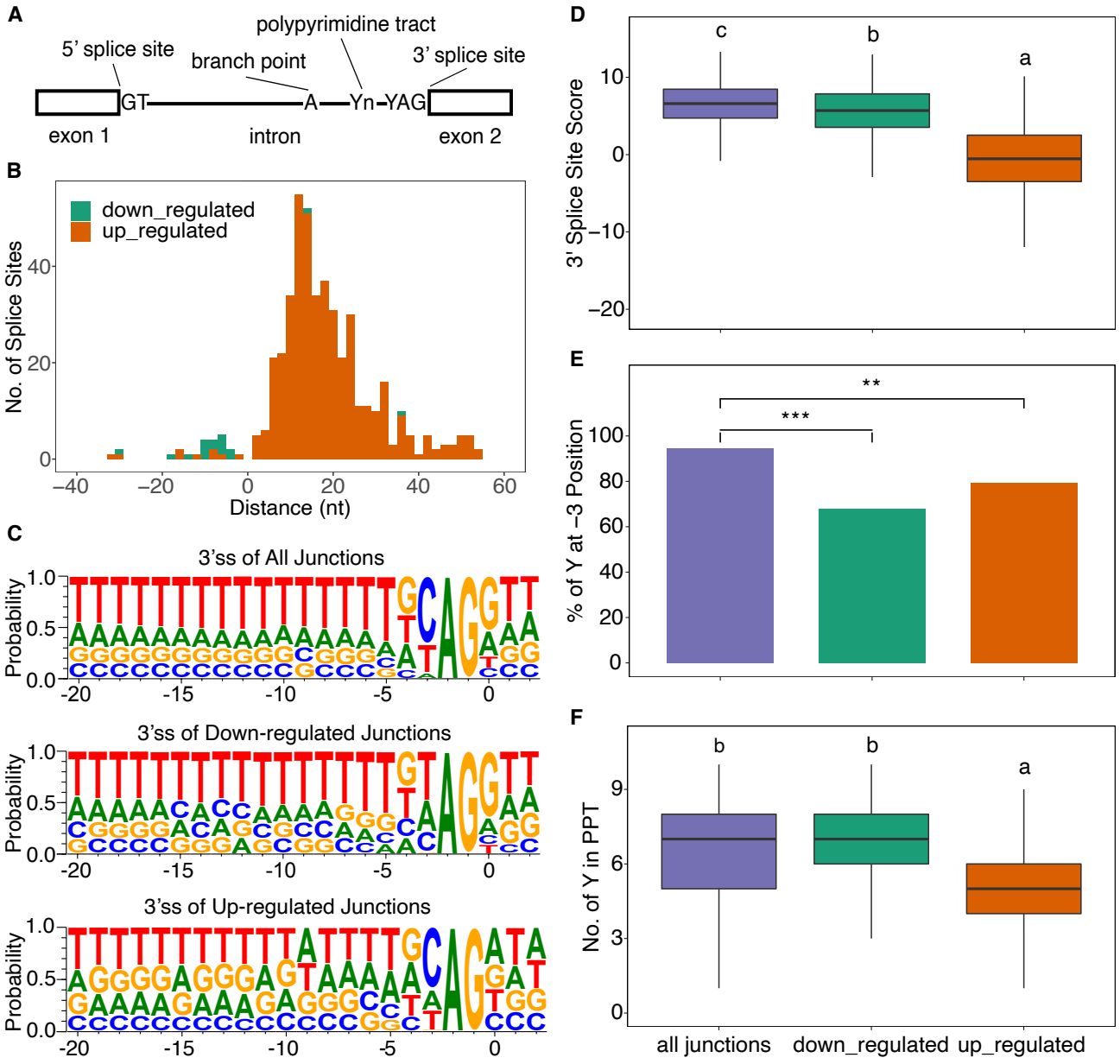

Figure 3. *XCT* is required for the fidelity of 3' splice site selection during pre-mRNA splicing. A, A schematic diagram showing the structure of a typical U2-type splice junction. Y: pyrimidine. B, Distribution of the distance between each pair of novel 3' splice site and its corresponding canonical 3' splice sites in *xct-2*. C, Pictograms showing the frequency of nucleotides in the 23-mers sequences flanking the 3' splice sites in all expressed, down-regulated and up-regulated splice junctions in *xct-2*. D, Maximum Entropy Model scores showing the strength of 3' splice sites of all expressed, down-regulated and up-regulated junctions in *xct-2*. E, Percentage of pyrimidines at the -3 position (i.e. the nucleotide preceding the AG at 3' splice site) in *xct-2*. F, Counts of pyrimidines in the Y10 polypyrimidine tract upstream of the 3' splice sites in *xct-2*. PPT, polypyrimidine tract. The lines in the boxplot represent the 75% quartile, median and 25% quartile of the data, respectively. Statistical significance in (D) and (F) was determined using linear regression model with junction class as a fixed effect and is shown in lower case letters (Tukey's multiple comparison test, $P < 0.05$). Statistical significance in (E) was determined by Fisher's exact test: **: $P < 0.01$; ***: $P < 0.001$.

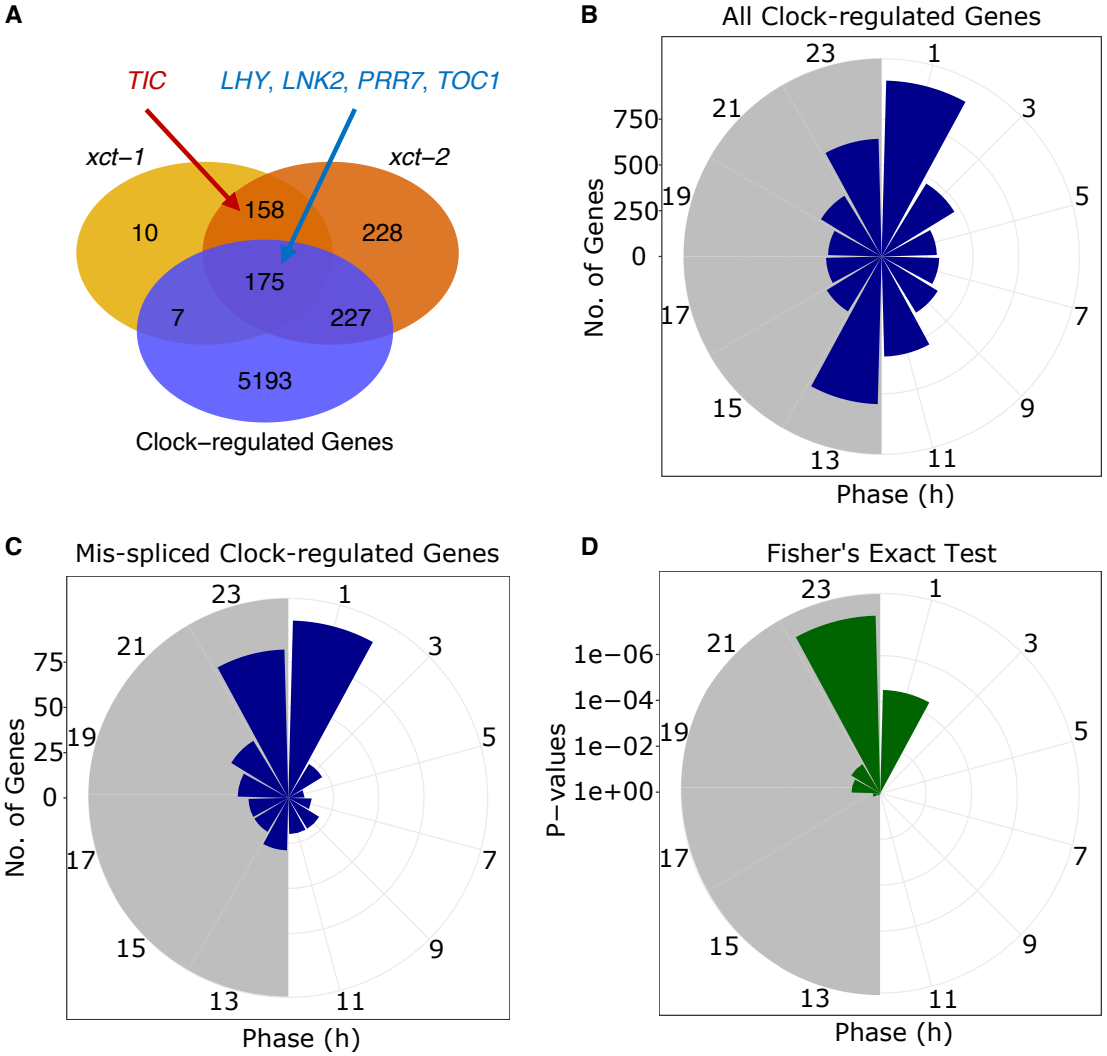

Figure 4. Dawn-phased genes are significantly enriched among all the circadian-clock-regulated genes that are aberrantly spliced in *xct-2*. A, Venn diagram showing the overlaps of aberrantly spliced genes in *xct-1*, *xct-2*, and total circadian-clock-regulated genes (Romanowski et al., 2020). The core circadian clock genes that are mis-spliced in both *xct* mutants are indicated in red (non-clock-regulated) and blue (clock-regulated) fonts. Only genes considered as 'detected' in all three RNA-Seq datasets are included. B and C, Phases of estimated peak expression of all circadian clock-regulated genes (Romanowski et al., 2020) that are detected (B) or significantly aberrantly spliced (C) in *xct-2*. The white and gray backgrounds represent the subjective day and subjective night, respectively. D, Distribution of *P*-values from Fisher's extact tests calculating whether the ratio of the number of aberrantly spliced genes in (B) to total clock-regulated genes in (A) is significantly higher than expected by chance in each 2-hour interval.

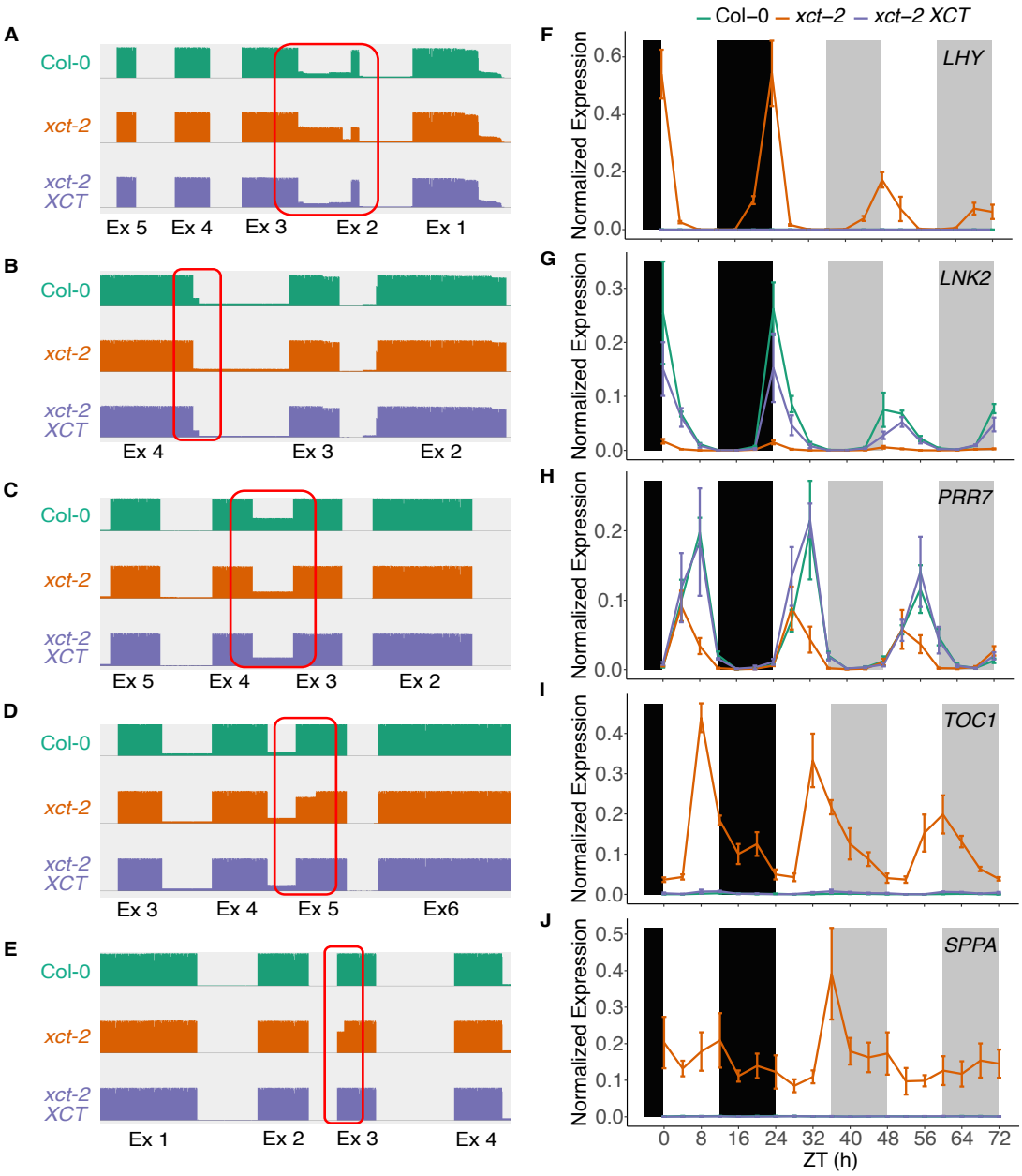

Figure 5. Time-course qRT-PCR experiments validate the role of *XCT* in regulating pre-mRNA splicing of clock-regulated and non-clock-regulated genes. A - E, Sashimi plots showing PacBio Iso-Seq reads mapped to *LHY* (A), *LNK2* (B), *PRR7* (C), *TOC1* (D) and *SPPA* (E) in Col-0 (teal), *xct-2* (orange) and *xct-2 XCT* (purple). The red rectangles highlight the aberrantly spliced exon-exon junctions that are examined by qRT-PCR in (F) - (J). F - J, Normalized expression of the aberrantly spliced isoforms of *LHY* (F), *LNK2* (G), *PRR7* (H), *TOC1* (I) and *SPPA* (J) in Col-0, *xct-2* and *xct-2 XCT*. Samples were collected every four hours over a 72-h window. Expression levels were examined by qRT-PCR using splice-junction-specific primers and normalized to *PP2A* and *IPP2*. Data points represent mean ± se from three independent biological replicates. For each isoform in each biological replicate, the normalized expression levels were relative to the highest expression levels of their total transcripts in Col-0 across all time points. Teal lines, wild type Col-0; orange lines, *xct-2* muants; purple lines, *xct-2 XCT*. Black background, dark period; white and gray background, light period during subjective day and night, respectively.

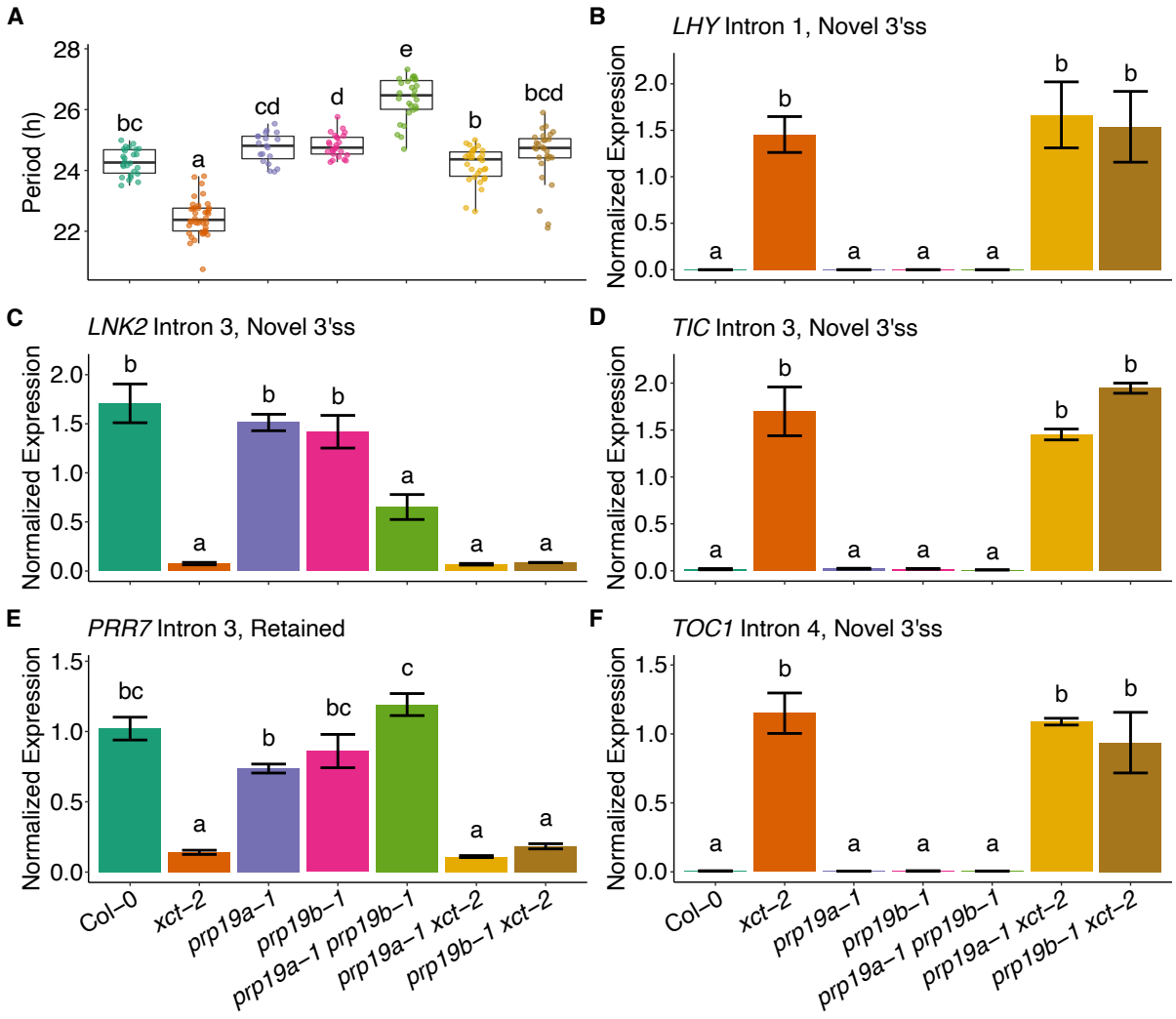

Figure 6. Loss of *PRP19* function suppresses the short circadian clock period phenotype but not the splicing defects of core clock genes in *xct-2*. A, Circadian periods of Col-0, *xct-2*, *prp19a-1*, *prp19b-1*, *prp19a-1 prp19b-1*, *prp19a-1 xct-2* and *prp19b-1 xct-2* plants. B - F, Normalized expression of the aberrantly spliced isoforms of *LHY*, *LNK2*, *TIC*, *PRR7* and *TOC1* in Col-0, *xct-2*, *prp19a-1*, *prp19b-1*, *prp19a-1 prp19b-1*, *prp19a-1 xct-2* and *prp19b-1 xct-2*. Samples were collected at the estimated peak expression time for each gene. Expression levels were examined by qRT-PCR using splice-junction-specific primers and normalized to *PP2A* and *IPP2*. Data points represent mean ± se from two independent biological replicates. Statistical significance was determined using linear regression model with genotype as a fixed effect and is shown in lower case letters (Tukey's multiple comparison test, $P < 0.05$).

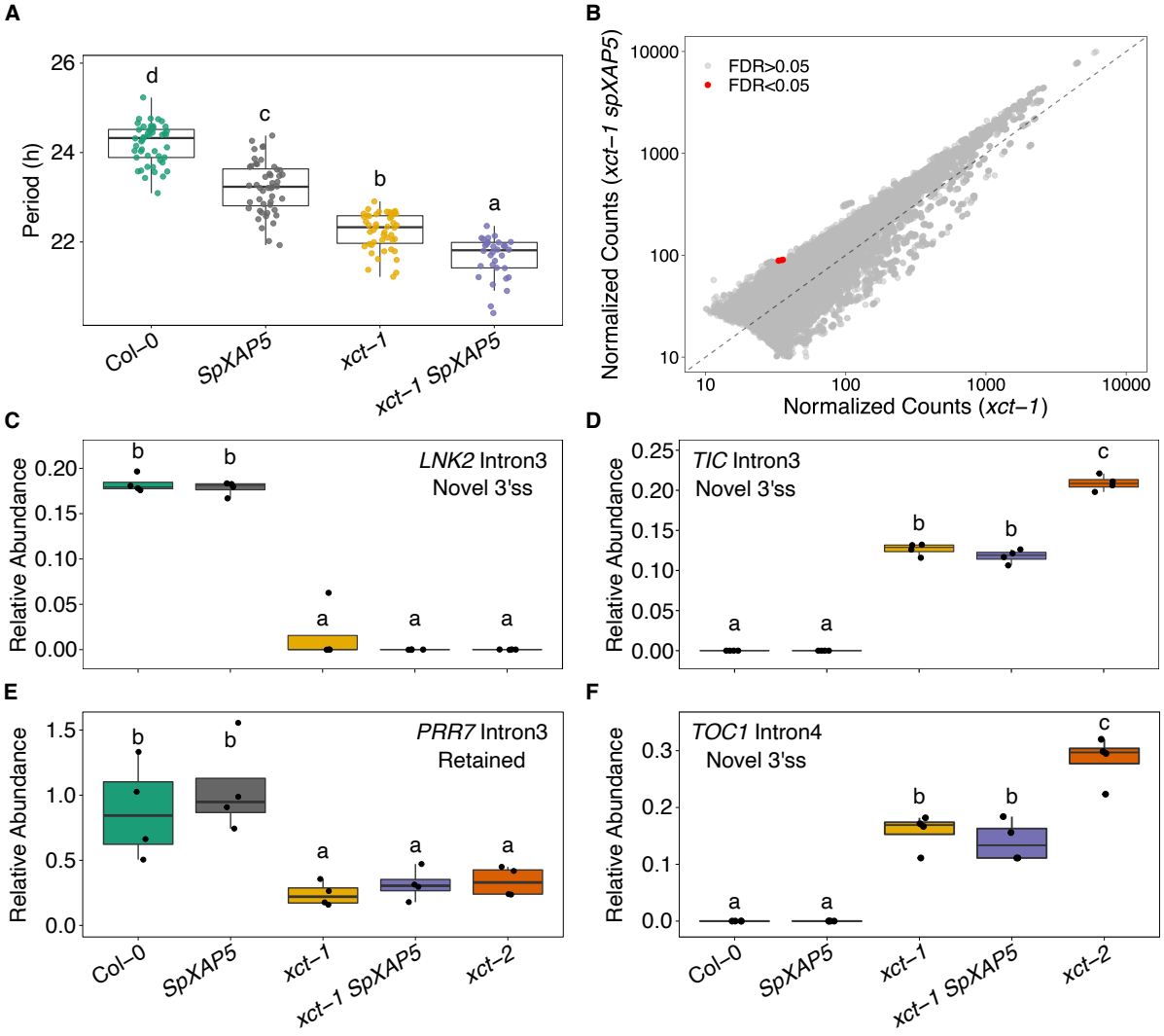

Figure 7. Expression of *S.pombe XAP5* in Arabidopsis causes a short circadian clock period phenotype without inducing pre-mRNA splicing defects. A, Circadian period of Col-0, *SpXAP5*, *xct-1* and *xct-1 SpXAP5* plants. B, A scatter plot showing differential splicing events in *xct-1 SpXAP5* compared with *xct-1* indicated by normalized PacBio Iso-Seq read counts of exons and intron elements. Differentially spliced elements that are statistically significant (FDR < 0.05) and insignificant (FDR > 0.05) are indicated with red and grey dots, respectively. The dashed line represents y=x. C - F, Relative abundance of aberrantly-spliced to fully-spliced isoforms of *LNK2*, *TIC*, *PRR7* and *TOC1* in Col-0, *SpXAP5*, *xct-1*, *xct-1 SpXAP5* and *xct-2*. Differentially spliced isoforms were amplified by semi-qRT-PCR using primers flanking the examined regions, followed by separation and quantification using a Lab Chip GX bioanalyzer. Statistical significance in (A) and (C) - (F) was determined using linear regression models with genotype as a fixed effect and is shown in lower case letters (Tukey's multiple comparison test, $P < 0.05$).