Arabidopsis U1 snRNP subunit LUC7 functions in alternative splicing and preferential removal of terminal introns

Marcella de Francisco Amorim[1,2,3], Eva-Maria Willing[4], Emese Xochitl Szabo[1,2,3,*], Irina Droste-Borel[5], Boris Maček[5], Korbinian Schneeberger[4] and Sascha Laubinger[1,2,3,*]

[1] Centre for Plant Molecular Biology (ZMBP), University of Tuebingen, Tuebingen, Germany

[2] Chemical Genomics Centre (CGC) of the Max Planck Society, Dortmund, Germany

[3] Max Planck Institute for Developmental Biology, Tuebingen, Germany

[4] Max Planck Institute for Plant Breeding Research (MPIPZ), Cologne, Germany

[5] Proteome Centre, University of Tuebingen, Tuebingen, Germany

* present address: Carl von Ossietzky University, Oldenburg, Germany

Corresponding author: Sascha Laubinger (sascha.laubinger@uol.de)

Keywords: U1 snRNP, Splicing, Non-sense mediated mRNA decay (NMD), Polyadenylation, RNA transport, Arabidopsis

Abstract

Introns are removed by the spliceosome, a large complex composed of five ribonucleoprotein subcomplexes (U snRNP). The U1 snRNP, which binds to 5' splice sites, plays an essential role in early steps of the splicing reaction. Here, we show that Arabidopsis LUC7 is a U1 snRNP subunit that affects constitutive and alternative splicing. Interestingly, LUC7 specifically promotes splicing of a subset of terminal introns. Splicing of LUC7-dependent terminal introns is a prerequisite for nuclear export and can be modulated by stress. Globally, intron retention under stress conditions occurs preferentially among first and terminal introns, uncovering an unknown bias for splicing regulation in Arabidopsis. Taken together, our study reveals that the Arabidopsis U1 snRNP component LUC7 is important for alternative splicing and removal of terminal introns and it suggests that Arabidopsis terminal introns fine-tune gene expression under stress conditions.

Introduction

Eukaryotic genes are often interrupted by non-coding sequences called introns that are removed from pre-mRNAs while the remaining sequence, the exons, are joined together. This process, called splicing, is an essential step before the translation of the mature mRNAs and it offers a wide range of advantages for eukaryotic organisms. For instance, alternative splicing allows the production of more than one isoform from a single gene expanding the genome coding capacity (Kornblihtt *et al.*, 2013; Reddy *et al.*, 2013). In plants, alternative splicing contributes to essentially all aspects of development and stress responses (Carvalho *et al.*, 2013; Staiger & Brown, 2013). Alternative splicing can also generate transcripts with premature termination codons (PTC) or/and a long 3'UTR, which may lead to RNA degradation via the nonsense-mediated decay (NMD) pathway (Drechsel *et al.*, 2013; Kalyna *et al.*, 2012; Shaul, 2015). Furthermore, splicing is coupled with other RNA processing events, such as 3'end formation and RNA transport to the cytosol (Kaida, 2016; Muller-McNicoll *et al.*, 2016).

Intron removal is catalyzed by a large macromolecular complex, the spliceosome, which is formed by five small ribonucleoprotein particles (snRNP): the U1, U2, U4, U5 and U6 snRNP. Each U snRNP contains a heteroheptameric ring of Sm or Lsm proteins, snRNP-specific proteins and a uridine-rich snRNA. Additional non-core spliceosomal proteins participate also during the splicing reaction affecting exon-intron recognition/definition and thus splicing efficiency. The canonical splicing cycle starts with binding of the U1 snRNP to the 5' splice site (5'ss), followed by association of auxiliary proteins such as U2AF to the pre-mRNA, which facilitate the recognition of the 3' splice site (3'ss). The thereby formed complex E recruits the U2 snRNP to generate complex A. In the next step, a trimeric complex consisting of U4/U5/U6 snRNPs joins to form complex B. Several rearrangements and ejection of

the U1 and U4 snRNP are necessary to generate a catalytically active splicing complex (Wahl *et al.*, 2009; Will & Luhrmann, 2011).

The fact that U1 snRNP is recruited to the 5'ss in the initial step of splicing suggests that this complex is necessary for correct splicing site selection. Indeed, it has been shown that U1-deficient zebrafish mutants accumulate alternative spliced transcripts, suggesting that the U1 snRNP fulfills regulatory roles in splice site selection (Rosel *et al.*, 2011). Although the spliceosome consists of stoichiometrically equal amounts of each subunit, the U1 snRNP is more abundant than all the other spliceosomal subcomplexes (Kaida, 2016; Kaida *et al.*, 2010). One reason for this is that the U1 snRNP executes splicing independent functions. For instance, metazoan U1 snRNP binds not only to the 5'ss, but also throughout the nascent transcript to block a premature cleavage and polyadenylation (Berg *et al.*, 2012; Kaida *et al.*, 2010). Furthermore, the U1 snRNP is also important to regulate promoter directionality and transcription (Almada *et al.*, 2013; Guiro & O'Reilly, 2015).

U1 snRNP complexes were purified and characterized from yeast and human. The U1 snRNP contains the U1 snRNA, Sm proteins, three U1 core proteins (U1-70K, U1-A and U1-C) and U1-specific accessory proteins, such as LUC7, PRP39 and PRP40. All these proteins are conserved in plants suggesting a U1 snRNP composition very similar to that of yeast and metazoans (Koncz *et al.*, 2012; Reddy *et al.*, 2013; Wang & Brendel, 2004). Interaction studies revealed that the U1 snRNP associates with serine-arginine (SR) proteins, indicating that the complex mechanisms for splicing site selection involve also non-snRNP proteins (Cho *et al.*, 2011; Golovkin & Reddy, 1998).

The function of the plant U1 snRNP is not well characterized. This might be due to the fact that in *Arabidopsis thaliana*, *U1-70K*, *U1-A* and *U1-C* are single copy genes and a complete knockout will probably cause lethality. On the other hand,

proteins such as PRP39, PRP40 and LUC7 come as small gene families, which will require the generation of multiple mutants for functional studies. Other factors, such as GEMIN2 or SRD2 are required for the functionality of all snRNPs, but not specifically for U1 function (Ohtani & Sugiyama, 2005; Schlaen *et al.*, 2015). Only a single U1 snRNP-specific gene, the *PRP39a* was characterized. *prp39a* mutants flower late due to increased expression of the flowering time regulator *FLOWERING LOCUS C* (*FLC*), but the mutant does not exhibit any additional developmental defects (Wang *et al.*, 2007). In a reverse genetic approach, *U1-70K* expression was specifically reduced in flowers by an antisense RNA and the resulting transgenic plants exhibit strong floral defects (Golovkin & Reddy, 2003). Despite evidences that U1 snRNP is essential for plant development, the functions of the U1 snRNP in regulating the transcriptome of plants are unknown.

Here, we report on the functional characterization of an *Arabidopsis* mutant impaired in U1 snRNP function. For this, we focused in this study on the U1 snRNP component LUC7, which we show to be essential for normal plant development. Our whole transcriptome analyses on *luc7* triple mutant show that impairments of LUC7s affect constitutive and alternative splicing. Surprisingly, our results reveal the existence of transcripts, in which terminal introns are preferentially retained in a LUC7-dependent manner and that these unspliced terminal introns cause a nuclear retention of the pre-mRNAs. Our findings are accompanied by the observation that splicing of first and last introns is preferentially regulated under stress conditions. Our results suggests that the plant U1 snRNP component LUC7 carries out a specialized function in the removal of terminal introns to prevent nuclear export of pre-mRNAs and this could be a mechanism to fine-tune gene expression under stress conditions.

Results

LUC7 proteins, a family of conserved nuclear zinc-finger / arginine-serine (RS) proteins, redundantly control plant development

LUC7 stands for *Lethal Unless CBC 7* and was first identified in a screen for synthetic lethality in a yeast strain lacking the cap-binding complex (CBC), which is involved in RNA processing (Fortes *et al.*, 1999b; Gonatopoulos-Pournatzis & Cowling, 2014; Sullivan & Howard, 2016). All LUC7 carry a C_3H and a C_2H_2-type zinc-finger, which are located in the conserved LUC7 domain. LUC7 proteins from higher eukaryotes usually contain an additional C-terminal Arginine/Serine-rich (RS) domain, which is known to be involved in mediating protein-protein interactions (Heim *et al.*, 2014; Puig *et al.*, 2007; Webby *et al.*, 2009). *Arabidopsis thaliana* encodes three *LUC7* genes (*AthLUC7A*, *AthLUC7B* and *AthLUC7RL*), which are separated in two clades: *LUC7A/B* and *LUC7RL* (Figure 1A and S1). AthLUC7RL is more similar to its yeast homologs and lacks a conserved stretch of 80 amino acids of unknown function present in AthLUC7A and AthLUC7B (Figure S1). A phylogenetic analysis revealed that algae contain a single *LUC7* gene belonging to the LUC7RL clade reinforcing the idea that LUC7RL proteins are closer to the ancestral LUC7 than LUC7A/B. In the moss *Physcomitrella* and in the fern *Selaginella* one can find proteins belonging to both clades, suggesting that the split into *LUC7RL* and *LUC7A/B* occurred early during the evolution of land plants.

In order to understand the function of the Arabidopsis U1 snRNP, we analyzed T-DNA insertion lines affecting *LUC7* genes (Figure 1B). Single and double *luc7* mutants were indistinguishable from wild-type plants (WT) (Figure S2). However, *luc7* triple mutant exhibit a wide range of developmental defects, including dwarfism and reduced apical dominance (Figure 1C-E). To test whether the impairment of *LUC7* functions were indeed responsible for the observed phenotypes, we

reintroduced a wild-type copy of *LUC7A*, *LUC7B* and *LUC7RL*, respectively, in the *luc7* triple mutant. Each of the *LUC7* genes was sufficient to restore the phenotype of the *luc7* triple mutant (Figure 1E). These results reveal that the phenotype observed in this mutant is attributable to the impairment of *LUC7* function and it suggests that *LUC7* genes act redundantly to control *Arabidopsis* growth and development.

LUC7 is a U1 snRNP component in plants

The composition of the U1 snRNP subcomplex is known in yeast and metazoans but not in plants (Koncz *et al.*, 2012; Will & Luhrmann, 2001). Therefore, we asked whether LUC7 is also an U1 component in Arabidopsis. Due to the fact that our genetic analyses of *luc7* mutants suggest that LUC7 proteins act largely redundant, we focused our further analyses mainly on a single LUC7 protein, LUC7A.

A protein that is part of the U1 complex is tightly associated with U1 specific components such as the U1 snRNA. To test whether LUC7 is found in a complex with the U1 snRNA, we performed RNA immunoprecipitation (RIP) experiments using a *luc7 triple* mutant carrying a functional *pLUC7A:LUC7A-YFP* rescue construct (Figure S3). LUC7A-YFP co-immunoprecipitated U1 snRNA, but not two unrelated, but abundant RNAs, U3 snoRNA and *ACTIN* mRNA (Figure 2A). Also small amounts of U2 snRNA associated with LUC7A-YFP, which is in agreement with the fact that U1 and U2 snRNP directly interact to form spliceosomal complex A (Figure 2A). These results strongly suggest that Arabidopsis LUC7 is a bona fide U1 snRNP component.

Next we analyzed the subcellular localization of LUC7A and its co-localization with a core U1 snRNP subunit. LUC7A localized to the nucleus, but not to the nucleolus in *Arabidopsis* plants containing the pLUC7A:LUC7A-YFP rescue construct (Figure 2B). In addition, LUC7A partially co-localized with U1-70K in the nucleoplasm

when transiently expressed in *Nicotiana benthamiana* (Figure 2C). Similar results were obtained for LUC7RL, the Arabidopsis LUC7 most distant in sequence to LUC7A (Figure 2C). Unlike U1-70K, LUC7A and LUC7RL did not form distinct speckles in the nucleus and were evenly localized in the nucleoplasm (Figure 2C).

To further test whether LUC7A associates *in planta* with known U1 snRNP core components, we purified LUC7A-containing complexes. For this, we used pLUC7A:LUC7A-YFP complemented lines and as controls wild-type plants and transgenic lines expressing free GFP. The mass spectrometry (MS) analysis revealed that LUC7 is indeed found in a complex with core U1 snRNP proteins U1-A and U1-70K (Figure 2D, Table S1). Furthermore, we detected peptides corresponding to the spliceosomal complex E components U2AF35 and U2AF65, further suggesting that LUC7 proteins are involved in very early steps of the splicing cycle (Figure 2D) Additional proteins known to be involved in splicing and general RNA metabolism including several serine-arginine (SR) proteins (SR30, SCL30A, SCL33), SR45 and SERRATE (SE) were found in LUC7A-containing complexes (Figure 2D). Interestingly, LUC7A was also associated with regulatory proteins, among them several kinases (Figure 2D, Table S1). Taken together, our data show that Arabidopsis LUC7 is an U1 snRNP protein and imply that LUC7 function is regulated by kinases.

***LUC7* effects on the Arabidopsis coding and non-coding transcriptome**

In order to identify misregulated and misspliced genes in luc7 mutants, we performed an RNA-sequencing (RNA-seq) analysis with three biological replicates. We decided to use 7 days old WT and luc7 triple mutant seedlings. At this age, luc7 triple mutant and WT seedlings are morphologically similar and therefore, changes in transcript levels and splicing patterns most likely reflect changes caused by LUC7s

impairments and are not due to different contribution of tissues caused by, for instance, a delay in development or/and different morphology (Figure S4).

An analysis of differentially expressed genes revealed that 840 genes are up- and 703 are downregulated in *luc7* mutants when compared to WT. The majority of genes that change expression were protein-coding genes (Figure 3A). Nevertheless, non-coding RNAs genes (ncRNAs) were significantly enriched among the *LUC7* affected genes ($p < 0.05$, hypergeometric test), although the overall number of ncRNA affected in *luc7* triple mutants is relatively small (Figure 3A, B). Previous studies implied that the U1 snRNP regulates microRNA (miRNA) biogenesis (Bielewicz *et al.*, 2013; Knop *et al.*, 2016; Schwab *et al.*, 2013; Stepien *et al.*, 2017). However, the expression of *MIRNA* genes was not affected in *luc7* triple mutants (Figure 3A). In addition, quantification of mature miRNA levels revealed that all tested miRNAs did not change abundance in *luc7* triple mutant (Figure 3C). These results show that *LUC7* genes affect the expression of protein-coding genes and a subset of ncRNAs, but are not involved in the miRNA pathway.

Arabidopsis *LUC7* function is important for constitutive and alternative splicing

Because LUC7 proteins are U1 snRNP components, we ask whether misspliced transcripts accumulate in the *luc7* triple mutant. In total, we identified 645 differential splicing events in *luc7* triple mutant compared to WT. We detected a large number of intron retention events (Figure 4A). RT-PCR experiments with oligonucleotides flanking selected intron retentions events confirmed the RNA-seq data (Figure 4B). These results suggest that lack of the U1 snRNP component LUC7 impairs intron recognition. Interestingly, we also identified a large number of exons that are included in the *luc7* triple mutant when compared to WT, as well as cases of

alternative 5' and 3' splice site selection (Figure 4A-F). Some of these affected splicing events generate transcript variants that did not exist in WT (e.g. *At2g32700* Figure 4C). On the contrary, in other cases the *luc7* triple mutant lacked specific mRNA isoforms, which exist in wild-type plants (e.g. *At1g10980*, *At4g32060*), or the ratio of two different isoforms was altered in *luc7* triple mutant when compared to WT (e.g. *At3g17310*, *At5g16715*, *At5g48150*, *At2g11000*) (Figure 4D-F). These results show that Arabidopsis U1 snRNP proteins LUC7 are involved in constitutive and alternative splicing.

Next, we checked whether splicing changes observed in *luc7* triple mutant are actually due to the loss of only a specific *LUC7* gene or whether *LUC7* genes act redundantly. To test this, we analyzed the splicing pattern of several mRNAs in *luc7* single, double and triple mutants. Some splicing defects were detectable even in *luc7* single mutants (Figure S5), but the degree of missplicing increased in *luc7* double and triple mutants suggesting that LUC7 proteins act additively on these introns (e.g. *At5g16715*). Some splicing defects occurred only in *luc7* triple mutants, implying that LUC7 proteins act redundantly to ensure splicing of these introns (e.g. *At1g60995*). Other splicing defects might more likely be due to the lack of LUC7A/B or LUC7RL. For instance, intron removal of *At2g42010* more strongly relied on *LUC7RL*, while removal of an intron in *At5g41220* preferentially depends on LUC7A/LUC7B (Figure S5). These findings suggest that Arabidopsis *LUC7* genes function redundantly, additively or specifically to ensure proper splicing.

LUC7 proteins are preferentially involved in the removal of terminal introns

In yeast, LUC7 connect the CBC with the U1 snRNP and this interaction is important for correct 5' splicing site selection (Fortes *et al.*, 1999a). In plants, the CBC associates with SE, which plays a role in the splicing of cap-proximal first introns in

Arabidopsis (Laubinger *et al.*, 2008; Raczynska *et al.*, 2010; Raczynska *et al.*, 2013). Thus, to test the relationship between LUC7 and the CBC/SE, we analyzed the splicing patterns of LUC7 dependent introns in *cbc* mutants (*cbp20* and *cbp80*) and *se-1* by RT-PCR. All tested introns retained in *luc7* triple mutant were correctly spliced in *cbc* and *se* mutants (Figure 5A). Conversely, first introns that are partially retained in *cbp20, cbp80* and *se-1* mutants were completely removed in the *luc7* triple mutant (Figure 5B). These observations suggest that the functions of LUC7 and CBC/SE in splicing of the selected introns do not overlap.

Next, we asked whether LUC7 has a preference for promoting splicing of cap-proximal first introns as it has the CBC/SE complex. We classified retained introns in *luc7* triple mutants according to their position within the gene (first, middle or last introns). Only genes with at least 3 introns were considered for this analysis. We found a significant increase in retained last introns, but not first introns, in *luc7* triple mutants (Figure 5C). Retention of terminal introns in *luc7* triple mutants was confirmed by RT-PCR analysis (Figure 5D). Although the total number of retained introns is higher in middle introns, the relative amount of retained middle introns in *luc7* triple mutant was significantly reduced (Figure 5C). In summary, our data revealed that (i) CBC/SE acts independently of LUC7 in splicing of cap-proximal introns and that (ii) LUC7 proteins play an important role for the removal of certain terminal introns.

Splicing of LUC7-dependent terminal introns occurs independently of polyadenylation

The removal of terminal introns in yeast and metazoans can be tightly linked to mRNA 3'end formation (Bentley, 2014; Cooke & Alwine, 2002; Cooke *et al.*, 1999; Nesic & Maquat, 1994; Wong *et al.*, 2016). In some cases, polyadenylation precedes

splicing of terminal introns and the polyadenylation machinery plays an important role in promoting splicing efficiency of terminal introns (Rigo & Martinson, 2008, 2009). To investigate a putative link in between splicing of LUC7-dependent terminal introns and polyadenylation, we isolated poly(A)+ and poly(A)- RNAs from WT and *luc7* triple mutant and analyzed the splicing patterns of several LUC7-dependent terminal introns (Figure 6A).

In WT, we found that unspliced isoforms of *At5g49840* and *At1g01860* accumulated in poly(A)+ RNA fractions indicating that some polyadenylated mRNAs still contained the LUC7 dependent terminal intron (Figure 6A). On the other hand, the terminal introns of *At2g41560*, *At1g70480* and *At5g41220* were efficiently removed in polyadenylated transcripts and high amounts of spliced mRNAs accumulated in poly(A)- fractions. These observations suggest that these terminal introns were efficiently removed before addition of the polyA-tail (Figure 6A). From these results we conclude that LUC7-dependent introns can be removed before or after polyadenylation. When we compared the splicing patterns in WT and *luc7* triple mutant, we found that the splicing efficiency of some introns in *luc7* mutants was already reduced in poly(A)- fractions (e.g. *At1g01860*). The splicing efficiency of other terminal introns was affected only in poly(A)+ fractions (e.g. *At1g70480*). These results further suggest that certain mRNAs require LUC7 function for efficient intron removal before, other mRNAs after the addition of the poly(A)-tail.

mRNAs harboring unspliced LUC7-dependent terminal introns remain in the nucleus and escape NMD

When introns are retained, the resulting mRNA can contain a premature stop codon and a long 3'UTR, which are hallmarks of NMD targets (Drechsel *et al.*, 2013; Kalyna *et al.*, 2012; Shaul, 2015). To check whether mRNAs containing a retained LUC7-

dependent terminal intron are NMD substrates, we analyzed their splicing patterns in two mutants impaired in NMD, *lba-1* and *upf3-1*. If unspliced isoforms were indeed NMD targets, we would expect their abundance to be increased in NMD mutants. Interestingly, we did not observe any change between WT and *upf* mutants (Figure 6B). Thus, we conclude that retained LUC7-dependent terminal introns do not trigger degradation via the NMD pathway.

NMD occurs in the cytoplasm and RNAs can escape NMD by not being transported from the nucleus to the cytosol (Gohring *et al.*, 2014). We therefore checked in which cellular compartment mRNAs with spliced and unspliced LUC7-dependent terminal introns accumulate. To do this, we isolated total, nuclear and cytosolic fractions from wild-type and *luc7* triple mutant plants and performed RT-PCR analyses (Figure 6C). Spliced mRNA isoforms accumulated in the cytosol, whereas mRNAs containing the unspliced terminal introns were found in nuclear fractions (Figure 6C). These results indicate that retention of terminal introns correlates with trapping mRNAs in the nucleus and suggest that splicing of LUC7-dependent terminal introns is essential for mRNA transport to the cytosol.

Splicing of LUC7-dependent terminal introns can be modulated by stress

Our results revealed that a subset of terminal introns requires LUC7 proteins for efficient splicing, and that splicing of these introns is a prerequisite for nuclear export. This mechanism could serve as a nuclear quality control step to prevent that unspliced mRNAs are exported prematurely. Interestingly, a GO analysis of genes containing LUC7 dependent terminal introns indicated an enrichment for stress related genes (Figure S6). This prompted us to speculate that nuclear retention of mRNAs could be exploited as a regulatory mechanism to fine-tune gene expression under stress conditions. To test this hypothesis, we decided to check the splicing of

LUC7-dependent terminal introns in WT under stress condition. We chose cold stress because it was suggested that U1 snRNP functionality is impaired under cold condition (Schlaen *et al.*, 2015). To quantify the amount of unspliced isoform in cold condition, we designed qPCR-primers specific to unspliced isoform and total RNA and calculate the ratio unspliced/total of LUC7-dependent terminal introns in four genes (*At1g70480*, *At2g41560*, *At5g44290* and *At5g41220*). Three of these genes significantly accumulated unspliced transcripts in responses to cold treatment demonstrating that cold stress modulates the splicing efficiency of LUC7-dependent terminal introns (Figure 7A).

Cold and salt stress preferentially affects splicing of first and terminal introns

To investigate whether terminal intron retention is a general feature of plant stress transcriptomes, we performed a global analysis of intron retention under stress condition. We made use of publically available RNA-seq data sets from Arabidopsis plants treated with cold and salt stress (Ding *et al.*, 2014; Schlaen *et al.*, 2015). We identified stress-regulated intron retention events, filtered for genes containing three or more introns, and assigned retained introns based on their position within the transcript (first, middle or terminal intron). The relative amount of first, middle and terminal introns among retained introns under stress conditions was compared to the relative amount of introns among all expressed genes. We found that up- and down-regulated intron retention events during cold and salt stress were significantly enriched in first and terminal introns (Figure 7B-C, Figure S7). These results show that first and terminal introns are preferentially subjected to alternative splicing regulation under stress conditions. We observed a higher proportion of retained first and last introns in two different stresses condition and varying stress intensities and durations (Figure 7B-C, Figure S7). Thus, our data strongly suggest that regulated

347 splicing of first and terminal introns widely contributes to shaping plant
348 transcriptomes under stress conditions.

Discussion

Functions of the Arabidopsis U1 snRNP component LUC7

For this study, we generated an Arabidopsis mutant deficient in the U1 snRNP component LUC7 and dissected the genome-wide effects of LUC7 impairment on the Arabidopsis transcriptome. Our results show that LUC7 proteins are *bona-fide* U1 components acting mainly redundantly. The reduction of U1 function in the *luc7* triple mutant affects constitutive splicing. A large number of introns are retained in *luc7* triple mutant, suggesting that without a proper recognition of the 5'ss, splicing of the affected introns is impaired. Our results also show that exon-skipping events are impaired in *luc7* triple mutant, revealing that a functional plant U1 snRNP is essential for exon definition. In addition, we show that *luc7* triple mutant affect alternative splicing also by influencing events of alternative 5' and 3' splice site. This implies that the U1 snRNP does not only affect 5' splice site usage, it might also indirectly regulate usage of 3' splice sites via its interaction with U2AFs and the U2 snRNP (Hoffman & Grabowski, 1992; Shao *et al.*, 2012). The functions of LUC7 proteins on the Arabidopsis transcriptome are likely to be underestimated, because misspliced mRNAs in *luc7* mutants might contain hallmarks of NMD and are therefore rapidly turned over and escape detection. Analysis of *luc7* mutants combined with mutations in NMD factors would help to uncover the full set of splicing events affected by LUC7. One has also to consider that U1 snRNP independent splicing has been described in animals, indicating that not all introns require the U1 complex for efficient intron removal (Fukumura *et al.*, 2009). The degree of U1-independent splicing in plants remains to be elucidated.

Duplications among genes encoding for U1 snRNP proteins, such as the LUC7 genes, might suggest that U1 accessory genes have undergone sub- and neofunctionalization. Furthermore, the Arabidopsis genome encodes 14 potential U1

375 snRNAs, which slightly differ in sequence (Wang & Brendel, 2004). Therefore, the
376 plant U1 snRNP presumably does not exist as a single complex, but exists as
377 different sub-complexes exhibiting distinct specificities and functions. In metazoans,
378 the existence of at least four different U1 snRNP subcomplex has been suggested
379 (Guiro & O'Reilly, 2015; Hernandez *et al.*, 2009). Specific combinations of plant U1
380 protein family members and U1 snRNAs could generate an even higher number of
381 such U1 subcomplexes, which could be responsible for specific splicing events. Our
382 results show that LUC7 can act redundantly, but can also fulfill specific functions,
383 suggesting that LUC7 complexes specifically act on certain pre-mRNAs. In this
384 regard, it is important to note that an additional short stretch of aminoacids separates
385 the two zinc-finger RNA binding domains in LUCA and LUC7B. Changing the space
386 in between RNA binding domains affects substrate specificities and could explain
387 different functions among LUC7 proteins (Chen & Varani, 2013).

388 Interestingly, *luc7* triple mutant showed a significant higher retention rate of
389 terminal introns compared to first or middle introns. This was surprising because
390 LUC7 was initially found to act in concert with the CBC, a complex involved in the
391 removal of cap-proximal first introns, but not of last introns (Lewis *et al.*, 1996). Often,
392 the removal of terminal introns is coupled to polyadenylation (Cooke & Alwine, 2002;
393 Cooke *et al.*, 1999; Rigo & Martinson, 2008). However, our analysis suggests that
394 splicing of LUC7-dependent terminal introns occurs in some cases independently of
395 polyadenylation because some introns were removed after the addition of the poly(A)
396 tail. Interestingly, we detected the pre-mRNA cleavage/polyadenylation factor
397 AT4G25550 as part of LUC7A complexes, suggesting that an interaction between
398 LUC7 and 3' end formation complexes may contribute to the specific functions of
399 LUC7 in terminal intron splicing.
400

401 **Potential functions of terminal intron retention in plants**

402 Targeting a transcript to NMD can be triggered by two mechanisms: For the first one,
403 it requires the deposition of an exon-junction-complex (EJC) downstream of the
404 retained intron which causes a premature stop codon. For the second one, mRNAs
405 containing a long 3'UTR (≥300 - 350 nt) are degraded via the NMD pathway
406 (Drechsel *et al.*, 2013; Kalyna *et al.*, 2012; Kervestin & Jacobson, 2012; Shaul,
407 2015). Retained terminal introns are then special because their detection solely relies
408 on the length of the 3'UTR. Transport of mRNAs containing unspliced terminal
409 introns, especially those which feature a relatively short 3'UTR, can be detrimental
410 and has to be tightly controlled. Interestingly, we found that splicing of LUC7-
411 dependent terminal introns is required for transport of mRNAs from the cytosol to the
412 nucleus. The fact that we cannot detect unspliced transcript in the cytosol suggests a
413 nuclear retention mechanism for such mRNAs. One possibility is that LUC7
414 dependent terminal introns might contain binding sites for specific trans-regulatory
415 factors that upon binding inhibit export. Polypyrimidine tract-binding protein 1 (PTB1)
416 is a prime candidate for such a trans-regulatory protein, because binding of PTB1 to
417 introns represses nuclear export of certain RNAs (Roy *et al.*, 2013; Yap *et al.*, 2012).

418 Nuclear retention of unspliced mRNAs is not limited to terminal introns and
419 might be a much more general mechanism to escape NMD and to regulate gene
420 expression (Wong *et al.*, 2016). In plants, some specific transcript isoforms have
421 been detected only in the nucleus, but not in the cytosol (Gohring *et al.*, 2014). Also
422 in metazoans, intron retention might have a more general role in regulating gene
423 expression (Naro *et al.*, 2017; Pimentel *et al.*, 2016; Yap *et al.*, 2012). The so-called
424 detained introns are evolutionary conserved, NMD insensitive and retained in the
425 nucleus (Boutz *et al.*, 2015). The functional importance of intron retention was also

426 suggested in the fern *Marsilea vestita*, in which many mRNAs contain introns that are
427 only spliced shortly before gametophyte development (Boothby *et al.*, 2013).

428 We found that LUC7-dependent splicing of terminal introns can be modulated
429 by cold stress. Because retention of these introns causes nuclear trapping, it is
430 prompting to speculate that environmental cues affect splicing and nuclear retention
431 of mRNAs. Such a mechanism would regulate the amount of translatable mRNAs in
432 the cytosol in a cost-efficient and rapid manner. Because the RS domains of LUC7
433 proteins are phosphorylated and we identified several kinases as LUC7A interactors,
434 stress-induced changes in phosphorylation might play a role in regulating LUC7
435 proteins and U1 snRNP function (Durek *et al.*, 2010; Heazlewood *et al.*, 2008). This
436 idea is supported by the observation that stress signaling triggered by the
437 phytohormone ABA results in phosphorylation of several splicing factors (Umezawa
438 *et al.*, 2013; Wang *et al.*, 2013)

439 In general, our analysis suggests that first and terminal introns are hotspots
440 for regulated splicing under stress conditions. First and last introns are close to the 5'
441 cap and the polyA tail, respectively. These positions might offer more possibilities for
442 splicing regulation via crosstalk between the splicing machinery and factors involved
443 in capping and cleavage/polyadenylation. We are just at the beginning of
444 understanding the importance of intron retention in gene regulation. Future
445 identification of cis- and trans-regulatory factors involved in the regulation of terminal
446 intron splicing will shed additional light on this layer of gene expression.

447

448 **Material and Methods**

449

450 **Plant material and growth conditions**

451 All mutants were in the Columbia-0 (Col-0) background. *luc7a-1* (SAIL_596_H02)
452 and *luc7a-2* (SAIL_776_F02), *luc7b-1* (SALK_144681), *luc7rl-1* (SALK_077718) and
453 *luc7rl-2* (SALK_130892C) were isolated by PCR-based genotyping (Table S2). *luc7*
454 double and triple mutants were generate by crossing individual mutants. All other
455 mutants used in this study (*abh1-285*, *cbp20-1*, *se*-1, *lba-1* and *upf3-1*) were
456 described elsewhere (Hori & Watanabe, 2005; Laubinger *et al.*, 2008; Papp *et al.*,
457 2004; Prigge & Wagner, 2001; Yoine *et al.*, 2006). The line expressing GFP was
458 generated using the vector pBinarGFP and was kindly provided by Dr. Andreas
459 Wachter (Wachter *et al.*, 2007). For complementation analyses, *pLUC7A:LUC7A-*
460 *FLAG*, *pLUC7B:LUC7B-FLAG*, *pLUC7RL:LUC7RL-FLAG* and *pLUC7A:LUC7A-YFP*
461 constructs were introduced in *luc7* triple mutant by Agrobacterium-mediated
462 transformation (Clough & Bent, 1998). All plants were grown on soil in long days
463 conditions (16-h light/8-h dark) at 20°C/18°C day/night. The size of *luc7* mutants was
464 assessed by measuring the longest rosette leaf after 21 days. For all molecular
465 studies, seeds were surface-sterilized, plated on 1/2 MS medium with 0.8%
466 phytoagar and grown for 7 days in continuous light at 22°C. For the cold treatment,
467 plates with *Arabidopsis* seedlings were transferred to ice-water for 60 min.

468

469 **Plasmid constructions and transient expression analyses**

470 For the expression of C-terminal FLAG- and YFP-tagged LUC7 proteins expressed
471 from their endogenous regulatory elements, 2100 bp, 4120 bp and 2106 bp upstream
472 of the ATG start codon of *LUC7A*, *LUC7B* and *LUC7RL*, respectively, to the last
473 coding nucleotide were PCR-amplified and subcloned in pCR8/GW/TOPO®

(Invitrogen). Oligonucleotides are listed in Table S2). Entry clones were recombined with pGWB10 and pGWB540 using Gateway LR clonase II (Invitrogen) to generate binary plasmids containing *pLUC7A:LUC7A-FLAG*, *pLUC7B:LUC7B-FLAG*, *pLUC7RL:LUC7RL-FLAG* and *pLUC7A:LUC7A-YFP*. For the co-localization studies, entry vector containing the coding sequence of U1-70k was recombined with pGWB654 for the expression of *p35S:U1-70k-mRFP* (Nakagawa *et al.*, 2007). Agrobacterium-mediated transient transformation of *Nicotiana benthamiana* plants was conducted as following. Overnight Agrobacterium culture were diluted in 1:6 and grown for 4 hours at 28°C. After centrifugation, pellets were ressuspended in infiltration medium (10mM $MgCl_2$, 10mM MES-KOH pH 5.6, 100 µM Acetosyringone). The OD 600nm was adjusted to 0.6-0.8 and samples were mixed when required. *N. benthamiana* were infiltrated and subcellular localization was checked after 3 days. Subcellular localization of fluorescent proteins was analyzed by confocal microscopy using a Leica TCS SP8.

Phylogenetic analysis

AthLUC7A (AT3G03340) protein sequence was analyzed in Interpro (https://www.ebi.ac.uk/interpro/) to retrieve the Interpro ID for the conserved Luc7-related domain (IPR004882). The sequence for *Saccharomyces cerevisiae* (strain ATCC 204508_S288c) was obtained in Interpro. Plants sequences were extracted using BioMart selecting for the protein domain IPR004882 on Ensembl Plants (http://plants.ensembl.org/). The following genomes were included in our analyses: *Amborella trichopoda* (AMTR1.0 (2014-01-AGD)); *Arabidopsis thaliana* (TAIR10 (2010-09-TAIR10)); *Brachypodium distachyon* (v1.0); *Chlamydomonas reinhardtii* (v3.1 (2007-11-ENA)); *Physcomitrella patens* (ASM242v1 (2011-03-Phypa1.6)); *Selaginella moellendorffii* (v1.0 (2011-05-ENA)); *Oryza sativa Japonica* (IRGSP-1.0);

and *Ostreococcus lucimarinus* genes (ASM9206v1). The phylogenetic analysis was performed in Seaview (Version 4.6.1) using Muscle for sequence alignment. Maximum likehood (PhYML) was employed with 1000 bootstraps (Gouy *et al.*, 2010).

RNA extractions, RT-PCR and qRT-PCR

RNAs extractions were performed with Direct-zol™ RNA MiniPrep Kit (Zymo Research). Total RNAs were treated with DNAse I and cDNA synthesis carried out with RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) using usually oligo dT primers or a mixture of hexamer and miRNA-specific stem-loop primers (Table S2). Standard PCRs for the splicing analysis were performed with DreamTaq DNA Polymerase (Thermo Scientific). Quantitative RT-PCR (qRT-PCR) was performed using the Maxima SYBR Green (Thermo Scientific) in a Bio-Rad CFX 384. For all qPCR-primers, primer efficiencies were determined by a serial dilution of cDNA template. The relative expressions were calculated using the 2ˆ(-ΔΔCT) method with *PP2A* or *ACTIN* as controls. For the qRT-PCR to detect splicing ratio changes under cold condition, the ratio 2ˆ(-ΔCTunspliced)/2ˆ(-ΔCTtotal RNA) was calculated separately for each replicate and t-test was performed before calculating the relative to WT. Primers used for PCR and qRT-PCR are listed Table S2. For RNA-sequencing analysis, polyA RNAs were enriched from 4 µg of total RNAs using NEBNext Oligo $d(T)_{25}$ Magnetic Beads (New England Biolabs). The libraries were prepared using ScriptSeq™ Plant Leaf kit (Epicentre) following the manufacturer's instruction. Single end sequencing was performed on an Illumina HiSeq2000. Sequencing data were deposited at Gene Expression Omnibus under accession number GSE98779.

Isolation of poly(A)- and poly(A)+ fractions

RNAs were extracted from seedlings and treated with DNAseI as described above. After the DNAse treatment, samples were cleaned up using RNA Clean & Concentrator-5 (Zymo Research).Two replicates were performed 1.5 - 3 µg of total RNA was used for the fractionation. The poly(A) fractions were prepared using NEBNext Oligo $d(T)_{25}$ Magnetic Beads (New England Biolabs). Each fraction was purified twice. For the poly A(-) fractions, large sample volumes were concentrated with RNA Clean & Concentrator-5. cDNA synthesis was carried out with random hexamer primers as described above.

Subcellular fractionation

Two grams of seedlings were ground in N_2 liquid and resuspended in 4 ml of Honda buffer (0.44 M sucrose, 1.25% Ficoll 400, 2.5% Dextran T40, 20 mM HEPES KOH pH 7.4, 10 mM $MgCl_2$, 0.5% Triton X-100, 5 mM DTT, 1 mM PMSF, protease inhibitor cocktail [ROCHE] supplemented with 40U/ml of Ribolock®). The homogenate was filtered through 2 layer of Miracloth, which was washed with 1ml of Honda Buffer. From the filtrate, 300 µl was removed as “total” fraction and kept on ice. Filtrates were centrifuged at 1,500 *g* for 10 min, 4°C for pelleting nuclei and supernatants were transferred to a new tube. Supernatants were centrifuged at 13 000 x g, 4 °C, 15 min and 300 µl were kept on ice as cytoplasmic fraction. Nuclei pellets were washed five times in 1 ml of Honda buffer (supplemented with 8U/ml of Ribolock®, centrifugation at 1,800 *g* for 5 min. The final pellet was resuspended in 300 µl of Honda buffer. To all the fractions (total, cytoplasmic and nuclei), 900 µl of TRI Reagent (Sigma) was added. After homogenization, 180 µl of chloroform was added and samples were incubated at room temperature for 10 min. After centrifugation at 10 0000 rpm for 20 min, 4°C, the aqueous phase were transferred to a new tube and RNA extracted with Direct-zol™ RNA MiniPrep Kit (Zymo Research).

The organic phase was collected and proteins were isolated according to manufacturer's instructions (TRI Reagent). cDNA synthesis with random primes was performed as above. Proteins extracted were analyzed by standard western blot techniques using the following antibodies: H3 (~ 17KDa / ab 1791, Abcam) and 60S ribosomal (~ 23,7-29KDa / L13, Agrisera).

RNA immunoprecipitation

RNA immunoprecipitation (RIP) using WT and a *pLUC7A:LUC7A-eYFP* rescue line was performed as described elsewhere with minor modifications (Rowley *et al.*, 2013; Xing *et al.*, 2015). Isolated nuclei were sonicated in nuclear lysis buffer in a Covaris E220 (Duty Cycle: 20%; Peak intensity: 140; Cycles per Burst: 200; Cycle time: 3'). RNAs were extracted using RNeasy Plant Mini Kit (QIAGEN) following the manufacturer's instructions. The RNA were treated with DNAseI (Thermo Scientific) and samples were split in half for the (-)RT reaction. cDNA synthesis were perform with SuperScript™ III Reverse Transcriptase (Invitrogen). qRT-PCRs were performed with QuantiNova™ SYBRR Green PCR (QIAGEN).

RNA-seq libraries: Mapping, differential expression analysis and splicing analysis

RNA-seq reads for each replicate were aligned against the *Arabidopsis thaliana* reference sequence (TAIR10) using tophat (v2.0.10, -p2, -a 10, -g 10, -N 10, --read-edit-dist 10, --library-type fr-secondstrand, --segment-length 31, -G TAIR10.gff). Next, cufflinks (version 2.2.1) was used to extract FPKM counts for each expressed transcript generating a new annotation file (transcripts.gtf), where the coordinates of each expressed transcript can be found. Cuffcompare (version 2.2.1) was then used to generate a non-redundant annotation file containing all reference transcripts in

addition to new transcripts expressed in at least one of the nine samples (cuffcmp.combined.gtf). The differential expression analysis was performed with cuffdiff (version 2.2.1) between wt/*luc7* triple using the annotation file generated by cuffcompare (FDR<2 and FC>0,05). For the splicing analysis, the same alignment files generated by tophat and annotation files generated by cuffcompare (cuffcmp.combined.gtf) were used as input for MATS (version 3.0.8) in order to test for differentially spliced transcripts (Shen *et al.*, 2014).

Global analysis of intron regulation under stress conditions

For the analyses of intron retention under stress conditions, published data sets were analyzed (accession numbers SRP035234 and SRP049993) (Ding *et al.*, 2014; Schlaen *et al.*, 2015). Reads were aligned to the *Arabidopsis thaliana* Ensembl3 33 genome and to the annotation GTF file (ftp://ftp.ensemblgenomes.org/pub/release-33/plants/fasta/arabidopsis_thaliana/dna/Arabidopsis_thaliana.TAIR10.dna.toplevel.fa.gz, ftp://ftp.ensemblgenomes.org/pub/release-33/plants/gtf/arabidopsis_thaliana/Arabidopsis_thaliana.TAIR10.33.gtf.gz) using TopHat2 applying following parameters: tophat2 -p 10 -i 10 -I 1000 -G Arabidopsis_thaliana.TAIR10.33.gtf Arabidopsis_thaliana.TAIR10.dna.toplevel.fa. After alignment, mock-treated samples were used to generate an expressed background for the respective dataset using featureCounts from the Rsubread package (Kersey *et al.*, 2016; Kim *et al.*, 2013; Liao *et al.*, 2013). Read numbers and gene lengths per genes (featureCounts -T 6 -R -p -F GTF -J -G Arabidopsis_thaliana.TAIR10.dna.toplevel.fa -a Arabidopsis_thaliana.TAIR10.33.gtf) were collected and TPM values were calculated using an in-house script. Log_2 transformed values of expressed genes were visualized with ggplot2 version 2.1.0,

and based on the density plot, threshold of expressed genes was defined as $TPM_{expressed} > 0.6$.

Intron retention events were identified using rMATS with the following parameters: python RNASeq-MATS.py -b1 untreated.bam -b2 treated.bam -gtf Arabidopsis_thaliana.TAIR10.33.gtf -o output_dir -t paired -len 101 (Shen *et al.*, 2014). After filtering the outputs (p value < 0.05, FDR < 0.05), we categorize introns based on their position and annotation. In case of a few ambiguous hits, they were manually recategorized. For the categorization of introns in first, middle and last introns, we used the GTF annotation file (Ensembl 33) and selected genes with 3 or more introns and TPM > 0.6. The intron distribution of all expressed genes served as a background reference, to which the distribution of retained introns under stress conditions was compared. To test for significance of changes in intron distribution, Fisher's exact test was employed since we assumed a normal distribution.

GO Analysis

GO analysis was performed in Bar Utoronto (http://bar.utoronto.ca/ntools/cgi-bin/ntools_classification_superviewer.cgi).

Protein complex purification and mass spectrometry (MS) Analysis

LUC7A immunoprecipitation was performed using a complemented line *pLUC7A:LUC7A-eYFP* (line 20.3.1) and a transgenic *p35S:GFP* and WT as controls. Four independent biological replicates were performed. Seedlings (4 g) were ground in N_2 liquid and respuspended in 1 volumes of extraction buffer (50 mM Tris-Cl pH 7.5, 100 mM NaCl, 0,5% Triton X-100, 5% Glycerol, 1mM PMSF, 100 µM MG132, Complete Protease Inhibitor Cocktail EDTA-free [Roche] and Plant specific protease Inhibitor, Sigma P9599). After thawing, samples were incubated on ice for 30 min,

centrifuged at 3220 rcf for 30 min at 4°C and filtrated with two layers of Miracloth. For each immunoprecipitation, 20 µl of GFP-trap (Chromotek) was washed twice with 1 ml of washing buffer (50 mM Tris-Cl pH 7.5, 100 mM NaCl, 0.2% Triton X-100) and once with 0.5 ml of IP buffer. For each replicate, the same amount of plant extracts (~5ml) were incubated with GFP-trap and incubated on a rotating wheel at 4°C for 3 hours. Samples were centrifuged at 800-2000 rcf for 1-2 min and the supernatant discarded. GFP-beads were resuspended in 1 ml of washing buffer, transferred into a new tube and washed 4 to 5 times. Then, beads were ressuspended in ~40 µl of 2x Laemmili Buffer and incubated at 80°C for 10 min. Short gel purifications (SDS-PAGE) were performed and gels slices were digested with Trypsin. LC-MS/MS analyses were performed in two mass spectrometer. Samples from R10 to R14 were analyzed on a Proxeon Easy-nLC coupled to Orbitrap Elite (method: 90min, Top10, HCD). Samples from R15 to R17 were analysed on a Proxeon Easy-nLC coupled to OrbitrapXL (method: 90min, Top10, CID). Samples from R18 to R20 analysis on a Proxeon Easy-nLC coupled to OrbitrapXL (method: 130min, Top10, CID). All the replicates were processed together on MaxQuant software (Version 1.5.2.8. with integrated Andromeda Peptide search engine) with a setting of 1% FDR and the spectra were searched against an *Arabidopsis thaliana* Uniprot database (UP000006548_3702_complete_20151023.fasta). All peptides identified are listed in Supplementary Table S1 and raw data were deposited publically (accession PXD006127).

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft DFG (to S.L., LA2633-4/1), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

(CAPES - Brazil) for doctoral fellowship (to M.d.F.A.), the Max Planck Society (to K.S.), the Max Planck Society Chemical Genomics Centre (CGC) through its supporting companies AstraZeneca, Bayer CropScience, Bayer Healthcare, Boehringer-Ingelheim and Merck (to S.L). We are grateful to Andreas Wachter (ZMBP, University of Tuebingen, Germany) and members of the lab for critical reading of the manuscript, Christa Lanz for her invaluable help with Illumina sequencing, and Johanna Schröter and her team for excellent care of our plants, Anja Hoffmann for excellent technical assistance, and Andreas Wachter (ZMBP, University of Tuebingen, Germany), Tsuyoshi Nakagawa (Department of Molecular and Functional Genomics, Center for Integrated Research in Science, Shimane University, Matsue, Japan) and the Notthingham Arabidopsis Stock Centre for providing seeds and DNA constructs.

Competing interests

The authors declare no competing interests.

Figure legends

Figure 1: *Arabidopsis* LUC7 proteins redundantly control plant development

A: Phylogenetic analysis of LUC7 proteins in the plant kingdom using *Saccharomyces cerevisiae* as an external group.

B: Exon/intron structure of *Arabidopsis thaliana LUC7A*, *LUC7B* and *LUC7RL*. Dotted lines indicate the positions of T-DNA insertions.

C: WT and *luc7* triple mutant plants 21 days (left) and 45 days (right) after germination.

D: Length of the longest rosette leaf of 21 days-old WT, *luc7* single, double and triple mutant plants growing under long day conditions. Leaves of 10-15 individual plants were measured. Dots indicate individual data points. Error bars denote the SEM.

E: Complementation of *luc7a,b,rl* mutants by *LUC7A*, *LUC7B* and *LUC7RL* genomic rescue constructs. Transformation of an "empty" binary vector served as a control. Two independent transgenic lines for each construct are shown.

Figure 2: *Arabidopsis* LUC7 is an U1 snRNP component

A: RNA immunoprecipitation using a *pLUC7A:LUC7A-YFP, luc7a,b,rl* complemented line. Proteins were immunoprecipitated using GFP-specific affinity matrix and RNAs were extracted from the input and the immunoprecipitated. U1, U2, U3 snRNAs and *ACTIN* RNA were quantified using qRT-PCR. Enrichment of the respective RNA in *LUC7A:LUC7A-YFP luc7a,b,rl* transgenic lines was calculated over WT (negative control). Error bars denote the range of two biological replicates.

B: Subcellular localization of LUC7A in *pLUC7A:LUC7A-YFP luc7a,b,rl* in Arabidopsis transgenic plants. Roots of 9 day-old seedlings were analyzed using confocal microscopy. Scale bar indicates 25 µm.

C: U1-70K-mRFP and LUC7A-YFP or LUC7RL-YFP proteins were transiently expressed in *N. benthamiana*. The subcellular localization of mRFP and YFP fusion proteins was analyzed using confocal microscopy. Scale bars indicate 10 µm and 25 µm for upper and lower panel, respectively.

D: List of some LUC7-associated proteins identified by MS.

Figure 3: Mutations in *LUC7* results in misexpression of protein-coding and non-coding genes, but not in *MIRNA* genes

A: Differentially expressed genes in *luc7a,b,rl* mutant compared to WT.

B,C: qRT-PCR analysis of selected ncRNA **(B)** and miRNAs **(C)** in WT and *luc7a,b,rl*. Error bars denote the SEM (n=3).

Figure 4: Global analysis of splicing defects in *luc7* triple mutant.

A: Classification of splicing events changes in *luc7* triple mutant compared to WT.

B-F: Coverage plots and RT-PCR validation experiments for selected splicing events in WT and *luc7* triple mutant. Genomic DNA (gDNA) or water (-) served as a control. Primers positions are indicated with gray arrows. IR, intron retention; ES, exon skipping; Alt.3'SS, alternative 3'splicing site; Alt.5'SS, alternative 5'splicing site.

Figure 5: LUC7 proteins have a pronounced effect on terminal intron splicing

A: RT-PCR analysis of LUC7-dependent introns in WT, *luc7* triple mutant, *cbp80*, *cbp20* and *se-1* mutants.

B: RT-PCR analysis of CBC/SE-dependent introns in WT, *luc7* triple mutant, *cbp80*, *cbp20* and *se-1* mutants.

C: Classification of intron retention according to the intron position (first, middle, or last). Only genes with 3 or more introns were considered for this analysis. Terminal introns are significantly enriched among introns retained in *luc7* triple mutants (Fischer test; two-side $p < 0.0016$).

D: RT-PCR analysis of genes carrying retained terminal introns in WT and *luc7* triple mutants.

Figure 6: mRNAs containing retained LUC7-dependent terminal introns are NMD-insensitive and remain nuclear.

A: Splicing patterns of mRNAs carrying LUC7-dependent terminal introns in poly(A)+ and poly(A)- fractions.

B: RT-PCR analysis of LUC7-dependent terminal introns in WT and NMD mutants (*lba1* and *upf3-1*).

C: Splicing patterns of mRNAs isolated from total (T), cytosolic (C) and nuclear (N) fractions.

Figure 7: Splicing of LUC7 dependent terminal introns can be modulated by stress.

A: Seven days old WT seedlings were exposed to cold for 60 min. Splicing ratios (unspliced/total RNA) of four genes featuring a LUC7-dependent terminal intron was analyzed by qPCR. T-test was performed before calculating the relative to WT (* p=0,0208, ** p =0,0095, *** p = 0.004).

B-C: Classification of intron retention events in cold- (B) and salt-treated plants (C) The relative amount of retained/removed first, middle, and last intron in stress samples was compared to the relative amount of first, middle, and last intron among all expressed genes (normal distribution). Only genes with 3 or more introns were considered for this analysis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

Literature

Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B., & Sharp, P. A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature, 499*(7458), 360-363. doi: 10.1038/nature12349

Bentley, D. L. (2014). Coupling mRNA processing with transcription in time and space. *Nat Rev Genet, 15*(3), 163-175. doi: 10.1038/nrg3662

Berg, M. G., Singh, L. N., Younis, I., Liu, Q., Pinto, A. M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L., & Dreyfuss, G. (2012). U1 snRNP determines mRNA length and regulates isoform expression. *Cell, 150*(1), 53-64. doi: 10.1016/j.cell.2012.05.029

Bielewicz, D., Kalak, M., Kalyna, M., Windels, D., Barta, A., Vazquez, F., Szweykowska-Kulinska, Z., & Jarmolowski, A. (2013). Introns of plant pri-miRNAs enhance miRNA biogenesis. *EMBO Rep, 14*(7), 622-628. doi: 10.1038/embor.2013.62

Boothby, T. C., Zipper, R. S., van der Weele, C. M., & Wolniak, S. M. (2013). Removal of retained introns regulates translation in the rapidly developing gametophyte of Marsilea vestita. *Dev Cell, 24*(5), 517-529. doi: 10.1016/j.devcel.2013.01.015

Boutz, P. L., Bhutkar, A., & Sharp, P. A. (2015). Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev, 29*(1), 63-80. doi: 10.1101/gad.247361.114

Carvalho, R. F., Feijao, C. V., & Duque, P. (2013). On the physiological significance of alternative splicing events in higher plants. *Protoplasma, 250*(3), 639-650. doi: 10.1007/s00709-012-0448-9

Chen, Y., & Varani, G. (2013). Engineering RNA-binding proteins for biology. *FEBS J, 280*(16), 3734-3754. doi: 10.1111/febs.12375

Cho, S., Hoang, A., Sinha, R., Zhong, X. Y., Fu, X. D., Krainer, A. R., & Ghosh, G. (2011). Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc Natl Acad Sci U S A, 108*(20), 8233-8238. doi: 10.1073/pnas.1017700108

Clough, S. J., & Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J, 16*(6), 735-743.

Cooke, C., & Alwine, J. C. (2002). Characterization of specific protein-RNA complexes associated with the coupling of polyadenylation and last-intron removal. *Mol Cell Biol, 22*(13), 4579-4586.

Cooke, C., Hans, H., & Alwine, J. C. (1999). Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. *Mol Cell Biol, 19*(7), 4971-4979.

Ding, F., Cui, P., Wang, Z., Zhang, S., Ali, S., & Xiong, L. (2014). Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in Arabidopsis. *Bmc Genomics, 15*, 431. doi: 10.1186/1471-2164-15-431

Drechsel, G., Kahles, A., Kesarwani, A. K., Stauffer, E., Behr, J., Drewe, P., Ratsch, G., & Wachter, A. (2013). Nonsense-mediated decay of alternative precursor

mRNA splicing variants is a major determinant of the Arabidopsis steady state transcriptome. *Plant Cell, 25*(10), 3726-3742. doi: 10.1105/tpc.113.115485

Durek, P., Schmidt, R., Heazlewood, J. L., Jones, A., MacLean, D., Nagel, A., Kersten, B., & Schulze, W. X. (2010). PhosPhAt: the Arabidopsis thaliana phosphorylation site database. An update. *Nucleic Acids Res, 38*(Database issue), D828-834. doi: 10.1093/nar/gkp810

Fortes, P., Bilbao-Cortes, D., Fornerod, M., Rigaut, G., Raymond, W., Seraphin, B., & Mattaj, I. W. (1999a). Luc7p, a novel yeast U1 snRNP protein with a role in 5' splice site recognition. *Genes Dev, 13*(18), 2425-2438.

Fortes, P., Kufel, J., Fornerod, M., Polycarpou-Schwarz, M., Lafontaine, D., Tollervey, D., & Mattaj, I. W. (1999b). Genetic and physical interactions involving the yeast nuclear cap-binding complex. *Mol Cell Biol, 19*(10), 6543-6553.

Fukumura, K., Taniguchi, I., Sakamoto, H., Ohno, M., & Inoue, K. (2009). U1-independent pre-mRNA splicing contributes to the regulation of alternative splicing. *Nucleic Acids Res, 37*(6), 1907-1914. doi: 10.1093/nar/gkp050

Gohring, J., Jacak, J., & Barta, A. (2014). Imaging of endogenous messenger RNA splice variants in living cells reveals nuclear retention of transcripts inaccessible to nonsense-mediated decay in Arabidopsis. *Plant Cell, 26*(2), 754-764. doi: 10.1105/tpc.113.118075

Golovkin, M., & Reddy, A. S. (1998). The plant U1 small nuclear ribonucleoprotein particle 70K protein interacts with two novel serine/arginine-rich proteins. *Plant Cell, 10*(10), 1637-1648.

Golovkin, M., & Reddy, A. S. (2003). Expression of U1 small nuclear ribonucleoprotein 70K antisense transcript using APETALA3 promoter suppresses the development of sepals and petals. *Plant Physiol, 132*(4), 1884-1891.

Gonatopoulos-Pournatzis, T., & Cowling, V. H. (2014). Cap-binding complex (CBC). *Biochem J, 457*(2), 231-242. doi: 10.1042/BJ20131214

Gouy, M., Guindon, S., & Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol, 27*(2), 221-224. doi: 10.1093/molbev/msp259

823 Guiro, J., & O'Reilly, D. (2015). Insights into the U1 small nuclear ribonucleoprotein
824 complex superfamily. *Wiley Interdisciplinary Reviews-Rna, 6*(1), 79-92. doi:
825 Doi 10.1002/Wrna.1257
826 Heazlewood, J. L., Durek, P., Hummel, J., Selbig, J., Weckwerth, W., Walther, D., &
827 Schulze, W. X. (2008). PhosPhAt: a database of phosphorylation sites in
828 Arabidopsis thaliana and a plant-specific phosphorylation site predictor.
829 *Nucleic Acids Res, 36*(Database issue), D1015-1021. doi:
830 10.1093/nar/gkm812
831 Heim, A., Grimm, C., Muller, U., Haussler, S., Mackeen, M. M., Merl, J., Hauck, S.
832 M., Kessler, B. M., Schofield, C. J., Wolf, A., & Bottger, A. (2014). Jumonji
833 domain containing protein 6 (Jmjd6) modulates splicing and specifically
834 interacts with arginine-serine-rich (RS) domains of SR- and SR-like proteins.
835 *Nucleic Acids Res, 42*(12), 7833-7850. doi: 10.1093/nar/gku488
836 Hernandez, H., Makarova, O. V., Makarov, E. M., Morgner, N., Muto, Y., Krummel, D.
837 P., & Robinson, C. V. (2009). Isoforms of U1-70k control subunit dynamics in
838 the human spliceosomal U1 snRNP. *PLoS One, 4*(9), e7202. doi:
839 10.1371/journal.pone.0007202
840 Hoffman, B. E., & Grabowski, P. J. (1992). U1 snRNP targets an essential splicing
841 factor, U2AF65, to the 3' splice site by a network of interactions spanning the
842 exon. *Genes Dev, 6*(12B), 2554-2568.
843 Hori, K., & Watanabe, Y. (2005). UPF3 suppresses aberrant spliced mRNA in
844 Arabidopsis. *Plant J, 43*(4), 530-540. doi: 10.1111/j.1365-313X.2005.02473.x
845 Kaida, D. (2016). The reciprocal regulation between splicing and 3'-end processing.
846 *Wiley Interdiscip Rev RNA, 7*(4), 499-511. doi: 10.1002/wrna.1348
847 Kaida, D., Berg, M. G., Younis, I., Kasim, M., Singh, L. N., Wan, L., & Dreyfuss, G.
848 (2010). U1 snRNP protects pre-mRNAs from premature cleavage and
849 polyadenylation. *Nature, 468*(7324), 664-668. doi: 10.1038/nature09479
850 Kalyna, M., Simpson, C. G., Syed, N. H., Lewandowska, D., Marquez, Y., Kusenda,
851 B., Marshall, J., Fuller, J., Cardle, L., McNicol, J., Dinh, H. Q., Barta, A., &
852 Brown, J. W. (2012). Alternative splicing and nonsense-mediated decay
853 modulate expression of important regulatory genes in Arabidopsis. *Nucleic*
854 *Acids Res, 40*(6), 2454-2469. doi: 10.1093/nar/gkr932
855 Kersey, P. J., Allen, J. E., Armean, I., Boddu, S., Bolt, B. J., Carvalho-Silva, D.,
856 Christensen, M., Davis, P., Falin, L. J., Grabmueller, C., Humphrey, J.,

857 Kerhornou, A., Khobova, J., Aranganathan, N. K., Langridge, N., Lowy, E.,
858 McDowall, M. D., Maheswari, U., Nuhn, M., Ong, C. K., Overduin, B., Paulini,
859 M., Pedro, H., Perry, E., Spudich, G., Tapanari, E., Walts, B., Williams, G.,
860 Tello-Ruiz, M., Stein, J., Wei, S., Ware, D., Bolser, D. M., Howe, K. L.,
861 Kulesha, E., Lawson, D., Maslen, G., & Staines, D. M. (2016). Ensembl
862 Genomes 2016: more genomes, more complexity. *Nucleic Acids Res, 44*(D1),
863 D574-580. doi: 10.1093/nar/gkv1209
864 Kervestin, S., & Jacobson, A. (2012). NMD: a multifaceted response to premature
865 translational termination. *Nat Rev Mol Cell Biol, 13*(11), 700-712. doi:
866 10.1038/nrm3454
867 Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013).
868 TopHat2: accurate alignment of transcriptomes in the presence of insertions,
869 deletions and gene fusions. *Genome Biol, 14*(4), R36. doi: 10.1186/gb-2013-
870 14-4-r36
871 Knop, K., Stepien, A., Barciszewska-Pacak, M., Taube, M., Bielewicz, D., Michalak,
872 M., Borst, J. W., Jarmolowski, A., & Szweykowska-Kulinska, Z. (2016). Active
873 5' splice sites regulate the biogenesis efficiency of Arabidopsis microRNAs
874 derived from intron-containing genes. *Nucleic Acids Res*. doi:
875 10.1093/nar/gkw895
876 Koncz, C., Dejong, F., Villacorta, N., Szakonyi, D., & Koncz, Z. (2012). The
877 spliceosome-activating complex: molecular mechanisms underlying the
878 function of a pleiotropic regulator. *Front Plant Sci, 3*, 9. doi:
879 10.3389/fpls.2012.00009
880 Kornblihtt, A. R., Schor, I. E., Allo, M., Dujardin, G., Petrillo, E., & Munoz, M. J.
881 (2013). Alternative splicing: a pivotal step between eukaryotic transcription
882 and translation. *Nat Rev Mol Cell Biol, 14*(3), 153-165. doi: 10.1038/nrm3525
883 Laubinger, S., Sachsenberg, T., Zeller, G., Busch, W., Lohmann, J. U., Ratsch, G., &
884 Weigel, D. (2008). Dual roles of the nuclear cap-binding complex and
885 SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis
886 thaliana. *Proc Natl Acad Sci U S A, 105*(25), 8795-8800. doi:
887 10.1073/pnas.0802493105
888 Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C., & Mattaj, I. W. (1996). A
889 nuclear cap-binding complex facilitates association of U1 snRNP with the cap-

proximal 5' splice site. *Genes & Development, 10*(13), 1683-1698. doi: 10.1101/gad.10.13.1683

Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res, 41*(10), e108. doi: 10.1093/nar/gkt214

Muller-McNicoll, M., Botti, V., de Jesus Domingues, A. M., Brandl, H., Schwich, O. D., Steiner, M. C., Curk, T., Poser, I., Zarnack, K., & Neugebauer, K. M. (2016). SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev, 30*(5), 553-566. doi: 10.1101/gad.276477.115

Nakagawa, T., Suzuki, T., Murata, S., Nakamura, S., Hino, T., Maeo, K., Tabata, R., Kawai, T., Tanaka, K., Niwa, Y., Watanabe, Y., Nakamura, K., Kimura, T., & Ishiguro, S. (2007). Improved Gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. *Biosci Biotechnol Biochem, 71*(8), 2095-2100. doi: 10.1271/bbb.70216

Naro, C., Jolly, A., Di Persio, S., Bielli, P., Setterblad, N., Alberdi, A. J., Vicini, E., Geremia, R., De la Grange, P., & Sette, C. (2017). An Orchestrated Intron Retention Program in Meiosis Controls Timely Usage of Transcripts during Germ Cell Differentiation. *Dev Cell, 41*(1), 82-93 e84. doi: 10.1016/j.devcel.2017.03.003

Nesic, D., & Maquat, L. E. (1994). Upstream introns influence the efficiency of final intron removal and RNA 3'-end formation. *Genes Dev, 8*(3), 363-375.

Ohtani, M., & Sugiyama, M. (2005). Involvement of SRD2-mediated activation of snRNA transcription in the control of cell proliferation competence in Arabidopsis. *Plant J, 43*(4), 479-490. doi: 10.1111/j.1365-313X.2005.02469.x

Papp, I., Mur, L. A., Dalmadi, A., Dulai, S., & Koncz, C. (2004). A mutation in the Cap Binding Protein 20 gene confers drought tolerance to Arabidopsis. *Plant Mol Biol, 55*(5), 679-686. doi: 10.1007/s11103-004-1680-2

Pimentel, H., Parra, M., Gee, S. L., Mohandas, N., Pachter, L., & Conboy, J. G. (2016). A dynamic intron retention program enriched in RNA processing genes regulates gene expression during terminal erythropoiesis. *Nucleic Acids Res, 44*(2), 838-851. doi: 10.1093/nar/gkv1168

Prigge, M. J., & Wagner, D. R. (2001). The arabidopsis serrate gene encodes a zinc-finger protein required for normal shoot development. *Plant Cell, 13*(6), 1263-1279.

924 Puig, O., Bragado-Nilsson, E., Koski, T., & Seraphin, B. (2007). The U1 snRNP-
925 associated factor Luc7p affects 5' splice site selection in yeast and human.
926 *Nucleic Acids Res, 35*(17), 5874-5885. doi: 10.1093/nar/gkm505
927 Raczynska, K. D., Simpson, C. G., Ciesiolka, A., Szewc, L., Lewandowska, D.,
928 McNicol, J., Szweykowska-Kulinska, Z., Brown, J. W., & Jarmolowski, A.
929 (2010). Involvement of the nuclear cap-binding protein complex in alternative
930 splicing in Arabidopsis thaliana. *Nucleic Acids Res, 38*(1), 265-278. doi:
931 10.1093/nar/gkp869
932 Raczynska, K. D., Stepien, A., Kierzkowski, D., Kalak, M., Bajczyk, M., McNicol, J.,
933 Simpson, C. G., Szweykowska-Kulinska, Z., Brown, J. W., & Jarmolowski, A.
934 (2013). The SERRATE protein is involved in alternative splicing in Arabidopsis
935 thaliana. *Nucleic Acids Res*. doi: 10.1093/nar/gkt894
936 Reddy, A. S., Marquez, Y., Kalyna, M., & Barta, A. (2013). Complexity of the
937 alternative splicing landscape in plants. *Plant Cell, 25*(10), 3657-3683. doi:
938 10.1105/tpc.113.117523
939 Rigo, F., & Martinson, H. G. (2008). Functional coupling of last-intron splicing and 3'-
940 end processing to transcription in vitro: the poly(A) signal couples to splicing
941 before committing to cleavage. *Mol Cell Biol, 28*(2), 849-862. doi:
942 10.1128/mcb.01410-07
943 Rigo, F., & Martinson, H. G. (2009). Polyadenylation releases mRNA from RNA
944 polymerase II in a process that is licensed by splicing. *RNA, 15*(5), 823-836.
945 doi: 10.1261/rna.1409209
946 Rosel, T. D., Hung, L. H., Medenbach, J., Donde, K., Starke, S., Benes, V., Ratsch,
947 G., & Bindereif, A. (2011). RNA-Seq analysis in mutant zebrafish reveals role
948 of U1C protein in alternative splicing regulation. *EMBO J, 30*(10), 1965-1976.
949 doi: 10.1038/emboj.2011.106
950 Rowley, M. J., Bohmdorfer, G., & Wierzbicki, A. T. (2013). Analysis of long non-
951 coding RNAs produced by a specialized RNA polymerase in Arabidopsis
952 thaliana. *Methods, 63*(2), 160-169. doi: 10.1016/j.ymeth.2013.05.006
953 Roy, D., Bhanja Chowdhury, J., & Ghosh, S. (2013). Polypyrimidine tract binding
954 protein (PTB) associates with intronic and exonic domains to squelch nuclear
955 export of unspliced RNA. *FEBS Lett, 587*(23), 3802-3807. doi:
956 10.1016/j.febslet.2013.10.005

957 Schlaen, R. G., Mancini, E., Sanchez, S. E., Perez-Santangelo, S., Rugnone, M. L.,
958 Simpson, C. G., Brown, J. W., Zhang, X., Chernomoretz, A., & Yanovsky, M.
959 J. (2015). The spliceosome assembly factor GEMIN2 attenuates the effects of
960 temperature on alternative splicing and circadian rhythms. *Proc Natl Acad Sci*
961 *U S A, 112*(30), 9382-9387. doi: 10.1073/pnas.1504541112
962 Schwab, R., Speth, C., Laubinger, S., & Voinnet, O. (2013). Enhanced microRNA
963 accumulation through stemloop-adjacent introns. *EMBO Rep, 14*(7), 615-621.
964 doi: 10.1038/embor.2013.58
965 Shao, W., Kim, H. S., Cao, Y., Xu, Y. Z., & Query, C. C. (2012). A U1-U2 snRNP
966 interaction network during intron definition. *Mol Cell Biol, 32*(2), 470-478. doi:
967 10.1128/MCB.06234-11
968 Shaul, O. (2015). Unique Aspects of Plant Nonsense-Mediated mRNA Decay.
969 *Trends Plant Sci, 20*(11), 767-779. doi: 10.1016/j.tplants.2015.08.011
970 Shen, S., Park, J. W., Lu, Z. X., Lin, L., Henry, M. D., Wu, Y. N., Zhou, Q., & Xing, Y.
971 (2014). rMATS: robust and flexible detection of differential alternative splicing
972 from replicate RNA-Seq data. *Proc Natl Acad Sci U S A, 111*(51), E5593-
973 5601. doi: 10.1073/pnas.1419161111
974 Staiger, D., & Brown, J. W. (2013). Alternative splicing at the intersection of biological
975 timing, development, and stress responses. *Plant Cell, 25*(10), 3640-3656.
976 doi: 10.1105/tpc.113.113803
977 Stepien, A., Knop, K., Dolata, J., Taube, M., Bajczyk, M., Barciszewska-Pacak, M.,
978 Pacak, A., Jarmolowski, A., & Szweykowska-Kulinska, Z. (2017).
979 Posttranscriptional coordination of splicing and miRNA biogenesis in plants.
980 *Wiley Interdiscip Rev RNA, 8*(3). doi: 10.1002/wrna.1403
981 Sullivan, C., & Howard, P. L. (2016). The Diverse Requirements of ARS2 in nuclear
982 cap-binding complex-dependent RNA Processing. *Rna & Disease*. doi:
983 10.14800/rd.1376
984 Umezawa, T., Sugiyama, N., Takahashi, F., Anderson, J. C., Ishihama, Y., Peck, S.
985 C., & Shinozaki, K. (2013). Genetics and phosphoproteomics reveal a protein
986 phosphorylation network in the abscisic acid signaling pathway in Arabidopsis
987 thaliana. *Sci Signal, 6*(270), rs8. doi: 10.1126/scisignal.2003509
988 Wachter, A., Tunc-Ozdemir, M., Grove, B. C., Green, P. J., Shintani, D. K., &
989 Breaker, R. R. (2007). Riboswitch control of gene expression in plants by

splicing and alternative 3' end processing of mRNAs. *Plant Cell, 19*(11), 3437-3450. doi: 10.1105/tpc.107.053645

Wahl, M. C., Will, C. L., & Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell, 136*(4), 701-718. doi: 10.1016/j.cell.2009.02.009

Wang, B. B., & Brendel, V. (2004). The ASRG database: identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing. *Genome Biol, 5*(12), R102. doi: 10.1186/gb-2004-5-12-r102

Wang, C., Tian, Q., Hou, Z., Mucha, M., Aukerman, M., & Olsen, O. A. (2007). The Arabidopsis thaliana AT PRP39-1 gene, encoding a tetratricopeptide repeat protein with similarity to the yeast pre-mRNA processing protein PRP39, affects flowering time. *Plant Cell Rep, 26*(8), 1357-1366. doi: 10.1007/s00299-007-0336-5

Wang, P., Xue, L., Batelli, G., Lee, S., Hou, Y. J., Van Oosten, M. J., Zhang, H., Tao, W. A., & Zhu, J. K. (2013). Quantitative phosphoproteomics identifies SnRK2 protein kinase substrates and reveals the effectors of abscisic acid action. *Proc Natl Acad Sci U S A, 110*(27), 11205-11210. doi: 10.1073/pnas.1308974110

Webby, C. J., Wolf, A., Gromak, N., Dreger, M., Kramer, H., Kessler, B., Nielsen, M. L., Schmitz, C., Butler, D. S., Yates, J. R., 3rd, Delahunty, C. M., Hahn, P., Lengeling, A., Mann, M., Proudfoot, N. J., Schofield, C. J., & Bottger, A. (2009). Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science, 325*(5936), 90-93. doi: 10.1126/science.1175865

Will, C. L., & Luhrmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol, 13*(3), 290-301.

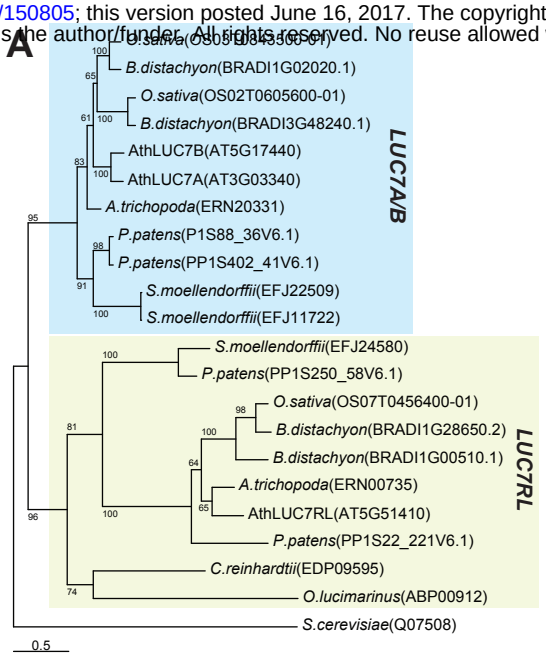
Will, C. L., & Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb Perspect Biol, 3*(7). doi: 10.1101/cshperspect.a003707

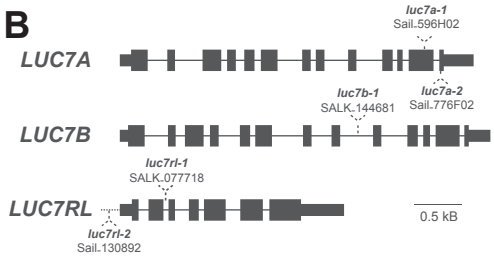
Wong, J. J., Au, A. Y., Ritchie, W., & Rasko, J. E. (2016). Intron retention in mRNA: No longer nonsense: Known and putative roles of intron retention in normal and disease biology. *Bioessays, 38*(1), 41-49. doi: 10.1002/bies.201500117

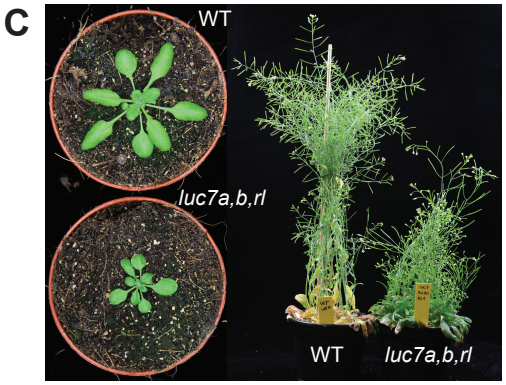
Xing, D., Wang, Y., Hamilton, M., Ben-Hur, A., & Reddy, A. S. (2015). Transcriptome-Wide Identification of RNA Targets of Arabidopsis SERINE/ARGININE-RICH45 Uncovers the Unexpected Roles of This RNA Binding Protein in RNA Processing. *Plant Cell*. doi: 10.1105/tpc.15.00641

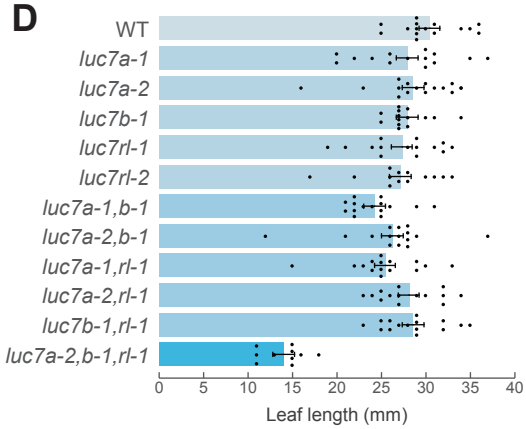
Yap, K., Lim, Z. Q., Khandelia, P., Friedman, B., & Makeyev, E. V. (2012). Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes Dev, 26*(11), 1209-1223. doi: 10.1101/gad.188037.112

Yoine, M., Ohto, M. A., Onai, K., Mita, S., & Nakamura, K. (2006). The lba1 mutation of UPF1 RNA helicase involved in nonsense-mediated mRNA decay causes pleiotropic phenotypic changes and altered sugar signalling in Arabidopsis. *Plant J, 47*(1), 49-62. doi: 10.1111/j.1365-313X.2006.02771.x









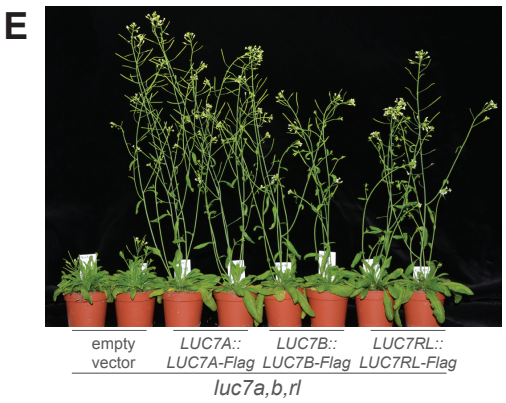


FIGURE 1

Figure 1: Arabidopsis LUC7 proteins redundantly control plant development

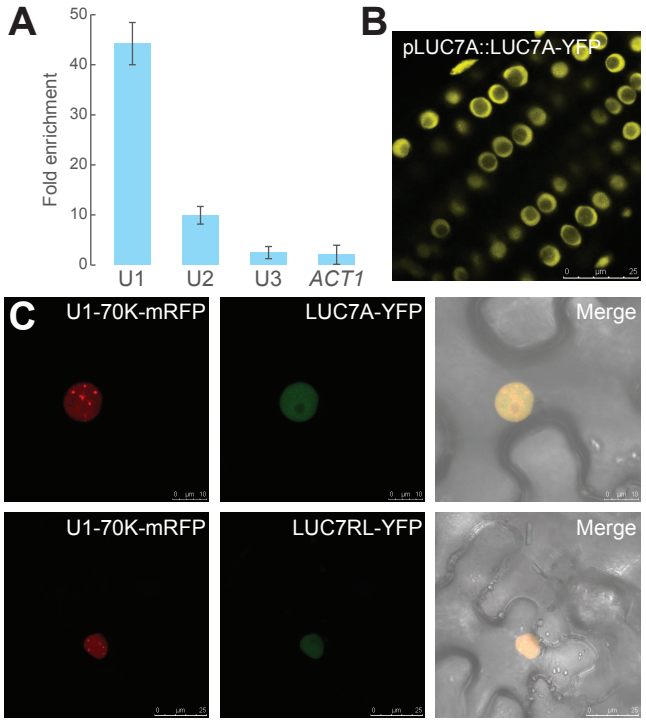
A: Phylogenetic analysis of LUC7 proteins in the plant kingdom using *Saccharomyces cerevisiae* as an external group.

B: Exon/intron structure of *Arabidopsis thaliana LUC7A*, *LUC7B* and *LUC7RL*. Dotted lines indicate the positions of T-DNA insertions.

C: WT and *luc7* triple mutant plants 21 days (left) and 45 days (right) after germination.

D: Length of the longest rosette leaf of 21 days-old WT, *luc7* single, double and triple mutant plants growing under long day conditions. Leaves of 10-15 individual plants were measured. Dots indicate individual data points. Error bars denote the SEM.

E: Complementation of *luc7a,b,rl* mutants by *LUC7A*, *LUC7B* and *LUC7RL* genomic rescue constructs. Transformation of an “empty” binary vector served as a control. Two independent transgenic lines for each construct are shown.



D

Identified protein	Aver. # of Peptides IP LUC7	Aver. # of Peptides IP WT	Seq.Coverage (%) IP LUC7	Seq.Coverage (%) IP WT	MW (kDa)
LUC7A	45.75	0	71.9	0	47.4
U1 snRNP components					
U1-70k	5	0	13.3	0	50.4
SmB-a ; SmB-b	2.5	0	9.8	0	27.0
U1A	2	0	7.7	0	28.1
Splicing-related					
U2AF35A;U2AF35B	3.5	1	11.9	4.3	34.6
SCL30A;SCL33	2.75	0	12.3	0	20.2
SR45	1.75	0	5.2	0	44.6
RSZ21;RSZ22;RSZ22A	1.75	0	8.0	0	22.5
SR30	1	0	3.9	0	29.1
RS2Z33	1	0	2.9	0	32.9
SCL30	0.75	0	2.8	0	29.6
U2AF65A	0.75	0	1.7	0	60.7
RNA metabolism					
SERRATE	7.75	0	12.0	0	81.1
Cleavage factor AT4G25550	1.75	0	11.0	0	22.8
ABH1	1	0	1.4	0	96.6
Kinases					
SRPK4	8.5	0	16.2	0	59.4
AT4G25550	8.25	0	15.9	0	60.7
SRPK3	4.25	0	8.9	0	61.2

FIGURE 2

Figure 2: Arabidopsis LUC7 is an U1 snRNP component

A: RNA immunoprecipitation using a *pLUC7A:LUC7A-YFP, luc7a,b,rl* complemented line. Proteins were immunoprecipitated using GFP-specific affinity matrix and RNAs were extracted from the input and the immunoprecipitated. U1, U2, U3 snRNAs and *ACTIN* RNA were quantified using qRT-PCR. Enrichment of the respective RNA in LUC7A:LUC7A-YFP luc7a,b,rl transgenic lines was calculated over WT (negative control). Error bars denote the range of two biological replicates.

B: Subcellular localization of LUC7A in *pLUC7A:LUC7A-YFP luc7a,b,rl* in Arabidopsis transgenic plants. Roots of 9 day-old seedlings were analyzed using confocal microscopy. Scale bar indicates 25 µm.

C: U1-70K-mRFP and LUC7A-YFP or LUC7RL-YFP proteins were transiently expressed in *N. benthamiana*. The subcellular localization of mRFP and YFP fusion proteins was analyzed using confocal microscopy. Scale bars indicate 10 µm and 25 µm for upper and lower panel, respectively.

D: List of some LUC7-associated proteins identified by MS.

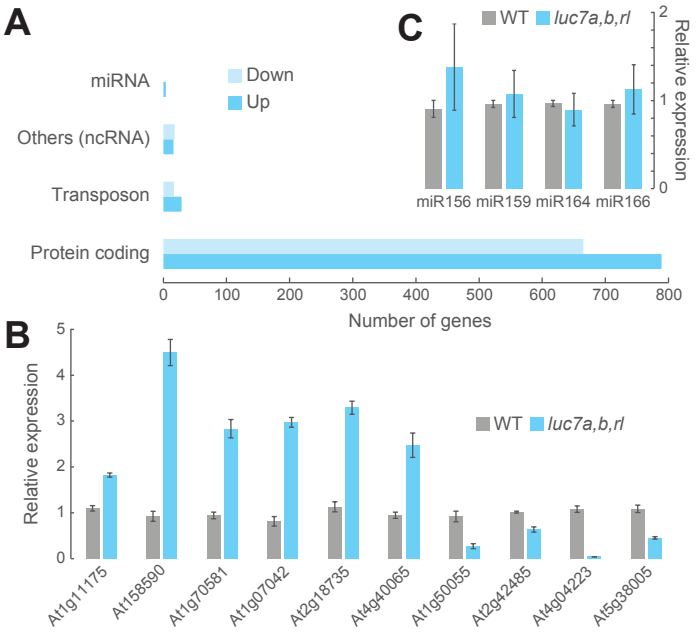


FIGURE 3

Figure 3: Mutations in *LUC7* results in misexpression of protein-coding and non-coding genes, but not in MIRNA genes

A: Differentially expressed genes in *luc7a,b,rl* mutant compared to WT.

B,C: qRT-PCR analysis of selected ncRNA (B) and miRNAs (C) in WT and *luc7a,b,rl.*
Error bars denote the SEM (n=3).

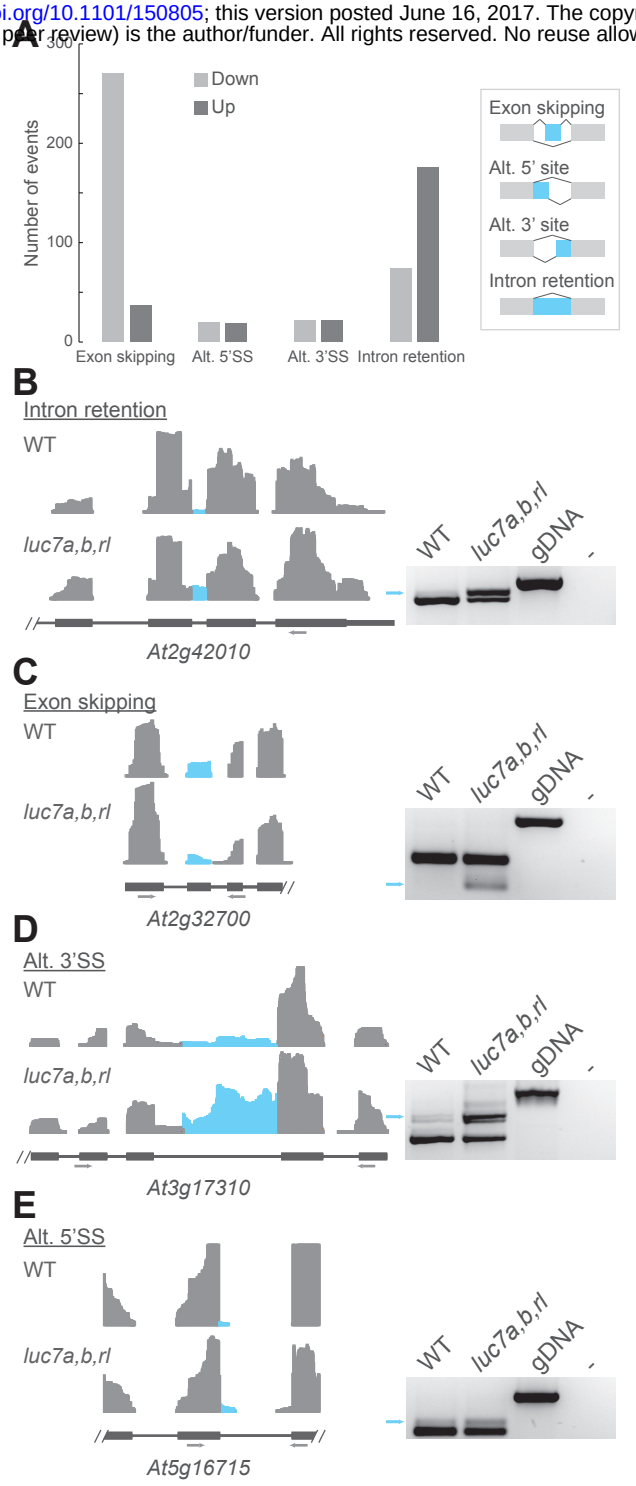


FIGURE 4

Figure 4: Global analysis of splicing defects in *luc7* triple mutant.

A: Classification of splicing events changes in *luc7* triple mutant compared to WT.

B-F: Coverage plots and RT-PCR validation experiments for selected splicing events in WT and *luc7* triple mutant. Genomic DNA (gDNA) or water (-) served as a control. Primers positions are indicated with gray arrows. IR, intron retention; ES, exon skipping; Alt.3'SS, alternative 3'splicing site; Alt.5'SS, alternative 5'splicing site.

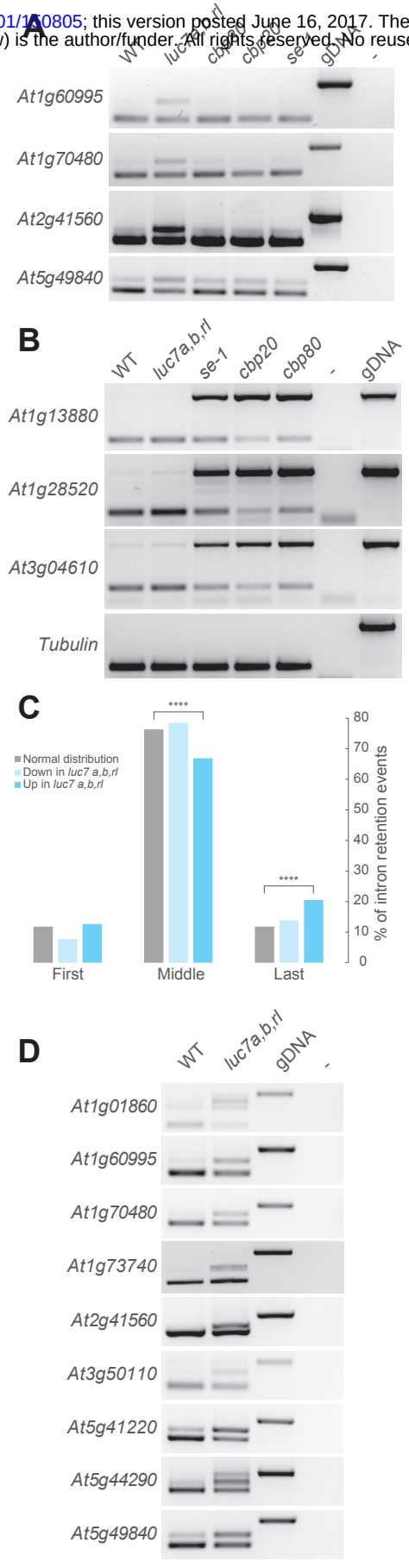


FIGURE 5

Figure 5: LUC7 proteins have a pronounced effect on terminal intron splicing

A: RT-PCR analysis of LUC7-dependent introns in WT, *luc7* triple mutant, *cbp80*, *cbp20* and *se-1* mutants.
B: RT-PCR analysis of CBC/SE-dependent introns in WT, *luc7* triple mutant, *cbp80*, *cbp20* and *se-1* mutants.
C: Classification of intron retention according to the intron position (first, middle, or last). Only genes with 3 or more introns were considered for this analysis. Terminal introns are significantly enriched among introns retained in *luc7* triple mutants (Fischer test; two-side $p < 0.0016$).
D: RT-PCR analysis of genes carrying retained terminal introns in WT and *luc7* triple mutants.

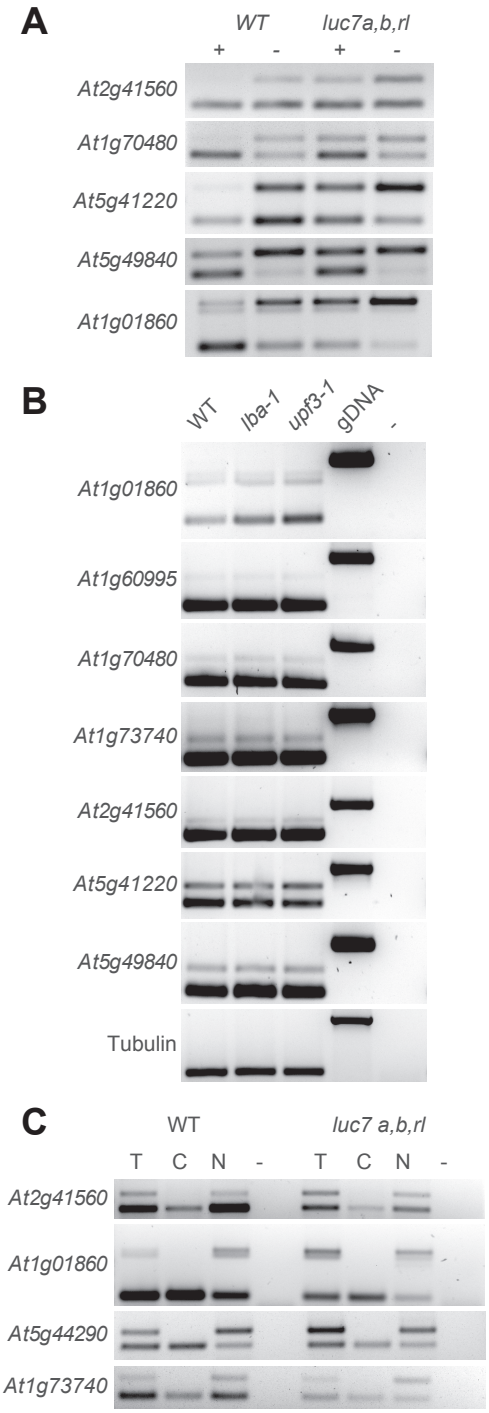


FIGURE 6

Figure 6: mRNAs containing retained LUC7-dependent terminal introns are NMD-insensitive and remain nuclear.

A: Splicing patterns of mRNAs carrying LUC7-dependent terminal introns in poly(A)+ and poly(A)- fractions.

B: RT-PCR analysis of LUC7-dependent terminal introns in WT and NMD mutants (*lba1* and *upf3-1*).

C: Splicing patterns of mRNAs isolated from total (T), cytosolic (C) and nuclear (N) fractions.

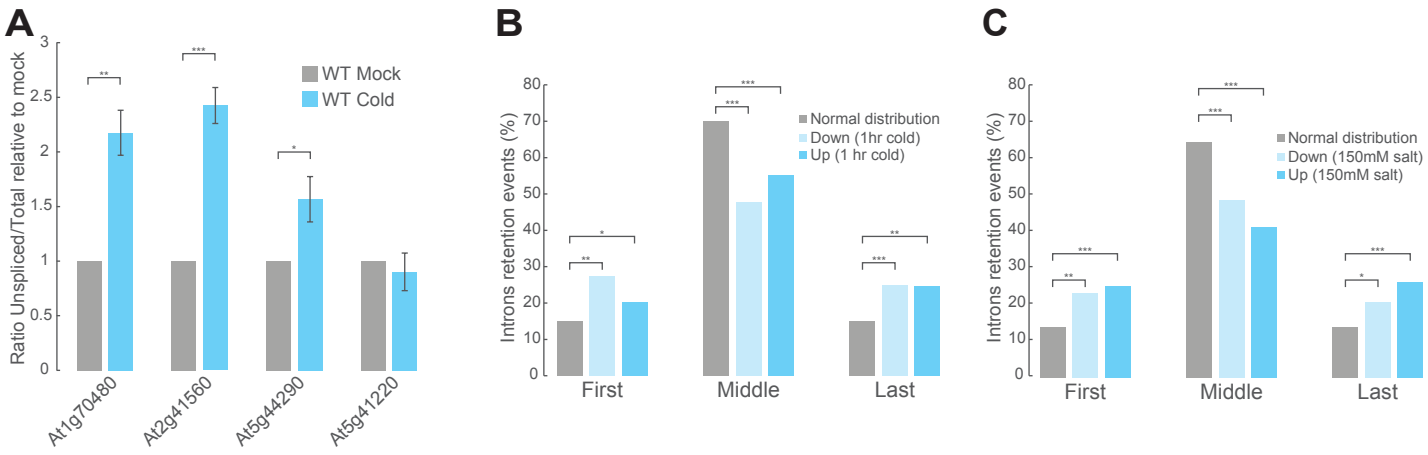


FIGURE 7

Figure 7: Splicing of LUC7 dependent terminal introns can be modulated by stress.

A: Seven days old WT seedlings were exposed to cold for 60 min. Splicing ratios (unspliced/total RNA) of four genes featuring a LUC7-dependent terminal intron were analyzed by qPCR. T-test was performed before calculating the relative to WT (* p=0,0208, ** p =0,0095, *** p = 0.004).

B-C: Classification of intron retention events in cold- (B) and salt-treated plants (C) The relative amount of retained/removed first, middle, and last intron in stress samples was compared to the relative amount of first, middle, and last intron among all expressed genes (normal distribution). Only genes with 3 or more introns were considered for this analysis.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$