1 **Title**

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

4

5 **Authors**

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

8

9 **Affiliations**

10 [1] Centre for Plant Molecular Biology (ZMBP), University of Tuebingen, Tuebingen,
11 Germany
12 [2] Chemical Genomics Centre (CGC) of the Max Planck Society, Dortmund, Germany
13 [3] Max Planck Institute for Developmental Biology, Tuebingen, Germany
14 [4] Max Planck Institute for Plant Breeding Research (MPIPZ), Cologne, Germany
15 [5] Proteome Centre, University of Tuebingen, Tuebingen, Germany

16

17 * present address: Carl von Ossietzky University, Oldenburg, Germany
18 Corresponding author: Sascha Laubinger (sascha.laubinger@uol.de)

19

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

Abstract

Introns are removed by the spliceosome, a large macromolecular 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 proteins are U1 snRNP accessory components important for plant development and stress resistance. Transcriptome analyses revealed that LUC7 proteins are not only important for constitutive splicing, but also affect hundreds of alternative splicing events. Interestingly, LUC7 proteins specifically promote splicing of a subset of terminal introns. Splicing of LUC7-dependent terminal introns is a prerequisite for nuclear export and some splicing events are modulated by stress in a LUC7-dependent manner. Taken together our results highlight the importance of the U1 snRNP component LUC7 in splicing regulation and suggest a previously unrecognized contribution of terminal introns splicing in plant stress responses.

38 **Introduction**

39 Eukaryotic genes are often interrupted by non-coding sequences called introns that
40 are removed from pre-mRNAs while the remaining sequence, the exons, are joined
41 together. This process, called splicing, is an essential step before the translation of
42 the mature mRNAs and it offers a wide range of advantages for eukaryotic
43 organisms. For instance, alternative splicing allows the production of more than one
44 isoform from a single gene expanding the genome coding capacity (Kornblihtt et al,
45 2013; Reddy et al, 2013). In plants, alternative splicing contributes to essentially all
46 aspects of development and stress responses (Carvalho et al, 2013; Staiger &
47 Brown, 2013). Alternative splicing can regulate gene expression by generating
48 transcripts with premature termination codons (PTC) or/and a long 3'UTR, which may
49 lead to RNA degradation via the nonsense-mediated decay (NMD) pathway
50 (Drechsel et al, 2013; Kalyna et al, 2012; Shaul, 2015). Furthermore, splicing can be
51 coupled with other RNA processing events, such as 3'end formation and RNA
52 transport to the cytosol (Kaida, 2016; Muller-McNicoll et al, 2016).

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

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

66 The fact that U1 snRNP is recruited to the 5'ss in the initial step of splicing
67 suggests that this complex is necessary for correct splicing site selection. Indeed, it
68 has been shown that U1-deficient zebrafish mutants accumulate alternative spliced
69 transcripts, suggesting that the U1 snRNP fulfills regulatory roles in splice site
70 selection (Rosel et al, 2011). Although the spliceosome consists of stoichiometrically
71 equal amounts of each subunit, the U1 snRNP is more abundant than all the other
72 spliceosomal subcomplexes (Kaida, 2016; Kaida et al, 2010). One reason for this is
73 that the U1 snRNP executes splicing independent functions. For instance, metazoan
74 U1 snRNP binds not only to the 5'ss, but also throughout the nascent transcript to
75 block a premature cleavage and polyadenylation (Berg et al, 2012; Kaida et al,
76 2010). Furthermore, the U1 snRNP is also important to regulate promoter
77 directionality and transcription in animals (Almada et al, 2013; Guiro & O'Reilly,
78 2015).

79 U1 snRNP complexes were purified and characterized in yeast and human.
80 The U1 snRNP contains the U1 snRNA, Sm proteins, three U1 core proteins (U1-
81 70K, U1-A and U1-C) and U1-specific accessory proteins, such as LUC7, PRP39
82 and PRP40. All these proteins are conserved in plants suggesting a U1 snRNP
83 composition very similar to the one in yeast and metazoans (Koncz et al, 2012;
84 Reddy et al, 2013; Wang & Brendel, 2004). Interaction studies revealed that U1
85 snRNP associates with serine-arginine (SR) proteins, indicating a complex
86 mechanism for splicing site selection that involves also non-snRNP proteins (Cho et
87 al, 2011; Golovkin & Reddy, 1998).

88 The function of the plant U1 snRNP is not well characterized. This might be
89 due to the fact that in *Arabidopsis thaliana* the core U1 snRNP components *U1-70K*,

90 *U1-A* and *U1-C* are single copy genes and a complete knockout most likely causes
91 lethality. On the other hand, proteins such as PRP39, PRP40 and LUC7 are encoded
92 by small gene families, which require the generation of multiple mutants for functional
93 studies. Other factors, such as GEMIN2 or SRD2 are required for the functionality of
94 all snRNPs, but not specifically for U1 function (Ohtani & Sugiyama, 2005; Schlaen et
95 al, 2015). Only a single mutant in a U1 snRNP-specific gene, *PRP39A,* was
96 characterized. *prp39a* mutants flower late due to increased expression of the
97 flowering time regulator *FLOWERING LOCUS C* (*FLC*), but the mutant does not
98 exhibit any additional developmental defects (Wang et al, 2007). In a reverse genetic
99 approach, *U1-70K* expression was specifically reduced in flowers by an antisense
100 RNA and the resulting transgenic plants exhibit strong floral defects (Golovkin &
101 Reddy, 2003). Thus, despite evidences that U1 snRNP is essential for plant
102 development, the functions of the U1 snRNP in regulating the transcriptome of plants
103 are unknown.

104 Here, we report on the functional characterization of *Arabidopsis* mutants
105 impaired in U1 snRNP function. For this, we focused in this study on the U1 snRNP
106 components LUC7, which we show to be essential for normal plant development and
107 plant stress resistance. Our whole transcriptome analyses on *luc7* triple mutant show
108 that impairments of LUC7 proteins affect constitutive and alternative splicing.
109 Surprisingly, our results reveal the existence of transcripts, in which terminal introns
110 are preferentially retained in a LUC7-dependent manner and that these unspliced
111 terminal introns cause a nuclear retention of the pre-mRNAs. Our results suggests
112 that the plant U1 snRNP component LUC7 carries out a specialized function in the
113 removal of terminal introns to prevent nuclear export of pre-mRNAs and this could be
114 a mechanism to fine-tune gene expression under stress conditions.

115 **Results**

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

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

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

141 reintroduced a wild-type copy of *LUC7A*, *LUC7B* or *LUC7RL* in the *luc7* triple mutant.
142 Each of the *LUC7* genes was sufficient to restore the growth phenotype of the *luc7*
143 triple mutant (Figure 1E). These results reveal that the phenotype observed in this
144 mutant is attributable to the impairment of *LUC7* function and it suggests that *LUC7*
145 genes act redundantly to control *Arabidopsis* growth and development.
146

147 **LUC7 functions in the ABA pathway and is important for salt stress resistance**

148 Splicing is essential for plant stress resistance and mutants impaired in splicing often
149 react differently to stress and the stress hormone abscisic acid (ABA) (Filichkin et al,
150 2015; Zhan et al, 2015). In addition, global impairment of the splicing machinery
151 elicits ABA signaling (AlShareef et al, 2017; Ling et al, 2017). To test whether LUC7
152 is important for plant stress resistance and ABA-mediated stress signaling, we
153 analyzed growth parameters of WT, the *luc7* triple mutant and a *luc7* rescue line in
154 presence of exogenous ABA or salt. A cotyledon greening assay showed that *luc7*
155 triple mutants reacted hypersensitively to exogenous ABA (Figure 2A, B), suggesting
156 that *LUC7* plays an important role in the ABA pathway. Furthermore, salt in the
157 growth medium impaired root growth much more strongly in *luc7* triple mutant than in
158 WT or in a *luc7* rescue line (Figure 2C, D). These results imply that functional *LUC7*
159 proteins are required for plant stress resistance and ABA responses.
160

161 **LUC7 is a U1 snRNP component in plants**

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

167 A protein that is part of the U1 complex is tightly associated with U1 specific
168 components such as the U1 snRNA. To test whether LUC7 is found in a complex
169 with the U1 snRNA, we performed RNA immunoprecipitation (RIP) experiments using
170 a *luc7* triple mutant carrying *pLUC7A:LUC7A-YFP* rescue construct (Figure S3).
171 Immunoprecipitation of LUC7A-YFP enriched the U1 snRNA more than 40-fold, but
172 did not enrich two unrelated, but abundant RNAs, U3 snoRNA and *ACTIN* mRNA
173 (Figure 3A). Small amounts of U2 snRNA was also found associated with LUC7A,
174 which is in agreement with the fact that U1 and U2 snRNP directly interact to form
175 spliceosomal complex A (Figure 3A). However, the amount of recovered U2 snRNA
176 is more than four-fold lower than that of the U1 snRNA (Figure 3A). These results
177 strongly suggest that Arabidopsis LUC7 proteins are bona fide U1 snRNP
178 components.

179 Next we analyzed the subcellular localization of LUC7A and its co-localization
180 with a core U1 snRNP subunit. LUC7A localized to the nucleus, but not to the
181 nucleolus in *Arabidopsis* plants containing the *pLUC7A:LUC7A-YFP* rescue construct
182 (Figure 3B). In addition, LUC7A partially co-localized with U1-70K in the nucleoplasm
183 when transiently expressed in *Nicotiana benthamiana* (Figure 3C). Similar results
184 were obtained for LUC7RL, the Arabidopsis LUC7 most distant in sequence to
185 LUC7A (Figure 3C). Interestingly, unlike U1-70K, LUC7A and LUC7RL did not form
186 distinct speckles in the nucleus and were evenly localized in the nucleoplasm (Figure
187 3C). In plants, co-localizations studies in protoplasts have shown that also the core
188 U1 components do not completely colocalize (Lorkovic & Barta, 2008). These partial
189 colocalizations suggest that plant U1 snRNP proteins may fulfill additional functions
190 as it has been observed in other eukaryotes (Workman et al, 2014).

191

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

193 In order to identify misregulated and misspliced genes in *luc7* mutants, we performed
194 an RNA-sequencing (RNA-seq) analysis with three biological replicates. We decided
195 to use seven days old WT and *luc7* triple mutant seedlings. At this age, *luc7* triple
196 mutant and WT seedlings are morphologically similar and therefore, changes in
197 transcript levels and splicing patterns most likely reflect changes caused by
198 impairments of LUC7 proteins and are not due to different contribution of tissues
199 caused by, for instance, a delay in development or/and different morphology (Figure
200 S4).

201 An analysis of differentially expressed genes revealed that 840 genes are up-
202 and 703 are downregulated in *luc7* triple mutant when compared to WT. The majority
203 of genes that change expression were protein-coding genes (Figure 4A).
204 Nevertheless, non-coding RNAs (ncRNAs) were significantly enriched among the
205 *LUC7* affected genes ($p < 0.05$, hypergeometric test), although the overall number of
206 ncRNA affected in *luc7* triple mutant is relatively small (Figure 4A, B). Previous
207 studies implied that the U1 snRNP regulates microRNA (miRNA) biogenesis
208 (Bielewicz et al, 2013; Knop et al, 2016; Schwab et al, 2013; Stepien et al, 2017).
209 However, the expression of *MIRNA* genes was not affected in *luc7* triple mutants
210 (Figure 4A). In addition, quantification of mature miRNA levels revealed that all tested
211 miRNAs did not change abundance in *luc7* triple mutants (Figure 4C). These results
212 show that LUC7 proteins affect the expression of protein-coding genes and a subset
213 of ncRNAs, but are not involved in the miRNA pathway.

214

215 **Arabidopsis *LUC7* functions are important for constitutive and alternative**
216 **splicing**

217 Because LUC7 proteins are U1 snRNP components, we ask whether misspliced
218 transcripts accumulate in the *luc7* triple mutant. In total, we identified 645 differential

219 splicing events in *luc7* triple mutants compared to WT. We detected a large number
220 of intron retention events (Figure 5A). RT-PCR experiments with oligonucleotides
221 flanking selected intron retentions events confirmed the RNA-seq data (Figure 5B).
222 These results suggest that lack and/or impairment of the U1 snRNP component
223 LUC7 disturbs intron recognition and thus splicing. We also identified a large number
224 of exons skipping events in the *luc7* triple mutant. Exon skipping is also a major
225 outcome of impairing U1 snRNP function or binding in metazoans (Lorkovic & Barta,
226 2008; Rosel et al, 2011). These defects are most likely caused by interactions of the
227 U2 snRNP with U1 snRNPs associated to alternative 5'splice sites (Morcos, 2007).
228 Furthermore, we detected several cases of alternative 5' and 3' splice site selection
229 in *luc7* triple mutants (Figure 5A-F).

230 Some splicing defects observed in *luc7* triple mutants generated transcript
231 variants that did not exist in WT (e.g. *At2g32700*, Figure 5C). On the contrary, in
232 other cases the *luc7* triple mutant lacked specific mRNA isoforms, which exist in wild-
233 type plants (e.g. *At1g10980*, *At4g32060*), or the ratio of two different isoforms was
234 altered in *luc7* triple mutant when compared to WT (e.g. *At3g17310*, *At5g16715*,
235 *At5g48150*, *At2g11000*) (Figure 5D-F). These results show that Arabidopsis LUC7
236 proteins are involved in constitutive and alternative splicing.

237 Next, we checked whether splicing changes observed in *luc7* triple mutant are
238 actually due to the loss of only a specific *LUC7* gene or whether *LUC7* genes act
239 redundantly. To test this, we analyzed the splicing pattern of some mRNAs in *luc7*
240 single, double and triple mutants. Some splicing defects were detectable even in *luc7*
241 single mutants (Figure S5), but the degree of missplicing increased in *luc7* double
242 and triple mutants suggesting that LUC7 proteins act additively on these introns (e.g.
243 *At5g16715*). Some splicing defects occurred only in *luc7* triple mutants, implying that
244 LUC7 proteins act redundantly to ensure splicing of these introns (e.g. *At1g60995*).

245 Other splicing defects might more likely be due to the lack of LUC7A/B or LUC7RL.
246 For instance, intron removal of *At2g42010* more strongly relied on *LUC7RL*, while
247 removal of an intron in *At5g41220* preferentially depends on LUC7A/LUC7B (Figure
248 S5). These findings suggest that Arabidopsis *LUC7* genes function redundantly,
249 additively or specifically to ensure proper splicing.
250

251 **LUC7 proteins are preferentially involved in the removal of terminal introns**

252 In yeast, LUC7 connects the CBC with the U1 snRNP and this interaction is
253 important for the correct 5' splicing site selection (Fortes et al, 1999a). In plants, the
254 CBC associates with SE to ensure splicing of cap-proximal first introns (Laubinger et
255 al, 2008; Raczynska et al, 2010; Raczynska et al, 2013). To investigate the
256 relationship between LUC7 and the CBC/SE in plants, we analyzed the splicing
257 patterns of LUC7 dependent introns in *cbc* mutants (*cbp20* and *cbp80*) and *se-1* by
258 RT-PCR. All tested introns retained in *luc7* triple mutant were correctly spliced in *cbc*
259 and *se* mutants (Figure 6A). Conversely, first introns that are partially retained in
260 *cbp20, cbp80* and *se-1* mutants were completely removed in the *luc7* triple mutant
261 (Figure 6B). These observations suggest that the functions of LUC7 and CBC/SE in
262 splicing of the selected introns do not overlap.

263 Next, we asked whether LUC7 has a preference for promoting splicing of cap-
264 proximal first introns as it has the CBC/SE complex. We classified retained introns in
265 *luc7* triple mutant according to their position within the gene (first, middle or last
266 introns). Only genes with at least 3 introns were considered for this analysis. We
267 found a significant increase in retained last introns, but not first introns, in *luc7* triple
268 mutants (Figure 6C). Although the total number of retained introns is higher among
269 middle introns, the relative amount of retained middle introns in *luc7* triple mutant
270 was significantly reduced (Figure 6C). Retention of terminal introns in *luc7* triple

271 mutants was confirmed by RT-PCR analysis (Figure 6D). In summary, our data
272 revealed that (i) CBC/SE acts independently of LUC7 in splicing of cap-proximal
273 introns and that (ii) LUC7 proteins play an important role for the removal of certain
274 terminal introns.
275

276 **Analysis of LUC7 containing complexes suggests an association of LUC7 with**
277 **proteins involved in 3'end processing and polyA-binding**

278 In analogy to LUC7 function in yeast, in which it interacts with the CBC to ensure
279 splicing of cap-proximal introns, we hypothesized that Arabidopsis LUC7 might be
280 also connected to 3'end processing and/or polyadenylation machinery to ensure
281 splicing of terminal introns. To test with which proteins LUC7 associates *in planta*, we
282 purified LUC7A-containing complexes. For this, we used *pLUC7A:LUC7A-YFP*
283 complemented lines and as controls wild-type plants and transgenic lines expressing
284 free GFP (*p35S:GFP*). Immunopurifications (IPs) were carried out three to four time
285 independently. We observed that WT often produced more background in mass
286 spectrometry (MS) analyses than the 35S:GFP line and we therefore decided to use
287 WT as a more stringent control (Table S1). Among all the identified proteins we
288 considered those putative LUC7 interactors that were found in at least two
289 independent experiment and were at least three time more abundant in
290 *pLUC7A:LUC7A-YFP* IPs than in WT IPs. The MS analysis revealed that LUC7 is
291 found in a complex with core U1 snRNP proteins U1-A, U1-70K and SM proteins,
292 further underpinning that LUC7 is a U1 snRNP component (Table 1, Table S1).
293 Additional proteins known to be involved in splicing including several serine-arginine
294 (SR) proteins (SR30, SCL30A, SCL33), and regulatory proteins, among them several
295 kinases, were found to be associated with LUC7A complexes (Table 1, Table S1).

296 The idea that LUC7A containing complexes associate with proteins involved in
297 3'end processing/polyA-binding was supported by several potential LUC7 interactors.
298 We found LUC7 in complexes with RBP47C, a polyA-binding protein of unknown
299 function (Lorkovic et al, 2000), and LARP6, which is targeted to 3'ends of mRNA
300 through interaction with polyA-binding protein 2 (PAB2) (Merret et al, 2013).
301 Interestingly, we also found an SPT6-like transcription factor associated with LUC7
302 complexes. SPT6 binds the pol II C-terminal domain (CTD) phosphorylated at serine
303 2 (Ser2P), which accumulates at 3'end of genes (Kaplan et al, 2000; Sun et al,
304 2010). In addition, we detected peptides corresponding to the spliceosomal complex
305 E components U2AF35 and U2AF65, the latter of which has been shown to interact
306 with a 3'end processing factor to connect splicing and 3'end processing (Millevoi et
307 al, 2002; Millevoi et al, 2006). Taken together, the analysis of LUC7A complexes
308 suggests that LUC7A forms complexes with proteins involved in 3'end processing
309 and polyA-binding. None of these LUC7 complex components has been studied
310 functionally and it will be a major effort for future studies to determine the function of
311 these proteins in terminal intron splicing.
312

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

315 We were further interested in determining the characteristics and possible functions
316 of LUC7-dependent terminal introns. When introns are retained, the resulting mRNA
317 can contain a premature stop codon and a long 3'UTR, which are hallmarks of NMD
318 targets (Drechsel et al, 2013; Kalyna et al, 2012; Shaul, 2015). To check whether
319 mRNAs containing a retained LUC7-dependent terminal intron are NMD substrates,
320 we analyzed their splicing patterns in two mutants impaired in NMD, *lba-1* and *upf3-*
321 *1*. If unspliced isoforms were indeed NMD targets, we would expect their abundance

322 to be increased in NMD mutants. Interestingly, we did not observe any change
323 between WT and *upf* mutants (Figure 7A). Thus, we conclude that retained LUC7-
324 dependent terminal introns do not trigger degradation via the NMD pathway.

325 NMD occurs in the cytoplasm and RNAs can escape NMD by not being
326 transported from the nucleus to the cytosol (Gohring et al, 2014). We therefore
327 checked in which cellular compartment mRNAs with spliced and unspliced LUC7-
328 dependent terminal introns accumulate. To do this, we isolated total, nuclear and
329 cytosolic fractions from wild-type and *luc7* triple mutant plants and performed RT-
330 PCR analyses (Figure 7B). Purity of cytosolic and nuclear fractions was controlled by
331 immunoblot analysis using antibodies against histone H3 (specific for nuclear
332 fractions) and a 60S ribosomal protein (L13-1, specific for cytosolic fractions) (Figure
333 7C). Spliced mRNA isoforms accumulated in the cytosol, whereas mRNAs containing
334 the unspliced terminal introns were found in nuclear fractions (Figure 7B). These
335 results indicate that retention of terminal introns correlates with trapping mRNAs in
336 the nucleus and suggest that splicing of LUC7-dependent terminal introns is essential
337 for mRNA transport to the cytosol.

338

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

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

348 To test this hypothesis, we decided to check the splicing of LUC7-dependent
349 terminal introns in WT under stress condition. We chose cold stress because it was
350 suggested that U1 snRNP functionality is impaired under cold condition (Schlaen et
351 al, 2015). To quantify the amount of unspliced isoforms in cold condition, we
352 designed qPCR-primers specific to unspliced isoforms and the total mRNA pool and
353 calculate the relative amount of mRNA carrying unspliced LUC7-dependent terminal
354 introns compared to the total mRNA pool. mRNAs of *At1g70480*, *At2g41560* and
355 *At5g44290* significantly accumulated unspliced isoforms in responses to cold
356 treatment demonstrating that cold stress modulates the splicing efficiency of these
357 LUC7-dependent terminal introns (Figure 8A). Interestingly, the amount of unspliced
358 mRNA in *luc7* triple mutants does not differ significantly between mock and stress
359 conditions (Figure 8A). This observation suggests that LUC7 is directly involved in
360 the regulation of terminal intron splicing under stress conditions and that LUC7 might
361 be a target for stress response pathways.
362

363 **Discussion**

364 **Functions of the Arabidopsis U1 snRNP component LUC7**

365 For this study, we generated an Arabidopsis triple mutant deficient in the U1 snRNP
366 components LUC7 and dissected the genome-wide effects of LUC7s impairments on
367 the Arabidopsis transcriptome. Our results show that LUC7 proteins are *bona-fide* U1
368 components acting mainly redundantly. The reduction of U1 function in the *luc7* triple
369 mutant affects constitutive splicing. A large number of introns are retained in *luc7*
370 triple mutant, suggesting that without a proper recognition of the 5'ss, splicing of the
371 affected introns is impaired. Our results also show that exon-skipping events are
372 impaired in *luc7* triple mutant, revealing that a functional plant U1 snRNP is essential
373 for exon definition. In addition, we show that *luc7* triple mutant affect alternative

374 splicing also by influencing events of alternative 5' and 3' splice site. This implies that
375 the U1 snRNP does not only affect 5' splice site usage, it might also indirectly
376 regulate usage of 3' splice sites via its interaction with U2AFs and the U2 snRNP
377 (Hoffman & Grabowski, 1992; Shao et al, 2012). The functions of LUC7 proteins on
378 the Arabidopsis transcriptome are likely to be underestimated, because misspliced
379 mRNAs in *luc7* mutants might contain hallmarks of NMD and are therefore rapidly
380 turned over and escape detection. Analysis of *luc7* mutants combined with mutations
381 in NMD factors would help to uncover the full set of splicing events affected by LUC7.
382 One has also to consider that U1 snRNP independent splicing has been described in
383 animals, indicating that not all introns require the U1 complex for efficient intron
384 removal (Fukumura et al, 2009). The degree of U1-independent splicing in plants
385 remains to be elucidated.

386 Duplications among genes encoding for U1 snRNP proteins, such as the
387 LUC7 genes, open up a possibility for sub- and neofunctionalization of U1 accessory
388 proteins. Furthermore, the Arabidopsis genome encodes 14 potential U1 snRNAs,
389 which slightly differ in sequence (Wang & Brendel, 2004). Therefore, the plant U1
390 snRNP presumably does not exist as a single complex, but might exist as different
391 sub-complexes exhibiting distinct specificities and functions. In metazoans, the
392 existence of at least four different U1 snRNP subcomplex has been suggested (Guiro
393 & O'Reilly, 2015; Hernandez et al, 2009). Specific combinations of plant U1 protein
394 family members and U1 snRNAs could generate an even higher number of such U1
395 subcomplexes, which could be responsible for specific splicing events. Our results
396 show that LUC7 can act redundantly, but can also fulfill specific functions, suggesting
397 that LUC7 complexes specifically act on certain pre-mRNAs. In this regard, it is
398 important to note that an additional short stretch of amino acids separates the two
399 zinc-finger domains in LUCA and LUC7B (Figure S1). Changing the space in

400 between RNA binding domains affects substrate specificities and could explain
401 different specificities among LUC7 proteins (Chen & Varani, 2013).

402 Interestingly, *luc7* triple mutant showed a significant higher retention rate of
403 terminal introns compared to first or middle introns. This was surprising because
404 LUC7 was initially found to act in concert with the CBC, a complex involved in the
405 removal of cap-proximal first introns, but not of last introns (Lewis et al, 1996). Often,
406 the removal of terminal introns is coupled to polyadenylation (Cooke & Alwine, 2002;
407 Cooke et al, 1999; Rigo & Martinson, 2008). Interestingly, we detected components
408 involved in RNA 3'end processing or polyA-binding as part of LUC7A complexes,
409 suggesting that such interactions may contribute to the specific functions of LUC7 in
410 terminal intron splicing. Future studies will focus on the characterization and
411 functional analysis of these proteins in Arabidopsis.

412

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

414 Intron retention can target a transcript to NMD due to the generation of a premature
415 stop codon (PTC) that either produces a long 3'UTR (≥300 - 350 nt) or introns
416 located downstream (≥50 - 55 nt) (Drechsel et al, 2013; Kalyna et al, 2012; Kervestin
417 & Jacobson, 2012; Shaul, 2015). Retained terminal introns are then special because
418 their detection solely relies on the length of the 3'UTR. Transport of mRNAs
419 containing unspliced terminal introns, especially those harboring a relatively short
420 3'UTR, can be detrimental and has to be tightly controlled. Interestingly, we found
421 that splicing of LUC7-dependent terminal introns is required for transport of mRNAs
422 from the cytosol to the nucleus. The fact that we cannot detect unspliced transcript in
423 the cytosol suggests a nuclear retention mechanism for such mRNAs. One possibility
424 is that LUC7-dependent terminal introns might contain binding sites for specific trans-
425 regulatory factors that upon binding inhibit export. Polypyrimidine tract-binding

426 protein 1 (PTB1) is a candidate for such a trans-regulatory protein, because binding
427 of PTB1 to introns represses nuclear export of certain RNAs (Roy et al, 2013; Yap et
428 al, 2012).

429 Nuclear retention of unspliced mRNAs is not limited to terminal introns and
430 might be a much more general mechanism to escape NMD and to regulate gene
431 expression (Wong et al, 2016). In plants, some specific transcript isoforms have been
432 detected only in the nucleus, but not in the cytosol (Gohring et al, 2014). Also in
433 metazoans, intron retention might have a more general role in regulating gene
434 expression (Naro et al, 2017; Pimentel et al, 2016; Yap et al, 2012). The so-called
435 detained introns are evolutionary conserved, NMD insensitive and retained in the
436 nucleus (Boutz et al, 2015). The functional importance of intron retention was also
437 suggested in the fern *Marsilea vestita*, in which many mRNAs contain introns that are
438 only spliced shortly before gametophyte development (Boothby et al, 2013).

439 We found that LUC7-dependent splicing of terminal introns can be modulated
440 by cold stress. Because retention of these introns causes nuclear trapping, it is
441 prompting to speculate that environmental cues affect splicing and nuclear retention
442 of mRNAs. Such a mechanism would regulate the amount of translatable mRNAs in
443 the cytosol in a cost-efficient and rapid manner (Figure 8B). Since the regulation of
444 terminal splicing is lost in *luc7* mutants, one can expect that LUC7 function might be
445 regulated under stress conditions. Interestingly, the RS domains of LUC7 proteins
446 are phosphorylated and we identified three kinases as LUC7A interactors (Durek et
447 al, 2010; Heazlewood et al, 2008). In addition, stress signaling triggered by the
448 phytohormone ABA causes differential phosphorylation of several splicing factors
449 (Umezawa et al, 2013; Wang et al, 2013). Whether stress-induced changes in
450 phosphorylation play a role in regulating LUC7 proteins and whether the LUC7-

451 interacting kinases here identified are involved in this process remains to be
452 elucidated.

453 **Material and Methods**

454

455 **Plant material and growth conditions**

456 All mutants were in the Columbia-0 (Col-0) background. *luc7a-1* (SAIL_596_H02)
457 and *luc7a-2* (SAIL_776_F02), *luc7b-1* (SALK_144681), *luc7rl-1* (SALK_077718) and
458 *luc7rl-2* (SALK_130892C) were isolated by PCR-based genotyping (Table S2). *luc7*
459 double and triple mutants were generate by crossing individual mutants. All other
460 mutants used in this study (*abh1-285*, *cbp20-1*, *se*-1, *lba-1* and *upf3-1*) were
461 described elsewhere (Hori & Watanabe, 2005; Laubinger et al, 2008; Papp et al,
462 2004; Prigge & Wagner, 2001; Yoine et al, 2006). The line expressing GFP was
463 generated using the vector pBinarGFP and was kindly provided by Dr. Andreas
464 Wachter (Wachter et al, 2007). For complementation analyses, *pLUC7A:LUC7A-*
465 *FLAG*, *pLUC7B:LUC7B-FLAG*, *pLUC7RL:LUC7RL-FLAG* and *pLUC7A:LUC7A-YFP*
466 constructs were introduced in *luc7* triple mutant by Agrobacterium-mediated
467 transformation (Clough & Bent, 1998). All plants were grown on soil in long days
468 conditions (16-h light/8-h dark) at 20°C/18°C day/night. The size of *luc7* mutants was
469 assessed by measuring the longest rosette leaf after 21 days. For all molecular
470 studies, seeds were surface-sterilized, plated on 1/2 MS medium with 0.8%
471 phytoagar and grown for 7 days in continuous light at 22°C. For the cold treatment,
472 plates with *Arabidopsis* seedlings were transferred to ice-water for 60 min. For the
473 root growth assay, 4 days old seedlings growing on vertical plates were transferred to
474 mock plates or plates containing indicated amount of NaCl and grown for more 11
475 days vertically. Root growth rate per day was assessed by measuring in ImageJ the
476 root length in the days 2 and 9 after transfer. For ABA sensitivity assays, seedlings
477 were grown for 10 days on 1/2 MS plates supplemented with 0.8% phytoagar, 1%
478 sucrose and indicated amounts of ABA (+) (Sigma - A4906).

479

480 **Plasmid constructions and transient expression analyses**

481 For the expression of C-terminal FLAG- and YFP-tagged LUC7 proteins expressed
482 from their endogenous regulatory elements, 2100 bp, 4120 bp and 2106 bp upstream
483 of the ATG start codon of *LUC7A*, *LUC7B* and *LUC7RL*, respectively, to the last
484 coding nucleotide were PCR-amplified and subcloned in pCR8/GW/TOPO®
485 (Invitrogen). Oligonucleotides are listed in Table S2). Entry clones were recombined
486 with pGWB10 and pGWB540 using Gateway LR clonase II (Invitrogen) to generate
487 binary plasmids containing *pLUC7A:LUC7A-FLAG*, *pLUC7B:LUC7B-FLAG*,
488 *pLUC7RL:LUC7RL-FLAG* and *pLUC7A:LUC7A-YFP*. For the co-localization studies,
489 entry vector containing the coding sequence of U1-70k was recombined with
490 pGWB654 for the expression of *p35S:U1-70k-mRFP* (Nakagawa et al, 2007).
491 Agrobacterium-mediated transient transformation of *Nicotiana benthamiana* plants
492 was conducted as following. Overnight Agrobacterium culture were diluted in 1:6 and
493 grown for 4 hours at 28°C. After centrifugation, pellets were resuspended in
494 infiltration medium (10mM $MgCl_2$, 10mM MES-KOH pH 5.6, 100 µM Acetosyringone).
495 The OD 600nm was adjusted to 0.6-0.8 and samples were mixed when required. *N.*
496 *benthamiana* were infiltrated and subcellular localization was checked after 3 days.
497 Subcellular localization of fluorescent proteins was analyzed by confocal microscopy
498 using a Leica TCS SP8.

499

500 **Phylogenetic analysis**

501 AthLUC7A (AT3G03340) protein sequence was analyzed in Interpro
502 (https://www.ebi.ac.uk/interpro/) to retrieve the Interpro ID for the conserved Luc7-
503 related domain (IPR004882). The sequence for *Saccharomyces cerevisiae* (strain
504 ATCC 204508_S288c) was obtained in Interpro. Plants sequences were extracted

505 using BioMart selecting for the protein domain IPR004882 on Ensembl Plants
506 (http://plants.ensembl.org/). The following genomes were included in our analyses:
507 *Amborella trichopoda* (AMTR1.0 (2014-01-AGD)); *Arabidopsis thaliana* (TAIR10
508 (2010-09-TAIR10)); *Brachypodium distachyon* (v1.0); *Chlamydomonas reinhardtii*
509 (v3.1 (2007-11-ENA)); *Physcomitrella patens* (ASM242v1 (2011-03-Phypa1.6));
510 *Selaginella moellendorffii* (v1.0 (2011-05-ENA)); *Oryza sativa Japonica* (IRGSP-1.0);
511 and *Ostreococcus lucimarinus* genes (ASM9206v1). The phylogenetic analysis was
512 performed in Seaview (Version 4.6.1) using Muscle for sequence alignment.
513 Maximum likehood (PhYML) was employed with 1000 bootstraps (Gouy et al, 2010).
514

515 **RNA extractions, RT-PCR and qRT-PCR**

516 RNAs extractions were performed with Direct-zol™ RNA MiniPrep Kit (Zymo
517 Research). Total RNAs were treated with DNAse I and cDNA synthesis carried out
518 with RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) using usually
519 oligo dT primers or a mixture of hexamer and miRNA-specific stem-loop primers
520 (Table S2). Standard PCRs for the splicing analysis were performed with DreamTaq
521 DNA Polymerase (Thermo Scientific). Quantitative RT-PCR (qRT-PCR) was
522 performed using the Maxima SYBR Green (Thermo Scientific) in a Bio-Rad CFX 384.
523 For all qPCR-primers, primer efficiencies were determined by a serial dilution of
524 cDNA template. The relative expressions were calculated using the 2ˆ(-ΔΔCT)
525 method with *PP2A* or *ACTIN* as control. For the qRT-PCR to detect splicing ratio
526 changes under cold condition, the ratio $2^{(-\Delta CT_{unspliced})}/2^{(-\Delta CT_{total\ RNA})}$ was
527 calculated separately for each replicate and t-test was performed before calculating
528 the relative to WT mock. Oligonucleotides are listed in Table S2. For RNA-
529 sequencing analysis, polyA RNAs were enriched from 4 µg of total RNAs using
530 NEBNext Oligo $d(T)_{25}$ Magnetic Beads (New England Biolabs). The libraries were

531 prepared using ScriptSeq™ Plant Leaf kit (Epicentre) following the manufacturer's
532 instruction. Single end sequencing was performed on an Illumina HiSeq2000.
533 Sequencing data were deposited at Gene Expression Omnibus under accession
534 number GSE98779.
535
536 **Subcellular fractionation**
537 Two grams of seedlings were ground in N_2 liquid and resuspended in 4 ml of Honda
538 buffer (0.44 M sucrose, 1.25% Ficoll 400, 2.5% Dextran T40, 20 mM HEPES KOH
539 pH 7.4, 10 mM $MgCl_2$, 0.5% Triton X-100, 5 mM DTT, 1 mM PMSF, protease
540 inhibitor cocktail [ROCHE] supplemented with 40U/ml of Ribolock®). The
541 homogenate was filtered through 2 layer of Miracloth, which was washed with 1ml of
542 Honda Buffer. From the filtrate, 300 µl was removed as “total” fraction and kept on
543 ice. Filtrates were centrifuged at 1,500 *g* for 10 min, 4°C for pelleting nuclei and
544 supernatants were transferred to a new tube. Supernatants were centrifuged at 13
545 000 x g, 4 °C, 15 min and 300 µl were kept on ice as cytoplasmic fraction. Nuclei
546 pellets were washed five times in 1 ml of Honda buffer (supplemented with 8U/ml of
547 Ribolock®, centrifugation at 1,800 *g* for 5 min. The final pellet was resuspended in
548 300 µl of Honda buffer. To all the fractions (total, cytoplasmic and nuclei), 900 µl of
549 TRI Reagent (Sigma) was added. After homogenization, 180 µl of chloroform was
550 added and samples were incubated at room temperature for 10 min. After
551 centrifugation at 10 0000 rpm for 20 min, 4°C, the aqueous phase were transferred to
552 a new tube and RNA extracted with Direct-zol™ RNA MiniPrep Kit (Zymo Research).
553 The organic phase was collected and proteins were isolated according to
554 manufacturer’s instructions (TRI Reagent). cDNA synthesis with random primes was
555 performed as above. Proteins extracted were analyzed by standard western blot

556 techniques using the following antibodies: H3 (~ 17 KDa / ab 1791, Abcam) and 60S
557 ribosomal (~ 23,7-29 KDa / L13, Agrisera).
558

559 **RNA immunoprecipitation**

560 RNA immunoprecipitation (RIP) using WT and a *pLUC7A:LUC7A-eYFP* rescue line
561 was performed as described elsewhere with minor modifications (Rowley et al, 2013;
562 Xing et al, 2015). Isolated nuclei were sonicated in nuclear lysis buffer in a Covaris
563 E220 (Duty Cycle: 20%; Peak intensity: 140; Cycles per Burst: 200; Cycle time: 3').
564 RNAs were extracted using RNeasy Plant Mini Kit (QIAGEN) following the
565 manufacturer's instructions. The RNA were treated with DNAseI (Thermo Scientific)
566 and samples were split in half for the (-)RT reaction. cDNA synthesis were perform
567 with SuperScript™ III Reverse Transcriptase (Invitrogen). qRT-PCRs were performed
568 with QuantiNova™ SYBRR Green PCR (QIAGEN).
569

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

572 RNA-seq reads for each replicate were aligned against the *Arabidopsis thaliana*
573 reference sequence (TAIR10) using tophat (v2.0.10, -p2, -a 10, -g 10, -N 10, --read-
574 edit-dist 10, --library-type fr-secondstrand, --segment-length 31, -G TAIR10.gff). Next,
575 cufflinks (version 2.2.1) was used to extract FPKM counts for each expressed
576 transcript generating a new annotation file (transcripts.gtf), where the coordinates of
577 each expressed transcript can be found. Cuffcompare (version 2.2.1) was then used
578 to generate a non-redundant annotation file containing all reference transcripts in
579 addition to new transcripts expressed in at least one of the nine samples
580 (cuffcmp.combined.gtf). The differential expression analysis was performed with
581 cuffdiff (version 2.2.1) between wt/*luc7* triple using the annotation file generated by

582 cuffcompare (FDR<2 and FC>0,05). For the splicing analysis, the same alignment
583 files generated by tophat and annotation files generated by cuffcompare
584 (cuffcmp.combined.gtf) were used as input for MATS (version 3.0.8) in order to test
585 for differentially spliced transcripts (Shen et al, 2014).
586
587 **GO Analysis**
588 GO analysis was performed in Bar Utoronto (http://bar.utoronto.ca/ntools/cgi-
589 bin/ntools_classification_superviewer.cgi).
590
591 **Protein complex purification and mass spectrometry (MS) Analysis**
592 LUC7A immunoprecipitation was performed using a complemented line
593 *pLUC7A:LUC7A-eYFP* (line 20.3.1) and a transgenic *p35S:GFP* and WT as controls.
594 Four independent biological replicates were performed. Seedlings (4 g) were ground
595 in N_2 liquid and respuspended in 1 volumes of extraction buffer (50 mM Tris-Cl pH
596 7.5, 100 mM NaCl, 0,5% Triton X-100, 5% Glycerol, 1mM PMSF, 100 µM MG132,
597 Complete Protease Inhibitor Cocktail EDTA-free [Roche] and Plant specific protease
598 Inhibitor, Sigma P9599). After thawing, samples were incubated on ice for 30 min,
599 centrifuged at 3220 rcf for 30 min at 4°C and filtrated with two layers of Miracloth. For
600 each immunoprecipitation, 20 µl of GFP-trap (Chromotek) was washed twice with 1
601 ml of washing buffer (50 mM Tris-Cl pH 7.5, 100 mM NaCl, 0.2% Triton X-100) and
602 once with 0.5 ml of IP buffer. For each replicate, the same amount of plant extracts
603 (~5ml) were incubated with GFP-trap and incubated on a rotating wheel at 4°C for 3
604 hours. Samples were centrifuged at 800-2000 rcf for 1-2 min and the supernatant
605 discarded. GFP-beads were resuspended in 1 ml of washing buffer, transferred into a
606 new tube and washed 4 to 5 times. Then, beads were ressuspended in ~40 µl of 2x
607 Laemmili Buffer and incubated at 80°C for 10 min. Short gel purifications (SDS-

608 PAGE) were performed and gels slices were digested with Trypsin. LC-MS/MS
609 analyses were performed in two mass spectrometer. Samples from R10 to R14 were
610 analyzed on a Proxeon Easy-nLC coupled to Orbitrap Elite (method: 90min, Top10,
611 HCD). Samples from R15 to R17 were analysed on a Proxeon Easy-nLC coupled to
612 OrbitrapXL (method: 90min, Top10, CID). Samples from R18 to R20 analysis on a
613 Proxeon Easy-nLC coupled to OrbitrapXL (method: 130min, Top10, CID). All the
614 replicates were processed together on MaxQuant software (Version 1.5.2.8. with
615 integrated Andromeda Peptide search engine) with a setting of 1% FDR and the
616 spectra were searched against an *Arabidopsis thaliana* Uniprot database
617 (UP000006548_3702_complete_20151023.fasta). All peptides identified are listed in
618 Supplementary Table S1 and raw data were deposited publically (accession
619 PXD006127).
620

621 **Acknowledgments**

622 This work was supported by the Deutsche Forschungsgemeinschaft DFG (to S.L.,
623 LA2633-4/1), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
624 (CAPES - Brazil) for doctoral fellowship (to M.F.A.), the Max Planck Society (to K.S.),
625 the Max Planck Society Chemical Genomics Centre (CGC) through its supporting
626 companies AstraZeneca, Bayer CropScience, Bayer Healthcare, Boehringer-
627 Ingelheim and Merck (to S.L). We are grateful to Andreas Wachter (ZMBP, University
628 of Tuebingen, Germany) and members of the lab for critical reading of the
629 manuscript, Christa Lanz for her invaluable help with Illumina sequencing, and
630 Johanna Schröter and her team for excellent care of our plants, Anja Hoffmann for
631 excellent technical assistance, and Andreas Wachter (ZMBP, University of
632 Tuebingen, Germany), Tsuyoshi Nakagawa (Department of Molecular and Functional
633 Genomics, Center for Integrated Research in Science, Shimane University, Matsue,

634 Japan) and the Notthingham Arabidopsis Stock Centre for providing seeds and DNA
635 constructs.

636

637 **Author contributions**

638 M.d.F.A. and S.L. designed research. M.d.F.A., I.D-B. and S.L. performed
639 experiments. E.-M. W., M.d.F.A., I.D-B., B.M., K.S. and S.L. analyzed the data.
640 M.d.F.A. and S.L. wrote the manuscript with contributions from all authors.

641

642 **Conflict of interest**

643 The authors declare no conflict of interest.

644

645 **Figure legends**

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

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

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

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

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

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

659

660 **Figure 2: *Arabidopsis* LUC7 is involved in ABA signaling and salt stress**
661 **responses**
662 **A,B:** WT, *luc7* triple mutant and a *luc7* rescue line (*luc7a,b,rl; pLUC7A:LUC7A-YFP*)
663 were grown on half-strength MS plates containing 1% sucrose and indicated amount
664 of ABA. Seedling phenotypes (A) and quantification of seedlings with green
665 cotyledons (B) are shown. Green cotyledons were scored ten days after germination.
666 One of two biological replicates is shown.
667 **C,D:** WT, *luc7* triple mutant and a *luc7* rescue line (*luc7a,b,rl; pLUC7A:LUC7A-YFP*)
668 were germinated on half-strength MS vertical plates and seedling were transferred on
669 half-strength MS plates containing the indicated amount of NaCl. Plates were always
670 placed vertically and the root growth was scored over 7 days. Phenotypes (C) and
671 root length quantification (D) are shown.
672
673 **Figure 3: *Arabidopsis* LUC7 is an U1 snRNP component**
674 **A:** RNA immunoprecipitation using a *pLUC7A:LUC7A-YFP, luc7a,b,rl* complemented
675 line. Proteins were immunoprecipitated using GFP-specific affinity matrix and RNAs
676 were extracted from the input and the immunoprecipitated. U1, U2, U3 snRNAs and
677 *ACTIN* RNA were quantified using qRT-PCR. Enrichment of the respective RNA in
678 *LUC7A:LUC7A-YFP luc7a,b,rl* transgenic line was calculated over WT (negative
679 control). Error bars denote the range of two biological replicates.
680 **B:** Subcellular localization of LUC7A in *pLUC7A:LUC7A-YFP luc7a,b,rl* in
681 Arabidopsis transgenic plants. Roots of 9 day-old seedlings were analyzed using
682 confocal microscopy. Scale bar indicates 25 µm.
683 **C:** U1-70K-mRFP and LUC7A-YFP or LUC7RL-YFP proteins were transiently
684 expressed in *N. benthamiana*. The subcellular localization of mRFP and YFP fusion

685 proteins was analyzed using confocal microscopy. Scale bars indicate 10 µm and 25
686 µm for upper and lower panel, respectively.
687
688 **Figure 4: Mutations in *LUC7* result in misexpression of protein-coding and non-**
689 **coding genes, but not in *MIRNA* genes**
690 **A:** Differentially expressed genes in *luc7a,b,rl* mutant compared to WT.
691 **B,C:** qRT-PCR analysis of selected ncRNA **(B)** and miRNAs **(C)** in WT and
692 *luc7a,b,rl*. Error bars denote the SEM (n=3).
693
694 **Figure 5: Global analysis of splicing defects in *luc7* triple mutant.**
695 **A:** Classification of splicing events changes in *luc7* triple mutant compared to WT.
696 **B-F:** Coverage plots and RT-PCR validation experiments for selected splicing events
697 in WT and *luc7* triple mutant. Genomic DNA (gDNA) or water (-) served as a control.
698 Primers positions are indicated with gray arrows. IR, intron retention; ES, exon
699 skipping; Alt.3'SS, alternative 3'splicing site; Alt.5'SS, alternative 5'splicing site.
700
701 **Figure 6: LUC7 proteins have a pronounced effect on terminal intron splicing**
702 **A:** RT-PCR analysis of LUC7-dependent introns in WT, *luc7* triple mutant, *cbp80*,
703 *cbp20* and *se-1* mutants.
704 **B:** RT-PCR analysis of CBC/SE-dependent introns in WT, *luc7* triple mutant, *cbp80*,
705 *cbp20* and *se-1* mutants.
706 **C:** Classification of intron retention according to the intron position (first, middle, or
707 last). Only genes with 3 or more introns were considered for this analysis. Terminal
708 introns are significantly enriched among introns retained in *luc7* triple mutants
709 (Fischer test; two-side $p < 0.0016$).

710 **D:** RT-PCR analysis of genes carrying retained terminal introns in WT and *luc7* triple
711 mutants.
712
713 **Figure 7: mRNAs containing retained LUC7-dependent terminal introns are**
714 **NMD-insensitive and remain nuclear.**
715 **A:** RT-PCR analysis of LUC7-dependent terminal introns in WT and NMD mutants
716 (*lba1* and *upf3-1*).
717 **B:** Splicing patterns of mRNAs isolated from total (T), cytosolic (C) and nuclear (N)
718 fractions.
719 **C:** Immunoblot analysis of proteins isolated from total, cytosolic and nuclear
720 fractions. Blots were probed with antibodies against histone H3 and a ribosomal
721 protein, L13-1.
722
723 **Figure 8: Splicing of LUC7 dependent terminal introns can be modulated by**
724 **stress.**
725 **A:** Seven days old WT seedlings were exposed to cold for 60 min. Splicing ratios
726 (unspliced/total RNA) of four genes featuring a LUC7-dependent terminal intron was
727 analyzed by qPCR. A T-test was performed for statistical analysis.
728 **B:** Model for the proposed function of LUC7 in Arabidopsis.
729
730 **Table 1: List of selected potential LUC7A interacting proteins identified in an**
731 **immunoprecipitation experiment followed by MS analysis.**
732
733 **Literature**
734 Almada AE, Wu X, Kriz AJ, Burge CB, Sharp PA (2013) Promoter directionality is
735 controlled by U1 snRNP and polyadenylation signals. *Nature* **499:** 360-363

736

737 AlShareef S, Ling Y, Butt H, Mariappan KG, Benhamed M, Mahfouz MM (2017)
738 Herboxidiene triggers splicing repression and abiotic stress responses in plants.
739 *BMC genomics* **18:** 260

740

741 Berg MG, Singh LN, Younis I, Liu Q, Pinto AM, Kaida D, Zhang Z, Cho S, Sherrill-Mix
742 S, Wan L, Dreyfuss G (2012) U1 snRNP determines mRNA length and regulates
743 isoform expression. *Cell* **150:** 53-64

744

745 Bielewicz D, Kalak M, Kalyna M, Windels D, Barta A, Vazquez F, Szweykowska-
746 Kulinska Z, Jarmolowski A (2013) Introns of plant pri-miRNAs enhance miRNA
747 biogenesis. *EMBO reports* **14:** 622-628

748

749 Boothby TC, Zipper RS, van der Weele CM, Wolniak SM (2013) Removal of retained
750 introns regulates translation in the rapidly developing gametophyte of Marsilea
751 vestita. *Developmental cell* **24:** 517-529

752

753 Boutz PL, Bhutkar A, Sharp PA (2015) Detained introns are a novel, widespread
754 class of post-transcriptionally spliced introns. *Genes & development* **29:** 63-80

755

756 Carvalho RF, Feijao CV, Duque P (2013) On the physiological significance of
757 alternative splicing events in higher plants. *Protoplasma* **250:** 639-650

758

759 Chen Y, Varani G (2013) Engineering RNA-binding proteins for biology. *The FEBS*
760 *journal* **280:** 3734-3754

761

762 Cho S, Hoang A, Sinha R, Zhong XY, Fu XD, Krainer AR, Ghosh G (2011)
763 Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-
764 70K snRNP protein determines early spliceosome assembly. *Proc Natl Acad Sci U S*
765 *A* **108:** 8233-8238
766
767 Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-
768 mediated transformation of Arabidopsis thaliana. *The Plant journal : for cell and*
769 *molecular biology* **16:** 735-743
770
771 Cooke C, Alwine JC (2002) Characterization of specific protein-RNA complexes
772 associated with the coupling of polyadenylation and last-intron removal. *Molecular*
773 *and cellular biology* **22:** 4579-4586
774
775 Cooke C, Hans H, Alwine JC (1999) Utilization of splicing elements and
776 polyadenylation signal elements in the coupling of polyadenylation and last-intron
777 removal. *Molecular and cellular biology* **19:** 4971-4979
778
779 Drechsel G, Kahles A, Kesarwani AK, Stauffer E, Behr J, Drewe P, Ratsch G,
780 Wachter A (2013) Nonsense-mediated decay of alternative precursor mRNA splicing
781 variants is a major determinant of the Arabidopsis steady state transcriptome. *Plant*
782 *Cell* **25:** 3726-3742
783
784 Durek P, Schmidt R, Heazlewood JL, Jones A, MacLean D, Nagel A, Kersten B,
785 Schulze WX (2010) PhosPhAt: the Arabidopsis thaliana phosphorylation site
786 database. An update. *Nucleic acids research* **38:** D828-834
787

788 Filichkin S, Priest HD, Megraw M, Mockler TC (2015) Alternative splicing in plants:
789 directing traffic at the crossroads of adaptation and environmental stress. *Current*
790 *opinion in plant biology* **24:** 125-135
791
792 Fortes P, Bilbao-Cortes D, Fornerod M, Rigaut G, Raymond W, Seraphin B, Mattaj
793 IW (1999a) Luc7p, a novel yeast U1 snRNP protein with a role in 5' splice site
794 recognition. *Genes & development* **13:** 2425-2438
795
796 Fortes P, Kufel J, Fornerod M, Polycarpou-Schwarz M, Lafontaine D, Tollervey D,
797 Mattaj IW (1999b) Genetic and physical interactions involving the yeast nuclear cap-
798 binding complex. *Molecular and cellular biology* **19:** 6543-6553
799
800 Fukumura K, Taniguchi I, Sakamoto H, Ohno M, Inoue K (2009) U1-independent pre-
801 mRNA splicing contributes to the regulation of alternative splicing. *Nucleic Acids Res*
802 **37:** 1907-1914
803
804 Gohring J, Jacak J, Barta A (2014) Imaging of endogenous messenger RNA splice
805 variants in living cells reveals nuclear retention of transcripts inaccessible to
806 nonsense-mediated decay in Arabidopsis. *Plant Cell* **26:** 754-764
807
808 Golovkin M, Reddy AS (1998) The plant U1 small nuclear ribonucleoprotein particle
809 70K protein interacts with two novel serine/arginine-rich proteins. *Plant Cell* **10:** 1637-
810 1648
811

812 Golovkin M, Reddy AS (2003) Expression of U1 small nuclear ribonucleoprotein 70K
813 antisense transcript using APETALA3 promoter suppresses the development of
814 sepals and petals. *Plant physiology* **132:** 1884-1891
815
816 Gonatopoulos-Pournatzis T, Cowling VH (2014) Cap-binding complex (CBC). *The*
817 *Biochemical journal* **457:** 231-242
818
819 Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical
820 user interface for sequence alignment and phylogenetic tree building. *Molecular*
821 *biology and evolution* **27:** 221-224
822
823 Guiro J, O'Reilly D (2015) Insights into the U1 small nuclear ribonucleoprotein
824 complex superfamily. *Wires Rna* **6:** 79-92
825
826 Heazlewood JL, Durek P, Hummel J, Selbig J, Weckwerth W, Walther D, Schulze
827 WX (2008) PhosPhAt: a database of phosphorylation sites in Arabidopsis thaliana
828 and a plant-specific phosphorylation site predictor. *Nucleic acids research* **36:**
829 D1015-1021
830
831 Heim A, Grimm C, Muller U, Haussler S, Mackeen MM, Merl J, Hauck SM, Kessler
832 BM, Schofield CJ, Wolf A, Bottger A (2014) Jumonji domain containing protein 6
833 (Jmjd6) modulates splicing and specifically interacts with arginine-serine-rich (RS)
834 domains of SR- and SR-like proteins. *Nucleic Acids Res* **42:** 7833-7850
835

836 Hernandez H, Makarova OV, Makarov EM, Morgner N, Muto Y, Krummel DP,
837 Robinson CV (2009) Isoforms of U1-70k control subunit dynamics in the human
838 spliceosomal U1 snRNP. *PloS one* **4:** e7202
839
840 Hoffman BE, Grabowski PJ (1992) U1 snRNP targets an essential splicing factor,
841 U2AF65, to the 3' splice site by a network of interactions spanning the exon. *Genes*
842 *& development* **6:** 2554-2568
843
844 Hori K, Watanabe Y (2005) UPF3 suppresses aberrant spliced mRNA in Arabidopsis.
845 *The Plant journal : for cell and molecular biology* **43:** 530-540
846
847 Kaida D (2016) The reciprocal regulation between splicing and 3'-end processing.
848 *Wiley Interdiscip Rev RNA* **7:** 499-511
849
850 Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G (2010) U1
851 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*
852 **468:** 664-668
853
854 Kalyna M, Simpson CG, Syed NH, Lewandowska D, Marquez Y, Kusenda B,
855 Marshall J, Fuller J, Cardle L, McNicol J, Dinh HQ, Barta A, Brown JW (2012)
856 Alternative splicing and nonsense-mediated decay modulate expression of important
857 regulatory genes in Arabidopsis. *Nucleic Acids Res* **40:** 2454-2469
858
859 Kaplan CD, Morris JR, Wu C, Winston F (2000) Spt5 and spt6 are associated with
860 active transcription and have characteristics of general elongation factors in D.
861 melanogaster. *Genes & development* **14:** 2623-2634

862

863 Kervestin S, Jacobson A (2012) NMD: a multifaceted response to premature
864 translational termination. *Nature reviews Molecular cell biology* **13:** 700-712

865

866 Knop K, Stepien A, Barciszewska-Pacak M, Taube M, Bielewicz D, Michalak M,
867 Borst JW, Jarmolowski A, Szweykowska-Kulinska Z (2016) Active 5' splice sites
868 regulate the biogenesis efficiency of Arabidopsis microRNAs derived from intron-
869 containing genes. *Nucleic Acids Res*

870

871 Koncz C, Dejong F, Villacorta N, Szakonyi D, Koncz Z (2012) The spliceosome-
872 activating complex: molecular mechanisms underlying the function of a pleiotropic
873 regulator. *Frontiers in plant science* **3:** 9

874

875 Kornblihtt AR, Schor IE, Allo M, Dujardin G, Petrillo E, Munoz MJ (2013) Alternative
876 splicing: a pivotal step between eukaryotic transcription and translation. *Nature*
877 *reviews Molecular cell biology* **14:** 153-165

878

879 Laubinger S, Sachsenberg T, Zeller G, Busch W, Lohmann JU, Ratsch G, Weigel D
880 (2008) Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA
881 splicing and microRNA processing in Arabidopsis thaliana. *Proc Natl Acad Sci U S A*
882 **105:** 8795-8800

883

884 Lewis JD, Izaurralde E, Jarmolowski A, McGuigan C, Mattaj IW (1996) A nuclear
885 cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5'
886 splice site. *Genes & Development* **10:** 1683-1698

887

888 Ling Y, Alshareef S, Butt H, Lozano-Juste J, Li L, Galal AA, Moustafa A, Momin AA,
889 Tashkandi M, Richardson DN, Fujii H, Arold S, Rodriguez PL, Duque P, Mahfouz MM
890 (2017) Pre-mRNA splicing repression triggers abiotic stress signaling in plants. *The*
891 *Plant journal : for cell and molecular biology* **89:** 291-309
892
893 Lorkovic ZJ, Barta A (2008) Role of Cajal bodies and nucleolus in the maturation of
894 the U1 snRNP in Arabidopsis. *PLoS One* **3:** e3989
895
896 Lorkovic ZJ, Wieczorek Kirk DA, Klahre U, Hemmings-Mieszczak M, Filipowicz W
897 (2000) RBP45 and RBP47, two oligouridylate-specific hnRNP-like proteins interacting
898 with poly(A)+ RNA in nuclei of plant cells. *RNA* **6:** 1610-1624
899
900 Merret R, Martino L, Bousquet-Antonelli C, Fneich S, Descombin J, Billey E, Conte
901 MR, Deragon JM (2013) The association of a La module with the PABP-interacting
902 motif PAM2 is a recurrent evolutionary process that led to the neofunctionalization of
903 La-related proteins. *RNA* **19:** 36-50
904
905 Millevoi S, Geraghty F, Idowu B, Tam JL, Antoniou M, Vagner S (2002) A novel
906 function for the U2AF 65 splicing factor in promoting pre-mRNA 3'-end processing.
907 *EMBO reports* **3:** 869-874
908
909 Millevoi S, Loulergue C, Dettwiler S, Karaa SZ, Keller W, Antoniou M, Vagner S
910 (2006) An interaction between U2AF 65 and CF I(m) links the splicing and 3' end
911 processing machineries. *The EMBO journal* **25:** 4854-4864
912

913 Muller-McNicoll M, Botti V, de Jesus Domingues AM, Brandl H, Schwich OD, Steiner
914 MC, Curk T, Poser I, Zarnack K, Neugebauer KM (2016) SR proteins are NXF1
915 adaptors that link alternative RNA processing to mRNA export. *Genes &*
916 *development* **30:** 553-566
917
918 Nakagawa T, Suzuki T, Murata S, Nakamura S, Hino T, Maeo K, Tabata R, Kawai T,
919 Tanaka K, Niwa Y, Watanabe Y, Nakamura K, Kimura T, Ishiguro S (2007) Improved
920 Gateway binary vectors: high-performance vectors for creation of fusion constructs in
921 transgenic analysis of plants. *Bioscience, biotechnology, and biochemistry* **71:** 2095-
922 2100
923
924 Naro C, Jolly A, Di Persio S, Bielli P, Setterblad N, Alberdi AJ, Vicini E, Geremia R,
925 De la Grange P, Sette C (2017) An Orchestrated Intron Retention Program in Meiosis
926 Controls Timely Usage of Transcripts during Germ Cell Differentiation.
927 *Developmental cell* **41:** 82-93 e84
928
929 Ohtani M, Sugiyama M (2005) Involvement of SRD2-mediated activation of snRNA
930 transcription in the control of cell proliferation competence in Arabidopsis. *The Plant*
931 *journal : for cell and molecular biology* **43:** 479-490
932
933 Papp I, Mur LA, Dalmadi A, Dulai S, Koncz C (2004) A mutation in the Cap Binding
934 Protein 20 gene confers drought tolerance to Arabidopsis. *Plant molecular biology*
935 **55:** 679-686
936

937 Pimentel H, Parra M, Gee SL, Mohandas N, Pachter L, Conboy JG (2016) A dynamic
938 intron retention program enriched in RNA processing genes regulates gene
939 expression during terminal erythropoiesis. *Nucleic Acids Res* **44:** 838-851
940
941 Prigge MJ, Wagner DR (2001) The arabidopsis serrate gene encodes a zinc-finger
942 protein required for normal shoot development. *Plant Cell* **13:** 1263-1279
943
944 Puig O, Bragado-Nilsson E, Koski T, Seraphin B (2007) The U1 snRNP-associated
945 factor Luc7p affects 5' splice site selection in yeast and human. *Nucleic Acids Res*
946 **35:** 5874-5885
947
948 Raczynska KD, Simpson CG, Ciesiolka A, Szewc L, Lewandowska D, McNicol J,
949 Szweykowska-Kulinska Z, Brown JW, Jarmolowski A (2010) Involvement of the
950 nuclear cap-binding protein complex in alternative splicing in Arabidopsis thaliana.
951 *Nucleic Acids Res* **38:** 265-278
952
953 Raczynska KD, Stepien A, Kierzkowski D, Kalak M, Bajczyk M, McNicol J, Simpson
954 CG, Szweykowska-Kulinska Z, Brown JW, Jarmolowski A (2013) The SERRATE
955 protein is involved in alternative splicing in Arabidopsis thaliana. *Nucleic Acids Res*
956
957 Reddy AS, Marquez Y, Kalyna M, Barta A (2013) Complexity of the alternative
958 splicing landscape in plants. *Plant Cell* **25:** 3657-3683
959
960 Rigo F, Martinson HG (2008) Functional coupling of last-intron splicing and 3'-end
961 processing to transcription in vitro: the poly(A) signal couples to splicing before
962 committing to cleavage. *Molecular and cellular biology* **28:** 849-862

963

964 Rosel TD, Hung LH, Medenbach J, Donde K, Starke S, Benes V, Ratsch G, Bindereif
965 A (2011) RNA-Seq analysis in mutant zebrafish reveals role of U1C protein in
966 alternative splicing regulation. *The EMBO journal* **30:** 1965-1976

967

968 Rowley MJ, Bohmdorfer G, Wierzbicki AT (2013) Analysis of long non-coding RNAs
969 produced by a specialized RNA polymerase in Arabidopsis thaliana. *Methods* **63:**
970 160-169

971

972 Roy D, Bhanja Chowdhury J, Ghosh S (2013) Polypyrimidine tract binding protein
973 (PTB) associates with intronic and exonic domains to squelch nuclear export of
974 unspliced RNA. *FEBS letters* **587:** 3802-3807

975

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

980

981 Schwab R, Speth C, Laubinger S, Voinnet O (2013) Enhanced microRNA
982 accumulation through stemloop-adjacent introns. *EMBO reports* **14:** 615-621

983

984 Shao W, Kim HS, Cao Y, Xu YZ, Query CC (2012) A U1-U2 snRNP interaction
985 network during intron definition. *Molecular and cellular biology* **32:** 470-478

986

987 Shaul O (2015) Unique Aspects of Plant Nonsense-Mediated mRNA Decay. *Trends*
988 *Plant Sci* **20:** 767-779

989

990 Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y (2014) rMATS:
991 robust and flexible detection of differential alternative splicing from replicate RNA-
992 Seq data. *Proc Natl Acad Sci U S A* **111:** E5593-5601

993

994 Staiger D, Brown JW (2013) Alternative splicing at the intersection of biological
995 timing, development, and stress responses. *Plant Cell* **25:** 3640-3656

996

997 Stepien A, Knop K, Dolata J, Taube M, Bajczyk M, Barciszewska-Pacak M, Pacak A,
998 Jarmolowski A, Szweykowska-Kulinska Z (2017) Posttranscriptional coordination of
999 splicing and miRNA biogenesis in plants. *Wiley Interdiscip Rev RNA* **8**

1000

1001 Sullivan C, Howard PL (2016) The Diverse Requirements of ARS2 in nuclear cap-
1002 binding complex-dependent RNA Processing. *Rna & Disease*

1003

1004 Sun M, Lariviere L, Dengl S, Mayer A, Cramer P (2010) A tandem SH2 domain in
1005 transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-
1006 terminal repeat domain (CTD). *The Journal of biological chemistry* **285:** 41597-41603

1007

1008 Umezawa T, Sugiyama N, Takahashi F, Anderson JC, Ishihama Y, Peck SC,
1009 Shinozaki K (2013) Genetics and phosphoproteomics reveal a protein
1010 phosphorylation network in the abscisic acid signaling pathway in Arabidopsis
1011 thaliana. *Sci Signal* **6:** rs8

1012

1013 Wachter A, Tunc-Ozdemir M, Grove BC, Green PJ, Shintani DK, Breaker RR (2007)
1014 Riboswitch control of gene expression in plants by splicing and alternative 3' end
1015 processing of mRNAs. *The Plant cell* **19:** 3437-3450
1016
1017 Wahl MC, Will CL, Luhrmann R (2009) The spliceosome: design principles of a
1018 dynamic RNP machine. *Cell* **136:** 701-718
1019
1020 Wang BB, Brendel V (2004) The ASRG database: identification and survey of
1021 Arabidopsis thaliana genes involved in pre-mRNA splicing. *Genome biology* **5:** R102
1022
1023 Wang C, Tian Q, Hou Z, Mucha M, Aukerman M, Olsen OA (2007) The Arabidopsis
1024 thaliana AT PRP39-1 gene, encoding a tetratricopeptide repeat protein with similarity
1025 to the yeast pre-mRNA processing protein PRP39, affects flowering time. *Plant cell*
1026 *reports* **26:** 1357-1366
1027
1028 Wang P, Xue L, Batelli G, Lee S, Hou YJ, Van Oosten MJ, Zhang H, Tao WA, Zhu
1029 JK (2013) Quantitative phosphoproteomics identifies SnRK2 protein kinase
1030 substrates and reveals the effectors of abscisic acid action. *Proc Natl Acad Sci U S A*
1031 **110:** 11205-11210
1032
1033 Webby CJ, Wolf A, Gromak N, Dreger M, Kramer H, Kessler B, Nielsen ML, Schmitz
1034 C, Butler DS, Yates JR, 3rd, Delahunty CM, Hahn P, Lengeling A, Mann M,
1035 Proudfoot NJ, Schofield CJ, Bottger A (2009) Jmjd6 catalyses lysyl-hydroxylation of
1036 U2AF65, a protein associated with RNA splicing. *Science* **325:** 90-93
1037

1038 Will CL, Luhrmann R (2001) Spliceosomal UsnRNP biogenesis, structure and
1039 function. *Current opinion in cell biology* **13:** 290-301
1040
1041 Will CL, Luhrmann R (2011) Spliceosome structure and function. *Cold Spring Harbor*
1042 *perspectives in biology* **3**
1043
1044 Wong JJ, Au AY, Ritchie W, Rasko JE (2016) Intron retention in mRNA: No longer
1045 nonsense: Known and putative roles of intron retention in normal and disease
1046 biology. *BioEssays : news and reviews in molecular, cellular and developmental*
1047 *biology* **38:** 41-49
1048
1049 Workman E, Veith A, Battle DJ (2014) U1A regulates 3' processing of the survival
1050 motor neuron mRNA. *The Journal of biological chemistry* **289:** 3703-3712
1051
1052 Xing D, Wang Y, Hamilton M, Ben-Hur A, Reddy AS (2015) Transcriptome-Wide
1053 Identification of RNA Targets of Arabidopsis SERINE/ARGININE-RICH45 Uncovers
1054 the Unexpected Roles of This RNA Binding Protein in RNA Processing. *Plant Cell*
1055
1056 Yap K, Lim ZQ, Khandelia P, Friedman B, Makeyev EV (2012) Coordinated
1057 regulation of neuronal mRNA steady-state levels through developmentally controlled
1058 intron retention. *Genes Dev* **26:** 1209-1223
1059
1060 Yoine M, Ohto MA, Onai K, Mita S, Nakamura K (2006) The lba1 mutation of UPF1
1061 RNA helicase involved in nonsense-mediated mRNA decay causes pleiotropic
1062 phenotypic changes and altered sugar signalling in Arabidopsis. *The Plant journal :*
1063 *for cell and molecular biology* **47:** 49-62

Zhan X, Qian B, Cao F, Wu W, Yang L, Guan Q, Gu X, Wang P, Okusolubo TA, Dunn SL, Zhu JK, Zhu J (2015) An Arabidopsis PWI and RRM motif-containing protein is critical for pre-mRNA splicing and ABA responses. *Nature communications* **6:** 8139

Table 1: List of selected potential LUC7A interacting proteins identified in a co-immunoprecipitation experiment followed by MS analysis.

Identified protein	Aver. # of Peptides		Seq.Coverage (%)		MW (kDa)
	IP LUC7	IP WT	IP LUC7	IP WT	
U1 snRNP components					
LUC7A	45.8	0	71.9	0	47.4
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	2.8	0	12.3	0	20.2
SR45	1.8	0	5.2	0	44.6
RSZ21;RSZ22;RSZ22A	1.8	0	8.0	0	22.5
SR30	1	0	3.9	0	29.1
RS2Z33	1	0	2.9	0	32.9
SCL30	0.8	0	2.8	0	29.6
U2AF65A	0.8	0	1.7	0	60.7
Kinases					
SRPK4	8.5	0	16.2	0	59.4
SRPK3	4.3	0	8.9	0	61.2
Proteins linked to 3'end processing/polyA binding					
SPT6L (AT1g63210, AT1g65440)	8.5	1	4.8	0.9	185.0
Polyadenylate-binding protein RBP47C	1.5	0.3	4.0	0.7	48.6
La-related protein 6B (LARP6)	2.3	0.7	5.5	1.9	60.6

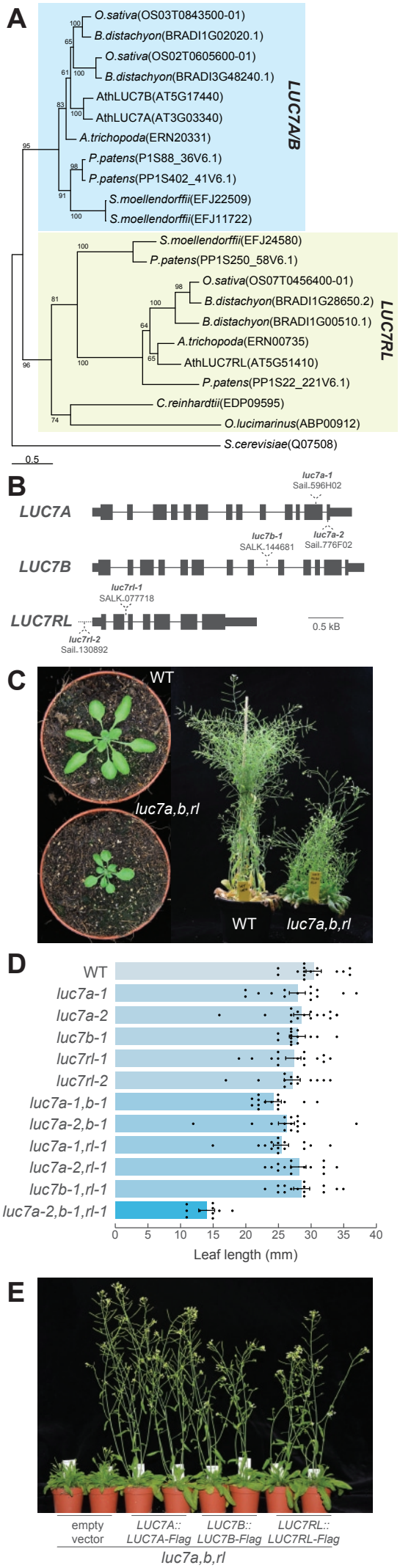

Figure 1: *Arabidopsis* LUC7 proteins redundantly control plant development

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 1

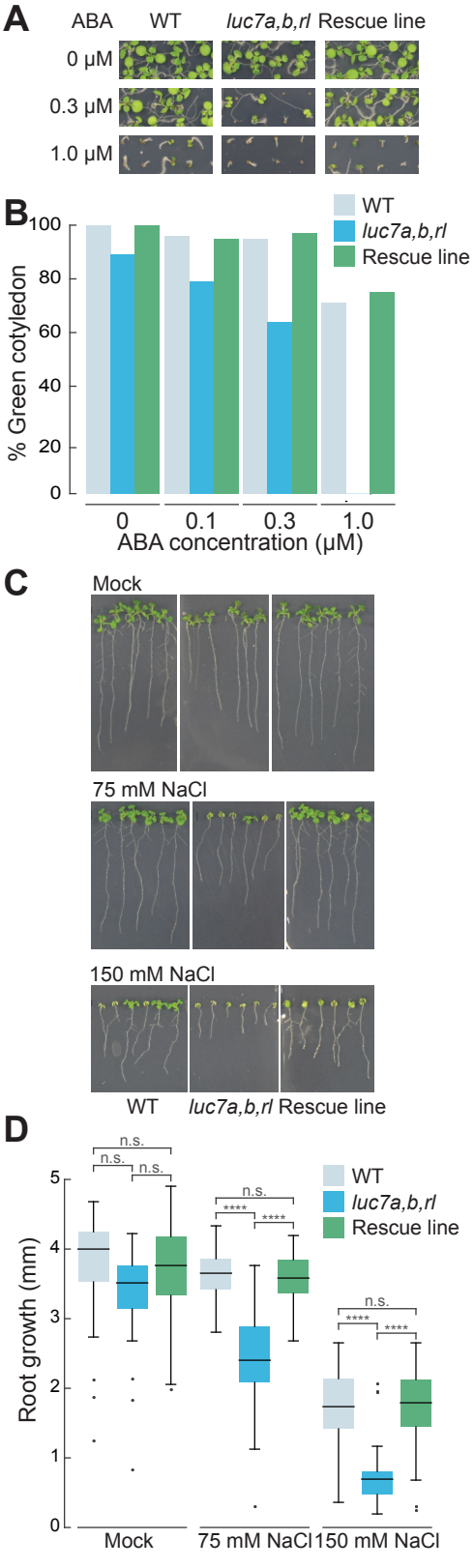

FIGURE 2

Figure 2: *Arabidopsis* LUC7 is involved in ABA signaling and salt stress responses

A,B: WT, *luc7* triple mutant and a *luc7* rescue line (*luc7a,b,rl; pLUC7A:LUC7A-YFP*) were grown on half-strength MS plates containing 1% sucrose and indicated amount of ABA. Seedling phenotypes (A) and quantification of seedlings with green cotyledons (B) are shown. Green cotyledons were scored ten days after germination. One of two biological replicates is shown.

C,D: WT, *luc7* triple mutant and a *luc7* rescue line (*luc7a,b,rl; pLUC7A:LUC7A-YFP*) were germinated on half-strength MS vertical plates and seedling were transferred on half-strength MS plates containing the indicated amount of NaCl. Plates were always placed vertically and the root growth was scored over 7 days. Phenotypes (C) and root length quantification (D) are shown.

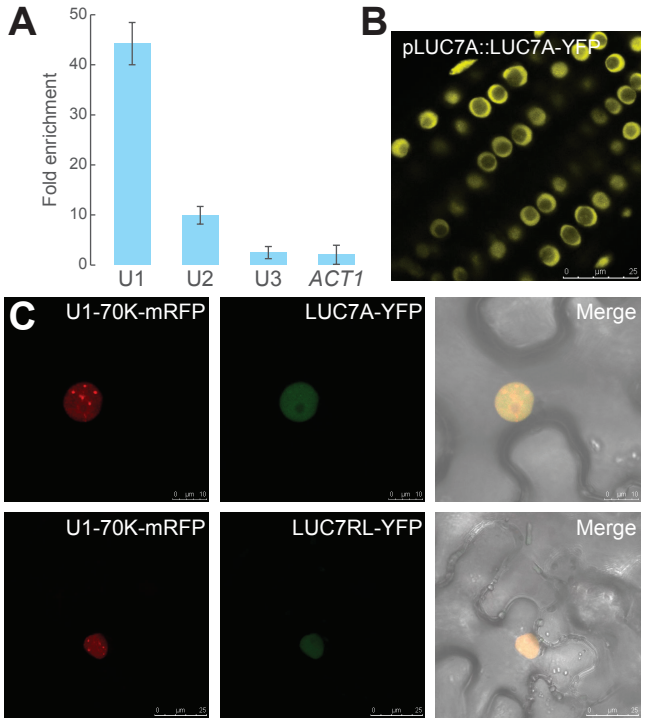

FIGURE 3

Figure 3: *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 line 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.

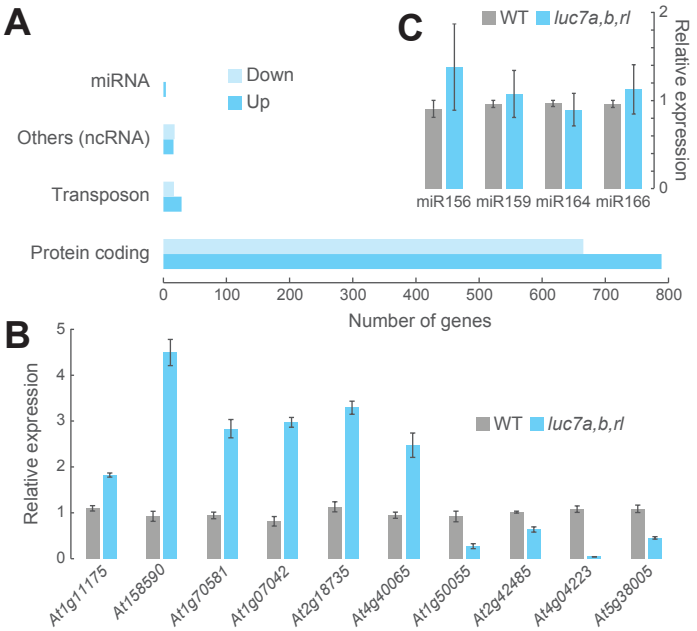

FIGURE 4

Figure 4: Mutations in *LUC7* result 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).

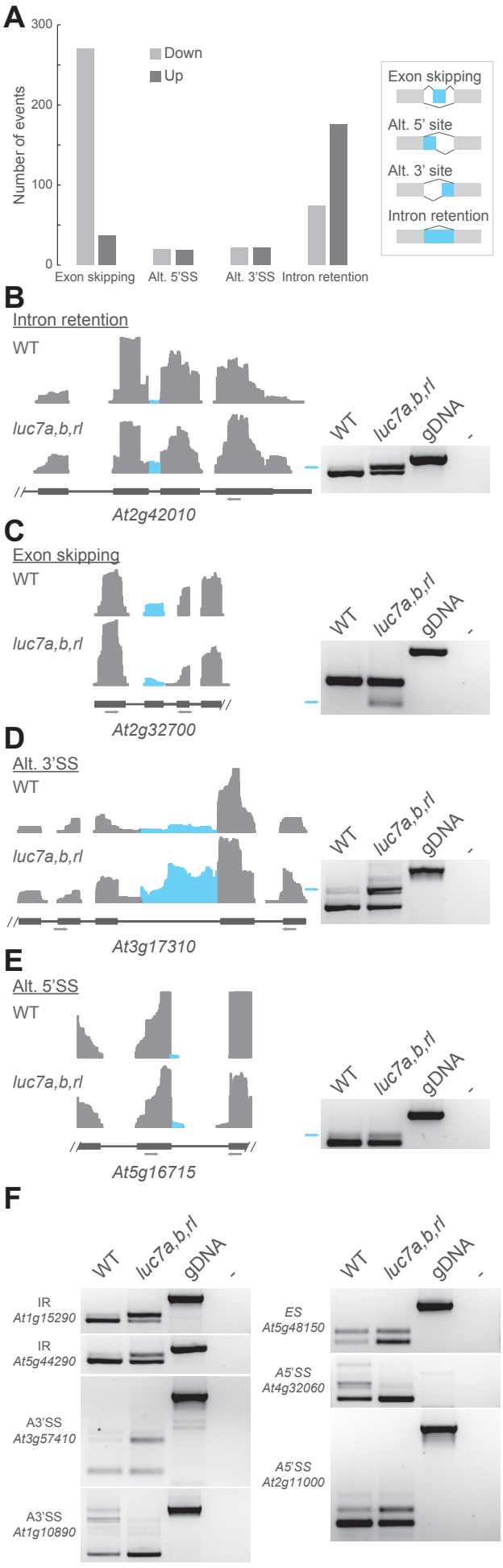

Figure 5: Global analysis of splicing defects in *luc7* triple mutant.
A: Classification of splicing events changes in *luc7* triple mutant compared to WT.
B-F: Coverage plots and RT-PCR validation experiments for selected splicing events in WT and *luc7* triple mutant. Genomic DNA (gDNA) or water (-) 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

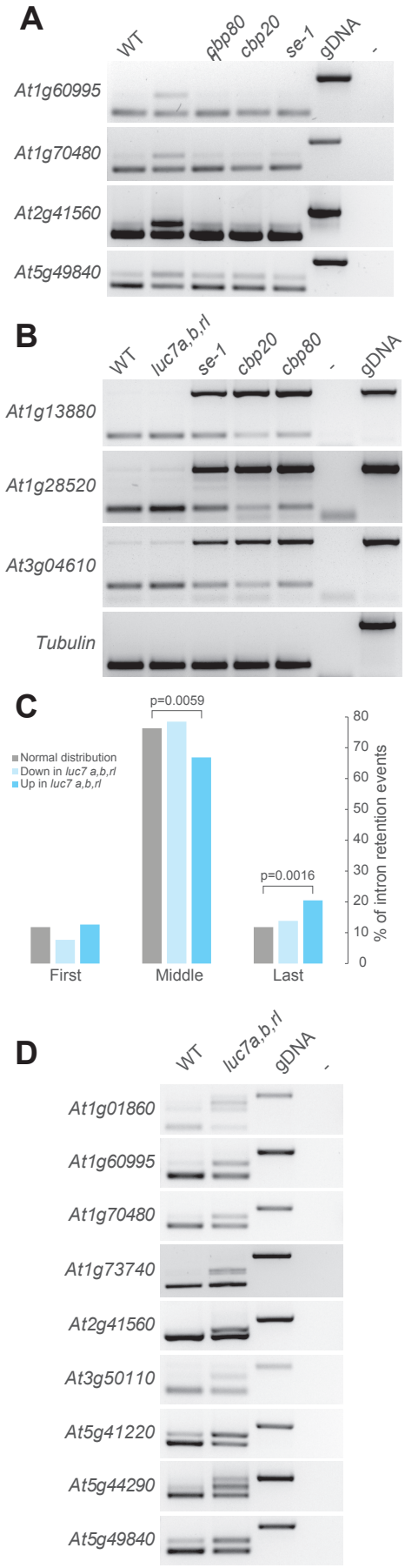

FIGURE 6

Figure 6: 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.

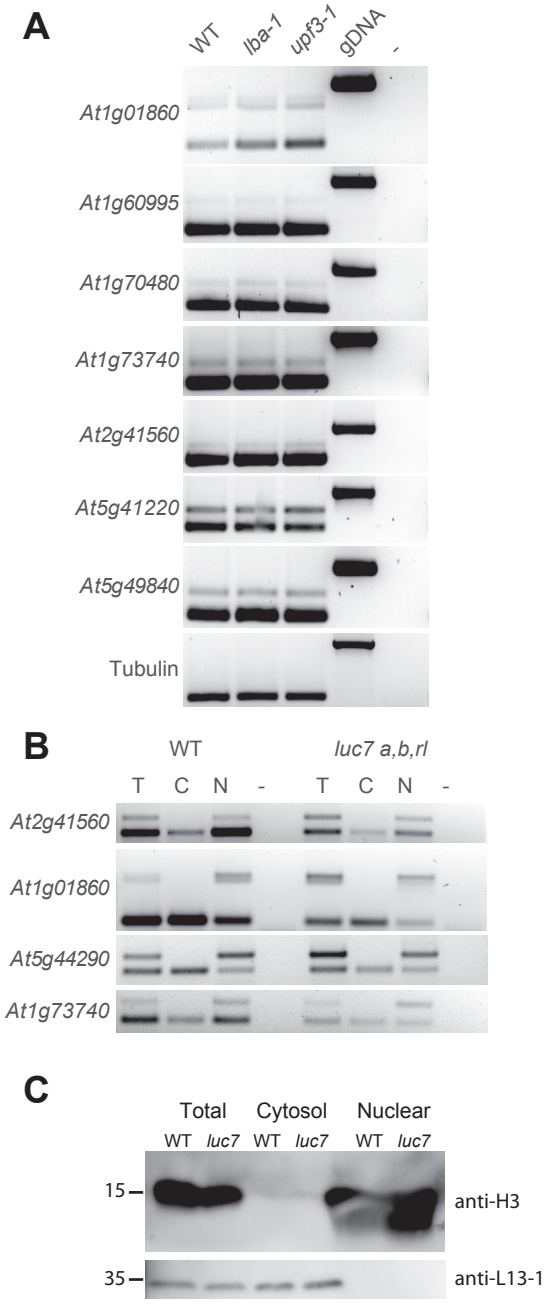

FIGURE 7

Figure 7: mRNAs containing retained LUC7-dependent terminal introns are NMD-insensitive and remain nuclear.
A: RT-PCR analysis of LUC7-dependent terminal introns in WT and NMD mutants (*lba1* and *upf3-1*).
B: Splicing patterns of mRNAs isolated from total (T), cytosolic (C) and nuclear (N) fractions.
C: Immunoblot analysis of proteins isolated from total, cytosolic and nuclear fractions. Blots were probed with antibodies against histone H3 and a ribosomal protein, L13-1.

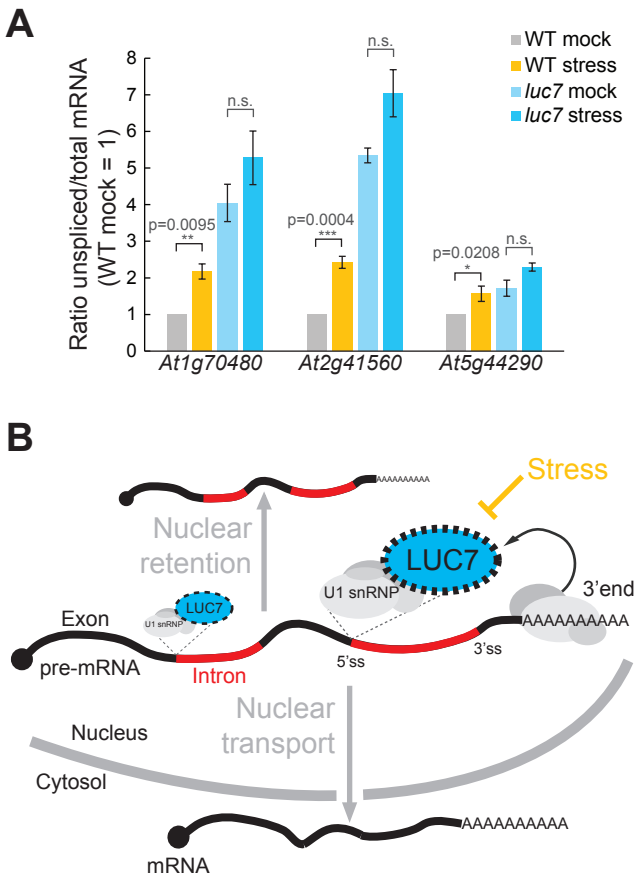

FIGURE 8

Figure 8: 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. A T-test was performed for statistical analysis.
B: Model for the proposed function of LUC7 in Arabidopsis.