1   **Title:**

2   **Global transcriptome analysis reveals circadian control of splicing events in *Arabidopsis***

3   ***thaliana***

4   Andrés Romanowski[1, *], Rubén G. Schlaen[1, †], Soledad Perez-Santangelo[1, †], Estefanía Mancini and

5   Marcelo J. Yanovsky[1, *]

6   [1] Comparative Genomics of Plant Development, Fundación Instituto Leloir, Instituto de

7   Investigaciones Bioquímicas Buenos Aires (IIBBA) - Consejo Nacional de Investigaciones Científicas

8   y Técnicas (CONICET), C1405BWE Buenos Aires, Argentina.

9   [†] Authors contributed equally to this work.

10   * To whom correspondence should be addressed. Tel: +54-11-5238-7500 (#3103); Email:

11   myanovsky@leloir.org.ar

12   Correspondence may also be addressed to Andres Romanowski, Tel: +54-11-5238-7500 (#2103);

13   Email: aromanowski@leloir.org.ar

14   Present Address:

15   Soledad Perez-Santangelo, Department of Biochemistry, Division of Health Sciences, University of

16   Otago, Dunedin, New Zealand.

17   Estefanía Mancini, Centre for Genomic Regulation (CRG), Parc de Recerca Biomedica de Barcelona

18   (PRBB), Barcelona, Spain.

19

20   **SUMMARY**

21   • The circadian clock of *Arabidopsis thaliana* controls many physiological and molecular

22   processes, allowing plants to anticipate daily changes in their environment. However,

23   developing a detailed understanding of how oscillations in mRNA levels are connected to

24   oscillations in post-transcriptional processes, such as splicing, has remained a challenge.

25   • Here we applied a combined approach using deep transcriptome sequencing and

26   bioinformatics tools to identify novel circadian regulated genes and splicing events.

27   • Using a stringent approach, we identified 300 intron retention, 8 exon skipping, 79 alternative

28   3' splice site usage, 48 alternative 5' splice site usage, and 350 multiple (more than one event

29   type) annotated events under circadian regulation. We also found 7 and 721 novel alternative

30   exonic and intronic events. Depletion of the circadian regulated splicing factor AtSPF30

31   homolog, resulted in the disruption of a subset of clock controlled splicing events.

32   • Altogether, our global circadian RNA-seq coupled with an *in silico*, event centred, splicing

33   analysis tool offers a new approach for studying the interplay between the circadian clock and

34   the splicing machinery at a global scale. The identification of many circadian regulated

35 splicing events broadens our current understanding of the level of control that the circadian
36 clock has over this posttranscriptional regulatory layer.

37 **Key words:** circadian clock, circadian rhythms, alternative splicing, splicing factors, SPF30,
38 Arabidopsis thaliana

39

40 **INTRODUCTION**

41 Land plants have evolved in a world that spins around its axis with a period close to 24 h.
42 Consequently, they are subject to daily changes in the environment, such as changes in light quality /
43 intensity and temperature. The cycling nature of these changes has allowed for the evolution of
44 endogenous circadian timekeeping mechanisms that allow plants, and other organisms, to predict the
45 timing of these changes, which contributes to the optimization of their growth and development
46 (Green *et al.*, 2002; Dunlap *et al.*, 2004; Dodd *et al.*, 2005).
47 In *Arabidopsis thaliana*, the molecular circadian clock consists of an interlocked series of
48 transcriptional-translational feedback loops that cycle with a period close to 24 h (Harmer *et al.*, 2001;
49 McClung, 2014; Millar, 2016; Hernando *et al.*, 2017). The basic layout of this clock consists of the
50 morning MYB transcription factor elements CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) and LATE
51 ELONGATED HYPOCOTYL (LHY). These elements peak in the morning and repress the
52 transcription of TIMING OF CAB EXPRESSION 1 (TOC1), also known as PSEUDO RESPONSE
53 REGULATOR 1 (PRR1). TOC1, in turn, peaks in the evening and represses CCA1/LHY. Another loop
54 involves the downregulation of CCA1/LHY throughout the rest of the day by the consecutive
55 repressive action of PRR9, PRR7 and later PRR5. During the evening, the elements EARLY
56 FLOWERING 3 (ELF3), ELF4 and LUX ARRHYTHMO (LUX) associate into an "evening complex"
57 (EC), repressing PRR9 and LUX. Additionally, several activators have been described. In particular,
58 the CCA1/LHY homologs reveille 8 (RVE8), RVE6 and RVE4 are morning factors that drive the
59 expression of PRR5 and TOC1. Together with the RVEs, the NIGHT LIGHT-INDUCIBLE AND
60 CLOCK-REGULATED (LNK) 1-4 family of genes also contribute to activate the expression of PRR5
61 and ELF4 (McClung, 2014).
62 This circadian clock integrates environmental inputs to fine tune its oscillations (entrainment) and then
63 controls the timing of several molecular, physiological and developmental outputs, like stomatal
64 opening, leaf movement, hormone synthesis and growth among others (McClung, 2006; Millar, 2016).
65 One of the many ways that this can be accomplished is through the control of gene expression, a
66 process that involves many different steps (Romanowski & Yanovsky, 2015). One such step is the
67 processing of precursor mRNAs (pre-mRNAs). Plant pre-mRNAs are composed of continuous
68 nucleotide segments known as exons and introns. In order to produce a mature functional mRNA,
69 exonic segments are kept and introns are removed through a process called splicing.
70 The process of splicing occurs at the spliceosome, a dynamic protein-RNA complex composed of
71 small nuclear ribonucleoprotein particles (snRNPs) and auxiliary proteins that assemble at exon-intron
72 boundaries known as donor 5'-splice site (ss) and acceptor 3'-ss. The selection and recognition of

73 these splice sites occur through the interaction of spliceosome components with *cis* acting sequences

74 and the *trans* acting proteins. Alternative splicing (AS) occurs through the alternative selection of 5'ss,

75 3'ss, exon inclusion or skipping and intron retention (Kornblihtt *et al.*, 2013). Through this process,

76 different functional mRNAs can arise from the transcription of the same gene, ultimately expanding

77 the repertoire of protein products that can be encoded by a single gene (Kornblihtt *et al.*, 2013; Saldi

78 *et al.*, 2016).

79 The first evidence of crosstalk between AS and the circadian clock was described in our lab in 2010.

80 A mutation in the clock regulated protein arginine methyltransferase 5 (PRMT5) was found to alter the

81 AS of the clock component PRR9, causing a significant alteration in period length (Sanchez *et al.*,

82 2010). Since then, additional work has shown that AS is a mechanism that can couple changes in

83 environmental temperature to circadian timing by altering the splicing of clock components (James *et*

84 *al.*, 2012a; James *et al.*, 2012b; Perez-Santangelo *et al.*, 2013; Filichkin *et al.*, 2015). Also, mutations

85 in the spliceosome components SKIP and STIPL1 can affect circadian timing (Jones *et al.*, 2012;

86 Wang *et al.*, 2012). However, no comprehensive survey has been done to date to understand how

87 widespread clock regulated AS is. In this work, we used mRNA-seq and a custom *in house* developed

88 bioinformatics pipeline to perform a genome wide analysis on circadian regulation of AS in *A. thaliana*

89 plants.

90

91 **MATERIALS AND METHODS**

92 **Plant materials**

93 All *Arabidopsis thaliana* plants used in this study were of the Columbia (Col-0) ecotype. *AT2G02570*

94 mutant alleles *atspf30-1* (SALK_052016C) and *atspf30-2* (SALK_081292C) were obtained from the

95 Arabidopsis Biological Resource Center (ABRC).

96 **Growth conditions and tissue collection for RNA-seq studies**

97 Col-0 and *atspf30-1* seeds were sown on Murashige and Skoog medium (GIBCO BRL) and cold

98 stratified (4ºC, darkness) for 3 days. For non-circadian experiments, the MS plates were then

99 transferred to 22 °C and continuous light (LL; 50 μmol·m^{-2}·s^{-1} of white light). For circadian

100 experiments, the MS plates were transferred to 22 °C under LD12:12 (12 h light/12 h dark cycles; 50

101 μmol·m^{-2}·s^{-1} of white light) conditions for 11 days and then transferred to continuous light (LL; 50

102 μmol·m^{-2}·s^{-1} of white light).

103 Samples containing 15-30 seedlings were flash frozen in liquid nitrogen. For circadian studies, the

104 samples were collected at the timepoints indicated in Fig. 1A, with 2 biological replicates per timepoint.

105 For non-circadian studies, 3 biological replicates of Col-0 and *atspf30-1* were sampled. After all

106 samples were collected, total RNA was extracted using the Trizol method according to the

107 manufacturer's protocol (ThermoFisher) and quantified using a NanoDrop spectrophotometer

108 (ThermoFisher). Total RNA integrity was further checked by agarose gel electrophoresis.

109 **Library preparation and RNA sequencing**

110 Libraries were prepared following the TruSeq RNA v2 Sample Preparation Guide (Illumina). Briefly, 3

111 μg of total RNA was polyA-purified and fragmented, and first-strand cDNA synthesized by reverse

112 transcriptase (SuperScript II; Invitrogen) and random hexamers. This was followed by RNA

113　degradation and second-strand cDNA synthesis. End repair process and addition of a single A

114　nucleotide to the 3′ ends allowed ligation of multiple indexing adapters. Then, an enrichment step of

115　12 cycles of PCR was performed. Library validation included size and purity assessment with the

116　Agilent 2100 Bioanalyzer and the Agilent DNA 1000 kit (Agilent Technologies). Samples were diluted

117　to 10nM and pooled to create the 12X multiplexed DNA libraries, and 12–15 pmol of material was

118　loaded onto each lane. Template hybridization, extension of the template, amplification of the

119　template, cluster generation, sequencing primer addition, and single end chemistry were performed

120　using C-BOT (formerly known as Cluster Station). Base calls were made using CASAVA.

121　The multiplexed samples were loaded onto two lanes on an Illumina HiSeq 2500 sequencer, providing

122　100-bp single-end reads.

123　The RNA-seq fastq files supporting the conclusions of this article are available in the ArrayExpress

124　(Kolesnikov *et al*., 2015) database at EMBL-EBI (www.ebi.ac.uk/arrayexpress), under accession

125　numbers E-MTAB-7933 (circadian dataset) and E-MTAB-8667 (atspf30-1 dataset).

126　**Processing of RNA reads and alignment**

127　Reads were quality-filtered using the standard Illumina process and demultiplexed with two allowed

128　barcode mismatches. Sequence files were generated in FASTQ format. The Bowtie/TopHat suite

129　(Trapnell *et al.*, 2009) was used to map reads to the *A. thaliana* TAIR10 reference genome. Along

130　with the prebuilt *A. thaliana* index, the reference genome was downloaded from ENSEMBL

131　(December 2014). Default values for TopHat parameters were used with the exception of maximum

132　intron length parameter, which was set to a value of 5,000 nt following estimated values, as previously

133　reported (Hong *et al.*, 2006). Table S1 provides a summary table of main read count statistics.

134　**RNA sequencing analysis**

135　Read counts were calculated for genes, exon bins, AS bins and intron bins with the ASpli package

136　(Mancini *et al.*, 2019) using custom scripts written in R (64-bit, version 3.6.0). Multidimensional scaling

137　(MDS) plots of distances between gene expression profiles were generated with the limma package

138　(Ritchie *et al.*, 2015), using custom script written in R. Coverage plots were generated using the IGB

139　Browser (Freese *et al.*, 2016).

140　Detection of cycling was performed using JTK_Cycle (Hughes *et al*., 2010) implemented in R, using

141　24±4 h periods as limits. Obtained p-values were FDR corrected. Recalculated circadian phases were

142　obtained by multiplying the JTK obtained phase (LAG) by 24 and dividing that number by the JTK

143　obtained period (PER). Relative amplitude estimates were calculated by dividing the subtraction of the

144　absolute peak expression and through levels, by the sum of through level plus one (to avoid divisions

145　by zero). Heatmaps were generated using scripts for R. PSI/PIR indexes were calculated as

146　previously described (Pervouchine *et al.*, 2013; Gueroussov *et al.*, 2015) using ASpli implemented in

147　R. GO analyses were performed with the Virtual Plant web service (Katari *et al.*, 2010). All code and

148　analyses are available on demand.

149　**Semiquantitative and reverse transcriptase PCR (sqRT-PCR)**

150　Plants were grown and sampled as described above (see 'Growth conditions and tissue collection for

151　RNA-seq studies'). Total RNA was obtained from these samples using TRIzol reagent (Invitrogen).

152　One microgram of RNA was treated with RQ1 RNase-Free DNase (Promega) and subjected to retro

153   transcription with Superscript III (Invitrogen) and oligo dT according to the manufacturer's instructions.

154   Amplification of cDNA was carried out for 30 cycles to measure splicing at the linear phase. RT-PCR

155   products were electrophoresed in 8% SDS-PAGE gels and detected with EtBr. Primers used for

156   amplification of each gene are described in Table S2.

157   **Quantitative reverse transcriptase PCR (qRT-PCR)**

158   cDNAs obtained for 'Semiquantitative and reverse transcriptase PCR (sqRT-PCR)' were also used for

159   quantitative RT-PCR amplification. Synthesized cDNAs were amplified with FastStart Universal SYBR

160   GreenMaster (Roche) using the Mx3000P Real Time PCR System (Agilent Technologies) cycler. The

161   PP2A (AT1G13320) transcript was used as a reference gene.

162   Quantitative RT-PCR quantification was conducted using the standard curve method as described in

163   the Methods and Applications Guide from Agilent Technologies. Primers and targets for gene

164   expression and splicing events validations are described in Table S2.

165   **Circadian Leaf Movement Analysis**

166   For leaf movement analysis, Col-0 and *atspf30-1* plants were entrained under a 16 h light : 8 h dark

167   cycle, transferred to continuous white fluorescent light, and the position of the first pair of leaves was

168   recorded every 2 h for 5–6 d using Image J software (http://imagej.nih.gov). Period estimates were

169   calculated using Brass 3.0 software (Biological Rhythms Analysis Software System;

170   http://www.amillar.org). Briefly, raw data was imported; baseline detrended and analysed using the

171   FFT-NLLS algorithm, as described previously (Plautz *et al.*, 1997). Statistical analysis was performed

172   using a two tailed Student's t test, after testing for normal distribution of the data.

173   **Flowering Time Analysis**

174   For flowering time analysis, Col-0, *atspf30-1* and *atspf30-2* plants were grown under long days (16 h

175   light : 8 h dark), short days (8 h light : 16 h dark) or continuous light at a constant temperature of

176   22 °C. Flowering time was estimated by counting the number of rosette leaves at the time of bolting.

177   Statistical analysis was performed using Dunnet's multiple comparisons test using Col-0 as a control,

178   after testing for normal distribution of the data.

179   **Phylogenetic Tree Reconstruction**

180   Splicing Factor 30 (SPF30) and Survival Motor Neuron (SMN) protein sequences from *Homo sapiens*,

181   *Rattus norvergicus*, *Mus musculus*, *Danio rerio*, *C. elegans* and *D. melanogaster*. Homolog

182   sequences from *A. thaliana*, *Capsella rubella*, *Solanum lycopersicum*, *Oryza sativa*, *Sorghum bicolor*,

183   were identified with BlastP and downloaded from GenBank. *Physcomitrella patens* sequences were

184   obtained from PEATMoss PpGML DB v1.6 (Fernandez-Pozo *et al.*, 2019). Using these sequences as

185   input, a phylogenetic tree was reconstructed using SeaView 4.7 (Gouy *et al.*, 2010). Sequences were

186   aligned with Clustal Omega 1.2.0 (Sievers *et al.*, 2011) with default options. PhyML 3.1 (Guindon *et*

187   *al.*, 2010) was used for the phylogenetic tree reconstruction, using default parameters except model

188   was set to LG amino-acid replacement matrix (Le & Gascuel, 2008), and bootstrapping was set to

189   1000 replicates.

190

191   **RESULTS**

192 **Global RNA-seq of the circadian transcriptome reveals known and novel circadian regulated**
193 **genes.**

194 To analyse the circadian regulation of gene expression, we performed a global transcriptome deep
195 sequencing of 15-day old *A. thaliana* plants grown under a light/dark (LD) 12 h : 12 h photoperiod,
196 transferred to constant light conditions (LL), and harvested after 24 h in LL every 4 h for a total of 2
197 days, with two biological replicates (Fig. 1a). This approach, in contrast to previous studies that used
198 microarray-based approaches, allowed us to analyse genes which are not represented in microarrays.
199 The data generated here represent more than 361 million short nucleotide reads (100bp length) and
200 36.1 Gb of successfully aligned sequence (Table S1). This constitutes the deepest sequencing of a
201 plant circadian transcriptome to date.

202 Gene expression was computed with the ASpli R package (Mancini *et al.*, 2019) using the AtRTD2
203 annotation (Zhang *et al.*, 2017) by obtaining a table of gene counts, which represent the total number
204 of reads aligning to each gene (34,212 annotated AtRTD2 genes). On average, the reproducibility
205 between replicate samples (i.e., samples from the same time point) was excellent, with Pearson
206 correlation average R^2 values approaching 0.99 (see Supplemental Fig. 1 online). Using a
207 Multidimensional scaling (MDS) plot we also verified that samples corresponding to the same
208 timepoint on the two consecutive free running days (LL day 2 and LL day 3) were similar (Fig. S1).
209 We then obtained the number of reads per gene length (read density) for each gene, and only those
210 that had a value higher than 0.05 in any of the time points throughout the time course were chosen for
211 further analysis (18,503 genes).

212 In order to improve the power of subsequent rhythmic statistical tests, we applied a pre-screening
213 step to exclude any obviously non-rhythmic transcripts, as previously suggested (Keegan *et al.*, 2007).
214 Only the genes that showed a variation throughout the time-course (13,256 genes; negative binomial
215 generalized linear model (GLM) model fit, $p < 0.05$ and $q < 0.10$) were then subjected to circadian
216 rhythmicity analysis. To this end, several statistical approaches have been described (Wichert *et al.*,
217 2004; Wijnen *et al.*, 2005; Michael *et al.*, 2008; Deckard *et al.*, 2013). In this work, we applied the
218 Jonckheere-Terpstra-Kendall (JTK) algorithm, which has been reported to identify rhythmic genes
219 with enhanced confidence, improved computational efficiency and more resistance to outliers than
220 other algorithms (Hughes *et al.*, 2010). Using this method, a total of 9,127 genes were found to be
221 rhythmic and exhibited a circadian expression pattern (JTK_cycle, $p < 0.01$ and $q < 0.01$; Fig. 1b, Table
222 S3) under our assay conditions. To verify the validity of our approach, we examined the expression
223 pattern of known core clock genes in our data set and validated them by qPCR (Fig. 1d-k). As
224 expected, the qPCR reproduced the patterns observed on our RNA-seq data (Fig. 1d-k). As expected
225 for circadian datasets, peak amplitude on day 3 is slightly lower than peak amplitude on day 2,
226 indicating a progressive decrease of amplitude in constant conditions (Table S3).

227 We then performed a Gene Ontology (GO) analysis of this dataset (Fig. S2, Table S4) which showed
228 that clock controlled genes (CCGs) were involved in diverse biological processes. These included, but
229 were not limited to, immune response, hormone signalling, light signalling, metabolism and

230   photosynthesis, which is in accordance to previous reports (Covington *et al.*, 2008; Hsu & Harmer,

231   2012). Amplitude histogram analysis reveals that most transcripts oscillate with a relative amplitude

232   below 3, while a minority of circadian transcripts oscillate at high amplitude. This latter set includes

233   the core circadian clock genes (Fig. S3), and is consistent to what is known in mammals and fruit flies

234   (Hughes *et al.*, 2009; Hughes *et al.*, 2012).

235   A phase distribution analysis of the circadian regulated genes showed that that 25.9% (2364 genes)

236   exhibit a peak phase expression at LL10-12, while 24.66% (2251 genes) exhibit a peak phase

237   expression at LL22-LL0. These correspond to the day-to-night and night-to-day transitions of the LD

238   12 h : 12 h entraining photoperiod, respectively (Fig. 1b). Genes that peaked at LL10-12 of the free

239   running cycle were involved in 96 biological processes (Fisher's Exact Test, $p < 0.05$), including

240   response to stress, immune response, lipid metabolism, signalling and protein processes (Fig. S4a).

241   Genes that peaked at LL22-0 were involved in 146 biological processes (Fisher's Exact Test, $p <$

242   0.05), including RNA modification/processing, glucosinolate biosynthesis, carbohydrate metabolism,

243   mitochondrial/intracellular transport and ribosome biogenesis (Fig. S4b)

244   Upon comparison of our dataset with previous published global *A. thaliana* circadian gene expression

245   datasets (Covington *et al.*, 2008; Hsu & Harmer, 2012) (Table S5), we found an overlap of 41.67%

246   (3,803 genes) and 49.76% (4,542 genes) to the Covington + Edwards combined gene lists (Covington

247   *et al.*, 2008) (DS1) and Hsu (Hsu & Harmer, 2012) (DS2) datasets, respectively. A total of 2,763

248   genes (30.27%) were found to cycle in all three datasets (Fig. 2a). By combining the three datasets,

249   we find that over 37.29% of *A. thaliana* genes (12,579 / 34,212 AtRTD2 genes) have been described

250   to be circadian regulated to this date, in global transcriptomic studies (Table S5).

251   **Novel circadian transcripts**

252   Our dataset also revealed novel circadian regulated genes that had not been identified by the two

253   previous global studies (Fig. 2a). This was to be expected, to a certain degree, because of the

254   difference in genome coverage of the different technologies used (RNA-seq, ATH1 microarray and

255   AGRONOMICS1 tiling array). However, hypothetically there should be no difference in the possible

256   number of positive genes between our RNA-seq approach and the AGRONOMICS1 platform

257   (Rehrauer *et al.*, 2010) (Fig. S5). Nevertheless, we found 3545 circadian genes that are absent from

258   previous global studies. We will refer to these genes as novel circadian genes. GO analysis of this

259   subset revealed an enrichment in terms related to translation, ribosome biogenesis and RNA

260   metabolism (Fig. 2b, Table S6). It should be noted that some of the genes in this subset have been

261   previously identified as circadian regulated in non-global studies. Such is the case, for example, of the

262   widely used circadian reporter *CAB2* (Millar & Kay, 1991).

263   Interestingly, 146 transcription factors (TFs) spanning many families, including bZIP, bHLH, MADS

264   box related, WRKY, MYB and NAC, are part of the novel CCGs subset (Fig. 2c, Table S3). The RNA-

265   seq data of the expression pattern of these TFs genes can be seen in Fig. 2d. Further study of these

266   genes might lead to new insight into how the circadian clock regulates gene expression.

267   **The circadian clock regulates alternative splicing events**

268   In this work, we took advantage of the accuracy of RNA-seq technology and then applied a custom
269   pipeline designed to specifically detect and quantify splicing events. To do so, we aligned the RNA-
270   seq reads to the TAIR10 reference genome to generate the BAM files required as input for the *in*
271   *house* developed ASpli Bioconductor R package. ASpli was designed by our lab for the specific
272   purpose of detecting and quantifying splicing events (Mancini *et al.*, 2016; Mancini *et al.*, 2019). This
273   package has successfully been employed to study alternative splicing in several studies spanning
274   different organisms (De Maio *et al.*, 2016; Mancini *et al.*, 2016; Beckwith *et al.*, 2017; Xin *et al.*, 2017).
275   In summary, when used to detect changes in splicing events, this algorithm uses a known genome
276   annotation to categorise the genome in discrete exon and intron "bins", and then quantifies the counts
277   that each bin has over time and normalizes it to gene expression. After obtaining the normalized bin
278   counts, we proceeded to analyse the rhythmicity of splicing events as explained for gene expression
279   (Mancini *et al.*, 2016; Mancini *et al.*, 2019). Using this method, coupled to the AtRTD2 annotation
280   (Zhang *et al.*, 2017), 3976 clock controlled splicing events (CCEs) were discovered. In particular, 300
281   intron retention (IR), 40 exon skipping, 78 alternative 5' splice site (Alt5'ss), 113 alternative 3' splice
282   site (Alt3'ss), 470 multiple (involved simultaneously in more than one AS event type), 22 undefined,
283   and 1720 novel exonic and 1233 novel intronic events were found to be rhythmic, and exhibited a
284   circadian expression pattern (JTK_cycle, $p<0.01$ and $q<0.01$) with relative amplitude > 0.15 under our
285   assay conditions. With the purpose of increasing the stringency of our approach, we applied an extra
286   filtering step and kept only those splicing events with evidence of splice-junction reads. To do so, we
287   calculated the Percentage of Spliced In (PSI) and the Percentage of Intron Retention (PIR) indexes as
288   previously described (Gueroussov *et al.*, 2015) for exon and intron bins, respectively. We kept the
289   subset of circadian splicing events that showed more than a 5% variation in PSI/PIR (dPSI/PIR)
290   across all the analysed timepoints. This resulted in a total of 1513 events corresponding to 1169
291   genes, split into the following categories: 79 Alt3'ss, 48 Alt5'ss, 300 IR, 8 ES, 350 multiple (Fig. 3a),
292   and 7 novel exonic and 721 novel intronic (Fig. 3b) (Table S7). It is interesting to note that this study
293   also revealed that 617 CCGs are also clock regulated at the level of alternative use of exons and/or
294   introns throughout the day (Fig. S6a, Table S7). On the other hand, there also are 552 genes whose
295   expression is not circadian regulated, but their exon/intron use is (Fig. S6a, Table S7), and 8510
296   CCGs that had no detectable CCEs. In total, 652 genes had novel exonic/intronic events, 69 Alt3'ss,
297   48 Alt5'ss, 286 IR, 8 ES and 292 multiple (Fig. S6b), including several splicing factors and
298   spliceosome components.

299   We then performed a Gene Ontology (GO) analysis of this dataset (Table S8) which showed that
300   clock controlled splicing events (CCEs) included genes involved in primary metabolism (GO:0044283),
301   nitrogen metabolism (GO:0006807 and GO:0034641), nucleic acid metabolism (GO:0006139 and
302   GO:0090304), RNA/mRNA processing (GO:0016070, GO:0006396, GO:0006397, GO:0016071), lipid
303   metabolism (GO:0006644, GO:0009245, GO:0046493 and GO:0008654), flowering (GO:0009909
304   and GO:0009910), and RNA splicing (GO:0000375, GO:0000377, GO:0008380).

305   In order to verify the validity of our approach, we performed an independent experiment and
306   examined the expression pattern of a couple of splicing events from our data set and validated them
307   by semi-quantitative (sq)RT-PCR, followed by SDS-PAGE. Results were then further corroborated by
308   qPCR, obtaining in both cases results that matched the patterns produced by our RNA-seq analysis
309   pipeline data (Fig. 3e-h, Fig. S7a-b). We validated the intron retention splicing event of the COL2
310   gene (IR - AT3G02380:E002 – width 143bp, Table S7), homolog of the photoperiod and flowering
311   regulator CONSTANS. This event had been previously described (Hazen *et al.*, 2009) and was
312   therefore a perfect positive control. According to the bioinformatics analysis of our pipeline, the IR
313   precedes the subjective day and has a peak at the day/night transition, as can be seen in the RNA-
314   seq coverage plots (Fig. S7a) and PIR pattern (Fig. 3e). This event also fulfils our stringent criteria of
315   having reads that span the intron-exon. The analysis by sqRT-PCR followed by SDS-PAGE (Fig. S7b)
316   and qPCR (Fig. 3f) demonstrate the validity of the bioinformatics analysis used in this work.
317   Additionally, our results exhibit a peak IR at CT36, which is in complete accordance to what was
318   previously described (Hazen *et al.*, 2009), and further validates our analysis pipeline. Additionally, we
319   also validated an IR event on the A/N-InvC gene (IR - AT3G06500:E015 – width 81bp, Fig. 3g-h).
320   This gene is expressed in roots, aerial parts (shoots and leaves) and flowers, and encodes an
321   alkaline/neutral invertase which localizes in mitochondria. Mutations of this gene have been shown to
322   have reduced shoot growth (Martin *et al.*, 2013). Retention of this intron causes inclusion of an in-
323   frame Opal stop codon (at position 542), leading to a truncated protein with 123 less amino acid
324   residues.

325   **Several splicing modulators are under circadian clock control**

326   We reasoned that this circadian control of alternative splicing could be due, at least in part, to an
327   underlying circadian regulation of the splicing machinery. To test this hypothesis, we compiled a list of
328   all known splicing modulators, as previously described (Perez-Santangelo *et al.*, 2013). This list
329   contains includes SR, hnRNPs, core spliceosomal proteins, regulators of spliceosome assembly,
330   components of specific snRNPs and some kinases that regulate SR protein function. We then
331   intersected it with the Covington + Edwards dataset (DS1), the Hsu & Harmer dataset (DS2) and our
332   own. Out of the 426 splicing related genes, 31 were rhythmic in all three datasets (Fig. 4a). Because
333   different research groups, each with their own environmental conditions, obtained these datasets
334   independently, we reasoned that these 31 splicing modulators would be the most robustly rhythmic
335   ones.

336   One of these genes, AT2G02570, has been previously described as the homolog of the mammalian
337   Survival of Motor Neuron (SMN), which controls spliceosome assembly (Kroiss *et al.*, 2008). However,
338   our reciprocal BlastP search revealed that AT2G02570 is more closely related to Splicing Factor 30
339   (SPF30, e.g. NP_005862.1) (Fig. S8). SPF30 is an SMN paralogue required for spliceosome
340   assembly, which as SMN also contains a Tudor domain (Cote & Richard, 2005; Liu *et al.*, 2012; Mier
341   & Perez-Pulido, 2012). This domain has affinity to Sm ribonucleoproteins that form a ring involved in
342   the splicing process (Cote & Richard, 2005). Based on increasing evidence indicating that alternative
343   splicing is regulated not only by SR and hnRNP auxiliary factors, but also by changes in the levels or

344   activities of core spliceosomal components or proteins that modulate the kinetics of spliceosome

345   assembly (Saltzman *et al.*, 2011; Papasaikas *et al.*, 2015), we decided to further investigate the

346   biological relevance of AT2G02570, which will be called AtSPF30 throughout the text.

347   **AtSPF30 affects a subset of clock-regulated events**

348   In order to investigate the role of AtSPF30 on the modulation of CCEs, we first validated its rhythmic

349   transcriptional pattern by qPCR (Fig. 4b-d) and isolated two different T-DNA insertion mutants

350   (*atspf30-1* and *atspf30-2*, Fig. 5a). Plants homozygous for either mutant allele were morphologically

351   very similar to the wild type (Fig. 5a). Characterization of two clock regulated outputs, leaf movement

352   and flowering, revealed no significant difference in leaf movement periodicity between Col-0 and

353   either mutant (Fig. 5b) and a consistent slightly early flowering phenotype across the different

354   photoperiods tested (LL, LD and SD, *p<0.05 ANOVA followed by Dunnett's *post hoc* test against Col-

355   0; Fig. 5c).

356   RNA-seq analysis of *atspf30-1* mutant plants grown in constant light revealed 286 altered alternative

357   splicing events when compared to Col-0 plants grown in the same conditions (|FC| > 1.5, p<0.05,

358   q<0.05 and dPSI/PIR > 5%; Fig. 6a, Table S9). In wild-type plants, the most abundant AS events

359   were those associated with the use of alternative 3′ splice sites (Alt3'ss; 33%), followed by intron

360   retention (IR; 32%), alternative 5′ splice sites (Alt5'ss; 22%), and exon skipping (ES; 13%). However,

361   among the AS events altered in the *atspf30-1* mutant, we found an increase in the proportion of IR

362   events, a decrease in 3'alt events and 5'alt events relative to their frequency in wild type plants (Fig.

363   6a-b). Interestingly, 32 of these events were also among our PSI/PIR filtered subset of CCEs (Fig. 6c,

364   Table S10). Upon closer inspection of the dPSI/PIR of *spf30-1* vs Col-0 of the 32 events in common

365   between both datasets (Table S10), we hypothesised that intron retention above the level of wild type

366   would also be expected on *atspf30-1* along an entire circadian period. With this in mind, we selected

367   an IR event that was both, affected by a mutation in AtSPF30 and clock controlled, with a peak phase

368   at CT16, for validation (Fig. 6d-f). qPCR quantification of a new set of circadian timecourse samples

369   showed that lack of functional AtSPF30 affected the overall inclusion of the of AT5G17670:E006

370   along the entire circadian day without affecting the rhythmic pattern (Fig. 6f). Hence, in the case of

371   this splicing event, AtSPF30 is modulating both the baseline level of intron retention and its overall

372   amplitude.

373   **DISCUSSION**

374   AS is highly prevalent in plants, affecting pre-mRNAs of more than 60% of intron-containing genes

375   (Reddy *et al.*, 2013). Many genes undergo AS in response to development and changes in the

376   environment, such as in light and temperature (James *et al.*, 2012a; James *et al.*, 2012b; Reddy *et al.*,

377   2013; Kwon *et al.*, 2014). For some time now, evidence has been accumulating pointing to the fact

378   that AS can affect clock components and that mutations affecting spliceosome related components

379   can alter the circadian period (Sanchez *et al.*, 2010; Jones *et al.*, 2012; Perez-Santangelo *et al.*, 2013;

380   Cui *et al.*, 2014; Filichkin *et al.*, 2015). Furthermore, studies performed in fruit flies, mice and human

381 cell lines have identified a role for the circadian clock in the control of the splicing process (Hughes *et*
382 *al.*, 2012; Preussner *et al.*, 2017; Genov *et al.*, 2019). However, there have been no global
383 transcriptomic studies that allowed for understanding the extent of clock control on AS on any plant
384 species. There is perhaps one interesting exception: a global study limited to the analysis of cycling
385 introns using GeneChip AtTILE1 tiling array technology (Hazen *et al.*, 2009). A total of 499 genes
386 were reported to contain cycling introns at the time. In this study, we used NGS Illumina mRNA-seq
387 technology, which allowed us to avoid the technical issues inherent to microarray probe performance
388 such as cross-hybridization, non-specific hybridization and limited detection range of individual probes
389 (Agarwal *et al.*, 2010). By combining the higher resolution of mRNA-seq with an *in house* developed
390 bioinformatics pipeline, we were able to find 300 intron retention, 79 alternative 3' splice site usage ,
391 48 Alt 5' splice site usage, 8 ES and 350 multiple (complex) events (Fig. 3a) that are under circadian
392 regulation. We also found 7 and 721 novel alternative exonic and intronic events, respectively.
393 Interestingly, of the 646 genes that were found to contain cycling introns (Table S7), only 35
394 overlapped with the 499 reported by Hazen et al (7%, 35 out of 499).

395 In addition to documenting CCEs, we found that a clock-controlled splicing factor, AT2G02570,
396 affected a subset of AS events when mutated. This gene had been previously identified as an SMN
397 homolog (Kroiss *et al.*, 2008). SMN encodes the survival motor neuron protein, a spliceosome
398 complex constituent which, when mutated, causes autosomal recessive spinal muscular atrophy
399 (Talbot *et al.*, 1998; Kroiss *et al.*, 2008). However, our bioinformatics analysis revealed that
400 AT2G02570 is more closely related to SMN's paralogue SPF30 (Talbot *et al.*, 1998) (Fig. S8). Using
401 the ASpli package, we identified 286 altered AS events in the AtSPF30 mutant *atspf30-1*. This is in
402 accordance with previous reports that showed that changes in the levels or activities of core
403 spliceosomal components or proteins that modulate the kinetics of spliceosome assembly can
404 modulate alternative splicing (Saltzman *et al.*, 2011; Papasaikas *et al.*, 2015).

405 A comparison between the CCEs datasets and the *atspf30-1* dataset revealed that 32 events of the
406 286 events affected on an AtSPF30 mutant were among our PSI/PIR filtered subset of CCEs.
407 Furthermore, a closer inspection of these splicing events showed that it preferentially affected IR
408 events, by increasing overall retention level (Table S10). Indeed, we observed this effect upon
409 validation of one of the shared events. qPCR analysis showed that, albeit with a similar pattern to
410 what was observed for Col-0, IR of AT5G17670:E015 was higher at all assayed timepoints (Fig. 6f).
411 This led us to believe that circadian control of AS is multifactorial because even though SPF30 is
412 contributing by modulating the amplitude and overall level (i.e. higher IR), the underlying oscillation
413 continues when SPF30 is absent.

414 In addition to AtSPF30, 30 other splicing related proteins were found to be under circadian regulation
415 in all known circadian datasets (Fig. 4a). The contribution of each of these robust cycling factors on
416 circadian control of AS remains to be determined and should be studied further. Prospectively,
417 modulation of these circadian controlled splicing related factors could lead to time specific fine tuning
418 of splicing events. Although many non-redundant splicing factors affect the clock when mutated
419 (Sanchez *et al.*, 2010; Staiger & Green, 2011; Jones *et al.*, 2012; Wang *et al.*, 2012; Romanowski &

420 Yanovsky, 2015; Schlaen *et al.*, 2015; Hernando *et al.*, 2017; Mateos *et al.*, 2018), we found no
421 evidence that circadian modulation of splicing factors and related proteins led to AS changes in core
422 clock components, but rather their effect was as an output (CCGs) mainly affecting other outputs
423 (CCEs).

424 Splicing factors were also subject to circadian splicing events. In this sense, our stringent analysis
425 revealed 57 splicing related genes that are also regulated at the splicing level (Table S7). A future
426 careful and detailed study of the biological relevance of these splicing events will help us gain a better
427 understanding of how the clock regulates alternative splicing, and vice versa.

428 It has been previously shown that core clock components are highly sensitive and can undergo AS
429 upon environmental perturbations (Filichkin *et al.*, 2010; James *et al.*, 2012a; James *et al.*, 2012b;
430 Syed *et al.*, 2012; Perez-Santangelo *et al.*, 2013; Staiger & Brown, 2013; Filichkin *et al.*, 2015; Mateos
431 *et al.*, 2018). Furthermore, environmental disruptions appear to modulate AS of core clock
432 components more than the clock itself. Taken together, this suggests that AS modulation of the
433 central clock operates mainly via its input pathways (after an environmental perturbation), and
434 circadian control of AS is mostly reserved to modulation of output pathways. However, we must point
435 out that the number of CCEs reported in this work is very conservative and will likely increase in the
436 future, as sequencing depth of circadian studies increases. The advent of novel long read
437 technologies like those provided by Pacific Biosystems or Oxford Nanopore will likely allow for reliable
438 transcript isoform level quantification at the global scale in the very near future.

439

440 **ACKNOWLEDGEMENTS**

448

449 **AUTHOR CONTRIBUTIONS**

450 AR, RGS, SPS and MY planned and designed the research; AR, RGS, SPS performed the
451 experiments; AR, RGS and SPS analysed the experimental data; AR and EM performed the
452 bioinformatics analysis, and AR and MY wrote the manuscript.

453

454 **REFERENCES**

455   Agarwal A, Koppstein D, Rozowsky J, Sboner A, Habegger L, Hillier LW, Sasidharan R, Reinke V,
456   Waterston RH, Gerstein M. 2010. Comparison and calibration of transcriptome data from
457   RNA-Seq and tiling arrays. *BMC Genomics* **11**: 383.
458   Beckwith EJ, Hernando CE, Polcownuk S, Bertolin AP, Mancini E, Ceriani MF, Yanovsky MJ. 2017.
459   Rhythmic Behavior Is Controlled by the SRm160 Splicing Factor in Drosophila melanogaster.
460   *Genetics* **207**(2): 593-607.
461   Cote J, Richard S. 2005. Tudor domains bind symmetrical dimethylated arginines. *J Biol Chem*
462   **280**(31): 28476-28483.
463   Covington MF, Maloof JN, Straume M, Kay SA, Harmer SL. 2008. Global transcriptome analysis
464   reveals circadian regulation of key pathways in plant growth and development. *Genome Biol*
465   **9**(8): R130.
466   Cui Z, Xu Q, Wang X. 2014. Regulation of the circadian clock through pre-mRNA splicing in
467   Arabidopsis. *J Exp Bot* **65**(8): 1973-1980.
468   De Maio FA, Risso G, Iglesias NG, Shah P, Pozzi B, Gebhard LG, Mammi P, Mancini E, Yanovsky MJ,
469   Andino R, et al. 2016. The Dengue Virus NS5 Protein Intrudes in the Cellular Spliceosome
470   and Modulates Splicing. *PLoS Pathog* **12**(8): e1005841.
471   Deckard A, Anafi RC, Hogenesch JB, Haase SB, Harer J. 2013. Design and analysis of large-scale
472   biological rhythm studies: a comparison of algorithms for detecting periodic signals in
473   biological data. *Bioinformatics* **29**(24): 3174-3180.
474   Dodd AN, Salathia N, Hall A, Kevei E, Toth R, Nagy F, Hibberd JM, Millar AJ, Webb AA. 2005. Plant
475   circadian clocks increase photosynthesis, growth, survival, and competitive advantage.
476   *Science* **309**(5734): 630-633.
477   Dunlap JC, Loros JJ, DeCoursey PJ. 2004. *Chronobiology : biological timekeeping*. Sunderland, Mass.:
478   Sinauer Associates.
479   Fernandez-Pozo N, Haas FB, Meyberg R, Ullrich KK, Hiss M, Perroud PF, Hanke S, Kratz V, Powell AF,
480   Vesty EF, et al. 2019. PEATmoss (Physcomitrella Expression Atlas Tool): a unified gene
481   expression atlas for the model plant Physcomitrella patens. *Plant J.*
482   Filichkin SA, Cumbie JS, Dharmawardhana P, Jaiswal P, Chang JH, Palusa SG, Reddy AS, Megraw M,
483   Mockler TC. 2015. Environmental Stresses Modulate Abundance and Timing of Alternatively
484   Spliced Circadian Transcripts in Arabidopsis. *Mol Plant* **8**(2): 207-227.
485   Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, Fox SE, Wong WK, Mockler TC. 2010.
486   Genome-wide mapping of alternative splicing in Arabidopsis thaliana. *Genome Res* **20**(1): 45-
487   58.
488   Freese NH, Norris DC, Loraine AE. 2016. Integrated genome browser: visual analytics platform for
489   genomics. *Bioinformatics* **32**(14): 2089-2095.
490   Genov N, Basti A, Abreu M, Astaburuaga R, Relogio A. 2019. A bioinformatic analysis identifies
491   circadian expression of splicing factors and time-dependent alternative splicing events in the
492   HD-MY-Z cell line. *Sci Rep* **9**(1): 11062.
493   Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: A multiplatform graphical user interface
494   for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**(2): 221-224.
495   Green RM, Tingay S, Wang ZY, Tobin EM. 2002. Circadian rhythms confer a higher level of fitness to
496   Arabidopsis plants. *Plant Physiol* **129**(2): 576-584.
497   Gueroussov S, Gonatopoulos-Pournatzis T, Irimia M, Raj B, Lin ZY, Gingras AC, Blencowe BJ. 2015.
498   An alternative splicing event amplifies evolutionary differences between vertebrates.
499   *Science* **349**(6250): 868-873.
500   Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and
501   methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML
502   3.0. *Syst Biol* **59**(3): 307-321.
503   Harmer SL, Panda S, Kay SA. 2001. Molecular bases of circadian rhythms. *Annu Rev Cell Dev Biol* **17**:
504   215-253.

505 **Hazen SP, Naef F, Quisel T, Gendron JM, Chen H, Ecker JR, Borevitz JO, Kay SA. 2009.** Exploring the
506   transcriptional landscape of plant circadian rhythms using genome tiling arrays. *Genome Biol*
507   **10**(2): R17.
508 **Hernando CE, Romanowski A, Yanovsky MJ. 2017.** Transcriptional and post-transcriptional control
509   of the plant circadian gene regulatory network. *Biochim Biophys Acta Gene Regul Mech*
510   **1860**(1): 84-94.
511 **Hong X, Scofield DG, Lynch M. 2006.** Intron size, abundance, and distribution within untranslated
512   regions of genes. *Mol Biol Evol* **23**(12): 2392-2404.
513 **Hsu PY, Harmer SL. 2012.** Circadian phase has profound effects on differential expression analysis.
514   *PLoS One* **7**(11): e49853.
515 **Hughes ME, DiTacchio L, Hayes KR, Vollmers C, Pulivarthy S, Baggs JE, Panda S, Hogenesch JB. 2009.**
516   Harmonics of circadian gene transcription in mammals. *PLoS Genet* **5**(4): e1000442.
517 **Hughes ME, Grant GR, Paquin C, Qian J, Nitabach MN. 2012.** Deep sequencing the circadian and
518   diurnal transcriptome of Drosophila brain. *Genome Res* **22**(7): 1266-1281.
519 **Hughes ME, Hogenesch JB, Kornacker K. 2010.** JTK_CYCLE: an efficient nonparametric algorithm for
520   detecting rhythmic components in genome-scale data sets. *J Biol Rhythms* **25**(5): 372-380.
521 **James AB, Syed NH, Bordage S, Marshall J, Nimmo GA, Jenkins GI, Herzyk P, Brown JW, Nimmo HG.**
522   **2012a.** Alternative splicing mediates responses of the Arabidopsis circadian clock to
523   temperature changes. *Plant Cell* **24**(3): 961-981.
524 **James AB, Syed NH, Brown JW, Nimmo HG. 2012b.** Thermoplasticity in the plant circadian clock:
525   how plants tell the time-perature. *Plant Signal Behav* **7**(10): 1219-1223.
526 **Jones MA, Williams BA, McNicol J, Simpson CG, Brown JW, Harmer SL. 2012.** Mutation of
527   Arabidopsis spliceosomal timekeeper locus1 causes circadian clock defects. *Plant Cell* **24**(10):
528   4066-4082.
529 **Katari MS, Nowicki SD, Aceituno FF, Nero D, Kelfer J, Thompson LP, Cabello JM, Davidson RS,**
530   **Goldberg AP, Shasha DE, et al. 2010.** VirtualPlant: a software platform to support systems
531   biology research. *Plant Physiol* **152**(2): 500-515.
532 **Keegan KP, Pradhan S, Wang JP, Allada R. 2007.** Meta-analysis of Drosophila circadian microarray
533   studies identifies a novel set of rhythmically expressed genes. *PLoS Comput Biol* **3**(11): e208.
534 **Kolesnikov N, Hastings E, Keays M, Melnichuk O, Tang YA, Williams E, Dylag M, Kurbatova N,**
535   **Brandizi M, Burdett T, et al. 2015.** ArrayExpress update--simplifying data submissions.
536   *Nucleic Acids Res* **43**(Database issue): D1113-1116.
537 **Kornblihtt AR, Schor IE, Allo M, Dujardin G, Petrillo E, Munoz MJ. 2013.** Alternative splicing: a
538   pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol* **14**(3):
539   153-165.
540 **Kroiss M, Schultz J, Wiesner J, Chari A, Sickmann A, Fischer U. 2008.** Evolution of an RNP assembly
541   system: a minimal SMN complex facilitates formation of UsnRNPs in Drosophila
542   melanogaster. *Proc Natl Acad Sci U S A* **105**(29): 10045-10050.
543 **Kwon YJ, Park MJ, Kim SG, Baldwin IT, Park CM. 2014.** Alternative splicing and nonsense-mediated
544   decay of circadian clock genes under environmental stress conditions in Arabidopsis. *BMC*
545   *Plant Biol* **14**: 136.
546 **Le SQ, Gascuel O. 2008.** An improved general amino acid replacement matrix. *Mol Biol Evol* **25**(7):
547   1307-1320.
548 **Liu K, Guo Y, Liu H, Bian C, Lam R, Liu Y, Mackenzie F, Rojas LA, Reinberg D, Bedford MT, et al. 2012.**
549   Crystal structure of TDRD3 and methyl-arginine binding characterization of TDRD3, SMN and
550   SPF30. *PLoS One* **7**(2): e30375.
551 **Mancini E, Iserte J, Yanovsky M, Chernomoretz A. 2019.** ASpli: Analysis of alternative splicing using
552   RNA-Seq. R package version 1.8.1. *Bioconductor* **Release (3.8)**.
553 **Mancini E, Sanchez SE, Romanowski A, Schlaen RG, Sanchez-Lamas M, Cerdan PD, Yanovsky MJ.**
554   **2016.** Acute Effects of Light on Alternative Splicing in Light-Grown Plants. *Photochem*
555   *Photobiol* **92**(1): 126-133.

556 **Martin ML, Lechner L, Zabaleta EJ, Salerno GL. 2013.** A mitochondrial alkaline/neutral invertase
557   isoform (A/N-InvC) functions in developmental energy-demanding processes in Arabidopsis.
558   *Planta* **237**(3): 813-822.
559 **Mateos JL, de Leone MJ, Torchio J, Reichel M, Staiger D. 2018.** Beyond Transcription: Fine-Tuning of
560   Circadian Timekeeping by Post-Transcriptional Regulation. *Genes (Basel)* **9**(12).
561 **McClung CR. 2006.** Plant circadian rhythms. *Plant Cell* **18**(4): 792-803.
562 **McClung CR. 2014.** Wheels within wheels: new transcriptional feedback loops in the Arabidopsis
563   circadian clock. *F1000Prime Rep* **6**: 2.
564 **Michael TP, Mockler TC, Breton G, McEntee C, Byer A, Trout JD, Hazen SP, Shen R, Priest HD,**
565   **Sullivan CM, et al. 2008.** Network discovery pipeline elucidates conserved time-of-day-
566   specific cis-regulatory modules. *PLoS Genet* **4**(2): e14.
567 **Mier P, Perez-Pulido AJ. 2012.** Fungal Smn and Spf30 homologues are mainly present in filamentous
568   fungi and genomes with many introns: implications for spinal muscular atrophy. *Gene* **491**(2):
569   135-141.
570 **Millar AJ. 2016.** The Intracellular Dynamics of Circadian Clocks Reach for the Light of Ecology and
571   Evolution. *Annu Rev Plant Biol* **67**: 595-618.
572 **Millar AJ, Kay SA. 1991.** Circadian Control of cab Gene Transcription and mRNA Accumulation in
573   Arabidopsis. *Plant Cell* **3**(5): 541-550.
574 **Mockler TC, Michael TP, Priest HD, Shen R, Sullivan CM, Givan SA, McEntee C, Kay SA, Chory J.**
575   **2007.** The DIURNAL project: DIURNAL and circadian expression profiling, model-based
576   pattern matching, and promoter analysis. *Cold Spring Harb Symp Quant Biol* **72**: 353-363.
577 **Papasaikas P, Tejedor JR, Vigevani L, Valcarcel J. 2015.** Functional splicing network reveals extensive
578   regulatory potential of the core spliceosomal machinery. *Mol Cell* **57**(1): 7-22.
579 **Perez-Santangelo S, Schlaen RG, Yanovsky MJ. 2013.** Genomic analysis reveals novel connections
580   between alternative splicing and circadian regulatory networks. *Brief Funct Genomics* **12**(1):
581   13-24.
582 **Pervouchine DD, Knowles DG, Guigo R. 2013.** Intron-centric estimation of alternative splicing from
583   RNA-seq data. *Bioinformatics* **29**(2): 273-274.
584 **Plautz JD, Straume M, Stanewsky R, Jamison CF, Brandes C, Dowse HB, Hall JC, Kay SA. 1997.**
585   Quantitative analysis of Drosophila period gene transcription in living animals. *J Biol Rhythms*
586   **12**(3): 204-217.
587 **Preussner M, Goldammer G, Neumann A, Haltenhof T, Rautenstrauch P, Muller-McNicoll M, Heyd**
588   **F. 2017.** Body Temperature Cycles Control Rhythmic Alternative Splicing in Mammals. *Mol*
589   *Cell* **67**(3): 433-446 e434.
590 **Reddy AS, Marquez Y, Kalyna M, Barta A. 2013.** Complexity of the alternative splicing landscape in
591   plants. *Plant Cell* **25**(10): 3657-3683.
592 **Rehrauer H, Aquino C, Gruissem W, Henz SR, Hilson P, Laubinger S, Naouar N, Patrignani A,**
593   **Rombauts S, Shu H, et al. 2010.** AGRONOMICS1: a new resource for Arabidopsis
594   transcriptome profiling. *Plant Physiol* **152**(2): 487-499.
595 **Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015.** limma powers differential
596   expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**(7): e47.
597 **Romanowski A, Yanovsky MJ. 2015.** Circadian rhythms and post-transcriptional regulation in higher
598   plants. *Front Plant Sci* **6**: 437.
599 **Saldi T, Cortazar MA, Sheridan RM, Bentley DL. 2016.** Coupling of RNA Polymerase II Transcription
600   Elongation with Pre-mRNA Splicing. *J Mol Biol* **428**(12): 2623-2635.
601 **Saltzman AL, Pan Q, Blencowe BJ. 2011.** Regulation of alternative splicing by the core spliceosomal
602   machinery. *Genes Dev* **25**(4): 373-384.
603 **Sanchez SE, Petrillo E, Beckwith EJ, Zhang X, Rugnone ML, Hernando CE, Cuevas JC, Godoy Herz MA,**
604   **Depetris-Chauvin A, Simpson CG, et al. 2010.** A methyl transferase links the circadian clock
605   to the regulation of alternative splicing. *Nature* **468**(7320): 112-116.

606 Schlaen RG, Mancini E, Sanchez SE, Perez-Santangelo S, Rugnone ML, Simpson CG, Brown JW,
607   Zhang X, Chernomoretz A, Yanovsky MJ. 2015. The spliceosome assembly factor GEMIN2
608   attenuates the effects of temperature on alternative splicing and circadian rhythms. *Proc*
609   *Natl Acad Sci U S A* **112**(30): 9382-9387.
610 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding
611   J, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments
612   using Clustal Omega. *Mol Syst Biol* **7**: 539.
613 Staiger D, Brown JW. 2013. Alternative splicing at the intersection of biological timing, development,
614   and stress responses. *Plant Cell* **25**(10): 3640-3656.
615 Staiger D, Green R. 2011. RNA-based regulation in the plant circadian clock. *Trends Plant Sci* **16**(10):
616   517-523.
617 Syed NH, Kalyna M, Marquez Y, Barta A, Brown JW. 2012. Alternative splicing in plants--coming of
618   age. *Trends Plant Sci* **17**(10): 616-623.
619 Talbot K, Miguel-Aliaga I, Mohaghegh P, Ponting CP, Davies KE. 1998. Characterization of a gene
620   encoding survival motor neuron (SMN)-related protein, a constituent of the spliceosome
621   complex. *Hum Mol Genet* **7**(13): 2149-2156.
622 Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq.
623   *Bioinformatics* **25**(9): 1105-1111.
624 Wang X, Wu F, Xie Q, Wang H, Wang Y, Yue Y, Gahura O, Ma S, Liu L, Cao Y, et al. 2012. SKIP is a
625   component of the spliceosome linking alternative splicing and the circadian clock in
626   Arabidopsis. *Plant Cell* **24**(8): 3278-3295.
627 Wichert S, Fokianos K, Strimmer K. 2004. Identifying periodically expressed transcripts in microarray
628   time series data. *Bioinformatics* **20**(1): 5-20.
629 Wijnen H, Naef F, Young MW. 2005. Molecular and statistical tools for circadian transcript profiling.
630   *Methods Enzymol* **393**: 341-365.
631 Xin R, Zhu L, Salome PA, Mancini E, Marshall CM, Harmon FG, Yanovsky MJ, Weigel D, Huq E. 2017.
632   SPF45-related splicing factor for phytochrome signaling promotes photomorphogenesis by
633   regulating pre-mRNA splicing in Arabidopsis. *Proc Natl Acad Sci U S A* **114**(33): E7018-E7027.
634 Zhang R, Calixto CPG, Marquez Y, Venhuizen P, Tzioutziou NA, Guo W, Spensley M, Entizne JC,
635   Lewandowska D, Ten Have S, et al. 2017. A high quality Arabidopsis transcriptome for
636   accurate transcript-level analysis of alternative splicing. *Nucleic Acids Res* **45**(9): 5061-5073.

637

638 **FIG. LEGENDS**

639 **Fig. 1. Global circadian gene expression.**

640 (a) Schematic of experimental conditions. (b) Phase sorted heatmap of global scaled logCPM gene
641 expression from the circadian timecourse experiment. (c) Phase histogram of global gene expression.
642 (d-k) RT-qPCR validation (black, solid line) and RNA-seq traces (grey, dotted line) of the core clock
643 genes (d) CCA1, (e) LHY, (f) RVE8, (g) PRR9, (h) PRR7, (i) PRR5, (j) TOC1, and (k) ELF3.
644 Expression values are normalised to maximum. Error bars represent SEM. CPM: counts per million.
645 CCA1: Clock Control Associated 1 (AT2G46830). LHY: Late Elongated Hypocotyl (AT1G01060).
646 RVE8: Reveille 8 (AT3G09600). PRR9: Pseudo Response Regulator 9 (AT2G46790). PRR7: Pseudo
647 Response Regulator 7 (AT5G02810). PRR5: Pseudo Response Regulator 5 (AT5G24470). TOC1:
648 Timing of CAB Expression 1 (AT5G61380). ELF3: Early Flowering 3 (AT2G25930).

649 **Fig. 2. Novel circadian genes and specific gene families.**

650   (a) Venn diagram showing RNA-seq circadian gene expression (RNA-seq LL) compared to the

651   Covington + Edwards combined dataset (DS1) (Covington *et al.*, 2008) and the Hsu & Harmer dataset

652   (DS2) (Hsu & Harmer, 2012). (b) Selected subset of significant Biological Process (BP) GO Terms of

653   novel circadian genes. Black line indicates enrichment above expected frequency. (c-d) Novel cycling

654   transcription factor genes. (c) Venn diagram showing the comparison between the novel genes of the

655   circadian dataset and all known transcription factor genes. (d) Phase sorted heatmap of scaled

656   logCPM gene expression of the 146 novel cycling transcription factor genes. DS1: Dataset 1. DS2:

657   Dataset 2. CPM: Counts per million.

658   **Fig. 3. Analysis of Clock Controlled Elements (CCEs)**.

659   (a) Distribution of clock controlled annotated events by type (alt 3'ss, alt 5'ss, ES, IR and multiple). (b)

660   Distribution of novel (non-annotated) clock controlled exonic and intronic events. (c-d) Phase sorted

661   heatmap of scaled logCPM clock controlled splicing event bins expression by type: (c) annotated

662   events (alt 3'ss, alt 5'ss, ES, IR and multiple), (d) non-annotated events (exonic and intronic). (e-f)

663   Validation of CCEs. (e) PIR normalised to maximum of COL2 E2. (f) qPCR validation of the COL2 E2

664   IR event (** $p<0.01$; Student's T-test). (g) PIR normalised to maximum of the A/N-INvC E15. (h) qPCR

665   validation of the A/N-INvC E15 IR event (** $p<0.01$; Student's T-test). Error bars represent SEM. alt

666   3'ss: alternative 3' splice site. alt 5'ss: alternative 5' splice site. ES: exon skipping. IR: intron retention.

667   CPM: counts per million. CCEs: clock controlled splicing events. PIR: Percent Intron Retention. COL2:

668   CONSTANS like 2 (AT3G02380). A/N-INvC: Alkaline/Neutral Invertase C (AT3G06500).

669   **Fig. 4. Circadian regulated splicing factors and AtSPF30**.

670   (a) Venn diagram showing the comparison between robust cycling genes (those that cycle on the

671   RNA-seq dataset of this work and also on DS1 and DS2) and splicing related genes. (b-d) AtSPF30

672   expression pattern from: (b) Mockler Diurnal web server, LL23_LDHH dataset (Mockler *et al.*, 2007),

673   (c) Normalised counts from the circadian RNA-seq (this work), and (d) qPCR data (n=3 biological

674   replicates per timepoint). Expression values are normalised to maximum. Error bars represent SEM.

675   AtSPF30: Arabidopsis thaliana Splicing Factor 30 (AT2G02570).

676   **Fig. 5. AtSPF30 alleles, circadian leaf movement and flowering phenotype**.

677   (a) Gene model of AtSPF30. White triangles indicate the T-DNA insertion sites for the mutant alleles

678   *atspf30-1* (SALK_052016C) and *atspf30-2* (SALK_081292C). Red and green arrows indicate primer

679   binding sites. Representative images of wild type (Col-0), atspf30-1 and atspf30-2 (from left to right)

680   are shown. (b) Graph showing circadian leaf angle plots for Col-0 (black) and atspf30-1 (blue), after

681   transfer to LL for 144 h (6 days). These experiments were repeated at least three times with similar

682   results. No differences in period were found among genotypes, as estimated by Fast Fourier

683   Transform-Non-Linear Least Squares (FFT-NLLS) algorithm with BRASS 3.0. software. (c) Flowering

684   time measured as the number of rosette leaves at 1 cm bolting in LL, SD and LD photoperiods (n = 3

685   independent experiments for LL and LD, and n = 4 independent experiments for SD). Error bars

686   represent SEM. * $p<0.05$ (ANOVA, followed by Dunnet's *post hoc* test against Col-0). AtSPF30:

687 Arabidopsis thaliana Splicing Factor 30 (AT2G02570). LL: continuous light. SD: short days. LD: long
688 days.

689 **Fig. 6. CCEs misregulated in atspf30-1**.

690 (a-b) Distribution of events affected in atspf30-1 plants. (a) Annotated events by type (alt 3'ss, alt 5'ss,
691 ES and IR) (b) Non-annotated events (exonic and intronic). (c) Venn diagram showing the comparison
692 between all clock-controlled elements (CCEs) and atspf30-1 affected splicing events. (d) Read
693 density histogram of RNA-seq data comparing Col-0 (blue) and atspf30-1 (black) reads for the
694 AT5G17670 gene locus region. The red rectangle area depicts the read density zone for the IR event.
695 The AT5G17670 gene model is shown at the bottom of the graph. (e) PIR plot of the
696 AT5G17670:E006 event. (f) qPCR of AT5G17670:E006 in Col-0 (black) and atspf30-1 (blue).
697 Expression values are normalised to maximum. Error bars represent SEM. alt 3'ss: alternative 3'
698 splice site. alt 5'ss: alternative 5' splice site. ES: exon skipping. IR: intron retention. PIR: percent
699 intron retention.

700

701 **SUPPORTING INFORMATION**

702 **Fig. S1:** Correlation between RNA-seq samples and Multidimensional Scaling (MDS) Plot.

703 **Fig. S2:** Biological Process (BP) Gene Ontology (GO) analysis of clock-controlled genes (CCGs).

704 **Fig. S3:** Amplitude histogram of clock-controlled genes (CCGs) and core clock components.

705 **Fig. S4:** Biological Process (BP) Gene Ontology (GO) analysis of CCGs corresponding to the two
706 largest phase groups.

707 **Fig. S5:** Comparison of genome coverage of RNA-seq, ATH1 microarray chip and AGRINOMICS1
708 technologies.

709 **Fig. S6:** CCGs are also clock regulated at the level of alternative use of exons and/or introns
710 throughout the day.

711 **Fig. S7**: Validation of the COL2 IR splicing event, accompanying Fig. 3.

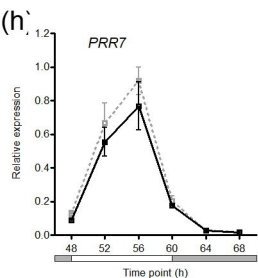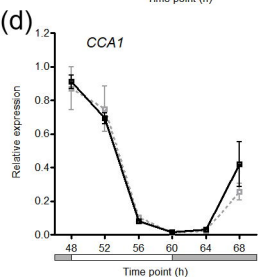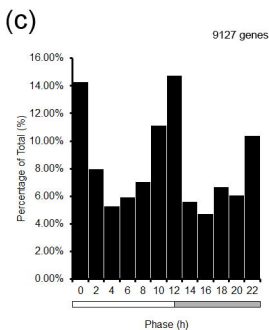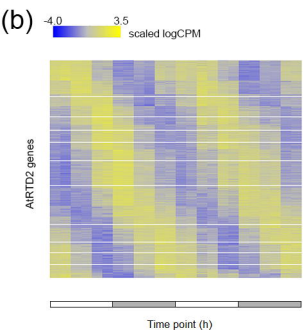712 **Fig. S8:** Phylogenetic analysis of the splicing modulator AT2G02570.
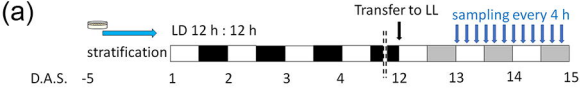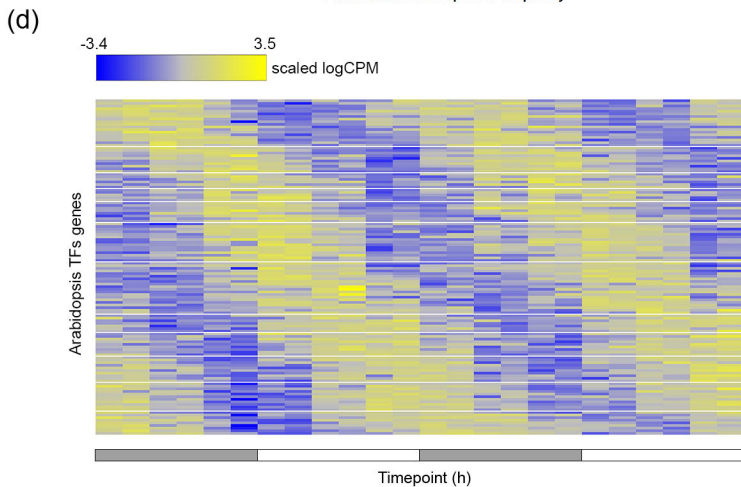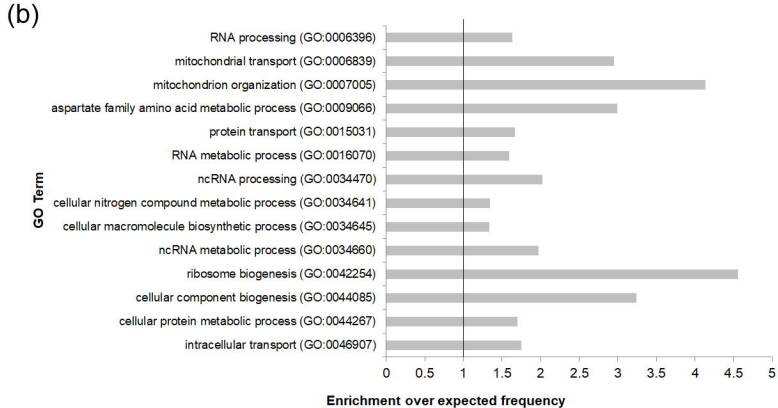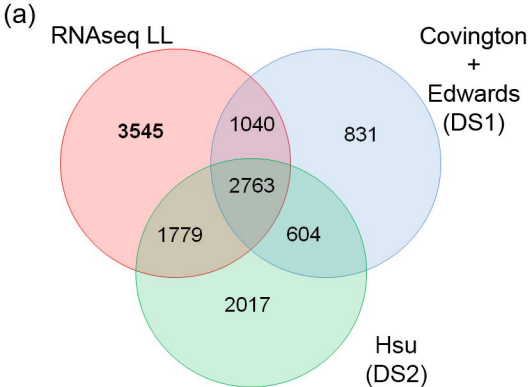
713 **Table S1** Mapping statistics.

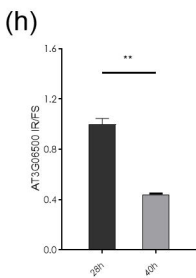714 **Table S2** List of primers used in this work.

715 **Table S3** Rhythmic clock-controlled genes (CCGs) by JTK.

716 **Table S4** GO terms of CCGs (all).

717   **Table S5** Circadian regulated genes in different datasets.

718   **Table S6** GO terms of novel CCGs (all).

719   **Table S7** Rhythmic clock-controlled splicing events (CCEs) by JTK.

720   **Table S8** GO terms of genes associated to CCEs (all).
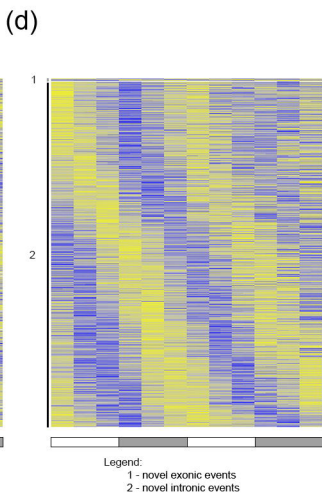
721   **Table S9** Splicing events misregulated in atspf30-1 mutant plants.

722   **Table S10** Misregulated events in atspf30-1 mutant plants that are also CCEs.

(a)

RNAseq LL — 3545, 1040, 2763, 1779, 604

Covington + Edwards (DS1) — 831

Hsu (DS2) — 2017

(b)

GO Term
RNA processing (GO:0006396)
mitochondrial transport (GO:0006839)
mitochondrion organization (GO:0007005)
aspartate family amino acid metabolic process (GO:0009066)
protein transport (GO:0015031)
RNA metabolic process (GO:0016070)
ncRNA processing (GO:0034470)
cellular nitrogen compound metabolic process (GO:0034641)
cellular macromolecule biosynthetic process (GO:0034645)
ncRNA metabolic process (GO:0034660)
ribosome biogenesis (GO:0042254)
cellular component biogenesis (GO:0044085)
cellular protein metabolic process (GO:0044267)
intracellular transport (GO:0046907)

Enrichment over expected frequency

(c)

RNAseq LL — 3399, 146

TFs — 1571

(d)

scaled logCPM −3.4 to 3.5

Arabidopsis TFs genes

Timepoint (h)

(a)

Number of events vs Type of event

- Alt3'ss: 79
- Alt5ss: 48
- IR: 300
- ES: 8
- Alt3'ss*: 83
- Alt5ss*: 44
- IR*: 174
- ES*: 49

multiple

(b)

Number of events vs Type of event

- Exon: 7
- Intron: 721

(c)

scaled logCPM: −3.01 to 2.72

Legend: Multiple:
1 - Alt5'ss   5 - Alt5'ss
2 - Alt3'ss   6 - Alt3'ss
3 - IR   7 - IR
4 - ES   8 - ES

(d)

Legend:
1 - novel exonic events
2 - novel intronic events

(e) COL2 Exon 2

P/R normalized to maximum vs Time point (h): 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64, 68

(f)

COL2 IR/FS: 28h, 36h ∗∗

(g) A/N-InvC Exon 15

P/R normalized to maximum vs Time point (h): 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64, 68

(h)

AT3G06500 IR/FS: 28h, 40h ∗∗

(a)

(b)

(c)

(a)

Number of events — Type of event: Alt3'ss (4), Alt5'ss (1), IR (95), ES (2), multiple: Alt3'ss* (1), Alt5'ss* (2), IR* (50), ES* (4)

(b)

Number of events — Type of event: Exon (0), Intron (127)

(c)

CCEs: 1481, 32, 254 — *spf30-1*

(d)

AT5G17670:E006

Col-0

spf30-1

(e)

AT5G17670:E006

P:R normalised to maximum — Time point (h)

(f)

AT5G17670:E006

IRtotal — Time point (h)

Col-0

spf30-1