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

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

[1] Comparative Genomics of Plant Development, Fundación Instituto Leloir, Instituto de Investigaciones Bioquímicas Buenos Aires (IIBBA) - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), C1405BWE Buenos Aires, Argentina.
[†] Authors contributed equally to this work.

* To whom correspondence should be addressed. Tel: +54-11-5238-7500 (#3103); Email: myanovsky@leloir.org.ar

Correspondence may also be addressed to Andres Romanowski, Tel: +54-11-5238-7500 (#2103); Email: aromanowski@leloir.org.ar

Present Address:
Soledad Perez-Santangelo, Department of Biochemistry, Division of Health Sciences, University of Otago, Dunedin, New Zealand.
Estefanía Mancini, Centre for Genomic Regulation (CRG), Parc de Recerca Biomedica de Barcelona (PRBB), Barcelona, Spain.

ABSTRACT

The circadian clock of *Arabidopsis thaliana* controls many physiological and molecular processes, allowing plants to anticipate daily changes in its environment. However, a detailed understanding of how oscillations in mRNA levels are connected to oscillations in post-transcriptional processes, such as splicing, has remained a challenge. Here we applied a combined approach of deep transcriptome sequencing and bioinformatics tools to identify novel circadian regulated genes and splicing events. Using a stringent approach, we identified 109 intron retention, 35 exon skipping, 42 alternative 5' splice site usage, and 63 alternative 3' splice site usage events under circadian regulation. We also found 24 and 1765 novel alternative exonic and intronic events. Depletion of the circadian regulated splicing factor SPF30 homolog, resulted in the disruption of a subset of clock controlled splicing events, mainly at CT16. Altogether, our global circadian RNAseq coupled with an *in silico*, event centered, splicing analysis tool offers a new approach for studying the interplay between the circadian clock and the splicing machinery at a global scale. The identification of a large number of circadian regulated splicing events, broadens our current understanding of the level of control that the circadian clock has over this posttranscriptional regulatory layer.

INTRODUCTION

Land plants on earth have evolved in a world that spins around its axis with a period close to 24 h. Therefore, they are subject to daily changes in the environment, such as changes in light quality / intensity and temperature. The cycling nature of these changes has allowed for the evolution of endogenous circadian timekeeping mechanisms that allow plants, and other organisms, to predict these changes, contributing to the optimization of their growth and development. (1-3).

In *Arabidopsis thaliana*, the molecular circadian clock consists of an interlocked series of transcriptional-translational feedback loops that cycle with a period close to 24 h (4-7). The basic

layout of this clock consists of the morning MYB transcription factor elements CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) and LATE ELONGATED HYPOCOTYL (LHY). These elements peak in the morning and repress the transcription of TIMING OF CAB EXPRESSION 1 (TOC1), also known as PSEUDO RESPONSE REGULATOR 1 (PRR1). TOC1 peaks in the evening and represses CCA1/LHY. Another loop involves the downregulation of CCA1/LHY throughout the rest of the day by the consecutive repression action of PRR9, PRR7 and later PRR5. During the evening, the elements EARLY FLOWERING 3 (ELF3), ELF4 and LUX ARRHYTHMO (LUX) associate into an "evening complex" (EC), repressing PRR9 and LUX. Additionally, several activators have been described. In particular, the CCA1/LHY homologs reveille 8 (RVE8), RVE6 and RVE4 are morning factors that drive the expression of PRR5 and TOC1. Together with the RVEs, the NIGHT LIGHT-INDUCIBLE AND CLOCK-REGULATED (LNK) 1-4 family of genes also contribute to activate the expression of PRR5 and ELF4 (7).

This circadian clock integrates environmental *inputs* to fine tune its oscillations (entrainment) and then controls the timing of several molecular, physiological and developmental outputs, like stomatal opening, leaf movement, hormone synthesis and growth among others (6,8). One of the many ways that this can be accomplished is through the control of gene expression, a process that involves many different steps (9). One of these steps is the processing of precursor mRNAs (pre-mRNAs). Plant pre-mRNAs are composed of continuous nucleotide segments known as exons and introns. In order to produce a mature functional mRNA, exonic segments are kept and introns are removed through a process called splicing.

The process of splicing occurs at the spliceosome, a dynamic protein-RNA complex composed of small nuclear ribonucleoprotein particles (snRNPs) and auxiliary proteins that assemble at exon-intron boundaries known as donor 5'-splice site (ss) and acceptor 3'-ss. The selection and recognition of this splice sites occurs through the interaction of spliceosome components with *cis* acting sequences and the *trans* acting proteins. AS occurs through the alternative selection of 5'ss, 3'ss, exon inclusion or skipping and intron retention (10). Through this process, different functional mRNAs can arise from the transcription of the same gene, ultimately expanding the repertoire of protein products that can be encoded by a single gene (10,11).

The first evidence of crosstalk between AS and the circadian clock was described in our lab in 2010. A mutation in the clock regulated protein arginine methyltransferase 5 (PRMT5) was found to alter the AS of the clock component PRR9, causing a significant alteration in period length (12). Since then, additional work has shown that AS is a mechanism that can couple changes in environmental temperature to circadian timing by altering the splicing of clock components (13-16). Also, mutations in the spliceosome components SKIP and STIPL1 can affect circadian timing (17,18). However, no comprehensive survey has been done to date to understand how widespread is clock regulated AS. In this work, we used mRNAseq and a custom *in house* developed bioinformatic pipeline to perform a genome wide analysis on circadian regulation of AS in *A. thaliana* plants.

MATERIAL AND METHODS

Plant materials

All *Arabidopsis thaliana* plants used in this study were of the Columbia (Col-0) ecotype. *AT2G02570* mutant alleles *atspf30-1* (SALK_052016C) and *atspf30-2* (SALK_081292C) were obtained from the Arabidopsis Biological Resource Center (ABRC).

Circadian Leaf Movement Analysis

For leaf movement analysis, plants were entrained under a 16 h light : 8 h dark cycle, transferred to continuous white fluorescent light, and the position of the first pair of leaves was recorded every 2 h for 5–6 d using Image J software (http://imagej.nih.gov). Period estimates were calculated using Brass 3.0 software (Biological Rhythms Analysis Software System; http://www.amillar.org). Briefly, raw data was imported; baseline detrended and analyzed using the FFT-NLLS algorithm, as described previously (19). Statistical analysis was performed using two tailed Student's t test.

Flowering Time Analysis

For flowering time analysis, plants were grown under long days (16 h light : 8 h dark), short days (8 h light : 16 h dark) or continuous light at a constant temperature of 22 °C. Flowering time was estimated by counting the number of rosette leaves at the time of bolting. Statistical analysis was performed using two tailed Student's t test.

Growth conditions and tissue collection for RNAseq studies

Col-0 and *atspf30-1* seeds were sown on Murashige and Skoog medium (GIBCO BRL) and cold stratified (4ºC, darkness) for 3 days. For non-circadian experiments, the MS plates were then transferred to 22 °C and continuous light (LL; 50 $\mu mol \cdot m^{-2} \cdot s^{-1}$ of white light). For circadian experiments, the MS plates were transferred to 22 °C under LD12:12 (12 h light/12 h dark cycles; 50 $\mu mol \cdot m^{-2} \cdot s^{-1}$ of white light) conditions for 11 days and then transferred to continuous light (LL; 50 $\mu mol \cdot m^{-2} \cdot s^{-1}$ of white light).

Samples containing 15-30 seedlings were flash frozen in liquid nitrogen. For circadian studies, the samples were collected at the timepoints indicated in Figure 1A. After all samples were collected, total RNA was extracted using the Trizol method according to the manufacturer's protocol (ThermoFisher) and quantified using a NanoDrop spectrophotometer (ThermoFisher). Total RNA integrity was further checked by agarose gel electrophoresis.

Library preparation and RNA sequencing

Libraries were prepared following the TruSeq RNA v2 Sample Preparation Guide (Illumina). Briefly, 3 µg of total RNA was polyA-purified and fragmented, and first-strand cDNA synthesized by reverse transcriptase (SuperScript II; Invitrogen) and random hexamers. This was followed by RNA degradation and second-strand cDNA synthesis. End repair process and addition of a single A nucleotide to the 3′ ends allowed ligation of multiple indexing adapters. Then, an enrichment step of 12 cycles of PCR was performed. Library validation included size and purity assessment with the Agilent 2100 Bioanalyzer and the Agilent DNA 1000 kit (Agilent Technologies). Samples were diluted to 10nM and pooled to create the 12X multiplexed DNA libraries, and 12–15 pmol of material was loaded onto each lane. Template hybridization, extension of the template, amplification of the template, cluster generation, sequencing primer addition, and single end chemistry were performed using C-BOT (formerly known as Cluster Station). Base calls were made using CASAVA.

The multiplexed samples were loaded onto two lanes on an Illumina HiSeq 2500 sequencer, providing 100-bp single-end reads.

Processing of RNA reads and alignment

Reads were quality-filtered using the standard Illumina process and demultiplexed with two allowed barcode mismatches. Sequence files were generated in FASTQ format. The Bowtie/TopHat suite (20) was used to map reads to the *A. thaliana* TAIR10 reference genome. Along with the prebuilt *A. thaliana* index, the reference genome was downloaded from ENSEMBL (December 2014). Default values for TopHat parameters were used with the exception of maximum intron length parameter, which was set to a value of 5,000 nt following estimated values, as previously reported (21). Supplemental Table 1 provides a summary table of main read count statistics.

RNA sequencing analysis

Read counts were calculated for genes, exon bins, AS bins and intron bins with the ASpli package (22) using custom scripts written in R. Coverage plots were generated using the IGB Browser (23).

Detection of cycling was performed using JTK_Cycle (24) implemented in R (64-bit, version 2.12.1). Relative amplitude estimates were calculated by dividing the subtraction of the absolute peak expression and through levels, by the sum of through level plus one (to avoid divisions by zero). Heatmaps were generated using scripts for R. PSI/PIR indexes were calculated as previously described (25,26) using custom scripts written in R. GO analyses were performed with the Virtual Plant web service (27). All code and analyses are available on demand.

Quantitative and reverse transcriptase PCR

For time course analysis, 10-d-old plants were grown under 12 h light : 12 h dark cycles at 22 °C (for gene expression analysis of splicing factors in wild-type plants) and then transferred for to continuous white light at 22 °C. Samples were collected every 4 h for 2 days, starting at 24 h, after transferring the plants to constant light conditions. Total RNA was obtained from these samples using TRIzol reagent (Invitrogen). One microgram of RNA was treated with RQ1 RNase-Free DNase (Promega) and subjected to retro transcription with Superscript III (Invitrogen) and oligo dT according to the manufacturer's instructions. Synthesized cDNAs were amplified with FastStart Universal SYBR GreenMaster (Roche) using the Mx3000P Real Time PCR System (Agilent Technologies) cycler. The PP2A (AT1G13320) transcript was used as a reference gene.

Quantitative RT-PCR quantification was conducted using the standard curve method as described in the Methods and Applications Guide from Agilent Technologies. Primers and targets for gene expression and splicing events validations are described in Supplemental Table 2.

Semiquantitative and reverse transcriptase PCR (sqRT-PCR)

cDNA from plants grown in 16-h light/8-h dark cycles and constant light were used for PCR amplification. Amplification of cDNA was carried out for 30 cycles to measure splicing at the linear phase. RT-PCR products were electrophoresed in 8% SDS-PAGE gels and detected with EtBr. Primers used for amplification of each gene are described in Supplemental Table 2.

RESULTS

Global RNAseq of the circadian transcriptome reveals known and novel circadian regulated genes.

To analyze the circadian regulation of gene expression, we performed a global transcriptome deep sequencing of 15-day old *A. thaliana* plants grown under a light/dark (LD) 12 h : 12 h photoperiod, transferred to constant light conditions (LL), and harvested after 24 h in LL every 4 h for a total of 2 days, with two biological replicates. This approach, in contrast to previous studies that used microarray-based approaches, allowed us to analyze genes which are not represented in microarrays. Briefly, total RNA was amplified, and ribosomal RNA was depleted (see Methods). Sequencing libraries were prepared from amplified RNA using standard techniques and then sequenced on an Illumina HiSeq 2500 high-throughput sequencer using two flow-cell lanes to generate 100-bp paired-end reads. On average, each sample generated an average of 16.6 million high quality short nucleotide reads were recovered for each sample (see Supplemental Table 1 online) and an average of 90.6% of reads were mapped to the TAIR10 genome sequence (see Supplemental Table 1 online), using the TopHat aligner (20). Taken together, these data represent more than 361 million short nucleotide reads (100bp length) and 36.1 Gb of successfully aligned sequence. This constitutes the deepest sequencing of a plant circadian transcriptome to the moment.

Gene expression was computed with the ASPli R package (22) by obtaining a table of gene counts, which represent the total number of reads aligning to each gene (33,602 annotated TAIR genes). On average, the reproducibility between replicate samples (i.e., samples from the same time point on different days) was excellent, with average R^2 values approaching 0.99 (see Supplemental Figure 1 online). We then obtained the number of reads per gene length (read density) for each gene, and only those that had a value higher than 0.05 in any of the time points throughout the time course, were further analysed (18,670 genes).

In order to improve the power of subsequent rhythmic statistical tests we applied a pre-screening step to exclude any obviously non-rhythmic transcripts, as previously suggested (28). Only the genes that showed a variation throughout the time-course (14,105 genes; negative binomial generalized linear model (GLM) model fit, $p<0.05$ and $q<0.10$) were then subjected to circadian rhythmicity analysis. To this end, several statistical approaches have been described (29-32). In this work we applied the JTK_cycle algorithm, which was reported to identify rhythmic genes with enhanced confidence, improved computational and more resistance to outliers than other algorithms (24). Using this method, a total of 5,118 genes were found to be rhythmic and exhibited a circadian expression pattern (JTK_cycle, $p<0.05$ and $q<0.15$; Figure 1A and see Supplemental Table 3 online), under our assay conditions. To verify the validity of our approach we examined the expression pattern of known core clock genes in our data set and validated them by qPCR (Figure 1C-J). As expected, the qPCR patterns matched the patterns produced by our RNAseq data. Interestingly, peak amplitude on day 3 is slightly lower than peak amplitude on day 2, indicating a progressive decrease of amplitude in constant conditions.

We then performed a Gene Ontology (GO) analysis of this dataset (Supplemental Figure 2 and Supplemental Table 4 online) which showed that clock controlled genes (CCGs) were involved in diverse biological processes, including but not limited to immune response, hormone signalling, light signalling, metabolism and photosynthesis, which is in accordance to previous reports (33,34). A phase distribution analysis of the circadian regulated genes was done and found that 36.23% of the genes exhibit a peak phase expression at LL12, which corresponds to the day to night transition of the LD 12 h : 12 h entraining photoperiod (Figure 1B). Genes that peaked at this particular time of the free running cycle were involved in 38 biological processes (Fisher's Exact Test, $p < 0.05$, see Supplemental Figure 4 online), including response to ethylene stimulus, sulfur compound biosynthesis, response to light, response to gibberellin, response to jasmonic acid.

Upon comparison of our dataset with previous published global *A. thaliana* circadian gene expression datasets (33,34) (see Supplemental Table 5 online) we found an overlap of 43.16% (2209 genes) and 57.67% (2952 genes) to the Covington (33) (DS1) and Hsu (34) (DS2) datasets, respectively. A total of 1831 genes (35.77%) were found to cycle in all three datasets (Figure 2A). By combining the three datasets we obtain that over 32% of *A. thaliana* genes (10822 / 33602 TAIR10 genes) have been described to be circadian regulated to this date, in global transcriptomic studies (see Supplemental Table 5 online).

Novel circadian transcripts

Our dataset also revealed novel circadian regulated genes that had not been identified by the two previous global studies (Figure 2A). This was to be expected, to a certain degree, because of the difference in genome coverage of the different technologies used (RNAseq, ATH1 microarray and AGRONOMICS1 tiling array). However, hypothetically there should be no difference in the possible number of positive genes between our RNAseq approach and the AGRONOMICS1 platform (35) (see Supplemental Figure 5 online). Nevertheless, we found 1788 circadian genes that are absent from previous global studies. We will refer to these genes as novel circadian genes. GO analysis of this subset revealed an enrichment in terms related to immunity and RNA metabolism (Figure 2B and see Supplemental Table 6 online). Some of the genes in this subset have been previously identified as circadian regulated in non-global studies. Such is the case, for example, of the widely used circadian reporter CAB2 (36), which further validates our approach.

Interestingly, many members of the B-Box (BBX) family are included in the novel CCGs subset. This family comprises a class of zinc-finger transcription factors containing one or two B-box motifs; and, in some cases contain a CCT (CONSTANS, CO-like, and TOC1) domain. These proteins are key factors in regulatory networks controlling growth and developmental processes, including photomorphogenesis, photoperiodic regulation of flowering, shade avoidance, and responses to biotic and abiotic stresses (37). The RNAseq data of the expression pattern of the novel BBX genes can be seen in Figure 2G. Other interesting genes that were represented in this novel subset are members of the bHLH and bZIP families of transcription factors (Figure 2D and 2F, respectively), pre-microRNA genes (Figure 2H), small auxin response (Figure 2E) and CCT domain containing genes (Figure 2C).

Further study of these genes might allow to gain new insight into how the circadian clock regulates gene expression, auxin responses and microRNA regulation.

The circadian clock regulates alternative splicing events

In this work, we took advantage of the accuracy of RNAseq technology and then applied a custom pipeline designed to specifically detect and quantify splicing events. In order to do so, we aligned the RNAseq reads to the TAIR10 reference genome to generate the BAM files required as input for the *in house* developed Aspli Bioconductor R package. Aspli was designed by our lab for the specific purpose of detecting and quantifying splicing events (22). Briefly, when used to detect changes in splicing events, this algorithm uses a known genome annotation to bin the genome in discrete exon and intron “bins”, and then quantifies the counts that each bin has over time and normalizes it to gene expression. After obtaining the normalized bin counts, we proceeded to analyze splicing events rhythmicity as explained for gene expression. Using this method, several clock controlled splicing events (CCEs) were discovered. In particular, 245 intron retention, 96 exon skipping, 128 5’ splice site (5’ss), 194 3’ splice site (3’ss), and 5096 novel exonic and 4652 novel intronic were found to be rhythmic and exhibited a circadian expression pattern (JTK_cycle, $p<0.05$ and $q<0.15$), under our assay conditions. With the purpose of eliminating false positives, we applied a more stringent filtering step and kept only those splicing events with splice-junction reads evidence. To do so, we calculated the Percentage of Spliced In (PSI) and the Percentage of Intron Retention (PIR) indexes as previously described (26) for exon and intron bins, respectively; and kept the subset of circadian splicing events that showed a variation in PSI/PIR between 5% and 95% across all the analyzed timepoints. After applying this filter 109 intron retention, 35 exon skipping, 42 5’ splice site (5’ss), 63 3’ splice site (3’ss) (Figura 3A), and 24 novel exonic and 1765 novel intronic (Figura 3B) events were further considered (see Supplemental Table 7 online). It is interesting to note that this study also reveals that 499 CCGs are also clock regulated at the level of alternative use of exons and/or introns throughout the day (see Supplemental Figure 6 and Supplemental Table 7 online). On the other hand, there also are 1201 genes whose expression is not circadian regulated, but their exon/intron use is (see Supplemental Figure 6 and Supplemental Table 7 online), and 4619 CCGs that had no detectable CCEs.

We then performed a Gene Ontology (GO) analysis of this dataset (see Supplemental Table 8 online) which showed that clock controlled splicing events (CCEs) corresponded to genes involved in post-embryonic development (GO:0009791), primary metabolism (GO:0044283), nitrogen metabolism (GO:0006807), nucleic acid metabolism (GO:0006139), microtubule dependent processes (GO:0007018) and mRNA processing (GO:0006397).

Validation of clock controlled alternative splicing events

In order to verify the validity of other novel genes we examined the expression pattern of a subset of splicing events from our data set and validated them by semi-quantitative (sq)RT-PCR followed by SDS-PAGE. Additionally, results were further corroborated by qPCR, obtaining in both cases results that matched the patterns produced by our RNAseq analysis pipeline data (see Supplemental Figure

7). We validated the intron retention splicing event of the COL2 gene, homolog of the photoperiod and flowering regulator CONSTANS, which was previously described (38) and was therefore a perfect positive control. According to the bioinformatics analysis of our pipeline, the IR precedes the subjective day and has a peak at the day/night transition, as can be seen in the RNAseq coverage plots (see Supplemental Figure 7A) and the normalized bin counts (see Supplemental Figure 7B). This event also fulfils our stringent criteria of having reads that span the intron-exon junction (see Supplemental Figure C). The analysis by sqRT-PCR followed by SDS-PAGE (see Supplemental Figure 7D) and qPCR (see Supplemental Figure 7E) demonstrate the validity of the bioinformatics analysis used in this work. Additionally, our results exhibit a peak IR at CT36, which is in complete accordance to what was previously described (38), further validating our analysis pipeline.

Several splicing modulators are under circadian clock control

We reasoned that this circadian control of alternative splicing could be due, at least in part, to an underlying circadian regulation of the splicing machinery. To test this hypothesis we compiled a list of all known splicing modulators (13) and intersected it with the Covington dataset (DS1), the Hsu & Harmer dataset (DS2) and our own. Out of the 27 genes that were shown to be clock controlled in at least one dataset, 11 were rhythmic in all three datasets (Figure 4A). Because different research groups, each with its own environmental conditions, obtained these datasets independently, we reasoned that these 11 splicing modulators would be the most robustly rhythmic ones.

One of these genes, AT2G02570, has been previously described as the homolog of the mammalian Survival of Motor Neuron (SMN), which controls spliceosome assembly (39). However, our reverse BlastP bioinformatics search revealed that AT2G02570 is more closely related to Splicing Factor 30 (SPF30) and might be instead a paralog to SMN (see Supplemental Figure 8).

Based on increasing evidence indicating that alternative splicing is regulated not only by SR and hnRNP auxiliary factors, but also by changes in the levels or activities of core spliceosomal components or proteins that modulate the kinetics of spliceosome assembly (40,41), we decided to further investigate the biological relevance of AT2G02570.

AtSPF30 affects a subset of clock-regulated events

In order to investigate the role of spf30 on the modulation of CCSEs we first validated its rhythmic transcriptional pattern by qPCR (Figure 4B-D) and isolated two different T-DNA insertion mutants (*atspf30-1* and *atspf30-2*, Figure 5A). Plants homozygous for either mutant allele were morphologically very similar to the wild type (Figure 5A). A physiological characterization of two clock regulated outputs, leaf movement and flowering, revealed no significant difference in leaf movement periodicity between Col-0 and either mutant (Figure 5B) and a consistent slightly early flowering phenotype across the different photoperiods tested (LL, SD and LD, *$p<0.05$ ANOVA followed by Dunnett's *post hoc* test against Col-0; Figure 5C).

An initial RNAseq analysis of *atspf30* mutant plants grown in constant light revealed 688 altered alternative splicing events when compared to Col-0 plants grown in the same conditions (Figure 6A). In wild-type plants, the most abundant AS events were those associated with the use of alternative 3′ splice sites (alt 3′; 33%), followed by intron retention (IR; 32%), alternative 5′ splice sites (alt 5′, 22%), and exon skipping (13%). However, among the AS events altered in the *atspf30* mutant, we found an increase in the proportion of IR events, a decrease in 3'alt events and almost no 5'alt events relative to their frequency in wild type plants (Figure 6B). Interestingly, 43 of these events were among our PSI/PIR filtered subset of CCEs (Figure 6A). A phase histogram of these events revealed that even though the majority was evenly spread throughout the day, there was a significant enrichment (Chi-squared test, $p<0.015$) for splicing events that exhibited a maximum of inclusion at CT16 (Figure 6C). This enrichment accounts for >30% of the total of AS events affected by AtSPF30. With this in mind, we selected an IR event with a peak phase at CT16 for validation (Figure 6D-F). qPCR quantification of a circadian timecourse showed that lack of functional AtSPF30 affected the overall inclusion of the retention of Intron 1 along the entire circadian day without affecting the rhythmic pattern (Figure 6E-F). Hence, in the case of this splicing event, AtSPF30 is modulating both the baseline level of intron retention and its overall amplitude.

DISCUSSION

AS is highly prevalent in plants, affecting pre-mRNAs of more than 60% of intron-containing genes (42). Many genes undergo AS in response to development and changes in the environment, such as light, cold or high temperature (14,15,42,43). For some time now, evidence has been accumulating pointing to the fact that AS can affect clock components and, that mutations affecting spliceosome related components can alter the circadian period (12,13,16,17,44). Furthermore, studies performed in fruit flies, mice and human cell lines have identified a role for the circadian clock in the control of the splicing process (45-47). Yet there have been no global transcriptomic studies that allowed for understanding the extent of clock control on AS on any plant species. There is perhaps one interesting exception, a global study limited to the analysis of cycling introns (38). This work was performed by Hazen et al. in 2009 using GeneChip AtTILE1 tiling array technology. A total of 499 genes were reported to contain cycling introns at the time. In this study, we used NGS Illumina mRNAseq technology. This allowed us to avoid the technical issues inherent to microarray probe performance such as cross-hybridization, non-specific hybridization and limited detection range of individual probes (48). By combining the higher resolution of mRNAseq with an *in house* developed bioinformatic pipeline we were able to find 109 intron retention, 35 exon skipping, 42 alternative 5' splice site usage, and 63 alternative 3' splice site usage events that are under circadian regulation. We also found 24 and 1765 novel alternative exonic and intronic events, respectively. Interestingly, of the 1615 genes that were found to contain cycling introns, only 76 overlapped with the 499 reported by Hazen et al (15%, 76 out of 499, see Supplemental Table 11).

In addition to documenting CCEs, we found that a clock-controlled homolog of the splicing factor SPF30 affected a subset of these when mutated. 43 events of the 688 events affected on an SPF30 mutant were among our PSI/PIR filtered subset of CCEs. Furthermore, a closer inspection of the

phase distribution of these splicing events showed that it preferentially affected intron retention events with a maximum of inclusion at CT16. This implies that the circadian control of alternative splicing is due, at least in part, to an underlying circadian regulation of the splicing machinery. In addition to SPF30, 26 other splicing factors were found to be under circadian regulation in at least one dataset. 10 of those were circadian in all known circadian datasets. The contribution of each of these robust cycling factors on circadian control of AS remains to be determined and should be further studied. Prospectively, modulation of these circadian controlled splicing factors could lead to time specific fine tuning of splicing events.

Splicing factors were also subject to circadian splicing events. In this sense, our stringent analysis revealed 59 splicing related genes that are also regulated at the splicing level. A future careful and detailed study of the biological relevance of these splicing events will help us gain a better understanding of how the clock regulates alternative splicing and viceversa.

The number of CCEs reported in this work is very conservative and will likely increase in the future, as sequencing depth of circadian studies increases. The advent of novel long read technologies like those provided by Pacific Biosystems or Oxford Nanopore will likely allow for reliable transcript isoform level quantification at the global scale in the very near future.

AVAILABILITY

The datasets supporting the conclusions of this article are included within the article and its additional files.

ACCESSION NUMBERS

The RNAseq datasets supporting the conclusions of this article are available in the ArrayExpress (49) database at EMBL-EBI (www.ebi.ac.uk/arrayexpress), under accession number E-MTAB-7933.

SUPPLEMENTARY DATA

Supplementary Data are available online.

ACKNOWLEDGEMENT

This work was supported by grants from Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) and the International Centre for Genetic Engineering and Biotechnology (ICGEB) to MY. AR was supported by postdoctoral fellowships from Comisión Nacional de Investigaciones Científicas (CONICET) and Fundación Bunge y Born (FBB). RGS was supported by Fundación Bunge y Born (FBB) and, SPS, MJDL and EM were supported by CONICET.

We wish to thank Dr. Julieta Lisa Mateos for fruitful discussions and critical reading of the manuscript.

FUNDING

This work was supported by grants from Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) and the International Centre for Genetic Engeneering and Biotechnology (ICGEB) (to M.J.Y).

CONFLICT OF INTEREST

Conflict of interest statement. None declared.

REFERENCES

1. Dunlap, J.C., Loros, J.J. and DeCoursey, P.J. (2004) *Chronobiology : biological timekeeping.* Sinauer Associates, Sunderland, Mass.
2. Dodd, A.N., Salathia, N., Hall, A., Kevei, E., Toth, R., Nagy, F., Hibberd, J.M., Millar, A.J. and Webb, A.A. (2005) Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science*, **309**, 630-633.
3. Green, R.M., Tingay, S., Wang, Z.Y. and Tobin, E.M. (2002) Circadian rhythms confer a higher level of fitness to Arabidopsis plants. *Plant physiology*, **129**, 576-584.
4. Hernando, C.E., Romanowski, A. and Yanovsky, M.J. (2017) Transcriptional and post-transcriptional control of the plant circadian gene regulatory network. *Biochimica et biophysica acta. Gene regulatory mechanisms*, **1860**, 84-94.
5. Harmer, S.L., Panda, S. and Kay, S.A. (2001) Molecular bases of circadian rhythms. *Annual review of cell and developmental biology*, **17**, 215-253.
6. Millar, A.J. (2016) The Intracellular Dynamics of Circadian Clocks Reach for the Light of Ecology and Evolution. *Annual review of plant biology*, **67**, 595-618.
7. McClung, C.R. (2014) Wheels within wheels: new transcriptional feedback loops in the Arabidopsis circadian clock. *F1000prime reports*, **6**, 2.
8. McClung, C.R. (2006) Plant circadian rhythms. *The Plant cell*, **18**, 792-803.
9. Romanowski, A. and Yanovsky, M.J. (2015) Circadian rhythms and post-transcriptional regulation in higher plants. *Frontiers in plant science*, **6**, 437.
10. Kornblihtt, A.R., Schor, I.E., Allo, M., Dujardin, G., Petrillo, E. and Munoz, M.J. (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nature reviews. Molecular cell biology*, **14**, 153-165.
11. Saldi, T., Cortazar, M.A., Sheridan, R.M. and Bentley, D.L. (2016) Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *Journal of molecular biology*, **428**, 2623-2635.
12. Sanchez, S.E., Petrillo, E., Beckwith, E.J., Zhang, X., Rugnone, M.L., Hernando, C.E., Cuevas, J.C., Godoy Herz, M.A., Depetris-Chauvin, A., Simpson, C.G. *et al.* (2010) A methyl transferase links the circadian clock to the regulation of alternative splicing. *Nature*, **468**, 112-116.
13. Perez-Santangelo, S., Schlaen, R.G. and Yanovsky, M.J. (2013) Genomic analysis reveals novel connections between alternative splicing and circadian regulatory networks. *Briefings in functional genomics*, **12**, 13-24.

14. James, A.B., Syed, N.H., Bordage, S., Marshall, J., Nimmo, G.A., Jenkins, G.I., Herzyk, P., Brown, J.W. and Nimmo, H.G. (2012) Alternative splicing mediates responses of the Arabidopsis circadian clock to temperature changes. *The Plant cell*, **24**, 961-981.
15. James, A.B., Syed, N.H., Brown, J.W. and Nimmo, H.G. (2012) Thermoplasticity in the plant circadian clock: how plants tell the time-perature. *Plant signaling & behavior*, **7**, 1219-1223.
16. Filichkin, S.A., Cumbie, J.S., Dharmawardhana, P., Jaiswal, P., Chang, J.H., Palusa, S.G., Reddy, A.S., Megraw, M. and Mockler, T.C. (2015) Environmental Stresses Modulate Abundance and Timing of Alternatively Spliced Circadian Transcripts in Arabidopsis. *Molecular plant*, **8**, 207-227.
17. Jones, M.A., Williams, B.A., McNicol, J., Simpson, C.G., Brown, J.W. and Harmer, S.L. (2012) Mutation of Arabidopsis spliceosomal timekeeper locus1 causes circadian clock defects. *The Plant cell*, **24**, 4066-4082.
18. Wang, X., Wu, F., Xie, Q., Wang, H., Wang, Y., Yue, Y., Gahura, O., Ma, S., Liu, L., Cao, Y. *et al.* (2012) SKIP is a component of the spliceosome linking alternative splicing and the circadian clock in Arabidopsis. *The Plant cell*, **24**, 3278-3295.
19. Plautz, J.D., Straume, M., Stanewsky, R., Jamison, C.F., Brandes, C., Dowse, H.B., Hall, J.C. and Kay, S.A. (1997) Quantitative analysis of Drosophila period gene transcription in living animals. *Journal of biological rhythms*, **12**, 204-217.
20. Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105-1111.
21. Hong, X., Scofield, D.G. and Lynch, M. (2006) Intron size, abundance, and distribution within untranslated regions of genes. *Molecular biology and evolution*, **23**, 2392-2404.
22. Mancini, E., Iserte, J., Yanovsky, M. and Chernomoretz, A. (2019) ASpli: Analysis of alternative splicing using RNA-Seq. R package version 1.8.1. *Bioconductor*, **Release (3.8)**.
23. Freese, N.H., Norris, D.C. and Loraine, A.E. (2016) Integrated genome browser: visual analytics platform for genomics. *Bioinformatics*, **32**, 2089-2095.
24. Hughes, M.E., Hogenesch, J.B. and Kornacker, K. (2010) JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *Journal of biological rhythms*, **25**, 372-380.
25. Pervouchine, D.D., Knowles, D.G. and Guigo, R. (2013) Intron-centric estimation of alternative splicing from RNA-seq data. *Bioinformatics*, **29**, 273-274.
26. Gueroussov, S., Gonatopoulos-Pournatzis, T., Irimia, M., Raj, B., Lin, Z.Y., Gingras, A.C. and Blencowe, B.J. (2015) An alternative splicing event amplifies evolutionary differences between vertebrates. *Science*, **349**, 868-873.
27. Katari, M.S., Nowicki, S.D., Aceituno, F.F., Nero, D., Kelfer, J., Thompson, L.P., Cabello, J.M., Davidson, R.S., Goldberg, A.P., Shasha, D.E. *et al.* (2010) VirtualPlant: a software platform to support systems biology research. *Plant physiology*, **152**, 500-515.
28. Keegan, K.P., Pradhan, S., Wang, J.P. and Allada, R. (2007) Meta-analysis of Drosophila circadian microarray studies identifies a novel set of rhythmically expressed genes. *PLoS computational biology*, **3**, e208.
29. Deckard, A., Anafi, R.C., Hogenesch, J.B., Haase, S.B. and Harer, J. (2013) Design and analysis of large-scale biological rhythm studies: a comparison of algorithms for detecting periodic signals in biological data. *Bioinformatics*, **29**, 3174-3180.
30. Wichert, S., Fokianos, K. and Strimmer, K. (2004) Identifying periodically expressed transcripts in microarray time series data. *Bioinformatics*, **20**, 5-20.
31. Wijnen, H., Naef, F. and Young, M.W. (2005) Molecular and statistical tools for circadian transcript profiling. *Methods in enzymology*, **393**, 341-365.
32. Michael, T.P., Mockler, T.C., Breton, G., McEntee, C., Byer, A., Trout, J.D., Hazen, S.P., Shen, R., Priest, H.D., Sullivan, C.M. *et al.* (2008) Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS genetics*, **4**, e14.

472 33. Covington, M.F., Maloof, J.N., Straume, M., Kay, S.A. and Harmer, S.L. (2008) Global
473 transcriptome analysis reveals circadian regulation of key pathways in plant growth and
474 development. *Genome biology*, **9**, R130.
475 34. Hsu, P.Y. and Harmer, S.L. (2012) Circadian phase has profound effects on differential
476 expression analysis. *PloS one*, **7**, e49853.
477 35. Rehrauer, H., Aquino, C., Gruissem, W., Henz, S.R., Hilson, P., Laubinger, S., Naouar, N.,
478 Patrignani, A., Rombauts, S., Shu, H. *et al.* (2010) AGRONOMICS1: a new resource for
479 Arabidopsis transcriptome profiling. *Plant physiology*, **152**, 487-499.
480 36. Millar, A.J. and Kay, S.A. (1991) Circadian Control of cab Gene Transcription and mRNA
481 Accumulation in Arabidopsis. *The Plant cell*, **3**, 541-550.
482 37. Gangappa, S.N. and Botto, J.F. (2014) The BBX family of plant transcription factors. *Trends in*
483 *plant science*, **19**, 460-470.
484 38. Hazen, S.P., Naef, F., Quisel, T., Gendron, J.M., Chen, H., Ecker, J.R., Borevitz, J.O. and Kay,
485 S.A. (2009) Exploring the transcriptional landscape of plant circadian rhythms using genome
486 tiling arrays. *Genome biology*, **10**, R17.
487 39. Kroiss, M., Schultz, J., Wiesner, J., Chari, A., Sickmann, A. and Fischer, U. (2008) Evolution of
488 an RNP assembly system: a minimal SMN complex facilitates formation of UsnRNPs in
489 Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United*
490 *States of America*, **105**, 10045-10050.
491 40. Saltzman, A.L., Pan, Q. and Blencowe, B.J. (2011) Regulation of alternative splicing by the
492 core spliceosomal machinery. *Genes & development*, **25**, 373-384.
493 41. Papasaikas, P., Tejedor, J.R., Vigevani, L. and Valcarcel, J. (2015) Functional splicing network
494 reveals extensive regulatory potential of the core spliceosomal machinery. *Molecular cell*, **57**,
495 7-22.
496 42. Reddy, A.S., Marquez, Y., Kalyna, M. and Barta, A. (2013) Complexity of the alternative
497 splicing landscape in plants. *The Plant cell*, **25**, 3657-3683.
498 43. Kwon, Y.J., Park, M.J., Kim, S.G., Baldwin, I.T. and Park, C.M. (2014) Alternative splicing and
499 nonsense-mediated decay of circadian clock genes under environmental stress conditions in
500 Arabidopsis. *BMC plant biology*, **14**, 136.
501 44. Cui, Z., Xu, Q. and Wang, X. (2014) Regulation of the circadian clock through pre-mRNA
502 splicing in Arabidopsis. *Journal of experimental botany*, **65**, 1973-1980.
503 45. Genov, N., Basti, A., Abreu, M., Astaburuaga, R. and Relogio, A. (2019) A bioinformatic
504 analysis identifies circadian expression of splicing factors and time-dependent alternative
505 splicing events in the HD-MY-Z cell line. *Sci Rep*, **9**, 11062.
506 46. Preussner, M., Goldammer, G., Neumann, A., Haltenhof, T., Rautenstrauch, P., Muller-
507 McNicoll, M. and Heyd, F. (2017) Body Temperature Cycles Control Rhythmic Alternative
508 Splicing in Mammals. *Molecular cell*, **67**, 433-446 e434.
509 47. Hughes, M.E., Grant, G.R., Paquin, C., Qian, J. and Nitabach, M.N. (2012) Deep sequencing
510 the circadian and diurnal transcriptome of Drosophila brain. *Genome Res*, **22**, 1266-1281.
511 48. Agarwal, A., Koppstein, D., Rozowsky, J., Sboner, A., Habegger, L., Hillier, L.W., Sasidharan, R.,
512 Reinke, V., Waterston, R.H. and Gerstein, M. (2010) Comparison and calibration of
513 transcriptome data from RNA-Seq and tiling arrays. *BMC genomics*, **11**, 383.
514 49. Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y.A., Williams, E., Dylag, M.,
515 Kurbatova, N., Brandizi, M., Burdett, T. *et al.* (2015) ArrayExpress update--simplifying data
516 submissions. *Nucleic acids research*, **43**, D1113-1116.
517 50. Mockler, T.C., Michael, T.P., Priest, H.D., Shen, R., Sullivan, C.M., Givan, S.A., McEntee, C.,
518 Kay, S.A. and Chory, J. (2007) The DIURNAL project: DIURNAL and circadian expression
519 profiling, model-based pattern matching, and promoter analysis. *Cold Spring Harbor*
520 *symposia on quantitative biology*, **72**, 353-363.

521

TABLE AND FIGURES LEGENDS

Table 1. Summary of splicing events.

Figure 1. Global circadian gene expression. (A) Schematic of experimental conditions. (B) Phase sorted heatmap of global gene expression from the circadian timecourse experiment. (C) Phase histogram of global gene expression. RT-qPCR validation (black, solid line) and RNAseq traces (gray, dotted line) of the core clock genes (D) CCA1, (E) LHY, (F) RVE8, (G) PRR9, (H) PRR7, (I) PRR5, (J) TOC1/PRR1, (K) ELF3. Expression values are normalised to maximum. Error bars represent SEM.

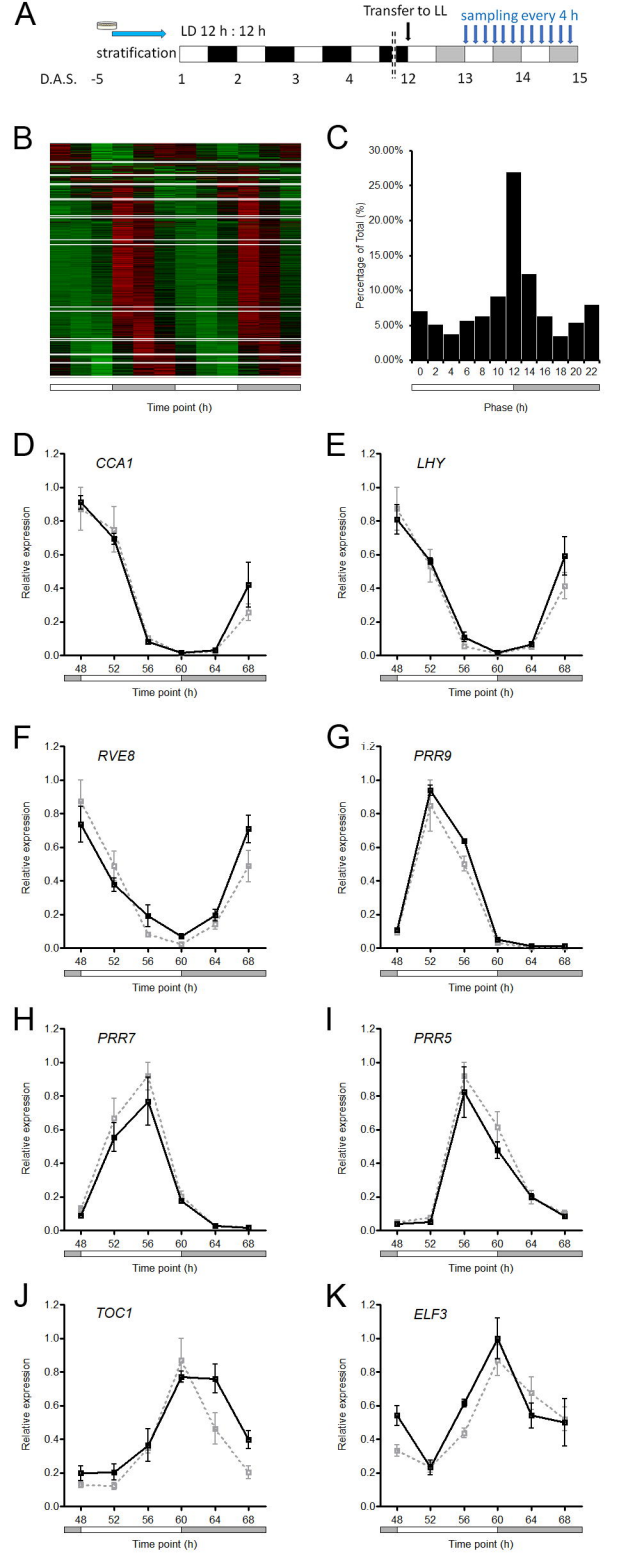
Figure 2. Novel circadian genes and specific gene families. (A) Venn diagram showing circadian gene expression compared to the Covington & Edwards combined dataset (DS1) (33) and the Hsu & Harmer dataset (DS2) (34). (B) Selected subset of significant Biological Process (BP) GO Terms of novel circadian genes. (C-H) Phase sorted heatmap of gene expression of (C) CCT domain containing genes, (D) bHLH Transcription Factor genes, (E) Small Auxin Response genes, (F) bZIP Transcription Factor genes, (G) BBX genes, and (H) pre-miRNA genes.

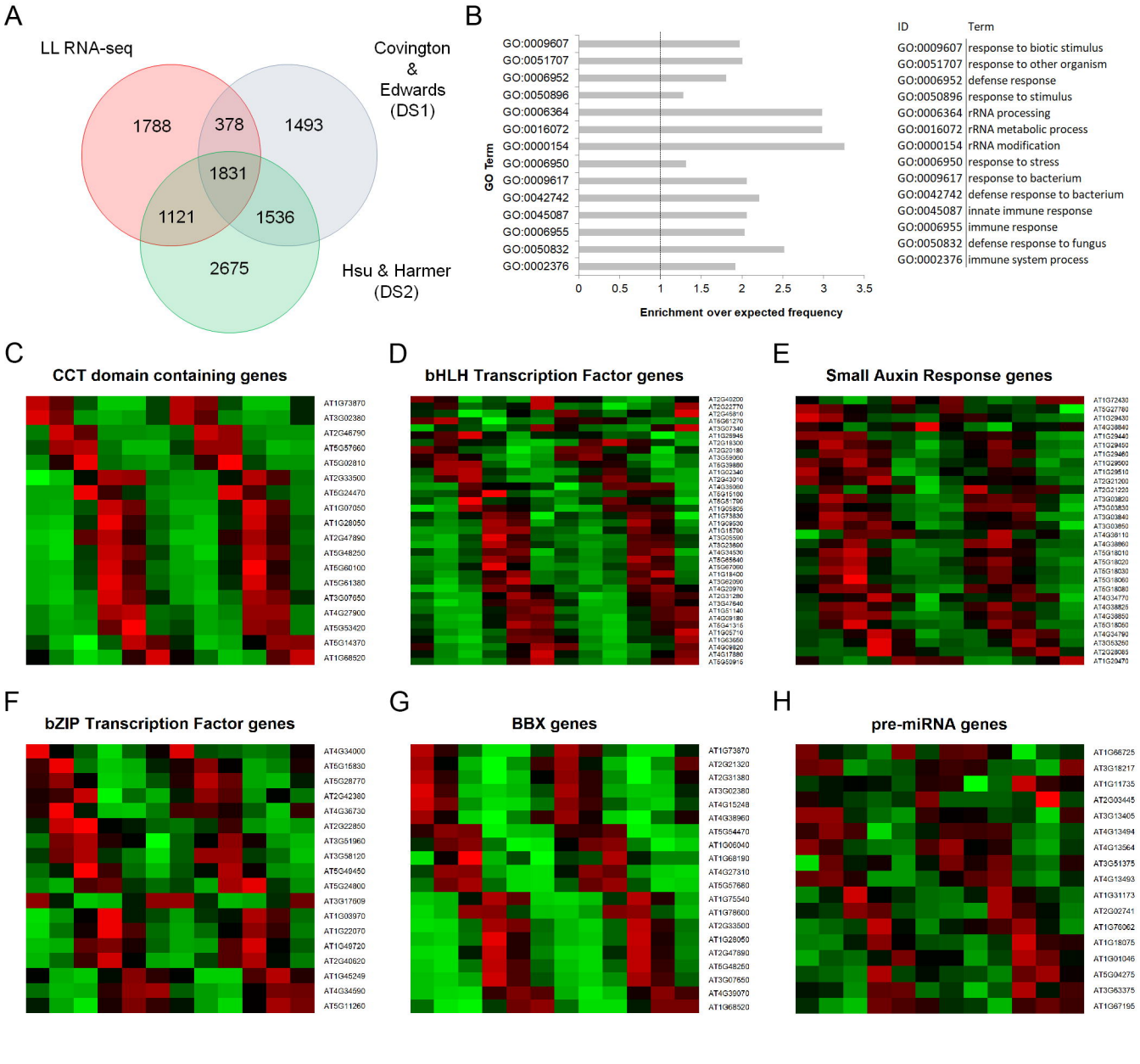
Figure 3. Analysis of Clock Controlled Elements (CCEs). (A) Distribution of clock controlled annotated events by type (alt 3'ss, alt 5'ss, ES and IR). (B) Distribution of novel clock controlled exonic and intronic events. (C-H) Phase sorted heatmap of clock controlled splicing event bins by type: (C) Alt 3'ss events, (D) Alt 5'ss events, (E) Exon Skipping events, (F) Intron Retention events, (G) Exon novel events, and (H) Intron novel events.

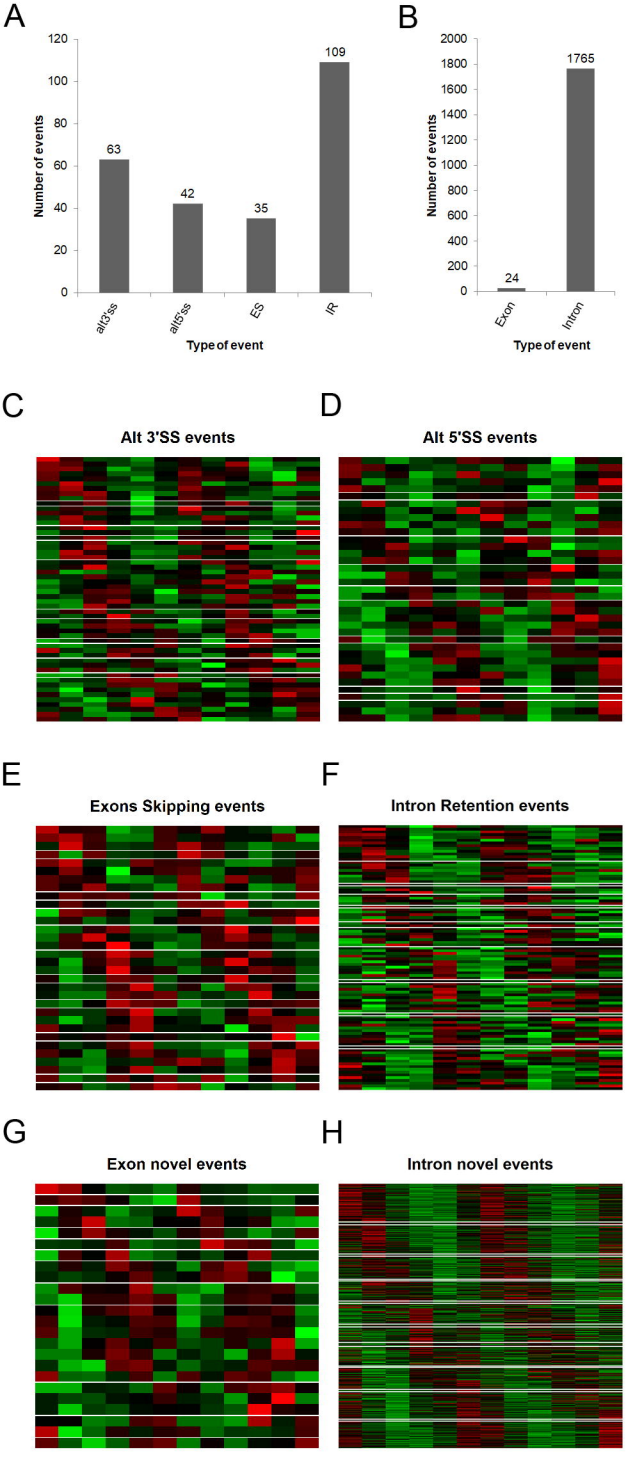
Figure 4. Circadian regulated splicing factors and SPF30. (A) Venn diagram showing the comparison between robust cycling genes (those that cycle on the RNAseq dataset of this work and also on DS1 and DS2) and splicing related genes. (B-D) SPF30 (AT2G02570) expression pattern from: (B) Mockler Diurnal web server, LL23_LDHH dataset (50), (C) Normalised counts from the circadian RNAseq (this work), and (D) qPCR data (n=3 biological replicates per timepoint). Expression values are normalised to maximum. Error bars represent SEM.

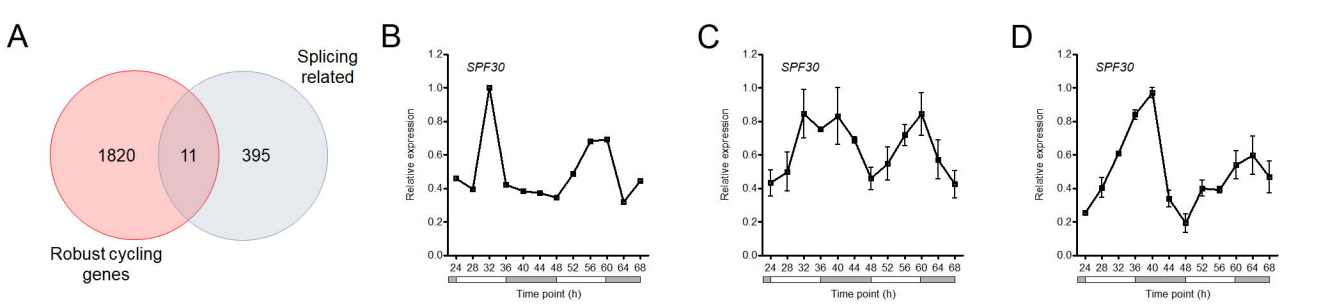
Figure 5. SPF30 Physiology. (A) Gene model of SPF30 (AT2G02570). White triangles indicate the the T-DNA insertion sites for the mutant alleles *atspf30-1* (SALK_052016C) and *atspf30-2* (SALK_081292C) . Red and green arrows indicate primer binding sites. Representative images of wild type (Col-0), atspf30-1 and atspf30-2 (from left to right) are shown. (B) Graph showing circadian leaf angle plots for Col-0 (black) and atspf30-1 (blue), after transfer to LL for 144 h (6 days). These experiments were repeated at least three times with similar results. No differences in period were found among genotypes, as estimated by Fast Fourier Transform-Non-Linear Least Squares (FFT-NLLS) algorithm with BRASS 3.0. software. (C) Flowering time measured as the number of rosette leaves at 1 cm bolting in LL, SD and LD photoperiods (n = 3 independent experiments for LL and LD, and n = 4 independent experiments for SD). Error bars represent SEM. *$p<0.05$ (ANOVA, followed by Dunnet's *post hoc* test against Col-0).

Figure 6. CCEs misregulated in atspf30-1. (A) Venn diagram showing the comparison between all clock-controlled elements (CCEs) and atspf30-1 affected splicing events. (B) Distribution of annotated events (56 out of 688 events) by type (alt 3'ss, alt 5'ss, ES and IR). (C) Phase histogram of CCEs and atspf30-1 affected splicing events. *$p<0.015$ N-1 Chi-squared test. (D) Read density histogram of RNAseq data comparing Col-0 (blue) and atspf30-1 (black) reads for the AT5G17670 gene locus region. The red rectangle area depicts the read density zone for intron 001. The AT5G17670 gene model is shown at the bottom of the graph. (E) PIR plot of the AT5G17670:001 event. (F) qPCR of AT5G17670:001 in Col-0 (black) and atspf30-1 (blue). Expression values are normalised to maximum. Error bars represent SEM.

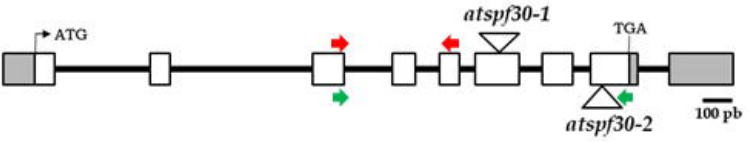

A
LD 12 h : 12 h
stratification
Transfer to LL
sampling every 4 h
D.A.S. -5 1 2 3 4 12 13 14 15
B
Time point (h)
C
Percentage of Total (%)
Phase (h)
D
CCA1
E
LHY
F
RVE8
G
PRR9
H
PRR7
I
PRR5
J
TOC1
K
ELF3
Relative expression
Time point (h)


A
LL RNA-seq
Covington & Edwards (DS1)
Hsu & Harmer (DS2)
1788
378
1493
1831
1121
1536
2675
B
GO Term
GO:0009607
GO:0051707
GO:0006952
GO:0050896
GO:0006364
GO:0016072
GO:0000154
GO:0006950
GO:0009617
GO:0042742
GO:0045087
GO:0006955
GO:0050832
GO:0002376
0
0.5
1
1.5
2
2.5
3
3.5
Enrichment over expected frequency
ID
Term
GO:0009607 response to biotic stimulus
GO:0051707 response to other organism
GO:0006952 defense response
GO:0050896 response to stimulus
GO:0006364 rRNA processing
GO:0016072 rRNA metabolic process
GO:0000154 rRNA modification
GO:0006950 response to stress
GO:0009617 response to bacterium
GO:0042742 defense response to bacterium
GO:0045087 innate immune response
GO:0006955 immune response
GO:0050832 defense response to fungus
GO:0002376 immune system process
C
CCT domain containing genes
D
bHLH Transcription Factor genes
E
Small Auxin Response genes
F
bZIP Transcription Factor genes
G
BBX genes
H
pre-miRNA genes


A
Number of events
120
100
80
60
40
20
0
63
42
35
109
alt3'ss
alt5'ss
ES
IR
Type of event
B
Number of events
2000
1800
1600
1400
1200
1000
800
600
400
200
0
24
1765
Exon
Intron
Type of event
C
Alt 3'SS events
D
Alt 5'SS events
E
Exons Skipping events
F
Intron Retention events
G
Exon novel events
H
Intron novel events

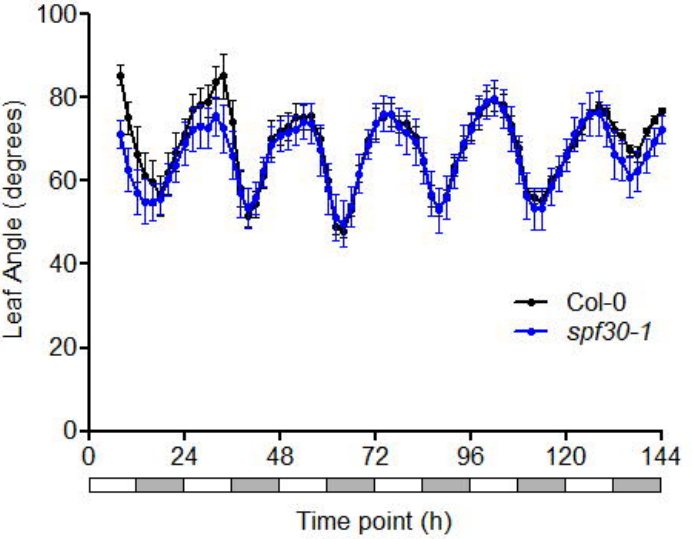

A
Splicing related
1820
11
395
Robust cycling genes
B
SPF30
Relative expression
Time point (h)
C
SPF30
Relative expression
Time point (h)
D
SPF30
Relative expression
Time point (h)

A

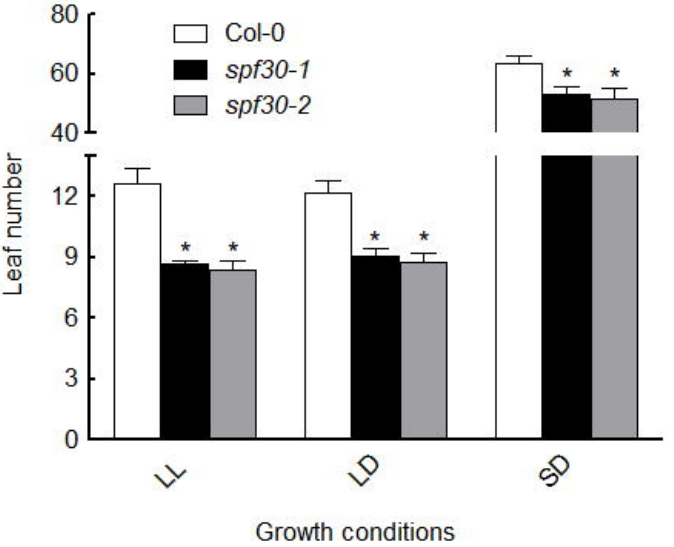

ATG
atspf30-1
TGA
atspf30-2
100 pb

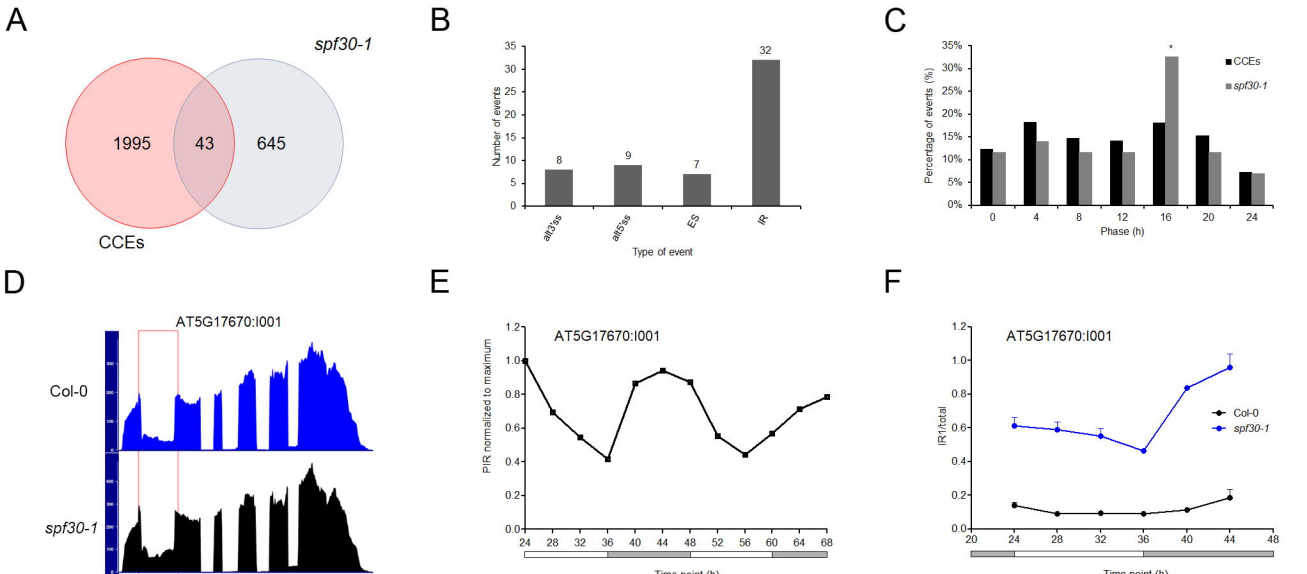
Col-0
atspf30-1
atspf30-2

B


Leaf Angle (degrees)
100
80
60
40
20
0
0
24
48
72
96
120
144
Time point (h)
Col-0
spf30-1

C


Leaf number
80
60
40
12
9
6
3
0
Col-0
spf30-1
spf30-2
*
LL
LD
SD
Growth conditions


A
spf30-1
1995
43
645
CCEs
B
Number of events
35
30
25
20
15
10
5
0
8
9
7
32
alt3'ss
alt5'ss
ES
IR
Type of event
C
Percentage of events (%)
35%
30%
25%
20%
15%
10%
5%
0%
CCEs
spf30-1
*
0
4
8
12
16
20
24
Phase (h)
D
AT5G17670:I001
Col-0
spf30-1
E
AT5G17670:I001
PIR normalized to maximum
1.2
1.0
0.8
0.6
0.4
0.2
0.0
24
28
32
36
40
44
48
52
56
60
64
68
Time point (h)
F
AT5G17670:I001
IR1/total
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Col-0
spf30-1
20
24
28
32
36
40
44
48
Time point (h)