



# A Role for Pre-mRNA-PROCESSING PROTEIN 40C in the Control of Growth, Development, and Stress Tolerance in *Arabidopsis thaliana*

## OPEN ACCESS

### Edited by:

Maria Kalyna,  
University of Natural Resources and  
Life Sciences Vienna, Austria

### Reviewed by:

Marjori Matzke,  
Academia Sinica, Taiwan  
Federico Damian Ariel,  
Instituto de Agrobiotecnología del  
Litoral (IAL), Argentina

### \*Correspondence:

Marcelo Javier Yanovsky  
mjyanovsky@gmail.com

<sup>†</sup>These authors contributed equally  
to this work.

<sup>‡</sup>These authors have contributed  
equally to this work as second authors.

### Specialty section:

This article was submitted to  
Plant Physiology,  
a section of the journal  
Frontiers in Plant Science

**Received:** 13 May 2019

**Accepted:** 22 July 2019

**Published:** 13 August 2019

### Citation:

Hernando CE, García Hourquet M,  
de Leone MJ, Careno D, Iserte J,  
Mora Garcia S and Yanovsky MJ  
(2019) A Role for Pre-mRNA-  
PROCESSING PROTEIN 40C in the  
Control of Growth, Development,  
and Stress Tolerance in  
*Arabidopsis thaliana*.  
Front. Plant Sci. 10:1019.  
doi: 10.3389/fpls.2019.01019

Carlos Esteban Hernando<sup>†</sup>, Mariano García Hourquet<sup>†</sup>, María José de Leone<sup>‡</sup>,  
Daniel Careno<sup>‡</sup>, Javier Iserte, Santiago Mora Garcia and Marcelo Javier Yanovsky\*

Comparative Genomics of Plant Development Laboratory, Instituto de Investigaciones Bioquímicas de Buenos Aires–Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina, Fundación Instituto Leloir, Buenos Aires, Argentina

Because of their sessile nature, plants have adopted varied strategies for growing and reproducing in an ever-changing environment. Control of mRNA levels and pre-mRNA alternative splicing are key regulatory layers that contribute to adjust and synchronize plant growth and development with environmental changes. Transcription and alternative splicing are thought to be tightly linked and coordinated, at least in part, through a network of transcriptional and splicing regulatory factors that interact with the carboxyl-terminal domain (CTD) of the largest subunit of RNA polymerase II. One of the proteins that has been shown to play such a role in yeast and mammals is pre-mRNA-PROCESSING PROTEIN 40 (PRP40, also known as CA150, or TCERG1). In plants, members of the PRP40 family have been identified and shown to interact with the CTD of RNA Pol II, but their biological functions remain unknown. Here, we studied the role of AtPRP40C, in *Arabidopsis thaliana* growth, development and stress tolerance, as well as its impact on the global regulation of gene expression programs. We found that the *prp40c* knockout mutants display a late-flowering phenotype under long day conditions, associated with minor alterations in red light signaling. An RNA-seq based transcriptome analysis revealed differentially expressed genes related to biotic stress responses and also differentially expressed as well as differentially spliced genes associated with abiotic stress responses. Indeed, the characterization of stress responses in *prp40c* mutants revealed an increased sensitivity to salt stress and an enhanced tolerance to *Pseudomonas syringae* pv. *maculicola* (*Psm*) infections. This constitutes the most thorough analysis of the transcriptome of a *prp40* mutant in any organism, as well as the first characterization of the molecular and physiological roles of a member of the PRP40 protein family in plants. Our results suggest that PRP40C is an important factor linking the regulation of gene expression programs to the modulation of plant growth, development, and stress responses.

**Keywords:** pre-mRNA processing protein 40, splicing, transcription, *Arabidopsis*, biotic stress, abiotic stress

## INTRODUCTION

Plants are permanently subjected to environmental changes that adversely affect their growth and development either in natural or agricultural settings. Accurate control of gene expression and precursor mRNAs (pre-mRNAs) processing is essential for the adaptation to rapid changes in the environment. Synthesis of mRNA by the RNA polymerase II (RNA Pol II) is coordinated with subsequent RNA processing events such as 5' capping, splicing, cleavage, and polyadenylation (Darnell, 2013). The largest subunit of the RNA Pol II couples transcription and pre-mRNA processing *via* its C-terminal domain (CTD). The CTD comprises tandem Tyr1–Ser2–Pro3–Thr4–Ser5–Pro6–Ser7 heptapeptide repeats that are highly conserved in eukaryotes (Allison et al., 1988; Nawrath et al., 1990) and interacts with capping, splicing, and polyadenylation factors, acting as a scaffolding platform for mRNA processing factors (Greenleaf, 1993; McCracken et al., 1997; Misteli and Spector, 1999; Buratowski, 2003; Phatnani and Greenleaf, 2006; Muñoz et al., 2010; Braunschweig et al., 2013).

Pre-mRNA splicing occurs in two sequential transesterification steps catalyzed by the spliceosome, a large dynamic ribonucleoprotein complex present in the nucleus. The major spliceosome consists of five small nuclear ribonucleoproteins (snRNPs) named U1, U2, U4, U5, and U6, along with ~200 accessory proteins (Wahl et al., 2009). The splicing reaction begins with the initial recognition of the 5' and 3' splice sites (5'SS and 3'SS, respectively) at the exon–intron boundaries, the branch point sequence (BPS), and the polypyrimidine tract (Py) (Wahl et al., 2009; Meyer et al., 2015); the completion of the splicing process involves the removal of an intron and the ligation of the resulting exons.

Splicing can be classified as constitutive splicing (CS) or alternative splicing (AS), depending on splice site choice and usage. In constitutive splicing, canonical splice sites are always used for a given transcript. In contrast, alternative splicing involves variable splice site choice and/or usage, giving place to multiple mRNAs variants from a single gene. Splice sites can be strong or weak depending on the degree in which their sequences diverge from consensus sequences, and this determines their affinities for the spliceosomal machinery. In general, strong splice sites lead to constitutive splicing through full usage of the site. On the other side, weak splice sites are usually associated with alternative splicing, and the frequency of usage of the alternative splice sites varies depending on cellular context and environmental conditions (Kornblihtt et al., 2013). Alternative splicing events comprise intron retention (IR), exon skipping (ES), and alternative 5' donor and 3' acceptor splice site selection (Alt 5'SS and Alt 3'SS, respectively).

In plants of the model species *Arabidopsis thaliana*, current estimates indicate that ~61% of the intron-containing genes undergo alternative splicing (Marquez et al., 2012), which proved to be essential for proper growth and development, as well as for optimal responses to environmental changes (Reddy et al., 2013; Staiger and Brown, 2013; Ding et al., 2014; Kwon et al., 2014; Shikata et al., 2014; Filichkin et al., 2015; Deng and Cao, 2017). One of the earliest events in spliceosome assembly is the formation of an RNA duplex between the 5' SS and the 5'

end of the U1 snRNP, an event involving the activity of multiple proteins that stabilize the complex. PRE-mRNA-PROCESSING PROTEIN 40 (PRP40) was first discovered in yeast and proved to be essential in the early steps of spliceosome complex formation (Kao and Siliciano, 1996). PRP40 harbors two WW domains in the amino terminus and four FF domain repeats in the carboxyl terminus, which have been well characterized as mediators of protein–protein interactions (Bedford and Leder, 1999; Macias et al., 2002). In yeast, as well as in mammals, PRP40 helps to define the bridging interaction that links both ends of the intron, by interacting simultaneously with the BRANCHPOINT BINDING PROTEIN (BBP) and the U1 snRNP (Abovich and Rosbash, 1997). *Saccharomyces cerevisiae* PRP40 (ScPRP40) and its three human homologues, HsPRPF40A, HsPRPF40B, and HsTCERG1/CA150, have been extensively studied in the last two decades. While HsPRPF40A and HsPRPF40B were characterized as essential components of the early spliceosome assembly process (Lin et al., 2004; Wahl et al., 2009; Becerra et al., 2015), HsTCERG1/CA150 was initially discovered as a transcriptional modulator and later linked to pre-mRNA splicing (Sune et al., 1997; Pearson et al., 2008; Munoz-Cobo et al., 2017). In fact, HsTCERG1/CA150 has been found associated with elongation factors and is present in a complex with the RNA polymerase II *via* the FF domains (Sune et al., 1997; Carty et al., 2000; Sanchez-Alvarez et al., 2006). HsTCERG1/CA150 has also been linked to RNA splicing through its WW domain 2 (WW2), which interacts with the splicing factors SF1, U2AF, and components of the SF3 complex (Goldstrohm et al., 2001; Lin et al., 2004). Furthermore, HsTCERG1/CA150 has been identified in highly purified spliceosomes in multiple studies (Makarov et al., 2002; Rappsilber et al., 2002; Deckert et al., 2006). As previously mentioned, the processes of transcription and pre-mRNA processing are coordinated by the CTD of the RNA pol II. The modular structure of HsTCERG1/CA150, containing the splicing-factor associating WW domains present in the N-terminus and the CTD-associating FF repeats in the C terminus, confers the ideal structure for a protein that couples transcription and splicing. In accordance with this model, both halves of HsTCERG1/CA150 have been shown to be essential for the assembly of higher-order transcription-splicing complexes (Sanchez-Alvarez et al., 2006).

Plant homologues of ScPRP40 were first identified in 2009 by Kang et al. by means of a bioinformatic screening of proteins that interact with the RNA Pol II CTD in *Arabidopsis*. Three proteins were thus identified: AtPRP40A, AtPRP40B, and AtPRP40C (Kang et al., 2009), which interacted with de RNA Pol II CTD (both phosphorylated and nonphosphorylated); at least for AtPRP40B, the authors showed that the WW domains at the amino terminus mediated this interaction (Kang et al., 2009). So far, a characterization of the biological and molecular roles of plant PRP40s is missing. Here, we decided to focus on the characterization of AtPRP40C using a reverse genetics approach and performed a physiological and molecular analysis of the function of this protein in the control of growth, development and stress tolerance, as well as in the regulation of gene expression. We found that AtPRP40C is involved in the regulation of flowering time and photomorphogenic responses. In addition, an RNA-seq

analysis revealed that AtPRP40C is associated with the proper control of expression and splicing of abiotic and biotic stress-related transcripts, and, indeed, physiological assays showed that *prp40c* mutants display altered tolerance to salt stress and *Pseudomonas syringae* pv. *maculicola* (*Psm*) infections. Our work is the first physiological and molecular characterization of a member of the PRP40 protein family in plants, which reveals an important role for PRP40C linking the regulation of gene expression and pre-mRNA splicing, and modulating plant growth, development, and stress responses.

## MATERIALS AND METHODS

### Plant Material

All of the *Arabidopsis* lines used in this study were on the Columbia (Col-0) ecotype background, and the *prp40c-1* (SALK\_148319) and *prp40c-2* (SALK-205357) mutants were obtained from the Arabidopsis Biological Research Center (ABRC). Genotypic verification of the mutants was performed via PCR analysis using primers as detailed in **Supplementary Table S1**. PRP40C expression levels in the mutant alleles, as well as in wild-type plants, were assessed by reverse transcription followed by a semiquantitative PCR (RT-PCR) using Actin 2 as expression control; primers used are described in the **Supplementary Table S1**.

### Growth Conditions

Plants were grown on soil at 22°C under long days (LD; 16-h light/8-h dark cycles; 80  $\mu\text{mol m}^{-2} \text{s}^{-1}$  of white light), 12:12 days (LD 12:12; 12-h light/12-h dark cycles; 80  $\mu\text{mol m}^{-2} \text{s}^{-1}$  of white light), short days (SD; 8-h light/16-h dark cycles; 140  $\mu\text{mol m}^{-2} \text{s}^{-1}$  of white light), or continuous light (LL; 50  $\mu\text{mol m}^{-2} \text{s}^{-1}$  of white light), depending on the experiment.

### Phylogenetic Analysis

Evolutionary analysis was conducted using the maximum likelihood method implemented in MEGA X (Kumar et al., 2018). Sequences were retrieved using TBLASTN with different query sequences from Phytozome (phytozome.jgi.doe.gov), Onekp project at China National GeneBank (db.cngb.org/blast/blast/tblastn/), www.fernbase.org, and the *Klebsormidium nitens* genome portal (<http://www.plantmorphogenesis.bio.titech.ac.jp/>). All sequences were manually inspected to search for annotation errors. Retrieved sequences with long stretches of indeterminate bases were excluded from the analysis. In the case of sequences derived from transcriptomic surveys (namely, Onekp), care was taken so that only species with the longest contigs for all the genes searched were included in the analysis.

### Flowering Time Analysis

For flowering time experiments, the plants were grown on soil at 22°C under standard long days (LD; 16-h light/8-h dark cycles; 80  $\mu\text{mol m}^{-2} \text{s}^{-1}$  of white light), 12:12 days (12-h light/12-h dark cycles; 80  $\mu\text{mol m}^{-2} \text{s}^{-1}$  of white light), continuous light (LL;

50  $\mu\text{mol m}^{-2} \text{s}^{-1}$  of white light) or short days (SD; 8-h light/16-h dark cycles; 140  $\mu\text{mol m}^{-2} \text{s}^{-1}$  of white light) depending on the experiment. Flowering time was estimated by counting the number of rosette leaves at the time of bolting. These experiments were performed in triplicate with  $n = 16$  for each genotype. The statistical analysis was done using a two-tailed Student's *t*-test.

### Hypocotyl Length Characterization

For hypocotyl length measurements, seedlings were grown on 0.8% agar under complete darkness, continuous red light (1  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ), continuous blue light (1  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ), continuous white light (LL), or under cycles of white light in SD or LD (all white light treatments 1  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ). The final length of the hypocotyls was measured 4 days after germination. Light effects on hypocotyl elongation under continuous red and blue light were calculated normalizing hypocotyl length under each light regime to the hypocotyl length of the same genotype under constant dark conditions. For the white light experiments, absolute values of hypocotyl elongation are shown. These experiments were performed in triplicate with  $n = 20$  for each genotype. The statistical analysis was done using a two-tailed Student's *t*-test.

### Circadian Leaf Movement Analysis

For leaf movement analysis, plants were grown under 16-h light/8-h dark cycles until the appearance of the first pair of leaves. This period is referred to as the entrainment period. In order to measure circadian rhythms in leaf movement, plants were transferred to continuous white light (20  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ) at 22°C. The position of the first pair of leaves was recorded every 2 h for 5–6 days using digital cameras, and the leaf angle was determined using ImageJ software (Schneider et al., 2012). Period estimates were calculated with Brass 3.0 software (Biological Rhythms Analysis Software System, available from <http://www.amillar.org>) and analyzed with fast Fourier transform nonlinear least squares (FFT-NLLS) using Brass 3.0 software. These experiments were performed in triplicate, with  $n = 8$  for each genotype. The statistical analysis was done using a two-tailed Student's *t*-test.

### Growth Conditions and Protocol Used for cDNA Library Preparation and High-Throughput Sequencing

Three biological replicates with seeds of wild-type (Col-0) and *prp40c-1* mutant allele were sown onto Murashige and Skoog medium containing 0.8% agarose, stratified for 4 days in the dark at 4°C and then grown at 22°C in continuous light. Whole plants were harvested after 12 days, and total RNA was extracted with RNeasy Plant Mini Kit (QIAGEN) following the manufacturer's protocols. To estimate the concentration and quality of samples, NanoDrop 2000c (Thermo Scientific) and the Agilent 2100 Bioanalyzer (Agilent Technologies) with the Agilent RNA 6000 Nano Kit were used, respectively. Libraries were prepared following the TruSeq RNA Sample Preparation Guide (Illumina). Briefly, 3  $\mu\text{g}$  of total RNA was polyA-purified and fragmented, first-strand cDNA synthesized by

reverse transcriptase (SuperScript II, Invitrogen) using 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 DNA1000 kit (Agilent Technologies). Samples were pooled to create six multiplexed DNA libraries, which were pair-end sequenced with an Illumina HiSeq 1500 at INDEAR Argentina, providing 100-bp pair-end reads. Three replicates for each genotype were sequenced. Sequencing data have been uploaded to the Gene Expression Omnibus database and are available under accession number (GSE129932).

### Processing of RNA Sequencing Reads

Sequence reads were mapped to *A. thaliana* TAIR10 (Lamesch et al., 2012) genome using TopHat v2.1.1 (Trapnell et al., 2009) with default parameters, except of maximum intron length set at 5,000. Count tables for different feature levels were obtained from bam files using custom R scripts and considering TAIR10 transcriptome.

### Differential Gene Expression Analysis

Before differential expression analysis, we decided to discard genes with fewer than 10 reads on average per condition. Differential gene expression was estimated using the edgeR package version 3.4.2 (Robinson et al., 2010), and resulting *p* values were adjusted using a false discovery rate (FDR) criterion (Benjamini and Hochberg, 1995). Genes with FDR values <0.05 and absolute log two-fold change > 0.58 were deemed differentially expressed. Overlapping analysis were performed using Venny (Oliveros, 2007).

### Differential Alternative Splicing

For the analysis of alternative splicing, the transcriptome was partitioned into subgenic joint features called "bins," as proposed on DEXseq (Anders et al., 2012). Because of our special interest in new intron retention events, not only exons but also introns were considered in our analysis. The transcriptome was partitioned into 281,321 bins; 152,631 corresponding exclusively to exonic regions, 120,717 to intronic regions, and 7,973 to DNA regions directly associated with alternatively spliced events. We labeled these three kinds of bins as exon-bin, intron-bin, or AS-bins, respectively. In addition, AS-bins were further classified as defined by Mancini et al. (2019). For our analysis, we discarded bins from monoexonic genes and with mean count values lower than five reads per condition. We used edgeR exact test for the identification of differential use of bins corresponding to AS events or introns and FDR-corrected *p* values. We also computed read densities to have a relationship between the bin and its corresponding gene. Only genes with read densities > 0.05 in all genotypes were used for the analysis. AS events as well as all introns with an absolute log<sub>2</sub> fold change (bin read density in the mutant/bin

read density in wild type) value >0.58, with FDR values <0.15 were deemed differentially spliced. Overlapping analysis was performed using Venny (Oliveros, 2007).

### Functional Category Enrichment Analysis

Functional categories associated with specific groups of genes were identified using the BioMaps tool from the virtual plant software (<http://virtualplant.bio.nyu.edu/cgi-bin/vpweb>). This tool allowed us to determine which functional categories were statistically overrepresented in particular lists of genes compared to the entire genome (Katari et al., 2010). We analyzed 14 functional categories of our interest, and for each one, we determined the genes in common with our data sets, finally calculating a representation factor and the probability of finding an overlap simply by chance. The representation factor is the number of overlapping genes divided by the expected number of overlapping genes drawn from two independent groups. A representation factor > 1 indicates more overlap than expected by chance for two independent groups of genes or events, a representation factor <1 indicates less overlap than expected. The probability of each overlapping was determined using the hypergeometric probability formula.

### Analysis of Splice-Site Sequences

To evaluate possible changes in the splice-site sequences of the most significantly affected splicing events in the *prp40c-1* mutants, we obtained the donor and acceptor splice site sequences of all the intron retention events that were deemed differentially spliced (absolute log<sub>2</sub> fold value >0.58, with FDR values <0.15) in *prp40c-1* mutants compared to wild-type plants and compared them to the consensus splice-site sequences of the total 30,142 introns sequenced in our RNA-seq experiment. The frequency of each nucleotide for each position was obtained using custom R scripts and was represented using the R package Seqlogo (Bembom, 2014). The over- or underrepresentation of a particular nucleotide relative to its genome-wide frequency was determined, and a *p* value for the analysis was obtained using the hypergeometric test. The custom R scripts used here are available upon request.

### Salt and Osmotic Stress Assays

For germination under salt stress assays, 40 seeds of wild-type (Col-0), *prp40c-1*, and *prp40c-2* plants, and three replicates of each, were sown on 3-mm filter paper (Whatman) and were imbibed in different concentrations of sodium chloride (NaCl) solution (30, 50, 70, and 90 mM) or distilled water. After sowing, the plates were irradiated with far-red light and stratified overnight. On the next day, the plates were irradiated with white light for 1 h and then transferred to dark conditions for 3 days at 22°C. A seed was considered as germinated when the embryonary root protruded through the seed envelope. These experiments were performed in triplicate. The statistical analysis was done using a two-way ANOVA. For growth and survival assays, seeds were germinated on 1/2 MS agar medium. In the case of the root growth assay, 4-day-old seedlings were transferred to 1/2 MS agar containing 100 mM of NaCl or

200 mM of D-mannitol (isosmotic concentrations) and grown vertically for 5 days. The plants were photographed after this period of growth, and their root length was assessed. These experiments were performed in triplicate with  $n = 30$  for each genotype. For the NaCl tolerance assay, 4-day-old seedlings were transferred from the germination medium to 1/2 MS agar containing 0 or 160 mM of NaCl, and the survival rate was determined 30 days after the seedlings were transferred. These experiments were performed in triplicate with  $n = 50$  for each genotype. The statistical analysis was done using a two-tailed Student's *t*-test.

## Evaluation of Cold Tolerance

Seeds were sown onto 1/2 MS agar medium, stratified for 3 days at 4°C, and grown at 22°C under continuous light. Four-day-old seedlings were transferred to 1/2 MS agar plates and grown vertically for 5 days at 22°C (control) or for 10 days at 10°C under continuous light. The plants were photographed after this period of growth, and their root length was assessed. These experiments were performed in triplicate with  $n = 30$  for each genotype.

## Pathogen Infection Assay

*Pseudomonas syringae* pv. *maculicola* ES4326 strains were grown at 28°C with King's B medium (20 g protease peptone, 1.5 g K<sub>2</sub>HPO<sub>4</sub>, 6.09 ml 1 M MgSO<sub>4</sub>, and 10 g glycerol per liter) supplemented with rifampicin (100mg/L) and kanamycin (50mg/L) for selection. Freshly cultured bacteria were collected and resuspended to a final concentration of OD<sub>600</sub> = 0.0002 in 10 mM MgSO<sub>4</sub>. The bacterial solution was then pressured infiltrated with a 1-ml needleless syringe into the abaxial side of the 8th to 10th leaves of 5- to 6-week-old plants grown under SD conditions. Bacterial growth assays were performed at 48 h postinfection (hpi). Leaves were surface washed in sterile water before carrying out bacterial counts. A disc was punched from each leaf, and then, the three discs corresponding to each plant were placed in 750 µl of 10 mM MgCl<sub>2</sub> and crushed to release the bacteria. The resulting solution was serially diluted and spot plated on KB plates containing the corresponding antibiotics. The plates were incubated at 28°C for 48 h before counting of the colonies. Log-transformed bacterial growth was statistically analyzed. All bacterial infection experiments were repeated at least three times. The statistical analysis was done using a two-tailed Student's *t*-test.

## PCR Alternative Splicing and qPCR Differential Expression Assessment

For alternative splicing and differential expression assessment, 12-day-old plants were grown in Murashige and Skoog 0.8% agar medium under continuous white light at 22°C. Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). One microgram of RNA was treated with RQ1 RNase-Free DNase (Promega, Madison, WI, USA) and subjected to retrotranscription with Super Script II Reverse Transcriptase (SSII RT) (Thermo Fisher Scientific, Waltham, MA, USA) and oligo-dT according to manufacturer's instructions. For alternative splicing, PCR amplification was performed using 1.5 U of Taq polymerase (Invitrogen). Primers used for amplification are detailed in

**Supplementary Table S1.** RT-PCR products were electrophoresed and detected by SYBR Green 2%. For qPCR differential expression assessment, cDNAs were then amplified with FastStart Universal SYBR Green Master (Roche, Basel, Switzerland) using the Mx3000P Real Time PCR System (Agilent Technologies, Santa Clara, CA, USA) cycler. The IPP2 (AT3G02780) transcript was used as a housekeeping gene. Quantitative RT-PCR (qRT-PCR) analysis was conducted using the 2- $\Delta\Delta$ CT method (Schmittgen and Livak, 2008). Primer sequences and conditions are available on **Supplementary Table S1.** Three biological samples were measured for each genotype. The statistical analysis was done using a two-tailed Student's *t*-test.

## RESULTS

### Phylogenetic Analysis of the Pre-mRNA Processing Protein 40 Family

We initially performed a phylogenetic analysis of the pre-mRNA processing protein 40 family (PRP40) (**Figure 1A**). We found that PRP40s from both human and *Arabidopsis* genomes cluster in two separate clades, one that includes the so-called PRP40A and B types and another that includes HsTCERG1/CA150 and AtPRP40C. In order to assess whether this clustering reflects an ancestral origin, or may be an effect of sequence affinity unrelated to descent, we retrieved and analyzed plant PRP40 sequences sampled from phylogenetically relevant species (**Figure 1B**). We found that members of the PRP40A/B clade are found in all Viridiplantae, both uni- and multicellular. However, the occurrence of A and B forms appear to be restricted to eudicots, since neither monocots nor basal angiosperms show this type of duplication. PRP40C-type sequences are also widespread among plants, usually found as single copy genes. Still, we could not find *bona fide* homologues in Chlorophytes; consequently, PRP40C genes appear to be restricted to streptophytes. Taken together, our analysis shows that 1) type A PRP40s appear to be an ancestral feature of eukaryotic organisms, 2) the occurrence of duplicated forms in this clade (types A and B) is lineage specific, and 3) genes of the HsTCERG1/PRP40C type may also have independent origins; in Viridiplantae, they appear to be an ancient innovation of streptophytes.

Thus, both AtPRP40C and HsTCERG1/CA150 appear to have diverged from the others members of the family in the respective organism. This observation, together with the fact that HsTCERG1/CA150 is a well-characterized protein that links transcription and pre-mRNA processing, suggested that this gene could have evolved to modulate gene expression networks in complex eukaryotic organisms and prompted us to analyze its biological roles and molecular functions in *A. thaliana*.

### PRP40C Mutants Display a Late Flowering Phenotype and Mild Alterations in Other Light Controlled Process

In order to characterize the role of AtPRP40C (hereinafter referred to as PRP40C) using a reverse genetics approach, we identified plants with T-DNA insertions in the *PRP40C* loci, *prp40c-1* (SALK\_148319), and *prp40c-2* (SALK\_205357). Both mutants

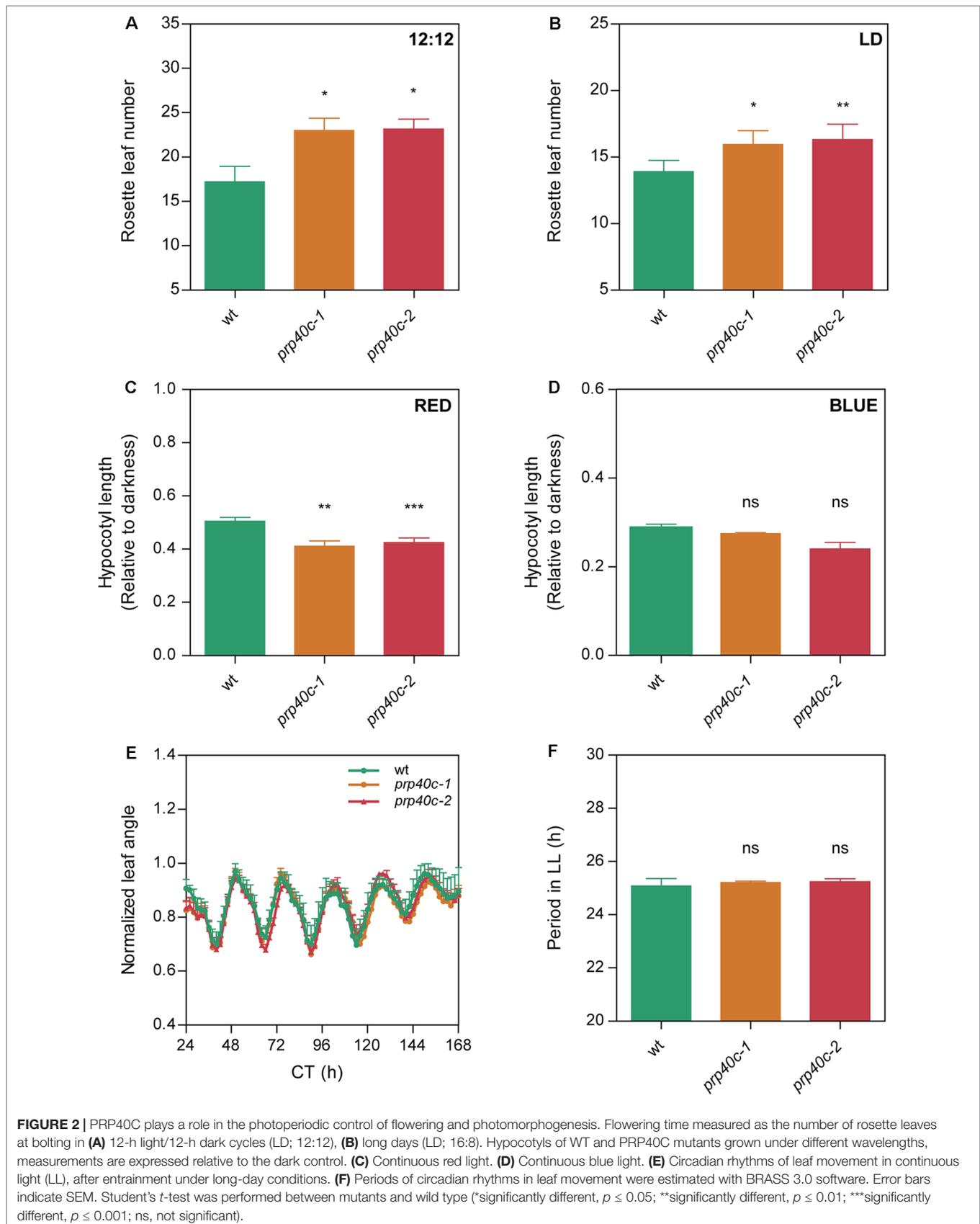


were genotyped to verify the homozygosity of the T-DNA insertion, and null expression of the wild-type mRNA was evaluated by RT-PCR in both mutant alleles. We then conducted an analysis of the effect of PRP40C in the control of several developmental and physiological processes. Initially, we did not observe severe phenotypic alterations in growth and morphology during the vegetative stage (**Supplementary Figure S1**). Then, we studied flowering time, a key developmental trait that marks the transition from the vegetative to the reproductive stage. We found that the *prp40c* mutants flower later than wild-type plants in both 12-h light/12-h dark and 16-h light/8-h dark long day photoperiods (**Figure 2A, B**). No delay in flowering time was observed in a short-day photoperiod or under continuous light (**Supplementary Figure S2**). Light perception and the circadian clock are two key factors that ensure proper flowering time regulation. To evaluate if any of these could be related to the late flowering phenotype, we analyzed photomorphogenic responses and clock function in the *prp40c* mutants. Seedling photomorphogenesis was assessed measuring the inhibition of hypocotyl elongation under different light treatments. Interestingly, *prp40c* mutants displayed shorter hypocotyls than wild-type plants under continuous red light, indicating hypersensitivity to this wavelength (**Figure 2C**). No differences were detected under continuous blue light (**Figure 2D**). We then assessed the effect of different photoperiods on hypocotyl elongation. No significant differences were detected between wild-type plants and *prp40c* mutants in any photoperiodic condition under whitelight (**Supplementary Figure S3**). Finally, we monitored circadian rhythms in leaf movements in wild-type plants as well as in *prp40c-1* and *prp40c-2* mutants. The rhythms observed for both mutant alleles were similar to those of wild-type plants, exhibiting a 24.5-h period (**Figure 2E, F**). These data suggest that the circadian function is unaffected, at least under standard growth conditions, in these plants. Collectively, our results indicate that PRP40C has a role in flowering time control, which could be associated, at least in part, with alterations in the red light signaling pathway.

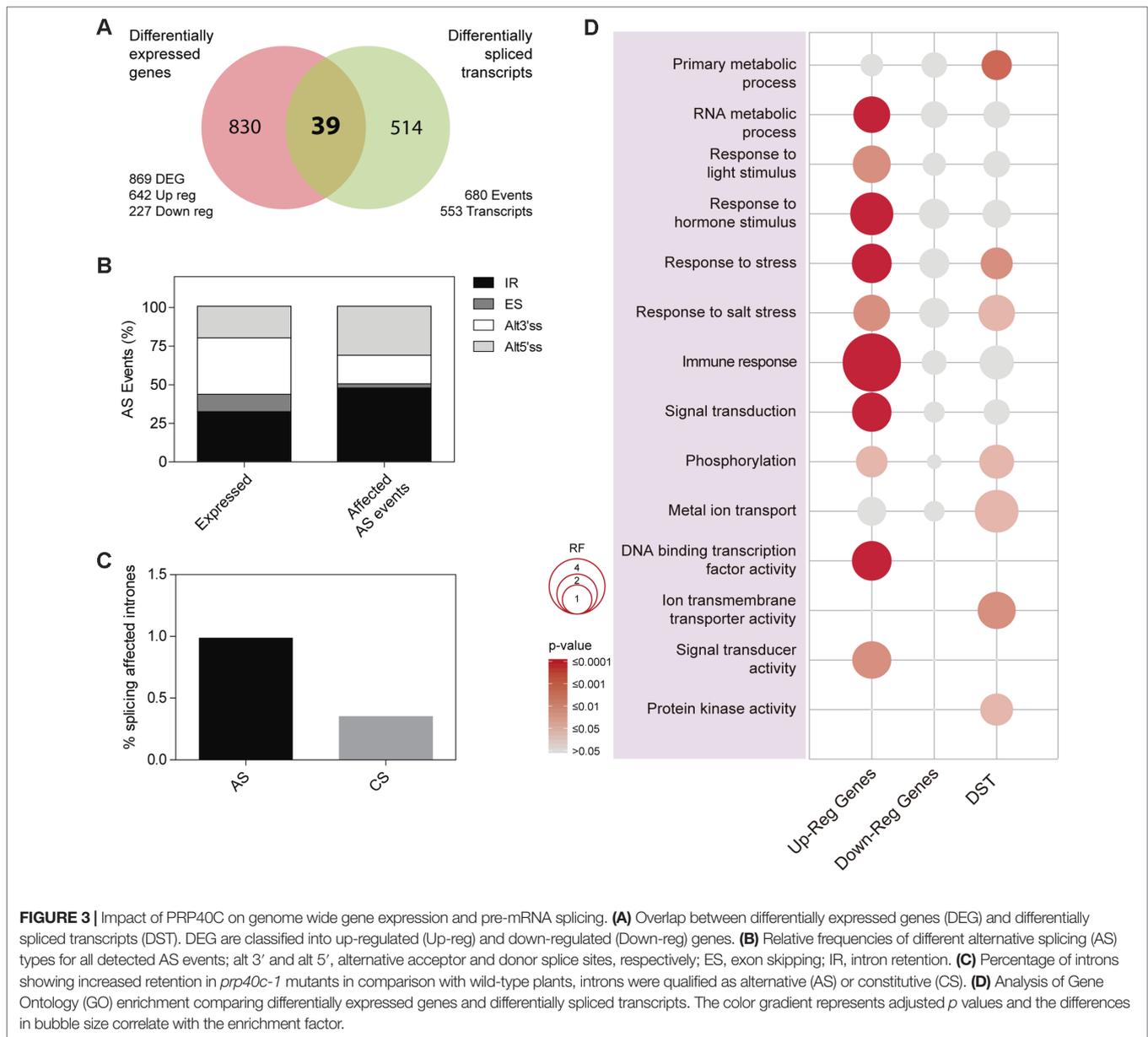
## Impact of PRP40C on Genome-Wide Gene Expression and Pre-mRNA Splicing

In order to study the regulatory impact of PRP40C on gene expression, we analyzed the transcriptome of wild-type and *prp40c-1* mutant plants grown under standard nonstressful conditions (continuous white light at 22°C) using RNA-seq. We found 869 differentially expressed genes (DEG) in *prp40c-1* mutants relative to wild-type plants, 642 overexpressed (73.9%) and 227 underexpressed (26.1%). To characterize the role of PRP40C in pre-mRNA splicing, we evaluated both constitutive (CS) as well as alternative splicing (AS) events, which included annotated or novel AS events. We found a total of 680 differentially spliced transcripts (DST) in *prp40c-1* mutants when compared to wild-type plants. These events occurred on 553 transcripts, indicating that only a few transcripts had more than one differentially spliced event. Interestingly, only 39 transcripts were differentially expressed and differentially spliced at the same time (**Figure 3A**). We next studied the abundance of the different annotated AS events among the differentially spliced transcripts. In wild-type plants, the most

abundant AS events were those associated with the use of alternative 3' splice sites (Alt 3'SS, 36.4%), followed by intron retention (IR, 32%), alternative 5' splice sites (Alt 5'SS, 20.3%), and exon skipping (ES, 11.2%). Among the annotated AS events affected in *prp40c-1*, we found an increase in the proportion of both IR events (IR, 47.4%) and alternative 5' splice sites (Alt 5'SS, 31.6%) and a decrease in alternative 3' splice sites (Alt 3'SS, 18.4%) and exon skipping events (ES, 2.6%), relative to their frequency in wild-type plants (**Figure 3B**). We also evaluated splicing of all introns present in expressed genes. Among the 105,555 introns analyzed, we detected alterations in splicing of 397 introns from which 18 were already annotated as alternatively spliced, and 379 had no previous evidence of being alternatively spliced (i.e., they were considered constitutively spliced introns). Interestingly, the proportion of increased intron retention events detected in the *prp40c-1* mutants among alternatively spliced introns was larger than the proportion observed among constitutively spliced introns (**Figure 3C**). This observation indicates that PRP40C acts as a splicing modulator rather than as an essential splicing factor. We also evaluated whether there was any change in the splice-site sequences of the intron retention events affected in *prp40c-1* mutants compared to the consensus sequence of all the introns belonging to genes expressed in our RNA-seq experiment. Interestingly, no significant deviation from the consensus sequence was observed for the donor splice site or for the acceptor splice site for the introns whose splicing was affected in *prp40c-1* mutants (**Supplementary Figure S4**). In order to get a broader landscape of the changes in the transcriptome induced by PRP40C, both the DEG and the DST were categorized into functional groups based on Gene Ontology (GO). Fourteen functional categories of our interest were examined in detail determining the representation factor for each category (data available in **Supplementary Table S2**). The representation factor (RF) is the number of overlapping genes observed in the selected group divided by the expected number of overlapping genes drawn randomly from two independent groups. Among the upregulated genes, we found a significant enrichment (representation factor > 1;  $p < 0.05$ ) for categories corresponding to RNA metabolic process, response to light stimulus, response to hormone stimulus, response to stress, response to salt stress, immune response, signal transduction, phosphorylation, sequence-specific DNA binding transcription factor activity, and signal transducer activity. For downregulated genes, no enrichment was found among the categories studied. Among the differentially spliced genes, significant enrichment was found for categories such as primary metabolic process, response to stress, response to salt stress, phosphorylation, metal ion transport, ion transmembrane transporter activity, and protein kinase activity (**Figure 3D**). None of the categories studied here displayed a statistically significant under-representation (representation factor < 1;  $p < 0.05$ ). These data taken together indicate that, instead of being a global transcription and pre-mRNA splicing regulator, PRP40C regulates transcription and splicing of a defined subset of transcripts particularly related to biotic and abiotic stress responses.



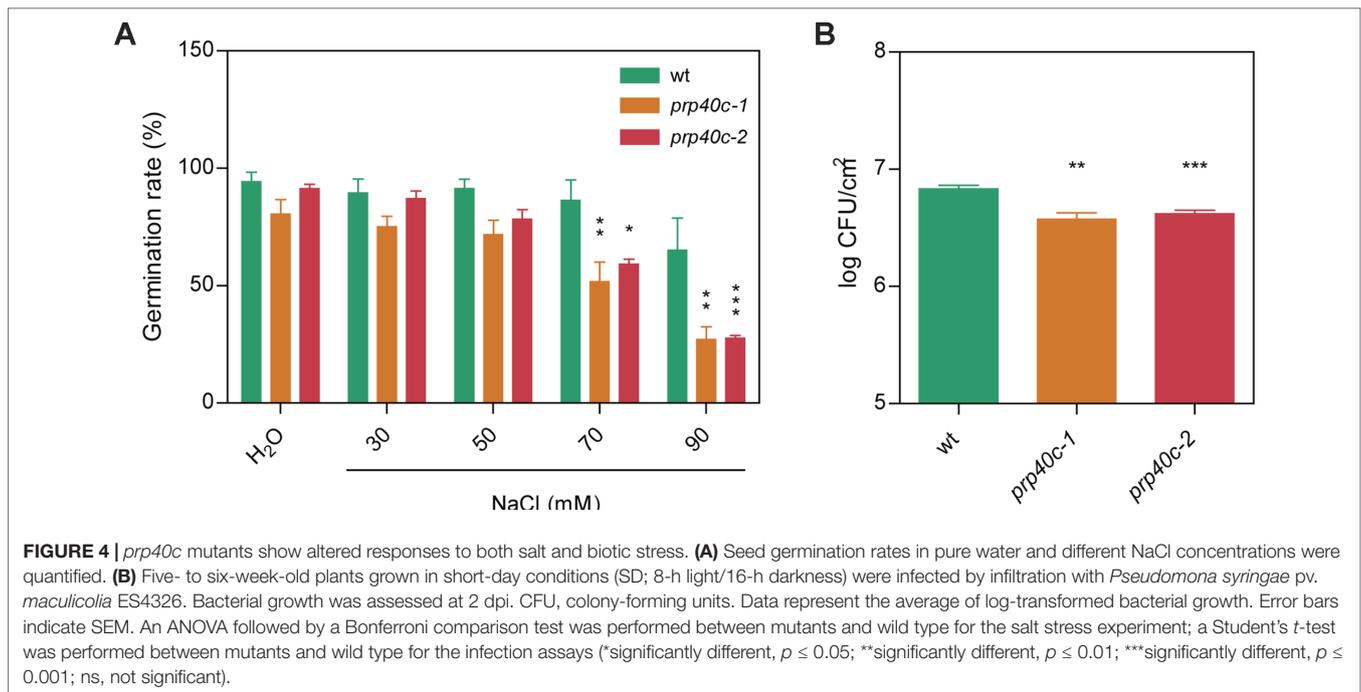
**FIGURE 2 |** PRP40C plays a role in the photoperiodic control of flowering and photomorphogenesis. Flowering time measured as the number of rosette leaves at bolting in (A) 12-h light/12-h dark cycles (LD; 12:12), (B) long days (LD; 16:8). Hypocotyls of WT and PRP40C mutants grown under different wavelengths, measurements are expressed relative to the dark control. (C) Continuous red light. (D) Continuous blue light. (E) Circadian rhythms of leaf movement in continuous light (LL), after entrainment under long-day conditions. (F) Periods of circadian rhythms in leaf movement were estimated with BRASS 3.0 software. Error bars indicate SEM. Student's *t*-test was performed between mutants and wild type (\*significantly different,  $p \leq 0.05$ ; \*\*significantly different,  $p \leq 0.01$ ; \*\*\*significantly different,  $p \leq 0.001$ ; ns, not significant).



## PRP40C Plays a Role in Both Salt and Biotic Stress Tolerance

The global transcriptome analysis revealed a significant enrichment for genes related to salt and biotic stress responses among the differentially expressed and differentially spliced transcripts in *prp40c-1* mutants. We therefore performed physiological analyses of the mutant plants under both stress conditions. We first compared germination rates between wild-type plants and *prp40c* mutants under salt stress. There was no difference in germination rates between wild-type, *prp40c-1*, and *prp40c-2* seeds germinated on filter paper embedded with water, whereas the germination rate significantly decreased in *prp40c* mutants when the seeds were placed on a sodium chloride (NaCl) solution (Figure 4A). We next examined the root growth of wild-type plants and *prp40c* mutants under salt (NaCl), mannitol, and cold stress conditions, as well

as survival rate under salt stress. No significant differences were found between genotypes in these experiments (Supplementary Figure S5). These results suggest that PRP40C is essential for proper germination under salt stress but is probably not involved in NaCl, mannitol, and cold tolerance in the early vegetative stage. In order to assess if *prp40c* mutants display alterations in response to biotic stress, we performed an infection assay with *Pseudomonas syringae* pv. *maculicola* ES4326. Leaves from 5-week-old wild-type and *prp40c* plants were pressure infiltrated with the virulent bacterium. Bacterial growth assays at 2 days postinfection (2 dpi) revealed a significant degree of enhanced resistance in both *prp40c-1* and *prp40c-2* mutant alleles when compared to wild-type plants (Figure 4B). Finally, we confirmed by RT-PCR some of the alterations in splicing detected in transcripts related to abiotic and biotic stress in the RNA-seq experiment (Figure 5).



## DISCUSSION

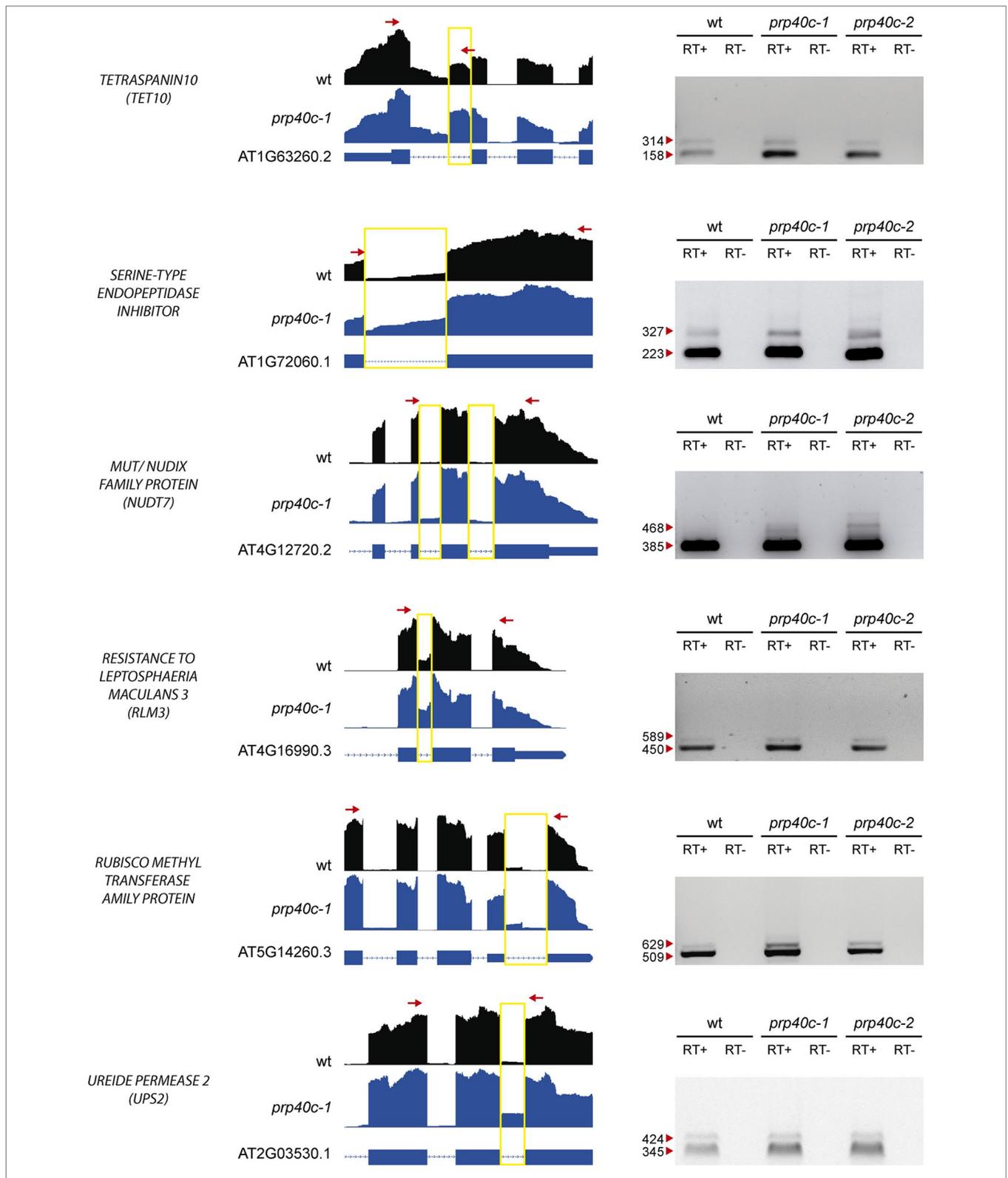
Proper control of gene expression and pre-mRNAs processing is crucial for eukaryotic organisms, and, indeed, many pre-mRNA processing events occur cotranscriptionally. PRP40 proteins appear to couple transcript elongation to the control of pre-mRNA splicing, but the global extent of their contribution to the regulation of gene expression networks, as well as their physiological roles, have only recently started to be characterized in higher eukaryotes (Munoz-Cobo et al., 2017). The founding member of this family (ScPRP40) was identified in yeast (Kao and Siliciano, 1996) and shown to be an essential splicing factor. Interestingly, PRP40 constitutes a small gene family in most eukaryotes, suggesting that duplication of the ancestral PRP40 gene followed by sequence divergence might have contributed to the evolution of novel molecular and/or physiological roles for some members of this protein family (Becerra et al., 2016). A comparison between PRP40s from *Saccharomyces cerevisiae*, *Homo sapiens*, and *A. thaliana* showed that A-type PRP40s are probably an ancestral feature of eukaryotic cells. Interestingly, AtPRP40C clusters with human TCERG1/CA150, but this clustering most likely reflects the divergent sequences in this group compared to other family members, rather than a common evolutionary origin. Still, C-type PRP40s proved to be conserved across all plant species, including streptophyte algae. This prompted us to explore a possible role for AtPRP40C as a modulator of gene expression programs, which could act linking transcription and splicing to optimize plant growth and development.

Loss-of-function *prp40c* mutants were viable and showed no severe phenotypical alterations compared to wild-type plants, suggesting that indeed PRP40C is a modulator of gene expression networks rather than an essential splicing factor. Whether

other PRP40 genes are essential in *Arabidopsis* remains to be determined. It is possible that different members of this gene family play both redundant as well as specific functions in the control of gene expression and that the essentiality of the PRP40 gene family could only be observed in a triple mutant lacking all PRP40 homologues.

In support of a modulatory function for PRP40C, *prp40c* mutants displayed a late flowering phenotype only under long-day photoperiods, revealing that the mutant does not have a global developmental alteration, but rather a defect in the environmental control of specific developmental traits. Furthermore, the finding that *prp40c* mutants only flower late under long-day conditions reveals that the photoperiodic flowering pathway may be specifically affected by this mutation; this well-known signaling pathway is controlled by light perception and the circadian clock through their effects on gene expression. *prp40c* mutants in fact exhibit hypersensitivity to red light but not alterations in circadian clock function, suggesting that a defect in light signaling rather than in time measurement could be associated with its altered flowering time phenotype.

To analyze the molecular cause of the late flowering phenotype, we looked for genes related to the control of the photoperiodic flowering pathway among the DEG in the mutant. We found four candidates: AT1G01060 (*LHY*), AT2G46830 (*CCA1*), AT1G25560 (*TEM1*), and AT1G68840 (*TEM2*). Given that the circadian function was not affected in the mutant, the flowering phenotype was most likely due, at least in part, to the overexpression of *TEM1* and *TEM2*, rather than to changes in *LHY* and *CCA1* clock genes (Supplementary Figure S6). In fact, the transcription factors *TEM1* and *TEM2* are floral repressors of the photoperiodic and the gibberellin



**FIGURE 5 |** Analysis of alternative splicing events. RT-PCR validation of six events identified as alternatively spliced through RNA-seq. The read density map for each event evaluated is displayed (WT: black, *prp40c-1*: blue). A scheme describing each gene is displayed below the read density maps, with exons and introns displayed as boxes and lines, respectively. A yellow square encloses the measured event, while red arrows display the position of the oligos used for the RT-PCR measurement. An image of the agarose gel with the RT-PCR amplicons is displayed next to the read density maps. Red arrowheads indicate amplicon sizes. +, retrotranscriptase added; -, retrotranscriptase not added.

flowering induction pathways (Osnato et al., 2012; Matias-Hernandez et al., 2014).

Almost three quarters of the differentially expressed genes in *prp40c-1* mutants were upregulated when compared to the wild type, suggesting that PRP40C may be acting as a negative transcriptional regulator. It has been proposed that HsTCERG1/CA150 is a promoter-specific negative regulator of RNA Pol II transcription elongation, and it is possible that plant PRP40C proteins have a similar repressive role (see **Supplementary Table S2**) (Sune et al., 1997; Sune and Garcia-Blanco, 1999).

Regarding alterations in pre-mRNA splicing in *prp40c-1* mutants, we detected 680 events as differentially spliced between mutant and wild-type plants. Although this is a significant number of differentially spliced events, it is relatively small compared with the numbers of altered events that we have previously observed in other *Arabidopsis* mutants affected in core spliceosomal components, such as *lsm5* (6,049 events) and *lsm4* (6,377 events), or in mutants affected in spliceosome assembly, such as *gemin2* (6680 events) (Perez-Santangelo et al., 2014; Schlaen et al., 2015). Among the different types of annotated AS events, IR was clearly overrepresented in the group of events that were differentially spliced between *prp40c* mutants and wild-type plants. In addition, *prp40c* had a stronger impact on the splicing of alternative compared to constitutive introns. These results, taken together, suggest that AtPRP40C is a splicing modulator, which affects a specific subset of splicing events, rather than an essential splicing factor.

In contrast to what was reported for HsPRPF40B, we found no significant deviations in the sequences of the global consensus donor and acceptor splice sites compared to the sequences of donor or acceptor sites present in the differentially spliced events (i.e., the affected events were not associated with weak splice site signals). This observation suggests that AtPRP40C may contribute to the regulation of pre-mRNA splicing events through a mechanism that is at least partially different from its previously proposed role in the stabilization of weak RNA-RNA interactions (Becerra et al., 2015).

When DEG and the DST were categorized into functional groups several interesting results came to light. Among the DEG, upregulated genes were enriched in several GO categories (**Supplementary Table S2**). Some of the overrepresented categories may help to explain, at least in part, some of the phenotypes observed in the mutants, such as red light hypersensitivity (Response to light stimulus) and the splicing alterations (RNA metabolic process). This suggests that some of the splicing alterations observed in the *prp40c* mutants could take place through the differential expression of other splicing regulators rather than through a direct action of PRP40C on pre-mRNA processing.

Other GO terms significantly overrepresented were immune response (RF = 4.45), response to hormone stimulus (RF = 2.4), and response to stress (RF = 1.9). Particularly, we noticed that among the 34 upregulated genes related to

the immune response ontology, 21 (60%) were also related to signal transduction (RF = 1.9). Among the DST signal-transduction related genes that were affected, categories such as phosphorylation (RF = 1.45) and protein kinase activity (RF = 1.38) were significantly enriched. Abiotic stress-related GO categories such as salt stress (RF = 1.73), metal ion transport (RF = 2.46), and ion transmembrane transporter activity (RF = 1.80) were also significantly enriched among DST. The ion-transport related categories were enriched among the DST but not among DEG, suggesting a level of specificity for AtPRP40C in splicing regulation.

Both biotic and abiotic stresses were differentially affected in *prp40c* mutants when compared to wild-type plants. Interestingly, seed germination was affected in the mutants under salt stress, but growth was not affected by this stress at the vegetative stage. In addition, cold stress during the early vegetative stage did not affect *prp40c* root growth. According to the data available in the Bio-Analytic Resource for Plant Biology of the University of Toronto (Toronto BAR—www.bar.utoronto.ca/), AtPRP40C expression is highest at the dry seed stage (**Supplementary Table S3**), which may explain why PRP40C plays a role in salt tolerance during germination. Nonetheless, a more thorough evaluation of stress tolerance phenotypes throughout a variety of life stages is needed to unveil in more detail the regulatory role of PRP40C in abiotic stress responses. On the other hand, *prp40c* mutants proved to be more resistant to *P. syringae* infection. We found that 15 out of the 34 DEG (44%) grouped in the immune response GO were TIR-NBS-LRR (Toll interleukin1 receptor–nucleotide binding site–leu-rich repeat) class of R (resistance) proteins, and some of these TIR-NBS-LRR proteins are known to enhance disease resistance when overexpressed (Parker et al., 1996; Tang et al., 1999; Stokes et al., 2002). These data suggest that AtPRP40C may regulate this specific group of genes negatively in order to orchestrate an appropriate defense response. Interestingly, according to the GENEVESTIGATOR database, the expression of the *PRP40C* gene appears to be modulated by both abiotic (salt and cold) as well as biotic (*P. syringae*) stresses, suggesting that PRP40C acts as a node within the stress response gene regulatory network in *Arabidopsis* (**Supplementary Figure S7**) (Hruz et al., 2008).

In summary, our work constitutes the first physiological and molecular characterization of a member of the PRP40 protein family in plants and the most thorough characterization of the global effects of PRP40 on gene expression and pre-mRNA splicing in eukaryotic organisms. We provide evidence that PRP40C is an important factor linking the regulation of gene expression programs to the modulation of plant growth, development, and biotic and abiotic stress responses. We propose that *PRP40C* evolved from an ancestral *PRP40* gene that had an essential role in the control of pre-mRNA splicing, into a global modulator of gene expression and splicing that targets specific genes and processes to improve physiological adjustments to environmental challenges. Although it is tempting to speculate that PRP40C modulates gene expression by coupling the regulation of transcription rates to the control of

pre-mRNA splicing, its precise mechanism of action remains to be determined.

## DATA AVAILABILITY

The datasets generated for this study can be found in the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129932>).

## AUTHOR CONTRIBUTIONS

Most of the investigation and experiments were performed by CEH and MGH. Pathogen infection assay was performed by MJdL and PCR alternative splicing and qPCR differential expression assessment was performed by DC. RNA-Seq data analysis was performed by JI and sequence retrieval and phylogenetic analysis was performed by SMG. Writing, review, and editing were done by CEH, SMG, and MJY. All authors read and approved the final manuscript.

## REFERENCES

- Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* 89 (3), 403–412. doi: 10.1016/S0092-8674(00)80221-4
- Allison, L. A., Wong, J. K., Fitzpatrick, V. D., Moyle, M., and Ingles, C. J. (1988). The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mammals: a conserved structure with an essential function. *Mol. Cellular Biol.* 8 (1), 321–329. doi: 10.1128/MCB.8.1.321
- Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Res.* 22 (10), 2008–2017. doi: 10.1101/gr.133744.111
- Becerra, S., Andres-Leon, E., Prieto-Sanchez, S., Hernandez-Munain, C., and Sune, C. (2016). Prp40 and early events in splice site definition. *Wiley Interdiscip. Rev. RNA* 7 (1), 17–32. doi: 10.1002/wrna.1312
- Becerra, S., Montes, M., Hernandez-Munain, C., and Sune, C. (2015). Prp40 pre-mRNA processing factor 40 homolog B (PRPF40B) associates with SF1 and U2AF65 and modulates alternative pre-mRNA splicing *in vivo*. *RNA* 21 (3), 438–457. doi: 10.1261/rna.047258.114
- Bedford, M. T., and Leder, P. (1999). The FF domain: a novel motif that often accompanies WW domains. *Trends Biochem. Sci.* 24 (7), 264–265. doi: 10.1016/S0968-0004(99)01417-6
- Bombom, O. (2014). “seqLogo: Sequence Logos for DNA Sequence Alignments. R Package Version 1.32.1.” in *R Foundation for Statistical Computing*.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57 (1), 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Braunschweig, U., Gueroussou, S., Plocik, A. M., Graveley, B. R., and Blencowe, B. J. (2013). Dynamic integration of splicing within gene regulatory pathways. *Cell* 152 (6), 1252–1269. doi: 10.1016/j.cell.2013.02.034
- Buratowski, S. (2003). The CTD code. *Nat. Struct. Biol.* 10, 679. doi: 10.1038/nsb0903-679
- Carty, S. M., Goldstrohm, A. C., Sune, C., Garcia-Blanco, M. A., and Greenleaf, A. L. (2000). Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phosphoCTD of RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* 97 (16), 9015–9020. doi: 10.1073/pnas.160266597
- Darnell, J. E., Jr. (2013). Reflections on the history of pre-mRNA processing and highlights of current knowledge: a unified picture. *RNA* 19 (4), 443–460. doi: 10.1261/rna.038596.113
- Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C. L., Kastner, B., et al. (2006). Protein composition and electron microscopy structure of

## FUNDING

This research was funded by grants from Agencia Nacional de Promoción Científica y Tecnológica to MJY.

## ACKNOWLEDGMENTS

We thank the Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET) for the scholarships of MJL, CEH, DC, and JI and Agencia Nacional de Promoción Científica y Tecnológica for the scholarship of MGH.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01019/full#supplementary-material>

- affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.* 26 (14), 5528–5543. doi: 10.1128/MCB.00582-06
- Deng, X., and Cao, X. (2017). Roles of pre-mRNA splicing and polyadenylation in plant development. *Curr. Opin. Plant Biol.* 35, 45–53. doi: 10.1016/j.pbi.2016.11.003
- Ding, F., Cui, P., Wang, Z., Zhang, S., Ali, S., and Xiong, L. (2014). Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in *Arabidopsis*. *BMC Genom.* 15, 431. doi: 10.1186/1471-2164-15-431
- Filichkin, S., Priest, H. D., Megraw, M., and Mockler, T. C. (2015). Alternative splicing in plants: directing traffic at the crossroads of adaptation and environmental stress. *Curr. Opin. Plant Biol.* 24, 125–135. doi: 10.1016/j.pbi.2015.02.008
- Goldstrohm, A. C., Albrecht, T. R., Sune, C., Bedford, M. T., and Garcia-Blanco, M. A. (2001). The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol. Cell. Biol.* 21 (22), 7617–7628. doi: 10.1128/MCB.21.22.7617-7628.2001
- Greenleaf, A. L. (1993). Positive patches and negative noodles: linking RNA processing to transcription? *Trends Biochem. Sci.* 18 (4), 117–119. doi: 10.1016/0968-0004(93)90016-G
- Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., et al. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv. Bioinformatics* 2008, 420747. doi: 10.1155/2008/420747
- Kang, C. H., Feng, Y., Vikram, M., Jeong, I. S., Lee, J. R., Bahk, J. D., et al. (2009). *Arabidopsis thaliana* PRP40s are RNA polymerase II C-terminal domain-associating proteins. *Arch. Biochem. Biophys.* 484 (1), 30–38. doi: 10.1016/j.abb.2009.01.004
- Kao, H. Y., and Siliciano, P. G. (1996). Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.* 16 (3), 960–967. doi: 10.1128/MCB.16.3.960
- Katari, M. S., Nowicki, S. D., Aceituno, F. F., Nero, D., Kelfer, J., Thompson, L. P., et al. (2010). VirtualPlant: a software platform to support systems biology research. *Plant Physiol.* 152 (2), 500–515. doi: 10.1104/pp.109.147025
- 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. *Nat. Rev. Mol. Cell. Biol.* 14 (3), 153–165. doi: 10.1038/nrm3525
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* 35 (6), 1547–1549. doi: 10.1093/molbev/msy096
- Kwon, Y. J., Park, M. J., Kim, S. G., Baldwin, I. T., and Park, C. M. (2014). Alternative splicing and nonsense-mediated decay of circadian clock genes under environmental stress conditions in *Arabidopsis*. *BMC Plant Biol.* 14, 136. doi: 10.1186/1471-2229-14-136

- Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi: 10.1093/nar/gkr1090
- Lin, K. T., Lu, R. M., and Tarn, W. Y. (2004). The WW domain-containing proteins interact with the early spliceosome and participate in pre-mRNA splicing *in vivo*. *Mol. Cell. Biol.* 24 (20), 9176–9185. doi: 10.1128/MCB.24.20.9176-9185.2004
- Macias, M. J., Wiesner, S., and Sudol, M. (2002). WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Lett.* 513 (1), 30–37. doi: 10.1016/S0014-5793(01)03290-2
- Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M., et al. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* 298 (5601), 2205–2208. doi: 10.1126/science.1077783
- Mancini, E., Iserte, J., Yanovsky, M., and Chernomoretz, A. (2019). *ASpli: Analysis of alternative splicing using RNA-Seq*. 1.8.1. ed. Bioconductor: Bioconductor.
- Marquez, Y., Brown, J. W. S., Simpson, C., Barta, A., and Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res.* 22 (6), 1184–1195. doi: 10.1101/gr.134106.111
- Matias-Hernandez, L., Aguilar-Jaramillo, A. E., Marin-Gonzalez, E., Suarez-Lopez, P., and Pelaz, S. (2014). RAV genes: regulation of floral induction and beyond. *Ann. Bot.* 114 (7), 1459–1470. doi: 10.1093/aob/mcu069
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J. et al. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385 (6614), 357–361. doi: 10.1038/385357a0
- Meyer, K., Koester, T., and Staiger, D. (2015). Pre-mRNA splicing in plants: *in vivo* functions of RNA-binding proteins implicated in the splicing process. *Biomolecules* 5 (3), 1717–1740. doi: 10.3390/biom5031717
- Misteli, T., and Spector, D. L. (1999). RNA polymerase II targets pre-mRNA splicing factors to transcription sites *in vivo*. *Mol. Cell.* 3 (6), 697–705. doi: 10.1016/S1097-2765(01)80002-2
- Munoz-Cobo, J. P., Sanchez-Hernandez, N., Gutierrez, S., El Yousfi, Y., Montes, M., Gallego, C., et al. (2017). Transcriptional elongation regulator 1 affects transcription and splicing of genes associated with cellular morphology and cytoskeleton dynamics and is required for neurite outgrowth in neuroblastoma cells and primary neuronal cultures. *Mol. Neurobiol.* 54 (10), 7808–7823. doi: 10.1007/s12035-016-0284-6
- Muñoz, M. J., de la Mata, M., and Kornblihtt, A. R. (2010). The carboxy terminal domain of RNA polymerase II and alternative splicing. *Trends Biochem. Sci.* 35 (9), 497–504. doi: 10.1016/j.tibs.2010.03.010
- Nawrath, C., Schell, J., and Koncz, C. (1990). Homologous domains of the largest subunit of eucaryotic RNA polymerase II are conserved in plants. *Mol. Gen. Genet. MGG* 223 (1), 65–75. doi: 10.1007/BF00315798
- Oliveros, J. C. (2007). VENNY. An interactive tool for comparing lists with Venn Diagrams. <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.
- Osnato, M., Castillejo, C., Matias-Hernandez, L., and Pelaz, S. (2012). TEMPRANILLO genes link photoperiod and gibberellin pathways to control flowering in *Arabidopsis*. *Nat. Commun.* 3, 808. doi: 10.1038/ncomms1810
- Parker, J. E., Holub, E. B., Frost, L. N., Falk, A., Gunn, N. D., and Daniels, M. J. (1996). Characterization of eds1, a mutation in *Arabidopsis* suppressing resistance to *Peronospora parasitica* specified by several different RPP genes. *Plant Cell.* 8 (11), 2033–2046. doi: 10.1105/tpc.8.11.2033
- Pearson, J. L., Robinson, T. J., Munoz, M. J., Kornblihtt, A. R., and Garcia-Blanco, M. A. (2008). Identification of the cellular targets of the transcription factor TCERG1 reveals a prevalent role in mRNA processing. *J. Biol. Chem.* 283 (12), 7949–7961. doi: 10.1074/jbc.M709402200
- Perez-Santangelo, S., Mancini, E., Francey, L. J., Schlaen, R. G., Chernomoretz, A., Hogenesch, J. B., et al. (2014). Role for LSM genes in the regulation of circadian rhythms. *Proc. Natl. Acad. Sci. U. S. A.* 111 (42), 15166–15171. doi: 10.1073/pnas.1409791111
- Phatmani, H. P., and Greenleaf, A. L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* 20 (21), 2922–2936. doi: 10.1101/gad.1477006
- Rappsilber, J., Ryder, U., Lamond, A. I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res.* 12 (8), 1231–1245. doi: 10.1101/gr.473902
- Reddy, A. S., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell.* 25 (10), 3657–3683. doi: 10.1105/tpc.113.117523
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616
- Sanchez-Alvarez, M., Goldstrohm, A. C., Garcia-Blanco, M. A., and Sune, C. (2006). Human transcription elongation factor CA150 localizes to splicing factor-rich nuclear speckles and assembles transcription and splicing components into complexes through its amino and carboxyl regions. *Mol. Cell. Biol.* 26 (13), 4998–5014. doi: 10.1128/MCB.01991-05
- Schlaen, R. G., Mancini, E., Sanchez, S. E., Perez-Santangelo, S., Rugnone, M. L., Simpson, C. G., et al. (2015). The spliceosome assembly factor GEMIN2 attenuates the effects of temperature on alternative splicing and circadian rhythms. *Proc. Natl. Acad. Sci. U. S. A.* 112 (30), 9382–9387. doi: 10.1073/pnas.1504541112
- Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* 3 (6), 1101–1108. doi: 10.1038/nprot.2008.73
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9 (7), 671–675. doi: 10.1038/nmeth.2089
- Shikata, H., Hanada, K., Ushijima, T., Nakashima, M., Suzuki, Y., and Matsushita, T. (2014). Phytochrome controls alternative splicing to mediate light responses in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 111 (52), 18781–18786. doi: 10.1073/pnas.1407147112
- Staiger, D., and Brown, J. W. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell.* 25 (10), 3640–3656. doi: 10.1105/tpc.113.113803
- Stokes, T. L., Kunkel, B. N., and Richards, E. J. (2002). Epigenetic variation in *Arabidopsis* disease resistance. *Genes Dev.* 16 (2), 171–182. doi: 10.1101/gad.952102
- Sune, C., and Garcia-Blanco, M. A. (1999). Transcriptional cofactor CA150 regulates RNA polymerase II elongation in a TATA-box-dependent manner. *Mol. Cell. Biol.* 19 (7), 4719–4728. doi: 10.1128/MCB.19.7.4719
- Sune, C., Hayashi, T., Liu, Y., Lane, W. S., Young, R. A., and Garcia-Blanco, M. A. (1997). CA150, a nuclear protein associated with the RNA polymerase II holoenzyme, is involved in Tat-activated human immunodeficiency virus type 1 transcription. *Mol. Cell. Biol.* 17 (10), 6029–6039. doi: 10.1128/MCB.17.10.6029
- Tang, X., Xie, M., Kim, Y. J., Zhou, J., Klessig, D. F., and Martin, G. B. (1999). Overexpression of Pto activates defense responses and confers broad resistance. *Plant Cell.* 11 (1), 15–29. doi: 10.1105/tpc.11.1.15
- Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25 (9), 1105–1111. doi: 10.1093/bioinformatics/btp120
- Wahl, M. C., Will, C. L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136 (4), 701–718. doi: 10.1016/j.cell.2009.02.009

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Hernando, García Hourquet, de Leone, Careno, Iserte, Mora Garcia and Yanovsky. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.