





Arabidopsis thaliana GLYCINE RICH RNA-BINDING PROTEIN 7 interaction with its iCLIP target *LHCB1.1* correlates with changes in RNA stability and circadian oscillation

Martin Lewinski^{1,†} , Alexander Steffen^{1,†}, Nitin Kachariya^{2,3,†}, Mareike Elgner¹, Christoph Schmal⁴, Niki Messini^{2,3}, Tino Köster¹, Marlene Reichel¹, Michael Sattler^{2,3} , Kathi Zarnack⁵  and Dorothee Staiger^{1,*} 

¹RNA Biology and Molecular Physiology, Faculty of Biology, Bielefeld University, Bielefeld, Germany,

²Helmholtz Munich, Molecular Targets and Therapeutics Center, Institute of Structural Biology, Neuherberg 85764, Germany,

³Department of Bioscience, Bavarian NMR Center, Technical University of Munich, TUM School of Natural Sciences, Garching 85747, Germany,

⁴Institute for Theoretical Biology, Humboldt-Universität zu Berlin and Charité-Universitätsmedizin Berlin, Berlin, Germany, and

⁵Buchmann Institute for Molecular Life Sciences (BMLS) & Institute of Molecular Biosciences, Goethe University Frankfurt, Frankfurt, Germany

Received 25 August 2023; accepted 9 December 2023; published online 20 December 2023.

*For correspondence (e-mail dorothee.staiger@uni-bielefeld.de).

[†]These authors are equal first authors.

SUMMARY

The importance of RNA-binding proteins (RBPs) for plant responses to environmental stimuli and development is well documented. Insights into the portfolio of RNAs they recognize, however, clearly lack behind the understanding gathered in non-plant model organisms. Here, we characterize binding of the circadian clock-regulated *Arabidopsis thaliana* GLYCINE-RICH RNA-BINDING PROTEIN 7 (*AtGRP7*) to its target transcripts. We identified novel RNA targets from individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP) data using an improved bioinformatics pipeline that will be broadly applicable to plant RBP iCLIP data. 2705 transcripts with binding sites were identified in plants expressing *AtGRP7*-GFP that were not recovered in plants expressing an RNA-binding dead variant or GFP alone. A conserved RNA motif enriched in uridine residues was identified at the *AtGRP7* binding sites. NMR titrations confirmed the preference of *AtGRP7* for RNAs with a central U-rich motif. Among the bound RNAs, circadian clock-regulated transcripts were overrepresented. Peak abundance of the *LHCB1.1* transcript encoding a chlorophyll-binding protein was reduced in plants overexpressing *AtGRP7* whereas it was elevated in *atgrp7* mutants, indicating that *LHCB1.1* was regulated by *AtGRP7* in a dose-dependent manner. In plants overexpressing *AtGRP7*, the *LHCB1.1* half-life was shorter compared to wild-type plants whereas in *atgrp7* mutant plants, the half-life was significantly longer. Thus, *AtGRP7* modulates circadian oscillations of its *in vivo* binding target *LHCB1.1* by affecting RNA stability.

Keywords: *Arabidopsis*, iCLIP, RNA-binding protein, RNA-protein interaction, PureCLIP score, circadian.

INTRODUCTION

Higher plants rely on an endogenous timekeeper, the circadian clock, to optimally align their physiology and development to the periodic changes of day and night. The core clockwork operates in each cell (Nohales & Kay, 2016). Through transcriptional feedback loops, the clock proteins undergo self-sustained 24-hour (h) oscillations that persist with a period of approximately 24 h in the absence of entraining light–dark cycles. The rhythmically expressed clock proteins in turn cause circadian oscillations in the

transcriptional activity of large parts of the genome. More recently, regulation at the RNA level has been shown to make an important contribution to shaping the circadian transcriptome (Mateos et al., 2018; Romanowski & Yanovsky, 2015; Staiger & Green, 2011). In line with this, RNA-binding proteins (RBPs) have been identified that undergo circadian oscillations themselves (Staiger, 2001). Among those, the *AtGRP7* (*Arabidopsis thaliana* GLYCINE-RICH RNA-BINDING PROTEIN 7) transcript encoding an RBP with a single RNA recognition motif (RRM) with the

conserved ribonucleoprotein (RNP) motifs RNP1 and RNP2 oscillates with a peak at the end of the daily light period (Carpenter et al., 1994; Heintzen, Melzer, et al., 1994). *AtGRP7* undergoes negative autoregulation via alternative splicing and nonsense-mediated decay of the splice isoform and thus a decline in the mature mRNA (Heintzen et al., 1997; Staiger et al., 2003). This represents the first reported clock-regulated feedback loop operating at the RNA level. *AtGRP7* is also involved in innate immunity and flowering time control (Hackmann et al., 2014; Steffen et al., 2019). Furthermore, *AtGRP7* exhibits RNA chaperone and nuclear export function and promotes freezing tolerance (Kim et al., 2005, 2007, 2008). Collectively, *AtGRP7* is a key posttranscriptional regulator, and identification of its binding targets will provide insights into how *AtGRP7* mediates the various facets of RNA-based control.

Individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP) is a state-of-the-art method to determine the binding landscape of RBPs transcriptome-wide, originally developed in metazoans (Haberman et al., 2017; Hafner et al., 2021; König et al., 2010; Schwich et al., 2021). Essentially, *in vivo* RNA–protein interactions are preserved by irradiation with UV light that leads to covalent bonds between nucleobases and nearby proteins. RBPs and associated RNAs are immunoprecipitated from cell lysates and the identity of the bound RNAs is determined by high-throughput sequencing (Hafner et al., 2021; König et al., 2010; Köster & Staiger, 2020). Importantly, during library preparation reverse transcriptase stops at the crosslinked peptides remaining after protease digestion. Consequently, the sequencing reads start one nucleotide downstream of the crosslink position so that binding sites can be determined with nucleotide resolution.

We previously adapted iCLIP to plants to profile binding targets of *AtGRP7*, identifying transcripts bound *in vivo* by an *AtGRP7*-GREEN FLUORESCENT PROTEIN (GFP) fusion but not by an RNA-binding dead variant of *AtGRP7* or GFP alone (Meyer et al., 2017). Here, we set out to characterize binding sites and address the impact of *AtGRP7* binding on the fate of its targets.

As a first step, we improved the bioinformatics pipeline to re-evaluate our previous *AtGRP7* iCLIP reads (Meyer et al., 2017). Using this modified pipeline, we now uncovered 2705 transcripts with reproducible binding sites, expanding the number of targets about three-fold. Furthermore, we introduced a definition for binding sites and determined a conserved GUUUC sequence motif enriched around the *AtGRP7* binding sites. Analysis of RNA binding *in vitro* using nuclear magnetic resonance (NMR) spectroscopy and isothermal titration calorimetry (ITC) experiments confirmed a preference of *AtGRP7* for RNAs with a central three-uridine motif, consistent with the iCLIP peaks.

In an orthogonal transcriptome analysis in *AtGRP7* overexpression plants and loss-of-function mutants

performed at four time points throughout the circadian cycle, the iCLIP targets were found to be predominantly negatively regulated by *AtGRP7*. Among the bound transcripts, circadian transcripts were overrepresented. *LIGHT HARVESTING CHLOROPHYLL A/B-BINDING PROTEIN 1.1* (*LHCB1.1*) peak abundance was depressed upon overexpression of *AtGRP7* whereas the peak was elevated in the loss-of-function mutant. We confirmed that recombinant *AtGRP7* binds to a binding motif in the *LHCB1.1* 3' untranslated region (UTR) whereas mutation of the iCLIP peak sites reduced binding. In Arabidopsis plants constitutively overexpressing *AtGRP7*, *LHCB1.1* transcript stability was reduced whereas it was increased in the loss-of-function mutant, indicating that *AtGRP7* affects the *LHCB1.1* transcript by regulating its stability. Transient expression of *LHCB1.1* reporter constructs in *Nicotiana benthamiana* suggested that binding of *AtGRP7* to the 3' UTR contributes to *LHCB1.1* downregulation but mutation of iCLIP sites did not abolish regulation by *AtGRP7*, indicating redundancy of regulatory elements.

RESULTS

Determination of *AtGRP7 in vivo* target transcripts

To obtain insights into the transcriptome-wide binding landscape of the key posttranscriptional regulator *AtGRP7*, we previously had for the first time applied iCLIP to Arabidopsis plants (Meyer et al., 2017). Transgenic plants expressing *AtGRP7*-GFP in the *atgrp7-1* mutant were grown in light–dark cycles and subsequently released to continuous light. UV crosslinking was done in the subjective evening, the circadian maximum of *AtGRP7* expression (36 h after transfer to continuous light, LL36), and in the subjective morning, the circadian minimum of *AtGRP7* expression (LL24) (Meyer et al., 2017). RNA–protein complexes were recovered by immunoprecipitation using GFP Trap beads. Upon high-throughput sequencing of cDNA libraries generated from the bound transcripts, we identified 858 transcripts with crosslink sites at the very same nucleotide in at least four out of five biological LL36 replicates. As an RRM can bind to more than one nucleotide and thus the protein can be crosslinked to different nucleotides of the binding site, we reasoned that we may have excluded many RNA partners (Lunde et al., 2007).

To test whether we indeed underestimated the number of binding targets, we reanalyzed the original *AtGRP7* iCLIP reads using a substantially modified bioinformatics pipeline. We defined a consistent nomenclature where we denoted the genomic position one nucleotide upstream of the mapped read start (–1 position) as the *crosslink site* (Figure 1). For further processing, the alignment from each read was discarded and only the crosslink site was kept. Each read mapped onto a crosslink site was considered a *crosslink*. A *called peak* described a crosslink site that

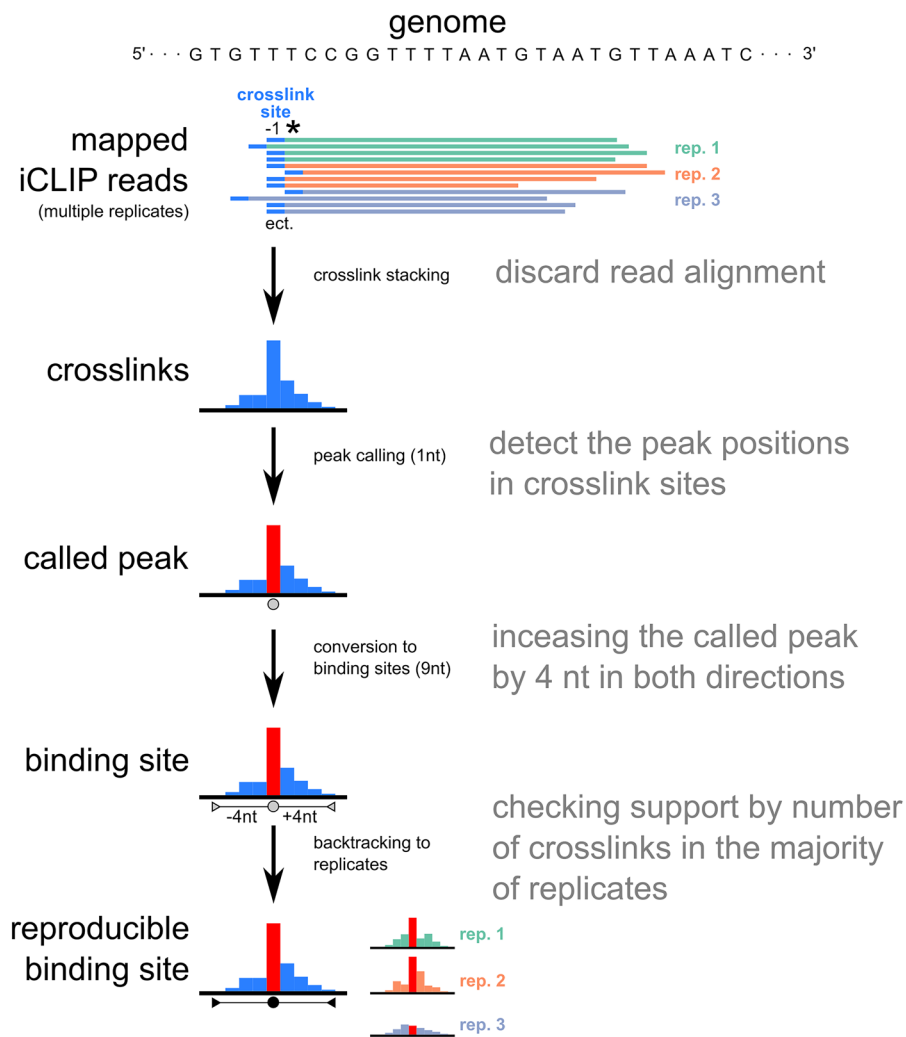


Figure 1. iCLIP binding site nomenclature. Overview of terms related to iCLIP reads and processing steps used in this publication. The first base of the mapped read (read start coordinate) is marked with an asterisk.

contained an enrichment of crosslinks in a close vicinity and was reported transcriptome-wide via a peak calling algorithm (Krakau et al., 2017). In the modified pipeline, we merged the biological replicates before peak calling to increase the signal strength (Busch et al., 2020). Subsequently, instead of determining the false discovery rate (FDR) by comparing the accumulation of AtGRP7 crosslinks with randomly shuffled crosslinks on each target transcript to identify significant peak positions (König et al., 2010; Meyer et al., 2017), we adapted the recently developed PureCLIP tool which is more sensitive (Krakau et al., 2017). PureCLIP trains a hidden Markov Model from genomic crosslink coordinates and uses the read fragment density to normalize for expression levels. The called peaks are reported by PureCLIP transcriptome-wide at single nucleotide resolution and correspond to the center of the RBP's binding site. As an RRM is expected to contact about two

to eight nucleotides (Corley et al., 2020; Lunde et al., 2007), we defined a *binding site* as a 9 nucleotide region for AtGRP7, where the called peak is in the center position (-4, ...0...+4). Finally, we introduced *reproducible binding sites* (Figure 1) where each binding site was tested whether it was supported by a critical amount of crosslinks in the majority of replicates, that is, contained a minimum amount of crosslinks overlapping the binding site. Thus, we checked for reproducibility among the replicates at the level of binding sites rather than the level of individual crosslink sites, as done previously (Meyer et al., 2017). In addition to these major changes, we introduced a few minor changes during read processing, as detailed in Experimental procedures.

Re-evaluation of our previously generated AtGRP7 LL36 iCLIP reads (Meyer et al., 2017) by these means resulted in an increased amount of uniquely mapped reads

and, thus, called peaks compared to the previous evaluation. The read counts and binding sites after each processing step are listed in Table S1. After removal of called peaks which were located at immediately adjacent nucleotides to reduce redundancy, we determined above 50 000 peaks. The coordinates were transformed to binding sites: The distribution of crosslinks at binding sites approximately follows a negative binomial distribution (Busch et al., 2020). Figure S1 (left) shows the numbers of crosslinks above and below the 10%, 20%, 30%, 40%, and 50% quantile of this distribution. We chose 30% as the filtering threshold to include binding sites that are detected in the majority of replicates (Figure S1, right) with the rationale that this threshold removes dominant binding sites supported by only a low number of crosslinks or a single replicate. After applying this filter, 27 931 binding sites appeared in at least four out of five replicates at LL36 (Table S2).

When considering binding sites in the LL24 iCLIP data (Meyer et al., 2017), which were supported by the majority of replicates (30% quantile and above), 2018 binding sites were identified (Table S2). To check whether AtGRP7 is bound to similar sites at LL36 and LL24, the position of the binding site peaks was compared. As shown in Figure S2, 1785 out of the 2018 binding site peaks detected at LL24 (88.5%) occurred within a 9- nucleotide window (−4, +4) of the binding site peaks at LL36, and the majority of these (1398) occurred at the identical position. Because of this strong overlap of the reproducible binding sites at LL36 and LL24, we merged both data sets into a single set of 28 551 reproducible binding sites.

To eliminate background binding, we determined whether the reproducible AtGRP7 binding sites also occurred in control plants expressing a mutated AtGRP7 variant with a conserved arginine within the RRM mutated to glutamine (AtGRP7 R⁴⁹Q-GFP) which strongly reduces RNA binding, or GFP alone (Schöning et al., 2007). 324 out of the 28 551 binding sites were found within a region of four nucleotides around the peaks site in the control plants and thus were removed, leaving 28 224 reproducible AtGRP7 binding sites. The read counts and binding sites after each processing step are listed in Table S1.

By overlapping the position of the reproducible binding sites with the Araport11 gene models of coding and non-coding genes, we identified 2705 target transcripts (Table S3). 84% (795/941) of the target transcripts previously determined for either LL36 or LL24 (Meyer et al., 2017) were again identified by the modified pipeline. For the remaining targets, we assume that the more stringent mapping criteria precluded their identification. On the other hand, we identified 1910 novel targets. Among the newly detected AtGRP7 targets, which displayed crosslinks but not significant crosslink sites in the previous analysis, we have now identified, for example, the plant

defensin *PDF1.2a*, ethylene response factor *RAP2.3/EBP*, dehydrin family protein *RAB18*, senescence-associated gene *SEN1*, and transcripts encoding a defensin-like protein and chaperone DnaJ-domain superfamily protein J8 as direct targets (Figure 2). These transcripts were found to be expressed at reduced levels in AtGRP7-ox plants in a microarray analysis (Streitner et al., 2010) and were now confirmed as direct targets of AtGRP7.

In order to get some information on the transcripts bound by AtGRP7, we performed a gene ontology (GO) term enrichment analysis of the updated AtGRP7 target gene set. Transcripts involved in various photosynthesis-related processes were enriched with high confidence, followed by transcripts involved in the response to cold or diverse other abiotic stresses (Figure S3, Table S4). AtGRP7 itself reacts strongly to external stimuli, and transcripts responding to stress and to abiotic stimuli were also prevalent among transcripts identified in a microarray analysis of differential expression in AtGRP7-ox plants (Streitner et al., 2010).

To validate the AtGRP7 iCLIP targets by an independent method, we previously had performed RNA immunoprecipitation (RIP)-seq on the *AtGRP7::AtGRP7-GFP* plants subjected to formaldehyde crosslinking and identified 2780 binding targets (Meyer et al., 2017). 1304 (46.91%) of these RIP targets were also identified among the iCLIP targets based on the new bioinformatics pipeline (Figure S4). In contrast, only 491 (17.66%) of the RIP binding targets were detected among the iCLIP targets determined by our previous pipeline, suggesting that the new pipeline provided higher sensitivity (Meyer et al., 2017).

AtGRP7 binding site location and binding motif

Next, we determined the localization of the reproducible binding sites within the transcripts. Most binding sites were found in the coding region (~42%) and 3' UTR (~38%) of protein-coding genes (Figure S5a). When considering the overall length of the features in the genome, binding sites were enriched in the 3' UTRs (about 3-fold) and in the 5' UTRs (2.4-fold) (Figure S5a). As PureCLIP assigns scores to each called peak relative to the strength of the crosslink signal, it is now possible to rank binding sites according to their PureCLIP score. As these scores can vary considerably within transcripts (Figure 2), we examined how the score values of reproducible binding sites were distributed among these regions (Figure S5b). Binding sites mapping to coding regions on average reported lower values compared to the ones mapping to 3' UTRs (*P*-value <0.001). An analysis of score distributions across the same protein-coding targets showed that with increasing peak scores the predominant location of AtGRP7 binding sites in the 3' UTRs increased (Figure S5c). This effect was even more prominent when normalizing to the overall size of the features in the genome (Figure S5d). Therefore, the AtGRP7

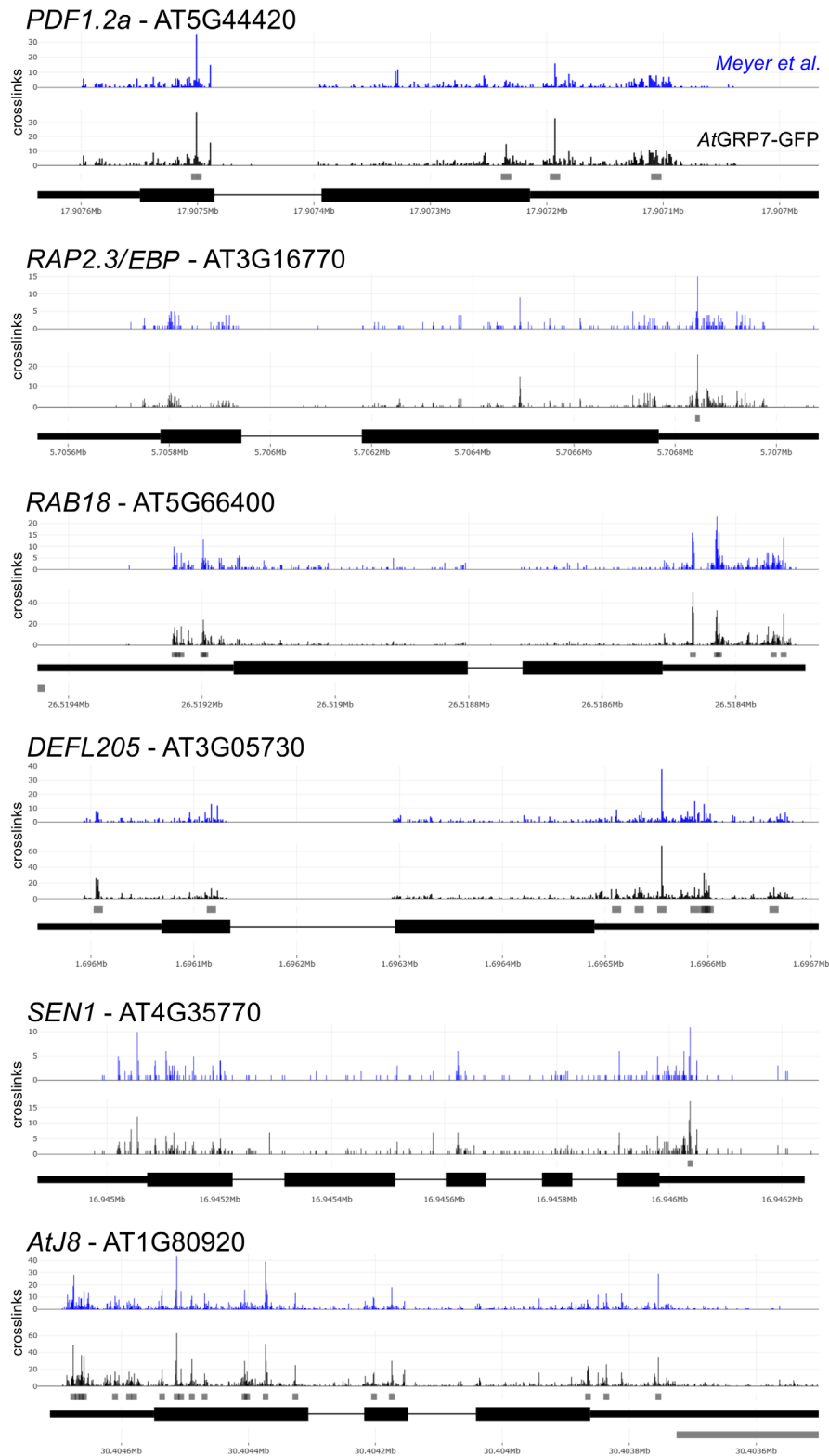


Figure 2. Comparison of crosslink and binding sites for selected iCLIP targets.

Each panel shows the SEQing tracks of reproducible AtGRP7 binding sites (Lewinski et al., 2020). The top track shows crosslinks from the first iteration of the bioinformatics pipeline at LL36 (blue) (Meyer et al., 2017) and the bottom track shows data from the improved pipeline (black). The gray bars below the crosslink track denote the position of reproducible binding sites. Gene models in block format are depicted below the data tracks. Thin bars represent 5' UTR and 3' UTR; thick bars denote coding regions and lines denote introns.

binding sites scoring highest resided mostly in the 3' UTRs.

To determine sequence motifs enriched at the reproducible AtGRP7 binding sites, the corresponding sequences were investigated using *pentamer* counts, *pentamer z*-scoring (Meyer et al., 2017), and machine learning (Bailey & Elkan, 1994). An initial analysis in the 5' UTR, coding sequence, intron, and 3' UTR separately revealed an over-representation of pyrimidine-rich sequences in each of these regions (Table S5). Therefore, we applied the motif discovery to the whole transcript instead of evaluating each region separately, as done previously (Meyer et al., 2017). The pentamer approach revealed a small set of subsequences enriched at the AtGRP7 reproducible binding sites, namely GUUUC, UUUCU, GUUUU, UUCUU, and UGUUU. The emerging pattern consisted of a guanine followed by multiple uracil residues and a cytosine (Figure 3a).

Then we applied machine learning on the identical set of sequences with STREME. The AtGRP7 iCLIP targets showed one significant motif at the reproducible binding

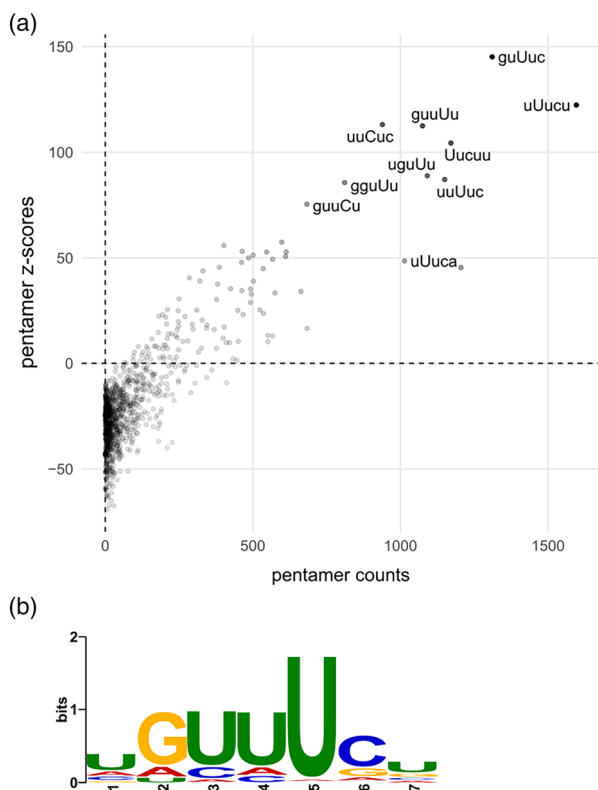


Figure 3. Sequence motifs enriched at the AtGRP7 reproducible binding sites.

(a) Scatterplot displaying pentamer counts (x-axis) and pentamer z-scores (y-axis). Pentamers in the upper right corner are enriched in comparison to a random background correction model. Letters in upper case mark the peak position of the reproducible binding site.

(b) Significantly enriched motif determined by STREME. The resulting motif was present in 13 904 out of 28 224 (49%) reproducible binding sites with a *P*-value of 2.6e-002.

sites (Figure 3b). The core of the motif, GUUUC, was identical to one of the top-ranked pentamers and partly overlapped with other top-ranked pentamers (Figure 3a). Therefore, both motif discovery approaches were in high agreement.

Overall, the improved pipeline is envisaged to be widely applicable to determine RNA binding motifs for other RBPs as well, leading to an increased understanding of *in vivo* RNA-protein interactions and, thus, the grammar of plant RNA-based regulation.

NMR titrations confirm that the AtGRP7 RRM binds to U-rich motifs

The AtGRP7 iCLIP peaks were found in U-rich regions. As UV crosslinking has an inherent bias to uridines due to the elevated photoreactivity of single-stranded (ss) U-rich regions, we tested whether AtGRP7 would also bind to U-rich regions *in vitro* (Haberman et al., 2017; Hafner et al., 2021). Therefore, NMR experiments were performed with the recombinantly expressed N-terminal half of AtGRP7 comprising the RRM (amino acid residues 1–91). The protein backbone NMR chemical shifts were assigned (Figure 4a), providing information about the secondary structure of the protein. We observed a β 1- α 1- β 2- β 3- α 2- β 4- β 5 topology (Figure 4b). This corresponds to a canonical RRM core, however, extended by an additional strand β 5, which was also predicted by AlphaFold2 (Clery et al., 2008; Daubner et al., 2013; Jumper et al., 2021). Such a C-terminal extension incorporating an additional fifth β strand compared to conventional RRMs was previously observed for the third RRM of polypyrimidine tract binding protein (PTB) and later also for the second RRM of PTB (Conte et al., 2000; Simpson et al., 2004). NMR steady-state $\{^1\text{H}\}$ - ^{15}N -heteronuclear nuclear Overhauser effect data demonstrated that the RRM (amino acid residues 8–85) has a rigid core with flexible N- and C-terminal residues (Figure 4c).

Next, to assess the RNA binding preference experimentally, NMR binding experiments were performed for the AtGRP7 RRM with a short RNA oligonucleotide 5' AGUUUCA 3' derived from the iCLIP motif (Figures 3b and 4d). The analysis showed significant NMR chemical shift perturbations (CSPs) that demonstrated strong binding to the 5' AGUUUCA 3' RNA by the AtGRP7 RRM (Figure 4d,e). Based on the spectral changes observed in the NMR titrations, the binding site involved residues in the conserved RNP1 and RNP2 motifs located on the β -sheets of the RRM domain, which collectively formed a positively charged surface region (Figure 4e–g).

As expected, RNA binding involved the β 1 and β 3 sheets with the conserved RNP1 and RNP2 sequence motifs as in canonical RNA recognition by RRM domains (Figure 4b,e). The β 4 strand showed only weak interaction with the RNA substrate, and structure prediction suggested that it was folded back behind the actual binding surface

(Figure 4f). Instead, the C-terminal $\beta 5$ strand exhibited strong CSPs upon RNA addition, suggesting that it may substitute for $\beta 4$ in RNA binding. This is in contrast to RRM3 of human PTB, where the $\beta 5$ strand provides an extension of the binding surface (Conte et al., 2000).

Using ITC, the RNA binding affinity for the 5' AGUUUCA 3' RNA oligonucleotide was determined with a dissociation constant K_D of 3 μM (Figure 4h, left panel). Mutation of the central 'UUU' to 'AAA' decreased the binding affinity by 2.6-fold to 8 μM (Figure 4h, middle panel). Taken together, these data showed that AtGRP7 recognized the RNAs with a central 'UUU' motif with low micromolar affinity.

To further investigate the RNA binding specificity of the AtGRP7 RRM, a series of short oligonucleotides with different sequences were designed and binding assessed by NMR. A 7-mer RNA 5' AGUUUCA 3' showed similar spectral changes to the corresponding single-stranded (ss) DNA 5' AGTTTCA 3' (Figure 4d,e; Figure S6a left, S6b top panel). As the binding affinity to the single-stranded DNA oligonucleotide was only weakly reduced compared to the RNA oligonucleotide (Figure 4h, right panel) we reasoned that the sequence-specific recognition, which involves the nucleobases is comparable to the corresponding RNA. Therefore, we explored the binding specificity using ssDNA oligos with different sequences.

We compared NMR titrations with the core 'TTT' motif and additional flanking adenosines, guanosines, and cytosines (Table S6). A, G, and C nucleotides flanking the core motif showed strong and similar CSPs with only minor variations of the CSP pattern (Figure S6a,b,e). In contrast, mutation of the core thymidine nucleotides to purines (5' AGAAACA 3') strongly reduced binding (Figure S6c left, S6d upper panel) and 5' AAAAAA 3' did not interact at all (Figure S6c right, S6d lower panel). However, the shorter version of 5-mer DNA oligonucleotides with a central TTT was bound with similar affinity, except for small deviations at the RNP2 region (Figure S7a,b). Again, exchanging the thymidines for adenosines (5' GAAAG 3') strongly reduced binding (Figure S7c left, S7d upper panel), whereas 5' AAAAA 3' did not interact at all (Figure S7c right, S7d lower panel, S7e). In summary, our *in vitro* NMR and ITC binding studies confirmed 'UUU' as a core recognition motif for the AtGRP7 RRM.

Impact of altered AtGRP7 level on iCLIP targets

To obtain insights into how AtGRP7 affected its target transcripts, we analyzed RNA-seq data of AtGRP7-ox plants and the *atgrp7-1 8i* mutant (the *atgrp7-1* mutant that harbors an *AtGRP8* RNAi construct to counteract elevated *AtGRP8* levels due to relief of repression) grown in parallel to the plants used for the iCLIP experiment (Meyer et al., 2017). In addition to plants harvested 36 h after transfer to LL (LL36) corresponding to the *AtGRP7* peak and at LL24 corresponding to the *AtGRP7* trough, we now

included plants harvested at LL18 corresponding to the falling phase and LL30 corresponding to the rising phase in the analysis. The read statistics are presented in Table S7. The correlation between the three biological replicates of each genotype showed a high agreement among the RNA-seq samples at each of the four time points (Figure S8).

For each time point, we determined genes differentially expressed (DEGs, $|\log_2$ fold change >1 , FDR <0.01) in the mutant or *AtGRP7*-ox plants relative to wild type (wt) (Table S8). In the *atgrp7-1 8i* mutant, 303 transcripts were differentially expressed at any time point. 75 (2.77%) of all 2705 iCLIP targets were differentially expressed at any time point. Of the 1910 iCLIP targets identified exclusively by the new pipeline, 57 (3.04%) were differentially expressed at any time point. In contrast, only 1.91% (18/941) of the iCLIP targets identified by our previous pipeline were differentially expressed at any time point. Thus, among the iCLIP targets we now identified a higher number of transcripts whose expression was affected by the loss of AtGRP7.

In the *AtGRP7*-ox plants, 1400 transcripts were differentially expressed at any time point. 171 (6.32%) of the 2705 iCLIP targets were DEGs. 101 of the 941 iCLIP targets identified by our previous pipeline were differentially expressed at any time point (10.73%) whereas 70 of the 1910 iCLIP targets identified by the new pipeline were differentially expressed at any time point (4.35%). This suggests that in addition to the large number of DEGs that have been identified as iCLIP targets already by the previous pipeline, few additional targets were revealed by the novel pipeline.

The temporal distribution of DEGs is displayed in Figure S9(a). When considering all DEGs, more genes were differentially expressed in the *AtGRP7*-ox plants than in the *atgrp7-1 8i* mutant. In the *AtGRP7*-ox plants, a similar number of transcripts was expressed at elevated or reduced levels compared to wt plants at LL18, and more transcripts were expressed at elevated levels than at reduced levels at LL24, LL30, and LL36. The DEGs detected in the *atgrp7-1 8i* plants were predominantly upregulated.

We then compared the distribution of DEGs with iCLIP targets identified by the previous bioinformatics pipeline (Figure S9b), genes which were newly discovered as *AtGRP7* targets by the updated pipeline (Figure S9c), and the total target set of the updated pipeline (Figure S9d). For the iCLIP targets identified by the previous pipeline, the few targets differentially expressed in the mutant were mostly upregulated whereas the targets were mostly downregulated in the *AtGRP7*-ox plants (Figure S9b). This points to a predominantly negative effect of *AtGRP7* on its direct targets in contrast to its overall effect on the transcriptome (Figure S9a). The newly identified *AtGRP7* targets were almost exclusively upregulated in the *atgrp7-1 8i* mutant and downregulated in *AtGRP7*-ox plants

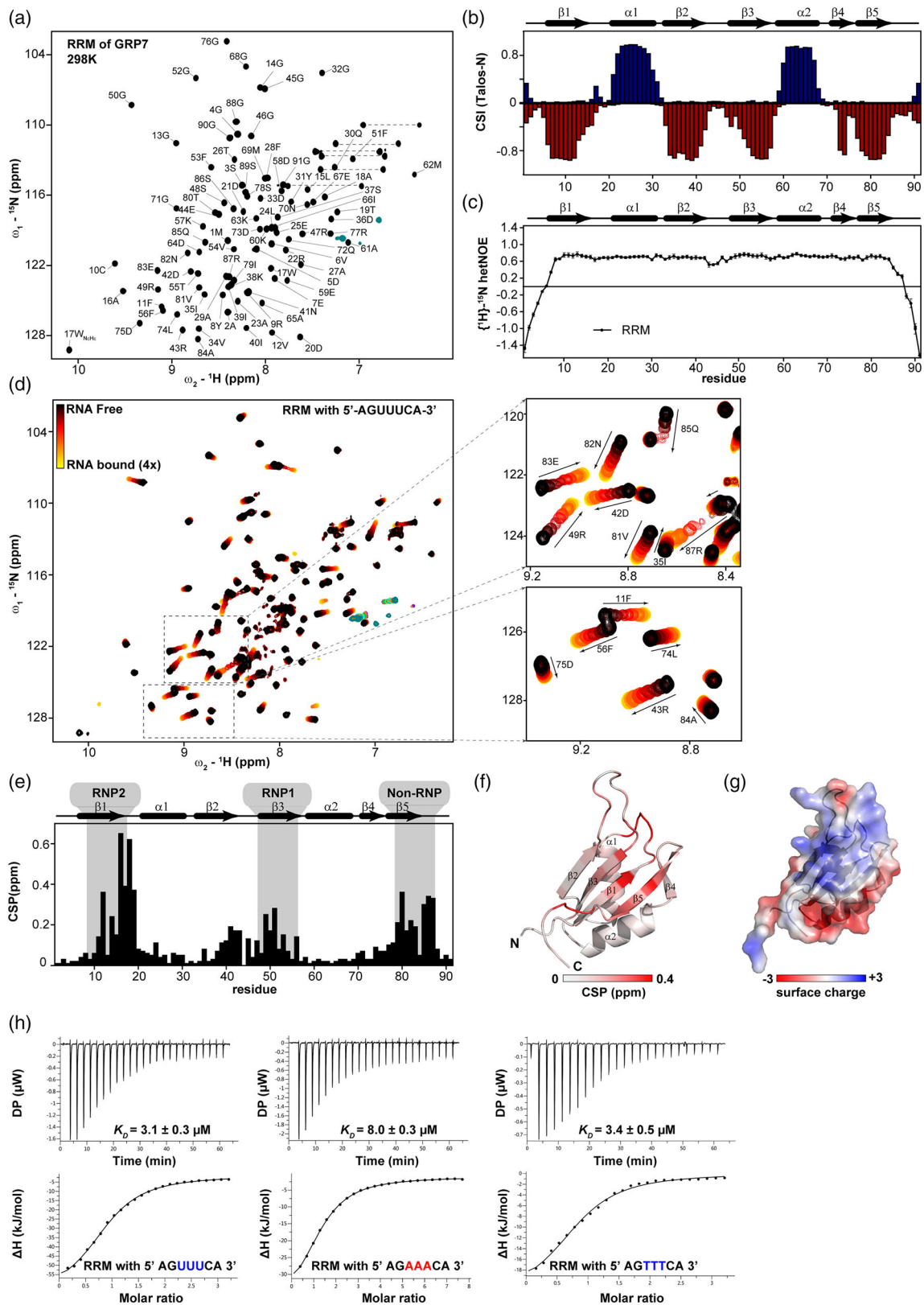


Figure 4. Binding of the canonical AtGRP7 RRM to the iCLIP motif.

- (a) Two dimensional (^1H)- ^{15}N correlation spectra of the AtGRP7 RRM and indicated with assignment of respective amide resonances.
- (b) TALOS-N based chemical shift index (CSI) of the AtGRP7 RRM backbone plotted against the amino acid residues. Helices and sheets are indicated in blue and red, respectively. Secondary structural elements are shown on top.
- (c) Steady-state (^1H)- ^{15}N -heteronuclear Nuclear Overhauser effect plot for the RRM.
- (d) Overlay of series of (^1H)- ^{15}N HSQC spectra of ^{15}N labeled RRM with increasing concentration of 5' AGUUUCA 3' RNA ligand are shown. The right panels provide zoom versions of the overlaid spectra.
- (e) Chemical shift perturbation (CSP) plot for the RRM amide resonances upon binding with a 4-fold excess of the 7-mer RNA oligonucleotide. Secondary structure elements are indicated on the top.
- (f) The CSPs are mapped onto the cartoon structure of the RRM. The gradient from white to red indicates weak to strong CSPs.
- (g) Adaptive Poisson-Boltzmann Solver (APBS)-derived electrostatic potential on the solvent-accessible surface area of the RRM is shown. Red to blue indicates the negative to positive surface charges, respectively.
- (h) Isothermal titration calorimetry (ITC) of binding of recombinant AtGRP7 RRM to a 7-mer RNA oligonucleotide 5' AGUUUCA 3' comprising the AtGRP7 consensus binding motif determined from iCLIP binding sites (left), a mutant version 5' AGAAACA 3' (middle), and a single-stranded DNA oligonucleotide 5' AGTTTCA 3' (right). The corresponding dissociation constant (K_D) is displayed in the upper panels.

(Figure S9c). Thus, newly identified targets were also subject to predominantly negative regulation through AtGRP7, validating a similar outcome of our previous and of our novel pipeline.

The impact of AtGRP7 on circadian-regulated transcripts

To determine the portion of clock-regulated transcripts among the iCLIP targets identified by the new pipeline, we used a recent extended circadian transcriptome as a benchmark (Romanowski et al., 2020). In this experiment, seedlings were grown on agar plates in entrainment conditions (12 h light-12 h dark), shifted to LL on day 11, and harvested every 4 h for 2 days. 9127 transcripts out of 18 503 transcripts expressed at any time point (49.3%) were found to be clock-controlled based on the Jonckheere-Terpstra-Kendall (JTK) algorithm (Hughes et al., 2010; Romanowski et al., 2020). Of the 2453 AtGRP7 iCLIP targets expressed also in this time course, 1746 (71.1%) were categorized as clock-regulated (Figure 5; Table S9). Thus, clock-controlled transcripts were significantly enriched among the iCLIP targets ($P = 0.0023$). We compared the list of clock-controlled AtGRP7 targets against a recently compiled set of clock genes (McClung, 2019). *EARLY FLOWERING 4 (ELF4)* and *FLAVIN-BINDING, KELCH REPEAT, F BOX 1 (FKF1)* were among the iCLIP target but not among transcripts differentially expressed in AtGRP7-ox or *atgrp7-1 8i* plants (Table S8), so that the relevance of AtGRP7 binding currently is unclear.

The peak times of all clock-controlled transcripts identified by Romanowski et al. were distributed around the clock, with 24.9% peaking around dawn (zt22-zt2) and 26.1% peaking around dusk (zt10-zt14) (Figure 5) (Romanowski et al., 2020). Of the 1746 iCLIP targets predicted to be clock-regulated, 30.8% showed a peak around dusk (zt10-zt14) whereas only 12.9% showed a peak around dawn (zt22-zt2). Overall, direct targets of AtGRP7 showed a similar phase distribution as the circadian transcriptome but were depleted in transcripts peaking around dawn ($P = 2.6\text{e-}16$).

To determine how AtGRP7 affected circadian oscillations of its targets, we monitored the expression of the morning-phased *LHCb1.1* gene encoding a chlorophyll-

binding protein in AtGRP7-ox plants and the *atgrp7-1 8i* line in a high-resolution time course (Figure 6) (Millar & Kay, 1991; Streitner et al., 2008). In Col wt plants, *LHCb1.1* peaked in the middle of the subjective day. Whereas the peak abundance was lower in the AtGRP7-ox plants, it was elevated in *atgrp7-1 8i* plants compared to wt plants.

To identify additional target transcripts with such reciprocal changes relative to wt plants in AtGRP7-ox plants and the *atgrp7-1 8i* mutant, respectively, we searched the RNA-seq data for the four time points (Table S8). Only few iCLIP targets showed statistically significant differential expression with a $|\log_2 \text{fold change}| > 1$ in opposite directions in AtGRP7-ox plants and the mutant, respectively (Table S10). We attribute this to the overall low number of iCLIP targets with altered expression in the loss-of-function mutant. We chose the *KETOACYL CoA SYNTHASE 8 (KCS8)*, and *ELP* encoding an extensin-like protein for validation.

Additionally, we manually searched for targets that would show a temporal expression pattern similar to *LHCb1.1* in the RNA-seq data sets (Table S7) similar to *LHCb1.1*. We chose *LHCb1.2* and *LHC1.3*, *J8* Chaperone DnaJ-domain superfamily protein, *AtAILP* encoding an aluminum-induced-like protein, *CIRCADIAN CLOCK REGULATED-LIKE (CCL)*, *DORMANCY/AUXIN ASSOCIATED FAMILY PROTEIN2 (DRM2)*, *ROTUNDIFOLIA LIKE 1 (RTFL17)*, and *BTB/POZ and TAZ domain-containing protein 2 (BT2)*. We monitored expression in AtGRP7-ox plants and the *atgrp7-1 8i* mutant grown in 12 h light-12 h dark cycles and sampled every 2 h (Figure S10). In addition, we included an *atgrp7-1 atgrp8-2* double mutant that we have recently generated by crossing the *atgrp7-1* T-DNA line to the *atgrp8-2* CRISPR mutant carrying a 29-nucleotide deletion in *AtGRP8* (Kasztelan et al, manuscript in preparation). In contrast to the *atgrp7 8i* line with the RNAi construct directed against *AtGRP8* that counteracts the upregulation of *AtGRP8* in the *atgrp7-1* T-DNA line, the double mutant lacks detectable *AtGRP8* protein.

LHCb1.1 showed an elevated peak both, in the *atgrp7-1 8i* mutant and the *atgrp7-1 atgrp8* double mutant (Figure S10a). Notably, for *LHCb1.2* and *LHCb1.3* peak

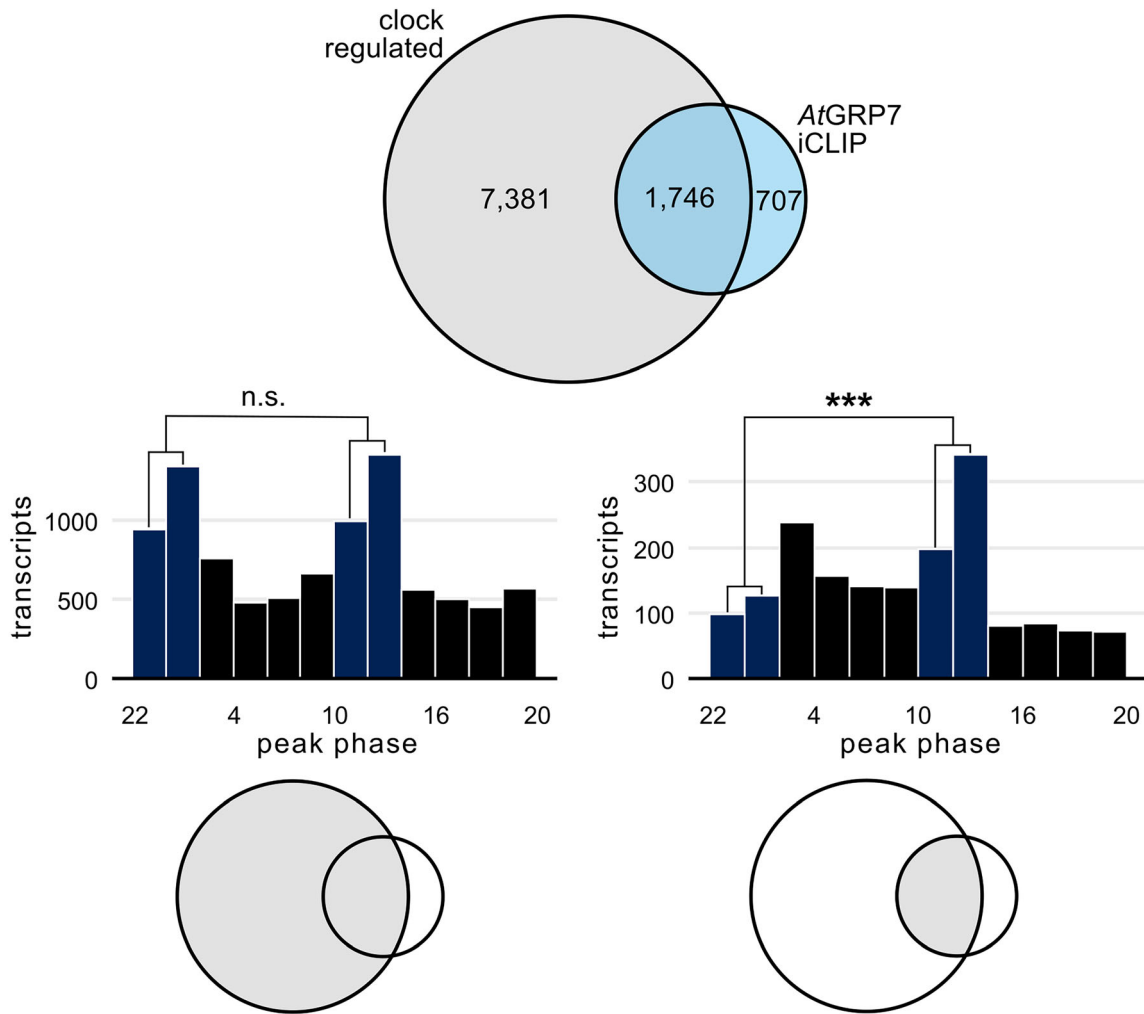


Figure 5. Impact of AtGRP7 on the circadian transcriptome. Venn Diagram of genes classified as circadian clock-regulated according to (Romanowski et al., 2020) and AtGRP7 iCLIP target genes. The phase histograms illustrate the peak distribution of all clock-regulated genes expressed in our RNA-seq data sets (left) and of the clock-regulated AtGRP7 iCLIP targets (right), respectively, based on the published phase distribution (Romanowski et al., 2020).

abundance was also reduced in the *AtGRP7-ox* plants but the expression in the mutants did not differ from wt (Figure S10b,c). Similarly, while *AtAILP1*, *ELP*, and *CCL* showed a reduced peak abundance in the *AtGRP7-ox* plants, expression in the *atgrp7-1 8i* mutant and the *atgrp7-1 atgrp8* double mutant was indistinguishable from wt (Figure S10d-f). In contrast, *J8* and *KCS8* showed a reduced peak abundance in the *AtGRP7-ox* plants and elevated expression in both, the *atgrp7-1 8i* mutant and the *atgrp7-1 atgrp8* double mutant (Figure S10g,h). *DRM2*, *RTFL17*, and *BT2* transcript abundance again was greatly reduced in *AtGRP7-ox* while it was indistinguishable from wt in both, the *atgrp7-1 8i* mutant and the *atgrp7-1 atgrp8* double mutant (Figure S10i,k,l). Thus, while elevated *AtGRP7* levels depress peak abundance of these transcripts, loss of *AtGRP7* in the *atgrp7-1 8i* line and the

atgrp7-1 atgrp8-2 mutant has no consequence, suggesting that redundant trans-acting RBPs impact on these transcripts.

AtGRP7 affects *LHCB1.1* RNA stability

As the *LHCB1.1* transcript was strongly affected by both, elevated levels of *AtGRP7* and loss of *AtGRP7*, respectively, we analyzed whether changes in RNA stability may contribute to these reciprocal changes. *AtGRP7-ox*, *atgrp7-1 atgrp8-2*, and wt plants were grown hydroponically, and transcription was inhibited by adding cordycepin at zt1, the beginning of the rising phase of *LHCB1.1* oscillation, and at zt6, the time of the peak. *LHCB1.1* RNA half-life was determined for the three biological replicates individually by assuming an exponential decay, that is, first-order degradation kinetics, and then averaged (Pett et al., 2021).

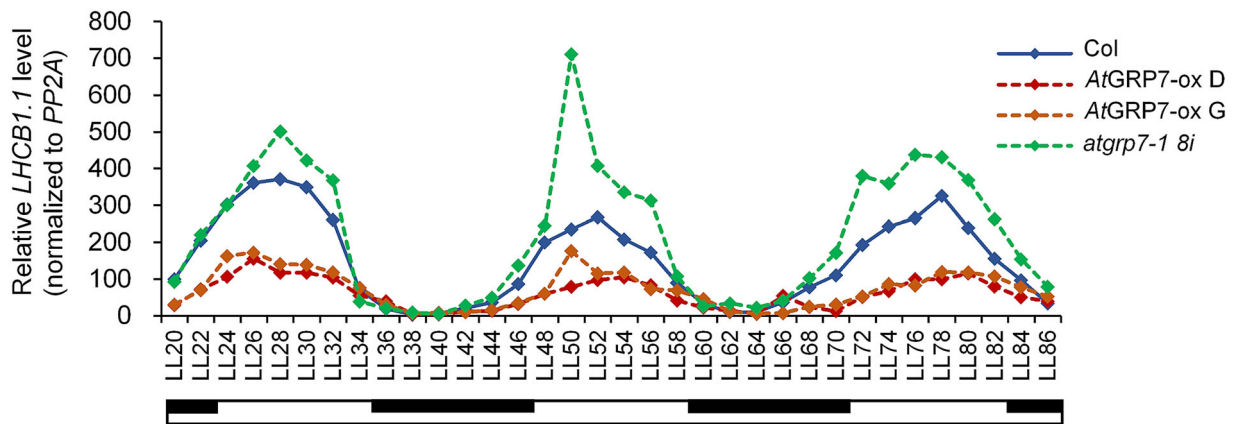


Figure 6. Circadian expression of *LHCb1.1* in plants with altered *AtGRP7* level.

Seedlings were grown in 12-h light –12-h dark cycles for 16 days, transferred to constant light (LL), and harvested at 2-h intervals. *LHCb1.1* transcript levels were determined in Col wt, two independent *AtGRP7* overexpressing lines (*AtGRP7-ox D*, *G*), and the *atgrp7-1 8i* mutant. The time course was performed twice with similar results. Open bar, constant light; inserted dark bar, subjective night.

In Col wt plants, the half-life at zt1 was 46.92 ± 4.48 min. In *AtGRP7-ox* plants, the *LHCb1.1* half-life was significantly shorter compared to wt plants (34.24 ± 4.5 min, $P = 0.0477$). In the absence of *AtGRP7* and *AtGRP8*, the half-life was significantly longer (91.81 ± 6.24 min, $P = 0.0012$) (Figure 7a–c,g). At zt6, the *LHCb1.1* half-life in wt plants was 37.16 ± 4.75 min, that is, 10 min shorter than at zt1. In *AtGRP7-ox* plants the half-life was somewhat shorter compared to wt plants (31.22 ± 2.92 min) whereas in the absence of *AtGRP7* and *AtGRP8*, it was longer (53.91 ± 9.4 min) (Figure 7d–f,h). Overall, while overexpression of *AtGRP7* led to a significant shortening of the half-life at zt1 and minor shortening at zt6, loss of *AtGRP7* and *AtGRP8* led to a significantly longer half-life in comparison to wt plants at zt1 and also at zt6, albeit less pronounced. Notably, *LHCb1.1* half-life became significantly different between zt1 and zt6 in the *atgrp7-1 atgrp8-2* mutant ($P = 0.0093$) pointing towards a specific time of day dependent effect in the mutant (Figure 7i).

Analysis of *AtGRP7* binding sites in the *LHCb1.1* UTRs

To begin to understand whether the iCLIP peaks identified in the *LHCb1.1* transcript may be responsible for the observed impact of *AtGRP7* on *LHCb1.1* oscillations, we first set out to validate binding of *AtGRP7* to the *LHCb1.1* 3' UTR, using electrophoretic mobility shift assays. In the *LHCb1.1* 3' UTR, five *AtGRP7* binding sites were detected, numbered consecutively from the stop codon towards the 3' end of the transcript (cf. Figure S11). Upon incubation of an RNA oligonucleotide 3' UTR-wt1 comprising binding site 1 with recombinant *AtGRP7*, complex formation was observed which was reduced by a 500-fold excess of homologous unlabelled oligonucleotide. The appearance of two retarded bands may point to binding of more than one protein. As *AtGRP7* iCLIP peaks were found to

comprise U-rich regions, we mutated the two 'TTT' motifs to 'AAA'. The corresponding oligonucleotide 3'UTR M1-1 competed only very weakly for *AtGRP7* binding (Figure 8a). For the oligonucleotide 3'UTR-wt3 comprising binding site 3 we observed a similar trend. Again, complex formation was reduced by a 500-fold excess of unlabelled oligonucleotide wt3. The oligonucleotide 3'UTR M3-2 with exchanges of the core 'UUU' motif and of a 'UUU' motif with substantial crosslinks immediately upstream of the predicted binding site to 'AAA' started to reduce complex formation at a 1000-fold molar excess, showing that mutation of this site also strongly reduced binding as for M1-1, albeit to a somewhat lesser extent (Figure 8b).

To address how binding of *AtGRP7* to the *LHCb1.1* transcript affects its regulation we set up quantitative transient expression of reporter constructs in *Nicotiana benthamiana*. The *GFP* reporter gene was expressed under control of the *LHCb1.1* promoter, including the UTRs (Figure S11) (Anderson et al., 1994; Leutwiler et al., 1986), and the *AtGRP7* cDNA driven by the constitutive CaMV promoter with duplicated enhancer was co-expressed. To prevent transgene-induced silencing, the silencing suppressor P19 was co-infiltrated in all experiments.

First, we performed a time course to monitor the expression of the transfected constructs. The expression of the *AtGRP7* effector strongly increased until day 2 after infiltration with only an incremental increase at day 3 (Figure S12a). Expression of the *GFP* reporter construct steadily increased from day 0 to day 3 (Figure S12b). To monitor the impact of *AtGRP7* on the reporter constructs at the earliest time point, we thus chose 2 days for the subsequent experiments.

Co-expression of the cDNA encoding the unrelated Arabidopsis Germin-like protein 3 (*AtGER3*) (Heintzen, Fischer, et al., 1994; Membre et al., 1997) served as control.

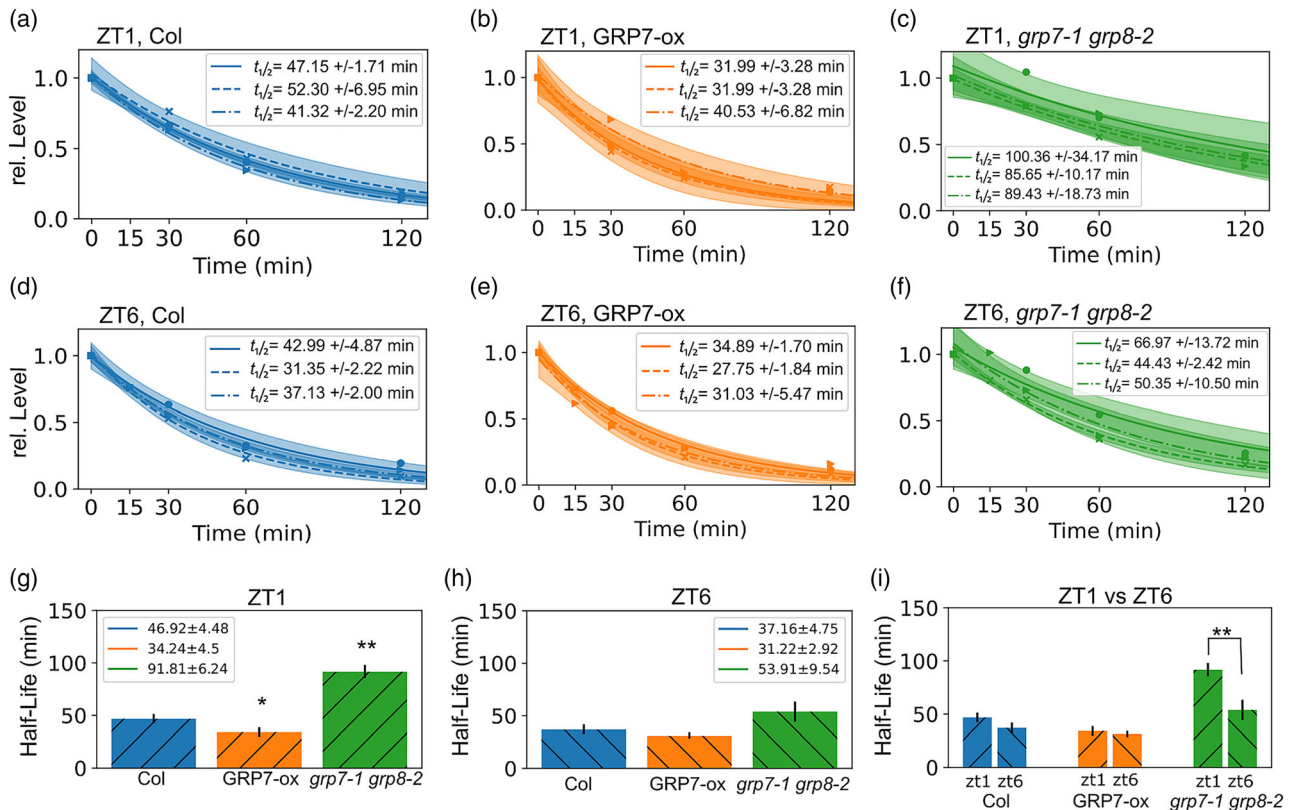


Figure 7. AtGRP7 dependent alteration of *LHCb1.1* half-life.

(a–c) Determination of *LHCb1.1* half-life at zt1 in Col wt (a), AtGRP7-ox (b) and *atgrp7-1 atgrp8-2* mutant (c) plants, assuming an exponential mRNA decay after application of cordycepin.

(d–f) Determination of *LHCb1.1* half-life at zt6 in Col wt (d), AtGRP7-ox (e), and *atgrp7-1 atgrp8-2* knockout mutant (f) plants, assuming an exponential mRNA decay after application of cordycepin.

(g–i) Bar plots depicting the determined half-lives together with the statistical error (SD) for all three genetic backgrounds at zt1 (g) and zt6 (h) and comparison between zt1 and zt6 (i) (* $P < 0.05$, ** $P < 0.01$, Student's *T* test).

Dose-dependent expression of the effector proteins was monitored by western blot (Figures S13–S15).

For the construct with the authentic 5' and 3' UTR (Figure S13a), co-infiltration of agrobacteria with the *AtGRP7* cDNA at increasing ODs led to a lower *GFP* transcript level in a dose-dependent manner in four independent experiments, whereas co-expression of *AtGER3* cDNA did not lead to reduced *GFP* transcript levels (Figure S13b–e). To determine statistical significance, we performed an outlier detection on the individual experiments and merged the data of the replicates. *GFP* transcript levels were significantly downregulated in plants infiltrated with agrobacteria carrying *AtGRP7* cDNA at ODs 0.05 and 0.1 ($P < 0.001$) (Figure S13f). The same was observed when the data was normalized to the mean expression of samples infiltrated with the reporter construct only (Figure S13g). Thus, the *Nicotiana benthamiana* transient expression system recapitulates the regulation of *LHCb1.1* by *AtGRP7*, as observed in Arabidopsis plants (cf. Figure 6; Figure S10a).

To test whether mutation of the iCLIP sites had an effect on regulation by *AtGRP7*, we mutated U-rich regions in the binding sites to adenines. The reporter construct with mutation of three binding sites in the 5' UTR (Figure S11b) also showed a significantly reduced level of the *GFP* transcript upon co-expression of agrobacteria with the *AtGRP7* cDNA at an OD of 0.05 and 0.1 ($P < 0.001$). This suggested that these motifs by themselves were not responsible for *LHCb1.1* regulation (Figure S14). The reporter constructs with mutation of the four binding sites 1 to 4 in the 3' UTR (Figure S11c) showed a somewhat reduced downregulation with *AtGRP7* at OD 0.05 ($P = 0.002$) and at OD 0.1 ($P < 0.001$) (Figure S15). However, when we applied normalization to the mean, again *AtGRP7* at the two highest expression levels exerted significant downregulation of the *GFP* reporter ($P < 0.001$) (Figure S15h). Overall, our mutations of the iCLIP peaks in the binding sites were not sufficient to abrogate the downregulation of *LHCb1.1* promoter-driven *GFP* expression by *AtGRP7* in the *N. benthamiana* system, pointing to redundant regulatory elements.

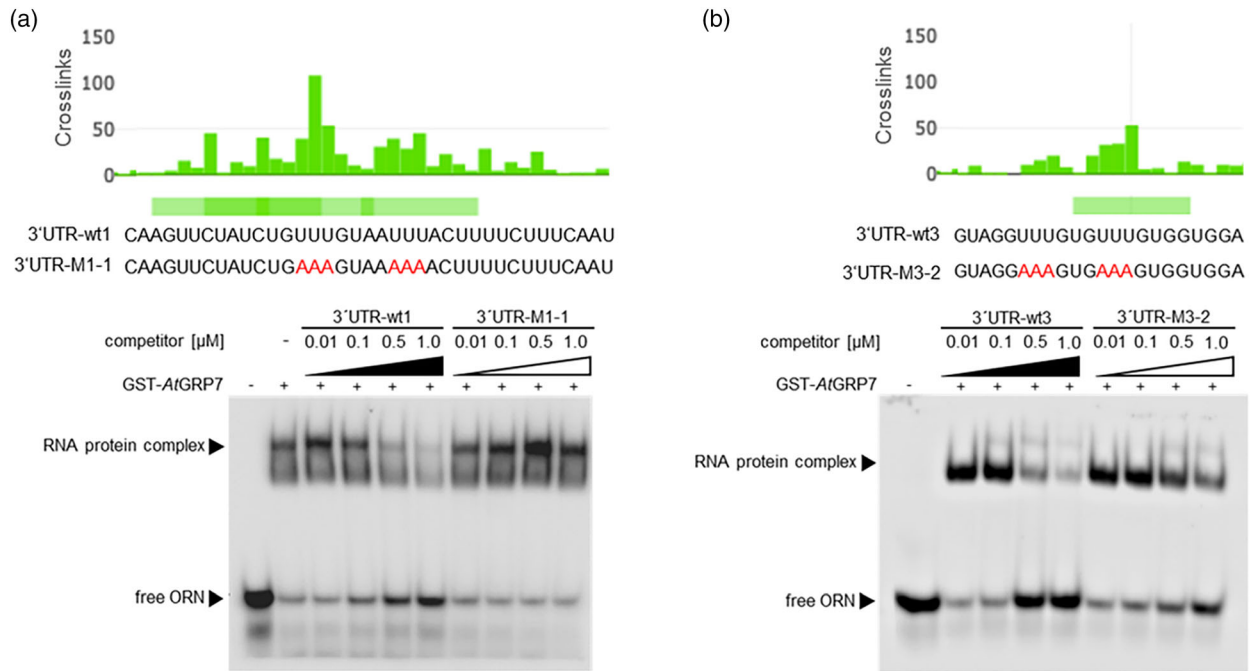


Figure 8. Binding of AtGRP7 to the *LHC1.1* 3'UTR.

(a) An RNA band shift of *LHC1.1* 3'UTR site wt1. 1 nM biotinylated RNA oligonucleotide wt1 was incubated with 4 μM GST-AtGRP7 in the presence of 1 μg yeast t-RNA. For the competition, increasing amounts of unlabeled oligo wt1 (black triangle) and oligo M1-1 with two 'UUU' motifs mutated to 'AAA' (white triangle) were added.

(b) RNA band shift of *LHC1.1* 3'UTR site wt3. 1 nM biotinylated RNA oligonucleotide wt3 was incubated with 4 μM GST-AtGRP7 in the presence of 1 μg yeast t-RNA. For competition, increasing amounts of unlabeled oligo wt3 (black triangle) and oligo M3-2 with two 'UUU' motifs mutated to 'AAA' mutated (white triangle) were added.

DISCUSSION

Here, we provide insights into in vivo binding of AtGRP7 and the consequences for the fate of its target transcripts.

Improvement of bioinformatics pipeline for Arabidopsis iCLIP data

Using iCLIP, we previously determined 858 transcripts where AtGRP7 was crosslinked to the very same nucleotide in at least four out of five biological replicates at LL36 (Meyer et al., 2017). As the RRM does not only contact a single nucleotide (Lunde et al., 2007) and thus the protein can be crosslinked to more than one neighboring nucleotide, we likely excluded many RNA partners. Therefore, we modified several steps of our analysis pipeline to improve the discovery of target RNAs, binding sites, and binding motifs with a particular focus on the peak-calling steps and binding site definition.

In the updated pipeline, we merged the biological replicates before peak calling instead of working with signals from individual replicates. This increased the number of reads on specific genomic coordinates, thereby increasing the protein footprints on the target RNAs to discover otherwise undetected peaks. A further improvement was the use of the more sensitive PureCLIP peak calling algorithm

that utilizes a hidden Markov Model (Krakau et al., 2017) instead of random shuffling and FDR (König et al., 2010; Meyer et al., 2017), as the latter required more reads in a given region to successfully call peaks and contains a random factor which reduces reproducibility. In this way, we could detect binding sites of AtGRP7 with a lower read count and therefore increase the sensitivity (Figure 2).

Another important change was the definition of what constituted an AtGRP7 binding site. Rather than considering each significant crosslink site (FDR <5%) as a distinct binding site, spanning a single nucleotide, we defined binding sites of 9 nucleotides, as an RRM is suggested to bind to 4–8 nucleotides (Lunde et al., 2007) where a called peak position (single nucleotide) represents the center of the binding site flanked by four nucleotides in both directions to cover the entire area known to interact with RNA (Lunde et al., 2007). This novel definition led to an improved procedure to test for reproducibility among replicates, as observed by Busch and co-workers (Busch et al., 2020). Crosslinks above a minimal threshold in each replicate had to be present at the binding site in the majority of replicates to consider it as reproducible and include it in further analyses. An example of how this novel binding site definition fits to the signals detected in iCLIP is

demonstrated on *RD29A* (Figure S16). This shows that the binding site width now fits the crosslink distribution gathered from iCLIP and spans over the entire binding signal.

The move from crosslink to binding sites also highly reduced the noise for motif discovery (Figure 3). Instead of multiple significant motifs for the different transcript regions, only a single consensus motif (GUUUC) was identified at the binding sites. The GUUUC sequence pattern was previously discovered as well in Meyer et al., where it is identified as a part of the significant MEME motifs and enriched pentamers in 3' UTR (Meyer et al., 2017). Additionally, the most abundant binding site pentamers (GUUUC, UUUUC, GUUUU, UUCUU, and UGUUU) from the updated analysis also match the ones identified before but did not rank as prominent in the previous approach. This can be explained by the higher noise levels in the data that originate from the larger binding area width (21 nt instead of 9 nt) and the consideration of all significant crosslink sites as independent binding sites in contrast to isolated peak positions in the updated analysis.

With the increased sensitivity of this novel pipeline, we identified 2705 transcripts with reproducible binding sites. Among those were 795 of the 941 *AtGRP7* iCLIP target transcripts determined with the previous pipeline (84.48%). Also, we detected a higher number of the transcripts we had previously identified by RIP-seq of *AtGRP7*-GFP. iCLIP displays a bias for highly expressed transcripts which is less pronounced for RIP-seq (Figure S17). Thus, our new pipeline seems to be more sensitive and will be widely applicable to determine binding targets for plant RBPs.

Among the newly determined iCLIP targets, we found a higher percentage of transcripts that were differentially expressed in the *atgrp7-1 8i* mutant relative to wt, suggesting that the improved bioinformatics pipeline indeed identified relevant targets. Furthermore, the newly detected iCLIP targets also showed a predominantly negative regulation by *AtGRP7*, in contrast to all genes differentially expressed in *AtGRP7* overexpressing or mutant plants.

For numerous transcripts shown to be downregulated in *AtGRP7*-ox plants in a previous microarray analysis (Streitner et al., 2010), we now find that they are direct targets (Figure 2). Notably, we did not determine binding sites on the *PATHOGENESIS RELATED 1 (PR1)* transcript. This is in line with our previous findings that *PR1* transcript levels were elevated likely indirectly through transcriptional activation of the *PR1* promoter in response to high *AtGRP7* levels, as constitutive and inducible *AtGRP7* overexpression leads to elevated levels of a *PR1* promoter- β glucuronidase reporter transcript, in line with the enhanced resistance to infection with *Pseudomonas syringae* in *AtGRP7*-ox plants (Hackmann et al., 2014).

Validation of *AtGRP7* binding sites

For *AtGRP7*, we determined a distinct trend toward binding in the 3' UTR. Importantly, the inclusion of the PureCLIP

peak calling algorithm in our pipeline enabled the ranking of the binding sites. We found that not only the number of binding sites were increased in the 3'UTR, but also a higher peak score was observed (Figure S5a,b). This was particularly obvious when only the highest-scored binding sites per transcript were considered, as visualized in Figure S5(d).

The move from crosslink to binding sites, that is, merging single nucleotide crosslink sites at called peaks to nine nucleotide wide regions, also highly reduced the noise for motif discovery (Figure 3). Instead of multiple significant motifs for the different transcript regions, only a single consensus motif (GUUUC) was identified at the binding sites. To independently validate *AtGRP7* binding to this motif, we performed NMR experiments. The RRM displayed a β 1- α 1- β 2- β 3- α 2- β 4- β 5 fold with a rigid core and flexible N- and C-terminal residues. RNA binding involved the β -sheet with the conserved RNP1 and RNP2 sequence motifs involved in canonical RNA recognition by RRM domains. Furthermore, the C-terminal β 5 strand exhibited strong CSPs upon RNA addition, indicating an expanded RNA interaction domain. An RNA oligonucleotide harboring the GUUUC iCLIP motif with flanking adenosine residues displayed a dissociation constant K_D of 3 μ M, while mutation of the central 'UUU' to 'AAA' decreased the binding affinity to 8 μ M. This validated the importance of uridine residues in the binding site, indicating that the motif did not merely reflect the bias of UV crosslinking towards single-stranded regions rich in U residues. Furthermore, NMR titrations comparing the potential contributions by nucleotides flanking the GUUUC consensus motif revealed that binding was only weakly affected by nucleotides flanking the core motif (Figures S6 and S7).

Although a single significant core sequence was determined, only about 7% of the reproducible binding sites contained the exact GUUUC consensus motif, mounting to 16% if GUUC and GUUUUC were also considered. Thus, the target sequences may consist of variable parts that were not found by conventional *de novo* motif discovery, as the uridine stretch could be variable in length. Also, *AtGRP7* binding may not rely on a sequence motif alone but also take structural information into account. As U-rich regions generally tend to form fewer secondary structures, the binding motif might reflect the RBPs preference for unstructured regions. Previously, an *in vitro* experiment had unveiled preferential binding of *AtGRP7* to single-stranded RNA (Schüttelpelz et al., 2008). Resolving the interdependence of *AtGRP7* binding and RNA secondary structure *in vivo* remains the next challenging step.

Among the *AtGRP7* target transcripts, we observed narrow, sharp iCLIP peaks but also broader ones. This may point to multiple overlapping binding sites and may partly be due to the fact that the motif UUU provides the possibility that *AtGRP7* binds in different registers. The appearance

of consecutive binding sites along the transcripts may also be an indication of a low-affinity scanning mode of an RBP, whereas narrow and clustered crosslink sites represent high-affinity binding sites (Chakrabarti et al., 2018).

RNA-binding domains overall display a lower specificity and affinity compared to DNA-binding domains. RBPs achieve an increase in specificity and affinity through cooperative binding of multiple RNA-binding domains. Accordingly, the presence of multiple binding sites increases the RBP's affinity for a target RNA (Helder et al., 2016; Stitzinger et al., 2023). Although AtGRP7 harbors only one conventional RNA-binding domain, the N-terminal RRM, the C-terminal glycine-rich stretch represents an intrinsically disordered region (IDR) (Xu et al., 2022). As IDRs can mediate protein-protein interactions, different AtGRP7 molecules could be brought into proximity. Tandem binding sites may thus act as sequence-specific seeds to induce local AtGRP7 condensation on RNAs (Chakrabortee et al., 2016). Similar to what is observed for an RBP with different RNA-binding domains, this increase in the number of RNA-binding domains by oligomerization is expected to increase the affinity (Stitzinger et al., 2023).

AtGRP7 affects the stability of the *LHC1.1* transcript

Previously, clock-regulated transcripts were found to be prevalent among the AtGRP7 iCLIP targets based on a comparison with circadian microarray data of 7 days old seedlings grown on medium containing 3% sucrose in 12 h light-12 h dark cycles at 22°C and released into LL (Covington & Harmer, 2007; Perez-Santángelo et al., 2014). It has become clear that sucrose has a large impact on circadian oscillations (Dalchau et al., 2011; James et al., 2008; Johnson et al., 1995). Therefore, we used the recent study by Yanovsky and co-workers which more closely resembled our growth conditions as a benchmark to determine how many iCLIP targets identified by the new pipeline are clock-regulated (Romanowski et al., 2020). With 71% of the AtGRP7 iCLIP targets being classified as clock-regulated, clock-controlled transcripts were significantly enriched.

An extended time course of *LHC1.1* expression showed that its peak abundance in the middle of the subjective day was reduced in AtGRP7-ox plants whereas it was elevated in loss-of-function mutants.

Thus, *LHC1.1* differs from other clock-regulated transcripts we have assayed so far including *DRM2* and *CCL* that showed a reduced peak abundance in response to elevated AtGRP7 levels whereas in *atgrp7-1 8i* the expression was indistinguishable from wt plants (Meyer et al., 2017). We previously attributed this to a partially redundant function of AtGRP8. However, as we now find wt levels of *CCL* and *DRM2* in the *atgrp7-1 atgrp8-2* double mutant, other proteins apart from AtGRP8 may also act redundantly to AtGRP7.

Apart from AtGRP7 itself with the expected highly elevated level in AtGRP7-ox plants and almost undetectable expression in the *atgrp7-1 8i* mutant, *LHC1.1* showed the strongest reciprocal changes in both, wt plants and the mutant (Table S10). In contrast, *LHC1.2* and *LHC1.3* showed reduced peak abundance in response to elevated AtGRP7 levels but were similar to wt in *atgrp7-1 8i* and the *atgrp7-1 atgrp8-2* double mutant. In contrast, for *J8* and *KCS8* we observed reduced peak abundance in AtGRP7-ox plants and elevated levels in both in loss-of-function mutants.

We hypothesized that AtGRP7 may exert this regulatory function by affecting RNA stability. Monitoring the decay of the *LHC1.1* transcript upon cordycepin treatment, we observed that at zt1, the rising phase of the *LHC1.1* oscillation, the half-life was significantly longer in the *atgrp7-1 atgrp8-2* mutant ($P = 0.0012$) than in wt plants and shorter in AtGRP7-ox plants ($P = 0.0477$). This suggests a so far unknown impact of transcript stability on the shape of the *LHC1.1* oscillation pattern in addition to the well described rhythmic transcriptional activation (Anderson et al., 1994; Anderson & Kay, 1995; Millar & Kay, 1991). The importance of RNA stability for circadian oscillations has been documented before in Arabidopsis, as disruption of RNA decay pathways led to circadian defects and RBPs have been implicated (Carenno et al., 2022; Lidder et al., 2005; Nolte & Staiger, 2015; Staiger, 2001).

AtGRP7 also affects the stability of its own transcript (Staiger et al., 2003). Elevated levels of AtGRP7 lead to the activation of a cryptic 5' splice site in the middle of the intron, leading to an alternative splice isoform with a much shorter half-life than the mature mRNA. As *LHC1.1* does not have introns, the impact on *LHC1.1* transcript stability must occur via different means. Notably, *LHC1.1* has been found to strongly accumulate in mutants defective in XRN4, a 5' to 3' exoribonuclease that is involved in cytoplasmic decay of mRNAs (Nagarajan et al., 2019). In the future, it will be interesting to determine whether AtGRP7 is part of a ribonucleoprotein recruiting nucleases such as XRN4 to selected target transcripts (Parker & Song, 2004).

Transient expression of an *LHC1.1-GFP* reporter RNA with authentic 5' and 3' UTRs in *N. benthamiana* recapitulated the reduced expression upon co-expression of increasing amounts of AtGRP7, while increasing amounts of the unrelated AtGER3 protein did not, indicating direct interaction of AtGRP7 with its target even in a distantly related species. However, mutational analysis of several AtGRP7 iCLIP sites in the 5' or 3' UTRs revealed no major impact on the AtGRP7 mediated downregulation of the GFP reporter level, which points to redundant regulatory elements. In the 3' UTR, iCLIP identified another extended region, site 5, with a series of closely spaced peaks immediately upstream of the polyadenylation site (Figure S11c). However, in

electrophoretic mobility shift assays, binding of recombinant AtGRP7 to this region was barely detectable (data not shown). When we introduced U to A mutations in this region in addition to the four mutated iCLIP peaks in the 3' UTR, we obtained partial alternative polyadenylation of the reporter transcript, likely due to an overlap with sequence motifs critical for polyA site determination. This precludes any interpretation as to whether AtGRP7 may have an effect on this region. Thus, the precise interplay of regulatory regions mediating *LHCB1.1* downregulation by AtGRP7 remains to be determined.

EXPERIMENTAL PROCEDURES

Plant material

The *AtGRP7::AtGRP7-GFP* line, the *AtGRP7::AtGRP7 R⁴⁹Q-GFP* line expressing an RNA binding mutant version with Arg⁴⁹ exchanged for Gln and the *AtGRP7::GFP* line expressing GFP only under the control of the AtGRP7 promoter have been described (Streitner et al., 2012). AtGRP7-ox plants express the AtGRP7 coding sequence under control of the Cauliflower Mosaic Virus (CaMV) 35S promoter with the duplicated enhancer (Heintzen et al., 1997). The *atgrp7-1 8i* line expresses an RNAi construct against AtGRP8 to counteract elevated AtGRP8 levels due to relief of repression by AtGRP7 in the *atgrp7-1* T-DNA line (Streitner et al., 2008). The *atgrp7-1 atgrp8* double mutant was generated by crossing *atgrp7-1* with the *atgrp8-2* mutant generated by CRISPR/Cas9 which will be described elsewhere. It harbors a 29 bp deletion 16 nucleotides downstream of the ATG start codon, leading to loss of detectable AtGRP8 protein (Kasztelan et al., manuscript in preparation).

Seeds were surface-sterilized and sown on half-strength MS (Murashige-Skoog) (Duchefa) plates (Staiger et al., 1999). Plants were grown in 12 h light/12 h dark cycles at 20°C in Percival incubators (CLF laboratories) followed by free run under continuous light (LL). Aerial tissue was harvested at subjective dawn or dusk for iCLIP, RIP, and RNA-seq or as indicated in the figure legends. Above 10 plants were bulked for each sample per replicate.

iCLIP

The preparation of the iCLIP libraries and the first bioinformatics determination of target transcripts using the iCount tool (König et al., 2010) have been described (Meyer et al., 2017). Here, we reevaluated the iCLIP data. Sequenced reads from all samples were investigated after processing with fastqc 0.11.5. Adapters at the 3' end were trimmed using cutadapt version 1.16. The demultiplexing of the samples was performed using flexbar 3.4.0 (Dodt et al., 2012; Roehr et al., 2017) and the -bk parameter to conserve the barcode information for further steps. Reads with a length below 24 nucleotides were discarded. Barcodes were trimmed and saved to the *read id* field. The processed high-quality reads were mapped to the TAIR10 genome with STAR version 2.6.0a allowing a maximum of 2 mismatches. In the mapping process we disabled the function to clip reads at the 5' end in order to gain higher alignment scores (soft-clipping) and to avoid shifting the position of the crosslink site (-alignEndsType Extend5pOfRead). PCR duplicates, which are by-products of the library amplification step that are derived from the same RNA molecule and share the identical unique molecular identifiers (UMIs), were removed by grouping reads by mapping start and removing reads with the same start position and random barcode. Peak calling was done using PureCLIP 1.0.4 with only uniquely mapped reads. From adjacent called peaks solely the peak with the

highest PureCLIP-score was used for further processing. Binding sites were defined as *called peak positions* (single nucleotide resolution) which were extended by 4 nt in both directions. The extension of the peak positions was computed using bedtools version 2.27.1. Binding sites with a single crosslink position within nine nucleotides were not considered further. Reproducibility among replicates was determined by overlapping binding sites with crosslinks of each sequenced replicate. A 30% quantile cutoff was set to filter out binding sites with few reads overlapping the corresponding binding sites at LL36 and LL24. Only binding sites with crosslink counts above the threshold, which also appear in four out of five replicates from the LL36 sample and two out of three from the LL24 samples, were termed reproducible binding sites. This workflow, starting from quality control up to reproducible binding sites, is described in detail in the online repository at <https://github.com/malewinski/iCLIP-AtGRP7-pipeline>. After subtraction of overlapping binding sites from the AtGRP7-R⁴⁹Q and GFP controls, the reproducible binding sites from both samples, LL36 and LL24, were merged into one set of AtGRP7 reproducible binding sites.

Motif discovery

The nine-nucleotide long sequence at each binding site was extracted by applying bedtools getfasta in strand-specific mode (Quinlan & Hall, 2010) together with the TAIR10 genome. Motif discovery was split into two separate approaches. The first aimed at *k*-mer distributions around the called peak position in the binding site region, and the second one applied Expectation maximization with STREME 5.4.1 (Bailey, 2021) to identify significant motifs. Both approaches used an identical set of sequences for motif discovery.

RNA bandshift assays

RNA oligonucleotides biotinylated at the 5' end were obtained from Metabion (Planegg, Germany). The sequences of all oligonucleotides are listed in Table S11. Recombinant Glutathione S-transferase (GST)-AtGRP7-fusion protein was expressed in *E. coli* and purified as previously described (Schöning et al., 2008). 1 nM of biotinylated probe was incubated with 4 μM of GST-AtGRP7-protein in the presence of 4.6 μM of yeast tRNA. For competition, unlabeled oligonucleotides (ORNs) were added as well as ORNs with mutated binding sites. Incubation was for 30 min. Subsequently, complexes were separated on 6% polyacrylamide gels in 40 mM Tris-acetate/1.8 mM EDTA buffer and transferred onto a nylon membrane by capillary blot and UV-crosslinked with 120 000 μJ cm⁻² in a Stratalinker (Stratagene, LA Jolla, CA, USA). Detection with Pierce™ Chemiluminescent Nucleic Acid Module Kit (Thermo Fisher, Waltham, MA, USA) was performed according to the manufacturer's instructions. Chemiluminescence signals were detected with the Fusion FX system (Vilber Lourmat, Eberhardzell, Germany) and quantification was carried out by ImageJ.

Protein expression and purification for NMR

The sequence encoding the AtGRP7 RRM (amino acid residues 1–91) was subcloned into the bacterial expression vector pETM11 (EMBL) encoding a N-terminal six Histidine tag followed by TEV cleavage site using primers listed in Table S8. Recombinant proteins were expressed in *E. coli* BL21 (DE3) strain (Novagen Merck Millipore, Darmstadt, Germany). Cells were grown in an M9 minimal medium supplemented with ¹⁵N nitrogen and/or ¹³C glucose as a sole nitrogen and carbon source, respectively. The cells were grown at 37°C to an optical density (at 600 nm) of 0.7–0.8. Protein expression was induced by the addition of 0.5 mM IPTG at 18°C overnight. Cells were harvested, resuspended in buffer A (50 mM

Tris-HCl pH 8, 150 mM NaCl, 1 mM TCEP, 10 mM imidazole) supplemented with protease inhibitors, and lysed by French press. The cleared lysate was passed through Ni²⁺-NTA resin using a gravity column. The resin was washed using buffer A with increasing NaCl concentrations and the protein was eluted with elution buffer (50 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM DTT, 250 mM imidazole). The eluted protein sample was dialyzed against dialysis buffer (20 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM DTT) overnight at 4°C in the presence of His-tagged TEV protease and passed through the Ni²⁺-NTA resin to remove the His-tag, TEV protease and uncleaved protein. The protein was further purified by anion exchange chromatography and size-exclusion chromatography. The final protein sample was in NMR buffer (20 mM sodium phosphate pH 6.5, 50 mM NaCl, 1 mM DTT). For NMR measurement, 5% D₂O was added to the protein sample to lock the external magnetic field.

NMR spectroscopy

For NMR backbone chemical shift assignments, the conventional 2D and 3D NMR experiments were recorded as: ¹H-¹⁵N heteronuclear single quantum coherence (HSQC), HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, HNN, (H)CC(CO)NH and H(CCCO)NH (Kumar et al., 2009; Sattler et al., 1999). All spectra were measured on Bruker spectrometers at proton Larmor frequency 500, 600, 800, 900, or 950 MHz equipped with cryogenic or room temperature probe. All spectra were measured at 25°C temperature using 3, 5 mm, or Shigemi tube. Spectra were processed with a shifted sine-bell window function and zero filling before Fourier transformation by Topspin 3.5p16 Bruker software or NMRPipe (Panchal et al., 2001). Proton chemical shifts were referenced against 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) in direct dimension. The spectra analysis was done using CCPN (v2.5) software (Vranken et al., 2005). Secondary structure propensity was derived from backbone chemical shifts assignments using Talos-N (Shen & Bax, 2013).

For NMR binding experiments, ¹H-¹⁵N HSQC experiments were measured in titration series using 50 μM ¹⁵N-labeled protein with increasing concentration of unlabelled synthetic RNA (Dharmacon, USA) or single-stranded DNA (Eurofins Genomics, Ebersberg, Germany) oligonucleotides. Spectra were analyzed by CCPN (v2.5) and chemical shift perturbations (CSPs) were calculated using the equation: $CSP, \Delta\Delta\delta = [(\Delta\delta^1H)^2 + (\Delta\delta^{15N}/5)^2]^{1/2}$. The CSPs were mapped on the RRM model derived from AlphaFold2 and further analyzed by PyMOL by Schrodinger tool. Cumulative CSPs were calculated by average maximum shifts observed on RNP2 regions of the RRM. Protein electrostatic surface potentials were extracted using the APBS Electrostatic tool (Jurrus et al., 2018). Steady-state {¹H}-¹⁵N heteronuclear NOE experiments were recorded with 120 and 144 msec acquisition times in direct and indirect dimensions and 3 sec of recycle delays for protein alone and in the presence of 4-fold excess of RNA (Noggle & Schirmer, 1971). The spectra were split into the reference and NOE spectra using Bruker au-program, further processed and analyzed by CCPN.

Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) measurements were carried out at 25°C using a MicroCal PEAQ-ITC (Malvern Panalytical, Kassel, Germany). Before the measurements, the protein sample was dialyzed overnight in ITC buffer (20 mM NaPO₄ pH 6.5, 50 mM NaCl, 1 mM TCEP). The ITC cell was filled with 15 μM protein concentration and the RNA or ssDNA oligonucleotide was added to the syringe with the respective concentration. The RNA or ssDNA to protein titrations were performed with 25 points of

1.5 μl injections and with a 150 sec interval. All measurements were conducted in duplicates and analyzed by MicroCal PEAQ-ITC analysis software v1.0.0.1259. For analysis, binding curves were fitted to one-site binding mode, and thermodynamic parameters were derived.

RNA-seq

Library preparation and sequencing

Library preparation and sequencing were done as in (Meyer et al., 2017). Sequencing files were examined using fastQC 0.11.5 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Each sample file was processed with sickle version 1.33 (<https://github.com/najoshi/sickle>) in single-end mode and the filtering options -q 20 and -l 30 to discard low-quality reads and sequences below a length of 30. The trimmed sequencing files were then inspected with fastQC.

Transcript quantification

Salmon version 0.91 (Patro et al., 2017) was applied to determine the expression of each *Arabidopsis thaliana* transcripts in TPM (transcripts per million) of the RNA-seq experiment in conjunction with the AtRTD2-QUASI annotation (Zhang et al., 2017). The index for salmon was created using k-mers of length 31 in quasi-mapping mode (-type quasi -k 31). While quantifying salmon was running with the correction for sequence bias (--seqBias) enabled and unstranded library (-l U) options for all samples. Additionally, the mean fragment lengths and standard deviations were provided for each sample individually using the parameters --fldMean and --fldSD for quantification with single-end reads.

Differential gene expression analysis

The expression tables from salmon were processed further using R version 3.4.0 and several commonly used packages. The import of the quantification tables from salmon into R was performed using the R package tximport 1.4.0 using the lengthScaledTPM option. Read counts were transformed to counts per million (CPM) and normalized to library sizes using the trimmed mean of M values (TMM) with the R package edgeR version 3.18.1. Lowly expressed genes not achieving a CPM value >1 in at least three samples, taking into account the mean-variance relationship (Law et al., 2016), were removed from the downstream analysis. Log₂ CPM values of each gene and sample were used to check for possible batch effects using the dimensionality reduction method Iso-map (Tenenbaum et al., 2000) from the vegan 2.5-2 package. Contrast groups were set up comparing the wt against the *atgrp7 8i* mutant or *AtGRP7-ox* line individually at each of the time points (Col-LL18 to *atgrp7-1 8i*-LL18 and so on). Differentially expressed genes were determined with limma version 3.32.10. Genes with an FDR <0.01 (1%) and an absolute log₂ fold change value >1 were considered differentially expressed.

Real-time PCR

For the *LHCb1.1* time course experiment, RNA was retrotranscribed with oligo (dT) primers, and expression was normalized to *PP2A*. For assessment of rhythmic target genes, RNA was retrotranscribed with random hexamer primers and normalized to *ACT2*. Reverse transcription was performed with native AMV reverse transcriptase (Roboklon, Berlin, Germany) according to the manufacturer's instruction. Transcript levels were determined by Real-time PCR on a CFX96 Real-time cycler (BioRad) using iTAQ SYBRGreen SuperMix (BioRad, Feldkirchen, Germany) and primers listed in Table S11.

RNA stability

Seeds were germinated in liquid half-strength MS medium supplemented with 0.5 g MES/L and incubated on a rotary shaker in 12 h light-12 h dark at 20°C. At day 10, the medium was replaced by PIPES buffer (1 mM Pipes-KOH pH 6.3, 1 mM sodium citrate, 1 mM KCl, 15 mM sucrose) containing 0.6 mM cordycepin or DMSO only at zt1 or zt6, as indicated. Plants were subject to vacuum infiltration twice for 3 min and harvested at the time points indicated. RNA was isolated with TriReagent, treated with RQ DNase (Promega, Walldorf, Germany), and precipitated overnight with LiCl. Reverse transcription was performed using random hexamer primers and native AMV reverse transcriptase (Roboklon). Real-time PCR was done as described above using primers listed in Table S11. *ACT2* served as control.

Quantification of RNA half-lives

Half-lives were determined by fitting an exponentially decaying function $f(t) = a \cdot \exp(-\log(2) \cdot t/b)$ to the data obtained from the RNA stability experiments as described in the previous paragraph and shown in Figure 7(a–c) and (e–g). Here, a denotes the value of $f(t)$ at time $t = 0$ h while b corresponds to the RNA half-life of interest. Initial values for parameters a and b used for the non-linear regression are determined from a linear fit to the corresponding log-transformed data. Non-linear regressions are performed by the *curve_fit* function of the Scientific Python (SciPy) package. Error propagation has been handled via the Python *uncertainties* package.

Transient expression in *Nicotiana benthamiana*

A reporter gene was designed comprising the *LHCB1.1* promoter starting from position –199, the *LHCB1.1* 5' UTR, the *eGFP* coding sequence with an engineered NcoI site at the ATG, followed by the *LHCB1.1* 3' UTR (*P_{prom}LHCB_5'UTR_eGFP_3'UTR*). To obtain *P_{prom}LHCB_5'UTR_{mut}_eGFP_3'UTR* the region corresponding to the iCLIP peaks in the 5' UTR was mutated from AATCACTCTC (binding site 1) to AATGGAAGCTC, CAAGTTAGT (binding site 2) to CAAAAAAGT, and AAGTTTCAA (binding site 3) to AAGAAACCA. To create *P_{prom}LHCB_5'UTR_eGFP_3'UTR_{mut1-4}* the regions corresponding to the iCLIP peaks in the 3' UTR was mutated from AGTCTATCTGTTTGTAATTTACTT (binding site 1) to AGTTC-TATCTGAAAGTAAAAACTT, TGAATTCGA (binding site 2) to TGAAAACGA, TGTTTGTGG (binding site 3) to TGAAGTGG and GAGACTTTCAG (binding site 4) to GAGACAAACAG. The constructs were obtained as synthetic genes in the vector pUCIDT (IDT, Integrated DNA Technologies). The cassettes were moved to the binary vector pHPT1 using flanking HindIII and SmaI sites (Figure S11) (Schöning et al., 2008).

The constructs expressing the *AtGRP7*, *AtGER3*, and DRONPA-s cDNAs under control of the CaMV promoter have been described (Heintzen et al., 1997; Lummer et al., 2011). For normalization of reporter gene expression, we used the expression of the hygromycin B phosphotransferase gene present on the *GFP* reporter construct and the *N. benthamiana F-Box* gene. In the initial experiments, an additional plasmid carrying DRONPA-s was co-infiltrated for normalization (Figures S13b–d, S14b–d, S15b). However, normalization to the hygromycin B phosphotransferase gene present on the plasmid carrying the reporter gene constructs proved more consistently and was then used for quantification.

All constructs were introduced into *Agrobacterium tumefaciens* GV3101 (pMP90) (Koncz & Schell, 1986). *Agrobacterium* cultures were grown to an OD₆₀₀ of 1.0 and independently centrifuged at 4000 g for 15 min, resuspended in infiltration

medium (10 mM MES-KOH pH 5.7, 10 mM MgCl₂, 150 μM acetosyringone) and adjusted to an OD₆₀₀ of 0.1 (Lummer et al., 2011). Cultures expressing the *LHCB1.1-GFP* reporter constructs were mixed with an equal volume of *Agrobacterium* expressing the viral silencing suppressor P19 from tomato bushy stunt virus and varying amounts of the effectors *AtGRP7* or *AtGER3*, and incubated at 4°C for 1 h. Multiple individual leaves (4–5 leaf stage, ca. 10 cm high) were mechanically infused by pressing the tip of the syringe against the abaxial surface of the leaf and applying gentle pressure to the plunger. Two days after infiltration leaves were harvested, quick-frozen in liquid N₂, and ground to a fine powder.

For real-time PCR, total RNA was isolated from each infiltrated leaf using TriReagent, treated with RQ DNase (Promega), and precipitated with isopropanol. Reverse transcription was performed with oligo(dT) primers and transcript levels were determined by real-time PCR as described above, using primers listed in Table S11. Each reporter construct was tested in 4–5 independent experiments. Prior to merging the data for statistical analysis, outliers were eliminated by calculating the fourth spread. Samples deviating more than 1.5 × fourth spread from the mean of the respective treatment were considered outliers. A Kruskal-Wallis test was used to test for statistical significance (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, n.s., not significant).

To control the level of the effector proteins in the transiently transformed *N. benthamiana* leaves via immunoblot analysis, protein extracts were prepared either from powder of each infiltrated leaf or powder pooled from the three leaves infiltrated with the same concentration of *Agrobacterium*. Western blot analysis with anti-peptide antibodies against *AtGRP7* was done as described (Lummer et al., 2011). Expression of *AtGER3* was monitored with an antibody against mustard germin-like protein (Heintzen, Fischer, et al., 1994). Amidoblack staining of the membrane served as loading control.

ACKNOWLEDGEMENTS

We are grateful to Elisabeth Klemme and Kristina Neudorf for expert technical assistance. We thank Dr. Katja Meyer for providing samples from the circadian time course experiment and Sam Asami and Gerd Gemmecker for support with NMR experiments. We acknowledge NMR measurements at the Bavarian NMR Center (BNMRZ) at the Technical University of Munich, Germany. We acknowledge SBGrid and the NMRbox server for providing access to NMR software. We acknowledge Dr. Christa Lanz and Prof. Dr. Detlef Weigel for use of the sequencing facilities at the Genomics Center Max-Planck-Institute for Developmental Biology, Tübingen. We thank Daniel Careno, Dr. Julieta Mateos, and Dr. Christian Heintzen for helpful comments on the manuscript. This work was supported by the DFG through STA653/14-1 to DS. CS acknowledges support by the DFG through SCHM3362/2-1 and SCHM3362/4-1 (project numbers 414704559 and 511886499). MS and KZ acknowledge support by the DFG SPP1935, project number 273941853, and MS is supported by the DFG GRK1721, project number 178567888. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data reported here are deposited in the Gene Expression Omnibus database with accession numbers GSE99427 (iCLIP data), GSE99615 (RNA-seq data), GSE99464 (RIP-seq data), and GSE99616 (SuperSeries). The workflow on how

to determine reproducible binding sites from the AtGRP7-GFP iCLIP samples is described at <https://github.com/malewins/iCLIP-AtGRP7-pipeline>. NMR chemical shifts of RRM have been deposited in the Biological Magnetic Resonance Data Bank (BMRB) with the BMRB ID 52245.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Binding site reproducibility in the biological replicates LL36.

Bar chart displaying the distribution of crosslink events at the binding sites of the five AtGRP7 iCLIP LL36 samples. The applied filtering threshold is highlighted in green. Binding sites with more crosslink events than the threshold in each replicate are considered reproducible binding sites.

Figure S2. Peak distances of LL24 and LL36.

Bar chart (gray) and density plot (black) of AtGRP7 binding site peaks overlapping in the LL24 and LL36 samples. The center position (zero) corresponds to the binding site peaks of the LL36 sample.

Figure S3. Most significantly enriched GO terms (biological process) of AtGRP7 iCLIP targets.

Points indicate the $-\log_{10}$ transformed adjusted *P*-values for corresponding terms. The relative term coverage (point size) is defined by the percentage of intersecting AtGRP7 targets among the total set of genes associated with this term. Only terms with a total count smaller than 1000 entries are displayed.

Figure S4. Overlap between iCLIP targets and RIP-seq targets.

Venn diagram showing the overlap between the AtGRP7 iCLIP targets and the transcripts genes determined by RIP at LL36 and LL24 (Meyer et al., 2017).

Figure S5. Distribution of AtGRP7 binding sites within the transcripts.

(a) Percentage of reproducible binding sites in different transcript regions (dark gray bars) compared to the percentage of cumulative length of the indicated region (light gray bars) based on representative gene models from Araport11.

(b) Mean peak scores determined by PureCLIP from reproducible binding sites mapped to the gene (5' UTR, CDS, intron, and 3' UTR) based on representative gene models from Araport11. Error bars indicate the standard error of PureCLIP scores in the corresponding regions. Peak scores of binding sites mapping to 3' UTRs were significantly higher than ones mapping to CDS (*P*-value < 0.001, Welch Two Sample *t*-test).

(c) Distribution of binding sites in different transcript regions (horizontal lines) in relation to the score of the peak. The colored lines correspond to reproducible binding sites mapping to the 5' UTR (blue), coding region (green), introns (purple), and 3' UTR (red), respectively. The *x*-axis describes the filtering threshold, that is, how the distribution within the transcripts forms after removing the lowest-scored binding sites (e.g., 10%, 20%, or 30%) in comparison to the highest measured peak score of a transcript. The outer left position (0%) considers all binding sites in the transcripts, whereas in contrast the outer right position (100%) considers only the binding site with the highest score for each transcript. The *y*-axis shows the percentage of reproducible binding sites matching the filtering threshold of the *x* axis.

(d) Enrichment of reproducible binding site distribution normalized by the cumulative length of the indicated region in the

representative gene models of Araport11. The dotted line marks an expected distribution according to the cumulative length of the region.

Figure S6. NMR titrations of the AtGRP7 RRM with 7-mer single-stranded DNA oligonucleotide variants.

(a, c) Overlay of ^1H - ^{15}N HSQC spectra of the AtGRP7 RRM with increasing the concentration of respective 7-mer ssDNA oligonucleotides are shown. Black and yellow resonances represent the free and bound forms, respectively.

(b, d) Chemical shift perturbations (CSPs) of the AtGRP7 RRM upon binding of the 7-mer ssDNA oligonucleotide plotted against the individual amino acid residues. Secondary structure elements are indicated on top. Higher CSPs regions, RNP1, RNP2, and non-RNP regions with high CSPs are highlighted with gray color.

(e) Cumulative CSPs of binding of the 7-mer single-stranded DNA oligonucleotides with 'TTT' and 'AAA' motifs binding to the AtGRP7 RRM, based on panels B and D.

Figure S7. NMR titrations of the AtGRP7 RRM with stepwise addition of respective 5-mer ssDNA oligonucleotide variants.

(a, c) Overlay of ^1H - ^{15}N HSQC spectra of the AtGRP7 RRM with stepwise addition of respective 5-mer ssDNA oligonucleotides.

(b, d) CSPs of the AtGRP7 RRM upon binding of the 5-mer single-stranded DNA oligonucleotide plotted against the individual amino acid residues. Secondary structure elements, RNPs, and non-RNP regions with high CSPs are indicated.

(e) Cumulative CSPs of the 5-mer single-stranded DNA oligonucleotides with 'TTT' and 'AAA' motifs binding to the AtGRP7 RRM.

Figure S8. RNA-seq sample correlation.

Pearson correlation matrix of the three biological replicates for wt plants (a), AtGRP7-ox plants (b), and *atgrp7-1 8i* mutant plants (c) at the time points LL18, LL24, LL30, and LL36. Quantified TPM values of expressed genes with CPM > 1 in at least 3 of the 36 RNA-seq sample data sets were considered.

Figure S9. Negative effect of AtGRP7 on direct targets.

(a) Differentially expressed genes (DEGs, $|\log_2 \text{FC}| > 1$, FDR < 0.01) at each time point. The number of up-regulated (blue) and down-regulated (red) genes in *atgrp7-1 8i* and AtGRP7-ox plants relative to wt plants at LL18, LL24, LL30 and LL36 are shown.

(b) Differential expression of iCLIP targets identified by the previous bioinformatics pipeline (Meyer et al., 2017).

(c) Differential expression of the iCLIP targets only identified by the novel pipeline.

(d) Differential expression of all iCLIP targets identified by the novel pipeline (including those also identified by Meyer et al., 2017).

The Venn diagrams show the iCLIP target sets which were intersected with the DEGs.

Figure S10. Expression patterns of selected AtGRP7 target genes.

Col wt plants, the AtGRP7 overexpressing line D, the *atgrp7-1 8i* mutant, and the *atgrp7-1 atgrp8-2* mutant were grown in 12-h light – 12-h dark cycles for 14 days and harvested at 2-h intervals. Transcript levels were quantified by RT-qPCR and normalized to *ACTIN-2*.

A *LHCB1.1*, B *LHCB1.2*, C *LHCB1.3*, D *AtAILP1*, E *ELP*, F *CCL*, G *J8*, H *KCS8*, I *DRM2*, K *RTFL17*, L *BT2*.

Figure S11. *LHCB1.1* reporter gene constructs for transient expression in *N. benthamiana*.

(a) Reporter construct *P_{rom}LHCB_5'UTR_eGFP_3'UTR* consisting of the *LHCB1.1* promoter, 5'UTR, and 3'UTR. The coding sequence was replaced by *eGFP*.

(b) Reporter construct *Prom*LHCB_5'UTR_{mut}eGFP_3'UTR with mutated AtGRP7 binding sites 1–3 in the 5' UTR.

(c) Reporter construct *Prom*LHCB_5'UTR_eGFP_3'UTR_{mut1-4} with mutated AtGRP7 binding sites 1–4 in the 3' UTR. Mutations are denoted by red crosses. The wt sequence is given below the construct with AtGRP7 iCLIP peaks denoted as green bars. Mutated bases are given in red above the sequence. The start and stop codons are marked with red boxes. The polyA site in close proximity to binding site 5 is denoted with a red arrow.

Figure S12. Time course of AtGRP7 and GFP expression after infiltration.

Agrobacteria carrying the *AtGRP7* cDNA under control of the 35S promoter and the duplicated enhancer and the reporter construct *Prom*LHCB_5'UTR_eGFP_3'UTR along with the silencing inhibitor P19 were infiltrated in *N. benthamiana* leaves at an OD₆₀₀ of 0.1 each. Leaves were harvested 0, 1, 2, and 3 days post infiltration.

(a) Immunoblot with antibody against AtGRP7 (top). Amidoblack (AB) staining served as loading control (bottom).

(b) GFP level quantified with RT-qPCR.

Figure S13. Transient expression of *Prom*LHCB_5'UTR_eGFP_3'UTR with the wild type *LHCB1.1* sequence motifs.

Agrobacteria containing the *LHCB1.1*-GFP reporter genes were infiltrated into *N. benthamiana* leaves along with the agrobacteria expressing either the *AtGRP7* cDNA or the *AtGER3* cDNA at corresponding ODs, as indicated.

(a) Schematic representation of *Prom*LHCB_5'UTR_eGFP_3'UTR sequence.

(b–e) Transcript levels were quantified by qPCR, normalized to the *N. benthamiana* F-box gene/*hygromycin phosphotransferase* (*hph*). Transcript levels are expressed relative to *hph*.

Immunoblot analysis of the infiltrated leaves probed with an antibody against AtGRP7 or mustard germin-like protein (Heintzen et al., 1994). Staining of the membrane with amidoblack (AB) served as loading control.

(f–g) Merged data from individual experiments for statistical analysis. Outliers were removed and data was either non-normalized or normalized to the mean GFP expression in the respective '1' sample. A Kruskal–Wallis test was used to test for statistical significance (**P* < 0.05, ***P* < 0.01, ****P* < 0.001, n.s., not significant).

Figure S14. Transient expression of *Prom*LHCB_5'UTR_{mut}eGFP_3'UTR with mutated AtGRP7 iCLIP peaks in the 5' UTR.

(a) Schematic representation of *Prom*LHCB_5'UTR_{mut}eGFP_3'UTR with mutated AtGRP7 iCLIP peaks in the 5' UTR. The wt sequence is given below the construct with the AtGRP7 iCLIP binding sites in bold and underlined. Mutated bases in the reporter construct are given in red below the sequence.

(b–e) Transcript levels were quantified by qPCR, normalized to the *N. benthamiana* F-box gene/*hygromycin phosphotransferase* (*hph*). Transcript levels are expressed relative to *hph*.

Immunoblot analysis of the infiltrated leaves probed with an antibody against AtGRP7 or mustard germin-like protein (Heintzen et al., 1994). Staining of the membrane with amidoblack (AB) served as loading control.

(f,g) Merged data from individual experiments for statistical analysis. Outliers were removed and data was either non-normalized or normalized to the mean GFP expression in the respective '1' sample. A Kruskal–Wallis test was used to test for statistical significance (**P* < 0.05, ***P* < 0.01, ****P* < 0.001, n.s., not significant).

Figure S15. Transient expression of *Prom*LHCB_5'UTR_eGFP_3'UTR_{mut1-4} with four mutated AtGRP7 iCLIP peaks in the 3' UTR.

(a) Schematic representation of *Prom*LHCB_5'UTR_eGFP_3'UTR_{mut1-4} with four mutated AtGRP7 iCLIP peaks in the 3' UTR. The wt sequence is given below the construct with the AtGRP7 iCLIP binding sites in bold and underlined. Mutated bases in the reporter construct are given in red below the sequence.

(b–f) Transcript levels were quantified by qPCR, normalized to the *N. benthamiana* F-box gene/*hygromycin phosphotransferase* (*hph*). Transcript levels are expressed relative to *hph*.

Immunoblot analysis of the infiltrated leaves probed with an antibody against AtGRP7 or mustard germin-like protein (Heintzen et al., 1994). Staining of the membrane with amidoblack (AB) served as loading control.

(g,h) Merged data from individual experiments for statistical analysis. Outliers were removed and data was either non-normalized or normalized to the mean GFP expression in the respective '1' sample. A Kruskal–Wallis test was used to test for statistical significance (**P* < 0.05, ***P* < 0.01, ****P* < 0.001, n.s., not significant).

Figure S16. AtGRP7 crosslinks and reproducible binding sites on *RD29A*.

(a) SEQing tracks of AtGRP7 crosslinks and reproducible binding sites determined at LL36 on *RD29A* (AT5G52310) (top). The data track (black) shows cumulated crosslink data from the improved pipeline. Grey bars below the crosslink track mark the positions of reproducible binding sites. The gene model in block format is depicted at the bottom. Thin bars represent 5' UTR and 3' UTR; thick bars denote exons and lines denote introns. The numbers on the x-axis refer to the chromosomal position.

(b) Magnification of AtGRP7 binding in the 3' UTR. Called peak positions (center of reproducible binding sites) are marked with asterisks. The genomic sequence of the corresponding region is shown above the SEQing track.

Figure S17. Correlation of iCLIP and RIP-seq targets with expression level.

Violin plots with 25%, 75% quantiles, median (black) and mean (gray) of log₂ Transcripts per Million (TPM) values from expressed genes (RNA-seq, gray), AtGRP7 iCLIP (red) and AtGRP7 RIP-seq (orange).

Table S1. Preprocessing of the iCLIP sequencing libraries and mapping statistics.

Table S2. Reproducible binding sites of AtGRP7.

Genomic coordinates of AtGRP7 reproducible binding sites after subtraction of GFP and AtGRP7 R⁴⁹Q-GFP controls. The score columns represent PureCLIP scores separately for the LL24 and LL36 time points.

Table S3. AtGRP7 iCLIP targets.

AtGRP7 iCLIP target genes after subtraction of GFP and AtGRP7 R⁴⁹Q-GFP controls.

The gene feature, in which reproducible binding sites were detected, is marked for each target gene.

Table S4. GO Terms Biological Processes.

Enrichment of GO biological process terms of the 2705 AtGRP7 iCLIP targets. Statistics were generated using the g:GOST web interface from the g:Profiler website (<https://biit.cs.ut.ee/gprofiler/gost>). The relative term coverage was computed by dividing the intersection count by the total Term count.

Table S5. AtGRP7 sequence motifs in protein-coding gene features.

Table S6. Overview of the CSPs observed for DNA oligonucleotides.

Table S7. Preprocessing of the RNA-seq libraries and mapping statistics.

RNA-seq read statistics of the Col-2, *atgrp7-1 8i*, and AtGRP7-ox libraries at the time points LL18, LL24, LL30, and LL36, respectively.

Table S8. Genes differentially expressed in *atgrp7-1 8i* and AtGRP7-ox plants in comparison to wt.

Table S9. Circadian AtGRP7 iCLIP targets.

Comparison of 18 503 transcripts classified as clock-regulated by Romanowski et al. (2020) with the AtGRP7 iCLIP targets also expressed in the Romanowski dataset revealed 1746 iCLIP targets predicted to be clock-regulated.

Table S10. Genes with reciprocal changes in *atgrp7-1 8i* and AtGRP7-ox plants.

Differentially expressed genes (FDR < 0.01) with inverse changes in *atgrp7-1 8i* and AtGRP7-ox plants compared to wt at LL18, LL24, LL30, or LL36, respectively.

Table S11. Oligonucleotides used.

REFERENCES

- Anderson, S.L. & Kay, S.A. (1995) Functional dissection of circadian clock and phytochrome-regulated transcription of the *Arabidopsis* CAB2 gene. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 1500–1504.
- Anderson, S.L., Teakle, G.R., Martino-Catt, S.J. & Kay, S.A. (1994) Circadian clock- and phytochrome-regulated transcription is conferred by a 78 bp cis-acting domain of the *Arabidopsis* CAB2 promoter. *The Plant Journal*, **6**, 457–470.
- Bailey, T.L. (2021) STREME: accurate and versatile sequence motif discovery. *Bioinformatics*, **37**, 2834–2840.
- Bailey, T.L. & Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings/International Conference on Intelligent Systems for Molecular Biology*, **2**, 28–36.
- Busch, A., Brüggemann, M., Ebersberger, S. & Zarnack, K. (2020) iCLIP data analysis: a complete pipeline from sequencing reads to RBP binding sites. *Methods*, **178**, 49–62.
- Careno, D.A., Perez Santangelo, S., Macknight, R.C. & Yanovsky, M.J. (2022) The 5′-3′ mRNA decay pathway modulates the plant circadian network in *Arabidopsis*. *Plant and Cell Physiology*, **63**, 1709–1719.
- Carpenter, C.D., Kreps, J.A. & Simon, A.E. (1994) Genes encoding glycine-rich *Arabidopsis thaliana* proteins with RNA-binding motifs are influenced by cold treatment and an endogenous circadian rhythm. *Plant Physiology*, **104**, 1015–1025.
- Chakrabarti, A., Haberman, N., Praznik, A., Luscombe, N. & Ule, J. (2018) Data science issues in studying protein–RNA interactions with CLIP technologies. *Annual Review of Biomedical Data Science*, **1**, 235–261.
- Chakrabortee, S., Kayatekin, C., Newby, G.A., Mendillo, M.L., Lancaster, A. & Lindquist, S. (2016) Luminidependens (LD) is an *Arabidopsis* protein with prion behavior. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 6065–6070.
- Clery, A., Blatter, M. & Allain, F.H. (2008) RNA recognition motifs: boring? Not quite. *Current Opinion in Structural Biology*, **18**, 290–298.
- Conte, M.R., Grüne, T., Ghuman, J., Kelly, G., Ladas, A., Matthews, S. et al. (2000) Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. *The EMBO Journal*, **19**, 3132–3141.
- Corley, M., Burns, M.C. & Yeo, G.W. (2020) How RNA-binding proteins interact with RNA: molecules and mechanisms. *Molecular Cell*, **78**, 9–29.
- Covington, M.F. & Harmer, S.L. (2007) The circadian clock regulates auxin signaling and responses in *Arabidopsis*. *PLoS Biology*, **5**, e222.
- Dalchau, N., Baek, S.J., Briggs, H.M., Robertson, F.C., Dodd, A.N., Gardner, M.J. et al. (2011) The circadian oscillator gene GIGANTEA mediates a long-term response of the *Arabidopsis thaliana* circadian clock to sucrose. *Proceedings of the National Academy of Sciences of the United States of America*, **22**, 5104–5109.
- Daubner, G.M., Cléry, A. & Allain, F.H.T. (2013) RRM–RNA recognition: NMR or crystallography...and new findings. *Current Opinion in Structural Biology*, **23**, 100–108.
- Dodt, M., Roehr, J.T., Ahmed, R. & Dieterich, C. (2012) FLEXBAR—flexible barcode and adapter processing for next-generation sequencing platforms. *Biology*, **1**, 895–905.
- Haberman, N., Huppertz, I., Attig, J., König, J., Wang, Z., Hauer, C. et al. (2017) Insights into the design and interpretation of iCLIP experiments. *Genome Biology*, **18**, 7.
- Hackmann, C., Korneli, C., Kutyniok, M., Köster, T., Wiedenlubbber, M., Müller, C. et al. (2014) Salicylic acid-dependent and -independent impact of an RNA-binding protein on plant immunity. *Plant, Cell & Environment*, **37**, 696–706.
- Hafner, M., Katsantoni, M., Köster, T., Marks, J., Mukherjee, J., Staiger, D. et al. (2021) CLIP and complementary methods. *Nature Reviews Methods Primers*, **1**, 20.
- Heintzen, C., Fischer, R., Melzer, S., Kappeler, S., Apel, K. & Staiger, D. (1994) Circadian oscillations of a transcript encoding a germin-like protein that is associated with cell walls in young leaves of the long-day plant *Sinapis alba* L. *Plant Physiology*, **106**, 905–915.
- Heintzen, C., Melzer, S., Fischer, R., Kappeler, S., Apel, K. & Staiger, D. (1994) A light- and temperature-entrained circadian clock controls expression of transcripts encoding nuclear proteins with homology to RNA-binding proteins in meristematic tissue. *The Plant Journal*, **5**, 799–813.
- Heintzen, C., Nater, M., Apel, K. & Staiger, D. (1997) AtGRP7, a nuclear RNA-binding protein as a component of a circadian-regulated negative feedback loop in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 8515–8520.
- Helder, S., Blythe, A.J., Bond, C.S. & Mackay, J.P. (2016) Determinants of affinity and specificity in RNA-binding proteins. *Current Opinion in Structural Biology*, **38**, 83–91.
- Hughes, M.E., Hogenesch, J.B. & Kornacker, K. (2010) JTK_CYCLE: an efficient non-parametric algorithm for detecting rhythmic components in genome-scale datasets. *Journal of Biological Rhythms*, **25**, 372–380.
- James, A.B., Monreal, J.A., Nimmo, G.A., Kelly, C.L., Herzyk, P., Jenkins, G.I. et al. (2008) The circadian clock in *Arabidopsis* roots is a simplified slave version of the clock in shoots. *Science*, **322**, 1832–1835.
- Johnson, C.H., Knight, M., Kondo, T., Masson, P., Sedbrook, J., Haley, A. et al. (1995) Circadian oscillations of cytosolic and chloroplastic free calcium in plants. *Science*, **269**, 1863–1865.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O. et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
- Jurus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L.E. et al. (2018) Improvements to the APBS biomolecular solvation software suite. *Protein Science*, **27**, 112–128.
- Kim, J.S., Jung, H.J., Lee, H.J., Kim, K.A., Goh, C.H., Woo, Y. et al. (2008) Glycine-rich RNA-binding protein7 affects abiotic stress responses by regulating stomata opening and closing in *Arabidopsis thaliana*. *The Plant Journal*, **55**, 455–466.
- Kim, J.S., Park, S.J., Kwak, K.J., Kim, Y.O., Kim, J.Y., Song, J. et al. (2007) Cold shock domain proteins and glycine-rich RNA-binding proteins from *Arabidopsis thaliana* can promote the cold adaptation process in *Escherichia coli*. *Nucleic Acids Research*, **35**, 506–516.
- Kim, Y.O., Kim, J.S. & Kang, H. (2005) Cold-inducible zinc finger-containing glycine-rich RNA-binding protein contributes to the enhancement of freezing tolerance in *Arabidopsis thaliana*. *The Plant Journal*, **42**, 890–900.
- Koncz, C. & Schell, J. (1986) The promoter of T_L-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. *Molecular & General Genetics*, **204**, 383–396.
- König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B. et al. (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature Structural & Molecular Biology*, **17**, 909–915.
- Köster, T. & Staiger, D. (2020) Plant individual nucleotide resolution cross-linking and immunoprecipitation to characterize RNA-protein complexes. *Methods in Molecular Biology*, **2166**, 255–268.
- Krakau, S., Richard, H. & Marsico, A. (2017) PureCLIP: capturing target-specific protein–RNA interaction footprints from single-nucleotide CLIP-seq data. *Genome Biology*, **18**, 240.
- Kumar, D., Chugh, J. & Hosur, R.V. (2009) Generation of serine/threonine check points in HN(C)N spectra. *Journal of Chemical Sciences*, **121**, 955–964.
- Law, C.W., Alhamdoosh, M., Su, S., Smyth, G.K. & Ritchie, M.E. (2016) RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research*, **5**, 1408.

- Leutwiler, L.S., Meyerowitz, E.M. & Tobin, E.M. (1986) Structure and expression of three light-harvesting chlorophyll a/b-binding protein genes in *Arabidopsis thaliana*. *Nucleic Acids Research*, **14**, 4051–4064.
- Lewinski, M., Bramkamp, Y., Köster, T. & Staiger, D. (2020) SEQing: web-based visualization of iCLIP and RNA-seq data in an interactive python framework. *BMC Bioinformatics*, **21**, 113.
- Lidder, P., Gutierrez, R.A., Salome, P.A., McClung, C.R. & Green, P.J. (2005) Circadian control of messenger RNA stability. Association with a sequence-specific messenger RNA decay pathway. *Plant Physiology*, **138**, 2374–2385.
- Lummer, M., Humpert, F., Steuwe, C., Schüttel, M., Sauer, M. & Staiger, D. (2011) Reversible photoswitchable DRONPA-s monitors nucleocytoplasmic transport of an RNA-binding protein in transgenic plants. *Traffic*, **12**, 693–702.
- Lunde, B.M., Moore, C. & Varani, G. (2007) RNA-binding proteins: modular design for efficient function. *Nature Reviews. Molecular Cell Biology*, **8**, 479–490.
- Mateos, J.L., de Leone, M.J., Torchio, J., Reichel, M. & Staiger, D. (2018) Beyond transcription: fine-tuning of circadian timekeeping by post-transcriptional regulation. *Genes (Basel)*, **9**, 616.
- McClung, C.R. (2019) The plant circadian oscillator. *Biology*, **8**, 14.
- Membre, N., Berna, A., Neutelings, G., David, A., David, H., Staiger, D. et al. (1997) cDNA sequence, genomic organization and differential expression of three *Arabidopsis* genes for germin/oxalate oxidase-like proteins. *Plant Molecular Biology*, **35**, 459–469.
- Meyer, K., Köster, T., Nolte, C., Weinholdt, C., Lewinski, M., Grosse, I. et al. (2017) Adaptation of iCLIP to plants determines the binding landscape of the clock-regulated RNA-binding protein AtGRP7. *Genome Biology*, **18**, 204.
- Millar, A.J. & Kay, S. (1991) Circadian control of *cab* gene transcription and mRNA accumulation in *Arabidopsis*. *Plant Cell*, **3**, 541–550.
- Nagarajan, V.K., Kukulich, P.M., von Hagel, B. & Green, P.J. (2019) RNA degradomes reveal substrates and importance for dark and nitrogen stress responses of *Arabidopsis* XRN4. *Nucleic Acids Research*, **47**, 9216–9230.
- Noggle, J.H. & Schirmer, R.E. (1971) *The nuclear overhauser effect: chemical applications*. New York, NY-London, UK: Academic Press.
- Nohales, M.A. & Kay, S.A. (2016) Molecular mechanisms at the core of the plant circadian oscillator. *Nature Structural & Molecular Biology*, **23**, 1061–1069.
- Nolte, C. & Staiger, D. (2015) RNA around the clock – regulation at the RNA level in biological timing. *Frontiers in Plant Science*, **6**, 311.
- Panchal, S.C., Bhavesh, N.S. & Hosur, R.V. (2001) Improved 3D triple resonance experiments, HN and HN(C)N, for HN and 15N sequential correlations in (13C, 15N) labeled proteins: application to unfolded proteins. *Journal of Biomolecular NMR*, **20**, 135–147.
- Parker, R. & Song, H. (2004) The enzymes and control of eukaryotic mRNA turnover. *Nature Structural & Molecular Biology*, **11**, 121–127.
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, **14**, 417–419.
- Perez-Santángelo, 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. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 15166–15171.
- Pett, J.P., Westermarck, P.O. & Herzel, H. (2021) Simple kinetic models in molecular chronobiology. *Methods in Molecular Biology*, **2130**, 87–100.
- Quinlan, A.R. & Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
- Roehr, J.T., Dieterich, C. & Reinert, K. (2017) Flexbar 3.0 – SIMD and multi-core parallelization. *Bioinformatics*, **33**, 2941–2942.
- Romanowski, A., Schlaen, R.G., Perez-Santángelo, S., Mancini, E. & Yanovsky, M.J. (2020) Global transcriptome analysis reveals circadian control of splicing events in *Arabidopsis thaliana*. *The Plant Journal*, **103**, 889–902.
- Romanowski, A. & Yanovsky, M.J. (2015) Circadian rhythms and post-transcriptional regulation in higher plants. *Frontiers in Plant Science*, **6**, 437.
- Sattler, M., Schleucher, J. & Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **34**, 93–158.
- Schöning, J.C., Streitner, C., Meyer, I.M., Gao, Y. & Staiger, D. (2008) Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in *Arabidopsis*. *Nucleic Acids Research*, **36**, 6977–6987.
- Schöning, J.C., Streitner, C., Page, D.R., Hennig, S., Uchida, K., Wolf, E. et al. (2007) Autoregulation of the circadian slave oscillator component AtGRP7 and regulation of its targets is impaired by a single RNA recognition motif point mutation. *The Plant Journal*, **52**, 1119–1130.
- Schüttel, M., Schöning, J.C., Doose, S., Neuweiler, H., Peters, E., Staiger, D. et al. (2008) Changes of conformational dynamics of mRNA upon AtGRP7 binding studied by fluorescence correlation spectroscopy. *Journal of the American Chemical Society*, **130**, 9507–9513.
- Schwich, O.D., Blümel, N., Keller, M., Wegener, M., Setty, S.T., Brunstein, M.E. et al. (2021) SRSF3 and SRSF7 modulate 3'UTR length through suppression or activation of proximal polyadenylation sites and regulation of CFIm levels. *Genome Biology*, **22**, 82.
- Shen, Y. & Bax, A. (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of Biomolecular NMR*, **56**, 227–241.
- Simpson, P.J., Monie, T.P., Szendroi, A., Davydova, N., Tyzack, J.K., Conte, M.R. et al. (2004) Structure and RNA interactions of the N-terminal RRM domains of PTB. *Structure*, **12**, 1631–1643.
- Staiger, D. (2001) RNA-binding proteins and circadian rhythms in *Arabidopsis thaliana*. *Philosophical transactions of the Royal Society of London, Series B: Biological sciences*, **356**, 1755–1759.
- Staiger, D., Apel, K. & Trepp, G. (1999) The Atger3 promoter confers circadian clock-regulated transcription with peak expression at the beginning of the night. *Plant Molecular Biology Reporter*, **40**, 873–882.
- Staiger, D. & Green, R. (2011) RNA-based regulation in the plant circadian clock. *Trends in Plant Science*, **16**, 517–523.
- Staiger, D., Zecca, L., Wiecek Kirk, D.A., Apel, K. & Eckstein, L. (2003) The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA. *The Plant Journal*, **33**, 361–371.
- Steffen, A., Elgner, M. & Staiger, D. (2019) Regulation of flowering time by the RNA-binding proteins AtGRP7 and AtGRP8. *Plant and Cell Physiology*, **60**, 2040–2050.
- Stitzinger, S.H., Sohrabi-Jahromi, S. & Söding, J. (2023) Cooperativity boosts affinity and specificity of proteins with multiple RNA-binding domains. *NAR Genomics and Bioinformatics*, **5**, lqad059.
- Streitner, C., Danisman, S., Wehrle, F., Schöning, J.C., Alfano, J.R. & Staiger, D. (2008) The small glycine-rich RNA-binding protein AtGRP7 promotes floral transition in *Arabidopsis thaliana*. *The Plant Journal*, **56**, 239–250.
- Streitner, C., Hennig, L., Korneli, C. & Staiger, D. (2010) Global transcript profiling of transgenic plants constitutively overexpressing the RNA-binding protein AtGRP7. *BMC Plant Biology*, **10**, 221.
- Streitner, C., Köster, T., Simpson, C.G., Shaw, P., Danisman, S., Brown, J.W.S. et al. (2012) An hnRNP-like RNA-binding protein affects alternative splicing by in vivo interaction with target transcripts in *Arabidopsis thaliana*. *Nucleic Acids Research*, **40**, 11240–11255.
- Tenenbaum, J.B., Silva, V.D. & Langford, J.C. (2000) A global geometric framework for nonlinear dimensionality reduction. *Science*, **290**, 2319–2323.
- Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M. et al. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
- Xu, F., Wang, L., Li, Y., Shi, J., Staiger, D., Chen, W. et al. (2022) The receptor kinase FER mediates phase separation of glycine-rich RNA-binding protein 7 to confer temperature resilience in *Arabidopsis*. *bioRxiv*, 2022.2003.2006.483201.
- Zhang, R., Calixto, C.P., Marquez, Y., Venhuizen, P., Tzioutziou, N.A., Guo, W. et al. (2017) A high quality *Arabidopsis* transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Research*, **45**, 5061–5073.