A single *cis*-element that controls cell-type specific expression in *Arabidopsis*

Jana Kneřová[1*], Patrick J. Dickinson[1*], Marek Szecówka[1], Steven J. Burgess[1], Hugh Mulvey[1], Anne-Maarit Bågman[2], Allison Gaudinier[2], Siobhan M. Brady[2] and Julian M. Hibberd[1]

[1]Department of Plant Sciences, Downing Street, University of Cambridge, Cambridge CB2 3EA, UK.
[2]Department of Plant Biology and Genome Center, UC Davis, California, 95616.

JK - j.knerova@gmail.com
PJD - pd373@cam.ac.uk
MS - szecowka@protonmail.com
SJB - sburgess011@gmail.com
HM - hughmulvey@gmail.com
AMB - ambagman@ucdavis.edu
AG - agaudinier@ucdavis.edu
SMB - sbrady@ucdavis.edu
JMH (corresponding) - jmh65@cam.ac.uk

* These authors contributed equally

Short title: Cell-type specific gene expression in leaves.

Keywords: Cell-type specific gene expression, *cis*-element, Arabidopsis, bundle sheath.

One sentence summary: A single *cis*-element that is necessary and sufficient to drive expression in a specific cell-type is identified upstream of the *MYB76* transcription factor.

Abstract

In multicellular organisms the specification of distinct tissues within organs allows compartmentation of complex processes. However, the mechanisms that allow gene expression to be restricted to such tissues are poorly understood. To better understand this process, we focused on bundle sheath expression of the gene encoding the MYB76 transcription factor in *Arabidopsis thaliana*. Functional and computational analyses were combined to identify a seven-nucleotide motif within a DNaseI hypersensitive site in the *MYB76* promoter that is necessary and sufficient to direct gene expression to bundle sheath cells. Thus, combining information from DNaseI hypersensitivity assays with classical truncation analysis allowed the rapid identification of a single *cis*-element that governs cell type-specificity. This motif is conserved in the Brassicaceae, acts to positively regulate gene expression, and is recognised *in planta* by two DREB transcription factors. In contrast to previous studies, these data indicate that the patterning of gene expression to specific cell types can be mediated by relatively simple interactions between *cis*-elements and transcription factors. Moreover, as the element in the *MYB76* promoter is short and can be oligomerized to tune expression levels it is well-suited for use in synthetic biology applications that require tissue specific expression in plants.

Introduction

Multicellular organisms consist of distinct tissues carrying out diverse and specialised functions. These tissues are defined by differences in function associated with their protein content, which in turn is determined by specific patterns of gene expression. Despite the importance of tissue specific gene expression, our understanding of how spatial patterning of gene expression is controlled is poor. In roots of *Arabidopsis thaliana* expression of *SHORTROOT (SHR)* is confined to the vasculature by a complex network of both activators and repressors (Sparks et al., 2016). Regulatory elements that are sufficient to drive hormone-responsive gene expression have also been described. For example, the synthetic promoter DR5 was created by multimerizing a mutated auxin response element and fusing this to the minimal CaMV35S (Ulmasov et al., 1997, 1995), and an oligomerised abscisic acid responsive element reports known centres of ABA signalling in roots (Wu et al., 2018). In both cases, these elements respond to hormonal and developmental triggers through diverse upstream signalling pathways (Wu et al., 2018). In comparison, our understanding of tissue specific gene expression in leaves is rudimentary, with mechanisms limiting gene expression to specific tissues being best understood in C_4 species where photosynthesis is typically compartmentalised between mesophyll and bundle sheath strands. For example, expression of the *Glycine decarboxylase P subunit (GLDPA)* in the bundle sheath and veins of C_4 *Flaveria bidentis* is due to interplay between regulatory regions (Wiludda et al., 2012). Activity of the distal promoter is high but not tissue-specific, however in the presence of a proximal promoter, expression in the bundle sheath is brought about by transcripts derived from the distal promoter being degraded in mesophyll cells through nonsense-mediated RNA decay of incompletely spliced transcripts (Engelmann et al., 2008; Wiludda et al., 2012). Similarly, two submodules in a distal region of the *Phosphoenolpyruvate carboxylaseA1 (PpcA1)* promoter from C_4 *Flaveria trinervia* are sufficient to confer mesophyll specificity that is enhanced by interaction with sequence in the proximal promoter (Gowik et al., 2004; Akyildiz et al., 2007). In addition to promoter sequences, other genic regions contain *cis*-elements that generate tissue-specific gene expression. For example, preferential expression of the *CARBONIC ANHYDRASE2, CARBONIC ANHYDRASE4* and *PYRUVATE,ORTHOPHOSPHATE DIKINASE* genes in mesophyll cells of the C_4 species *Gynandropsis gynandra* is mediated by a nine base pair motif present in both 5' and 3' untranslated regions (Williams et al., 2016). Moreover, preferential expression of *NAD-ME1&2* genes in the bundle sheath of *G. gynandra* is associated with two motifs known as Bundle Sheath Modules (BSM) 1a and 1b that co-operatively restrict gene expression to this tissue. BSM1a and BSM1b represent duons

because they are located in coding sequence and so determine amino acid composition as well as gene expression (Brown et al., 2011; Reyna-Llorens et al., 2018). In summary, tissue-specific expression can be generated through multiple mechanisms. However, single *cis*-elements that generate tissue specific patterning of gene expression via interaction with their cognate transcription factors have not yet been identified.

To better understand the regulation of gene expression in specific cell-types we focussed on the bundle sheath in the C_3 species *A. thaliana*. The bundle sheath represents about 15% of cells in leaves of *A. thaliana* (Kinsman and Pyke, 1998) and has been proposed to play important roles in hydraulic conductance (Shatil-Cohen et al., 2011), transport of metabolites (Leegood, 2008), as well as storage of carbohydrates (Koroleva et al., 1997), ions (Williams et al., 2018) and water (Sage, 2001; Griffiths et al., 2013). A number of findings are also consistent with bundle sheath cells being involved in sulphur metabolism and glucosinolate biosynthesis. First, the promoter of the *SULPHUR TRANSPORTER2.2* generates preferential expression in the bundle sheath sheath (Takahashi et al., 2000; Kirschner et al., 2018) and secondly, compared with the whole leaf, transcripts encoding enyzmes of sulphur metabolism are more abundant on bundle sheath ribosomes (Aubry et al., 2014). Notably, so called S-cells that accumulate high levels of sulphur lie just inside the abaxial bundle sheath and are thought to be specialised for defence against insects penetrating the phloem (Koroleva et al., 2010). It is therefore possible that the bundle sheath synthesises sulphur-rich compounds and transports them for storage in the S-cells. Glucosinolate biosynthesis is regulated by MYB domain transcription factors. One such transcription factor is MYB76, which contains the canonical R2R3 MYB DNA-binding domain (Stracke et al., 2001) and acts with MYB28, MYB29 and three basic helix-loop-helix transcription factors to regulate the methionine-derived glucosinolate biosynthetic pathway (Gigolashvili et al., 2007; Sønderby et al., 2007, 2010; Li et al., 2013; Schweizer et al., 2013; Malitsky et al., 2008). Transcripts of *MYB76* and other MYB domain transcription factors involved in glucosinolate biosynthesis are preferentially associated with bundle sheath ribosomes (Aubry et al., 2014) and the *MYB76* promoter has been reported to generate vascular specific expression (Tamara et al., 2007); however, how their expression is restricted to the vasculature is not known.

As there are few examples of *cis*-elements that are necessary and sufficient for cell type-specific expression in plants, we sought to use *MYB76* as a model for this process. A classical truncation analysis of *MYB76* was combined with computational interrogation of transcription factor binding sites to identify a 250-nucleotide region required for expression in the *A. thaliana* bundle sheath. Within this region a seven base pair motif was identified

that is both necessary and sufficient for patterning gene expression in this cell type. This sequence is bound by two transcription factors in the DREB family. The short *cis*-element from *MYB76* therefore provides insight into regulatory mechanisms that direct gene expression to specific cells of plants, identifies a single DNA motif that could be used to engineer expression of genes required for the efficient C_4 pathway into bundle sheath cells of C_3 species, and provides proof-of-principle that by combining classical truncation and computational analysis, short regions of DNA that are suitable for use in synthetic biology to pattern gene expression in specific cell types can be identified.

Results

The MYB76 *promoter contains a positive regulator directing expression to the bundle sheath*

Expression patterns can be determined by *cis*-elements in promoters, UTRs, exons, introns or downstream 3' regions (Ali and Taylor, 2001; Brown et al., 2011; Kajala et al., 2011; Williams et al., 2016; Gallegos and Rose, 2017). Therefore, to gain insight into the spatial patterning of *MYB76* a translational fusion between the *MYB76* genomic sequence and the *uidA* reporter was generated. Transgenic lines harbouring this genomic fusion showed preferential accumulation of GUS in the bundle sheath (Figure 1A, Supplemental Figure 1). A construct consisting of nucleotides -1725 to +279 relative to the predicted translational start site was also sufficient to direct GUS to bundle sheath strands (Figure 1B, Supplemental Figure 2). Use of the fluorometric 4-MethylUmbelliferyl β-D-Glucuronide (MUG) assay showed that GUS accumulation was lower when nucleotides +280 to +1254 were included (Figures 1A and 1B). We conclude that nucleotides -1725 to +279 of *MYB76* contain one or more elements that are sufficient to generate expression in the bundle sheath. Moreover, the full genomic sequence of *MYB76* contains regulators that quantitatively repress expression, but these do not override the ability of the promoter and 5' end of the gene to generate GUS accumulation in bundle sheath cells.

To investigate the elements driving *MYB76* expression in the bundle sheath, additional 5' deletions were generated (Figures 1C to 1E). Removal of nucleotides -1725 to -1264 did not impact on GUS accumulation in the bundle sheath (Figure 1C, Supplemental Figure 3) however once nucleotides -1264 to -796 were removed GUS accumulation in the bundle sheath was no longer detectable (Figure 1D, Supplemental Figure 4) and removal of another 500 base pairs had no further impact on the spatial pattern of GUS accumulation (Figure 1E, Supplemental Figure 5). Quantification *via* MUG assays (Figures 1A to 1E right) showed that removal of nucleotides -1725 to -1264 reduced accumulation of the reporter, and MUG was no longer detectable once sequence upstream of nucleotide -796 was absent. Overall, these data indicate that *MYB76* contains at least one site within the gene that represses expression, a region between nucleotides -1725 and -1264 that acts to enhance expression, and a region between nucleotides -1264 and -796 that directs expression to the bundle sheath.

A DNaseI hypersensitive site within the MYB76 *promoter is necessary and sufficient to direct expression to the* A. thaliana *bundle sheath*

The DNaseI enzyme preferentially cuts accessible DNA and so can be used to define sequences available for transcription factor binding. To complement our truncation analysis, an existing dataset that defined DNaseI Hypersensitive Sites (DHS) in *A. thaliana* (Zhang et al., 2012) was interrogated. Two DHS were detected upstream of MYB76 in both flower tissue and leaves, but only one (from nucleotides -909 to -654) overlapped with the region required for expression in the bundle sheath (Figure 2A). Removing this entire region abolished accumulation of GUS (Figure 2B, Supplemental Figure 6) whereas fusing it to the minimal CaMV35S promoter led to accumulation of GUS in the bundle sheath (Figure 2C, Supplemental Figure 7). Furthermore, oligomerizing two copies of the DHS upstream of the minimal CaMV35S promoter resulted in very strong accumulation of GUS in the bundle sheath (Figure 2D, Supplemental Figure 8). We conclude that sequence within this DHS is both necessary and sufficient for preferential expression of *MYB76* in the bundle sheath. Combined with the truncation analysis indicating that nucleotides -1264 to -796 are required for expression in the bundle sheath (Figures 1C and 1D), these data show that a positive regulator of bundle sheath expression is located between nucleotides -909 (the start of the DHS) and -796 upstream of *MYB76*.

To provide further insight into the DHS upstream of *MYB76* we assessed the extent to which sequence in this region was conserved both within accessions of *A. thaliana*, and more broadly within the Brassicaceae. Interrogation of the 1001 Genome database (Consortium, 2016) indicated that these accessions possess ten single nucleotide polymorphisms of which only four were located between nucleotides -909 and -796 that are necessary for expression in the bundle sheath (Figure 3A). An alignment of this region to five other species in the Brassicaceae identified two broad regions (from -909 to -870, and from -826 to -792) that appeared reasonably conserved (Figure 3B). Analysis of the whole promoter in these species confirmed this and indicated that other than the ~250 nucleotides directly upstream of the predicted translational start site, these two regions are located in sequence that is more conserved than other parts of the promoter (p-value < 0.0001, Figure 3C).

A motif within the DHS is necessary and sufficient for bundle sheath-preferential expression

Two orthogonal approaches were combined to identify a seven-nucleotide sequence in the DHS upstream of *MYB76* that is sufficient to specify bundle sheath preferential

expression. First, DNaseI-SEQ data can be used to identify specific protein binding footprints known as Digital Genomic Footprints (DGF) that represent short DNA motifs bound by transcription factors (Hesselberth et al., 2009). Re-analysis of the Zhang *et al* (2012) dataset identified two such DGF between nucleotides -909 and -796 upstream of *MYB76* (Figure 4A). Second, phylogenetic footprinting identified two motifs (Figure 3B) shared by *MYB76* as well as the promoters of *SCR, SULTR2.2* and *GLDP* that have previously been reported to generate expression in the *A. thaliana* bundle sheath (Takahashi et al., 2000; Wysocka-Diller et al., 2000; Engelmann et al., 2008; Kirschner et al., 2018). Site directed mutagenesis of the first of these motifs (TGGGCA) had no impact on accumulation of GUS in the bundle sheath (Supplemental Figure 9). The second of these motifs (TGCACCG) overlapped with the CGAAAAACGTGCAC sequence defined by one DGF (Figure 4C) and also coincided with the truncation that abolished expression in the bundle sheath (Figure 1D). Substitution of the motif TGCACCG with random sequence abolished GUS accumulation in all twelve independent lines inspected (Figure 4D, Supplemental Figure 10). To test whether this sequence is sufficient to restrict expression to the bundle sheath, it was combined with ten upstream and ten downstream nucleotides within the context of the endogenous *MYB76* promoter, oligomerized and fused to GUS. This construct generated preferential expression in the bundle sheath (Figure 4E, Supplemental Figure 11). We conclude that this motif, identified from deletion analysis, phylogenetic footprinting and DNaseI sequencing, is both necessary and sufficient to generate bundle sheath specific expression.

DREB2A and DREB26 bind the region of MYB76 that drives expression in the bundle sheath

In order to identify transcription factors that interact with the region of the *MYB76* promoter that is necessary and sufficient for bundle sheath expression a Yeast One-Hybrid screen was performed (Gaudinier et al., 2011, 2017; Reece-Hoyes et al., 2011). This identified thirteen transcription factors that interacted with the DHS (Figure 5A). Binding sites for eight of these have been defined using DAP-seq (O'Malley et al., 2016) and so we scanned the DHS for these motifs using the FIMO tool (Grant et al., 2011). This identified putative binding sites ($p<0.001$) in the DHS for DREB26, DREB2A, DF1 and MYB73 (Figure 5B, Supplemental Table 1). The putative DF1 and MYB73 binding sites are downstream of the sequence required to generate bundle sheath expression in *A. thaliana* (Figure 5B). But, the TOMTOM algorithm (Gupta et al., 2007) indicated that the TGCACCG sequence, shown to be necessary for expression in the bundle sheath (Figure

4D), is similar to the DREB26 and DREB2A binding sites but not to those of DF1 or MYB73 (Figure 5C). In fact, the binding sites of DREB26 and DREB2A overlap with this TGCACCG motif (Figure 5B) which is shared with the *SULTR2;2, SCR* and *GLDP* promoters that also generate bundle sheath expression in *A. thaliana.* Although not linked to cell type-specific gene expression, interaction between DREB26 and the *MYB76* promoter has been reported previously (Li et al., 2013). To confirm that these DREB transcription factors interact with the DHS *in planta* we performed a *trans*-activation assay in *Nicotiana benthamiana*. Co-infiltration of DREB2A and DREB26 with the DHS fused to the GUS reporter drove GUS accumulation in leaves of *N. benthamiana* (Figure 5D). We therefore conclude that both DREB2A and DREB26 can interact with sequence centred on the TGCACCG motif which generates cell type-specific expression in the *A. thaliana* bundle sheath.

Discussion

Deletion and truncation analyses have been used for decades to better understand the role of *cis*-elements in gene expression. However, there are relatively few examples of such elements that have been shown to underpin spatial patterning of gene expression in plants (Akyildiz et al., 2007; Brown et al., 2011; Wiludda et al., 2012; Reyna-Llorens et al., 2018; Williams et al., 2016). DNaseI-seq analysis has recently been adopted to define interactions between DNA sequence and transcription factors on a genome-wide basis but the binding sites of some transcription factors remain undefined, and individual genes are typically bound by multiple transcription factors. It is therefore not always possible to link individual binding sites either to a particular transcription factor or to a specific function. However, through identification of DHS sites that are distant from gene bodies and functional testing in *A. thaliana*, multiple enhancers of gene expression have been identified (Zhu et al., 2015; Sullivan et al., 2014). Using an analogous approach that combines deletion analysis with candidate transcription factor binding sites predicted from DNaseI-seq (Zhang et al., 2012), we were able to identify one sequence motif that is both necessary and sufficient to drive gene expression in the bundle sheath of *A. thaliana*. This finding implies that bringing these two approaches together allows relatively rapid progress in identifying specific *cis*-elements responsible for determining a particular pattern of gene expression.

Within *MYB76*, a single *cis*-element that is necessary and sufficient to direct bundle sheath expression in *A. thaliana* was identified. This contrasts with previous work that indicated more complex regulatory networks underpin tissue specific gene expression. This includes combinatorial interactions between multiple activators and repressors that restrict *SHR* expression to the root vasculature of *A. thaliana* (Sparks et al., 2017), two duons in the coding sequence of *NAD-ME* in *G. gynandra* that act together to repress expression in mesophyll cells (Reyna-Llorens et al., 2018), and in the case of *GLDPA* gene from *F. bidentis*, nonsense-mediated RNA decay of incompletely spliced transcripts from a constitutive distal promoter in mesophyll cells that is mediated by a more proximal promoter (Wiludda et al., 2012). To our knowledge, the simplest regulatory system that has been shown to generate tissue specific patterning of gene expression is the mesophyll enhancing module in the *PpcA1* promoter of C_4 *Flaveria trinervia*. This sequence is made up of two sub-modules located in the promoter (Akyildiz et al., 2007). Thus, in all cases multiple *cis*-elements, presumably bound by multiple transcription factors, are responsible for patterning gene expression. In contrast, the TGCACCG motif that is both necessary and sufficient for expression in the *A. thaliana* bundle sheath is a relatively simple module.

From our data we propose that *MYB76* patterning is mediated by the interaction of TGCACCG motif with transcription factors belonging to the DREB family. We show that DREB2A and DREB26 bind the *MYB76* DHS both in yeast and *in planta* and that the TGCACCG *cis*-element overlaps with putative binding sites for these transcription factors. Although transcripts of *DREB2A* and *DREB26* are not specifically associated with ribosomes in the bundle sheath (Aubry et al., 2014), it is possible that neighbouring sequence to the TGCACCG motif is bound by a different transcription factor which is itself bundle sheath specific and that this facilitates the binding of DREB2A and DREB26 to activate *MYB76* expression *in vivo*. But, no annotated transcription factor binding sites are close to the TGCACCG motif. It is also possible that a related DREB transcription factor with similar binding specificity to DREB2A and DREB26 that is specific to the *A. thaliana* bundle sheath activates *MYB76* in this cell type. Of the 57 annotated DREB transcription factors, six are preferentially expressed in the bundle sheath (log2 BS/35s > 1, PPDE > 0.9) (Supplemental Data 1) (Aubry et al., 2014). However, none of these six transcription factors were identified in the yeast one hybrid screen, and even if they did bind *in vivo*, because DREB2A and DREB26 transcripts both accumulate in mesophyll cells, it is not clear how this would lead to bundle sheath specific expression of *MYB76*. An alternate mechanism that might explain our findings is that binding of DREB2A and/or DREB26 patterns *MYB76* expression directly because they are limited to the bundle sheath by post-transcriptional and/or post-translational mechanisms. In fact, post-transcriptional and post-translational regulation has been reported for DREB2A (Agarwal et al., 2017) with post-transcriptional regulation by alternative splicing shown in grasses (Matsukura et al., 2010; Egawa et al., 2006; Feng et al., 2007) and *A. thaliana* (Vainonen et al., 2012). Moreover, whilst overexpression of *DREB2A* in *A. thaliana* does not affect the expression of target genes (Liu et al., 1998) an isoform lacking key phosphorylation sites activates the majority of DREB2A target genes (Sakuma et al., 2006). As far as we are aware there are no reports of post-translational regulation of DREB26 activity and this may be an interesting area for future research. As the evidence above suggests that DREB2A and DREB26 bind the *MYB76* DHS to activate expression and they have binding sites overlapping the TGCACCG motif, we favour a model where DREB2A and/or DREB26 are specifically activated post-transcriptionally or post-translationally in the bundle sheath.

The comparatively simple architecture associated with the TGCACCG motif is relevant to bioengineering and use in synthetic biology applications. Generating synthetic promoters is one of the major requirements for synthetic biology (Dey et al., 2015). Short synthetic promoters have a number of advantages over the long promoter fragments

currently available to direct gene expression to bundle sheath cells (Takahashi et al., 2000; Wysocka-Diller et al., 2000). These include reducing the likelihood of homology-based gene silencing if used more than once in any construct (Bhullar et al., 2003), and decreasing the chances of leakiness or off-target gene expression associated with use of full-length promoter fragments (Hernandez-Garcia and Finer, 2014). Oligomerization of *cis*-elements to achieve higher expression levels is a common strategy when creating synthetic promoters (Dey et al., 2015). As the *MYB76* DHS is short and can be oligomerized to tune expression levels, it appears to be particularly suitable for use in synthetic promoters.

Overall, we conclude that in addition to distant DHS from gene bodies being excellent candidates for enhancer elements (Zhu et al., 2015), for genes that are preferentially expressed in specific cell types, DHS also represent good candidates for *cis*-elements controlling the patterning of gene expression. Moreover, as with hormone responsiveness (Ulmasov et al., 1997, 1995; Wu et al., 2018), our data indicate that cell-type specific gene expression can be mediated by a simple element in *cis*, and that the *MYB76* DHS is suitable to drive or manipulate gene expression in the bundle sheath. Uses could include the targeted manipulation of storage carbohydrates and ions (Koroleva et al., 1997; Williams et al., 2018), the control of hydraulic conductance (Sage 2001; Griffiths et al., 2013; Shatil-Cohen et al., 2011), the transport of metabolites in and out of veins (Leegood, 2008), responses to high light episodes (Fryer et al., 2003), and improvements to photosynthesis (Wang et al., 2017) in these cells.

Methods

Plant Growth

Seed of *A. thaliana* was washed in 70% (v/v) ethanol for 5 minutes, sterilised in a solution of 10% (v/v) Sodium Hypochlorite and 0.1% (v/v) Tween20 for 20 minutes and rinsed five times in sterile distilled water prior to being selected on 0.5% (w/v) Murrasige & Skoog medium (pH 5.8), 1% (w/v) agar, 50mg l^{-1} Kanamycin and 100mg l^{-1} Timentin. After 3 days of stratification in the dark at 4°C, tissue culture plates were transferred to a 16 hour photoperiod growth chamber with a light intensity of 200 µmol m^{-2} s^{-1} photon flux density, 65% relative humidity and a temperature cycle of 24°C (day) and 20°C (night). After 12 days, transformed seedlings were transferred onto 1:1 Levington M3 high nutrient compost and Sinclair fine Vermiculite soil mixture and grown for another 2-3 weeks before analysis. *N. benthamiana* plants used for transient assays were grown from seed in pots containing the same soil mixture in a 16 hour photoperiod, at 200 µmol m^{-2} s^{-1} photon flux density, 60% relative humidity and 22°C.

Generation of constructs, stable plant transformation and computational analysis

The full length *MYB76* gene as well as the promoter alone were amplified from *A. thaliana Col-0* genomic DNA and then fused to *uidA*. The minimal CaMV35S promoter was synthesised and fused to *MYB76 DHS* by polymerase chain reaction (PCR). Deletion of the DHS within the promoter was achieved by PCR fusion of the 5' end of the promoter with the 3' end of the promoter prior to being cloned into the pENTR/D TOPO vector. Each forward primer contained a CACC overhang to ensure directional cloning. A Gateway LR reaction was performed to transfer the relevant inserts into a modified pGWB3 vector (Nakagawa et al., 2007) that contained an intron within the *uidA* sequence. The *MYB76* genomic *DNA::uidA* and 2xDHSCaMV35SMin::uidA constructs were made using Golden Gate technology (Weber et al., 2011). To mutate the TGCACCG motif within the *MYB76* promoter the QuikChange Lightning Site-Directed Mutagenesis (Agilent Technologies) was used. All constructs were then placed into *Agrobacterium tumefaciens* strain GV3101 and introduced into *A. thaliana* Col-0 by floral dipping (Clough and Bent, 1998).

The MEME tool from The Multiple Em for Motif Elucidation (MEME) suite v.4.8.1. (Bailey et al., 2009) was applied to search for conserved motifs within promoter sequences of genes expressed in the *A. thaliana* bundle sheath. Maximum length of the motif was set to 8 nucleotides, both strands of the sequence were searched and each motif had to be present in every sequence. The alignment of the positive regulatory region amongst the Brassicaceae was generated using the MUSCLE algorithm (Edgar, 2004) and edited

manually. The online tool Evolutionary Analysis of Regulatory Sequences (EARS) (Emma et al., 2010) was used, with window size set to 100 nucleotides and the significance threshold set at p = 0.0001 to predict regulatory regions within non-coding sequences of *MYB76* homologues in Brassicaceae.

Histochemical analysis of GUS localisation and quantitative analysis of MUG

GUS staining of at least nine T1 plants of each *uidA* fusion construct was performed according to (Jefferson et al., 1987). Analysis of multiple T1 plants allows position effects associated with transgene insertion to be averaged out. The staining solution contained 0.1 M Na_2HPO_4 pH 7.0, 2 mM Potassium ferricyanide, 2 mM Potassium ferrocyanide, 10 mM EDTA pH 8.0, 0.06% (v/v) Triton X-100 and 0.5 mg ml^{-1} X-gluc. Leaves from three-week old plants were vacuum-infiltrated three times in GUS solution for one minute and then incubated at 37°C for between 3 and 72 hrs depending on the strength of the promoter being assessed. Next, stained samples were fixed in 3:1 (v/v) ethanol:acetic acid for 30 minutes at room temperature, cleared in 70% (v/v) ethanol at 37°C and then placed in 5 M NaOH for 2 hrs. Samples were stored in 70% (v/v) ethanol at 4°C. To quantify reporter accumulation from each promoter a quantitative assay that assesses the rate of MUG conversion to 4-methylumbulliferone (MU) was performed (Jefferson et al., 1987) on between 10 and 25 lines. Tissue was frozen in liquid nitrogen, homogenised and soluble protein extracted in 5 volumes of Protein extraction buffer (1 mM $MgCl_2$, 100 mM NaCl, 50 mM Tris pH7.8). 15 µl of protein extract was incubated with 60 µl of MUG at 37 ºC for one, two, three and four hours respectively. The reaction was stopped after each time point by addition of 75 µl 200 mM anhydrous sodium carbonate. GUS activity was analysed via measurements of fluorescence of MU at 455 nm after excitation at 365 nm. The concentration of MU/unit fluoresence in each sample was interpolated using a concentration gradient of MU over a linear range.

Yeast One-Hybrid analysis and interaction assays in *Nicotiana benthamiana*

Regions screened for transcription factor binding via Yeast One-Hybrid were first inserted into pENTR 5'TOPO TA entry vector (Thermofisher) and subsequently placed into the pMW2 and pMW3 destination vectors containing *Histidine* and *β-Galactosidase* marker genes respectively (Deplancke et al., 2006). The enhanced Yeast One-Hybrid screen against a complete collection of 2000 *A. thaliana* transcription factors was undertaken as described previously (Gaudinier et al., 2017, 2011; Pruneda-Paz et al., 2014).

To test interactions between promoter regions and transcription factors *in planta* transient infiltration of *N. benthamiana* was performed. Overnight cultures of *Agrobacterium tumefaciens* strain GV3101 carrying a promoter::*uidA* fusion, as well as a transcription factor under control of the CaMV35S promoter and the *P19* gene to suppress RNA silencing activity were pelleted and then re-suspended in infiltration media (10 mM $MgCl_2$, 10 mM MES-KOH pH 5.6 and 150 µM acetosyringone). The optical density of cultures was assessed at 600 nm and adjusted to 0.3 for promoter constructs, 1.0-2.0 for transcription factors (Ma et al., 2013) and 0.5 for the *P19* construct. Cultures were incubated at 28˚C for two to four hours and then mixed in 1:1:1 ratio. The abaxial side of leaves from three-week old plants were inoculated with a 1 ml syringe. Leaves were analysed 72 hrs after inoculation.

Supplemental Legends

Supplemental Figure 1: The genomic sequence of MYB76 fused to GUS generates expression in the bundle sheath of Arabidopsis. Images from eleven independent transgenic lines Leaves were stained for 72 hours except line 11 which was stained for 48hrs. Scale bars represent 100 µm.

Supplemental Figure 2: Nucleotides -1725 to +279 relative to the predicted translational start site of *MYB76* generate preferential expression in the bundle sheath. Images from twelve independent transgenic lines. Leaves were stained for 30 hours. Scale bars represent 100 µm.

Supplemental Figure 3: Nucleotides -1264 to +279 relative to the predicted translational start site of *MYB76* generate preferential expression in the bundle sheath. Images from twelve independent transgenic lines. Leaves were stained for 48 hours. Scale bars represent 100 µm.

Supplemental Figure 4: Nucleotides -796bp to +279 relative to the predicted translational start site of *MYB76* do not generate preferential expression in the bundle sheath. Images from ten independent transgenic lines. Leaves were stained for 30 hours. Scale bars represent 100 µm.

Supplemental Figure 5: Nucleotides -294bp to +279 relative to the predicted translational start site of *MYB76* do not generate preferential expression in the bundle sheath. Images from ten independent transgenic lines. Leaves were stained for 30 hours. Scale bars represent 100 µm.

Supplemental Figure 6: Deleting the *MYB76* DHS abolishes accumulation of GUS in the bundle sheath. Images from five independent transgenic lines. Leaves were stained for 48 hours. Scale bars represent 100 µm.

Supplemental Figure 7: The *MYB76* DHS combined with the minimal 35SCaMV promoter generates preferential expression in the bundle sheath. Images from twelve independent transgenic lines. Leaves were stained for 72 hours. Scale bars represent 100 µm.

Supplemental Figure 8: Oligomerizing the *MYB76* DHS combined with the minimal 35SCaMV promoter generates strong bundle sheath preferential expression. Images from twelve independent transgenic lines. Leaves were stained for 3 hours. Scale bars represent 100 µm.

Supplemental Figure 9: Mutation of the TGGGCA motif does not abolish accumulation of GUS from the bundle sheath. Images from twelve independent transgenic lines. Leaves were stained for 48 hours. Scale bars represent 100 µm.

Supplemental Figure 10: Mutation of the TGCACCG motif leads to loss of GUS in the bundle sheath. Images from twelve independent transgenic lines. Leaves were stained for 48 hours. Scale bars represent 100 µm.

Supplemental Figure 11: Two copies of the TGCACCG motif combined with ten upstream and ten downstream nucleotides within the context of the native *MYB76* promoter fused to the minimal 35SCaMV promoter generate preferential expression in the bundle sheath. Images from seven independent transgenic lines. Leaves were stained for 30 hours. Scale bars represent 100 µm.

Acknowledgements: JK was supported a Derek Brewer Studentship from Emmanuel College Cambridge, PJD by Advanced ERC Grant 694733 Revolution, and MS by BBSRC grant RG75840, SJB was supported by the 3to4 grant from the EU and BB/I002243 from the BBSRC to JMH. SB is funded by an HHMI Faculty Scholar Fellowship.

Contributions: JK, PJD, MS, SJB and HM conducted the analysis *in planta*. A-M B and AG conducted yeast-one hybrid. JK, PJD, SMB and JMH designed and wrote the manuscript and prepared the figures.

References

Agarwal, P.K., Gupta, K., Lopato, S., and Agarwal, P. (2017). Dehydration responsive element binding transcription factors and their applications for the engineering of stress tolerance. J. Exp. Bot. 68: 2135–2148.

Akyildiz, M., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., and Westhoff, P. (2007). Evolution and function of a *cis*-regulatory module for mesophyll-specific gene expression in the C_4 dicot *Flaveria trinervia*. Plant Cell 19: 3391–3402.

Ali, S. and Taylor, W.C. (2001). The 3' non-coding region of a C_4 photosynthesis gene increases transgene expression when combined with heterologous promoters. Plant Mol. Biol. 46: 325–333.

Aubry, S., Smith-Unna, R.D., Boursnell, C.M., Kopriva, S., and Hibberd, J.M. (2014). Transcript residency on ribosomes reveals a key role for the *Arabidopsis thaliana* bundle sheath in sulfur and glucosinolate metabolism. Plant J. 78: 659–673.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME Suite: tools for motif discovery and searching. Nucleic Acids Res. 37: W202–W208.

Bhullar, S., Chakravarthy, S., Advani, S., Datta, S., Pental, D., and Burma, P.K. (2003). Strategies for Development of Functionally Equivalent Promoters with Minimum Sequence Homology for Transgene Expression in Plants: cis-Elements in a Novel DNA Context versus Domain Swapping. Plant Physiol. 132: 988 LP-998.

Brown, N.J., Newell, C.A., Stanley, S., Chen, J.E., Perrin, A.J., Kajala, K., and Hibberd, J.M. (2011). Independent and Parallel Recruitment of Preexisting Mechanisms Underlying C_4 Photosynthesis. Science. 331: 1436–1439.

Clough, S.J. and Bent, A.F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J. 16: 735–743.

Consortium, 1001 Genomes (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell 166: 481–491.

Deplancke, B. et al. (2006). A Gene-Centered *C. elegans* Protein-DNA Interaction Network. Cell 125: 1193–1205.

Dey, N., Sarkar, S., Acharya, S., and Maiti, I.B. (2015). Synthetic promoters *in planta*. Planta 242: 1077–1094.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–1797.

Egawa, C., Kobayashi, F., Ishibashi, M., Nakamura, T., Nakamura, C., and Takumi, S. (2006). Differential regulation of transcript accumulation and alternative splicing of a

DREB2 homolog under abiotic stress conditions in common wheat. Genes Genet. Syst. 81: 77–91.

Emma, P., Peter, K., Alexander, T., Isabelle, C., and Sascha, O. (2010). Evolutionary analysis of regulatory sequences (EARS) in plants. Plant J. 64: 165–176.

Engelmann, S., Wiludda, C., Burscheidt, J., Gowik, U., Schlue, U., Koczor, M., Streubel, M., Cossu, R., Bauwe, H., and Westhoff, P. (2008). The gene for the P-subunit of glycine decarboxylase from the C_4 species *Flaveria trinervia*: analysis of transcriptional control in transgenic *Flaveria bidentis* (C_4) and *Arabidopsis* (C_3). Plant Physiol. 146: 1773–1785.

Feng, Q., Masayuki, K., Yoh, S., Kyonoshin, M., Yuriko, O., Phan, T.L.-S., Kazuo, S., and Kazuko, Y.-S. (2007). Regulation and functional analysis of *ZmDREB2A* in response to drought and heat stresses in *Zea mays* L. Plant J. 50: 54–69.

Fryer, M.J., Ball, L., Oxborough, K., Karpinski, S., Mullineaux, P.M., and Baker, N.R. (2003). Control of Ascorbate Peroxidase 2 expression by hydrogen peroxide and leaf water status during excess light stress reveals a functional organisation of *Arabidopsis* leaves. Plant J 33: 691–705.

Gallegos, J.E. and Rose, A.B. (2017). Intron DNA Sequences Can Be More Important Than the Proximal Promoter in Determining the Site of Transcript Initiation. Plant Cell.

Gaudinier, A. et al. (2011). Enhanced Y1H assays for Arabidopsis. Nat. Methods 8: 10.1038/nmeth.1750.

Gaudinier, A., Tang, M., Bågman, A.-M., and Brady, S.M. (2017). Identification of Protein–DNA Interactions Using Enhanced Yeast One-Hybrid Assays and a Semiautomated Approach BT - Plant Genomics: Methods and Protocols. In W. Busch, ed (Springer New York: New York, NY), pp. 187–215.

Gigolashvili, T., Yatusevich, R., Berger, B., Müller, C., and Flügge, U. (2007). The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in Arabidopsis thaliana. Plant J. 51: 247–261.

Gowik, U., Burscheidt, J., Akyildiz, M., Schlue, U., Koczor, M., Streubel, M., and Westhoff, P. (2004). *cis*-Regulatory elements for mesophyll-specific gene expression in the C_4 plant *Flaveria trinervia*, the promoter of the C_4 phosphoenolpyruvate carboxylase gene. Plant Cell 16: 1077–1090.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinforma. 27: 1017–1018.

Griffiths, H., Weller, G., Toy, L.F., and Dennis, R.J. (2013). You're so vein: bundle sheath physiology, phylogeny and evolution in C_3 and C_4 plants. Plant Cell Env. 36: 249–261.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol. 8: R24.

Hernandez-Garcia, C.M. and Finer, J.J. (2014). Identification and validation of promoters and cis-acting regulatory elements. Plant Sci. 217–218: 109–119.

Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S., Fields, S., and Stamatoyannopoulos, J. a (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat. Methods 6: 283–9.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: ß-Glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. 6: 3901–3907.

Kajala, K., Williams, B.P., Brown, N.J., Taylor, L.E., and Hibberd, J.M. (2011). Multiple Arabidopsis genes primed for direct recruitment into C_4 photosynthesis. Plant J. 69: 47–56.

Kinsman, E.A. and Pyke, K.A. (1998). Bundle sheath cells and cell-specific plastid development in Arabidopsis leaves. Development 125: 1815–1822.

Kirschner, S., Woodfield, H., Prusko, K., Koczor, M., Gowik, U., Hibberd, J.M., and Westhoff, P. (2018). Expression of SULTR2;2, encoding a low-affinity sulphur transporter, in the Arabidopsis bundle sheath and vein cells is mediated by a positive regulator. J. Exp. Bot. http://dx.doi.org/10.1093/jxb/ery263.

Koroleva, O.A., Gibson, T.M., Cramer, R., and Stain, C. (2010). Glucosinolate-accumulating S-cells in Arabidopsis leaves and flower stalks undergo programmed cell death at early stages of differentiation. Plant J. 64: 456–469.

Koroleva, O., Farrar, J., Tomos, A., and Pollock, C. (1997). Patterns of solute in individual mesophyll, bundle sheath and epidermal cells of barley leaves induced to accumulate carbohydrate. New Phytol. 136: 97–104.

Leegood, R.C. (2008). Roles of the bundle sheath cells in leaves of C_3 plants. J. Exp. Bot. 59: 1663–1673.

Li, Y., Sawada, Y., Hirai, A., Sato, M., Kuwahara, A., Yan, X., and Hirai, M.Y. (2013). Novel Insights Into the Function of Arabidopsis R2R3-MYB Transcription Factors Regulating Aliphatic Glucosinolate Biosynthesis. Plant Cell Physiol. 54: 1335–1344.

Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998). Two Transcription Factors, DREB1 and DREB2, with an EREBP/AP2 DNA Binding Domain Separate Two Cellular Signal Transduction Pathways in Drought- and Low-Temperature-Responsive Gene Expression,

Respectively, in Arabidopsis. Plant Cell 10: 1391 LP-1406.

Ma, S., Shah, S., Bohnert, H.J., Snyder, M., and Dinesh-Kumar, S.P. (2013). Incorporating Motif Analysis into Gene Co-expression Networks Reveals Novel Modular Expression Pattern and New Signaling Pathways. PLoS Genet 9: e1003840.

Malitsky, S., Blum, E., Less, H., Venger, I., Elbaz, M., Morin, S., Eshed, Y., and Aharoni, A. (2008). The Transcript and Metabolite Networks Affected by the Two Clades of Arabidopsis Glucosinolate Biosynthesis Regulators. Plant Physiol. 148: 2021–2049.

Matsukura, S., Mizoi, J., Yoshida, T., Todaka, D., Ito, Y., Maruyama, K., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2010). Comprehensive analysis of rice DREB2-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. Mol. Genet. Genomics 283: 185–196.

Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., and Kimura, T. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. J Biosci Bioeng 104: 34–41.

O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell 165: 1280–1292.

Pruneda-Paz, J.L., Breton, G., Nagel, D.H., Kang, S.E., Bonaldi, K., Doherty, C.J., Ravelo, S., Galli, M., Ecker, J.R., and Kay, S.A. (2014). A Genome-Scale Resource for the Functional Characterization of Arabidopsis Transcription Factors. Cell Rep. 8: 622–632.

Reece-Hoyes, J.S., Diallo, A., Lajoie, B., Kent, A., Shrestha, S., Kadreppa, S., Pesyna, C., Dekker, J., Myers, C.L., and Walhout, A.J.M. (2011). Enhanced yeast one-hybrid (eY1H) assays for high-throughput gene-centered regulatory network mapping. Nat. Methods 8: 1059–1064.

Reyna-Llorens, I., Burgess, S.J., Reeves, G., Singh, P., Stevenson, S.R., Williams, B.P., Stanley, S., and Hibberd, J.M. (2018). Ancient duons may underpin spatial patterning of gene expression in C_4 leaves. Proc. Natl. Acad. Sci.: doi.org/10.1073/pnas.1720576115.

Sage RF. Environmental and evolutionary preconditions for the origin and diversification of the C4 photosynthetic syndrome. Plant Biol. 2001;3:202--13.

Sakuma, Y., Maruyama, K., Osakabe, Y., Qin, F., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2006). Functional Analysis of an Arabidopsis Transcription Factor, DREB2A, Involved in Drought-Responsive Gene Expression. Plant Cell 18: 1292–

1309.

Schweizer, F., Fernández-Calvo, P., Zander, M., Diez-Diaz, M., Fonseca, S., Glauser, G., Lewsey, M.G., Ecker, J.R., Solano, R., and Reymond, P. (2013). Arabidopsis Basic Helix-Loop-Helix Transcription Factors MYC2, MYC3, and MYC4 Regulate Glucosinolate Biosynthesis, Insect Performance, and Feeding Behavior. Plant Cell 25: 3117–3132.

Shatil-Cohen, A., Attia, Z., and Moshelion, M. (2011). Bundle-sheath cell regulation of xylem-mesophyll water transport via aquaporins under drought stress: a target of xylem-borne ABA? Plant J. 67: 72–80.

Sønderby, I.E., Burow, M., Rowe, H.C., Kliebenstein, D.J., and Halkier, B.A. (2010). A Complex Interplay of Three R2R3 MYB Transcription Factors Determines the Profile of Aliphatic Glucosinolates in Arabidopsis. Plant Physiol. 153: 348–363.

Sønderby, I.E., Hansen, B.G., Bjarnholt, N., Ticconi, C., Halkier, B.A., and Kliebenstein, D.J. (2007). A Systems Biology Approach Identifies a R2R3 MYB Gene Subfamily with Distinct and Overlapping Functions in Regulation of Aliphatic Glucosinolates. PLoS One 2: e1322.

Sparks, E.E. et al. (2017). Establishment of Expression in the SHORTROOT-SCARECROW Transcriptional Cascade through Opposing Activities of Both Activators and Repressors. Dev. Cell 39: 585–596.

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in Arabidopsis thaliana. Curr Opin Plant Biol 4: 447–56.

Sullivan, A.M. et al. (2014). Mapping and dynamics of regulatory DNA and transcription factor networks in *A. thaliana*. Cell Rep. 8: 2015–2030.

Takahashi, H., Watanabe-Takahashi, A., Smith, F.W., Blake-Kalff, M., HawkesfordM.J., and Saito, K. (2000). The roles of three functional sulphate transporters involved in uptake and translocation of sulphate in Arabidopsis thaliana. Plant J. 23: 171–182.

Tamara, G., Martin, E., Ruslan, Y., Caroline, M., and Ulf-Ingo, F. (2007). HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in Arabidopsis thaliana. New Phytol. 177: 627–642.

Ulmasov, T., Liu, Z.B., Hagen, G., and Guilfoyle, T.J. (1995). Composite structure of auxin response elements. Plant Cell 7: 1611 LP-1623.

Ulmasov, T., Murfett, J., Hagen, G., and Guilfoyle, T.J. (1997). Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. Plant Cell 9: 1963 LP-1971.

637 Vainonen, J.P., Jaspers, P., Wrzaczek, M., Lamminmäki, A., Reddy, R.A., Vaahtera, L.,
638 Brosché, M., and Kangasjärvi, J. (2012). RCD1–DREB2A interaction in leaf
639 senescence and stress responses in *Arabidopsis thaliana.* Biochem. J. 442: 573 LP-
640 581.
641 Wang, P., Khoshravesh, R., Karki, S., Tapia, R., Balahadia, C.P., Bandyopadhyay, A.,
642 Quick, W.P., Furbank, R., Sage, T.L., and Langdale, J.A. (2017). Re-creation of a Key
643 Step in the Evolutionary Switch from C_3 to C_4 Leaf Anatomy. Curr. Biol. 27: 3278–
644 3287.e6.
645 Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A Modular
646 Cloning System for Standardized Assembly of Multigene Constructs. PLoS One 6:
647 e16765.
648 Williams, B.P., Burgess, S.J., Reyna-Llorens, I., Knerova, J., Aubry, S., Stanley, S., and
649 Hibberd, J.M. (2016). An untranslated *cis*-element regulates the accumulation of
650 multiple C_4 enzymes in *Gynandropsis gynandra* mesophyll cells. Plant Cell 28: 454–
651 465.
652 Williams, M., Thomas, B., Farrar, J., and Pollock, C. (2018). Visualizing the distribution of
653 elements within barley leaves by energy dispersive X-ray image maps (EDX maps).
654 New Phytol. 125: 367–372.
655 Wiludda, C., Schulze, S., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., Bauwe, H.,
656 and Westhoff, P. (2012). Regulation of the photorespiratory *GLDPA* gene in C_4
657 *Flaveria*: an intricate interplay of transcriptional and posttranscriptional processes.
658 Plant Cell 24: 137–151.
659 Wu, R., Duan, L., Pruneda-Paz, J., Oh, D.-H., Pound, M.P., Kay, S.A., and Dinneny, J.R.
660 (2018). The 6xABRE synthetic promoter enables the spatiotemporal analysis of ABA-
661 mediated transcriptional regulation. Plant Physiol.
662 Wysocka-Diller, J.W., Helariutta, Y., Fukaki, H., Malamy, J.E., and Benfey, P.N. (2000).
663 Molecular analysis of SCARECROW function reveals a radial patterning mechanism
664 common to root and shoot. Development 127: 595–603.
665 Zhang, T., Marand, A.P., and Jiang, J. (2016). PlantDHS: a database for DNase I
666 hypersensitive sites in plants. Nucleic Acids Res. 44: D1148–D1153.
667 Zhang, W., Zhang, T., Wu, Y., and Jiang, J. (2012). Genome-Wide Identification of
668 Regulatory DNA Elements and Protein-Binding Footprints Using Signatures of Open
669 Chromatin in Arabidopsis. Plant Cell Online 24: 2719–2731.
670 Zhu, B., Zhang, W., Zhang, T., Liu, B., and Jiang, J. (2015). Genome-Wide Prediction and
671 Validation of Intergenic Enhancers in Arabidopsis Using Open Chromatin Signatures.

672 Plant Cell 27: 2415 LP-2426.

673 **Figure legends**

674 **Figure 1: The *MYB76* promoter contains a region necessary for bundle sheath**
675 **specific expression.** Schematics of deletion series (left), representative images of GUS in
676 *A. thaliana* leaves (centre) and quantification of MUG activity via the flourometric assay
677 (right). (A) The *MYB76* genomic sequence. (B) The *MYB76* promoter including the first two
678 exons and the first intron fused to the *uidA*. (C-E) A region between -1264 and -796bp is
679 responsible for GUS accumulation in the bundle sheath. Staining times are given in the top
680 right corner of each leaf image. The fluorometric MUG assay shows quantitative
681 repressors and enhancers are located in the gene and in the promoter respectively (A-C
682 right). X-axis shows GUS activity and individual biological replicates are ordered randomly
683 on the y-axis. n, number of replicates; *, p value < 0.05; ** p value < 0.001. Scale bars
684 represent 100µm.

685

686 **Figure 2: A DNaseI Hypersensitive Site in *pMYB76* is sufficient for expression in the**
687 **bundle sheath.** (A) The *MYB76* promoter contains a DNaseI Hypersensitive Site (DHS)
688 located between nucleotides -909 to -654. DHS shown as the DHS score (Zhang et al.,
689 2016) from flower buds (top) and leaves (bottom). Data from Zhang et al., (2012) and
690 visualisd with the IGV browser. (B) Deletion of the DHS abolishes GUS accumulation in
691 the bundle sheath. (C) The DHS fused to the minimal CaMV35S promoter is sufficient to
692 direct expression in the bundle sheath. (D) Oligomerising the DHS increases GUS
693 accumulation in the bundle sheath. Staining times are given in the top right corner of each
694 image. Scale bars represent 100µm.

695

696 **Figure 3: The DNaseI Hypersensitive Site is conserved in the Brassicaceae.** (A)
697 Single nucleotide polymorphisms (red) in the DHS region from 1135 accessions of *A.*
698 *thaliana*. The region necessary for bundle sheath expression is underlined. (B) Alignment
699 of nucleotides -909 to -796 of *MYB76* from *A. thaliana* to five additional species of the
700 Brassicaceae. (C) Conservation profile shows three sections of high nucleotide
701 conservation in the Brassicaceae, two of which coincide with the region in DHS that is
702 required for expression in the bundle sheath.

703

704 **Figure 4: A TGCACCG motif within the DHS is necessary and sufficient for**
705 **preferential expression in the bundle sheath** (A) Two digital footprints (DGF) were
706 predicted based on the position of DNaseI digestion sites in the *pMYB76* DHS. (B) Two
707 motifs common to other promoters driving bundle sheath expression in *A. thaliana*. (C)

DGF II (underlined) overlaps with a TGCACCG motif (gold text). (D) Mutation of the TGCACCG motif leads to loss of GUS. (E) Oligomerization of the TGCACCG motif generates bundle sheath specific expression. GUS staining time was 48hrs. Scale bars represent 100µm.

Figure 5: DREB2A and DREB26 bind the *MYB76* DHS. Transcription factors that bind the DHS were investigated using Yeast One-Hybrid (Y1-H) and *in planta* promoter/TF interaction assays. (A) Schematic demonstrating transcription factors that interact with the MYB76 DHS as indicated by Y1-H. Blue locus identifiers are DREB2A, DREB26, DF1 and MYB73. (B) MYB76 DHS sequence with the best matching sequence marked for the binding sites of DREB26 (bold), DREB2A (underlined reverse complement), MYB73 (italics) and DF1 (dashed underlined). The TGCACCG motif is coloured in gold. (C) Comparison of the TGCACCG motif with the DAP-seq defined binding sites of DREB2A, DREB26, DF1 and MYB73. Numbers under motifs show the position of bases in the motif. Likelihood of matching the TGCACCG motif by chance is indicated by p-values. D) Three biological replicates of *N. benthamiana* co-infiltrated with candidate transcription factors and the *MYB76* DHS. Clockwise from top left: (1) *CaMV35S::DREB2A* and *pMYB76DHS:CaMV35SMin::uidA*, (2) *CaMV35S::DREB26* and *pMYB76DHS:CaMV35SMin::uidA,* (3) *CaMV35S::uidA* and (4) *pMYB76DHS:CaMV35SMin::uidA*. Staining time was 48hrs. Scale bars represents 2cm.

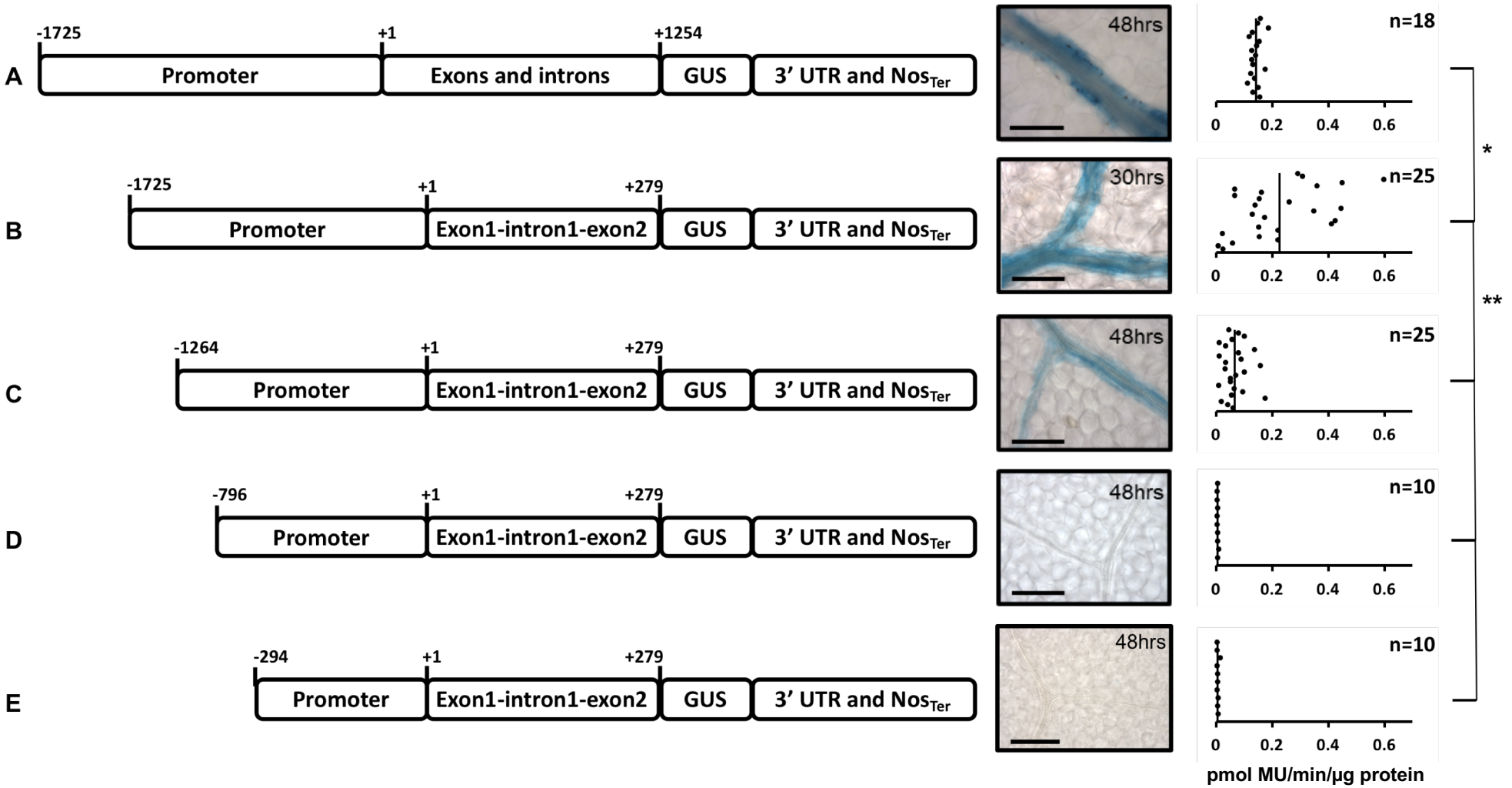


Figure 1: The *MYB76* promoter contains a region necessary for bundle sheath specific expression. Schematics of deletion series (left), representative images of GUS in *A. thaliana* leaves (centre) and quantification of MUG activity via the flourometric assay (right). (A) The *MYB76* genomic sequence. (B) The *MYB76* promoter including the first two exons and the first intron fused to the *uidA*. (C-E) A region between -1264 and -796bp is responsible for GUS accumulation in the bundle sheath. Staining times are given in the top right corner of each leaf image. The fluorometric MUG assay shows quantitative repressors and enhancers are located in the gene and in the promoter respectively (A-C right). X-axis indicates GUS activity and individual biological replicates are ordered randomly on the y axis. n, number of replicates; *, p value < 0.05; ** p value < 0.001. Scale bars represent 100µm.

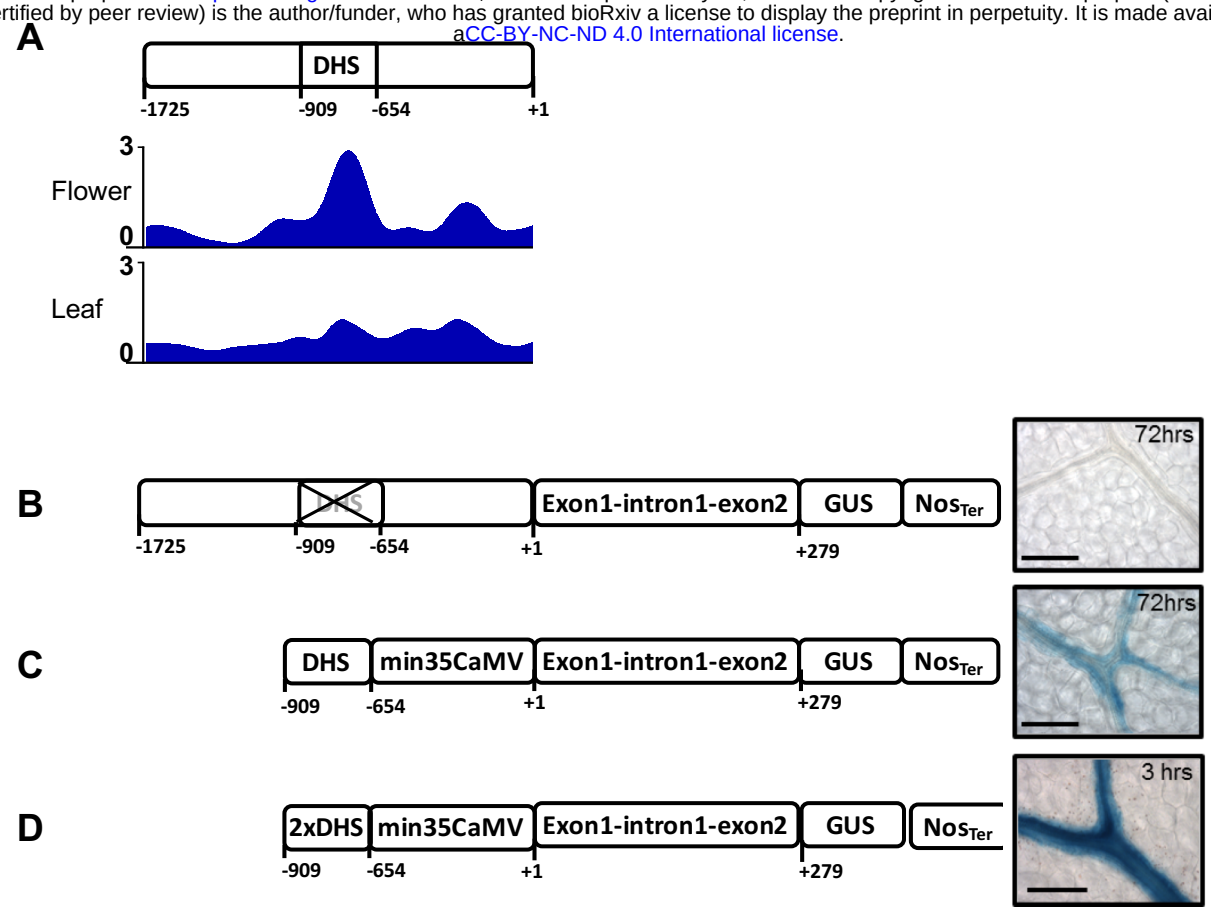
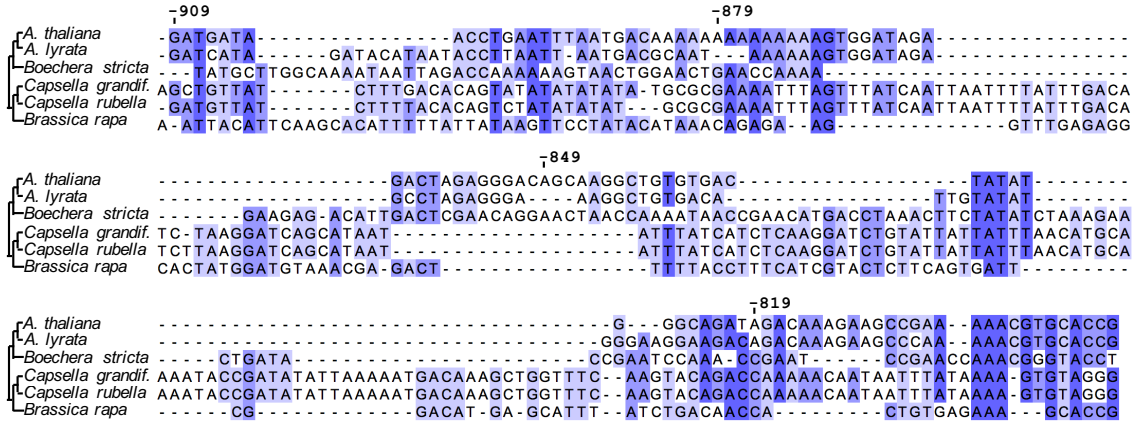


Figure 2: A DNaseI Hypersensitive Site in *pMYB76* is sufficient for expression in the bundle sheath. (A) The *MYB76* promoter contains a single DNaseI Hypersensitive Site (DHS) located from nucleotides -909 to -654. DHS shown as the DHS score (Zhang et al., 2015) from flower buds (top) and leaves (bottom). Data from Zhang et al (2012) and visualised with the IGV browser. (B) Deletion of the DHS abolishes GUS accumulation in the bundle sheath. (C) The DHS fused to the minimal CaMV35S promoter is sufficient to direct expression in the bundle sheath. (D) Oligomerising the DHS increases GUS accumulation in the bundle sheath. Staining times are given in the top right corner of each image. Scale bars represent 100μm.

A

```
-909   GATGATACA CCTGAATTTA ATGACAAAAA AAAAAAAAAG TGGATAGACA CTAGAGGGAC
-849   AGCAAGGCTG TGTGACATAT ATGGGCAGAT AGACAAAGAA GCCGAAAAAC GTGCACCGTC
-789   CAAGATTCTG GCTACTATAC CTAATTTCCT TCCCGCAGGG ACTTGACAAA TATCACTATC
-729   TGCCATTTTT AGTTTTATTT TGTATTGGTG TCAAAGAATT GAAATAATGA ACAACGGTCG
-669   TAAAAAGATG TAAATG
```

B



C

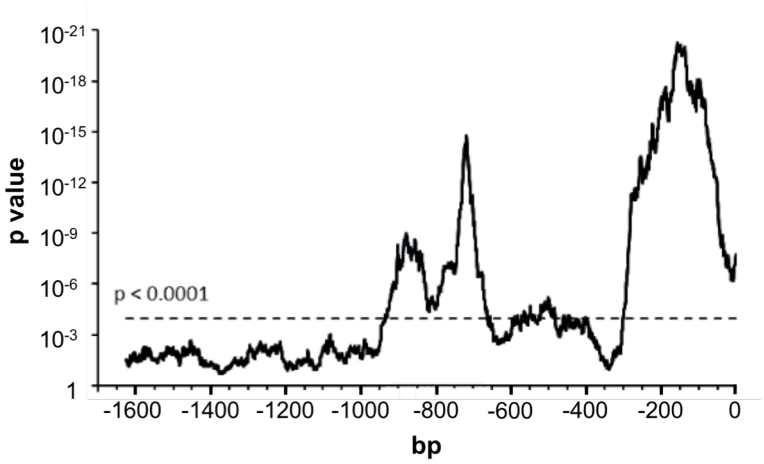


Figure 3: The DNaseI Hypersensitive Site is conserved in the Brassicaceae. (A) Single nucleotide polymorphisms (red) in the DHS region from 1135 accessions of *A. thaliana*. The region necessary for bundle sheath expression is underlined. (B) Alignment of nucleotides -909 to -796 of *MYB76* from *A. thaliana* to five additional species of the Brassicaceae. (C) Conservation profile shows three sections of high nucleotide conservation in the Brassicaceae, two of which coincide with the region in DHS that is required for expression in the bundle sheath.

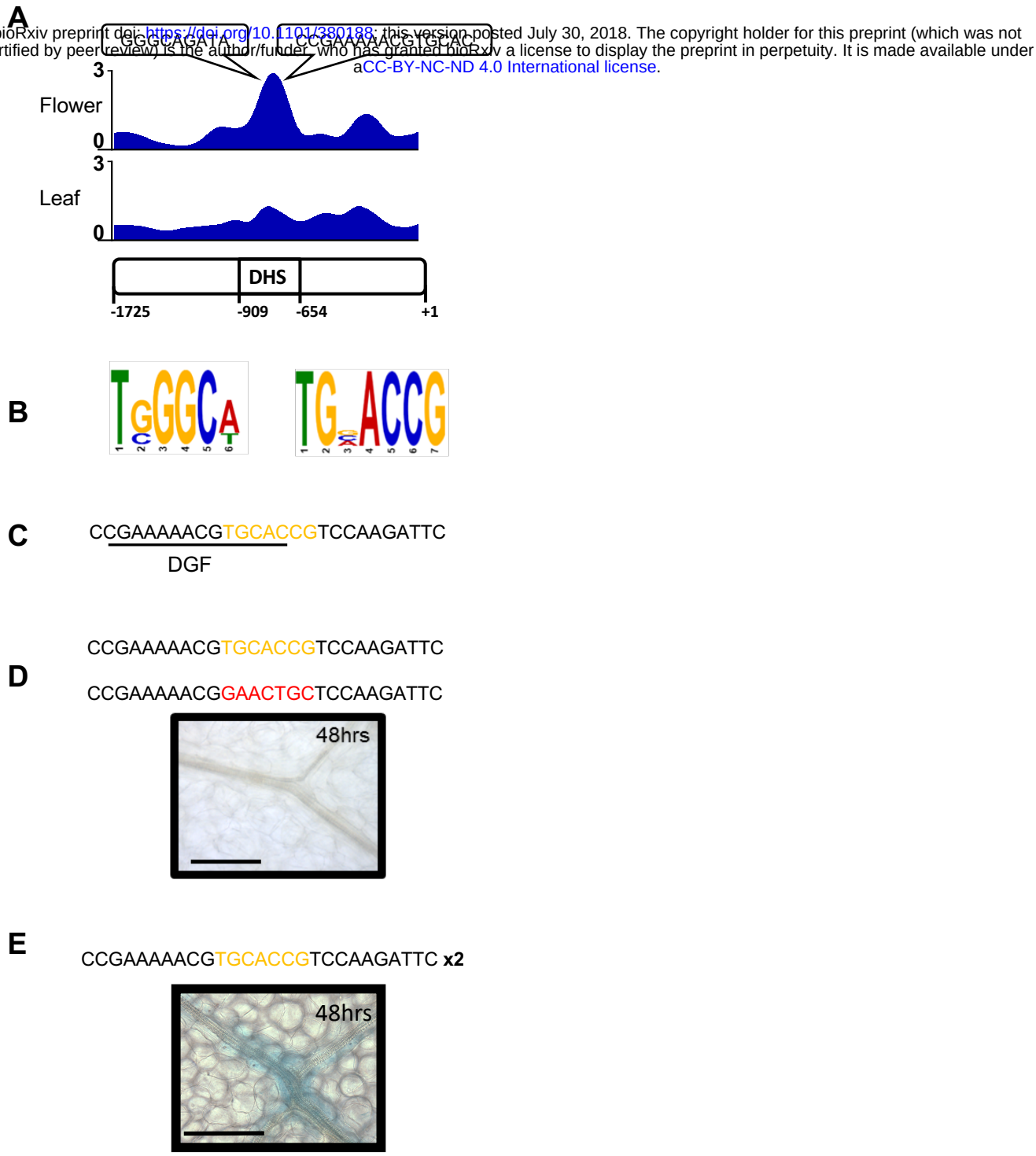
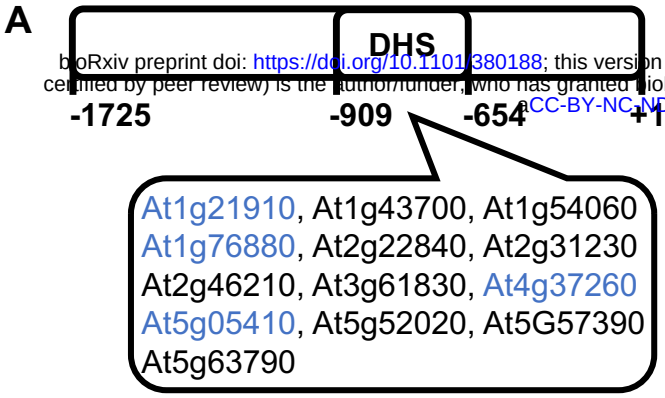


Figure 4: A TGCACCG motif within the DHS is necessary and sufficient for preferential expression in the bundle sheath (A) Two digital footprints (DGF) were predicted based on the position of DNaseI digestion sites in the *pMYB76* DHS. DHS shown as the DHS score (Zhang et al., 2015) from flower buds (top) and leaves (bottom). (B) Two motifs common to other promoters driving bundle sheath expression in *A. thaliana*. (C) DGF II (underlined) overlaps with a TGCACCG motif (gold text). (D) Mutation of the TGCACCG motif leads to loss of GUS. (E) Oligomerization of the TGCACCG motif generates bundle sheath specific expression. GUS staining time was 48hrs. Scale bars represent 100μm.



B

GATGATAACACCTGAATTTAATGACAAAAAAAAAAAA

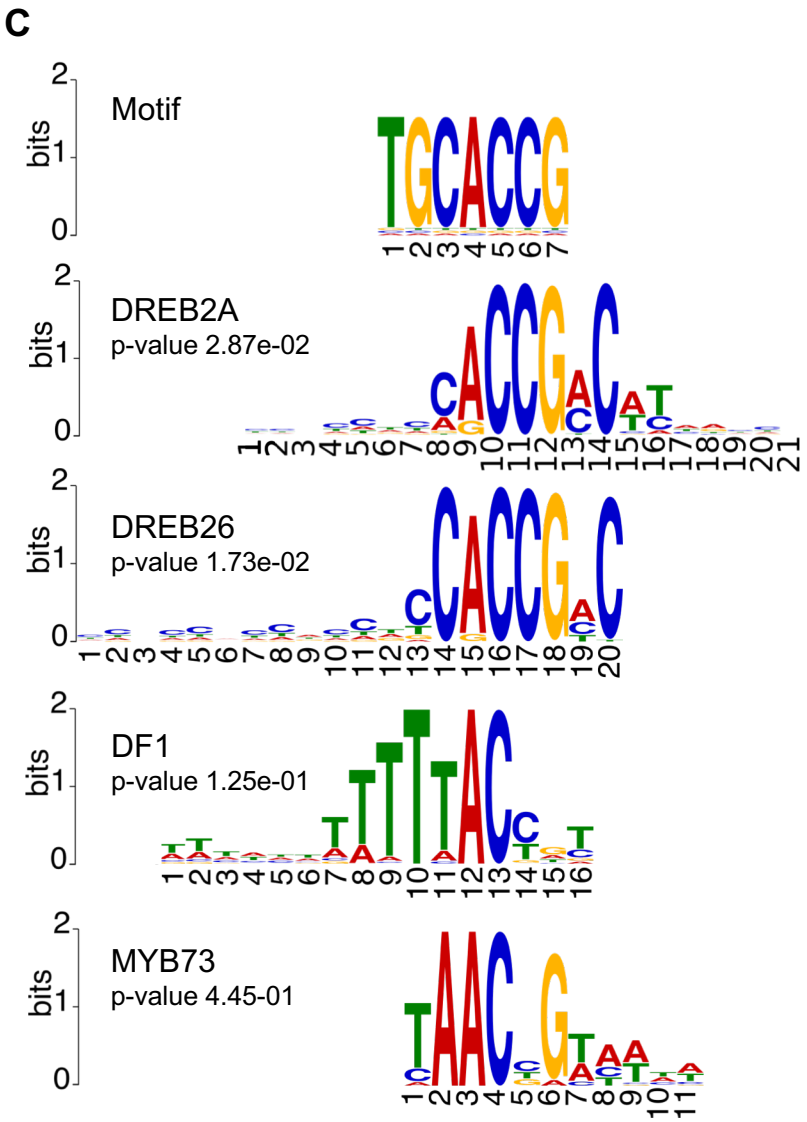
ACATATATGGGCAGATAGACAAAGAA**GCCGAAAAACG**

TGCACCGTCCAAGATTCTGGCTACTATACCTAATTTC

CTTCCCGCAGGGACTTGACAAATATCACTATCTGCCA

TTTTTAGTTTTATTTTGTATTGGTGTCAAAGAATTGA

AATAATGAA*CAACGGTCGTAAAAAGATGTAAATG*



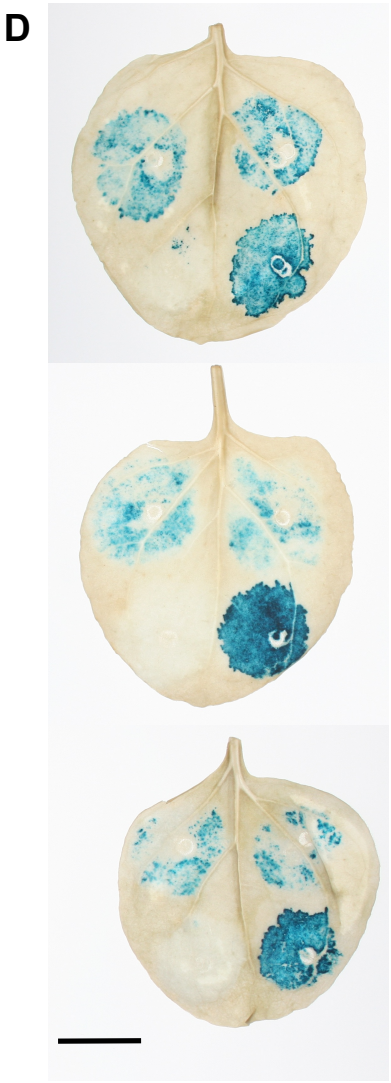


Figure 5: DREB2A and DREB26 bind the *MYB76* DHS. Transcription factors binding the DHS were identified using Yeast One-Hybrid (Y1-H) and assays *in planta*. (A) Schematic demonstrating transcription factors that interact with the MYB76 DHS *in* Y1-H. Blue locus identifiers are DREB2A, DREB26, DF1 and MYB73. (B) MYB76 DHS sequence marked with the best matching sequence for DREB26 (bold), DREB2A (underlined reverse complement), MYB73 (italics) and DF1 (dashed underlined) binding sites. The TGCACCG motif is coloured in gold. (C) Comparison of the TGCACCG motif with DAP-seq defined binding sites of DREB2A, DREB26, DF1 and MYB73. Numbers under motifs show the position of bases in the motif. Likelihood of matching the TGCACCG motif by chance is indicated by p-value. D) Three biological replicates of *N. benthamiana* co-infiltrated with candidate transcription factors and the *MYB76* DHS. Within each leaf clockwise from top left: (1) *CaMV35S::DREB2A* and *pMYB76DHS:CaMV35SMin::uidA*, (2) *CaMV35S::DREB26* and *pMYB76DHS:CaMV35SMin::uidA,* (3) *CaMV35S::uidA* and (4) *pMYB76DHS:CaMV35SMin::uidA*. Staining time was 48hrs. Scale bars represents 2cm.