A transcription factor quartet orchestrating bundle sheath expression in rice

Lei Hua, Na Wang, Susan Stanley, Ruth Donald, Kumari Billakurthi, Ana Rita Borba and Julian M. Hibberd

Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, United Kingdom

Email for contact: jmh65@cam.ac.uk

Keywords: *Oryza sativa*; C_4 rice; Bundle Sheath; Sulfite Reductase; Promoter; cis-elements

Abstract

The C_4 pathway has evolved in over sixty lineages of plants and improves photosynthetic efficiency by ~50%. One unifying character of C_4 plants is photosynthetic activation of a compartment such as the bundle sheath, but gene regulatory networks controlling this cell type are poorly understood. For example, in *Arabidopsis thaliana* a bipartite MYC-MYB transcription factor module restricts gene expression to this cell type but in grasses the regulatory logic allowing bundle sheath gene expression has not been defined. Using the global staple and C_3 crop rice we identified the *SULFITE REDUCTASE* promoter as sufficient for strong expression in this cell type. This promoter encodes an intricate *cis*-regulatory logic with multiple activators and repressors acting in a combinatorial manner. However, within this landscape we identified a distal enhancer comprising a module activated by quartet of transcription factors from the G2-like, MYB-related, IDD and bZIP families. This module is necessary and sufficient to pattern gene expression to the rice bundle sheath, and oligomerisation allowed activity to be increased 23-fold. In summary, we identify a short tunable enhancer patterning expression to the rice bundle sheath that we anticipate will be useful for engineering this cell type. The transcription factor module responsible for this patterning is distinct from that previously reported in *A. thaliana*.

Introduction

Multicellularity allows increased specialisation between cell-types, and since the colonisation of land the complexity of photosynthetic structures found in plants appears to have increased. For example, in the liverwort *Marchantia polymorpha* the photosynthetic thallus contains seven cell types of which mature somatic cells and initials at the base of chambers contain the most chloroplasts (Wang et al. 2023). Leaves of the angiosperm *Arabidopsis thaliana* possess at least nine anatomically distinct cell types and single-cell sequencing has identified an additional eight cell populations (Wang et al. 2021). Similarly, leaves of rice display a complex structure featuring at least ten cell types, with five more cell clusters predicted from single-cell sequencing (Wang et al. 2021). Photosynthesis gene expression in species such as rice is induced by light in all major cell types of the leaf and whilst this response is greatest in spongy and palisade mesophyll cells it is also evident in guard, mestome and bundle sheath cells (Swift et al. 2023). The bundle sheath constitutes a ring of photosynthetic cells around the vasculature (Esau, 1965) whose patterns of gene expression appear to condition these cells to allow water transport from veins to mesophyll cells, as well as sulphur assimilation and nitrate reduction (Leegood 2008; Aubry et al. 2014b; Hua et al. 2021). In some species, the bundle sheath is also specialised to accumulate metabolites such as glucosinolates that deter attack (Rosa and Rodrigues 1999; Rask et al. 2000). And, strikingly in multiple other lineages this specialised cell type has been repurposed during evolution to become fully photosynthetic and allow the complex C_4 pathway to operate.

Although the bundle sheath is found in all angiosperms and associated with multiple processes fundamental to leaf function, the molecular mechanisms responsible for directing expression to this cell type are poorly understood (Hibberd and Covshoff 2010; Kajala et al. 2011a). For example, few promoters capable of directing expression to this cell type have been reported (Schuler et al. 2016; Cui 2021). Moreover, only one transcription factor network comprising a MYC-MYB module that is necessary and sufficient to drive gene expression to the bundle sheath has been defined(Dickinson et al. 2020a). To our knowledge, no transcription factors or *cis*-elements responsible for decoding their activity have been shown to be capable of driving gene expression to the bundle sheath of grasses. In fact, in cereal crops such as rice only two promoters that pattern gene expression to the bundle sheath have been reported. The first is the *PHOSPHOENOLPYRUVATE CARBOXYKINASE* promoter derived from *Zoysia japonica* (*ZjPCK)* (Nomura et al. 2005a; Wang et al. 2017a; Danila et al. 2022a), and the second a *GLYCINE DECARBOXYLASE P-SUBUNIT* from *Flaveria trinervia* (*FtGLDP)* (Lee et al. 2021). Both promoters direct expression in mature cells of developed leaves but are weak, and the regulatory logic allowing them to generate bundle sheath expression is not well understood.

Identifying how rice patterns gene expression to the bundle sheath would provide insight into the underlying gene regulatory networks but also facilitate engineering of this cell type. One such example is associated with attempts to install C_4 photosynthesis into rice (Mitchell and Sheehy 2006). Compared with the ancestral C_3 state, plants that use C_4 photosynthesis operate higher light, water

and nitrogen use efficiencies (Makino et al. 2003; Sage 2004a; Mitchell and Sheehy 2006). It is estimated that introducing C_4 photosynthetic metabolism into C_3 rice would allow a 50% increase in yield (Mitchell and Sheehy 2006; Hibberd et al. 2008). C_4 photosynthesis requires multiple photosynthesis genes to be expressed in the bundle sheath, including enzymes that decarboxylate C_4 acids to release CO_2 around RuBisCO, organic acid transporters, components of the Calvin-Benson-Bassham cycle, RuBisCO activase, and enzymes of starch biosynthesis (Kajala et al. 2011b; Aubry et al. 2014a; Ermakova et al. 2021). Whilst expression of genes can be driven from misexpression of transcription factors such as designer Transcription Activator-Like Effectors (dTALEs) (Danila et al. 2022b) this still demands a promoter able to direct robust expression. In addition to driving strong expression of structural photosynthesis genes in the rice bundle sheath, introducing the C_4 pathway into C_3 leaves demands an expansion of the chloroplast and mitochondrial compartments compared with the ancestral C_3 state (Sage 2004a; Wang et al. 2017b; Lee et al. 2021). Because biogenesis of these organelles takes place early in leaf development (Wang et al. 2017a; Lee et al. 2021) a promoter capable of driving expression at these early stages is also desirable. Thus, the exogenous *ZjPCK* and *FtGLDP* promoters currently limit our ability to engineer processes in the rice bundle sheath.

We hypothesized that analysis of endogenous patterns of gene expression in the rice bundle sheath would allow us to identify a strong and early-acting promoter for this cell type. Once such a promoter was identified we also hypothesised that it could be used to initiate an understanding of the *cis*-regulatory logic that allows gene expression to be patterned to this cell type in grasses. We tested twenty-five promoters from rice genes that transcriptome sequencing indicated were highly expressed in these cells. Of these, four promoters specified preferential expression in the bundle sheath, and one derived from the *SULFITE REDUCTASE* (*SiR*) gene generated strong bundle sheath expression from plastochron 3 leaves. Truncation analysis showed that bundle sheath expression pattern from the *SiR* promoter is mediated by a short distal enhancer. This bundle sheath module is cryptic until other enhancers acting to both constitutively activate and repress expression in mesophyll cells are removed. The enhancer is composed of a quartet of *cis*-elements recognised by their cognate transcription factors from the GLK2, MYBS1, IDD and bZIP families. These four transcription factors act synergistically and are sufficient to drive expression of the strong bundle sheath *SiR* promoter.

Results

The *SiR* promoter directs expression to the rice bundle sheath

To identify sequences allowing robust expression in rice bundle sheath cells we used data derived from laser capture microdissection of bundle sheath strands and mesophyll cells from mature leaves. Promoter sequence from seven of the most strongly expressed genes in bundle sheath strands (**Supplemental Figure 1A**) were cloned, fused to the β-glucoronidase (GUS) reporter and transformed into rice. Although five of these promoters (*MYELOBLASTOSIS*, *MYB*; *HOMOLOG OF E. COLI BOLA*, *bolA*; *GLUTAMINE SYNTHETASE 1*, *GS1*; *STRESS RESPONSEIVE PROTEIN*, *SRP*; *ACYL COA BINDING PROTEIN*, *ARP*) led to GUS accumulation, it was restricted to veins **(Supplemental Figure 1B, 1C)**. And, for the *SULFATE TRANSPORTER 3;1* and *3;3* (*SULT3;1* and *SULT3;3*) promoters, no staining was observed (**Supplemental Figure 1B, 1C**). The approach of cloning promoters from bundle sheath strands therefore appeared to be more efficient at identifying sequences capable of driving expression in veins. We therefore optimised a procedure allowing bundle sheath cells to be separated from veins (Hua and Hibberd, 2019) to produce high quality transcriptomes from mesophyll, bundle sheath and vascular bundles of rice (Hua et al. 2021). From these data eighteen genes whose transcripts were more abundant in bundle sheath cells compared with both veins and mesophyll cells were identified (**Supplemental Figure 2A**). When each of these promoters was fused to GUS and transformed into rice, those from *ATP-SULFURYLASE 1B*, *ATPS1b*; *SULFITE REDUCTASE*, *SiR*; *HIGH ARSENIC CONTENT 1.1*, *HAC1.1*; and *FERREDOXIN*, *Fd* were sufficient to generate expression in the bundle sheath (**Supplemental Figure 2B)**. However, *ATPS1b* and *Fd* also displayed weak activity in the mesophyll, and the *HAC1.1* promoter led to GUS accumulation in epidermal and vascular cells as well as the bundle sheath. Thus, only the *SiR* promoter drove strong expression in the bundle sheath with no GUS detected in other cells (**Supplemental Figure 2B, 2C**). An additional six promoters (*SOLUBLE INORGANIC PYROPHOSPHATASE*, *PPase*; *PLASMA MEMBRANE INTRINSIC PROTEIN 1;1*, Os*PIP1;1*; *PLASMA MEMBRANE INTRINSIC PROTEIN 1;3*, Os*PIP1;3*; *ACTIN-DEPOLYMERIZING FACTOR*, *ADF*; *PEPTIDE TRANSPORTER PTR2*, *PTR2*; *NITRATE REDUCTASE*, *NIA2*) generated expression in vascular bundles, and eight promoters produced no staining (**Supplemental Figure 2B, 2C**). In summary, most candidate promoters failed to generate expression that was specific to bundle sheath cells, but the region upstream of the rice *SiR* gene was able to do so. We therefore selected the *SiR* promoter for further characterization.

The *SiR* promoter drives strong and early expression in bundle sheath cells

Sequence upstream of the *SiR* gene comprising nucleotides -2571 to +42 relative to the predicted translational start site was sufficient to generate expression in the rice bundle sheath. To allow faster analysis of sequences responsible for this output we domesticated the sequence by removing four *Bsa*I and *Bpi*I sites such that it was compatible with the modular Golden Gate cloning system. When this modified SiR_{pro} sequence was placed upstream of the GUS reporter it also generated bundle

140 sheath preferential accumulation **(Figure 1A)**. Fusion to a nuclear-targeted mTurquoise2 fluorescent
141 protein confirmed that the *SiR* sequence was sufficient to direct expression to bundle sheath cells,
142 and also revealed expression in the longer nuclei of veinal cells **(Figure 1B)**. There were no
143 detectable differences in levels of expression between domesticated and non-domesticated
144 sequences **(Figure 1C)**. Compared with 0.58 nmol 4-MU/min/mg protein previously reported for the
145 $ZjPCK_{pro}$ (Emmerling 2018) activity from the *SiR* promoter was at least 36% higher. Designer
146 Transcription Activator-Like Effector (dTALEs) and cognate Synthetic TALE- Activated Promoters
147 (STAPs) allow multiple transgenes to be driven by a single promoter (Brückner et al., 2015; Danila
148 et al., 2022). To test whether bundle sheath expression mediated by the *SiR* promoter can be
149 maintained and strengthened by the dTALE-STAP system, it was placed upstream of the dTALE
150 sequence. STAP62 and STAP4 were fused to GUS and a chloroplast targeted mTurquoise2
151 respectively. Stable transformants harbouring these reporters showed bundle sheath specific
152 expression **(Supplemental Figure 3A, 3B)**. GUS activity mediated by dTALE and STAP62 was ~18-
153 fold higher than GUS driven by the endogenous *SiR* promoter (**Supplemental Figure 3C**). We
154 conclude that *SiR* is compatible with the dTALE-STAP system and its activity can be strengthened.
155 We also investigated when activity from the *SiR* promoter was first detected during leaf development
156 and discovered that GUS as well as fluorescence from mTurquoise2 were visible in 5-20mm long
157 fourth leaves at Plastochron 3 (**Supplemental Figure 4**). This was not the case for $ZjPCK_{pro}$ even
158 when a dTALE was used to amplify expression (**Supplemental Figure 4**). We conclude that the *SiR*
159 promoter initiates expression in the bundle sheath before the *ZjPCK* promoter, and that it is also able
160 to sustain higher levels of expression in this cell type.
161

162 **A distal enhancer and Y-patch necessary for expression in the bundle sheath**

163 Analysis of the *SiR* promoter sequence revealed a highly complex *cis* landscape **(Figure 1D)**
164 comprising at least 638 motifs from 56 transcription factor families (**Supplemental table 1**). We
165 therefore designed a 5' truncation series to investigate regions necessary for expression in the
166 bundle sheath (**Figure 1E**). Deleting nucleotides -2571 to -2180 and -1490 to -980 led to a
167 statistically significant reduction and then increase in MUG activity respectively but neither truncation
168 abolished preferential accumulation of GUS in the bundle sheath **(Figure 1E-F).** However, when
169 nucleotides -980 to -394 upstream of the predicted translational start site were removed GUS was
170 no longer detectable in bundle sheath cells (**Figure 1E-F**). Consistent with this, MUG assays showed
171 a statistically significantly reduction in activity when these nucleotides were absent **(Figure 1G)**.
172 Thus, nucleotides spanning -980 to -394 of the *SiR* promoter are necessary for bundle sheath
173 specific expression.

174 To test whether this region is sufficient for bundle sheath specific expression we linked it to the
175 minimal CaMV35S core promoter. Although weak GUS signal was detected in a few veinal cells it
176 failed to do so in the bundle sheath **(Figure 1E-G)**. We conclude that sequence in two regions of the
177 promoter (from -394 to +42 and from -980 to -394) interact to specify expression to the bundle sheath.

178 To better understand this interaction we next generated unbiased 5' and 3' deletions. This second
179 deletion series further reinforced the notion that the *SiR* promoter contains a complex *cis*-regulatory
180 landscape. For example, when nucleotides -980 to -829 were removed very weak GUS staining was
181 observed and the MUG assay confirmed that activity was significantly reduced to 11% of the full-
182 length promoter **(Figure 2, Supplemental Figure 5).** We conclude that nucleotides -980 to -829
183 from the *SiR* promoter are necessary for tuning expression in the leaf. However, when nucleotides -
184 829 to -700 were removed GUS appeared in mesophyll cells **(Supplemental Figure 5)**. Truncating
185 nucleotides -613 to -529 abolished GUS accumulation (**Supplemental Figure 5)**. The 3' deletion
186 that removed nucleotides -251 to +42 also stopped accumulation of GUS in both bundle sheath and
187 mesophyll cells (**Figure 2A-C, Supplemental Figure 5)**. Notably, when the distal region required for
188 bundle sheath expression (-980 to -829) was combined with nucleotides -251 to +42 these two
189 regions were sufficient for patterning to this cell type (**Figure 2**).
190 Having identified a region in the *SiR* promoter that was necessary and sufficient for patterning to
191 the bundle sheath, we next used phylogenetic shadowing and yeast one hybrid analysis to better
192 understand the *cis*-elements and *trans*-factors responsible. Analysis of *cis*-elements in the *SiR*
193 promoter that are highly conserved in grasses identified a short region located from nucleotides -588
194 to -539 that contained an *ETHYLENE INSENSITIVE3-LIKE 3* (*EIL3*) transcription factor binding site
195 (**Supplemental Figure 6A&B**). Whilst deletion of this motif had no detectable effect of patterning to
196 the bundle sheath (**Supplemental Figure 6C**) the level of expression was reduced (**Supplemental**
197 **Figure 6D**). We infer that the *EIL3* motif positively regulates activity of the *SiR* promoter but is not
198 responsible for cell specificity. These data are consistent with the promoter truncation analysis that
199 showed nucleotides -613 to -529 containing this motif were not required for bundle sheath specific
200 expression, but instead function as a constitutive activator (**Supplemental Figure 5**). When yeast
201 one hybrid was used to search for transcription factors capable of binding the *SiR* promoter, sixteen
202 were identified (**Supplemental Figure 7A, 7B**). For each, cognate binding sites were present. This
203 included TCP21 and OsOBF1 that can bind to TCP motifs and bZIP_A or Ocs elements respectively.
204 Consistent with the outcome of deleting the *EIL3* motif, three EIL transcription factors interacted with
205 nucleotides -899 to -500 (**Supplemental Figure 7B, 7C**). Examination of transcript abundance in
206 mature leaves showed that most of these transcription factors were expressed in both bundle sheath
207 and mesophyll cells (**Supplemental Figure 7D**) implying that combinatorial interaction with cell
208 specific factors may be required for bundle sheath specific expression from the *SiR* promoter.
209

210 **The distal enhancer contains a tetrapartite *cis*-regulatory for bundle sheath expression**

211 The truncation analysis above identified two short regions comprising nucleotides -980 to -829
212 and -251 to +42 that were necessary and sufficient for expression in the rice bundle sheath (**Figure**
213 **2**). Sequence spanning nucleotides -251 to +42 includes both the annotated 5' untranslated region
214 but also likely contains core promoter elements (**Supplemental Figure 8A**). Re-analysis of publicly
215 available data identified two major transcription start sites at positions -91 (TSS1) and -41 (TSS2)

216 (**Supplemental Figure 8A**). Although no canonical TATA-box was evident in this region, at position
217 -130 a TATA-box variant was detected (5'-ATTAAA-3') (Civáň and Švec 2009) that could be
218 responsible for transcription from TSS1. Moreover, upstream of TSS2 is a putative pyrimidine patch
219 (Y-patch) that represents an alternate but common TC-rich core promoter motif (Civáň and Švec
220 2009) (**Supplemental Figure 8A**). Scanning sequence from -251 to +42 for core promoter elements
221 also identified MTE (Motif Ten Element), BREu (TFIIB Recognition Element upstream) and DCE-S-
222 I (Downstream Core Element S-I) motifs associated with eukaryotic core promoters (**Supplemental**
223 **Figure 8B**). We therefore assume the region upstream of TSS1 and TSS2 contained the core
224 promoter elements. When consecutive deletions to this sequence were made, statistically significant
225 reductions in MUG activity were evident but there was no impact on accumulation of GUS in the
226 bundle sheath. Interestingly, when the Y-patch was retained but the TATA-box like motif removed,
227 low levels of GUS specific to the bundle sheath were apparent (**Supplemental Figure 8C, 8D, 8E,**
228 **8F**), and deletion of the Y-patch completely abolished GUS staining (**Supplemental Figure 8C, 8D,**
229 **8E, 8F**). Consistent with the Y-patch being important for bundle sheath expression, when core
230 promoters from other genes (*PIP1;1, NRT1.1A*) containing a Y-patch were linked to the distal
231 enhancer from *SiR* bundle sheath expression was detected **(Figure 3A,3B)**, but this was not the
232 case for genes with only a TATA-box **(Figure 3A,3B)**. GUS activity was higher from the *PIP1;1* core
233 promoter that contains more Y-patch motifs. Overall, we conclude that the TATA-box like motif is not
234 required for expression in the bundle sheath, but in contrast the Y-patch is necessary for this
235 patterning and in combination with a distal enhancer comprising nucleotides -980 to -829 is sufficient
236 for expression in this cell type.

237 We next assessed the distal enhancer for transcription factor binding sites. The FIMO algorithm
238 identified motifs associated with WRKY, G2-like, MYB-related, MADS, DOF, IDD, ARR, SNAC
239 (Stress-responsive NAC) families. PlantPAN (Chow et al. 2018), which includes historically validated
240 *cis*-elements, found an additional *Dc3 Promoter Binding Factor* (*DPBF*) binding site for group A bZIP
241 transcription factors **(Figure 3D)**. Seven consecutive deletions spanning this enhancer region and
242 hereafter termed subregions a-g were then generated **(Figure 3D)**. Deletion of subregions a, b, d
243 and f resulted in loss of GUS from bundle sheath cells, although veinal expression persisted when
244 subregions a, b and d were absent (**Figure 3E-3G**). MUG analysis showed that deletion of all four
245 regions significantly reduced promoter activity **(Figure 3G)**. In contrast, deletions of nucleotides -938
246 to -923 (subregion c), -904 to 873 (subregion e), and -853 to -829 (subregion g) had no impact on
247 either the pattern or level of expression. Subregions a, b, c and f that were necessary for expression
248 in the bundle sheath contained unique binding sites for WRKY, G2-like, MYB-related, IDD, NAC and
249 bZIP (DPBF) transcription factors. To examine the significance of these regions in the context of full-
250 length *SiR* promoter, consecutive deletions from subregion a to f were generated **(Supplemental**
251 **Figure 10A)**. Deletion of subregion a, d or f, led to GUS accumulating primarily in mesophyll cells
252 whereas removal of subregion b, c or e, caused GUS staining in both mesophyll cells and bundle
253 sheath cells **(Supplemental Figure 10B)**. No significant changes in GUS activity were observed in

254 these deletion lines **(Supplemental Figure 10C)**. We conclude that that the distal enhancer
255 generates expression in the bundle sheath due to four distinct sub-regions and that other nucleotides
256 between -980 to -853 function as repressors of mesophyll expression.
257
258 **G2-like, MYB-related, IDD and bZIP transcription factors activate the distal enhancer**
259 To gain deeper insight how the distal enhancer operates we employed multiple approaches
260 including transactivation assay, co-expression analysis and site directed mutagenesis. The distal
261 enhancer contained WRKY, G2-like, MYBR, IDD, SNAC and bZIP (DPBF) motifs **(Figure 4A)**. We
262 therefore cloned rice transcription factors from each family and used them as effectors in transient
263 assays. WRKY121, GLK2, MYBS1, IDD2/3/4/6/10, and bZIP9/10/11 transcription factors activated
264 expression from the bundle sheath enhancer, whereas the stress-responsive NAC transcription
265 factors did not **(Figure 4B, Supplemental Figure 11)**. We therefore conclude that the SNAC motif
266 is not important for activity of the bundle sheath enhancer. Effector assays using pairwise
267 combinations of transcription factors showed synergistic activation from the distal enhancer when
268 GLK2 and IDD3,4,6,10 were co-expressed **(Figure 4C)**.
269 Co-expression analysis using a cell-specific leaf developmental gradient dataset revealed that
270 GLK2, MYBS1 and IDD4,6,10 transcription factors that bind the G2-like, MYB-related and IDD motifs
271 respectively were more abundant in mesophyll cells **(Figure 4D)**. However, the WRKY121, IDD2,
272 and bZIP9/10/11 transcription factors correlated with *SiR* transcript abundance and were
273 preferentially expressed in bundle sheath cells **(Figure 4D)**. To test whether WRKY121, IDD2, and
274 bZIP9/10/11 are sufficient to pattern *SiR* expression to specific cells, we mis-expressed them in the
275 mesophyll in the presence of the *SiR* distal enhancer and Y-patch **(Supplemental Figure 12)**.
276 Presence of these three factors in mesophyll cells activated expression in this cell type **(Figure 4E)**
277 suggesting at least one directly interacts with the distal enhancer in bundle sheath cells to drive *SiR*
278 expression. We next mutated WRKY, MYB-related, IDD, and bZIP motifs. Although mutating the
279 WRKY did not affect bundle sheath expression pattern, mutations in the MYB-related, IDD, and bZIP
280 (DPBF) motifs abolished it **(Supplemental Figure 13)**, and so we conclude that the WRKY motif is
281 not necessary for bundle sheath expression.
282 In order to test whether the regions containing the G2-like, MYB-related IDD, and bZIP (DPBF)
283 sites are sufficient to pattern expression to rice bundle sheath cells we concatenated them and fused
284 them to the core promoter of *SiR* (**Figure 5A**). GUS staining was evident in the bundle sheath
285 (**Figure 5B**). Oligomerisation of this minimal module increased expression by 23-fold (**Figure 5C**).
286 Collectively our data indicate that transcription factors belonging to the G2-like, MYB-related, IDD
287 and bZIP (DPBF) families act cooperatively to decode distinct *cis*-elements in a distal enhancer of
288 the *SiR* promoter. This regulatory module is both necessary and sufficient for bundle sheath-specific
289 expression in rice.

290 **Discussion**

291 **Expression of multiple genes in the rice bundle sheath is not associated with close upstream**
292 **enhancers**

293 Gene expression is determined by interactions between elements in the core promoter allowing
294 basal levels of transcription (Juven-Gershon and Kadonaga 2009; Haberle and Stark 2018) with
295 more distal *cis*-regulatory modules (Spitz and Furlong 2012; Shlyueva et al. 2014; Ray-Jones and
296 Spivakov 2021). Such *cis*-regulatory modules include enhancers and silencers that act as hubs
297 receiving input from multiple transcription factors and so allow gene expression to respond spatially
298 and temporally to both internal and external stimuli (Li et al. 2007; Buecker and Wysocka 2012).
299 After 25 promoters were tested we discovered that the majority were not capable of driving
300 expression in the rice bundle sheath, and this included ten that generated no detectable activity of
301 GUS in leaves. In all cases we had cloned sequence between -3191 and -960 nucleotides upstream
302 of the predicted translational start site and so these data demonstrate that the core promoter and
303 any enhancers in these regions are not sufficient to direct expression to rice bundle sheath cells.
304 Combined with the paucity of previously reported promoters active in this cell type (Nomura et al.
305 2005b; Lee et al. 2021) these data argue either for long range upstream enhancers (Studer et al.
306 2011; Liu et al. 2015; Li et al. 2019; Yan et al. 2019; Zhao et al. 2022) or other regulatory mechanisms
307 being important to specify expression in the bundle sheath. Possibilities include transcription factor
308 binding sites in introns that impact on transcription start site and strongly enhance gene expression
309 (Rose et al. 2008; Gallegos and Rose 2019; Rose 2019) or in exons where because such sequences
310 specify amino acid sequence as well as binding of *trans*-factors, they have been termed duons
311 (Stergachis et al. 2013) Functional analysis showed that duons can pattern expression to the bundle
312 sheath of the C_4 plant *Gynandropsis gynandra* (Reyna-Llorens et al. 2018), and it is notable that a
313 genome-wide analysis of transcription factor binding sites in grasses revealed genes preferentially
314 expressed in bundle sheath cells tended to contain transcription factor binding sites in their coding
315 sequence (Burgess et al. 2019). It therefore appears possible that gene expression in the bundle
316 sheath is commonly encoded by non-canonical architecture perhaps based on duons rather than
317 more traditional enhancer elements upstream of the core promoter.

318 Despite the above, we discovered four promoters capable of driving expression in the rice bundle
319 sheath, and each was associated with a gene important in sulphur metabolism. For example, *ATPS*,
320 *SiR* and *Fd* all participate in the first two steps of sulphate reductive assimilation, while *HAC1;1*
321 encodes an arsenate reductase important in the detoxification of arsenate using glutathione that is
322 a product of sulphur assimilation. Collectively, these data further support the notion that the rice
323 bundle sheath cell is specialised in sulphur assimilation (Hua et al. 2021).

324

325 **Two distinct genetic networks governing expression in bundle sheath cells**

326 The only other promoter for which both *cis*-elements and *trans*-factors that are necessary and
327 sufficient to pattern bundle sheath expression have been reported is from *A. thaliana*. Here, a

328 bipartite MYC-MYB module upstream of the *MYB76* gene is responsible for this output (Dickinson et
329 al. 2020b). MYB76 forms part of a network governing glucosinolate biosynthesis in *A. thaliana*, and
330 so it is notable that the gene regulatory network we report in rice is also associated with sulphur
331 metabolism. However, rather than the bipartite transcription factor network that regulates bundle
332 sheath expression in *A. thaliana,* in rice we report a quartet of transcription factors controlling *SiR*
333 **(Figure 6A&6B)**. The enhancer controlling bundle sheath *SiR* expression in rice comprises four
334 distinct regions recognised by transcription factors belonging to the G2-like, MYB-related, IDD and
335 bZIP families **(Figure 6B)**. As loss of the MYB-related, IDD and bZIP motifs abolished expression,
336 this implies they act co-operatively a notion further supported by the fact that GLK2 and IDD3,4,6,10
337 synergistically activated promoter output in a transient assay **(Figure 4C)**. It is possible that other
338 motifs in the enhancer such as MADS, DOFs and ARR act as modulators to tune the level of bundle
339 sheath expression. In fact, single nuclei sequencing of rice and sorghum during photomorphogenesis
340 identified DOFs as important for bundle sheath expression (Swift, Luginbuhl et al. 2023).
341 Interestingly, the three other promoters (*ATPSb, Fd* and *HAC1;1*) we identified as being sufficient to
342 drive reporter expression to the rice bundle sheath, and also known bundle sheath promoters (*ZjPCK*
343 and *FtGLDP*) contain the cognate *cis*-elements for G2-like, MYB-related, IDD and bZIP transcription
344 factors (**Supplementary Figure 14**).

345 The distal enhancer in the *SiR* promoter operates in conjunction with the core promoter that
346 contains two transcription start sites, one with an upstream TATA-box and the other a TC-rich
347 element known as a pyrimidine (Y) patch **(Supplemental Figure 8A)**. The TATA-box is found in
348 metazoans and plants and allows recognition by the pre-initiation complex (Smale and Kadonaga
349 2003), but in plants computational analysis showed that many promoters lack a TATA-box and
350 instead contain a Y-patch (Yamamoto et al. 2007a, 2007b; Bernard et al. 2010). These genes tend
351 to be relatively steadily expressed and associated with protein metabolism (Bernard et al., 2010),
352 and presence of a Y-patch can increase core promoter strength (Jores et al. 2021). For *SiR,* whilst
353 the TATA-box is not required, the Y-patch is needed for expression in the bundle sheath.

354 This regulatory network comprising the Y-patch and distal enhancer enabling bundle sheath
355 expression of *SiR* is embedded within a complex *cis*-regulatory landscape upstream of *SiR* with
356 distinct regions encoding activating and repressing activities (**Figure 6**). For example, the distal
357 enhancer (nucleotides -980 to -829) activating expression in the bundle sheath is proximate to
358 sequence (nucleotides -829 to -700) that suppresses expression in the mesophyll. Notably, the distal
359 enhancer is both essential for mesophyll repression but also sufficient to drive bundle sheath specific
360 expression. In addition to controlling cell specificity, this complexity likely also facilitates the tuning
361 of expression to environmental conditions. For instance, the *EIL* motif (position -572 to -552) is
362 recognised by *ETHYLENE-INSENSITIVE LIKE* (*EIL*) transcription factors that respond to sulphur
363 deficiency (Maruyama-Nakashita et al. 2006; Dietzen et al. 2020). As transcripts encoding *EIL*
364 accumulate in both bundle sheath and mesophyll cells in response to sulphate deficiency it seems
365 likely that transcription factors repressing expression in these mesophyll cells respond in a dynamic

366 manner. In addition to *EIL*, the yeast one hybrid analysis identified seven other families of
367 transcription factor families that can bind the *SiR* promoter. Many play documented roles during
368 abiotic or biotic stress, with for example OBF1, ERF3, NAP and FLP acting during low-temperature
369 or drought responses (Shimizu et al. 2005; Zhang et al. 2013, 2022; Chen et al. 2014; Qu et al.
370 2022), while TCP21, EREBP1, ERF3, ERF72, ERF83 are involved in both abiotic and biotic stress
371 (Lin et al. 2007; Jisha et al. 2015; Zhang et al. 2016, 2018; Tezuka et al. 2019; Jung et al. 2021).
372 Consistent with previous *in silico* analysis (Kurt et al. 2022) the presence of multiple AP2/ERF and
373 EIL transcription factors binding sites suggests that *SiR* is likely subject to control from ethylene
374 signalling (Binder 2020) and also of transcription factors that respond to ABA and jasmonic acid
375 biosynthesis and signalling (Yaish et al. 2010; Chen et al. 2014; Jisha et al. 2015; Zhang et al. 2016,
376 2018). Together this implies that multiple phytohormone signalling pathways converge on the *SiR*
377 promoter. These data are similar to those reported for the *SHORTROOT* promoter in *A. thaliana*
378 roots where a complex network of activating and repressing *trans*-factors also tunes expression
379 (Sparks et al. 2016). Such complexity in the regulatory landscape may enable robustness in
380 maintaining cell specific gene expression, such that individual mutations are less likely to completely
381 abolish the required pattern of expression. It could also allow a fine tuning of gene expression both
382 temporally as well as spatially in response to environmental perturbations. As sulphite reductases
383 are preferentially expressed in the bundle sheath of most species analysed to date (Hua et al. 2021)
384 it will be interesting in the future to determine if this regulatory landscape is shared across these
385 species.
386
387 **Using the *SiR* promoter to engineer the rice bundle sheath**
388 In addition to bundle sheath cells being important for sulphur assimilation (Leegood 2008; Aubry
389 et al. 2014b; Hua et al. 2021) they have also been implicated in nitrate assimilation, and the control
390 of leaf hydraulic conductance and solute transport (Hua et al. 2021) and the systemic response to
391 high light (Xiong et al. 2021). Moreover, in one of the most striking examples of a cell type being
392 repurposed for a new function, bundle sheath cells have repeatedly been rewired to allow the
393 evolution of C_4 photosynthesis (Sage 2004b). To engineer these diverse processes, specific and
394 tuneable promoters for this cell are required. However, identification of sequence capable of driving
395 specific expression to bundle sheath strands has previously been limited to *A. thaliana* and C_4
396 species. For example, the *SCARECROW* (Cui et al. 2014), *SCL23* (Cui et al. 2014), *SULT2;2*
397 (Kirschner et al. 2018) and *MYB76* promoters (Dickinson et al. 2020a) derive from *A. thaliana*, whilst
398 the *Glycine Decarboxylase P-protein (GLDP*) promoter is from the C_4 dicotyledon *Flaveria trinervia*
399 (Engelmann et al. 2008; Wiludda et al. 2012). In rice, only the C_4 *Zoysia japonica PCK* and the C_4
400 *Flaveria trinervia GLDP* promoters are known to pattern expression to the bundle sheath (Nomura
401 et al. 2005c; Lee et al. 2021; Danila et al. 2022b). Both are capable of conditioning expression in this
402 cell type, but are weak, turn on at late leaf developmental stages and the molecular basis

403 underpinning their ability to restrict expression to the bundle sheath has not been defined. It has
404 therefore not been possible to rationally design or tune expression to this important cell type in rice.
405 From analysis of the ~2600 nucleotide *SiR* promoter we identify an enhancer comprising 81
406 nucleotides that is sufficient to drive expression to bundle sheath cells. Moreover, we show that
407 output can be tuned via two approaches. First, oligomerising the distal enhancer can drastically
408 increase expression. Second, combining it with different core promoters achieved the same output,
409 and this appears to correlate with length of the Y-patch in each promoter. Our identification of a
410 minimal promoter that drives expression in bundle sheath cells of rice now provides a tool to allow
411 this important cell type to be manipulated. Cell specific manipulation of gene expression has many
412 perceived advantages. For example, when constitutive promoters have been used to drive gene
413 expression gene silencing and reduction of plant fitness due to metabolic penalties can ensue (Glick
414 1995; Que et al. 1997). In contrast, tissue specific promoters allow targeted gene expression either
415 spatially or at particular developmental stages and so allow increased precision in trait engineering
416 and (Kummari et al. 2020). The *SiR* promoter and the bundle sheath *cis*-regulatory module that we
417 identify provides insights into mechanisms governing cell specific expression in rice, and may also
418 contribute to our ability to engineer and improve cereal crops.

419 **Materials and methods**

420 **Plant material and growth conditions**

421 Kitaake (*O. sativa* ssp. *japonica*) was transformed using *Agrobacterium tumefaciens* with the
422 following modifications as described previously (Hiei et al., 2008). Mature seeds were sterilized with
423 2.5% (v/v) sodium hypochlorite for 15 mins, and calli were induced on NB medium with 2 mg/L 2,4-
424 D at 30 °C in darkness for 3-4 weeks. Actively growing calli were then co-incubated with *A.*
425 *tumefaciens* strain LBA4404 in darkness at 25°C for 3 days, they were selected on NB medium
426 supplied with 35 mg/L hygromycin B for 4 weeks, and those that proliferated placed on NB medium
427 with 10 mg/L hygromycin B for 4 weeks at 28 °C under continuous light. Plants resistant to
428 hygromycin were planted in 1:1 mixture of top-soil and sand and placed in a greenhouse at the
429 Botanic Gardens, University of Cambridge under natural light conditions but supplemented with a
430 minimum light intensity of 390 μmol m^{-2} s^{-1}, a humidity of 60%, temperatures of 28°C and 23°C
431 during the day and night respectively, and a photoperiod of 12 h light, 12 h dark. Subsequent
432 generations were grown in a growth cabinet in 12 h light/12 h dark, at 28 °C, a relative humidity of
433 65%, and a photon flux density of 400 μmol m^{-2} s^{-1}.

434

435 **Cloning and construct preparation, and motif analysis**

436 The 2613-bp promoter DNA fragment of *SULFITE REDUCTASE* (*SiR,* MSU7 ID:
437 LOC_Os05g42350, RAP-DB ID: Os05g0503300) was originally amplified from Kitaake genomic
438 DNA, with forward primer (5'-3') "CACCATGCTTGACCATGTGGACTC" and reverse primer (5'-3')
439 "ACGGAACCCGTGGAACTC". Gel-purified PCR product was cloned into a Gateway pENTR™
440 vector to generate *pENTR-SiRpro* using pENTR™/D-TOPO™ Cloning Kit (Invitrogen), then the
441 promoter was recombined into the pGWB3 expression vector and fused with *GUS* gene using LR
442 reaction. The resultant vector was transformed into *A. tumefaciens* strain LBA4404 and transformed
443 into Kitaake. To engineer the *SiRpro* such that it is compatible with the Golden gate system, four
444 *Bsa*I or *Bpi*I restriction enzyme recognition sites at -214, -298, -1468, and -2309 were mutated from
445 T to A and cloned into the *pAGM9121* vector using Golden Gate level 0 cloning reactions and then
446 into a level 0 PU module EC14328, which was used for driving *kzGUS* (intronless GUS) and *H2B-*
447 *mTurquoise2* reporter genes via Golden gate reaction and using *Tnos* as a terminator. A five prime
448 deletion series was generated using EC14328 as the template and prepared as level 0 PU modules,
449 and a three prime deletion series prepared as level 0 P modules. The minimal *CaMV35S* promoter
450 was used as the U module, and they were linked with *kzGUS* and terminated with *Tnos*.

451 To test *SiRpro* in the dTALE/STAP system, the 42-bp coding region were excluded and the 2571-
452 bp resultant fragment placed into a level 0 PU module EC14330 and was used to drive *dTALE1*.
453 Two reporters were used. For the GUS reporter *kzGUS* was linked with *STAP62* and terminated with
454 *Tnos*. In the fluorescent reporter construct, a chloroplast targeting peptide fused to *mTurquoise 2*
455 was linked with *STAP 4* and terminated with *Tact2*. In both constructs, *pOsAct1* driving *HYG*

456 (Hygromycin resistant gene) was terminated with *Tnos* and used as the selection marker during rice
457 transformation.

458 The Find Individual Motif Occurrences (FIMO) tool (Grant et al. 2011) from the Multiple Em for Motif
459 Elucidation (MEME) suite v.5.4.0 (Bailey et al. 2015) was used to search for individual motifs within
460 the promoter sequences using default parameters with "--thresh" of "1e-3". Position weight matrix of
461 656 non-redundant plant motifs and 13 RNA polymerase II (POLII) core promoter motifs were
462 obtained from JASPAR (https://jaspar.elixir.no/downloads/) (Fornes et al. 2020). To cluster the
463 transcription factor binding motifs, the RSAT matrix-clustering tool (Castro-Mondragon et al. 2017) was
464 run on all 656 non-redundant plant motifs using the default parameters, which yield 51 motif clusters,
465 these clusters were further divided based on transcription factor families (Supplemental table 2).
466

467 **Analysis of GUS and fluorescent reporters**

468 GUS staining was performed as described previously (Jefferson et al., 1987) with the following
469 minor modifications. Leaf tissue was fixed in 90% (v/v) acetone overnight at 4 ^{o}C after washing with
470 100 mM phosphate buffer (pH 7.0). Leaf samples were transferred into 1 mg/ml 5-bromo-4-chloro-
471 3-indolyl glucuronide (X-Gluc) GUS staining solution, subjected to 2 mins vacuum infiltration 5 times,
472 and then incubated at 37 ^{o}C for between 1 and 168 hours. Chlorophyll was cleared further using 90%
473 (v/v) ethanol overnight at room temperature. Cross sections were prepared manually using a razor
474 blade and images were taken using an Olympus BX41 light microscopy. Quantification of GUS
475 activity was performed using a fluorometric MUG assay (Jefferson et al., 1987). ~200 mg mature
476 leaves from transgenic plants were frozen in liquid nitrogen and ground into fine powder with a
477 Tissuelyser (Qiagen). Soluble protein was extracted in 1 mL of 50 mM phosphate buffer (pH 7.0)
478 supplemented with 0.1% [v/v] Triton X-100 and cOmplete™ Protease Inhibitor Cocktail (half tablet
479 per 50 mL). Protein concentration then determined using a Qubit protein assay kit (Invitrogen). The
480 MUG fluorescent assay was performed in duplicates with 20 µl protein extract in MUG assay buffer
481 (50mM phosphate buffer (pH 7.0), 10 mM EDTA-Na_2, 0.1% [v/v] Triton X-100, 0.1% [w/v] N-
482 lauroylsarcosine sodium, 10 mM DTT, 2 mM 4-methylumbelliferyl-β-D-glucuronide (MUG)) in a 200
483 µl total volume. The reaction was conducted at 37 ˚C in GREINER 96 F-BOTTOM microtiter plate
484 using a CLARIOstar plate reader. 4-Methylumbelliferone (4-MU) fluorescence was recorded every 2
485 minutes for 20 cycles with excitation at 360 nm and emission detected at 450 nm. 4-MU
486 concentration was determined based on a standard curve of ten 4-MU standards placed in the same
487 plate. GUS enzymatic rates were calculated by averaging the slope of MU production from each of
488 the duplicate reactions.

489 In order to visualize mTurquoise2 mature leaves were dissected into 2-cm sections, leaf
490 epidermal cells were removed by scraping the leaf surface with a razor blade and then mounted with
491 deionized water. 5-mm and middle sections of 2-cm young tissue of the fourth leaves were dissected
492 and mounted with deionized water directly. Imaging was then performed using a Leica TCS SP8
493 confocal laser-scanning microscope using a 20x air objective. mTurquoise2 fluorescence was

excited at 442 nm with emission at 471–481 nm, chlorophyll autofluorescence was excited at 488 nm with emission at 672–692 nm.

Yeast one hybrid, protoplast isolation and transactivation assay

The yeast one hybridisation assay was performed by Hybrigenics (https://www.hybrigenics-services.com/). Fragments were synthesized and used as bait. Rice leaf and root cDNA libraries were used as prey. The number of clones screened and concentration of 3-AT were as follows: fragment 1, 70.2 million clones screened with 0 mM 3-AT; fragment 2, 61.5 million clones screened with 0 mM 3-AT; fragment 3, 68.4 million clones screened with 20 mM 3-AT; fragment 4, 57.4 million clones screened with 100 mM 3-AT; fragment 5, 94.2 million clones screened with 200 mM 3-AT.

Rice leaf protoplasts and PEG-mediated transformation were performed as described previously (Page et al. 2019). Golden gate level 1 modules for transformation were isolated using ZymoPURE™ II Plasmid Midiprep Kit, *ZmUBIpro::GUS-Tnos* was used as transformation control. Transcription factor coding sequences were amplified using rice leaf cDNA, with *Bsa*I and *Bpi*I sites mutated, and cloned into Golden gate SC level 0 modules. They were assembled into a level 1 module with a *ZmUBIpro* promoter and *Tnos* terminator module. Nucleotides -980 to -829 with the endogenous core promoter (nucleotide -250 to +42) were fused with the *LUC* reporter to generate output of transcription activity. In each transformation, 2 µg of transformation control plasmids, 5 µg of reporter plasmids, and 5 µg of effector plasmids per transcription factor were combined and mixed with 170 µl protoplasts. After incubation on the benchtop for overnight protein was extracted using passive lysis buffer, GUS activity was determined with 20 µl of protein sample and MUG fluorescent assay as described above, LUC activity was measured with 20 µl of protein sample and 100 µl of LUC assay reagent (Promega) using Clariostar plate reader. Transcriptional activity from the promoter was calculated as LUC luminescence/rate of MUG accumulation.

Acknowledgements

This work was funded by the Bill and Melinda Gates Foundation C_4 Rice grant awarded to the University of Oxford (2015-2019 (OPP1129902) and 2019-2024 (INV-002970). For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Author contributions

L.H. and J.M.H. conceived the work. J.M.H. guided execution of experiments and oversaw the project. L.H., N.W., S.S., R.D., K.B., and A.R.B. did the experiments and analysed the data. L.H. and J.M.H. wrote the manuscript with input from all authors.

Declaration of interests

The authors declare no competing interests.

532 **Figure legends**

533

534 **Figure 1. Nucleotides -980 to -394 of the *SiR* promoter are necessary for bundle sheath**
535 **expression. (A)** Domesticated *SiR* promoter generates strong GUS staining in bundle sheath. **(B)**
536 mTurquoise2 signal driven by the domesticated *SiR* promoter in nuclei (indicated by yellow arrows)
537 of bundle sheath cells (marked by yellow dashed lines) and vein cells in mature leaves, red indicates
538 chlorophyll autofluorescence. **(C)** The fluorometric 4-methylumbelliferyl-β-D-glucuronide (MUG)
539 assay shows no statistically significant difference between the endogenous and domesticated *SiR*
540 promoter activity. **(D)** Landscape of transcription factor binding sites in the *SiR* promoter using the
541 Find Individual Motif Occurrences (FIMO) program. The likelihood of match between 656 plant
542 nonredundant known transcription factor motifs in the *SiR* promoter is shown by transcription factor
543 families (Supplemental table 1). **(E)** Schematics of 5' truncations. **(F)** Representative images of leaf
544 cross sections from transgenic lines after GUS staining. Zoomed-in images of lateral veins shown in
545 right panels, bundle sheath cells were highlighted with white dash line, scale bars = 50 μm. **(G)**
546 Promoter activity determined by the fluorometric 4- methylumbelliferyl-β-D-glucuronide (MUG) assay.
547 Data subjected to a Kruskal-Wallis test, median indicates median value of catalytic rate of GUS, n
548 indicates total number of transgenic lines assessed.

549

550 **Figure 2. A distal enhancer and the core promoter are necessary and sufficient for bundle**
551 **sheath gene expression (A)** Schematics showing deletions of nucleotides -980 to -849 and -119 to
552 +42. **(B)** Representative image of leaf cross sections of transgenic lines after GUS staining. Zoomed-
553 in images of lateral veins shown in right panels, bundle sheath cells were highlighted with white dash
554 line, scale bars = 50 μm. **(C)** Promoter activity determined by the fluorometric 4-methylumbelliferyl-
555 β-D-glucuronide (MUG) assay, data subjected to a Kruskal-Wallis test, median indicates median
556 value of catalytic rate of GUS protein, n indicates total number of transgenic lines assessed.

557

558 **Figure 3. The Y-patch and four short regions in the distal enhancer are required for bundle**
559 **sheath specific expression. (A-C)** Nucleotides -980 and -829 from the *SiR* promoter pattern
560 expression to the bundle sheath when linked with the *PIP1;1* and *NRT1.1A* core promoters
561 containing Y-patches. **(A)** Prediction of Y-patch and TATA-box sequences in core promoters of
562 *PIP1;1*, *NRT1.1A*, *PIP1;3* and *ATPSb*. **(B)** Representative cross sections of transgenic rice leaves
563 after GUS staining, zoomed-in image of lateral veins shown in the right panel, bundle sheath cells
564 highlighted with white dashed lines, scale bars = 50 μm. **(C)** Promoter activity determined by the
565 fluorometric 4- methylumbelliferyl-β-D-glucuronide (MUG) assay. Data were subjected to a Kruskal-
566 Wallis test, median indicates median value of catalytic rate of GUS protein, n indicates total number
567 of transgenic lines assessed. **(D)** Schematics showing transcription factor binding sites between
568 nucleotides -980 and -829. **(E)** Schematics showing consecutive deletions between nucleotides -980
569 and -829 fused to the GUS reporter. **(F)** Representative images of cross sections from transgenic

570 lines after GUS staining, zoomed-in images of lateral veins shown in right panels, bundle sheath
571 cells highlighted with white dashed lines, scale bars = 50 μm. **(G)** Promoter activity determined by
572 the fluorometric 4-methylumbelliferyl-β-D-glucuronide (MUG) assay. Data were subjected to a
573 Kruskal-Wallis test, median indicates median value of catalytic rate of GUS protein, n indicates total
574 number of transgenic lines assessed.
575
576 **Figure 4. G2-like, MYB-related, IDD and bZIP transcription factors interact and activate the**
577 **distal enhancer. (A)** Schematics showing transcription factor binding sites between nucleotides -
578 980 and -829 which are likely required for bundle sheath specific expression. **(B)** Effector assays
579 showing that each transcription factor activates expression from the distal enhancer. **(C)** Effector
580 assays showing synergistic activation from the distal enhancer when GLK2 and IDD3,4,6,10 were
581 co-expressed. **(D)** Transcript abundance of transcription factors in bundle sheath strands (BSS) and
582 mesophyll (M) cells during maturation. Leaf developmental stage S2 to S7 represent base of the 4th
583 leaf at the 6th, 8th, 9th, 10th, 13th and 17th day after sowing. **(E)** Representative image of transgenic
584 lines expressing WRKY121, IDD2 and bZIP9 in mesophyll cells, zoomed-in image of lateral veins
585 shown in right panel, bundle sheath cells highlighted with white dashed lines.
586
587 **Figure 5. Oligomerisation of the distal enhancer increases bundle sheath expression. (A)**
588 Schematic showing the distal enhancer comprises three short regions and oligomerisation this
589 enhancer, fusing it to the *SiR* core promoter sequence. **(B)** Representative images of cross sections
590 from transgenic lines after GUS staining, zoomed-in image of lateral veins shown in right panel,
591 bundle sheath cells highlighted with white dashed lines, scale bars = 50 μm. **(C)** Promoter activity
592 determined by the fluorometric 4-methylumbelliferyl-β-D-glucuronide (MUG) assay. Data subjected
593 to a Kruskal-Wallis test, median indicates median value of catalytic rate of GUS protein, n indicates
594 total number of transgenic lines assessed.
595
596 **Figure 5. Model depicting mechanism underpinning bundle sheath expression from the *SiR***
597 **promoter in rice.** (**A&B**) Schematics illustrating the bundle sheath cis regulatory module (BS-CRM)
598 present in the distal enhancer, along with a Y-patch, and *trans*-acting factors that can bind and act
599 as activators. (**C**) Schematic indicating how the enhancer operates within a broader *cis*-regulatory
600 landscape including a constitutive activator and mesophyll repressor.

601 **Supplementary Figure legend titles**

602

603 **Supplemental Figure 1.** Analysis of seven rice promoters identified after analysis of transcripts that
604 accumulate preferentially in bundle sheath strands**.**

605 **Supplemental Figure 2.** Identification of bundle sheath specific promoters after analysis of
606 transcripts that accumulate preferentially in Kitaake bundle sheath compared with mesophyll cells.

607 **Supplemental Figure 3.** The domesticated *SiR* promoter combined with the dTALE/STAP system
608 drives strong expression in the rice bundle sheath.

609 **Supplemental Figure 4.** The rice *SiR* promoter drives expression in bundle sheath cells earlier than
610 the *Zoysia japonica PCK* promoter.

611 **Supplemental Figure 5.** Impact of 5' and 3' deletions between on patterning of GUS from the *SiR*
612 promoter.

613 **Supplemental Figure 6.** The evolutionally conserved *ETHYLENE INSENSITIVE3-LIKE* (*EIL*)
614 binding site regulates the expression level but not cell specificity of the *SiR* promoter.

615 **Supplemental Figure 7.** Identification of transcription factors interacting with the *SiR* promoter.

616 **Supplemental Figure 8.** Nucleotides -251 to -1 likely serve as a core promoter.

617 **Supplemental Figure 9.** Subregions in the distal enhancer not required for bundle sheath specific
618 expression between nucleotides -980 to -829.

619 **Supplemental Figure 10**. Regions between nucleotides -980 and -829 that in combination with
620 nucleotides -828 to -252 repress mesophyll expression.

621 **Supplemental Figure 11.** Summary of transactivation assay using rice protoplasts.

622 **Supplemental Figure 12.** Schematic showing mis-expression of WRKY121, IDD2 and bZIP9 in
623 mesophyll cells using the maize *PEPC* promoter driving dTALE1 in combination with STAP56, 45
624 and 62.

625 **Supplemental Figure 13.** Site-directed mutagenesis of transcription factor binding sites in
626 subregions a, b, d and f.

627 **Supplemental Figure 14.** G2-like, MYB-related, IDD and bZIP transcription factor binding sites
628 identified by FIMO program in *HAC1;1*, *ATPSb*, *Fd*, *ZjPCK* and *FtGLDP* promoters. Regions <300
629 bp containing all four motif families highlighted in red.

630 ## References

631 **Aubry S, Kelly S, Kümpers BMCC, Smith-Unna RD, and Hibberd JM**. Deep Evolutionary Comparison of Gene
632 Expression Identifies Parallel Recruitment of Trans-Factors in Two Independent Origins of C_4
633 Photosynthesis. PLoS Genet. 2014a:**10**(6):e1004365. https://doi.org/10.1371/journal.pgen.1004365

634 **Aubry S, Smith-Unna RD, Boursnell CM, Kopriva S, and Hibberd JM**. Transcript residency on ribosomes
635 reveals a key role for the *Arabidopsis thaliana* bundle sheath in sulfur and glucosinolate metabolism.
636 Plant Journal. 2014b:**78**(4):659–673. https://doi.org/10.1111/tpj.12502

637 **Bailey TL, Johnson J, Grant CE, and Noble WS**. The MEME Suite. Nucleic Acids Res. 2015:**43**:39–49.
638 https://doi.org/10.1093/nar/gkv416

639 **Bernard V, Brunaud V, and Lecharny A**. TC-motifs at the TATA-box expected position in plant genes: A
640 novel class of motifs involved in the transcription regulation. BMC Genomics. 2010:**11**(1):1–15.
641 https://doi.org/10.1186/1471-2164-11-166/TABLES/4

642 **Binder BM**. Ethylene signaling in plants. J Biol Chem. 2020:**295**(22):7710.
643 https://doi.org/10.1074/JBC.REV120.010854

644 **Buecker C and Wysocka J**. Enhancers as information integration hubs in development: lessons from
645 genomics. Trends Genet. 2012:**28**(6):276. https://doi.org/10.1016/J.TIG.2012.02.008

646 **Burgess SJ, Reyna-Llorens I, Stevenson SR, Singh P, Jaeger K, and Hibberd JM**. Genome-Wide Transcription
647 Factor Binding in Leaves from C3 and C4 Grasses. Plant Cell. 2019:**31**(10):2297–2314.
648 https://doi.org/10.1105/TPC.19.00078

649 **Castro-Mondragon JA, Jaeger S, Thieffry D, Thomas-Chollier M, and Van Helden J**. RSAT matrix-clustering:
650 dynamic exploration and redundancy reduction of transcription factor binding motif collections.
651 Nucleic Acids Res. 2017:**45**(13):e119–e119. https://doi.org/10.1093/NAR/GKX314

652 **Chen X, Wang Y, Lv B, Li J, Luo L, Lu S, Zhang X, Ma H, and Ming F**. The NAC family transcription factor
653 OsNAP confers abiotic stress response through the ABA pathway. Plant Cell Physiol. 2014:**55**(3):604–
654 619. https://doi.org/10.1093/PCP/PCT204

655 **Chow C-N, Lee T-Y, Hung Y-C, Li G-Z, Tseng K-C, Liu Y-H, Kuo P-L, Zheng H-Q, and Chang W-C**. PlantPAN3.0:
656 a new and updated resource for reconstructing transcriptional regulatory networks from ChIP-seq
657 experiments in plants. Nucleic Acids Res. 2018. https://doi.org/10.1093/nar/gky1081

658 **Civáň P and Švec M**. Genome-wide analysis of rice (Oryza sativa L. subsp. japonica) TATA box and Y Patch
659 promoter elements. Genome. 2009:**52**(3):294–297. https://doi.org/10.1139/G09-
660 001/SUPPL_FILE/G09-001.DOC

661 **Cui H**. Challenges and Approaches to Crop Improvement Through C3-to-C4 Engineering. Front Plant Sci.
662 2021:**12**:715391. https://doi.org/10.3389/FPLS.2021.715391/BIBTEX

663 **Cui H, Kong D, Liu X, and Hao Y**. SCARECROW, SCR-LIKE 23 and SHORT-ROOT control bundle sheath cell fate
664 and function in Arabidopsis thaliana. The Plant Journal. 2014:**78**(2):319–327.
665 https://doi.org/10.1111/TPJ.12470

666 **Danila F, Schreiber T, Ermakova M, Hua L, Vlad D, Lo SF, Chen YS, Lambret-Frotte J, Hermanns AS, Athmer**
667 **B, et al.** A single promoter-TALE system for tissue-specific and tuneable expression of multiple genes
668 in rice. Plant Biotechnol J. 2022a:**20**(9):1786–1806. https://doi.org/10.1111/PBI.13864

669 **Danila F, Schreiber T, Ermakova M, Hua L, Vlad D, Lo SF, Chen YS, Lambret-Frotte J, Hermanns AS, Athmer**
670 **B, et al.** A single promoter-TALE system for tissue-specific and tuneable expression of multiple genes
671 in rice. Plant Biotechnol J. 2022b. https://doi.org/10.1111/PBI.13864

Dickinson PJ, Kneřová J, Szecówka M, Stevenson SR, Burgess SJ, Mulvey H, Bågman AM, Gaudinier A, Brady SM, and Hibberd JM. A bipartite transcription factor module controlling expression in the bundle sheath of Arabidopsis thaliana. Nat Plants. 2020a:**6**(12):1468–1479. https://doi.org/10.1038/s41477-020-00805-w

Dickinson PJ, Kneřová J, Szecówka M, Stevenson SSR, Burgess SJ, Mulvey H, Bågman A-MM, Gaudinier A, Brady SM, and Hibberd JM. A bipartite transcription factor module controlling expression in the bundle sheath of Arabidopsis thaliana. Nat Plants. 2020b:**6**(12):1468–1479. https://doi.org/10.1038/s41477-020-00805-w

Dietzen C, Koprivova A, Whitcomb SJ, Langen G, Jobe TO, Hoefgen R, and Kopriva S. The Transcription Factor EIL1 Participates in the Regulation of Sulfur-Deficiency Response. Plant Physiol. 2020:**184**(4):2120–2136. https://doi.org/10.1104/pp.20.01192

Emmerling J. Studies into the Regulation of C4 Photosynthesis – Towards Factors Controlling Bundle Sheath Expression and Kranz Anatomy Development. Doctoral thesis, Heinrich Heine University, Düsseldorf. 2018:(September).

Engelmann S, Wiludda C, Burscheidt J, Gowik U, Schlue U, Koczor M, Streubel M, Cossu R, Bauwe H, and Westhoff P. The Gene for the P-Subunit of Glycine Decarboxylase from the C4 Species Flaveria trinervia: Analysis of Transcriptional Control in Transgenic Flaveria bidentis (C4) and Arabidopsis (C3). Plant Physiol. 2008:**146**(4):1773–1785. https://doi.org/10.1104/PP.107.114462

Ermakova M, Arrivault S, Giuliani R, Danila F, Alonso-Cantabrana H, Vlad D, Ishihara H, Feil R, Guenther M, Borghi GL, et al. Installation of C_4 photosynthetic pathway enzymes in rice using a single construct. Plant Biotechnol J. 2021:**19**(3):575–588. https://doi.org/10.1111/pbi.13487

Fornes O, Castro-Mondragon JA, Khan A, Van Der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranašić D, et al. JASPAR 2020: Update of the open-Access database of transcription factor binding profiles. Nucleic Acids Res. 2020:**48**(D1):D87–D92. https://doi.org/10.1093/nar/gkz1001

Gallegos JE and Rose AB. An intron-derived motif strongly increases gene expression from transcribed sequences through a splicing independent mechanism in Arabidopsis thaliana. Scientific Reports 2019 9:1. 2019:**9**(1):1–9. https://doi.org/10.1038/s41598-019-50389-5

Glick BR. Metabolic load and heterologous gene expression. Biotechnol Adv. 1995:**13**(2):247–261. https://doi.org/10.1016/0734-9750(95)00004-A

Grant CE, Bailey TL, and Noble WS. FIMO: Scanning for occurrences of a given motif. Bioinformatics. 2011:**27**(7):1017–1018. https://doi.org/10.1093/bioinformatics/btr064

Haberle V and Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. Nat Rev Mol Cell Biol. 2018:**19**(10):621. https://doi.org/10.1038/S41580-018-0028-8

Hibberd JM and Covshoff S. The Regulation of Gene Expression Required for C_4 Photosynthesis. Annu Rev Plant Biol. 2010:**61**(1):181–207. https://doi.org/10.1146/annurev-arplant-042809-112238

Hibberd JM, Sheehy JE, and Langdale JA. Using C4 photosynthesis to increase the yield of rice-rationale and feasibility. Curr Opin Plant Biol. 2008:**11**(2):228–231. https://doi.org/10.1016/j.pbi.2007.11.002

Hua L, Stevenson SR, Reyna-Llorens I, Xiong H, Kopriva S, and Hibberd JM. The bundle sheath of rice is conditioned to play an active role in water transport as well as sulfur assimilation and jasmonic acid synthesis. The Plant Journal. 2021:**107**(1):268–286. https://doi.org/10.1111/TPJ.15292

713 **Jisha V, Dampanaboina L, Vadassery J, Mithöfer A, Kappara S, and Ramanan R**. Overexpression of an
714 AP2/ERF Type Transcription Factor OsEREBP1 Confers Biotic and Abiotic Stress Tolerance in Rice. PLoS
715 One. 2015:**10**(6). https://doi.org/10.1371/JOURNAL.PONE.0127831

716 **Jores T, Tonnies J, Wrightsman T, Buckler ES, Cuperus JT, Fields S, and Queitsch C**. Synthetic promoter
717 designs enabled by a comprehensive analysis of plant core promoters. Nature Plants 2021 7:6.
718 2021:**7**(6):842–855. https://doi.org/10.1038/s41477-021-00932-y

719 **Jung SE, Bang SW, Kim SH, Seo JS, Yoon H Bin, Kim YS, and Kim JK**. Overexpression of OsERF83, a Vascular
720 Tissue-Specific Transcription Factor Gene, Confers Drought Tolerance in Rice. Int J Mol Sci.
721 2021:**22**(14). https://doi.org/10.3390/IJMS22147656

722 **Juven-Gershon T and Kadonaga JT**. Regulation of gene expression via the core promoter and the basal
723 transcriptional machinery. 2009. https://doi.org/10.1016/j.ydbio.2009.08.009

724 **Kajala K, Covshoff S, Karki S, Woodfield H, Tolley BJ, Dionora MJA, Mogul RT, Mabilangan AE, Danila FR,**
725 **Hibberd JM, et al.** Strategies for engineering a two-celled C4 photosynthetic pathway into rice. J Exp
726 Bot. 2011a:**62**(9):3001–3010. https://doi.org/10.1093/jxb/err022

727 **Kajala K, Covshoff S, Karki S, Woodfield H, Tolley BJ, Dionora MJA, Mogul RT, Mabilangan AE, Danila FR,**
728 **Hibberd JM, et al.** Strategies for engineering a two-celled C4photosynthetic pathway into rice. J Exp
729 Bot. 2011b:**62**(9):3001–3010. https://doi.org/10.1093/jxb/err022

730 **Kirschner S, Woodfield H, Prusko K, Koczor M, Gowik U, Hibberd JM, and Westhoff P**. Expression of
731 SULTR2;2, encoding a low-affinity sulphur transporter, in the Arabidopsis bundle sheath and vein cells
732 is mediated by a positive regulator. J Exp Bot. 2018:**69**(20):4897–4906.
733 https://doi.org/10.1093/jxb/ery263

734 **Kummari D, Palakolanu SR, Kishor PBK, Bhatnagar-Mathur P, Singam P, Vadez V, and Sharma KK**. An
735 update and perspectives on the use of promoters in plant genetic engineering. J Biosci. 2020:**45**(1):1–
736 24. https://doi.org/10.1007/S12038-020-00087-6/TABLES/7

737 **Kurt F, Filiz E, and Aydın A**. Sulfite Reductase (SiR) Gene in Rice (Oryza sativa): Bioinformatics and
738 Expression Analyses Under Salt and Drought Stresses. J Plant Growth Regul. 2022:**41**(6):2246–2260.
739 https://doi.org/10.1007/S00344-021-10438-8/FIGURES/9

740 **Lee DY, Hua L, Khoshravesh R, Giuliani R, Kumar I, Cousins A, Sage TL, Hibberd JM, and Brutnell TP**.
741 Engineering chloroplast development in rice through cell-specific control of endogenous genetic
742 circuits. Plant Biotechnol J. 2021:**19**(11):2291–2303. https://doi.org/10.1111/PBI.13660

743 **Leegood RC**. Roles of the bundle sheath cells in leaves of C_3 plants. J Exp Bot. 2008:**59**(7):1663–1673.
744 https://doi.org/10.1093/jxb/erm335

745 **Li E, Liu H, Huang L, Zhang X, Dong X, Song W, Zhao H, and Lai J**. Long-range interactions between proximal
746 and distal regulatory regions in maize. Nat Commun. 2019:**10**(1). https://doi.org/10.1038/S41467-
747 019-10603-4

748 **Li L, Zhu Q, He X, Sinha S, and Halfon MS**. Large-scale analysis of transcriptional cis-regulatory modules
749 reveals both common features and distinct subclasses. Genome Biol. 2007:**8**(6).
750 https://doi.org/10.1186/GB-2007-8-6-R101

751 **Lin R, Zhao W, Meng X, and Peng YL**. Molecular cloning and characterization of a rice gene encoding
752 AP2/EREBP-type transcription factor and its expression in response to infection with blast fungus and
753 abiotic stresses. Physiol Mol Plant Pathol. 2007:**70**(1–3):60–68.
754 https://doi.org/10.1016/J.PMPP.2007.06.002

Liu L, Du Y, Shen X, Li M, Sun W, Huang J, Liu Z, Tao Y, Zheng Y, Yan J, et al. KRN4 Controls Quantitative Variation in Maize Kernel Row Number. PLoS Genet. 2015:**11**(11):e1005670. https://doi.org/10.1371/JOURNAL.PGEN.1005670

Makino A, Sakuma H, Sudo E, and Mae T. Differences between Maize and Rice in N-use Efficiency for Photosynthesis and Protein Allocation.

Maruyama-Nakashita A, Nakamura Y, Tohge T, Saito K, and Takahashi H. *Arabidopsis* SLIM1 is a central transcriptional regulator of plant sulfur response and metabolism. Plant Cell. 2006:**18**(11):3235–3251. https://doi.org/10.1105/tpc.106.046458

Mitchell PL and Sheehy JE. Supercharging rice photosynthesis to increase yield. New Phytol. 2006:**171**(4):688–693. https://doi.org/10.1111/J.1469-8137.2006.01855.X

Nomura M, Higuchi T, Ishida Y, Ohta S, Komari T, Imaizumi N, Miyao-Tokutomi M, Matsuoka M, and Tajima S. Differential Expression Pattern of C_4 Bundle Sheath Expression Genes in Rice, a C_3 Plant. Plant Cell Physiol. 2005a:**46**(5):754–761. https://doi.org/10.1093/pcp/pci078

Nomura M, Higuchi T, Ishida Y, Ohta S, Komari T, Imaizumi N, Miyao-Tokutomi M, Matsuoka M, and Tajima S. Differential expression pattern of C4 bundle sheath expression genes in rice, a C3 plant. Plant Cell Physiol. 2005b:**46**(5):754–761. https://doi.org/10.1093/PCP/PCI078

Nomura M, Higuchi T, Ishida Y, Ohta S, Komari T, Imaizumi N, Miyao-Tokutomi M, Matsuoka M, and Tajima S. Differential expression pattern of C4 bundle sheath expression genes in rice, a C3 plant. Plant Cell Physiol. 2005c:**46**(5):754–761. https://doi.org/10.1093/PCP/PCI078

Page MT, Parry MAJ, and Carmo-Silva E. A high-throughput transient expression system for rice. Plant Cell Environ. 2019:**42**(7):2057–2064. https://doi.org/10.1111/PCE.13542

Qu X, Zou J, Wang J, Yang K, Wang X, and Le J. A Rice R2R3-Type MYB Transcription Factor OsFLP Positively Regulates Drought Stress Response via OsNAC. Int J Mol Sci. 2022:**23**(11). https://doi.org/10.3390/IJMS23115873

Que Q, Wang HY, English JJ, and Jorgensen RA. The Frequency and Degree of Cosuppression by Sense Chalcone Synthase Transgenes Are Dependent on Transgene Promoter Strength and Are Reduced by Premature Nonsense Codons in the Transgene Coding Sequence. Plant Cell. 1997:**9**(8):1357. https://doi.org/10.1105/TPC.9.8.1357

Rask L, Andréasson E, Ekbom B, Eriksson S, Pontoppidan B, and Meijer J. Myrosinase: Gene family evolution and herbivore defense in Brassicaceae. Plant Mol Biol. 2000:**42**(1):93–114. https://doi.org/10.1023/A:1006380021658/METRICS

Ray-Jones H and Spivakov M. Transcriptional enhancers and their communication with gene promoters. Cellular and Molecular Life Sciences. 2021:**78**(19–20):6453. https://doi.org/10.1007/S00018-021-03903-W

Reyna-Llorens I, Burgess SJ, Reeves G, Singh P, Stevenson SR, Williams BP, Stanley S, and Hibberd JM. Ancient duons may underpin spatial patterning of gene expression in C4 leaves. Proc Natl Acad Sci U S A. 2018:**115**(8):1931–1936. https://doi.org/10.1073/PNAS.1720576115/SUPPL_FILE/PNAS.1720576115.SD07.XLSX

Rosa EAS and Rodrigues PMF. Towards a more sustainable agriculture system: The effect of glucosinolates on the control of soil-borne diseases. Journal of Horticultural Science and Biotechnology. 1999:**74**(6):667–674. https://doi.org/10.1080/14620316.1999.11511170/ASSET//CMS/ASSET/B3CB692A-CEFD-4DF3-8FBE-D467F93CF12C/14620316.1999.11511170.FP.PNG

798 **Rose AB**. Introns as Gene Regulators: A Brick on the Accelerator. Front Genet. 2019:**9**(FEB).
799 https://doi.org/10.3389/FGENE.2018.00672

800 **Rose AB, Elfersi T, Parra G, and Korf I**. Promoter-proximal introns in Arabidopsis thaliana are enriched in
801 dispersed signals that elevate gene expression. Plant Cell. 2008:**20**(3):543–551.
802 https://doi.org/10.1105/TPC.107.057190

803 **Sage RF**. The evolution of C4 photosynthesis. New Phytologist. 2004a:**161**(2):341–370.

804 **Sage RF**. The evolution of C_4 photosynthesis. New Phytologist. 2004b:**161**(2):341–370.
805 https://doi.org/10.1111/j.1469-8137.2004.00974.x

806 **Schuler ML, Mantegazza O, and Weber APM**. Engineering C4 photosynthesis into C3 chassis in the
807 synthetic biology age. The Plant Journal. 2016:**87**(1):51–65. https://doi.org/10.1111/TPJ.13155

808 **Shimizu H, Sato K, Berberich T, Miyazaki A, Ozaki R, Imai R, and Kusano T**. LIP19, a basic region leucine
809 zipper protein, is a Fos-like molecular switch in the cold signaling of rice plants. Plant Cell Physiol.
810 2005:**46**(10):1623–1634. https://doi.org/10.1093/PCP/PCI178

811 **Shlyueva D, Stampfel G, and Stark A**. Transcriptional enhancers: from properties to genome-wide
812 predictions. Nat Rev Genet. 2014:**15**(4):272–286. https://doi.org/10.1038/NRG3682

813 **Smale ST and Kadonaga JT**. The RNA polymerase II core promoter. Annu Rev Biochem. 2003:**72**(Volume 72,
814 2003):449–479. https://doi.org/10.1146/ANNUREV.BIOCHEM.72.121801.161520/CITE/REFWORKS

815 **Sparks EE, Drapek C, Gaudinier A, Li S, Ansariola M, Shen N, Hennacy JH, Zhang J, Turco G, Petricka JJ, et**
816 **al.** Establishment of Expression in the SHORTROOT-SCARECROW Transcriptional Cascade through
817 Opposing Activities of Both Activators and Repressors. Dev Cell. 2016:**39**(5):585–596.
818 https://doi.org/10.1016/j.devcel.2016.09.031

819 **Spitz F and Furlong EEM**. Transcription factors: from enhancer binding to developmental control. Nat Rev
820 Genet. 2012:**13**(9):613–626. https://doi.org/10.1038/NRG3207

821 **Stergachis AB, Haugen E, Shafer A, Fu W, Vernot B, Reynolds A, Raubitschek A, Ziegler S, LeProust EM,**
822 **Akey JM, et al.** Exonic transcription factor binding directs codon choice and impacts protein evolution.
823 Science. 2013:**342**(6164):1367. https://doi.org/10.1126/SCIENCE.1243490

824 **Studer A, Zhao Q, Ross-Ibarra J, and Doebley J**. Identification of a functional transposon insertion in the
825 maize domestication gene tb1. Nat Genet. 2011:**43**(11):1160. https://doi.org/10.1038/NG.942

826 **Swift J, Luginbuehl LH, Schreier TB, Donald RM, Lee TA, Nery JR, Ecker JR, and Hibberd JM**. Single nuclei
827 sequencing reveals C4 photosynthesis is based on rewiring of ancestral cell identity networks. bioRxiv.
828 2023:2023.10.26.562893. https://doi.org/10.1101/2023.10.26.562893

829 **Tezuka D, Kawamata A, Kato H, Saburi W, Mori H, and Imai R**. The rice ethylene response factor OsERF83
830 positively regulates disease resistance to Magnaporthe oryzae. Plant Physiol Biochem. 2019:**135**:263–
831 271. https://doi.org/10.1016/J.PLAPHY.2018.12.017

832 **Wang L, Wan MC, Liao RY, Xu J, Xu ZG, Xue HC, Mai YX, and Wang JW**. The maturation and aging
833 trajectory of Marchantia polymorpha at single-cell resolution. Dev Cell. 2023:**58**(15):1429-1444.e6.
834 https://doi.org/10.1016/J.DEVCEL.2023.05.014

835 **Wang P, Khoshravesh R, Karki S, Tapia R, Balahadia CP, Bandyopadhyay A, Quick WP, Furbank R, Sage TL,**
836 **and Langdale JA**. Re-creation of a Key Step in the Evolutionary Switch from C_3 to C_4 Leaf Anatomy.
837 Current Biology. 2017a:**27**(21):3278-3287.e6. https://doi.org/10.1016/j.cub.2017.09.040

838 **Wang P, Khoshravesh R, Karki S, Tapia R, Balahadia CPP, Bandyopadhyay A, Quick WPP, Furbank R, Sage**
839 **TLTL, and Langdale JAJA**. Re-creation of a Key Step in the Evolutionary Switch from C_3 to C_4 Leaf
840 Anatomy. Current Biology. 2017b:**27**(21):3278-3287.e6. https://doi.org/10.1016/j.cub.2017.09.040

841 **Wang Y, Huan Q, Li K, and Qian W**. Single-cell transcriptome atlas of the leaf and root of rice seedlings.
842 Journal of Genetics and Genomics. 2021:**48**(10):881–898. https://doi.org/10.1016/J.JGG.2021.06.001

843 **Wiludda C, Schulze S, Gowik U, Engelmann S, Koczor M, Streubel M, Bauwe H, and Westhoff P**. Regulation
844 of the Photorespiratory GLDPA Gene in C4 Flaveria: An Intricate Interplay of Transcriptional and
845 Posttranscriptional Processes. Plant Cell. 2012:**24**(1):137. https://doi.org/10.1105/TPC.111.093872

846 **Xiong H, Hua L, Reyna-Llorens I, Shi Y, Chen KM, Smirnoff N, Kromdijk J, and Hibberd JM**. Photosynthesis-
847 independent production of reactive oxygen species in the rice bundle sheath during high light is
848 mediated by NADPH oxidase. Proc Natl Acad Sci U S A. 2021:**118**(25):e2022702118.
849 https://doi.org/10.1073/PNAS.2022702118/SUPPL_FILE/PNAS.2022702118.SD04.XLSX

850 **Yaish MW, El-Kereamy A, Zhu T, Beatty PH, Good AG, Bi YM, and Rothstein SJ**. The APETALA-2-like
851 transcription factor OsAP2-39 controls key interactions between abscisic acid and gibberellin in rice.
852 PLoS Genet. 2010:**6**(9). https://doi.org/10.1371/JOURNAL.PGEN.1001098

853 **Yamamoto Y, Ichida H, Abe T, Suzuki Y, Sugano S, and Obokata J**. Differentiation of core promoter
854 architecture between plants and mammals revealed by LDSS analysis. Nucleic Acids Res.
855 2007a:**35**(18):6219. https://doi.org/10.1093/NAR/GKM685

856 **Yamamoto YY, Ichida H, Matsui M, Obokata J, Sakurai T, Satou M, Seki M, Shinozaki K, and Abe T**.
857 Identification of plant promoter constituents by analysis of local distribution of short sequences. BMC
858 Genomics. 2007b:**8**(1):1–23. https://doi.org/10.1186/1471-2164-8-67/FIGURES/9

859 **Yan W, Chen D, Schumacher J, Durantini D, Engelhorn J, Chen M, Carles CC, and Kaufmann K**. Dynamic
860 control of enhancer activity drives stage-specific gene expression during flower morphogenesis. Nat
861 Commun. 2019:**10**(1). https://doi.org/10.1038/S41467-019-09513-2

862 **Zhang C, Ding Z, Wu K, Yang L, Li Y, Yang Z, Shi S, Liu X, Zhao S, Yang Z, et al.** Suppression of Jasmonic Acid-
863 Mediated Defense by Viral-Inducible MicroRNA319 Facilitates Virus Infection in Rice. Mol Plant.
864 2016:**9**(9):1302–1314. https://doi.org/10.1016/J.MOLP.2016.06.014

865 **Zhang C, Zhang J, Liu H, Qu X, Wang J, He Q, Zou J, Yang K, and Le J**. Transcriptomic analysis reveals the
866 role of FOUR LIPS in response to salt stress in rice. Plant Mol Biol. 2022:**110**(1–2):37–52.
867 https://doi.org/10.1007/S11103-022-01282-9/FIGURES/7

868 **Zhang H, Zhang J, Quan R, Pan X, Wan L, and Huang R**. EAR motif mutation of rice OsERF3 alters the
869 regulation of ethylene biosynthesis and drought tolerance. Planta. 2013:**237**(6):1443–1451.
870 https://doi.org/10.1007/S00425-013-1852-X

871 **Zhang X, Bao Y, Shan D, Wang Z, Song X, Wang Z, Wang J, He L, Wu L, Zhang Z, et al.** Magnaporthe oryzae
872 Induces the Expression of a MicroRNA to Suppress the Immune Response in Rice. Plant Physiol.
873 2018:**177**(1):352. https://doi.org/10.1104/PP.17.01665

874 **Zhao H, Yang M, Bishop J, Teng Y, Cao Y, Beall BD, Li S, Liu T, Fang Q, Fang C, et al.** Identification and
875 functional validation of super-enhancers in Arabidopsis thaliana. Proc Natl Acad Sci U S A.
876 2022:**119**(48):e2215328119.
877 https://doi.org/10.1073/PNAS.2215328119/SUPPL_FILE/PNAS.2215328119.SD02.XLSX

878

Figure 1

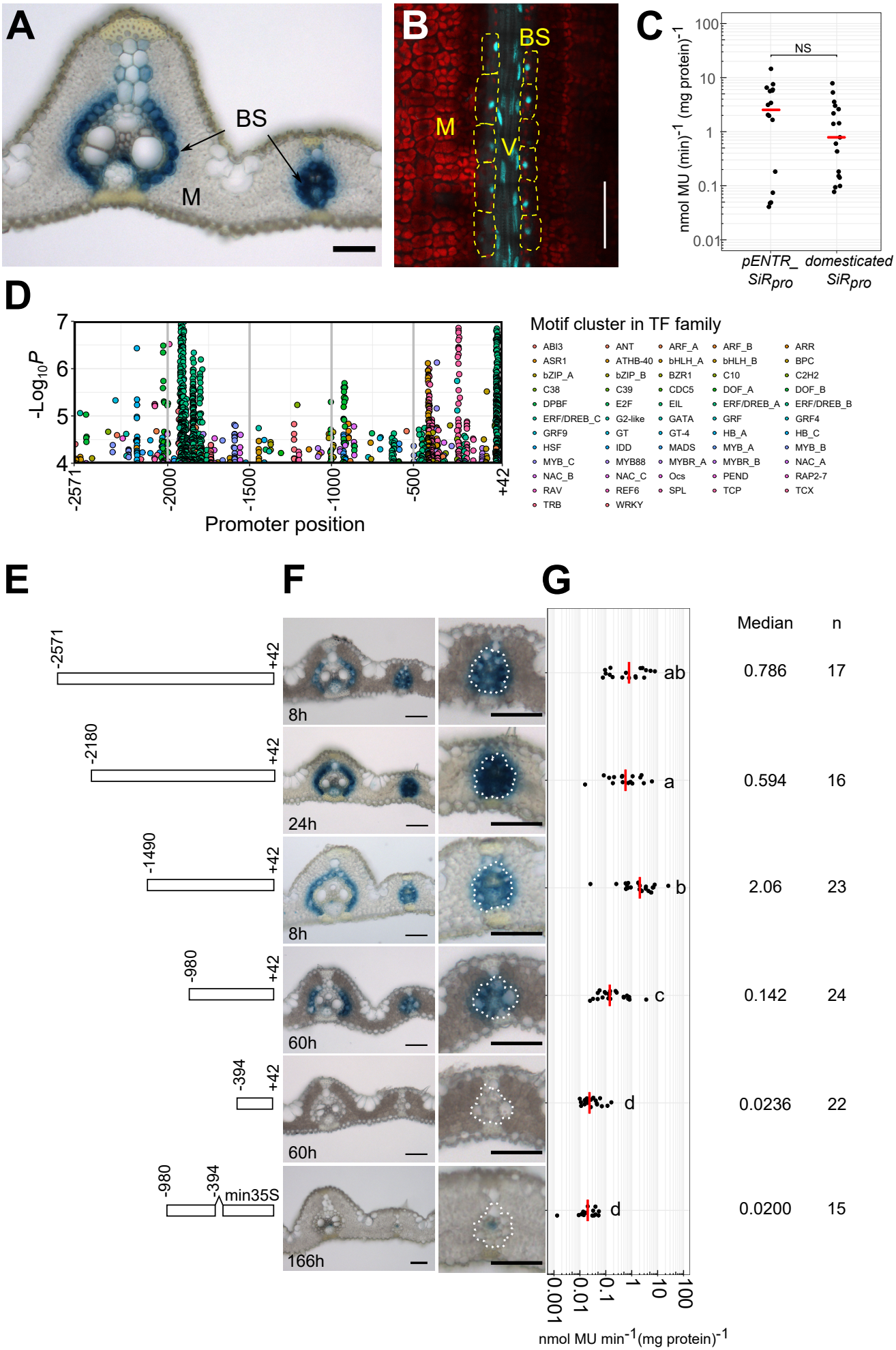

A
BS
M
B
BS
M
V
C
NS
nmol MU (min)$^{-1}$ (mg protein)$^{-1}$
100
10
1
0.1
0.01
pENTR_ SiR$_{pro}$
domesticated SiR$_{pro}$
D
-Log$_{10}$P
7
6
5
4
-2571
-2000
-1500
-1000
-500
+42
Promoter position
Motif cluster in TF family
ABI3
ASR1
bZIP_A
C38
DPBF
ERF/DREB_C
GRF9
HSF
MYB_C
NAC_B
RAV
TRB
ANT
ATHB-40
bZIP_B
C39
E2F
G2-like
GT
IDD
MYB88
NAC_C
REF6
WRKY
ARF_A
bHLH_A
BZR1
CDC5
EIL
GATA
GT-4
MADS
MYBR_A
Ocs
SPL
ARF_B
bHLH_B
C10
DOF_A
ERF/DREB_A
GRF
HB_A
MYB_A
MYBR_B
PEND
TCP
ARR
BPC
C2H2
DOF_B
ERF/DREB_B
GRF4
HB_C
MYB_B
NAC_A
RAP2-7
TCX
E
F
G
-2571
+42
8h
-2180
+42
24h
-1490
+42
8h
-980
+42
60h
-394
+42
60h
-980
-394
min35S
166h
Median
n
ab
0.786
17
a
0.594
16
b
2.06
23
c
0.142
24
d
0.0236
22
d
0.0200
15
0.001
0.01
0.1
1
10
100
nmol MU min^{-1}(mg protein)$^{-1}$

Figure 2

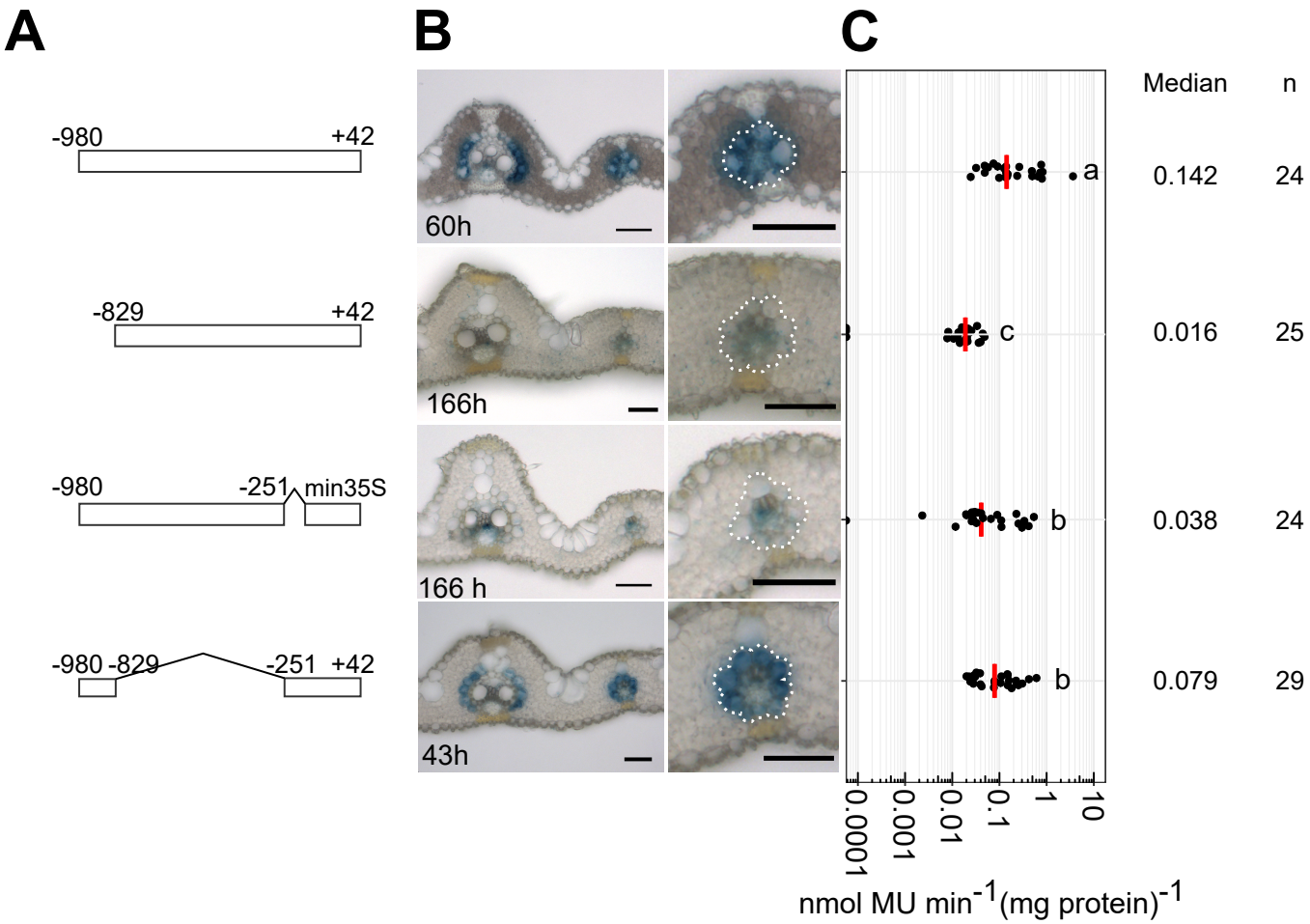

A
-980
+42
-829
+42
-980
-251
min35S
-980 -829
-251 +42
B
60h
166h
166 h
43h
C
Median
n
a
0.142
24
c
0.016
25
b
0.038
24
b
0.079
29
0.0001
0.001
0.01
0.1
1
10
nmol MU min-1(mg protein)-1

Figure 3

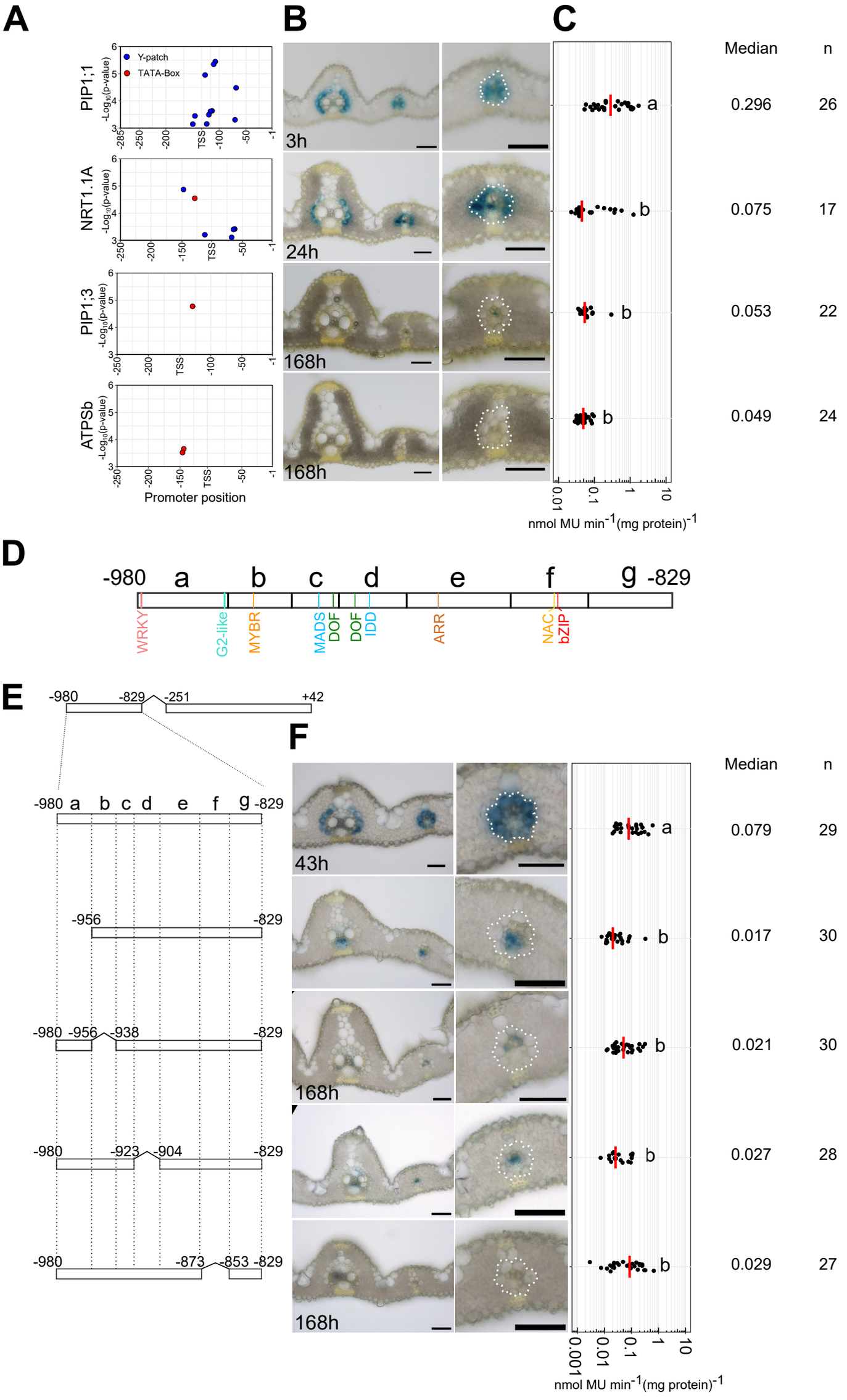

A
B
C
PIP1;1
NRT1.1A
PIP1;3
ATPSb
Y-patch
TATA-Box
-Log10(p-value)
Promoter position
TSS
3h
24h
168h
168h
Median
n
a
0.296
26
b
0.075
17
b
0.053
22
b
0.049
24
nmol MU min-1(mg protein)-1
D
-980
a
b
c
d
e
f
g
-829
WRKY
G2-like
MYBR
MADS
DOF
DOF
IDD
ARR
NAC
bZIP
E
-980
-829
-251
+42
-956
-938
-923
-904
-873
-853
F
43h
168h
168h
a
0.079
29
b
0.017
30
b
0.021
30
b
0.027
28
b
0.029
27
nmol MU min-1(mg protein)-1

Figure 4

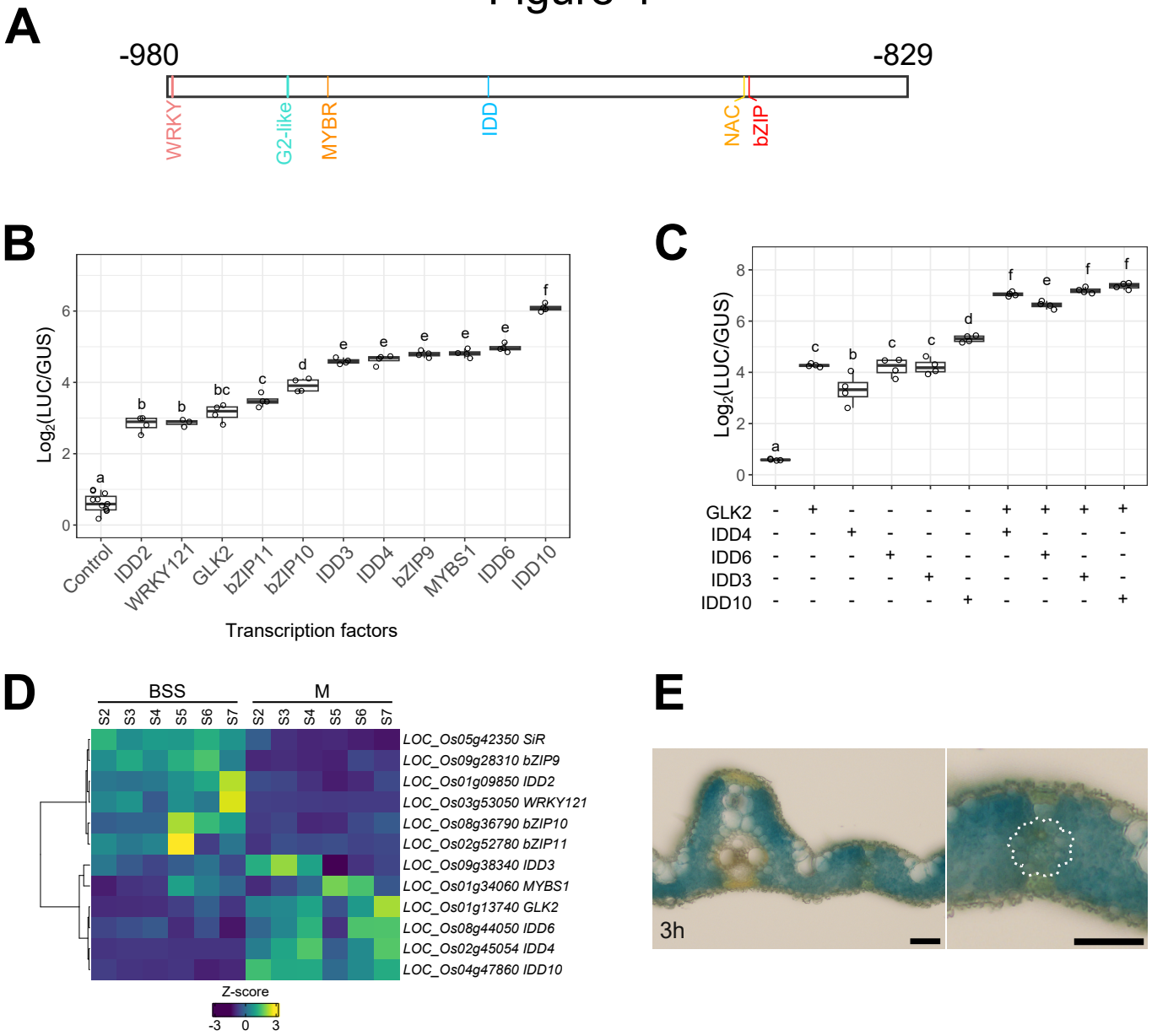

A
-980
-829
WRKY
G2-like
MYBR
IDD
NAC
bZIP
B
Log2(LUC/GUS)
Control
IDD2
WRKY121
GLK2
bZIP11
bZIP10
IDD3
IDD4
bZIP9
MYBS1
IDD6
IDD10
Transcription factors
C
Log2(LUC/GUS)
GLK2
IDD4
IDD6
IDD3
IDD10
D
BSS
M
S2
S3
S4
S5
S6
S7
LOC_Os05g42350 SiR
LOC_Os09g28310 bZIP9
LOC_Os01g09850 IDD2
LOC_Os03g53050 WRKY121
LOC_Os08g36790 bZIP10
LOC_Os02g52780 bZIP11
LOC_Os09g38340 IDD3
LOC_Os01g34060 MYBS1
LOC_Os01g13740 GLK2
LOC_Os08g44050 IDD6
LOC_Os02g45054 IDD4
LOC_Os04g47860 IDD10
Z-score
-3
0
3
E
3h

Figure 5

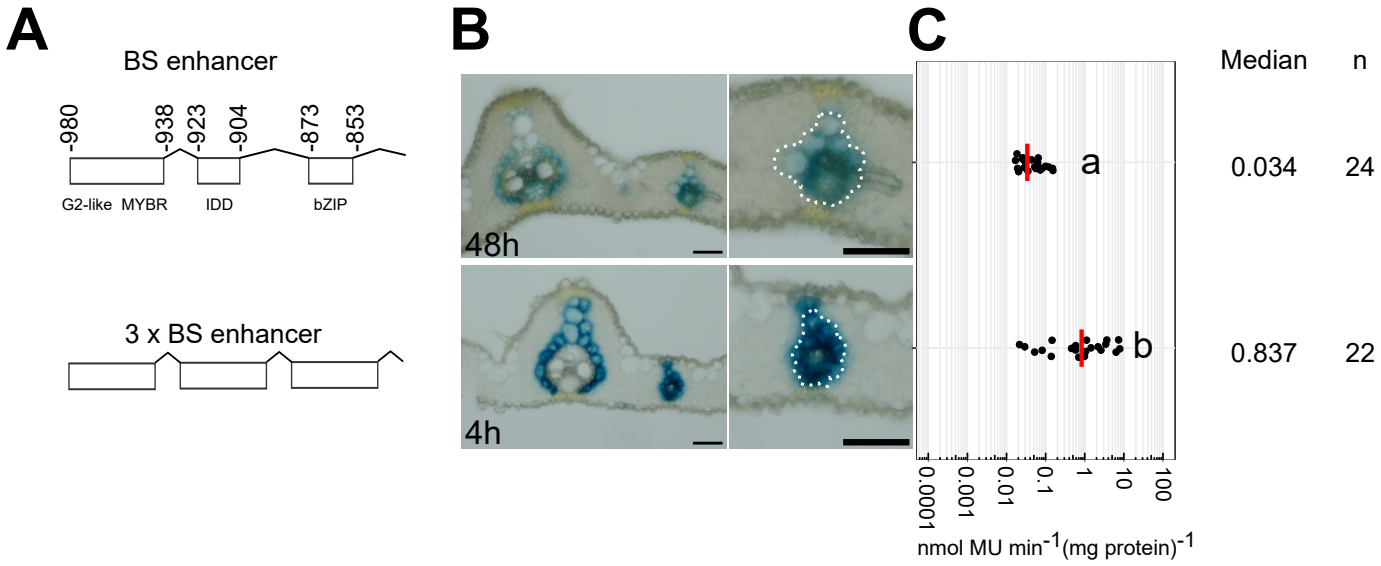
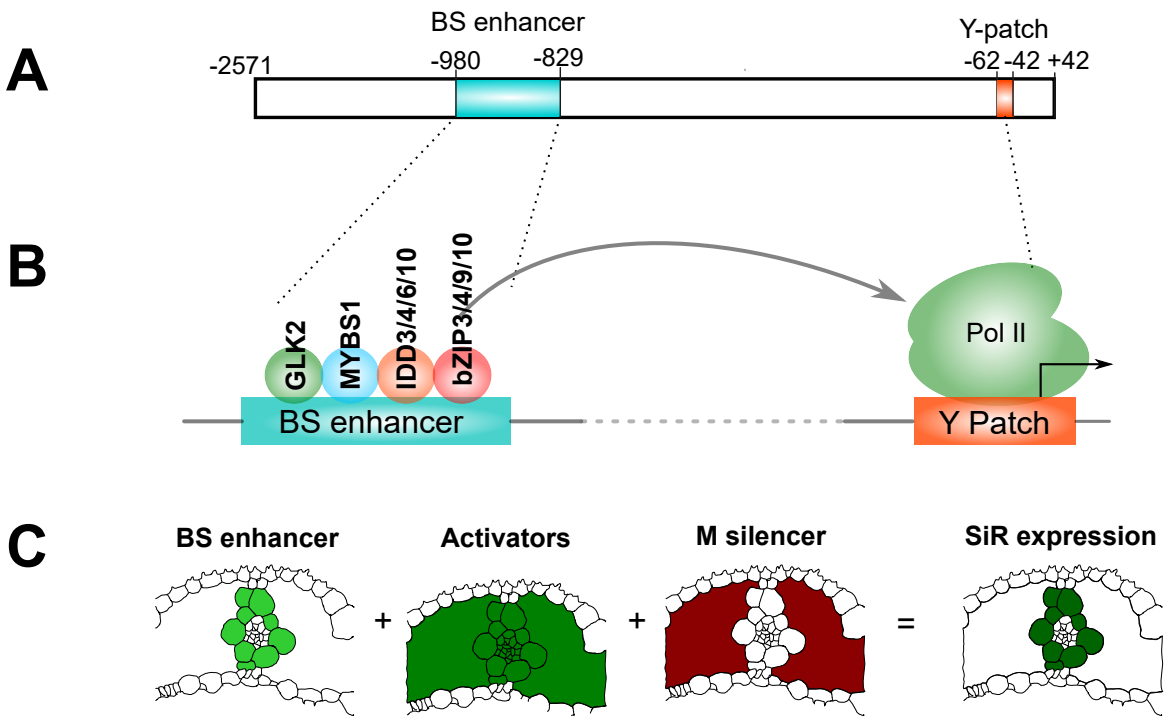

A
BS enhancer
-980
-938
-923
-904
-873
-853
G2-like
MYBR
IDD
bZIP
3 x BS enhancer
B
48h
4h
C
a
b
0.0001
0.001
0.01
0.1
1
10
100
nmol MU min-1(mg protein)-1
Median
n
0.034
24
0.837
22

Figure 6


A
BS enhancer
Y-patch
-2571
-980
-829
-62
-42
+42
B
GLK2
MYBS1
IDD3/4/6/10
bZIP3/4/9/10
BS enhancer
Pol II
Y Patch
C
BS enhancer
Activators
M silencer
SiR expression
+
+
=