Identification of plant enhancers and their constituent elements by STARR-seq in tobacco leaves

Tobias Jores[1], Michael W Dorrity[1], Josh T Cuperus[1]*, Stanley Fields[1,2]* and Christine Queitsch[1]*

[1] Department of Genome Sciences, University of Washington, Seattle, WA, USA.
[2] Department of Medicine, University of Washington, Seattle, WA, USA.

* Correspondence to:

Josh T Cuperus: cuperusj@uw.edu

Stanley Fields: fields@uw.edu

Christine Queitsch: queitsch@uw.edu

Introductory paragraph

In a time of climate change and increasing human population, crop plants with higher yields and improved response to abiotic stresses will be required to ensure food security. As many of the beneficial traits in domesticated crops are caused by mutations in *cis*-regulatory elements, especially enhancers, genetic engineering of such elements is a promising strategy for improving crops[1,2]. However, such engineering is currently hindered by our limited knowledge of plant *cis*-regulatory elements. Here, we adapted STARR-seq — a technology for the high-throughput identification of enhancers — for its use in transiently transfected tobacco leaves. We demonstrate that the optimal placement in the reporter construct of enhancer sequences from a plant virus, pea and wheat was just upstream of a minimal promoter, and that none of these four known enhancers was active in the 3′-UTR of the reporter gene. The optimized assay sensitively identified small DNA regions containing each of the four enhancers, including two whose activity was stimulated by light. Furthermore, we coupled the assay to saturation mutagenesis to pinpoint functional regions within an enhancer, which we recombined to create synthetic enhancers. Our results describe an approach to define enhancer properties that can be performed in any plant species or tissue transformable by *Agrobacterium* and that can use regulatory DNA derived from any plant.

Main

Enhancers can be identified by self-transcribing active regulatory region sequencing (STARR-seq), a massively parallel reporter assay[3]. Here, candidate enhancer sequences are inserted into the 3' untranslated region (3′-UTR) of a reporter gene under the control of a minimal promoter. If an insert has enhancer activity, it can upregulate its own transcription. The resulting transcript can be detected by next generation sequencing and linked to its corresponding enhancer element, which is incorporated within the mRNA. This method has been widely used in *Drosophila* and human cells. In plants, STARR-seq has been described in only two studies that apply the method to the monocot species rice and maize[4,5]. Both studies relied on species-specific protoplasts as recipient cells for the assay. Our approach bypasses the need for a species-specific protoplasting protocol by using transient expression of diverse STARR-seq libraries in tobacco leaves. As transcription factors are highly conserved among plant species[6,7], the versatile tobacco system can serve as a proxy for many plant species, including crops.

We used a green fluorescent protein (GFP) reporter gene under control of the Cauliflower mosaic virus 35S minimal promoter and the 35S core enhancer[8,9] (subdomains A1 and B1-3) to benchmark the assay. We systematically analyzed the position- and orientation-dependency of the enhancer (Fig. 1a). To this end, we used a more generalized version of STARR-seq in which we placed a barcode in the GFP open reading frame. This barcode is linked to the corresponding enhancer variant by next generation sequencing and serves as a readout for the activity of this variant. As expected, the 35S enhancer was active in either orientation and both up- and downstream of the reporter gene (Fig. 1b). However, in contrast to the mammalian system, when placed in the 3′-UTR the enhancer had almost no activity. Addition of a second copy of the enhancer in the 'downstream' and 'distal upstream' positions led on average to a 70% increase in transcript levels as compared to a single enhancer, while a second copy in the 'upstream' position increased transcript levels by only 30% (Fig. 1b). These observations suggest that the transcriptional activation caused by a single 35S enhancer directly upstream of the minimal promoter is already close to the maximum level detectable in our assay.

We observed the strongest activation of transcription with the enhancer immediately upstream of the minimal 35S promoter, and lower levels when the enhancer was placed about 1.5 kb away from the promoter as in the 'downstream' and 'distal upstream' constructs (Fig. 1a,b). To characterize the distance–activity relationship, we inserted the 35S enhancer at different positions within a 2 kb spacer upstream of the minimal promoter (Fig. 1c). Enhancer activity was strongest immediately upstream of the promoter. However, enhancer activity was greatly reduced by 500 bp or more of spacer between the enhancer and promoter (Fig. 1c). To test if the observed position-dependency is unique to the 35S enhancer, we assayed three additional enhancers derived from the *Pisum sativum* AB80 and rbcS-E9 genes and the wheat Cab-1 gene[10–12]. Similar to the 35S enhancer, these enhancers were orientation-independent and most active immediately upstream of the promoter, and did not activate transcription when placed in the 3′-UTR (Fig. 1d).

Although previous STARR-seq studies placed candidate enhancer fragments in the 3′-UTR[3–5], enhancers in this position were not active in our system. To test if enhancers in the 3′-UTR destabilize the mRNA by promoting nonsense-mediated decay, we inserted the unstructured region from the Turnip crinkle virus 3′-

UTR, shown to reduce nonsense-mediated decay[13], in between the stop codon and the 35S enhancer. However, insertion of this region further reduced transcript levels when the 35S enhancer was placed in the 3′-UTR (Supplementary Fig. 1a,b). We next asked whether insertion of the 35S enhancer in an intron, which would also be transcribed, could confer transcriptional activation, but found that it did not (Supplementary Fig. 1a,c). Furthermore, combining an upstream AB80 enhancer with a 35S enhancer within the 3′-UTR transcribed region considerably reduced transcription compared to that from the AB80 enhancer alone (Supplementary Fig. 1d). Taken together, these findings demonstrate that the 35S enhancer residing within the transcribed region is not active in our system. Therefore, for subsequent experiments, we placed the enhancer fragments directly upstream of the minimal promoter, barcoding the reporter amplicons to enable detection by RNA-seq. A similar approach with a barcode in the transcript was used in previous studies of enhancers in human cells[14,15].

The AB80, Cab-1 and rbcS-E9 enhancers are activated by light[10–12]. We tested their light-dependency in our assay system by placing the transformed plants in the dark prior to mRNA extraction. The AB80 and Cab-1 enhancers demonstrated decreased activity in the dark. Although the activity of the rbcS-E9 enhancer also showed a response to light, in this case the activity was higher in the dark (Supplementary Fig. 2). A previous study found higher expression of *Arabidopsis thaliana rbcS* genes in extracts from dark-grown plant cells compared to those from light-grown ones, with reversal of this tendency upon reconstitution of chromatin[16]. Much like this *in vitro* system, our *Agrobacterium*-mediated transient assay also lacks chromatin.

Next, we tested if the assay could detect enhancer signatures among randomly fragmented DNA sequences from a plasmid containing embedded enhancers. We constructed a plasmid harboring the 35S, AB80, Cab-1, and rbcS-E9 enhancers. We fragmented the plasmid using Tn5 transposase and inserted the fragments upstream of the 35S minimal promoter to generate a fragment library for use in the STARR-seq assay (Fig. 2a). The assay identified the known enhancers as the regions with highest enrichment values (Fig. 2b). As expected, the orientation in which the fragments were cloned into the STARR-seq plasmid did not affect their enrichment (Supplementary Fig. 3a). The assay was highly reproducible, with good correlation across replicates for the individual barcodes (Spearman's ρ = 0.80 –

0.82, Supplementary Fig. 3b). The correlation further improved if the enrichment of all barcodes linked to the same fragment was aggregated (Spearman's ρ = 0.80 – 0.85, Supplementary Fig. 3c). Replicate correlations were similar for all STARR-seq experiments in this study (Spearman's $\rho \geq 0.6$ for barcodes and ≥ 0.7 for fragments or variants).

We also used the fragment library in a STARR-seq experiment with plants kept in the dark prior to mRNA extraction to test for light-dependency. We observed the expected changes in enrichment (Supplementary Fig. 2), with the AB80 and Cab-1 enhancers less active and the rbcS-E9 enhancer more active in the light-deprived plants (Fig. 2b,c). We conclude that the STARR-seq assay established in this study can identify enhancers in a condition-specific manner.

To further reveal individual elements of enhancers, we repeated the screen with a second library that contained shorter fragments (median length 84 bp vs. 191 bp in the initial library, Fig. 2d). As these shorter fragments were, on average, well below the size of the full-length enhancers, they are unlikely to contain all the elements required for maximum activity. The shorter fragments split the peaks of the AB80 and Cab-1 enhancers into two subpeaks, suggesting that these enhancers contain at least two independent functional elements. The sole functional element of the rbcS-E9 enhancer resided in the 3′ half of the tested region (Fig. 2d,e).

Having established the capacity of the assay to distinguish enhancer subdomains, we tested its suitability for conducting saturation mutagenesis of *cis*-regulatory elements. To do so, we array-synthesized all possible single nucleotide substitution, deletion, and insertion variants of the 35S enhancer and of the minimal promoter as two separate variant pools, and subjected the two pools to STARR-seq. We first assayed the activity of variants of a 46 bp region containing the 35S minimal promoter, in constructs with and without an enhancer. The effects of the individual mutations were similar in both contexts (Supplementary Fig. 4). As expected, mutations that disrupt the TATA box (positions 16–22) had a strong negative impact on promoter activity, while most others had a weak effect or no effect (Fig. 3a,b, Supplementary Table 1).

In contrast to the minimal promoter, the 35S enhancer contained several mutation-sensitive regions (Fig. 3c,d, Supplementary Table 2). These regions co-localize with predicted transcription factor binding

sites[17,18]. Mutations in positions 116–135 were especially deleterious. This region, previously implicated in enhancer activity, can be bound by the tobacco activation sequence factor 1 (ASF-1), a complex containing the bZIP transcription factor TGA2.2 (ref. [8,9,19,20]). Similarly, we observed mutational sensitivity of the 35S enhancer in positions 95–115, which contain a binding site of the bHLH transcription factor complex ASF-2 (ref. [21]). A third mutation-sensitive region in positions 7–28 is predicted to be bound by ERF and TCP transcription factors.

To demonstrate that these mutation-sensitive regions possess enhancer activity, we split the enhancer into three fragments that span positions 1–30 (A), 60–105 (B), and 106–140 (C) (Fig. 3d). These fragments were cloned in one to four copies on average, in random order, and the enhancer activity of the resulting constructs was determined. Fragments A and C alone were sufficient to activate transcription, while fragment B was active only in the presence of a second fragment (Fig. 3e). In line with our observations from the enhancer mutagenesis, fragment C had the highest activity. The greater the number of fragments in a construct, the higher was its activity. However, even four fragments combined did not reach the level of transcription achieved with the full-length enhancer, indicating that the sequences excluded from the A, B and C fragments contribute to enhancer activity, either directly or by providing the correct spacing for the fragments (Fig. 3e). Although spacing may play a role in enhancer activity, the order of the fragments had only weak effects (Supplementary Fig. 5). Taken together, we demonstrate that this assay can identify functional enhancer elements that can be recombined to create synthetic enhancers of varying strength.

In this study, we developed a massively parallel reporter assay that can identify DNA regions with enhancer or promoter activity and can dissect these regions to pinpoint functional sequences with single nucleotide resolution. Moreover, the assay can detect enhancer condition-specificity (light dependence), although it does not recapitulate regulation mediated by chromatin and its modifications. Apart from enhancers and promoters, the assay potentially can be adapted for other types of *cis*-regulatory elements. The assay does not depend on efficient protoplasting and transformation protocols, which have been established only for a limited number of species and tissues. While these experiments used tobacco leaves, the assay can likely be performed in any species or tissue that can be transiently transformed by

Agrobacterium[22–25]. The assay is applicable to enhancer screens for any plant species to analyze plant gene regulation and to identify promising building blocks for future genetic engineering efforts. The data generated by these screens and subsequent saturation mutagenesis will enable deep learning approaches to identify defining characteristics of plant enhancers.

Methods

Plasmid construction and library creation

The STARR-seq plasmid used herein is based on the pGreen plasmid[26]. In its T-DNA region, it harbors a phosphinothricin resistance gene (BlpR) and the GFP reporter construct under control of the 35S minimal promoter and terminated by the polyA site of the *Arabidopsis thaliana* ribulose bisphosphate carboxylase small chain 1A gene. The synthetic 5′-UTR synJ[27] (ACACGCTGGAATTCTAGTATACTAAACC), an ATG start codon and a 15 bp random barcode (VNNVNNVNNVNNVNN) was cloned between the minimal promoter and the second codon of GFP. Enhancers or DNA fragments were inserted by Golden Gate cloning[28] into the indicated positions. For constructs with full-length enhancers (Fig. 1b-d and Supplementary Figs. 1 and 2), 5 – 10 uniquely barcoded variants were used. The sequences of the 35S, AB80, Cab-1, and rbcS-E9 enhancers are listed in Supplementary Table 3. These enhancers were inserted into the SacI, XbaI, XhoI, and SfoI sites of the plasmid pZS*11-yfp0 (ref. [29], Addgene #53241, https://www.addgene.org/53241/). The resulting plasmid (pZS*11_4enh) was fragmented with Nextera Tn5 transposase (Illumina) and the fragments were amplified with primers containing adapters suitable for Golden Gate cloning. The single nucleotide variants of the 35S promoter and enhancer were ordered as an oligonucleotide array from Twist Bioscience (Fig. 3a–d). The libraries were bottlenecked to approximately 10 barcodes per variant. Fragments A, B, and C (Fig. 3d) were ordered as oligonucleotide with 5′ GTGATG overhangs, mixed with Golden Gate cloning adaptors and ligated with T4 DNA ligase. The IV2 intron was inserted into GFP at position 103. The unstructured region of the Turnip crinkle virus 3′-UTR (TACGGTAATAGTGTAGTCTTCTCATCTTAGTAGTTAGCTCTCTCTTATATT) was inserted after the GFP stop codon. Next generation sequencing on an Illumina NextSeq platform was used to link the inserted fragments and barcodes. The STARR-seq plasmid libraries were introduced into *Agrobacterium tumefaciens* GV3101 strain harboring the helper plasmid pSoup[26] by electroporation.

Plant cultivation and transformation

Tobacco (*Nicotiana benthamiana*) was grown at 25 °C in a long-day photoperiod (16 h light and 8 h dark). Plants were transformed 3 – 4 weeks after germination. For transformation, an overnight culture of *Agrobacterium tumefaciens* was diluted into 50 mL YEP medium and grown at 28 °C to an OD of approximately 1. A 5 mL input sample of the cells was taken, and plasmids were isolated from it. The remaining cells were harvested and resuspended in 50 mL induction medium (M9 medium with supplemented 1 % glucose, 10 mM MES, pH 5.2, 100 µM $CaCl_2$, 2 mM $MgSO_4$, and 100 µM acetosyringone). After overnight growth, the bacteria were harvested, resuspended in infiltration solution (10 mM MES, pH 5.2, 10 mM $MgCl_2$, 150 µM acetosyringone, and 5 µM lipoic acid) to an OD of 1 and infiltrated into the first two mature leaves of 2 – 4 tobacco plants. The plants were further grown for 48 h under normal conditions or in the dark prior to mRNA extraction.

STARR-seq assay

For each experiment at least 3 independent biological replicates were performed. Depending on library size, two samples of 2 – 3 leaves were collected from a total of 2 – 4 plants. They were frozen in liquid nitrogen, ground in a mortar and immediately resuspended in 5 mL TRIzol (Thermo Fisher Scientific). The suspension was cleared by centrifugation (5 min at 4000 x g, 4 °C), and the supernatant was thoroughly mixed with 2 mL chloroform. After centrifugation (15 min at 4000 x g, 4 °C), the upper, aqueous phase was transferred to a new tube, mixed with 1 mL chloroform and centrifuged again (15 min at 4000 x g, 4 °C). 2.4 mL of the upper, aqueous phase were transferred to new tubes, and RNA was precipitated with 240 µL 8 M LiCl and 6 mL 100 % ethanol by incubation at -80 °C for 15 min. The RNA was pelleted (30 min at 4000 x g, 4 °C), washed with 2 mL 70% ethanol, centrifuged again (5 min at 4000 x g, 4 °C), and resuspended in 500 µL nuclease-free water. mRNAs were isolated from this solution using 100 µL magnetic Oligo$(dT)_{25}$ beads (Thermo Fisher Scientific) according to the manufacturer's protocol. The mRNAs were eluted in 40 µL. The two samples per library were pooled and supplemented with DNase I buffer, 10 mM $MnCl_2$, 2 µL DNase I (Thermo Fisher Scientific), and 1 µL RNaseOUT (Thermo Fisher Scientific). After 1 h incubation at 37 °C, 2 µL 20 mg/mL glycogen (Thermo Fisher Scientific), 10 µL 8 M LiCl, and 250 µL 100% ethanol were added to the samples. Following precipitation at -80 °C, centrifugation (30 min at 20,000 x g, 4 °C), and washing with 200 µL 70% ethanol (5 min at 20,000 x g,

4 °C), the pellet was resuspended in 100 µL nuclease-free water. Eight reactions with 5 µL mRNA each and a GFP construct-specific primer (GAACTTGTGGCCGTTTACG) were prepared for cDNA synthesis using SuperScript IV reverse transcriptase (Thermo Fisher Scientific) according to the manufacturer's instructions. Half of the reactions were used as no reverse transcription control and the enzyme was replaced with water. After cDNA synthesis, the reactions were pooled and purified with DNA Clean & Concentrator-5 columns (Zymo Research). The barcode region was amplified with 10 – 20 cycles of PCR and read out by next generation sequencing on an Illumina NextSeq platform.

Computational methods

Binding site motifs for *N. benthamiana* transcription factors were obtained from the PlantTFDB,[17] and FIMO[18] was used to predict their occurrence in the 35S core enhancer. Fragments of pZS*11_4enh were aligned to the reference sequence using Bowtie2 (ref. [30]). For analysis of the STARR-seq experiments, the reads for each barcode were counted in the input and cDNA samples. Barcode counts below 5 and barcodes present in only one of three replicates were discarded. Barcode enrichment was calculated by dividing the barcode frequency (barcode counts divided by all counts) in the cDNA sample by that in the input sample. For the pZS*11_4enh fragment library (Fig. 2 and Supplementary Fig. 3) and the mutagenesis (Fig. 3a–d) experiments, the enrichment of the fragments or variants was calculated as the median enrichment of all barcodes linked to them. Boxplots were created using all corresponding barcodes from all replicates performed and were normalized to the median enrichment of constructs without an enhancer. The enrichment coverage of pZS*11_4enh was calculated by summing up the enrichment of all fragments containing a given nucleotide and dividing this sum by the number of fragments. Nucleotides covered by less than 5 fragments were excluded from analysis. Light dependency of enhancers or enhancer fragments was calculated as $\log_2$ of the enrichment in the light condition divided by the enrichment from the dark condition.

Acknowledgements

We thank Dr. Cristina M. Alexandre for initial plasmid construction. This work was supported by NSF RESEARCH-PGR 1748843 grant to C.Q. and S.F.

Author contributions

T.J., J.T.C., S.F, M.W.D. and C.Q. conceived and interpreted experiments, and wrote the manuscript. T.J. conducted experiments and data analysis.

Competing interests

The authors declare no competing interests.

References

1. Swinnen, G., Goossens, A. & Pauwels, L. Lessons from Domestication: Targeting Cis-Regulatory Elements for Crop Improvement. *Trends Plant Sci.* **21**, 506–515 (2016).
2. Scheben, A., Wolter, F., Batley, J., Puchta, H. & Edwards, D. Towards CRISPR/Cas crops – bringing together genomics and genome editing. *New Phytol.* **216**, 682–698 (2017).
3. Arnold, C. D. *et al.* Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science* **339**, 1074–1077 (2013).
4. Sun, J. *et al.* Global Quantitative Mapping of Enhancers in Rice by STARR-seq. *Genomics Proteomics Bioinformatics* **17**, 140–153 (2019).
5. Ricci, W. A. *et al.* Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants* **5**, 1237–1249 (2019).
6. Lehti-Shiu, M. D., Panchy, N., Wang, P., Uygun, S. & Shiu, S.-H. Diversity, expansion, and evolutionary novelty of plant DNA-binding transcription factor families. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1860**, 3–20 (2017).
7. Wilhelmsson, P. K. I., Mühlich, C., Ullrich, K. K. & Rensing, S. A. Comprehensive Genome-Wide Classification Reveals That Many Plant-Specific Transcription Factors Evolved in Streptophyte Algae. *Genome Biol. Evol.* **9**, 3384–3397 (2017).

8. Fang, R. X., Nagy, F., Sivasubramaniam, S. & Chua, N. H. Multiple cis regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. *Plant Cell* **1**, 141–150 (1989).

9. Benfey, P. N., Ren, L. & Chua, N. H. Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. *EMBO J.* **9**, 1677–1684 (1990).

10. Simpson, J., Schell, J., Montagu, M. V. & Herrera-Estrella, L. Light-inducible and tissue-specific pea lhcp gene expression involves an upstream element combining enhancer- and silencer-like properties. *Nature* **323**, 551–554 (1986).

11. Nagy, F., Boutry, M., Hsu, M. Y., Wong, M. & Chua, N. H. The 5′-proximal region of the wheat Cab-1 gene contains a 268-bp enhancer-like sequence for phytochrome response. *EMBO J.* **6**, 2537–2542 (1987).

12. Fluhr, R., Kuhlemeier, C., Nagy, F. & Chua, N.-H. Organ-Specific and Light-Induced Expression of Plant Genes. *Science* **232**, 1106–1112 (1986).

13. May, J. P., Yuan, X., Sawicki, E. & Simon, A. E. RNA virus evasion of nonsense-mediated decay. *PLOS Pathog.* **14**, e1007459 (2018).

14. Kwasnieski, J. C., Mogno, I., Myers, C. A., Corbo, J. C. & Cohen, B. A. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc. Natl. Acad. Sci.* **109**, 19498–19503 (2012).

15. Inoue, F., Kreimer, A., Ashuach, T., Ahituv, N. & Yosef, N. Identification and Massively Parallel Characterization of Regulatory Elements Driving Neural Induction. *Cell Stem Cell* **25**, 713-727.e10 (2019).

16. Ido, A. *et al.* Arabidopsis Pol II-Dependent in Vitro Transcription System Reveals Role of Chromatin for Light-Inducible rbcS Gene Transcription. *Plant Physiol.* **170**, 642–652 (2016).

17. Tian, F., Yang, D.-C., Meng, Y.-Q., Jin, J. & Gao, G. PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res.* **48**, D1104–D1113 (2020).

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

19. Lam, E., Benfey, P. N., Gilmartin, P. M., Fang, R. X. & Chua, N. H. Site-specific mutations alter in vitro factor binding and change promoter expression pattern in transgenic plants. *Proc. Natl. Acad. Sci.* **86**, 7890–7894 (1989).

20. Niggeweg, R., Thurow, C., Kegler, C. & Gatz, C. Tobacco Transcription Factor TGA2.2 Is the Main Component of as-1-binding Factor ASF-1 and Is Involved in Salicylic Acid- and Auxin-inducible Expression of as-1-containing Target Promoters. *J. Biol. Chem.* **275**, 19897–19905 (2000).

21. Lam, E. & Chua, N. H. ASF-2: a factor that binds to the cauliflower mosaic virus 35S promoter and a conserved GATA motif in Cab promoters. *Plant Cell* **1**, 1147–1156 (1989).

22. Wroblewski, T., Tomczak, A. & Michelmore, R. Optimization of Agrobacterium-mediated transient assays of gene expression in lettuce, tomato and Arabidopsis. *Plant Biotechnol. J.* **3**, 259–273 (2005).

23. Andrieu, A. *et al.* An in planta, Agrobacterium-mediated transient gene expression method for inducing gene silencing in rice (Oryza sativa L.) leaves. *Rice* **5**, 23 (2012).

24. Zheng, L., Liu, G., Meng, X., Li, Y. & Wang, Y. A Versatile Agrobacterium-Mediated Transient Gene Expression System for Herbaceous Plants and Trees. *Biochem. Genet.* **50**, 761–769 (2012).

25. Bond, D. M. *et al.* Infiltration-RNAseq: transcriptome profiling of Agrobacterium-mediated infiltration of transcription factors to discover gene function and expression networks in plants. *Plant Methods* **12**, 41 (2016).

26. Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S. & Mullineaux, P. M. pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Mol. Biol.* **42**, 819–832 (2000).

27. Kanoria, S. & Burma, P. K. A 28 nt long synthetic 5′UTR (synJ) as an enhancer of transgene expression in dicotyledonous plants. *BMC Biotechnol.* **12**, 85 (2012).

28. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE* **3**, e3647 (2008).

29. Subramaniam, A. R., Pan, T. & Cluzel, P. Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria. *Proc. Natl. Acad. Sci.* **110**, 2419–2424 (2013).

30. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

Figure legends

Fig. 1 | The positioning of enhancers has a pronounced impact on their activity. a, Scheme of the tobacco STARR-seq assay. All constructs are driven by the 35S minimal promoter (green). Enhancers (blue) are inserted in the indicated orientation and position. Barcodes (shades of purple) are inserted in the GFP open reading frame. BlpR, phosphinothricin resistance gene; pA site, poly-adenylation site. **b**, STARR-seq was performed with constructs harboring a single or double (2×) 35S enhancer in the indicated positions and orientations. Plots show $\log_2$(enrichment) of recovered RNA barcodes compared to DNA input. Each boxplot represents three independent replicates combined, as do subsequent ones. **c**, After introduction of a 2 kb spacer upstream of the minimal promoter, the 35S enhancer was inserted at the indicated distance upstream of the minimal promoter, and STARR-seq was performed. **d**, The 35S and three known plant enhancers were introduced in either the forward (fwd) or reverse (rev) orientation at the indicated positions and the STARR-seq assay was performed.

Fig. 2 | The tobacco STARR-seq assay identifies condition-specific enhancers. a, A plasmid harboring the indicated enhancers was fragmented. The fragments were inserted in the upstream position of the STARR-seq construct and their activity was measured by the STARR-seq assay. **b**, Plants were grown for two days in normal light/dark cycles (light, black line) or completely in the dark (dark, blue line) prior to mRNA extraction. The $\log_2$(enrichment) of RNA expression over input of all fragments at each position was averaged. **c**, Light dependency ($\log_2(\text{enrichment}^{\text{light}}/\text{enrichment}^{\text{dark}})$ was determined for each base of the original plasmid. **d**, The STARR-seq assay was performed with plasmid fragment libraries with different fragment length distributions (see inset), and $\log_2$(enrichment) for each fragment library is shown across the whole plasmid. **e**, $\log_2$(enrichment) obtained from the library with shorter fragments is shown in more detail for regions of interest. Positions in the original plasmid that contain enhancers are shaded in gray.

Fig. 3 | Saturation mutagenesis identifies functional elements in the 35S promoter and enhancer. **a**, All possible single nucleotide variants of the 35S minimal promoter were inserted into constructs with or without the 35S enhancer. The enrichment of the individual promoter insertions, substitutions and deletions was measured by the STARR-seq assay, normalized to the wild type variant, and plotted as a heatmap. Missing values are shown in gray and wild type variants are marked with gray dot. **b**, A sliding average (window size = 4 bp) of the positional mean enrichment scores for all substitutions, insertions and deletions was determined. The TATA box is highlighted in gray. **c**, The enrichment of all possible single nucleotide insertions, substitutions and deletions of the 35S core enhancer was determined as in **a**. **d**, A sliding average (window size = 8 bp) of the positional mean enrichment was determined via the STARR-seq assay. Predicted binding sites for the transcription factors from the indicated families are highlighted in gray. **e**, Three fragments (A, B, C; see **d**) of the 35S enhancer were inserted into the STARR-seq plasmid in random number and order, and assayed for their enhancer activity.

Supplementary Fig. 1 | The 35S enhancer destabilizes mRNA. a, Scheme of the constructs used in this figure. All constructs contain a 35S minimal promoter (green) controlling expression of a GFP reporter gene with or without an intron. In some constructs, an unstructured region of the Turnip crinkle virus 3′-UTR (TCV-USR) was inserted after the stop codon. The 35S core enhancer was inserted into the indicated positions. **b**, Constructs with the TCV-USR inserted after the stop codon were subjected to the STARR-seq assay. **c**, STARR-seq with constructs harboring an intron in the GFP open reading frame. The 35S enhancer was inserted upstream (up) of the promoter or into the intron (intron). **d**, STARR-seq using constructs with the AB80 enhancer upstream of the minimal promoter and/or the 35S enhancer in the 3′-UTR.

Supplementary Fig. 2 | STARR-seq can identify light-dependency of plant enhancers. a, STARR-seq was performed with plants that were grown for two days in normal light/dark cycles (+ light) or completely in the dark (- light) prior to mRNA extraction. The enrichment was determined for the indicated

enhancers. **b**, Light dependency ($\log_2(\text{enrichment}^{light}/\text{enrichment}^{dark})$) was determined for the indicated enhancers.

Supplementary Fig. 3 | Activity in the STARR-seq assay is insensitive to orientation and reproducible. a, The data from the STARR-seq assay (Fig. 2b, light condition) was analyzed separately for fragments inserted in the forward or reverse orientation. **b**, Correlation (Spearman's ρ) between two (out of three) replicates for individual barcodes. **c**, Correlation (Spearman's ρ) between two (out of three) replicates for fragments (median enrichment of all linked barcodes).

Supplementary Fig. 4 | Activity of the promoter variants is correlated with and without an enhancer in the construct. The enrichment scores for 35S minimal promoter variants in constructs with or without an upstream 35S enhancer was compared. The correlation (Spearman's ρ) is indicated.

Supplementary Fig. 5 | The order of 35S enhancer fragments has a subtle influence on enhancer activity. Three fragments (A, B, C) of the 35S enhancer (see Fig. 3d) were inserted into the STARR-seq plasmid in random number and order, and assayed for their enhancer activity.

Supplementary tables

Supplementary Table 1 | Enrichment scores for 35S minimal promoter variants.

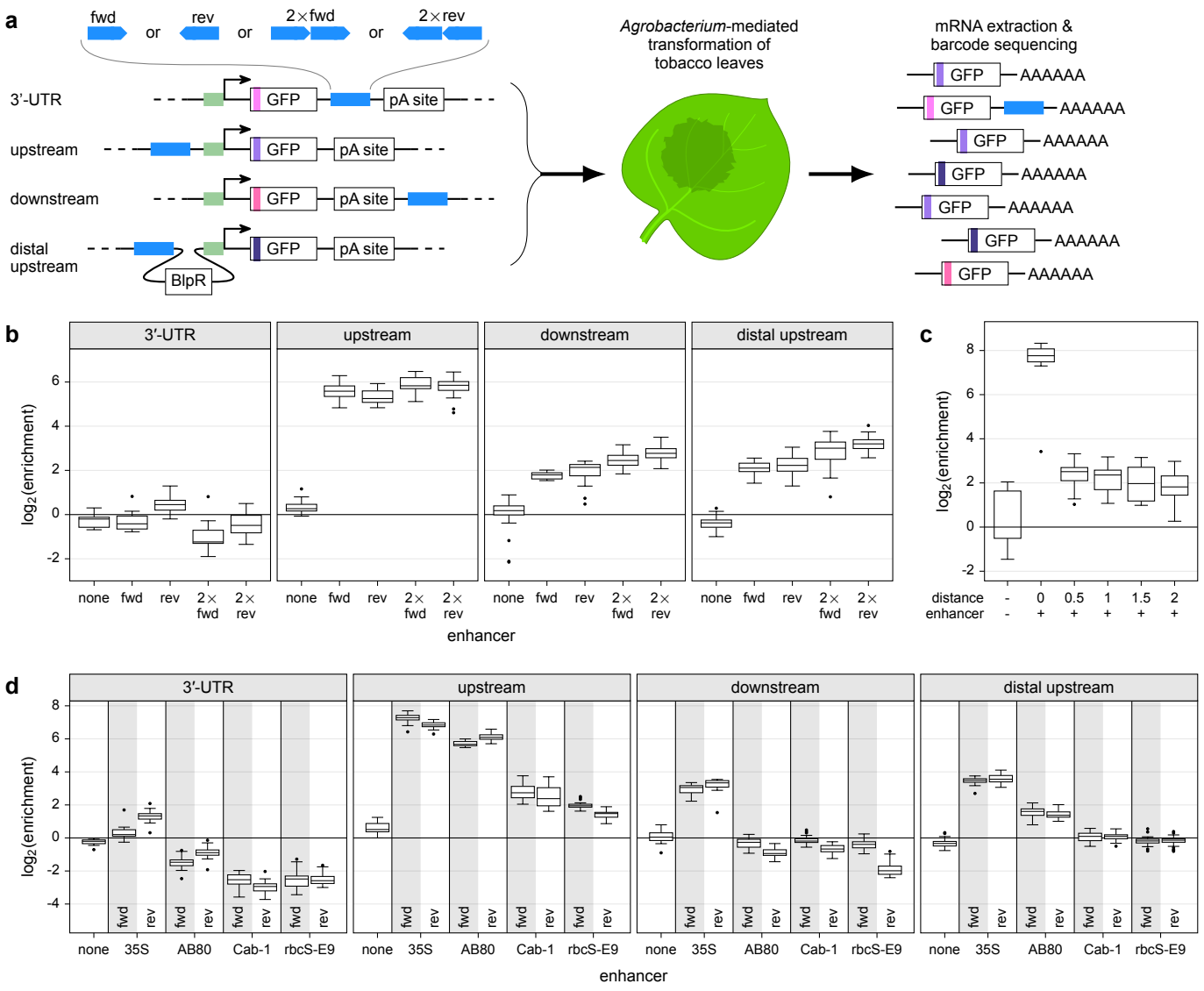
(in separate excel file)

Supplementary Table 2 | Enrichment scores for 35S core enhancer variants.

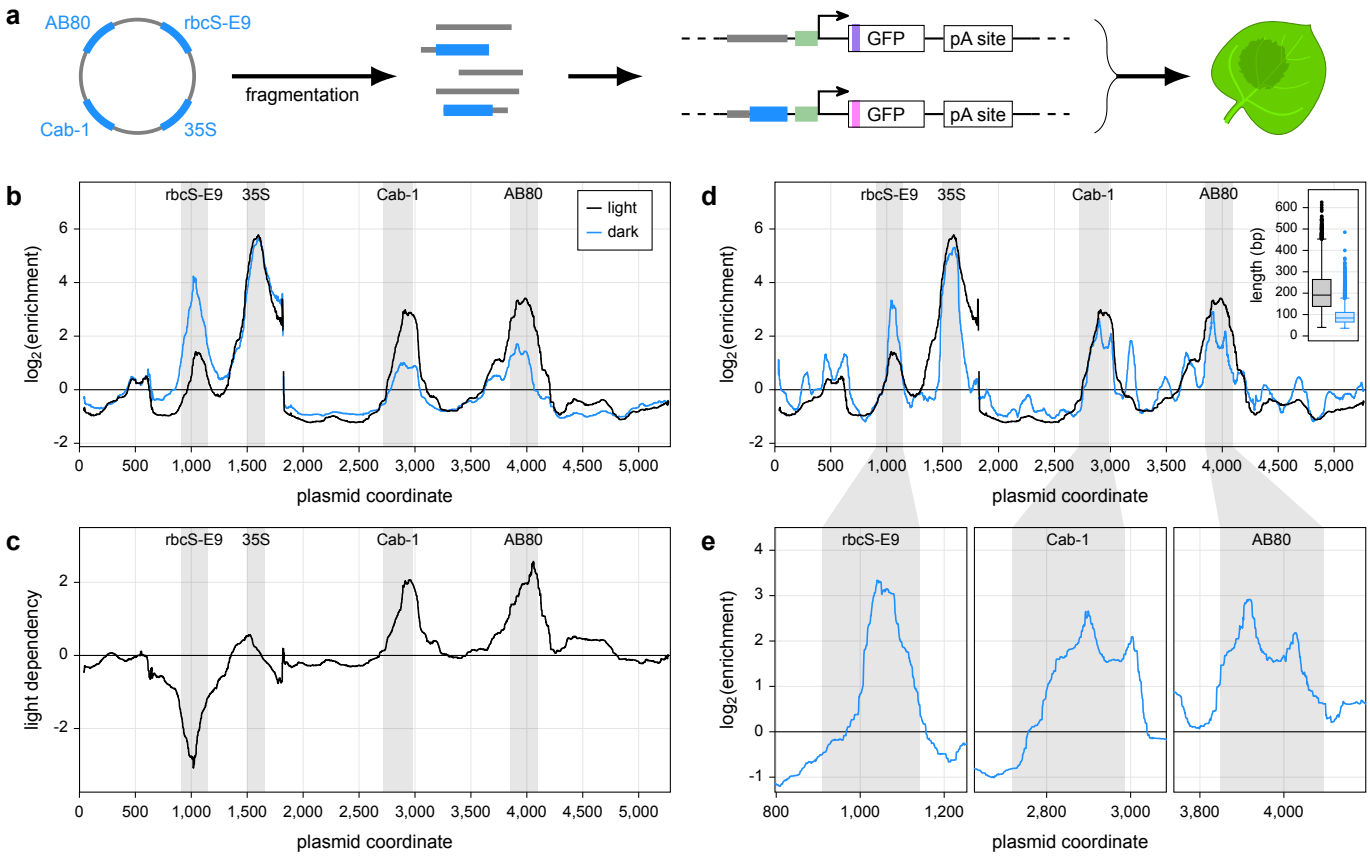
(in separate excel file)

Supplementary Table 3 | Enhancers sequences used in this study.

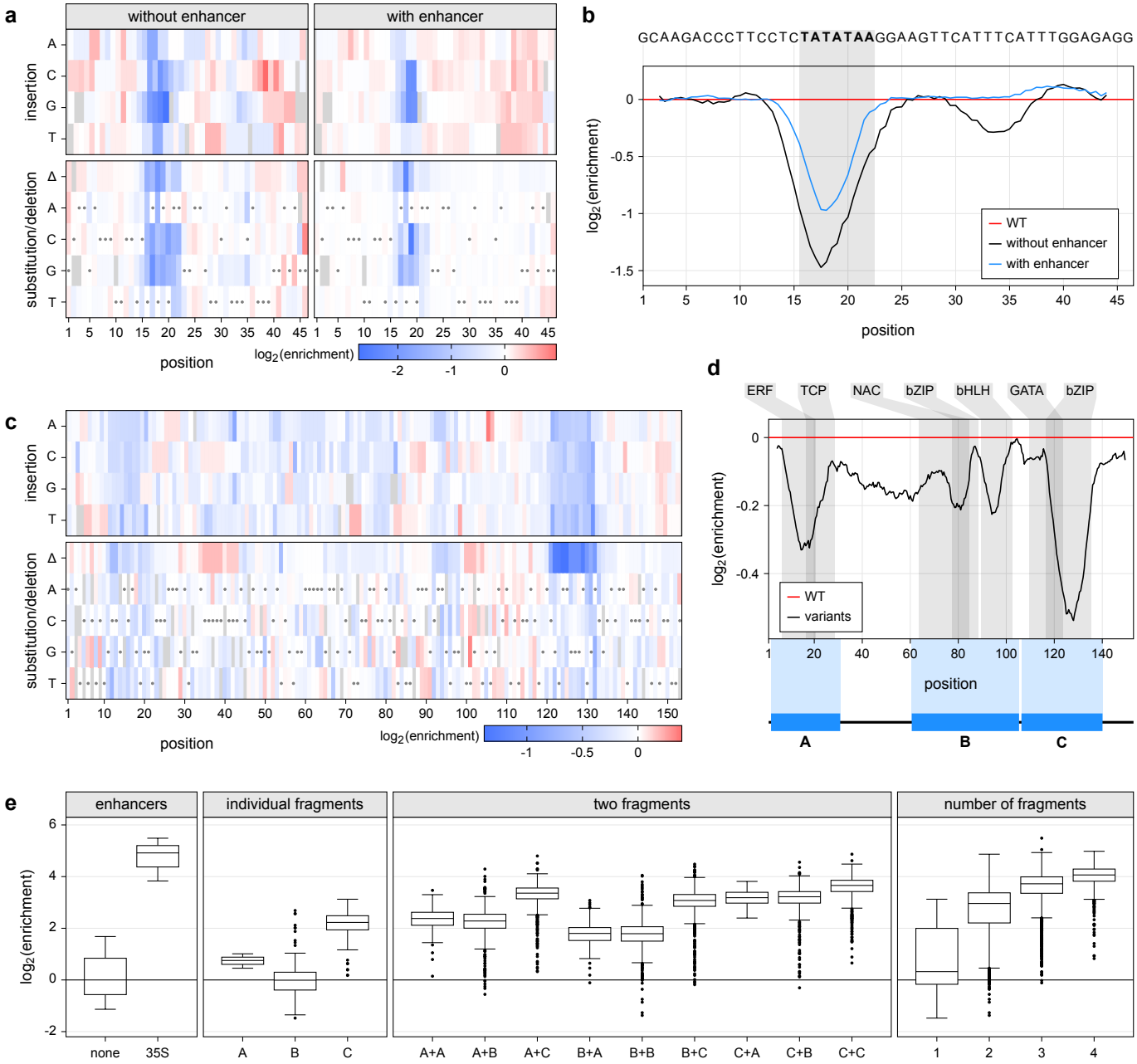
Enhancer	Species	Sequence	Ref.
35S	Cauliflower mosaic virus	AGATCTCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGACATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTC	8,9
AB80	*Pisum sativum*	CTGCCATAATGTCACAATTTTTCTCAAATCTTGTGGCTCTCAAACACTGTATAAAACACGACAAATGTGGACCCAAAATATATACCTTACACTTCTGAGTTAGAGAAGCAGAGCCCCATAATTAAGCCTATTTTATGAAAAAAATAATATTATGTTGAGTCATATATCCATAAGAATCCCCACAGTCACACATGGAAGAGCAGCATTGGATACAAATGATATGAAGATTTTGATCCATGAACGGATT	10
rbcS-E9	*Pisum sativum*	ACACTAACATCGAATGTACTTATCTCATTAGTTTAAATTATTGTTTGATCATGTTTAATCCTTATACTGTTGTTAGTTTTTTCAGTTAGCTTAGTGGGCATCTTACACGTGGCATTATTATCCTATTGGTGGCTAATGATAAGGTTAGCACACAAAACTTTTCAATCTTGTGTGGTTAATATGACTGCAAAGTTTATCATTTTCACAATCCAACAAACTGGTTCTAGGCAGTGG	12
Cab-1	*Triticum aestivum*	ATCGAATTTGTTGGGCAAGCGCCCAGTGTGATGTACGTTGGGAAAACTTGCAAGAGGATGCGACCAAATGAACTGGTAAACCATCCCGTGAGCGTGGCCTACACATTTTAAGCCAGCGGACTCTTTCGACTTGTCTTACAAAAGCTGGTCCAGTCACGAGCCTTAGCCCTAACCATAGCCACAAGTACATCTCATCCATTTAAGGCCTCTGCGTGCACCAATGGCATCCAAGCTGCAGATTTCTTTTCACCACCGTCTCTCTTGTCAG	11



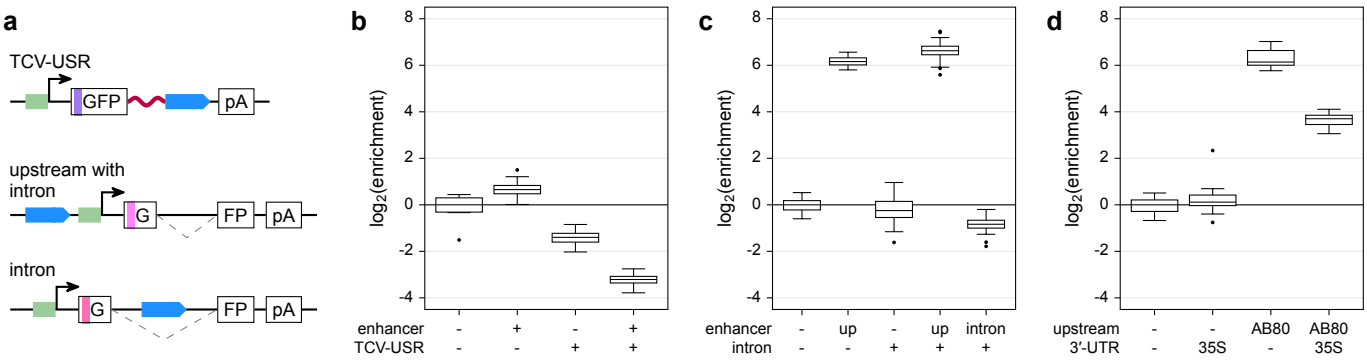
Jores *et al.*, Fig. 1



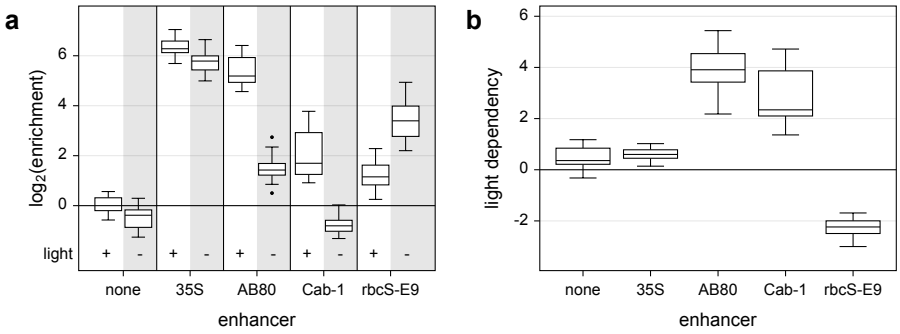
Jores *et al.*, Fig. 2



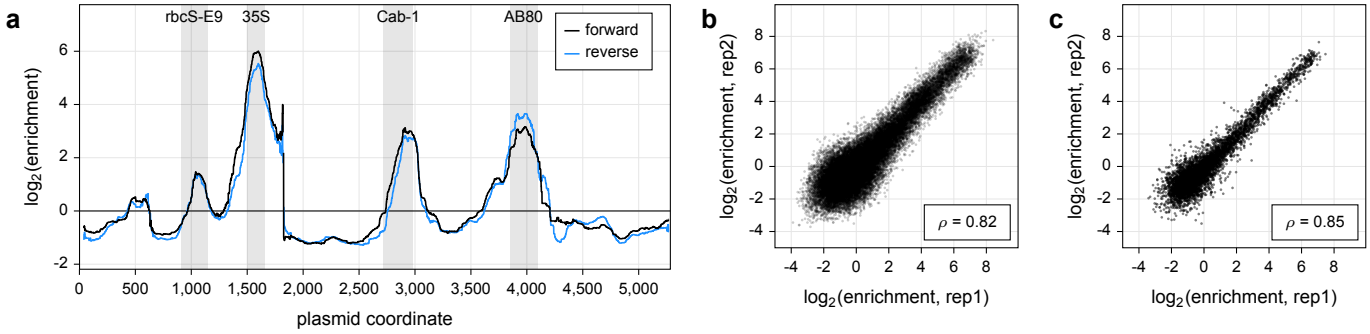
Jores *et al.*, Fig. 3



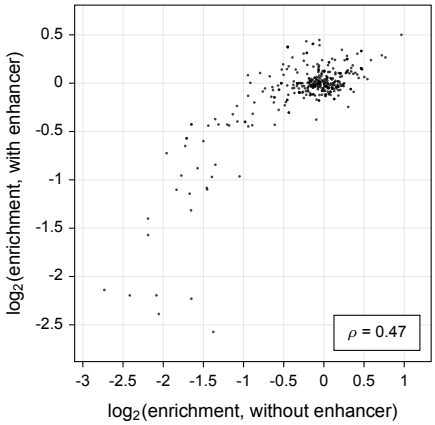
Jores *et al.*, Supplementary Fig. 1



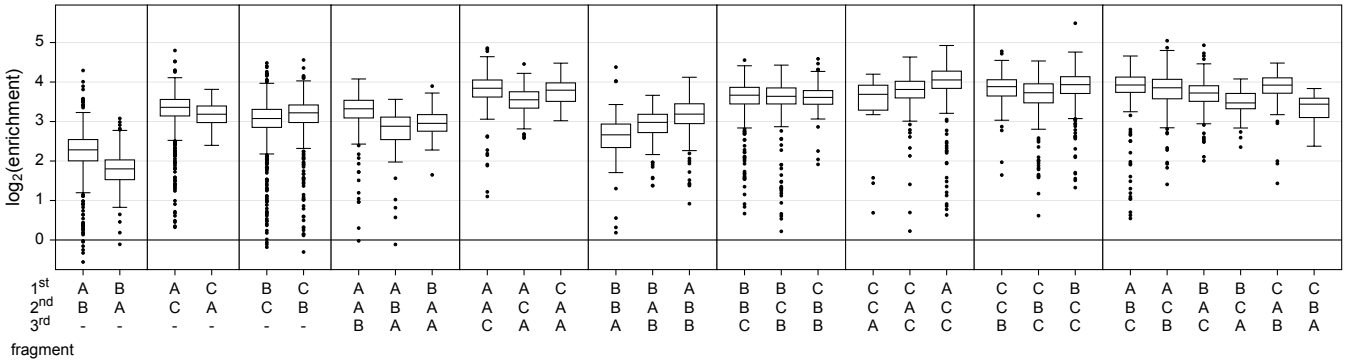
Jores *et al.*, Supplementary Fig. 2



Jores *et al.*, Supplementary Fig. 3



Jores *et al.*, Supplementary Fig. 4


log2(enrichment)
5
4
3
2
1
0
1st A B A C A C B A A B A A A C A B B A B B B C B C C C A C A C C C B C C B C A B C A B C A B C A B
2nd B A C A C B A B A A C A A B B A B B C B B A C C C B C C B C A B C A B C A B
3rd - - - - - - B A A C A A B A B C B B C C C