nature chemical biology

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41589-020-0547-4

In the format provided by the authors and unedited.

Design of orthogonal regulatory systems for modulating gene expression in plants

Michael S. Belcher[1,2,5], Khanh M. Vuu[1,3,5], Andy Zhou[1,3,4], Nasim Mansoori[1,3], Amanda Agosto Ramos[1,3], Mitchell G. Thompson[1,3,4], Henrik V. Scheller[1,2,3], Dominique Loqué[1,3] and Patrick M. Shih[1,3,4] ✉

[1]Feedstocks Division, Joint BioEnergy Institute, Emeryville, CA, USA. [2]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, USA. [3]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [4]Department of Plant Biology, University of California, Davis, Davis, CA, USA. [5]These authors contributed equally: Michael S. Belcher, Khanh M. Vuu.
✉e-mail: pmshih@ucdavis.edu

Supplementary Information

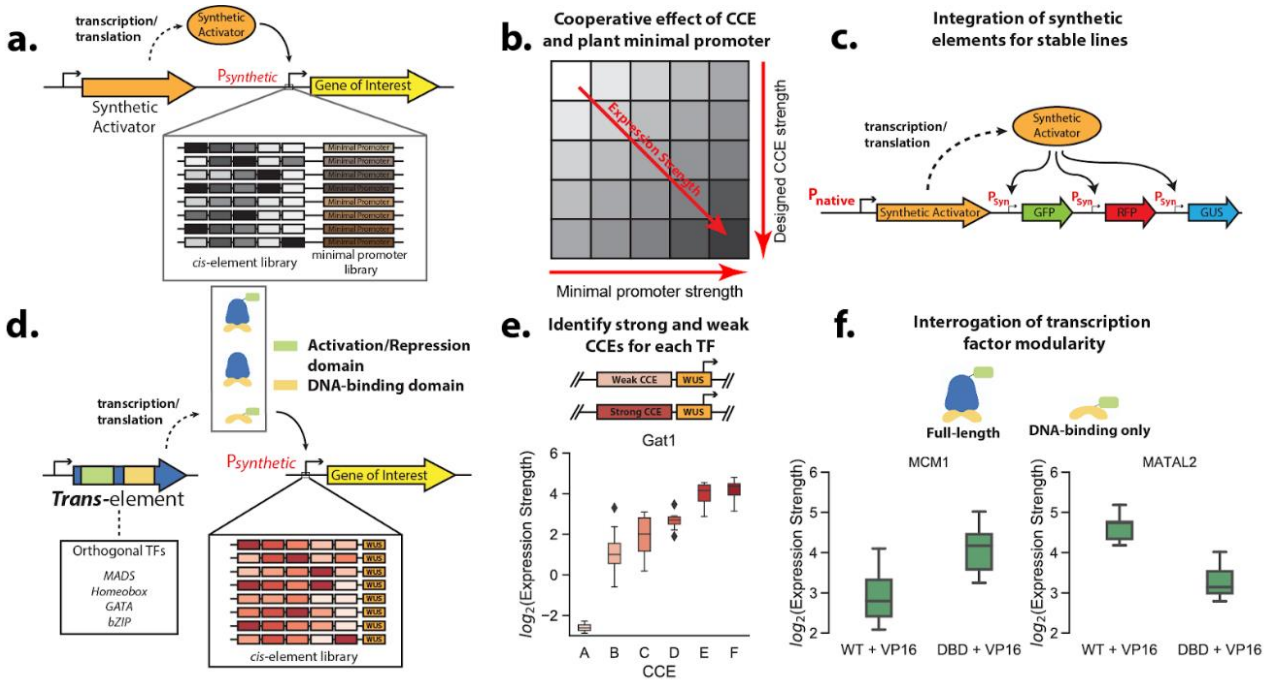
Design of orthogonal regulatory systems for modulating gene expression in plants

Authors: Michael S. Belcher[^,1,2], Khanh M. Vuu[^,1,3], Andy Zhou[1,3], Nasim Mansoori[1,3], Amanda Agosto Ramos[1,3], Mitchell G. Thompson[1,3,4], Henrik V. Scheller[1,2,3], Dominique Loqué[1,3], Patrick M. Shih[1,3,4]*

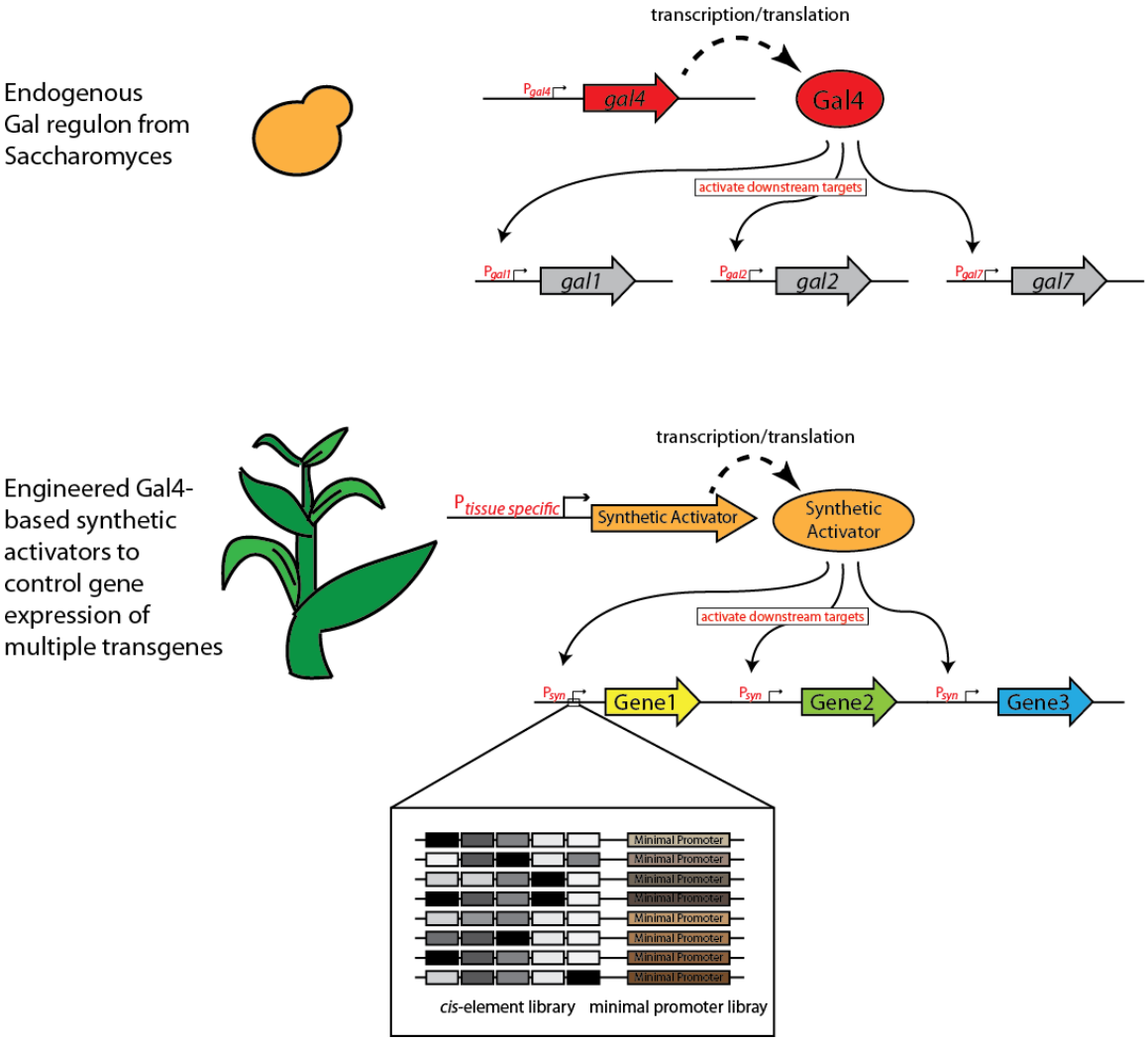
Affiliations:
[^] Authors contributed equally to this work.
[1] Feedstocks Division, Joint BioEnergy Institute, Emeryville, California, USA.
[2] Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, California, USA.
[3] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA.
[4] Department of Plant Biology, University of California, Davis, Davis, California, USA.

* Correspondence to pmshih@ucdavis.edu

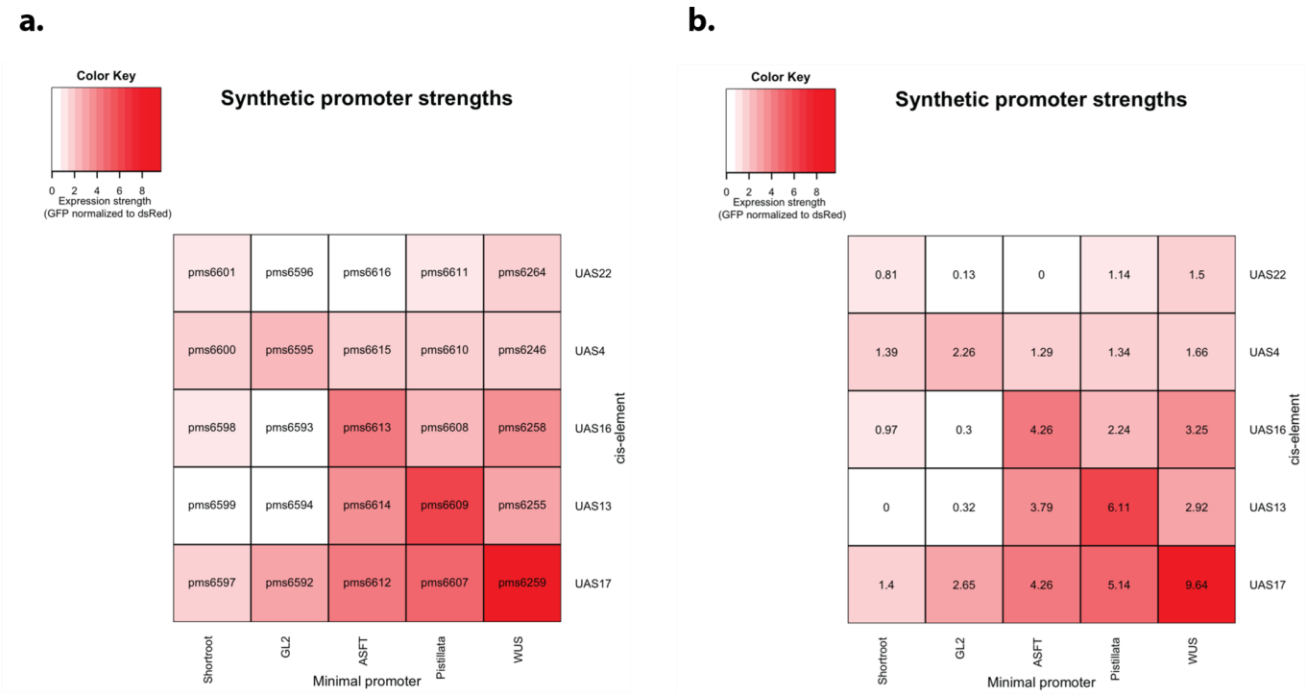
Supplementary Information - Table of Contents



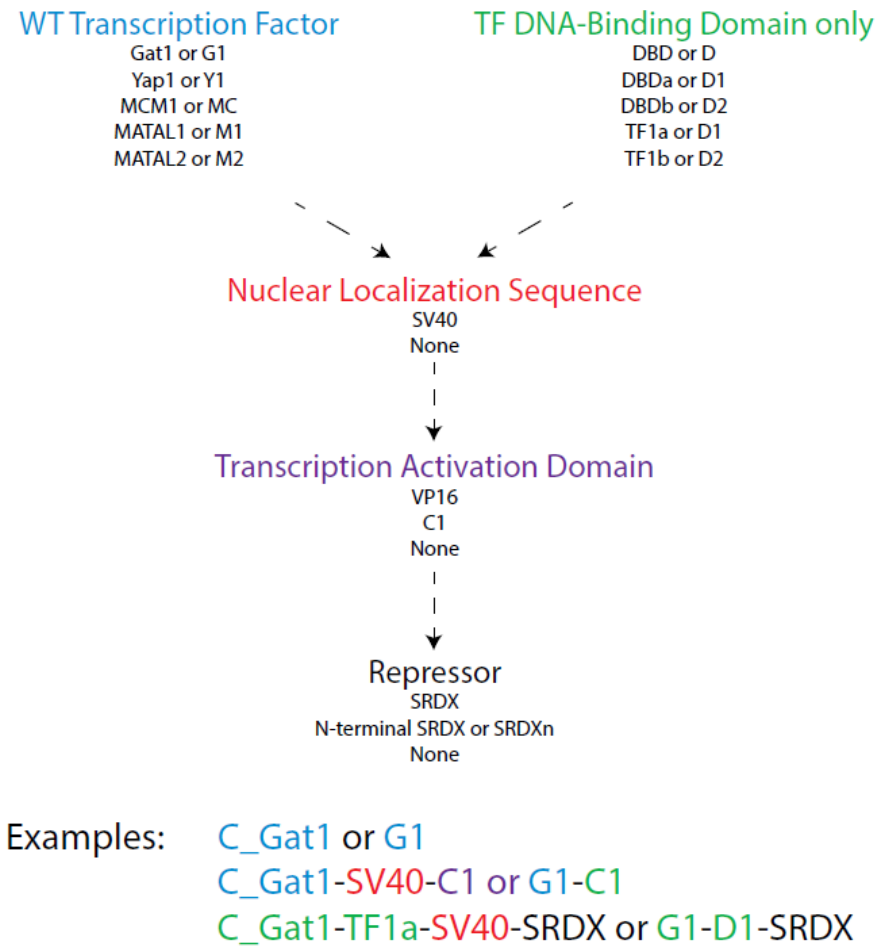
Supplementary Figure 1: Visual schematic of design principles and experimental workflow. a, A randomized library of CCEs and minimal promoters was used to generate a suite of synthetic promoters for the Gal4 expression system to identify contributions of each element to the strength of a synthetic promoter. **b,** A combinatorial library of five select CCEs and minimal promoters was generated for Gal4. We found that CCEs and minimal promoters have an additive contribution to gene expression. **c,** A single endogenous promoter, denoted as P_{Native}, was used to drive the expression of an orthogonal synthetic activator in a tissue specific or environmentally responsive manner. This synthetic activator then induced expression of three reporter genes to verify reporter presence corresponds with the tissue or condition tested. **d,** The effect of *trans*-element and CCE diversity on gene expression was examined using numerous TF families from *S. cerevisiae*. Increasing the diversity of TF families tested allows for an investigation into how our modifications behave in varying systems and expands our parts library drastically. **e,** Testing a single *trans*-activator with the corresponding suite of synthetic promoters verifies that the CCE has a direct effect on gene expression strength for the other TF families. This variation provides the capacity to control the relative expression strengths of multiple genes when coupled with a single *trans*-element and increases the sequence diversity of our parts. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5^{th} and 95^{th} percentile with values outside considered outliers (n = 8 independent biological replicates). **f,** The modularity of our *trans*-element design was investigated by fusing activation domains to the minimal DNA binding scaffold of each TF. We found that the fusion of an activation domain to the truncated TF, denoted as DBD, had varying effects on transcriptional output dependent on the family used for the basis of the design. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5^{th} and 95^{th} percentile with values outside considered outliers (n = 8 independent biological replicates).



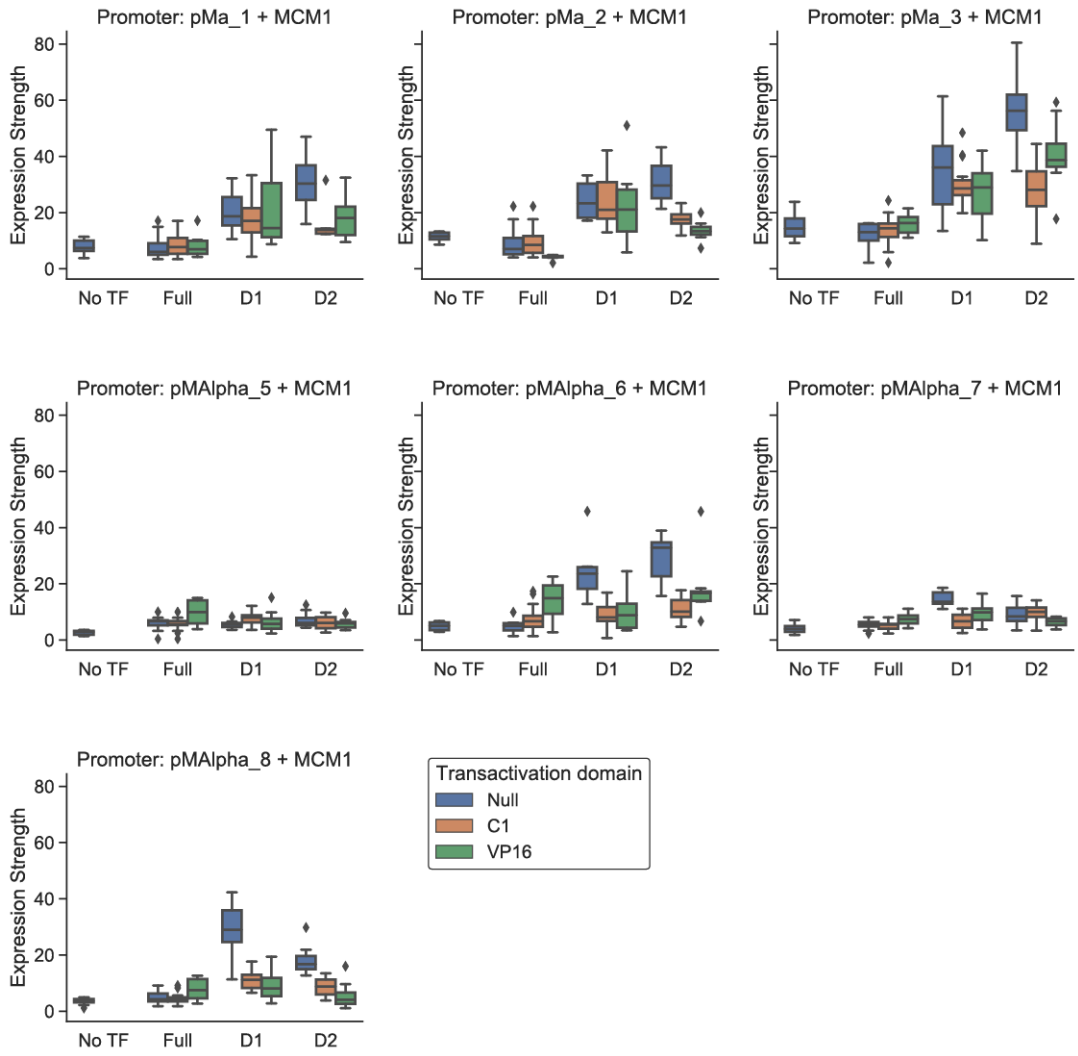
Supplementary Figure 2. Building a heterologous synthetic activator-promoter system using the GAL4 network for plant expression. Top, rationale of utilizing endogenous *cis*-elements from Gal promoters in the Saccharomyces Gal regulon. Various UAS/Gal *cis*-elements identified from varying Saccharomyces Gal promoter regions. **Bottom,** various cis-elements (UAS) are used to build a randomized cis-element library based on concatenation of five variants. These are fused to a library of plant minimal promoters, building a library of synthetic promoters that replicate the endogenous yeast Gal regulon. These promoters are then characterized and tested in plants as an orthogonal synthetic transcription system for controlling gene expression strength and tissue-specificity of stacked transgenes.



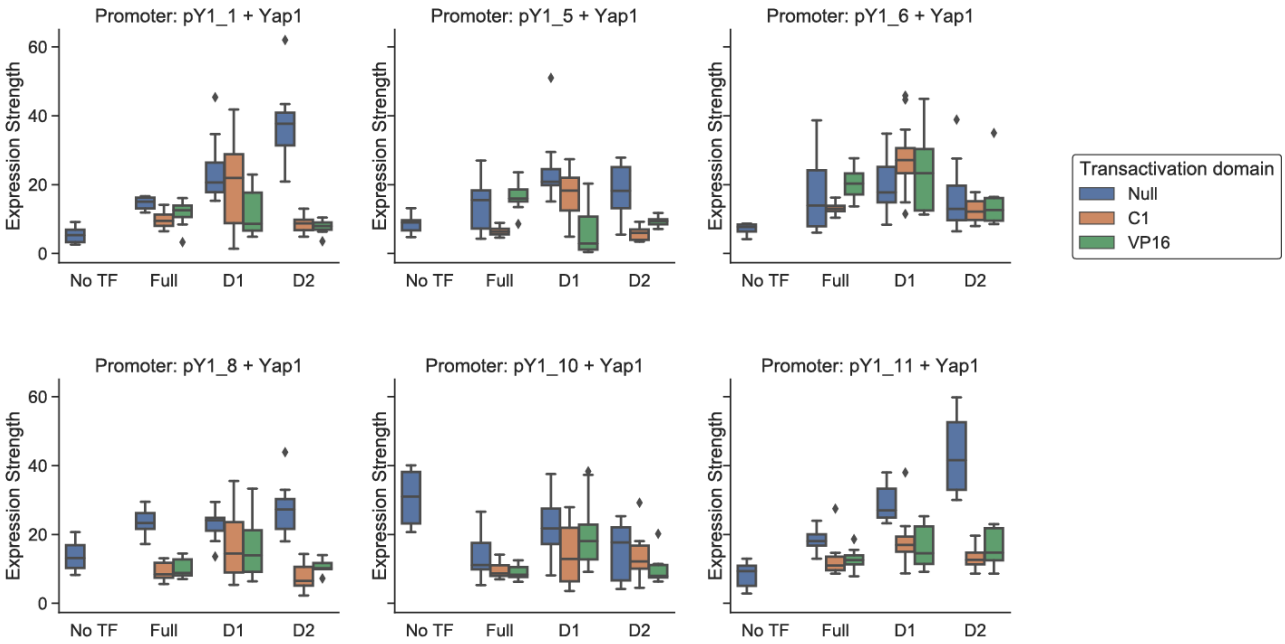
Supplementary Figure 3. Synthetic promoters designed to investigate the rational design and contributory effects of combinations of CCE and minimal promoter. a, Name of all constructs tested corresponding to the CCE and minimal promoters fused together to build each synthetic promoter. Heatmap corresponds to Figure 3b. **b,** Corresponding expression strength of all characterized promoters. Values represent the average normalized expression strength for each combination tested ($n \geq 8$ biological replicates).



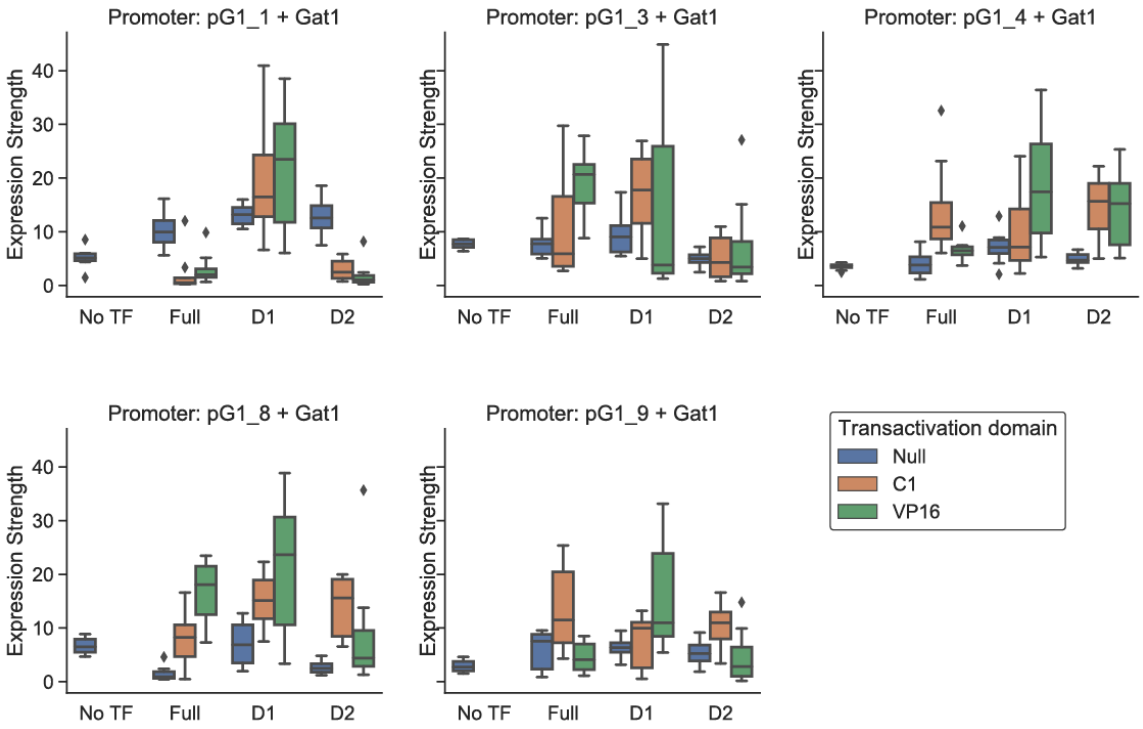
Supplementary Figure 4. Design and nomenclature of synthetic transcription factors *trans*-element library. Synthetic trans-element design and nomenclature follow the depicted pipeline. Transcription factors are initially selected as either full-length WT or only DNA-binding domain and subsequently fused with a selection of NLS, TAD, or repressor domains. Expression strengths for constructs tested with compatible synthetic promoters are listed in Table S2 and sequences are listed in Table S1.



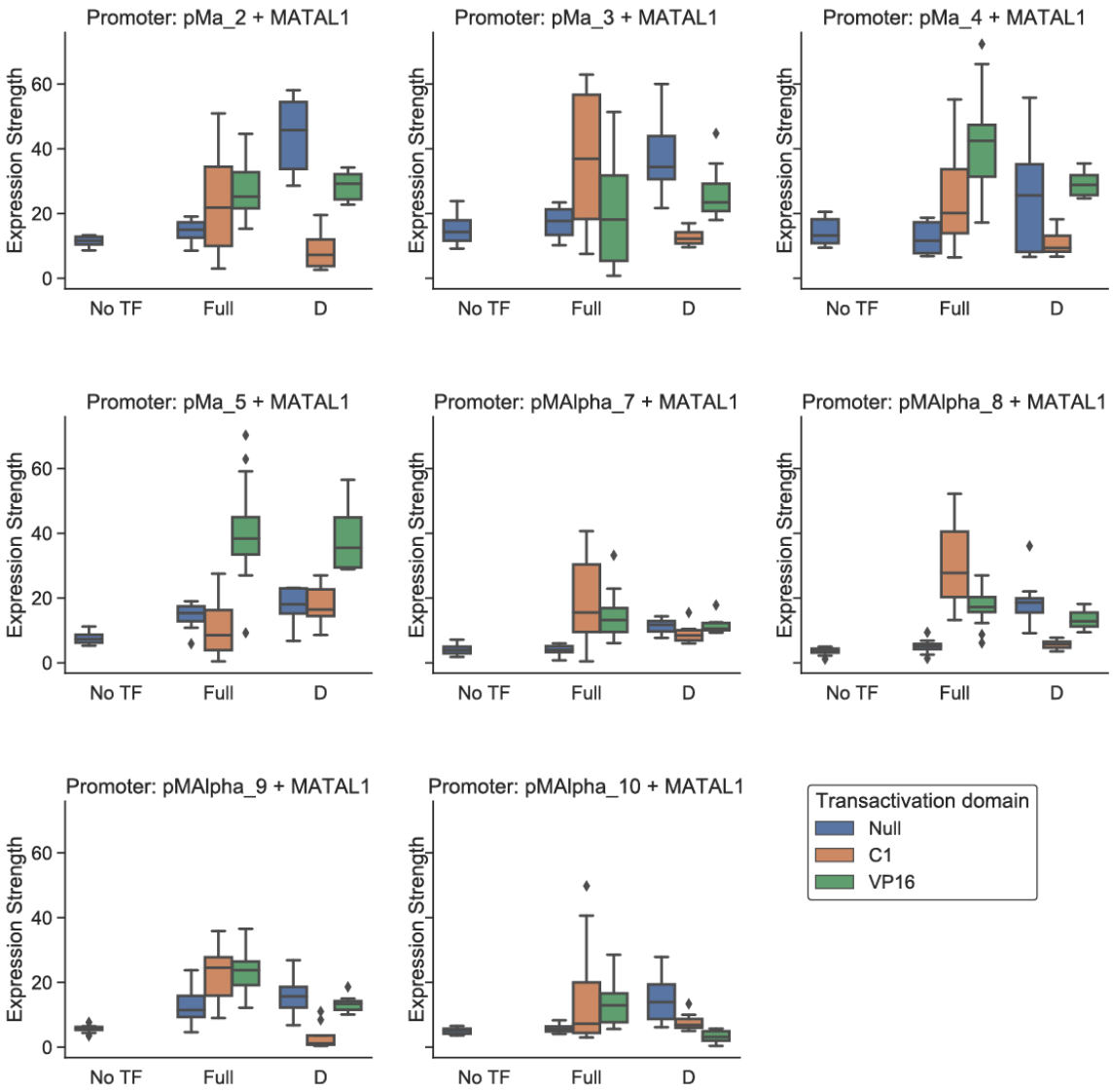
Supplementary Figure 5. Characterization of *trans*-elements for MCM1. Expression strength is determined via measurement of GFP fluorescence/RFP fluorescence. Each graph represents a single promoter's background expression with all variations of *trans*-element from the corresponding library. This includes the full-length TF both with and without each TAD, as well as each predicted DNA-binding domain (D1 or D2) both with and without each TAD. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5th and 95th percentile with values outside considered outliers (n = 8 independent biological replicates).



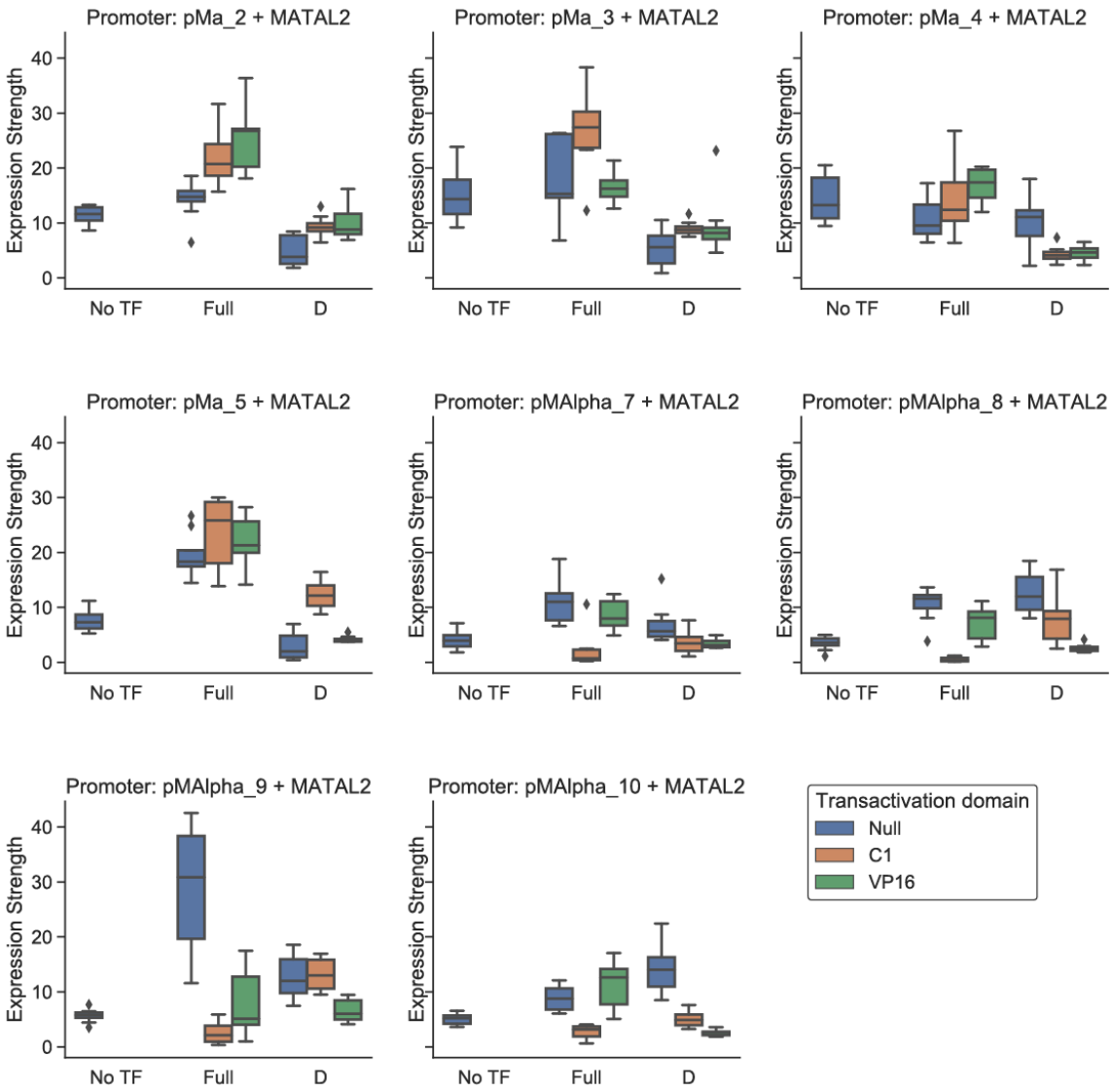
Supplementary Figure 6. Characterization of *trans*-elements for Yap1. Expression strength is determined via measurement of GFP fluorescence/RFP fluorescence. Each graph represents a single promoter's background expression with all variations of *trans*-element from the corresponding library. This includes the full-length TF both with and without each TAD, as well as each predicted DNA-binding domain (D1 or D2) both with and without each TAD. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5th and 95th percentile with values outside considered outliers (n = 8 independent biological replicates).



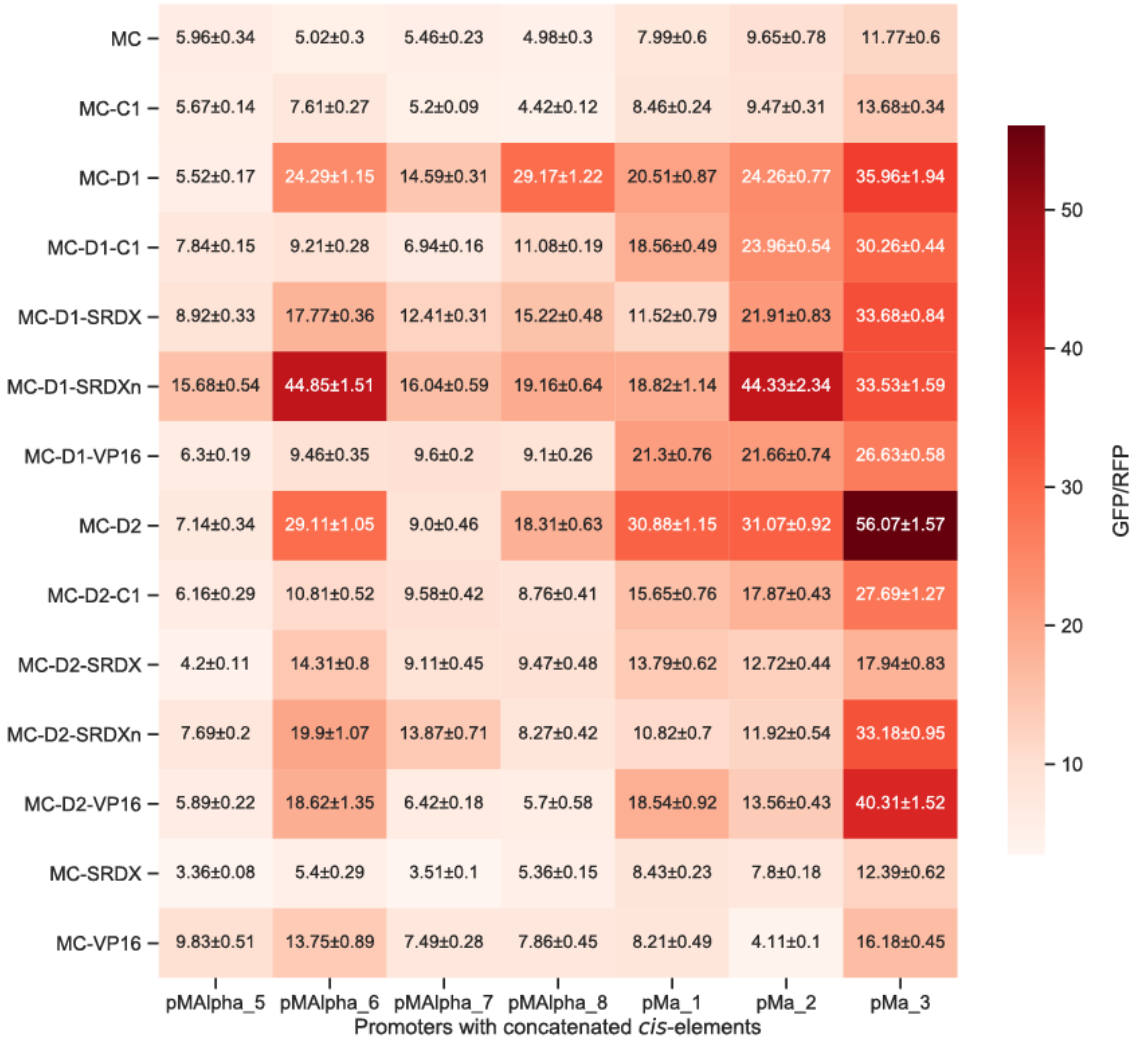
Supplementary Figure 7. Characterization of *trans*-elements for Gat1. Expression strength is determined via measurement of GFP fluorescence/RFP fluorescence. Each graph represents a single promoter's background expression with all variations of *trans*-element from the corresponding library. This includes the full-length TF both with and without each TAD, as well as each predicted DNA-binding domain (D1 or D2) both with and without each TAD. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5th and 95th percentile with values outside considered outliers (n = 8 independent biological replicates).



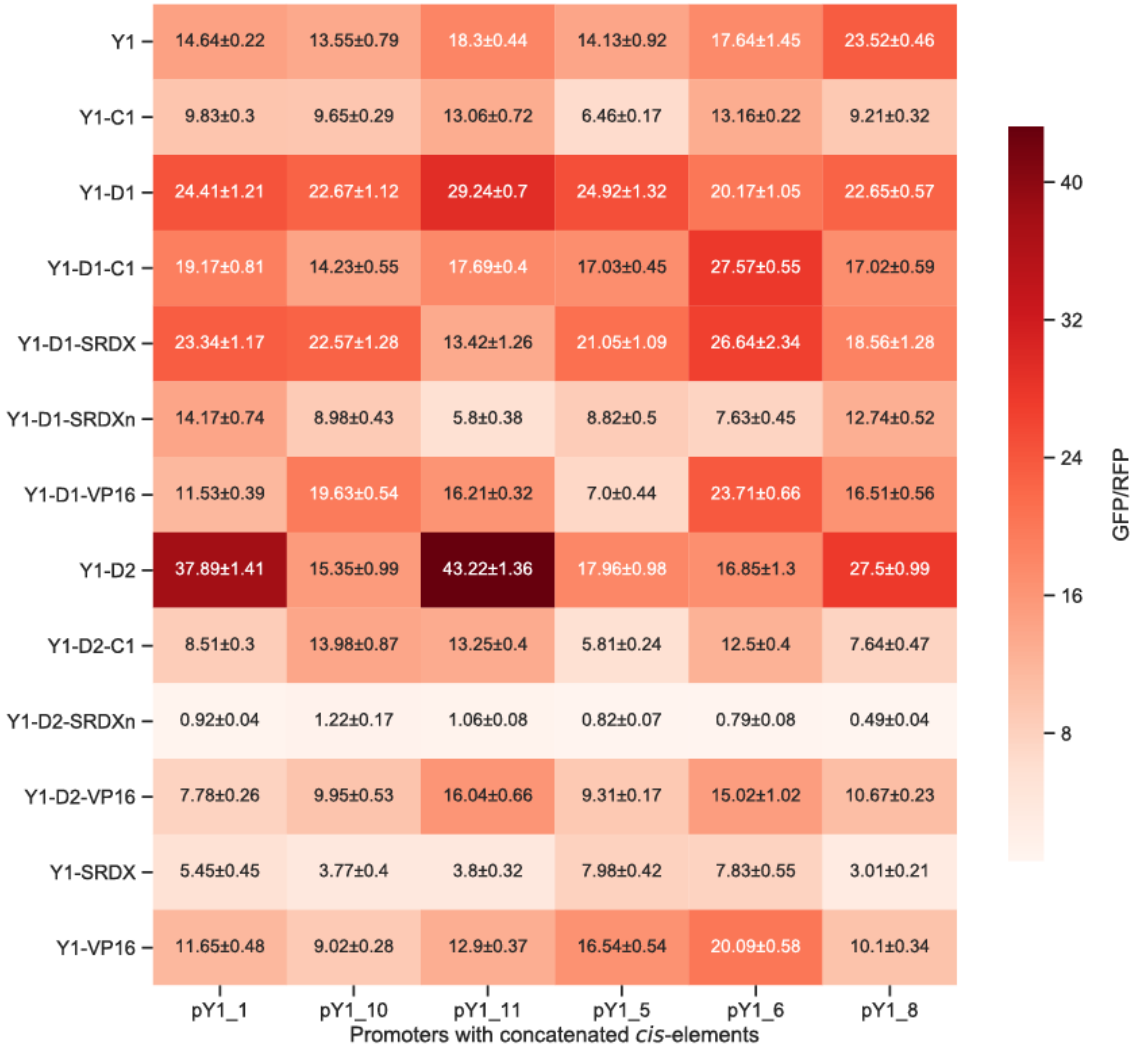
Supplementary Figure 8. Characterization of *trans*-elements for Mata1. Expression strength is determined via measurement of GFP fluorescence/RFP fluorescence. Each graph represents a single promoter's background expression with all variations of *trans*-element from the corresponding library. This includes the full-length TF both with and without each TAD, as well as the predicted DNA-binding domain (D) both with and without each TAD. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5th and 95th percentile with values outside considered outliers (n = 8 independent biological replicates).



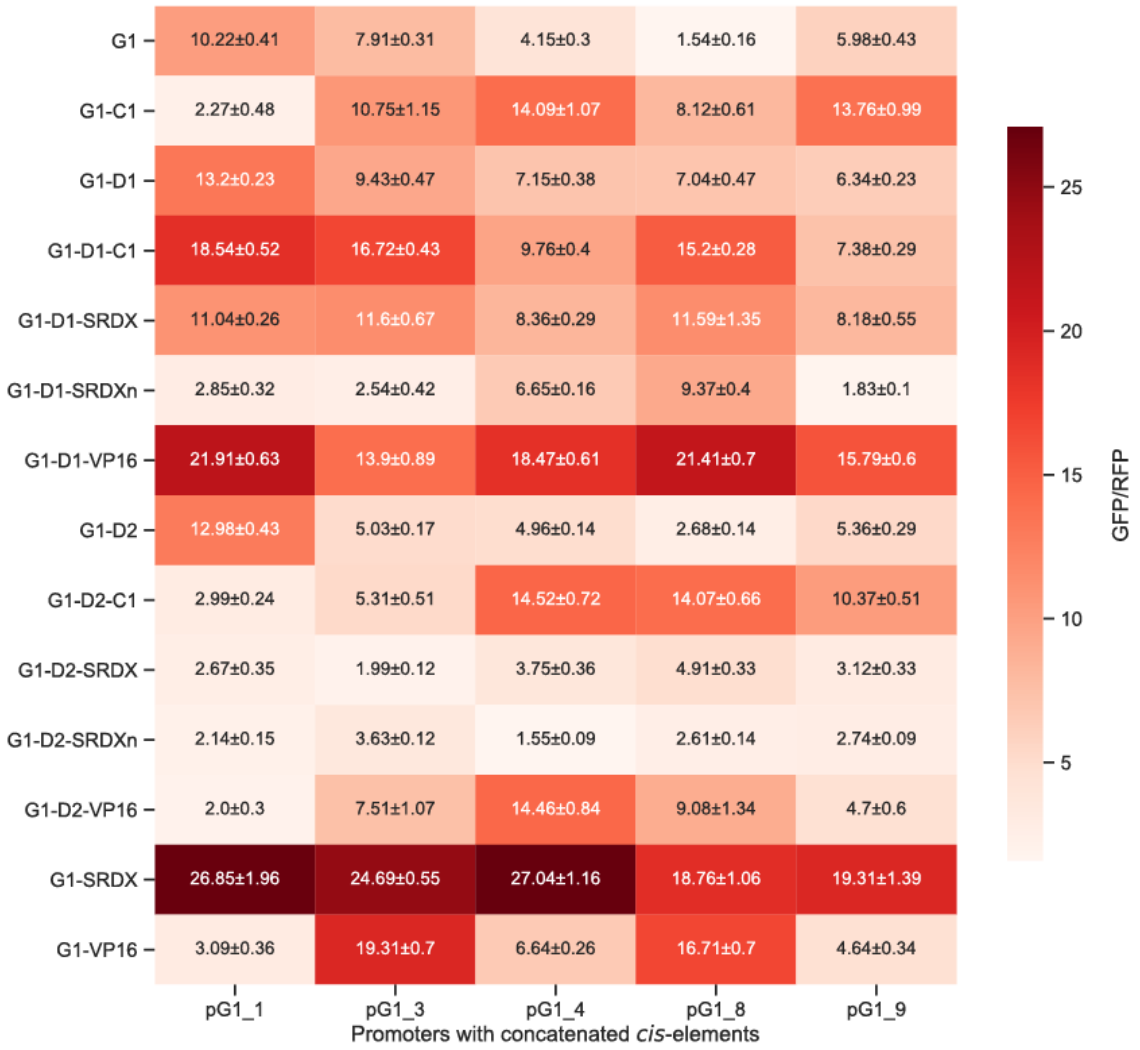
Supplementary Figure 9. Characterization of *trans*-elements for Matα2. Expression strength is determined via measurement of GFP fluorescence/RFP fluorescence. Each graph represents a single promoter's background expression with all variations of *trans*-element from the corresponding library. This includes the full-length TF both with and without each TAD, as well as the predicted DNA-binding domain (D) both with and without each TAD. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5th and 95th percentile with values outside considered outliers (n = 8 independent biological replicates).



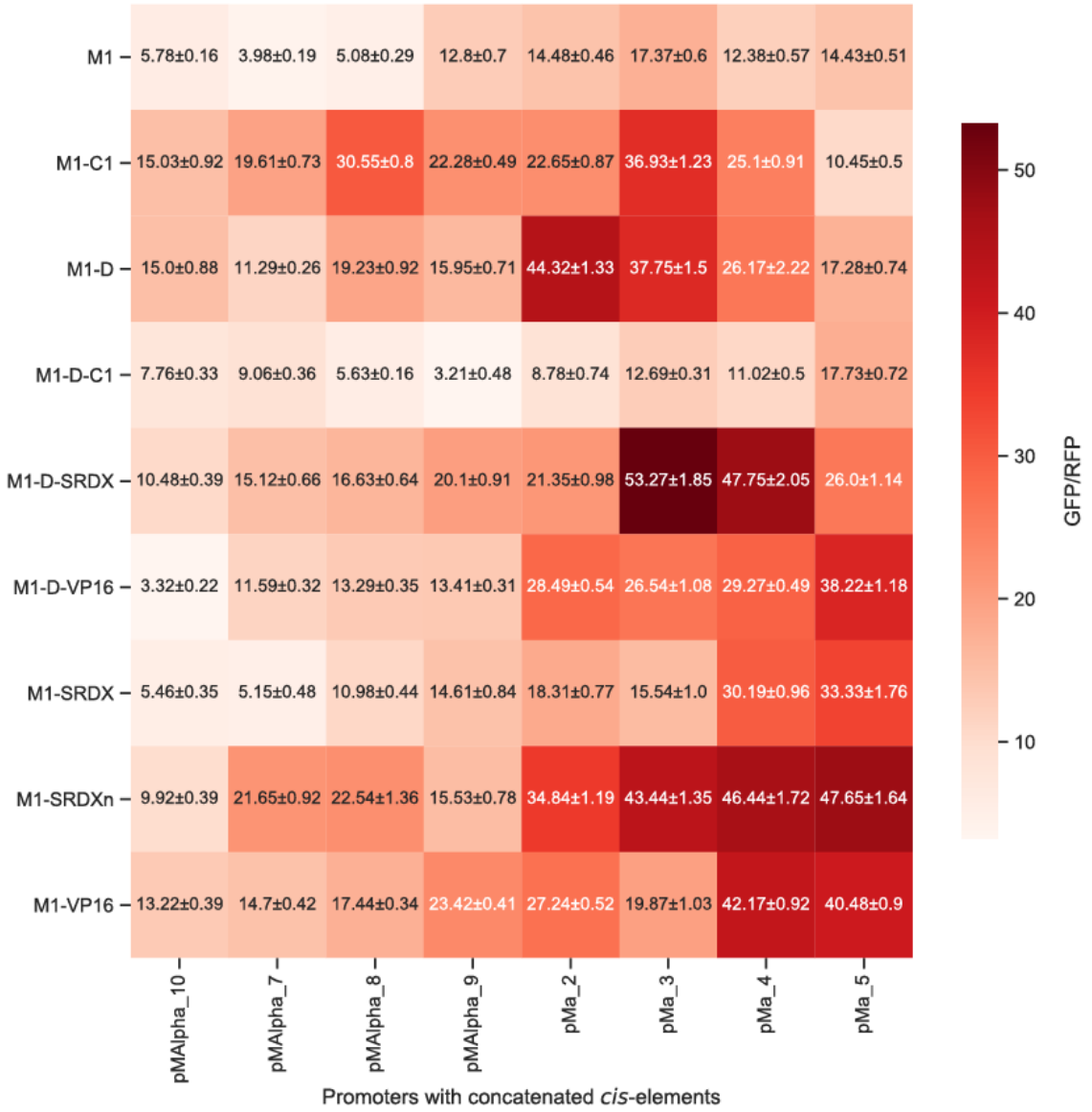
Supplementary Figure 10. Expression strengths of synthetic promoter and transcription factor pairs for the MCM1 transcription factor system. Combinations of MCM11 *trans*-elements and corresponding synthetic promoters (*trans*-elements are denoted on the y-axis and synthetic promoters are denoted on the x-axis). Expression strength is calculated by dividing GFP fluorescence from a control RFP fluorescence. All values shown are the average of 8 independent biological replicates including standard error.



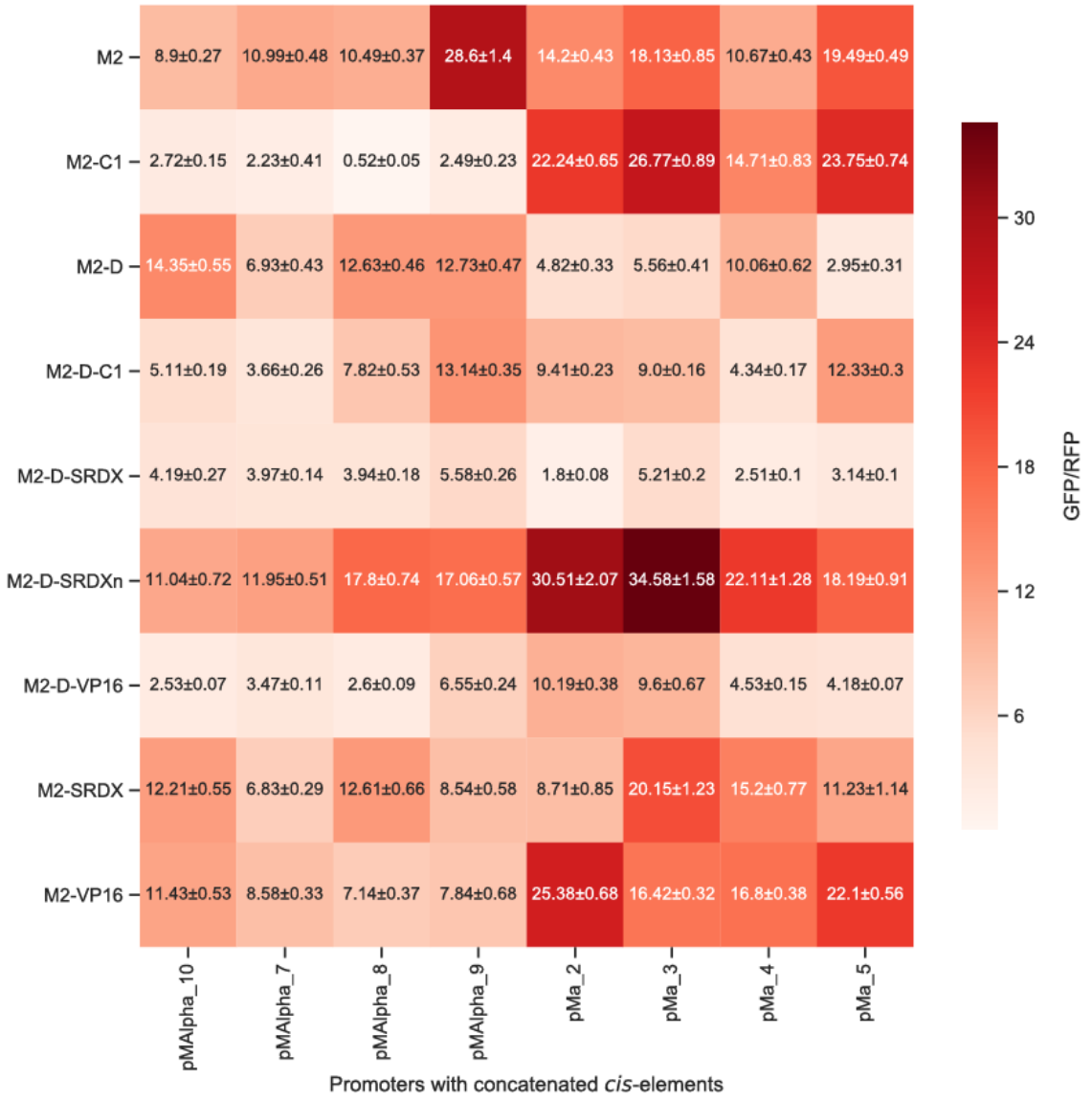
Supplementary Figure 11. Expression strengths of synthetic promoter and transcription factor pairs for the Yap1 transcription factor system. Combinations of Yap1 *trans*-elements and corresponding synthetic promoters (*trans*-elements are denoted on the y-axis and synthetic promoters are denoted on the x-axis). Expression strength is calculated by dividing GFP fluorescence from a control RFP fluorescence. All values shown are the average of 8 independent biological replicates including standard error.



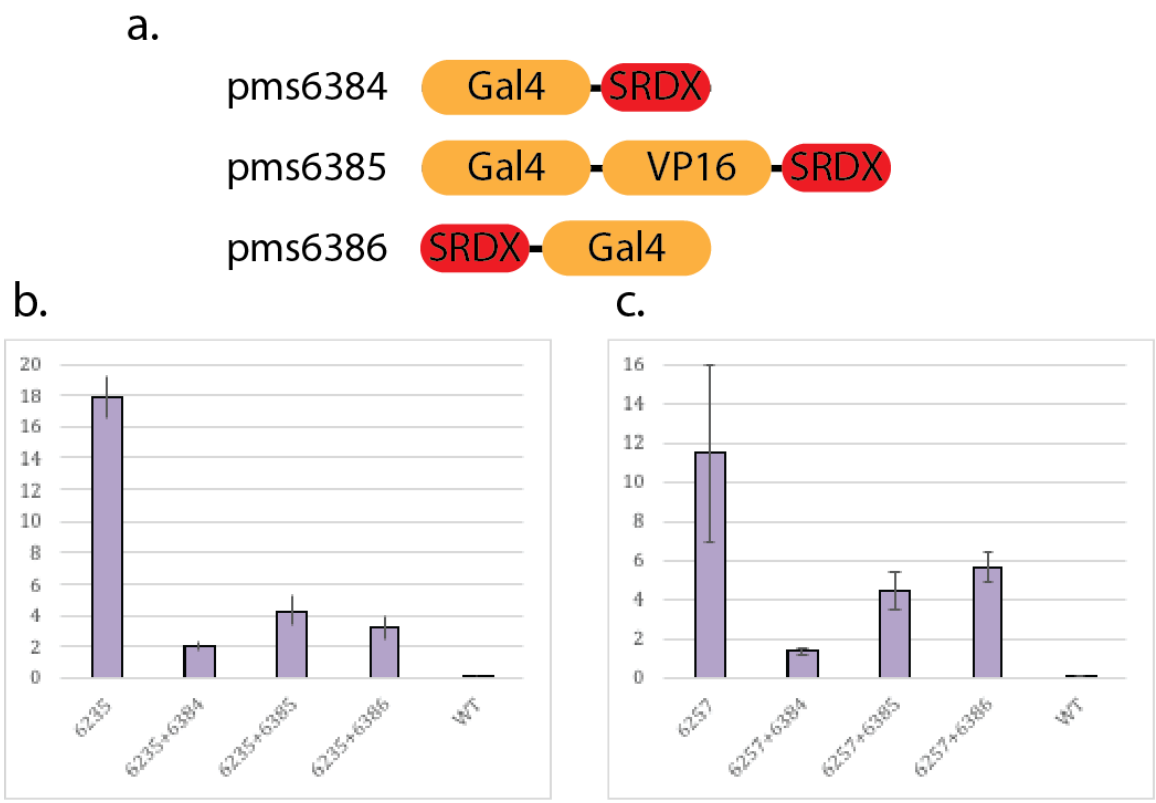
Supplementary Figure 12. Expression strengths of synthetic promoter and transcription factor pairs for the Gat1 transcription factor system. Combinations of Gat1 *trans*-elements and corresponding synthetic promoters (*trans*-elements are denoted on the y-axis and synthetic promoters are denoted on the x-axis). Expression strength is calculated by dividing GFP fluorescence from a control RFP fluorescence. All values shown are the average of 8 independent biological replicates including standard error.



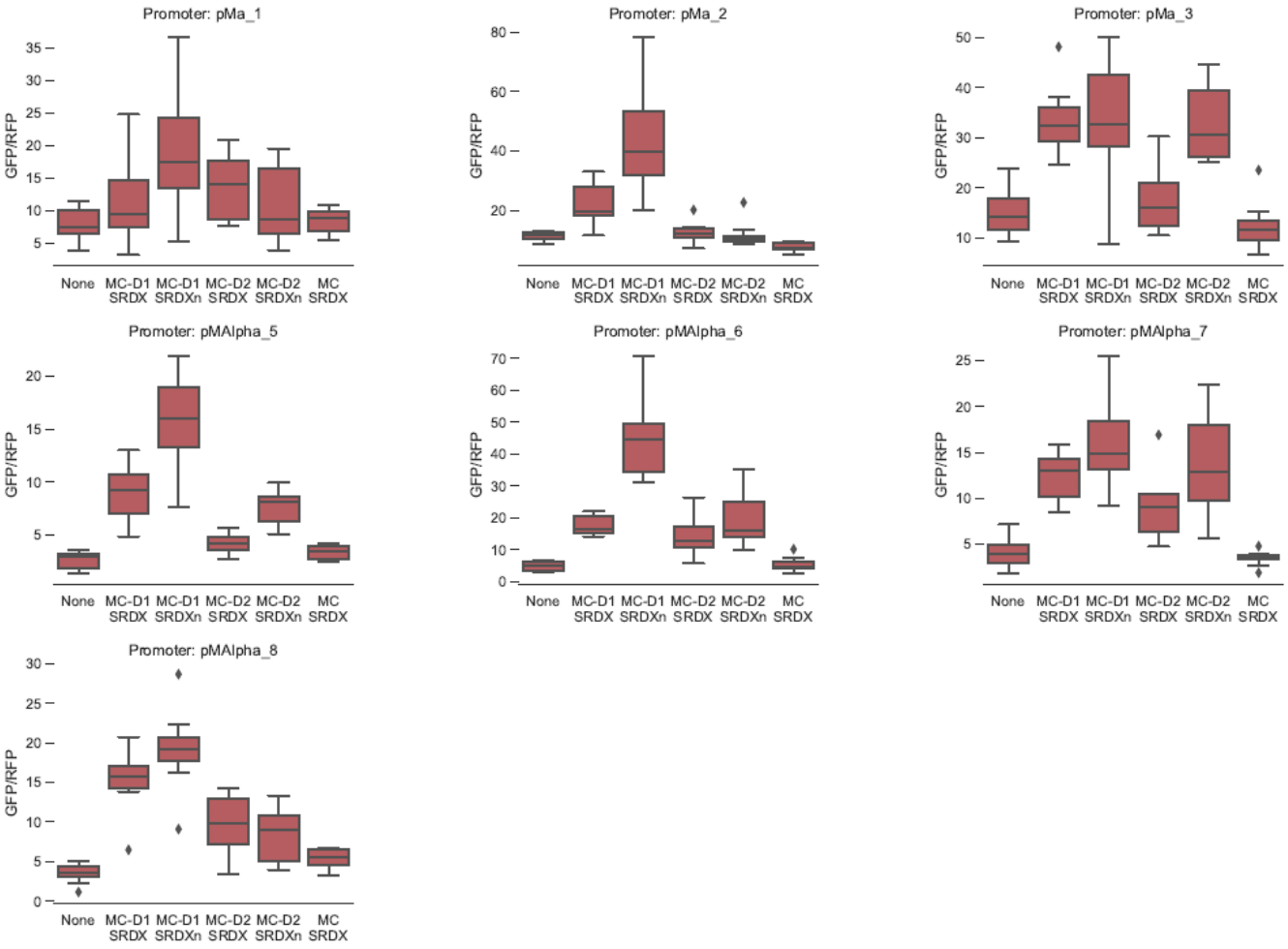
Supplementary Figure 13. Expression strengths of synthetic promoter and transcription factor pairs for the Mata1 transcription factor system. Combinations of MATAL1 *trans*-elements and corresponding synthetic promoters (*trans*-elements are denoted on the y-axis and synthetic promoters are denoted on the x-axis). Expression strength is calculated by dividing GFP fluorescence from a control RFP fluorescence. All values shown are the average of 8 independent biological replicates including standard error.



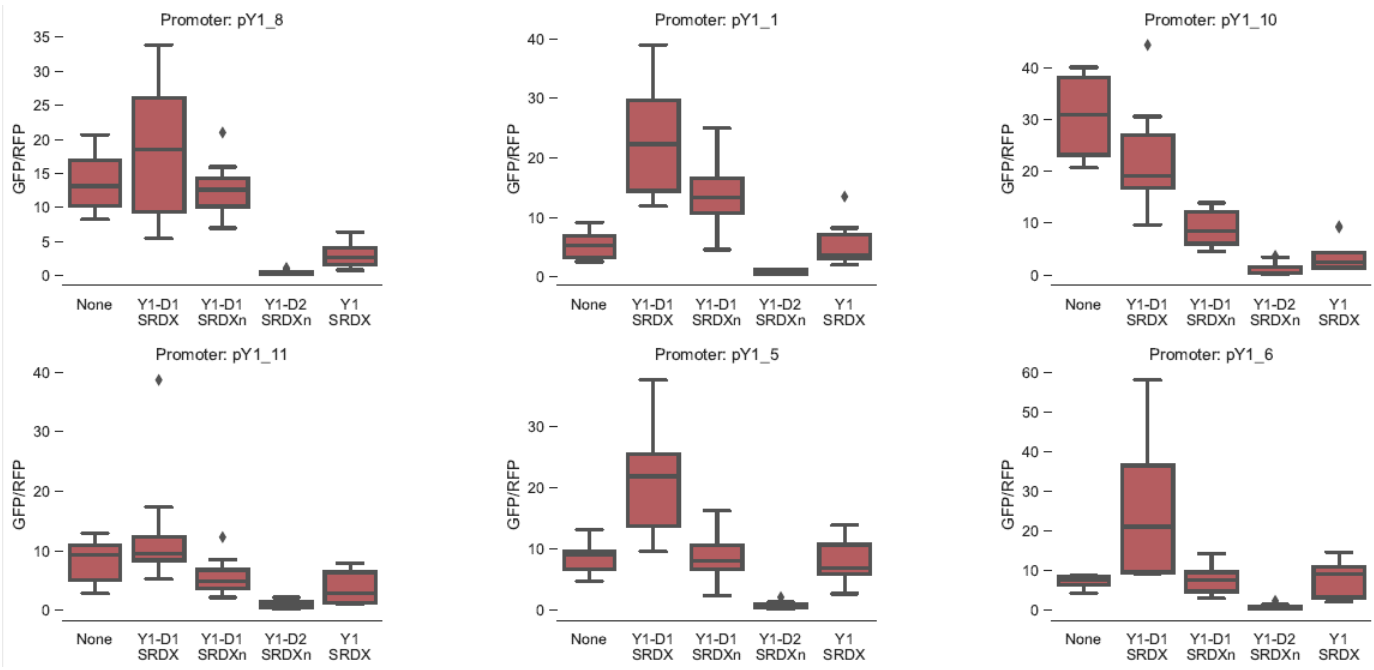
Supplementary Figure 14. Expression strengths of synthetic promoter and transcription factor pairs for the Matα2 transcription factor system. Combinations of MATAL2 *trans*-elements and corresponding synthetic promoters (*trans*-elements are denoted on the y-axis and synthetic promoters are denoted on the x-axis). Expression strength is calculated by dividing GFP fluorescence from a control RFP fluorescence. All values shown are the average of 8 independent biological replicates including standard error.



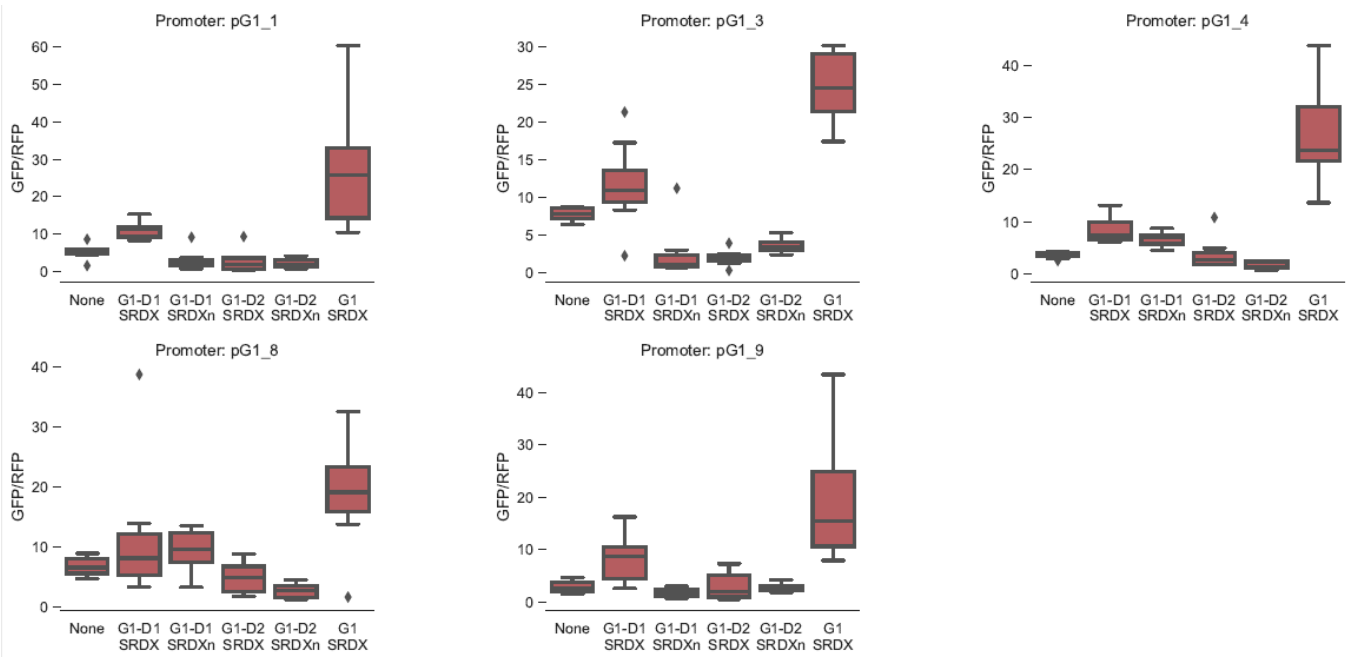
Supplementary Figure 15. Design and characterization of Gal4 repressor constructs. **a,** Various synthetic repressor fusion constructs were tested to optimize transcriptional repression. The SRDX repression domain was fused to the C- or N- terminus, with and without the VP16 domain. **b and c,** Two different synthetic promoters were expressed with the Gal4 synthetic activator, then co-infiltrated with the three different repressor constructs, which yielded a decrease in GFP expression (normalized to RFP). Construct pms6384 repressed gene expression the best for both synthetic promoters and was characterized as the most efficient synthetic repressor tested for the Gal4 system. Error bars represent the standard error for each condition tested (n = 8).



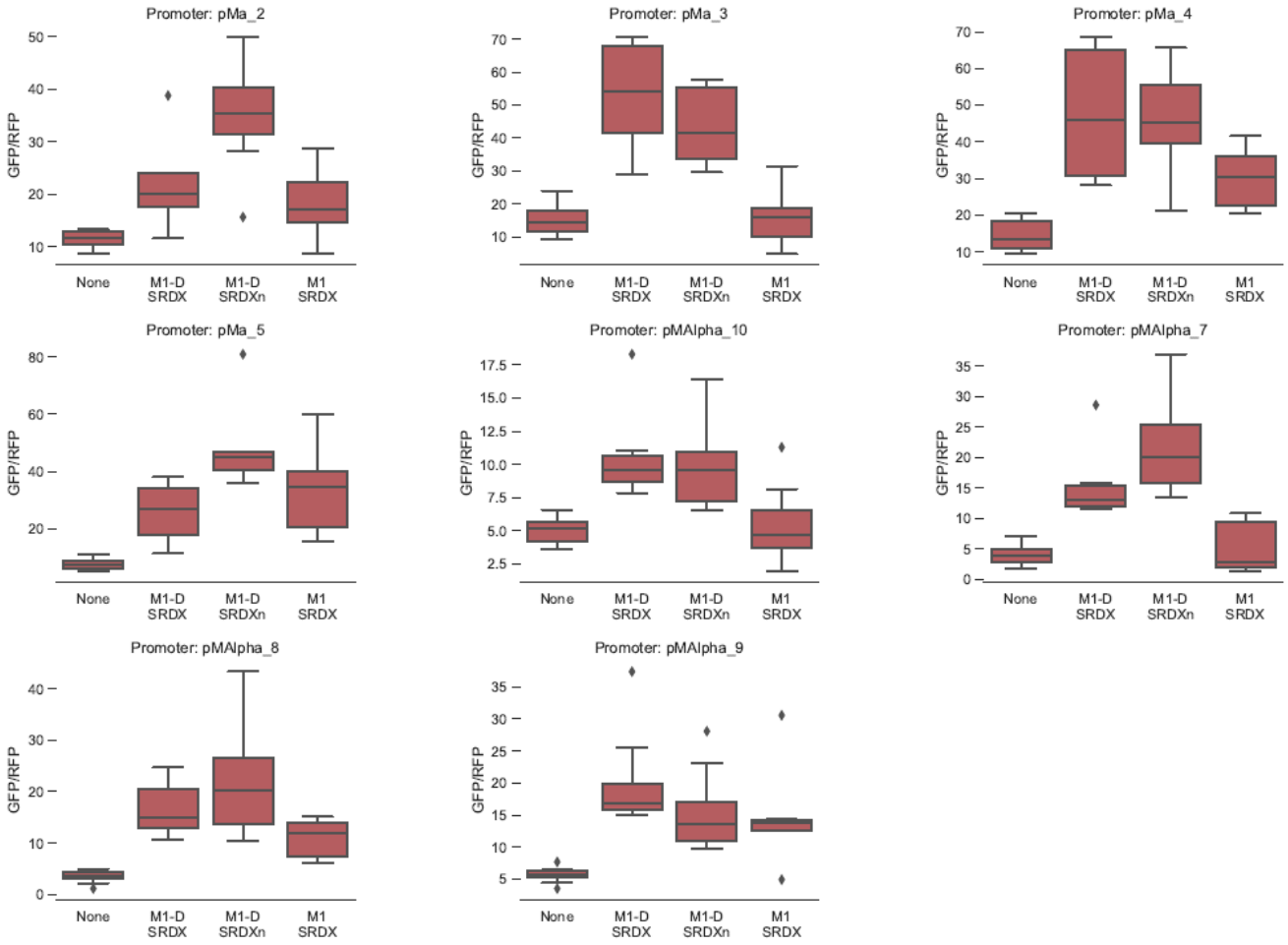
Supplementary Figure 16. Design and characterization of MCM1 repressor constructs. Various repressor constructs were generated by the fusion of the SRDX domain to either the C-terminus (SRDX) or N-terminus (SRDXn) of the full-length TF or the predicted DNA-binding domain for each TF. Repressors were tested for their capacity to reduce the background expression of each promoter, denoted as None on the x-axis. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5th and 95th percentile with values outside considered outliers (n = 8 independent biological replicates).



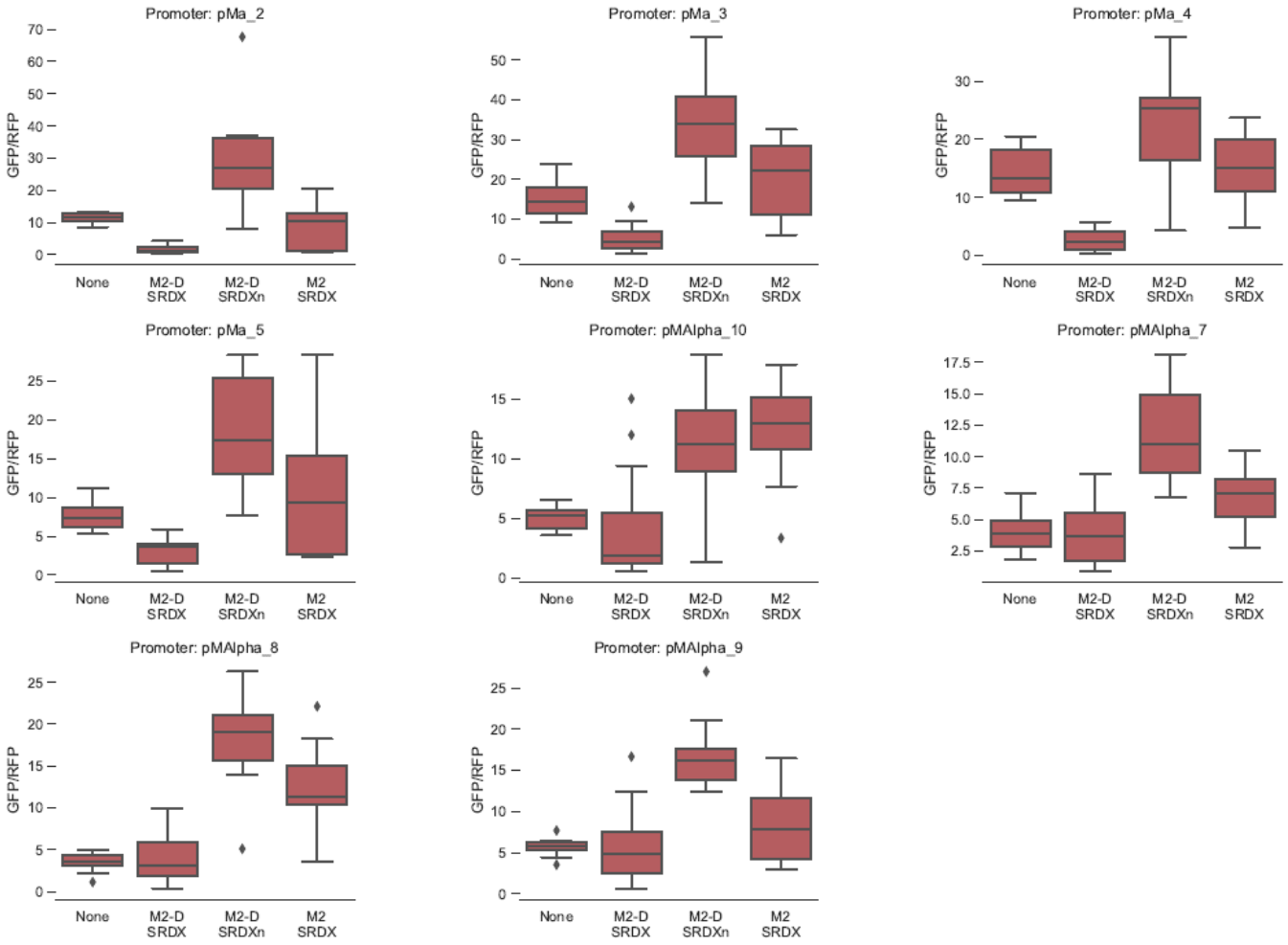
Supplementary Figure 17. Design and characterization of Yap1 repressor constructs. Various repressor constructs were generated by the fusion of the SRDX domain to either the C-terminus (SRDX) or N-terminus (SRDXn) of the full-length TF or the predicted DNA-binding domain for each TF. Repressors were tested for their capacity to reduce the background expression of each promoter, denoted as None on the x-axis. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5th and 95th percentile with values outside considered outliers (n = 8 independent biological replicates).



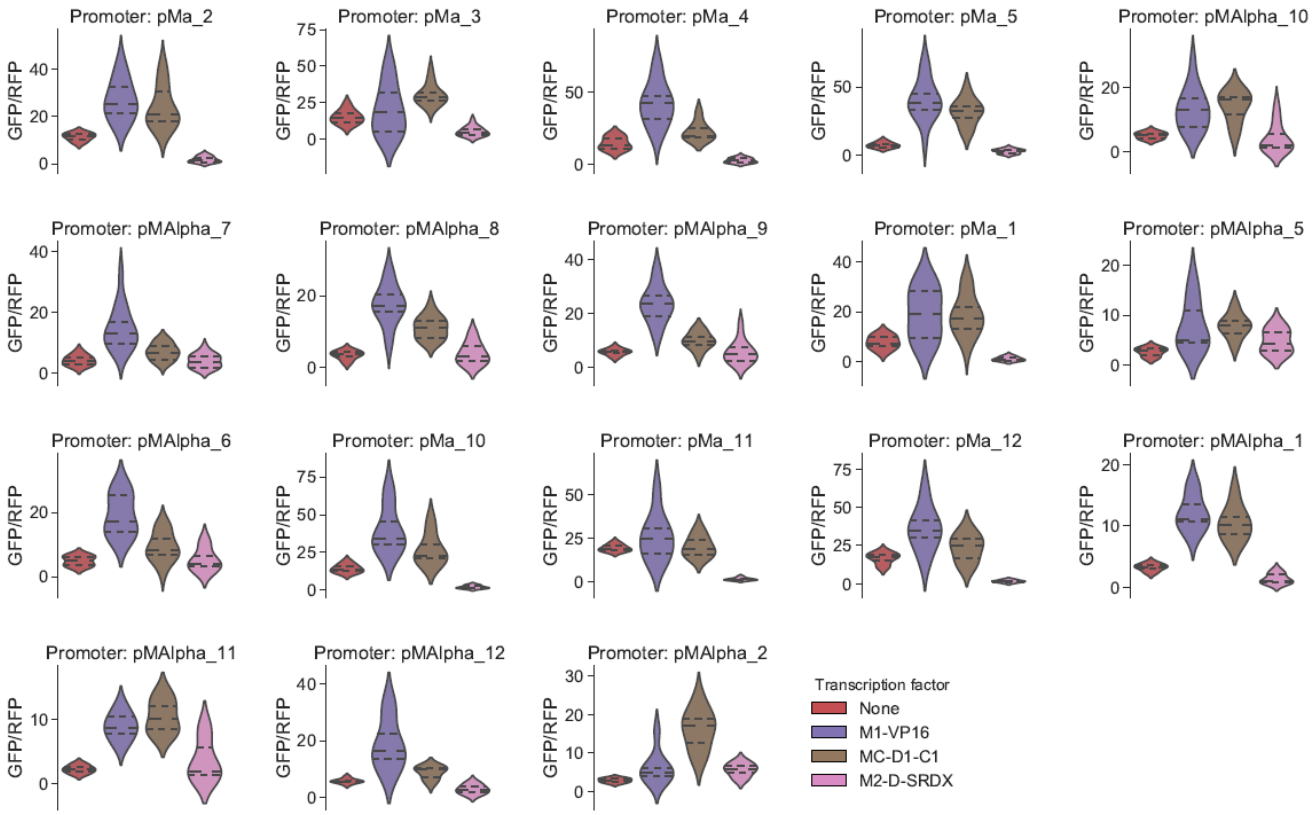
Supplementary Figure 18. Design and characterization of Gat1 repressor constructs. Various repressor constructs were generated by the fusion of the SRDX domain to either the C-terminus (SRDX) or N-terminus (SRDXn) of the full-length TF or the predicted DNA-binding domain for each TF. Repressors were tested for their capacity to reduce the background expression of each promoter, denoted as None on the x-axis. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5^{th} and 95^{th} percentile with values outside considered outliers (n = 8 independent biological replicates).



Supplementary Figure 19. Design and characterization of Mata1 repressor constructs. Various repressor constructs were generated by the fusion of the SRDX domain to either the C-terminus (SRDX) or N-terminus (SRDXn) of the full-length TF or the predicted DNA-binding domain for each TF. Repressors were tested for their capacity to reduce the background expression of each promoter, denoted as None on the x-axis. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5th and 95th percentile with values outside considered outliers (n = 8 independent biological replicates).



Supplementary Figure 20. Design and characterization of Matα2 repressor constructs. Various repressor constructs were generated by the fusion of the SRDX domain to either the C-terminus (SRDX) or N-terminus (SRDXn) of the full-length TF or the predicted DNA-binding domain for each TF. Repressors were tested for their capacity to reduce the background expression of each promoter, denoted as None on the x-axis. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5th and 95th percentile with values outside considered outliers (n = 8 independent biological replicates).



Supplementary Figure 21. Synthetic promoters that bind multiple TF families can be repressed and activated. Synthetic promoters that can bind multiple TF families broadens the expression range achievable with a single promoter and allows for more complex logic principles to be introduced. Additionally, the basal expression of each synthetic promoter can be repressed with one of the SRDX constructs. The capacity to modulate the expression strength of a single promoter with various *trans*-elements adds an additional level of control over gene expression. Standard violin plot represents the estimated kernel density with the median and interquartile range represented by dashed lines (n = 8 independent biological replicates).

Supplementary Table 1. Expression strengths of all promoters characterized in the initial promoter library described in Figure 1. *AtAct2* promoter was used to drive the synthetic activator.

Promoter name	Description	Construct Name	Expression Strength	Construct Name	Basal Expression Strength
pms5261	CCE01+A.t.Hp70	pms7526	2.12	pms6018	0.19
pms5267	CCE02+A.t.PRK	pms7527	4.10	pms6019	0.05
pms5263	CCE03+A.t.WUS	pms7528	4.53	pms6020	0.15
pms5264	CCE04+A.t.RGF1	pms7529	0.01	pms6021	0.01
pms5930	CCE05+D.c.Ext	pms5997	1.83	pms6022	0.43
pms5931	CCE06+G.m.RbcS	pms5998	0.02	pms6023	0.03
pms5932	CCE07+P.a.RbcS	pms5999	0.32	pms6024	0.04
pms5933	CCE08+T.a.RbcS	pms6000	0.01	pms6025	0.00
pms5934	CCE09+P.s.Cab	pms6001	0.72	pms6026	0.08
pms5936	CCE11+P.a.Cab	pms6003	0.04	pms6028	0.02
pms5937	CCE12+T.a.Glia	pms6004	2.54	pms6029	0.29
pms5938	CCE13+H.v.Hord	pms6005	1.64	pms6030	0.22
pms5939	CCE14+Z.m.Zein	pms6006	0.45	pms6031	0.07
pms5940	CCE15+P.s.Legu	pms6007	1.33	pms6032	0.17
pms5941	CCE16+R.c.Ricin	pms6008	0.00	pms6033	0.00
pms5942	CCE17+G.m.Leghe	pms6009	1.72	pms6034	0.05
pms5943	CCE18+G.m.Nod	pms6010	1.21	pms6035	0.03
pms5944	CCE19+G.m.HSP	pms6011	0.41	pms6036	0.02
pms5945	CCE20+A.t.Adh	pms6012	0.10	pms6037	0.02
pms5946	CCE21+A.m.Chalc	pms6013	0.67	pms6038	0.07
pms5948	CCE23+N.b.ATPs	pms6015	0.51	pms6040	0.06
pms5949	CCE24+Z.m.Triso	pms6016	0.12	pms6041	0.02
pms5950	CCE25+S.t.Prot	pms6017	0.74	pms6042	0.15

Supplementary Table 2. Expression strengths of all promoters characterized in the promoter library rationally designed to explore the additive effects of various DNA elements described in Figure 2. *AtAct2* promoter was used to drive the synthetic activator.

Promoter name	Description	Construct Name	Expression Strength	Construct Name	Basal Expression Strength
pms6189	CCE17+A.t.WUS	pms6259	9.64	pms6370	1.24
pms6188	CCE16+A.t.WUS	pms6258	3.25	pms6369	0.26
pms6185	CCE13+A.t.WUS	pms6255	2.92	pms6366	0.36
pms6176	CCE04+A.t.WUS	pms6246	1.66	pms6357	0.08
pms6194	CCE22+A.t.WUS	pms6264	1.50	pms6375	0.63
pms6542	CCE17+A.t.GL2	pms6592	2.65	pms6642	0.02
pms6543	CCE16+A.t.GL2	pms6593	0.30	pms6643	0.01
pms6544	CCE13+A.t.GL2	pms6594	0.32	pms6644	0.02
pms6545	CCE04+A.t.GL2	pms6595	2.26	pms6645	0.01
pms6546	CCE22+A.t.GL2	pms6596	0.13	pms6646	0.01
pms6547	CCE17+A.t.Srtrt	pms6597	1.40	pms6647	0.03
pms6548	CCE16+A.t.Srtrt	pms6598	0.97	pms6648	0.03
pms6549	CCE13+A.t.Srtrt	pms6599	0.00	pms6649	0.06
pms6550	CCE04+A.t.Srtrt	pms6600	1.39	pms6650	0.02
pms6551	CCE22+A.t.Srtrt	pms6601	0.81	pms6651	0.04
pms6557	CCE17+A.t.Pist	pms6607	5.14	pms6657	0.10
pms6558	CCE16+A.t.Pist	pms6608	2.24	pms6658	0.12
pms6559	CCE13+A.t.Pist	pms6609	6.11	pms6659	0.08
pms6560	CCE04+A.t.Pist	pms6610	1.34	pms6660	0.11
pms6561	CCE22+A.t.Pist	pms6611	1.14	pms6661	0.04
pms6562	CCE17+A.t.ASFT	pms6612	4.26	pms6662	0.04
pms6563	CCE16+A.t.ASFT	pms6613	4.26	pms6663	0.08
pms6564	CCE13+A.t.ASFT	pms6614	3.79	pms6664	0.05
pms6565	CCE04+A.t.ASFT	pms6615	1.29	pms6665	0.03
pms6566	CCE22+A.t.ASFT	pms6616	0.00	pms6666	0.04

Supplementary Table 3. Primer validation for primers used in qPCR experiment. All primer sets were validated with plasmid DNA. Primer sets were then tested with cDNA from a control plant infiltrated with an empty vector control to confirm no off-target amplification.

Target	Primer name	Sequence	R2	Efficiency	Intercept
GFP	GFP_F	AGCCACAACGTCTATATC	0.999	103.5%	37.69
	GFP_R	CTCAGGTAGTGGTTGTC			
mCherry	mCherry_F	TAATGCAGAAGAAGACCAT	0.999	102.0%	38.13
	mCherry_R	CTTCTTGGCCTTGTAGG			
VP16	VP16_F	ACTGATGTTTCTCTTGGA	0.999	103.0%	36.2
	VP16_R	CACCATATGGAGCTGAA			
C1	C1_F	CTCCAAAGGCTGTTAGAT	0.999	97.3%	40.39
	C1_R	AGAACCAACCCTCAATG			
Matα2	Matal2_F	AGTCATGGTTCGCTAAG	0.995	107.1%	35.86
	Matal2_R	GGAGCAATTGTAATAGTCTT			