

Supplemental Material to:

Dissection of thousands of cell type-specific enhancers identifies dinucleotide repeat motifs as general enhancer features

J. Omar Yáñez-Cuna¹, Cosmas D. Arnold¹, Gerald Stampfel¹, Łukasz M. Boryń¹,
Daniel Gerlach^{1,2}, Martina Rath¹, Alexander Stark^{1*}

¹Research Institute of Molecular Pathology (IMP), 1030 Vienna, Austria

²Current address: Boehringer Ingelheim RCV GmbH & Co KG, Vienna, Austria

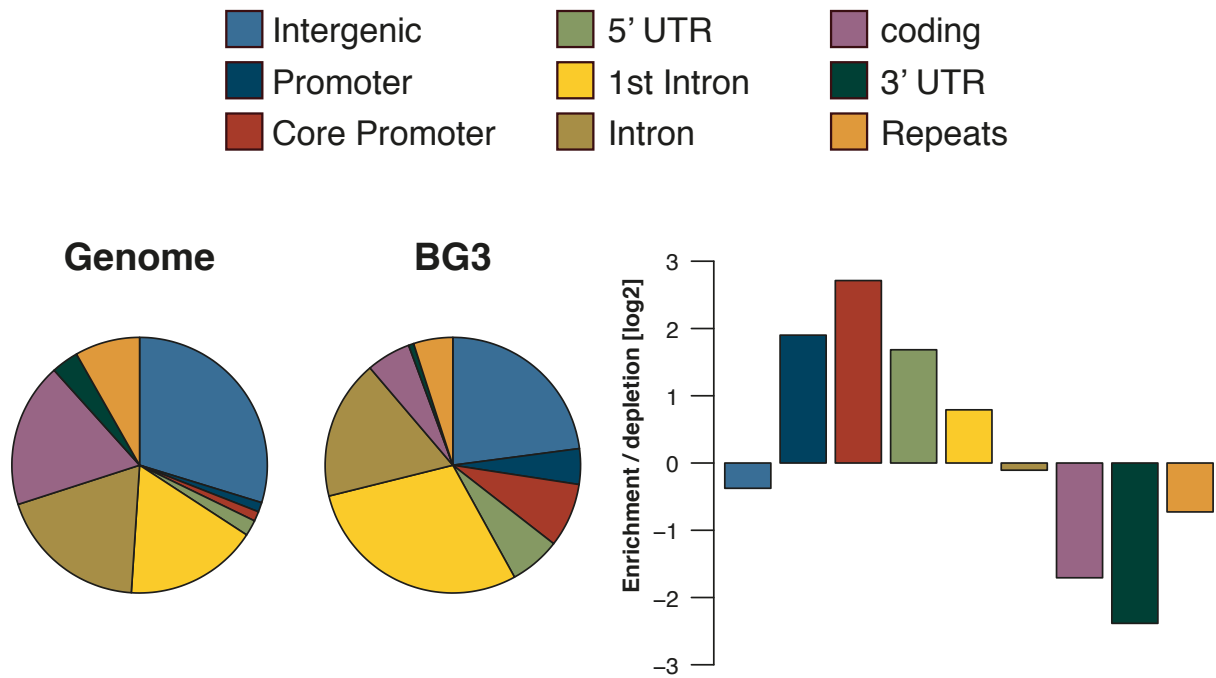
* To whom correspondence should be addressed (stark@starklab.org)

Supplemental Figures	2
<i>Supplemental Fig. S1. STARR-seq for BG3 cells</i>	2
<i>Supplemental Fig. S2. Gene Ontology (GO) analysis</i>	3
<i>Supplemental Fig. S3. TSS distal enhancers are also enriched in DRMs</i>	4
<i>Supplemental Fig. S4. Motif enrichments between enhancer classes</i>	5
<i>Supplemental Fig. S5. Sequence-based enhancer predictions</i>	6
<i>Supplemental Fig. S6. Mutations of non-important control motifs</i>	7
<i>Supplemental Fig. S7. Motif occurrences of enriched TF motifs in all four enhancer classes</i>	8
<i>Supplemental Fig. S8. DRM occurrences and lengths in all four enhancer classes</i>	9
<i>Supplemental Fig. S9. Increased evolutionary conservation of DRMs.</i>	10
<i>Supplemental Fig. S10. Conservation of DRM occurrences and lengths.</i>	11
<i>Supplemental Fig. S11. The AP-1 motif is required for broadly active enhancers</i>	13
<i>Supplemental Fig. S12. Synthetic enhancers</i>	14
Supplemental Tables	15
<i>Supplemental Table S1. SVM Predictions using different cross-validation schemes</i>	15
<i>Supplemental Table S2. Broad enhancers are a distinct enhancer class</i>	16
Reporter and primer sequences for the luciferase experiments	17
References	25

Supplemental Figures

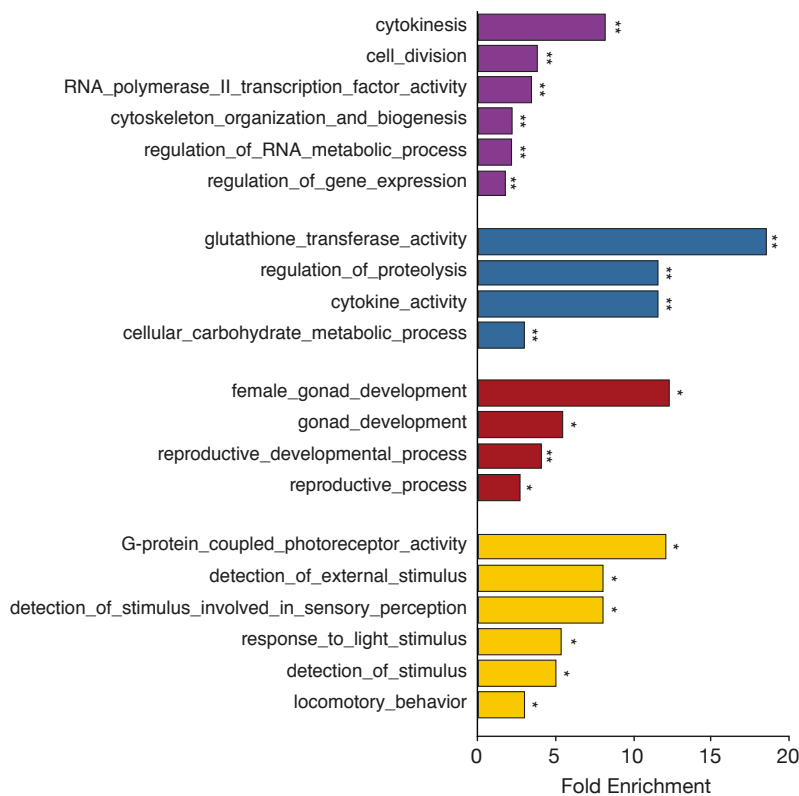
Supplemental Fig. S1. STARR-seq for BG3 cells

	Replicate 1	Replicate 2
Reads	20,892,038	16,506,577
Fragments	24,536	207,707



We performed STARR-seq for BG3 cells (Ui et al. 1994). Shown are the number of paired-end reads and collapsed fragments obtained for the two biological replicates (Top). The two biological replicates had a Pearson correlation coefficient of 0.75. Pie charts showing the distribution STARR-seq peaks called on the combined replicates in defined genomic regions and the relative sizes of these regions in the genome. The enrichment of peaks in the different genomic regions relative to the regions' sizes in the genome is quantified as fold enrichment (in log₂). All deep sequencing data are available at www.starklab.org and GEO (series accession number GSE49809).

Supplemental Fig. S2. Gene Ontology (GO) analysis

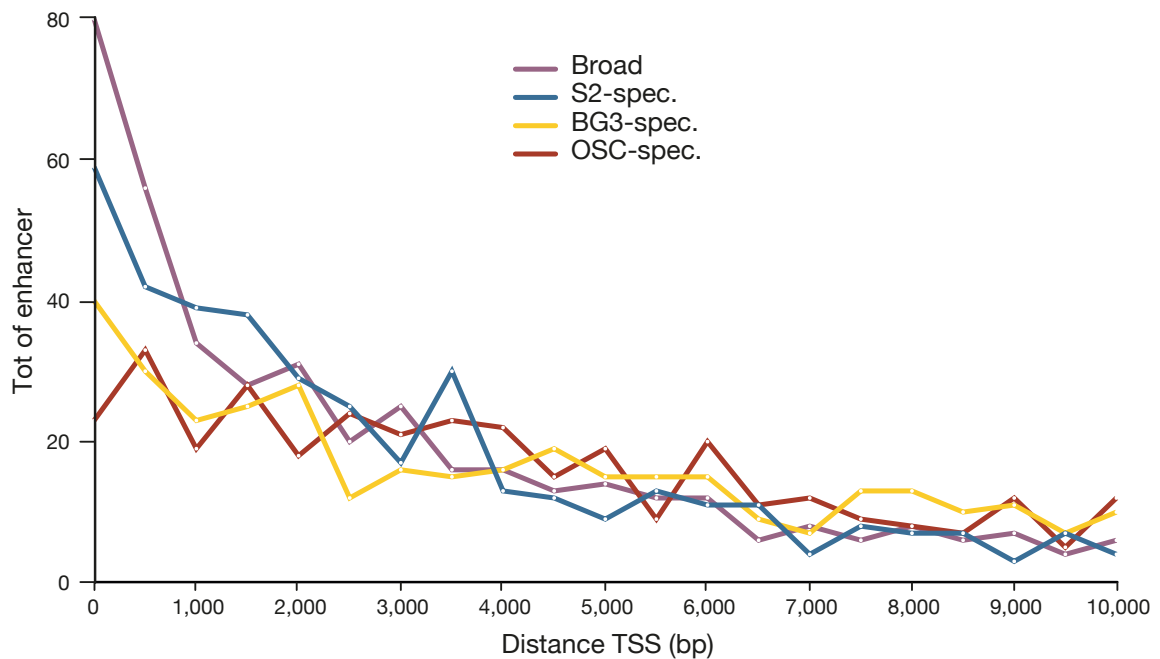


Genes associated to enhancers from different enhancer classes were enriched in particular GO categories. Shown are the GO categories that are most highly enriched for genes associated with a given enhancer class versus the other three enhancer classes (according to fold enrichment; categories are selected from the top 10, excluding non-specific categories; broad (purple), S2-specific (blue), OSC-specific (red) and BG3-specific (yellow) enhancers; * FDR adjusted p-value ≤ 0.05 ; ** FDR adjusted p-value ≤ 0.001). The table shows the same GO categories but with values for each of the enhancer classes.

	S2-cells		OSC-cells		BG3-cells		Ubi	
	log ₂ (FE)	p-value	log ₂ (FE)	p-value	log ₂ (FE)	p-value	log ₂ (FE)	p-value
reproductive_developmental_process	0.386	0.955	4.104	0.010	0.310	0.956	0.820	0.726
RNA_polymerase_II_transcription_factor_activity	0.713	0.852	0.274	0.992	0.537	0.931	3.470	0.000
gonad_development	0.386	0.919	5.472	0.031	0.504	1.000	0.820	0.727
female_gonad_development	0.464	1.000	12.312	0.025	0.806	1.000	0.683	0.796
regulation_of_gene_expression	0.853	0.821	0.739	0.919	0.665	0.966	1.794	0.001
reproductive_process	0.409	0.967	2.736	0.027	0.448	0.931	1.104	0.500
regulation_of_RNA_metabolic_process	0.662	0.969	0.586	0.975	0.743	0.889	2.190	0.000
glutathione_transferase_activity	18.542	0.000	0.513	1.000	0.504	1.000	0.256	1.000
cytokinesis	0.193	0.991	0.316	0.954	0.310	0.956	8.202	0.000
cell_division	0.244	0.995	0.912	0.653	0.403	0.952	3.845	0.001
cytoskeleton_organization_and_biogenesis	0.480	0.997	0.849	0.750	0.793	0.818	2.226	0.000
cellular_carbohydrate_metabolic_process	3.013	0.007	0.616	0.859	0.183	0.994	0.724	0.819
regulation_of_proteolysis	11.589	0.008	0.821	1.000	0.806	1.000	0.410	1.000
G-protein_coupled_receptor_activity	0.302	0.994	2.565	0.019	1.612	0.175	0.683	0.864
detection_of_external_stimulus	0.464	0.884	0.821	0.731	8.062	0.016	0.293	1.000
detection_of_stimulus	0.331	0.944	0.586	0.826	5.039	0.019	0.586	0.850
cytokine_activity	11.589	0.008	0.821	1.000	0.806	1.000	0.410	1.000
locomotory_behavior	0.632	0.844	0.684	0.793	3.023	0.042	0.559	0.888
detection_of_stimulus_involved_in_sensory_perception	0.386	1.000	1.026	0.665	8.062	0.016	0.410	0.908
response_to_light_stimulus	0.290	1.000	0.684	0.784	5.375	0.032	0.820	0.727

Supplemental Fig. S3. TSS distal enhancers are also enriched in DRMs

A.

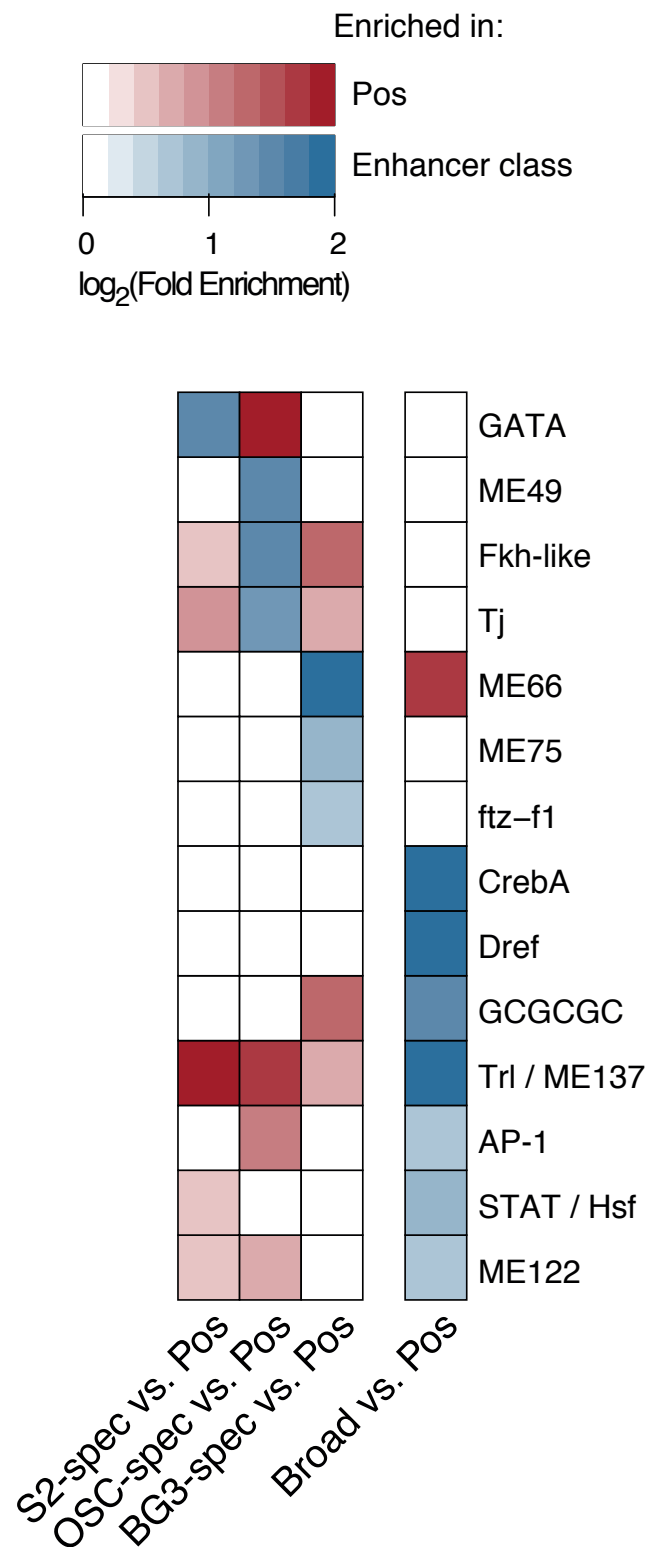


B.

		With Promoter		Without Promoter	
		FE	p-value	FE	p-value
S2-spec.	CA	1.154	4.479E-29	1.181	1.084E-30
	GA	0.872	6.298E-07	1.010	3.726E-09
	CG	0.941	0.001	1.000	0.000
	TA	-0.107	0.137	-0.250	0.006
BG3-spec.	CA	1.369	1.688E-43	1.492	1.044E-53
	GA	1.561	9.498E-24	1.312	7.392E-16
	CG	0.163	0.355	0.263	0.249
	TA	0.585	8.673E-13	0.672	8.057E-17
OSC-spec.	CA	0.925	7.462E-18	0.896	1.056E-16
	GA	0.905	2.036E-07	0.937	6.364E-08
	CG	0.941	0.001	0.604	0.034
	TA	-0.420	1.650E-05	-0.453	4.405E-06
Broad	CA	1.900	3.322E-100	1.935	3.355E-105
	GA	3.082	1.084E-155	3.171	1.834E-170
	CG	2.287	8.468E-24	2.263	3.844E-23
	TA	-0.150	0.062	-0.073	0.229

(A) Distribution of enhancer-to-TSS distances for enhancers from each of the four different classes (500 enhancers each). Broad enhancers (purple) are more frequently found close to TSS (e.g. 80 are within 500bp compared to 59 for S2-specific enhancers). **(B)** DRMs of the CA, GA, and CG type are enriched in enhancers, in particular broadly active enhancers when considering all enhancers (left columns) or only enhancers that were at least 100bp from the nearest TSS (i.e. excluding enhancers that overlapped 201bp-long windows centered on each TSS; right columns). Shown are fold-enrichment values (log2-scale) and hypergeometric p-values.

Supplemental Fig. S4. Motif enrichments between enhancer classes



Heatmap showing significant motifs (FDR adjusted p-value ≤ 0.01 and fold enrichment $\log_2 \geq 0.7$) when comparing the enhancers of each class (500 enhancers each) against a pool of enhancers from the other three classes (1500 enhancers; Pos).

Supplemental Fig. S5. Sequence-based enhancer predictions

A.

	vs. Neg	AUC	vs. Pos	AUC
S2-spec	75.4	0.84	65.3	0.70
BG3-spec	74.5	0.80	63.9	0.67
OSC-spec	76.5	0.85	71.6	0.79
Broad	81.0	0.90	68.7	0.74

B.

S2-specific

		Pred Pow Neg	Pred Pow Pos
GATA		0.0605	0.019
E-box		0.0315	0.017
ME139		0.0145	-
AP-1		0.0135	-
STARR-Mot2		0.006	-
Ptx1		-	0.008

BG3-specific

		Pred Pow Neg	Pred Pow Pos
STARR-Mot43		0.0215	-
AP-1		0.016	-
Trl		0.014	-
Hsf / STAT		0.012	-
ME140		0.011	-
CACA		0.006	-
Ptx		-	0.009

OSC-specific

		Pred Pow Neg	Pred Pow Pos
Fkh-like		0.038	0.0265
STARR-Mot65		0.0165	-
STARR-Mot88		0.013	-
CACA		0.006	-
Hsf / STAT		0.004	-
CACA		0.002	-
STARR-Mot36		-	0.007

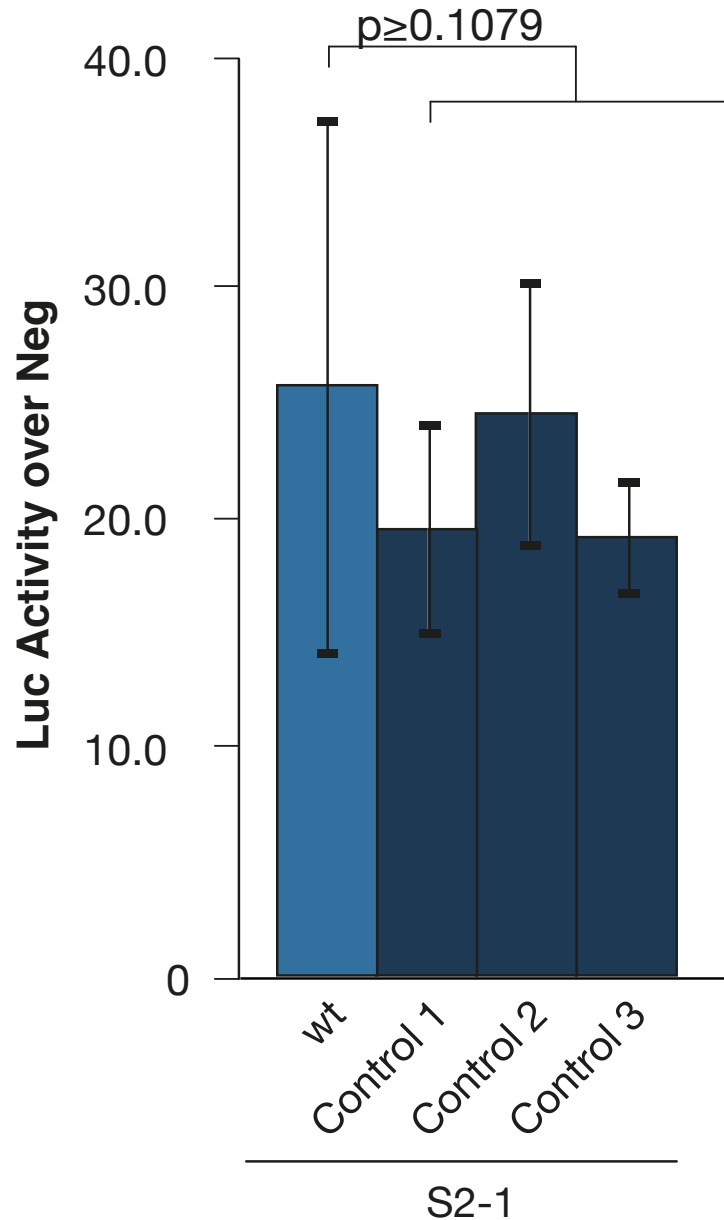
Broadly active

		Pred Pow Neg	Pred Pow Pos
AP-1		0.0205	0.032
CACA		0.0105	-
Trl		0.001	0.029
ME54		-	0.008
STARR-Mot2		-	0.0125

(A) Performance of the sequence-based enhancer predictions (binary classification) for S2-specific, BG3-specific, OSC-specific and Broad active enhancers against either negative regions (Neg) or against a pool with enhancers from the other three classes (Pos) (see text) based solely on motif occurrences. Shown are the accuracies (recall; light shading) and area under the ROC curve (AUC; grey shading) for each of the comparisons.

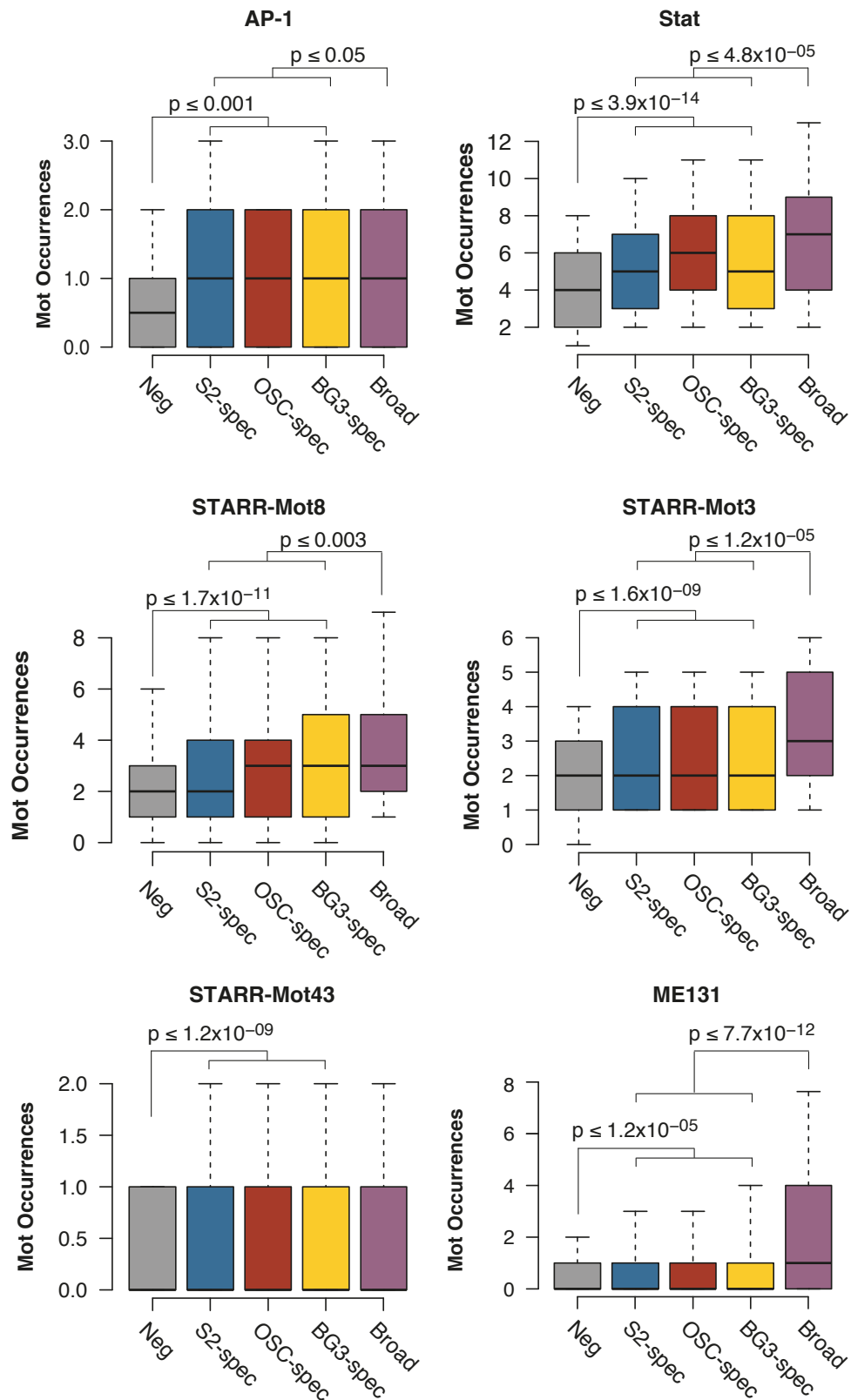
(B) Shown are the predictive power values (Pred Pow) for the motifs selected as the most discriminative features (calculated by difference in the accuracy in the presence or absence of the TF motif; see (Yanez-Cuna et al. 2012)) and that have a fold enrichment of 1.4 or higher for the S2-specific, BG3-specific, OSC-specific and broadly active enhancers.

Supplemental Fig. S6. Mutations of non-important control motifs



Luciferase assays in S2 cells of the wild-type enhancer S2-1 (light blue) and three mutant variants in which different control motifs were disrupted by point mutations. In Control 1 and 2, we mutated the sub-threshold motif matches of important motifs (AP-1 and GA [Trl/ME137]). In Control 3 we mutated the motif TorRE, which was enriched in negative regions and depleted in active enhancers. There was no significant drop in enhancer activity as indicated by the p-value (shown is the lowest [i.e. most significant] t-test p-value from comparing the wild-type enhancer against each of the controls). Error bars show the standard deviation of at least 3 biological replicates. DNA sequences of the wild-type and mutant enhancers are shown below (page 18).

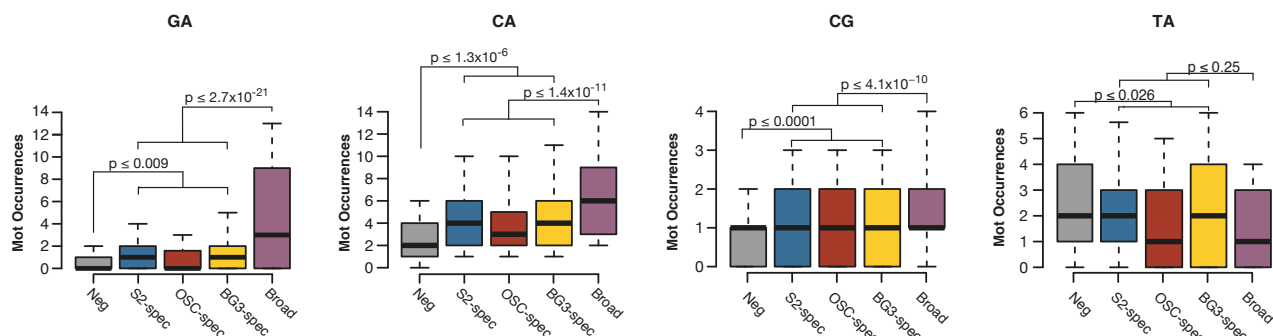
Supplemental Fig. S7. Motif occurrences of enriched TF motifs in all four enhancer classes



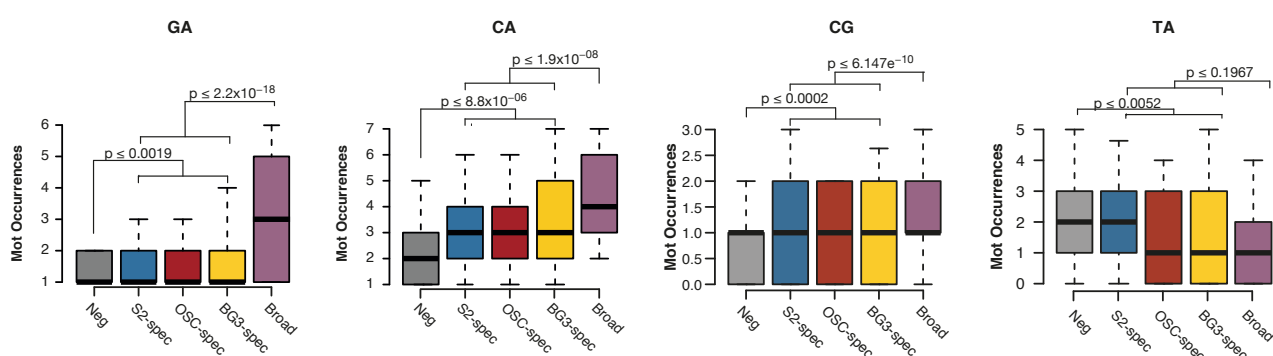
Boxplot of motif occurrences per enhancer as in Figure 4A in the 1,000 negative regions (gray), the top 500 S2-specific (blue), OSC-specific (red), BG3-specific (yellow) and Broad (purple) enhancers. The boxes indicate the median and the 25 and 75 percentiles and the whiskers mark the 10 and 90 percentiles.

Supplemental Fig. S8. DRM occurrences and lengths in all four enhancer classes

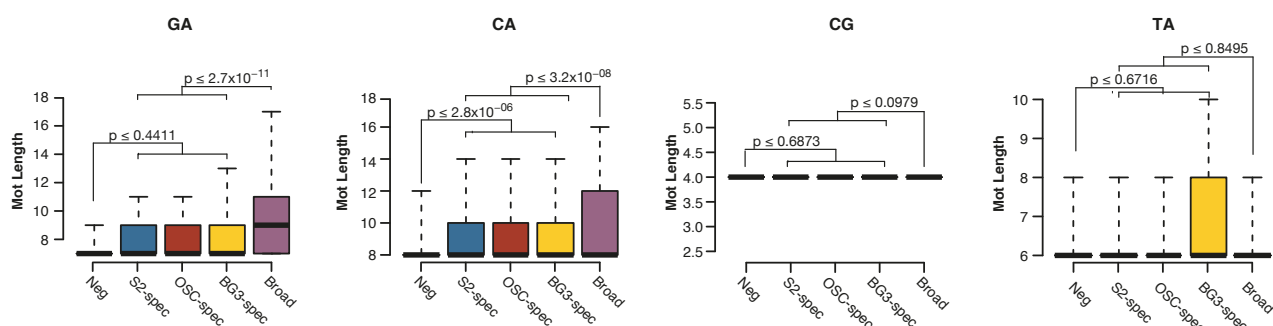
A.



B.

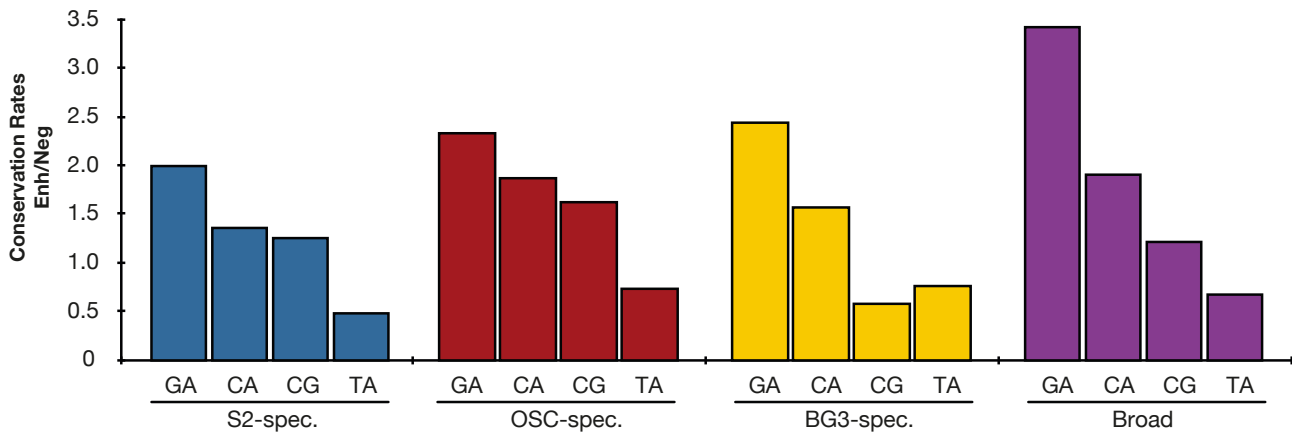


C.



Boxplots of **(A)** all DRM occurrences, **(B)** non-overlapping DRM occurrences per enhancer, and **(C)** lengths of all DRM instances. Shown are the 1,000 negative regions (Neg; gray), the top 500 S2-specific (blue), OSC-specific (red), BG3-specific (yellow) and Broad (purple) enhancers. Wilcoxon p-values for the comparison of inactive regions (Neg) against the cell type-specific enhancers and the cell type-specific against the broadly active enhancers. The following motifs were used (see Fig 2C for motif logos): ME137 (GA), STARRMot15 (CA), GCGC (CG) and TATA (TA) with a PWM cutoff ≤ 0.0039 .

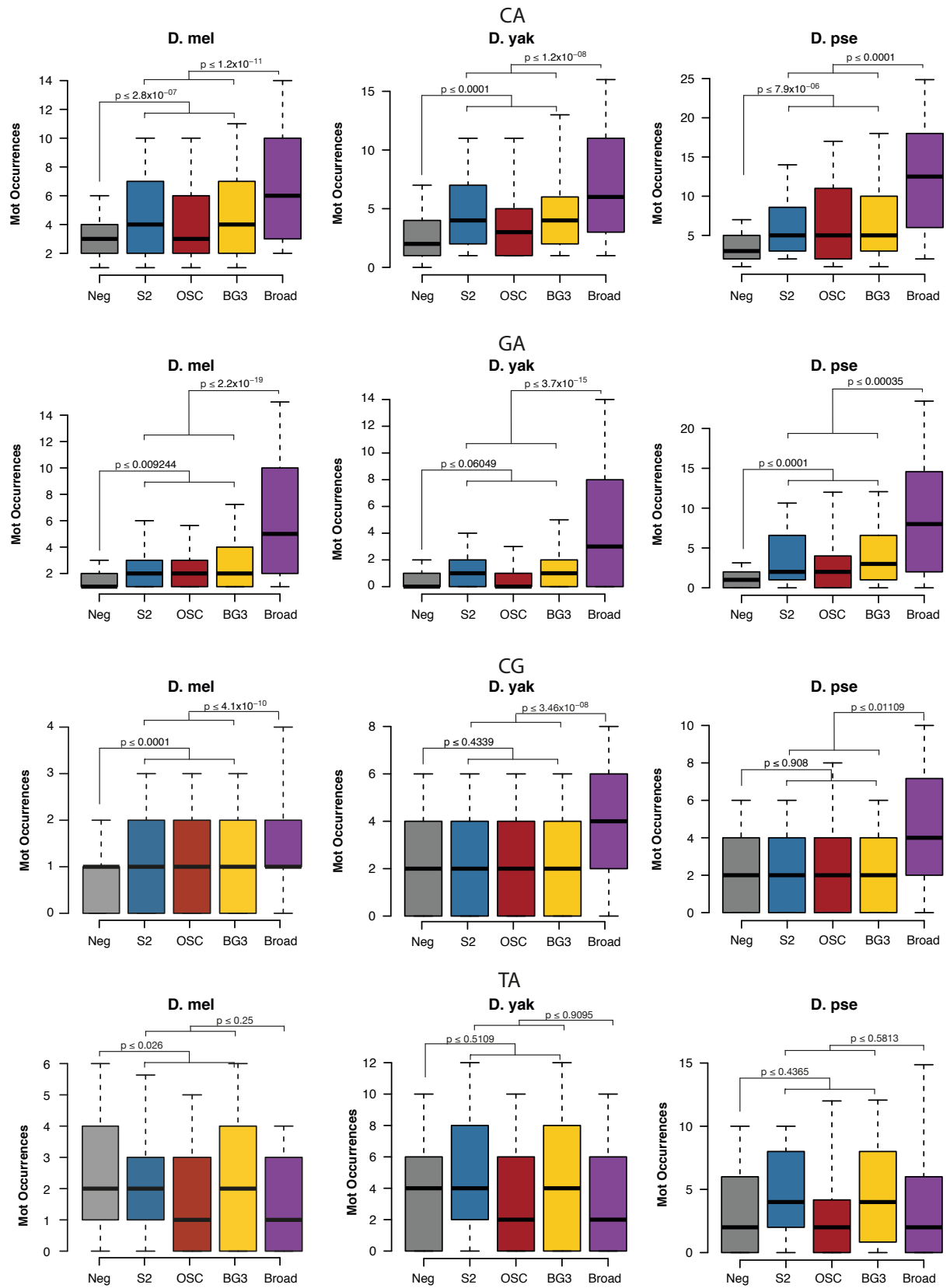
Supplemental Fig. S9. Increased evolutionary conservation of DRMs.



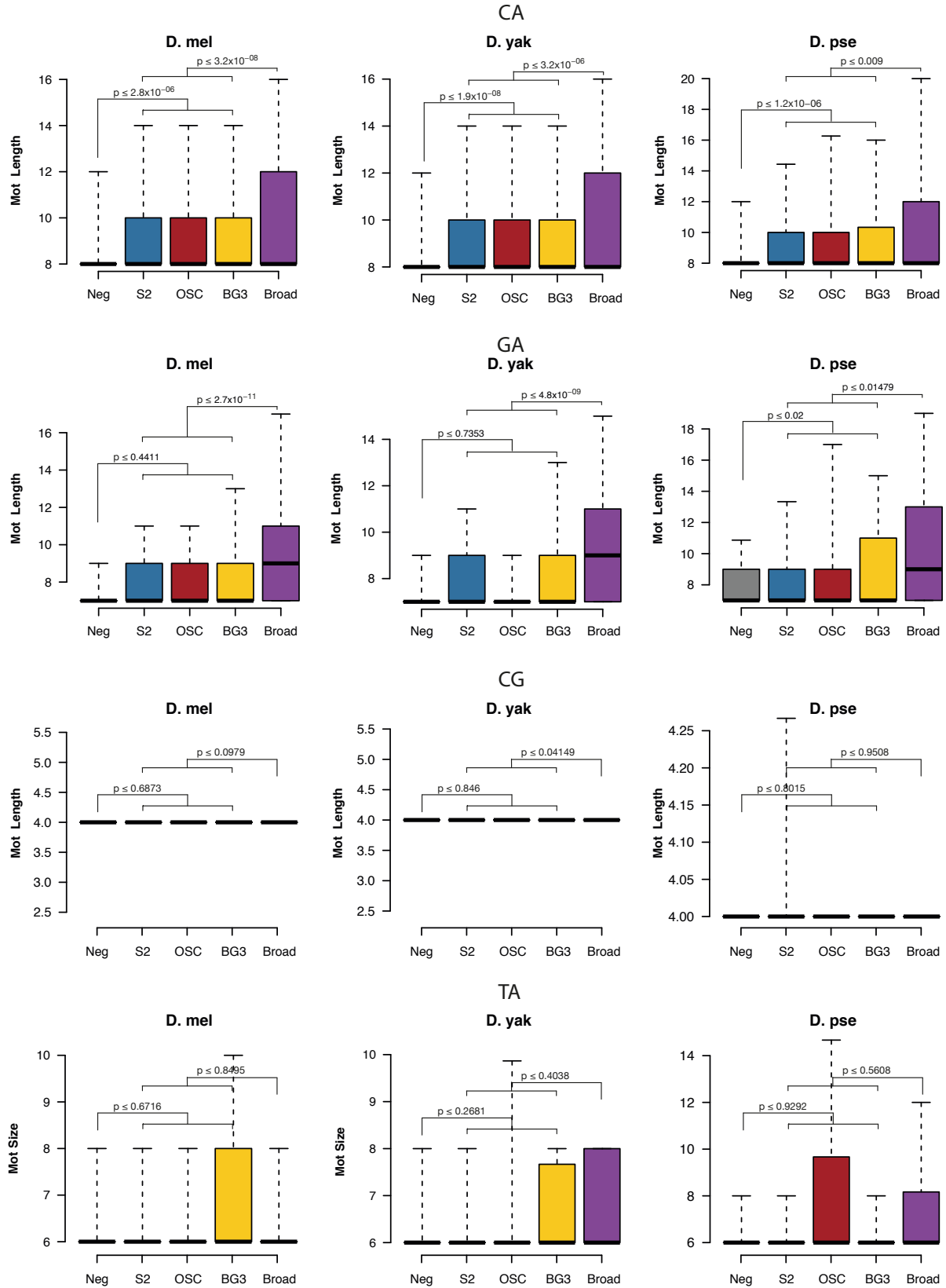
Conservation rates of the DRMs in each of the four different classes of active enhancers compared to their rates in negative regions (displayed is the ratio of (rate in enhancers / rate in negative regions). The rates were determined as the proportion of motifs with branch-length-scores [BLS] of 2.13 or 50% of the *Drosophila* phylogeny (Kheradpour et al. 2007).

Supplemental Fig. S10. Conservation of DRM occurrences and lengths.

A.

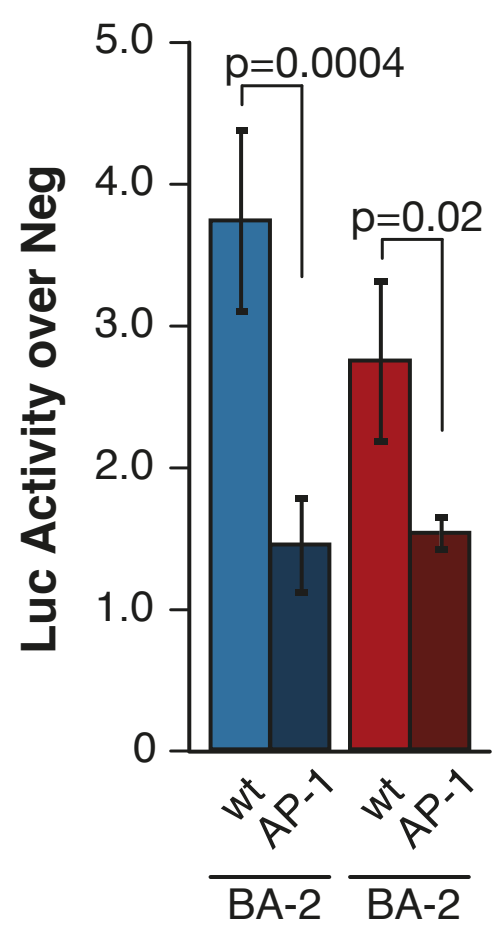


B.



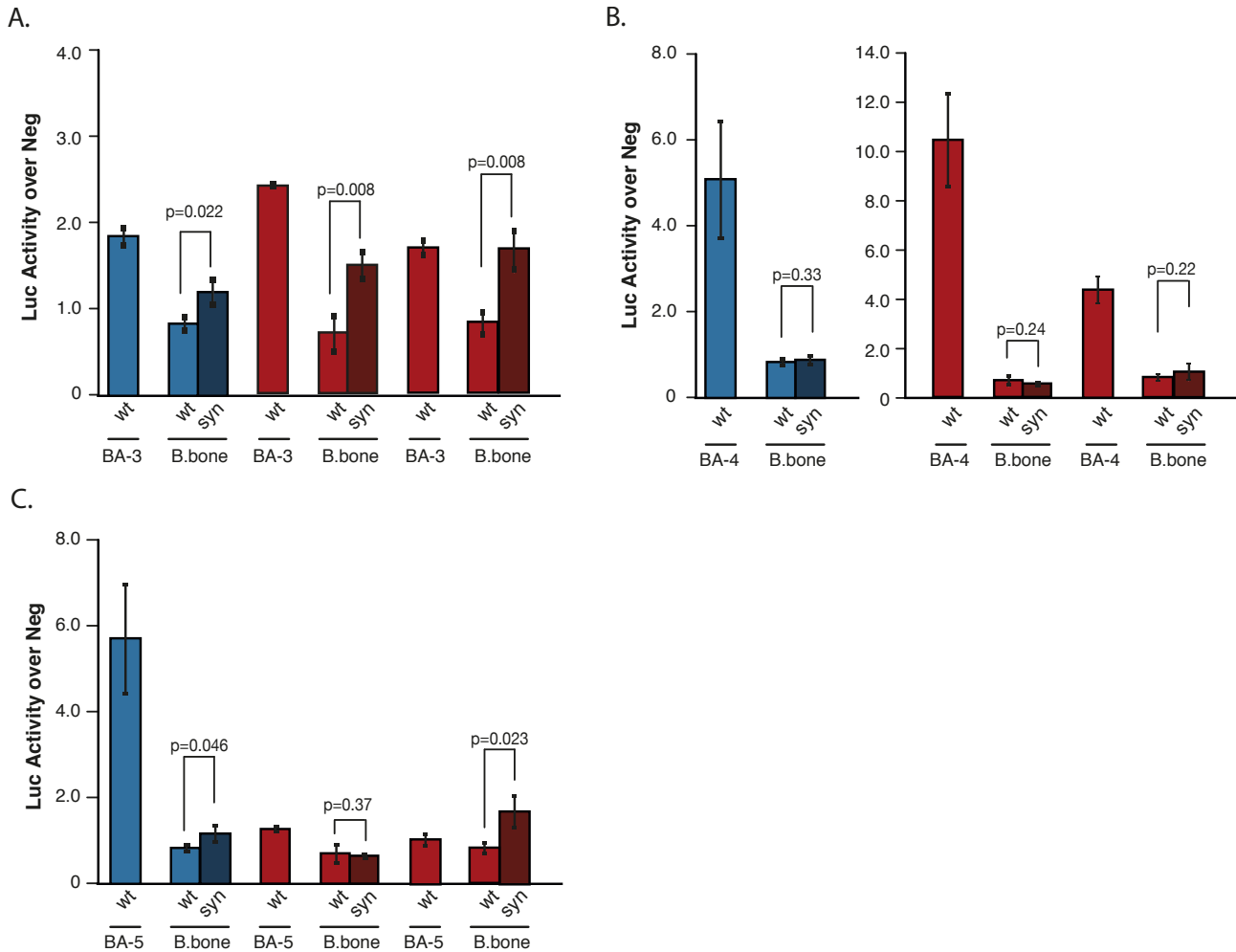
Boxplot of the motif lengths (A) and motif occurrences (B) for all DRM instances as in Supplemental Fig. S8 but comparing *D. melanogaster* (*D. mel*), *D. yakuba* (*D. yak*) and *D. pseudoobscura* (*D. pse*). The respective trends for DRM occurrences and lengths are conserved across species.

Supplemental Fig. S11. The AP-1 motif is required for broadly active enhancers



Luciferase assays in S2 cells (blue) and OSCs (red) of the broadly active enhancer 2 (BA-2; wt; light) and an AP-1 motif mutant variant (dark). Shown are t-test p-values and error bars that depict standard deviations from three to five biological replicates. DNA sequences of the wild-type and mutant enhancers are shown below (page 18).

Supplemental Fig. S12. Synthetic enhancers



(A-C) Synthetic enhancers created by copying DRMs and TF motifs from broadly active enhancers BA-3 to BA-5 into an inactive sequence (backbones; B.bone) as in main Fig. 4F. Shown are normalized luciferase activities from S2 cells (blue) and OSCs (red) for wildtype enhancer and backbone sequences (light) and synthetic enhancers (dark) **(A)** Synthetic enhancer based on BA-3 from which only GA and CA type DRMs were copied. **(B)** Synthetic enhancer based on BA-4 from which GA and CA type DRMs as well as E-box and GATA motifs were copied. **(C)** Synthetic enhancer based on BA-5 from which GA, CA, and CG type DRMs as well as a GATA motif were copied. For the OSC experiments (red) two biological replicates are shown separately and the standard deviations are based on three technical replicates each. P-values are from unpaired t-tests.

Supplemental Tables

Supplemental Table S1. SVM Predictions using different cross-validation schemes

Set 1	Set 2	Fold CV	Acc	SD
S2	Neg	LOOCV	0.757	0.008
		5	0.752	0.004
		10	0.761	0.003
		15	0.749	0.004
		20	0.754	0.006
S2	Pos	LOOCV	0.637	0.034
		5	0.661	0.031
		10	0.659	0.026
		15	0.652	0.029
		20	0.653	0.035
BG3	Neg	LOOCV	0.751	0.008
		5	0.741	0.020
		10	0.730	0.013
		15	0.748	0.016
		20	0.745	0.013
BG3	Pos	LOOCV	0.640	0.016
		5	0.612	0.024
		10	0.617	0.022
		15	0.627	0.007
		20	0.639	0.008
OSC	Neg	LOOCV	0.772	0.019
		5	0.760	0.013
		10	0.768	0.014
		15	0.772	0.015
		20	0.765	0.018
OSC	Pos	LOOCV	0.721	0.007
		5	0.698	0.004
		10	0.705	0.007
		15	0.710	0.009
		20	0.716	0.009
Broad	Neg	LOOCV	0.803	0.023
		5	0.802	0.012
		10	0.813	0.013
		15	0.818	0.011
		20	0.810	0.011
Broad	Pos	LOOCV	0.696	0.004
		5	0.688	0.005
		10	0.683	0.004
		15	0.685	0.009
		20	0.687	0.001

SVM prediction accuracies (Acc) using different cross-validation schemes: Leave-One-Out cross-validation as in the main text (LOOCV), 5-fold, 10-fold, 15-fold and 20-fold cross-validation. The accuracy was calculated as the average of three different predictions with bootstrapped test and training sets (SD: standard deviation). Predictions accuracies were constant across all cross-validation schemes.

Supplemental Table S2. Broad enhancers are a distinct enhancer class

A.

Set 1	Set 2	acc	roc
Broad	Negative	81.0	0.900
	S2-spec + OSC-spec + BG3-spec.	68.7	0.740
	(S2 and OSC)-specific	61.4	0.618
	(S2 and BG3)-specific	64.0	0.677
	(OSC and BG3)-specific	62.4	0.675

B.

Set 1	Set 2	acc	roc
(S2 and OSC)-specific	Negative	66.2	0.715
	S2-spec + OSC-spec	47.2	0.467
(S2 and BG3)-specific	Negative	70.4	0.783
	S2-spec + BG3-spec	53.8	0.533
(OSC and BG3)-specific	Negative	63.4	0.702
	OSC-spec + BG3-spec	54.1	0.555

(A) Broadly active enhancers can be successfully discriminated from negative regions, enhancers that are active in two but not all three cell types, and – most importantly – from a pool of enhancers that are specifically active in a single cell type (i.e. in either S2 cells, OSCs, or BG3 cells). The latter and the observation that DRMs are specifically highly enriched in broadly active enhancers argues that these enhancers constitute a distinct enhancer class. **(B)** Enhancers that are active in two but not all three cell types can be successfully discriminated from negative regions and broadly active enhancers (see A) but not from a pool of cell type specific enhancers that are active in the same two cell types. This suggests that enhancers that are active in two but not all three cell types carry signatures of cell type specific enhancers but do not constitute a distinct enhancer class.

Reporter and primer sequences for the luciferase experiments

Enhancer	Primers	5' - 3'
S2-1	Forward	<i>GCCCGGGTTATAAAATACAAAATTCGAAAA</i>
	Reverse	<i>TGCACAATTTCTGTGAAGTGACAGCC</i>
S2-2	Forward	<i>GCTTCCACATTGAATTCTCATCGTTTTG</i>
	Reverse	<i>CAAGCATAGAGTTTGCCAGAGAGCCG</i>
S2-3	Forward	<i>CCAATGAATGTTTTCCAAACCTATATTAGTCA</i>
	Reverse	<i>TGAAACAGCAACAACATAGAGCAGAA</i>
OSC-1	Forward	<i>TGCGTGTCCACTAAAAGCCGAAAAA</i>
	Reverse	<i>AGTCATCATCGTCATCGTCGCCTGT</i>
OSC-2	Forward	<i>AGAATGCTCAGTTTTTCATAGTATTCAAATTGT</i>
	Reverse	<i>ATTATTGCCGAATTACGCCGCC</i>
OSC-3	Forward	<i>TTTTGATAGCCCACGATGAGGGAG</i>
	Reverse	<i>CTGTTTGATATAGCCCCACAGCCAA</i>
BA-1	Forward	<i>ACGTATCAGCGTTGGCGAATAAAAA</i>
	Reverse	<i>AAAAACGGAGTGAAAAAGCCA</i>
BA-2	Forward	<i>CCATGCATATGCTTATCACAACGGA</i>
	Reverse	<i>TTTGGGTGCTTCAACTGCGGTTTTT</i>
BA-3	Forward	<i>AATTTGCATTGACTGCTGCCACAAC</i>
	Reverse	<i>CGCAATTGATTAAAGGCCAGCACTTTT</i>
B.bone	Forward	<i>CCAACACAAATCCCAAGAAAGAACACA</i>
	Reverse	<i>CCTGGTGGCTCCGCCTTCATC</i>
HsEnh-1	Forward	<i>CCCGTCGCGGTGGAGCTTAG</i>
	Reverse	<i>AACAGGGATTCCACCCGCCG</i>
HsEnh-2	Forward	<i>TCCATCACTGCCTTCAAGTATGATAAAATTGCT</i>
	Reverse	<i>TCACCAAGGAAAAGGTATTCTCCAA</i>
HsEnh-3	Forward	<i>GCCGCTCGCAACCCCGAC</i>
	Reverse	<i>TCTCGGATCTGCTGGAAGAATCAGC</i>

S2-1 wt

GCCCCGGGTTATAAAATACAAAATTTCGAAAATAGCACAACGCTTGGCTTTCGTAGCTAAACATCGAAAACGAA
ATCCAAAAACAAAGACTCAAAGCACCCCTGTAGCCCCGGCAATTGTGCGCAGTTACCCCAACAACACTACACACAT
TTTTTCAAACGACGAACAAAAGATAATGTCGACAACCTAAACCGCAGATATGCATATGTTTATGTACTCCGTAC
AAGTCGCTGTACTACGCACATACACACATAGTATATTTAACAAAACGGAGCCAAGTCGCGGAAGTAATTCTC
CGCATCTTATCGCACTTAGAGATTCCGCCCCACAATCCAAATCCATCCCGCATCATTTGGGAATGCAGACTGC
GTGCCAGATCTGGCCGTATGTATATCGTCTAAATTGATATTAGCGGACGAGAGCTACTTGACACTATTTCTC
AGTGGCTGTCACTTCACAGAAATTGTGCA

S2-1 GATA

GCCCCGGGTTATAAAATACAAAATTTCGAAAATAGCACAACGCTTGGCTTTCGTAGCTAAACATCGAAAACGAA
ATCCAAAAACAAAGACTCAAAGCACCCCTGTAGCCCCGGCAATTGTGCGCAGTTACCCCAACAACACTACACACAT
TTTTTCAAACGACGAACAAAGATGCAATGTCGACAACCTAAACCGCAGATATGCATATGTTTATGTACTCCGTAC
AAGTCGCTGTACTACGCACATACACACATAGTATATTTAACAAAACGGAGCCAAGTCGCGGAAGTAATTCTC
CGCATCTCTGGGCACTTAGAGATTCCGCCCCACAATCCAAATCCATCCCGCATCATTTGGGAATGCAGACTGC
GTGCCAGATCTGGCCGTATGTATATCGTCTAAATTGATATTAGCGGACGAGAGCTACTTGACACTATTTCTC
AGTGGCTGTCACTTCACAGAAATTGTGCA

S2-1 CA

GCCCCGGGTTATAAAATACAAAATTTCGAAAATAGCACAACGCTTGGCTTTCGTAGCTAAACATCGAAAACGAA
ATCCAAAAACAAAGACTCAAAGCACCCCTGTAGCCCCGGCAATTGTGCGCAGTTACCCCAACAACACTACGTCTAT
TTTTTCAAACGACGAACAAAAGATAATGTCGACAACCTAAACCGCAGATATGCATATGTTTATGTACTCCGTAC
AAGTCGCTGTACTACGCACCTAGGACACATAGTATATTTAACAAAACGGAGCCAAGTCGCGGAAGTAATTCTC
CGCATCTTATCGCACTTAGAGATTCCGCCCCACAATCCAAATCCATCCCGCATCATTTGGGAATGCAGACTGC
GTGCCAGATCTGGCCGTATGTATATCGTCTAAATTGATATTAGCGGACGAGAGCTACTTGACACTATTTCTC
AGTGGCTGTCACTTCACAGAAATTGTGCA

S2-1 CG

GCCCCGGGTTATAAAATACAAAATTTCGAAAATAGCACAACGCTTGGCTTTCGTAGCTAAACATCGAAAACGAA
ATCCAAAAACAAAGACTCAAAGCACCCCTGTAGCCCCGGCAATTGATCAAAGTTACCCCAACAACACTACACACAT
TTTTTCAAACGACGAACAAAAGATAATGTCGACAACCTAAACCGCAGATATGCATATGTTTATGTACTCCGTAC
AAGTCGCTGTACTACGCACATACACACATAGTATATTTAACAAAACGGAGCCAAGTATCCAAGTAATTCTC
CGCATCTTATCGCACTTAGAGATTCCGCCCCACAATCCAAATCCATCCCGCATCATTTGGGAATGCAGACTGC
GTGCCAGATCTGGCCGTATGTATATCGTCTAAATTGATATTAGCGGACGAGAGCTACTTGACACTATTTCTC
AGTGGCTGTCACTTCACAGAAATTGTGCA

S2-2 wt

GCTTCCACATTGAATTCTCATCGTTTTTGGTGGGTGAAAACGTGACACGTTCACTCTTTTCGCCGCTGGCTTA
TCACGTTGGCTTATTCCGTATCACATGTTGCCACAGCAGTTTCTATTTTCGACACACACGCCGCCCCCTTTT
GTGTCTTCGCCCCCTGTTTGTGTATTGCTCTCTCCCCGCCGCTCTTTGGCGTTGATAATGAACCGAGTGACA
TGGAGTGAAAGTGAGGCAATCATCGGCGGGCCAGCATTAGATCATATATACAGACAGAAACGACGACGACATC
ATCATGATGATCATGATCATCATCGTACACTCTGACTCATGAACCTCACTCGCCTCACTTGCACTGGACAGA
CTCTCCGGATATCTCTCGCTCTCGCTTGCTCGCTCGCTCCCTCGCCCGTCAGTAGATTCTCCAATTCTCCGG
CTCTCTGGCAAACCTCTATGCTTG

S2-2 GATA

GCTTCCACATTGAATTCTCATCGTTTTTGGTGGGTGAAAACGTGACACGTTCACTCTTTTCGCCGCTGGCTCA
TAACGTTGGCTTATTCCGTATCACATGTTGCCACAGCAGTTTCTATTTTCGACACACACGCCGCCCCCTTTT
GTGTCTTCGCCCCCTGTTTGTGTATTGCTCTCTCCCCGCCGCTCTTTGGCGTTTCGCATGAACCGAGTGACA
TGGAGTGAAAGTGAGGCAATCATCGGCGGGCCAGCATTAGATCATATATACAGACAGAAACGACGACGACATC
ATCATGATGATCATGATCATCATCGTACACTCTGACTCATGAACCTCACTCGCCTCACTTGCACTGGACAGA
CTCTCCGGATATCTCTCGCTCTCGCTTGCTCGCTCGCTCCCTCGCCCGTCAGTAGATTCTCCAATTCTCCGG
CTCTCTGGCAAACCTCTATGCTTG

S2-3 wt

CCAATGAATGTTTTCCAAACCTATATTAGTCAGTATATTTGGCCTATATATCAGTAAACAAATAAATAGCTA
AACGGTAGATTGTTGAAAATAATATCGCAATACGCCACTGTTTCATAACAGAAGCAGAAGTTCGATGCCGGCA
GGGAATGAATGCGAATTAAGCAATACATATGATATGCAAATATGAAGCGCAACAATTATAACACGCTCTTTT
GTTGATATTCTTTGTTTGAAGTGCAGTGCAGTAAAGATAAAGCTGAAAACGTGCCGCGCAGCTGTTGCGGCTT
CTCATATCTATTTATAGCGGTTTTATTGCAGCACAGCACCCACACGCTTTTGTGTGCTCTTCTCCATCCTC
CTTTGCTACTATTTTCATCTCATACGTGACTAGGCACGCATAGCATGGGAAGTGGTAGCCCTGTTGCACAGT
GCTTAACAACACTAAGTTTTCTGCTCTATGTTGTTGCTGTTTCA

S2-3 E-box

CCAATGAATGTTTTCCAAACCTATATTAGTCAGTATATTTGGCCTATATATCAGTAAACAAATAAATAGCTA
AACGGTAGATTGTTGAAAATAATATCGCAATACGCCACTGTTTCATAACAGAAGCAGAAGTTCGATGCCGGCA
GGGAATGAATGCGAATTAAGCAATAATTACAATATGCAAATATGAAGCGCAACAATTATAACACGCTCTTTT
GTTGATATTCTTTGTTTGAAGTGCAGTGCAGTAAAGATAAAGCTGAAAACGTGCCGCGGGGCTCTTGCGGCTT
CTCATATCTATTTATAGCGGTTTTATTGCAGCACAGCACCCAGGCGCTTTTGTGTGCTCTTCTCCATCCTC
CTTTGCTACTATTTTCATCTCATACGTGACTAGGCACGCATAGCATGGGAAGTGGTAGCCCTGTTGCACAGT
GCTTAACAACACTAAGTTTTCTGCTCTATGTTGTTGCTGTTTCA

OSC-1 wt

TGCGTGTCCACTAAAAGCCGAAAAAGAAGAGGAAGCGTAACGCGGCTCCATAAACAACGAAGGGCGCTCAGC
ACGCAAGTCGTGGAAGAACAATAAACACGAACCCCGAAAATCCAATGACTGTGCTCCGTCCAGCGGATTGAA
AGTGTGCTGGCTGGGTATGTGGATGTGGATTTGGATCGGGCACTGGAGGAGGAGGAGCTGTCCAAGTGGGGC
AGTCGCAAGGGCGAAGTGTTAGGGCGAAATTCGAGCAGCGTAAGCTGAAACCGAAACCAAAGCTGAAAACCTC
GAGCCGCGCTCGCTCAAAACTATAACATTTTTTCGTTTTATTCCCTTATTTTGTGTGTTTTATTTTTATGGC
CAGCACTGCATTTATTGCGACAGCCCCCGGGTATCGAGAACAGGCGACGATGACGATGATGACT

OSC-1 Fkh

TGCGTGTCCACTAAAAGCCGAAAAAGAAGAGGAAGCGTAACGCGGCTCCTAGGTGAACGAAGGGCGCTCAGC
ACGCAAGTCGTGGAAGACATTCGAAACGAACCCCGAAAATCCAATGACTGTGCTCCGTCCAGCGGATTGAA
AGTGTGCTGGCTGGGTATGTGGATGTGGATTTGGATCGGGCACTGGAGGAGGAGGAGCTGTCCAAGTGGGGC
AGTCGCAAGGGCGAAGTGTTAGGGCGAAATTCGAGCAGCGTAAGCTGAAACCGAAACCAAAGCTGAAAACCTC
GAGCCGCGCTCGCTCAAAACTATAACATTTTTTCGTTTTATTCCCTTATTTTGTGTGTTTTATTTTTATGGC
CAGCACTGCATTTATTGCGACAGCCCCCGGGTATCGAGAACAGGCGACGATGACGATGATGACT

OSC-2 wt

AGAATGCTCAGTTTTTTCATAGTATTCAAATTGTATTCTATTTTGTCCCCATGCCACATTCATATTCATATCT
GGGTTTTAGAAACCACCGAGTGGGTCTCAGCTTGCAATCATAGATTTGGGTTTTCATCTCGCCGTTTTCCGAG
AAGTGGCATCTTCTGTGTGGTCTGTGCGCATGGATAAATGGGCTTCCAGCGACGACGACCCCCGCGGGCAAC
AGTTTAAGAATCAAATAAACACTTCGATCCCGATCCCTCTGGACTTCGATAAGTTATGTCAACAAACTTCG
GTCGCGTAACGGCGCATGCAGGAGCTCTGGAGTAGTAGGAGCTCCAGCTGAAGTGCAGCATTGGGTTACAAT
CGTCCACATGGAGGAGATCAGCATCAGCCCCTAGGAGGGGGAGGGGGCGGTGTTGGAACTTTGGAGCAGTCC
AGCATTCAGTTGGGGCGGCGTAATTCGGCAATAAT

OSC-2 Fkh

AGAATGCTCAGTTTTTTCATAGTATTCAAATTGTATTCTATTTTGTCCCCATGCCACATTCATATTCATATCT
GGGTTTTAGAAACCACCGAGTGGGTCTCAGCTTGCAATCATAGATTTGGGTTTTCATCTCGCCGTTTTCCGAG
AAGTGGCATCTTCTGTGTGGTCTGTGCGCATGGATAAATGGGCTTCCAGCGACGACGACCCCCGCGGGCAAC
AGTTTAAGAATCAAATCAGGACTTCGATCCCGATCCCTCTGGACTTCGATAAGTTATGTCAACAAACTTCG
GTCGCGTAACGGCGCATGCAGGAGCTCTGGAGTAGTAGGAGCTCCAGCTGAAGTGCAGCATTGGGTTACAAT
CGTCCACATGGAGGAGATCAGCATCAGCCCCTAGGAGGGGGAGGGGGCGGTGTTGGAACTTTGGAGCAGTCC
AGCATTCAGTTGGGGCGGCGTAATTCGGCAATAAT

OSC-3 wt

TTTTGATAGCCACGATGAGGGAGGGCAACAGACGAGAAGGCCACAGGCGCTTAAATATTTGGGTAAAAGA
GAGTTGGGAATCAGGCCGGAATTGGAACCTCACTTTGGCCGCCGATGGGAGTTTGCAGAATTCAAAAATTT
CACACCATCTCACAACCTGTTTGTTTTTGCGCTGGAATGCAGTTTGAATGTTATTGCAGAGTGAATGTTTTTC
TGACTCACACTCACAATCCCAGCCTCATTTAAACGAAGCCAGAGTTTCGTATAAACATCTTCTGCAGTCGT
TTGATTCCCGATTATAAACAAATGCCTGCGAGCAAAATCCAAAAGACAAATAAACCAATGGCCAATAATAA
ACGGAGAATCAGTGGGATTCATCATGGTAATTATACAGGGATTTGGATTTGAGCAGCTTGGACTGTTGGCTG
TGGGGCTATATCAAACAG

OSC-3 Fkh

TTTTGATAGCCACGATGAGGGAGGGCAACAGACGAGAAGGCCACAGGCGCTTAAATATTTGGGTAAAAGA
GAGTTGGGAATCAGGCCGGAATTGGAACCTCACTTTGGCCGCCGATGGGAGTTTGCAGAATTCAAAAATTT
CACACCATCTCACAACCTGTTTGTTTTTGCGCTGGAATGCAGTTTGAATGTTATTGCAGAGTGAATGTTTTTC
TGACTCACACTCACAATCCCAGCCTCATTTAAACGAAGCCAGAGTTTCGTGCCATCATCTTCTGCAGTCGT
TTGATTCCCGATTATCGTGAATGCCTGCGAGCAAAATCCAAAAGACAAATAAACCAATGGCCAATAATAA
ACGGAGAATCAGTGGGATTCATCATGGTAATTATACAGGGATTTGGATTTGAGCAGCTTGGACTGTTGGCTG
TGGGGCTATATCAAACAG

BA-1 wt

ACGTATCAGCGTTGGCGAATAAAAAAGTGCAGTGCATACCCGGCGGACTTATCTAATAGTTTTCACTTTCAA
TGTCGAGCGAAAGCTCTGCTGCCCCTGACGCGGTGCTGCCTCGCTTATCGCGATTCGCAATTCGCAATTCG
CAATTCTCAATTCGCATTTACAGCGGAGCGGGACGGAAACAGGAGAGAGACAGAGAGAGAGGCGGCAAGAGA
GAGAGAGAGAGAGAGAGAGGGAGAGGTGAAATGCCCCGAAAGTTGTCTGTATTCAAATTATTTATTTTTTTT
TGGGATTTCTTTTTCTGCAAGATGTCCAGATGGGAACAGAGCGGAGCGACCGACCCCATCCCTGTTTACCCC
GCAACGCACCGTACCGTACAAGTGTCTTGGTATATTTCGCATTTGTTATAGTTTCGCTTTTTTTTCACTCCTCC
CTCCATTTGCGATGCGTTGGCTTTTTTCACTCCGTTTTT

BA-1 GA

ACGTATCAGCGTTGGCGAATAAAAAAGTGCAGTGCATACCCGGCGGACTTATCTAATAGTTTTCACTTTCAA
TGTCGAGCGAAAGCTCTGCTGCCCCTGACGCGGTGCTGCCTCGCTTATCGCGATTCGCAATTCGCAATTCG
CAATTCTCAATTCGCATTTACAGCGCTATTGGGACGGAAACAGTGCCTCCACAGCCTAGTAGGCGGCATGACG
ACAAGGCTGCTAACACGAGGGAGAGGTGAAATGCCCCGAAAGTTGTCTGTATTCAAATTATTTATTTTTTTT
TGGGATTTCTTTTTCTGCAAGATGTCCAGATGGGAACGCGTCGGAGCGACCGACCCCATCCCTGTTTACCCC
GCAACGCACCGTACCGTACAAGTGTCTTGGTATATTTCGCATTTGTTATAGTTTCGCTTTTTTTTCACTCCTCC
CTCCATTTGCGATGCGTTGGCTTTTTTCACTCCGTTTTT

BA-2 wt

CCATGCATATGCTTATCACAACGGAAGAACATCTTATTGCTGCTCTTTAAGGTTTTTTAAGGAATGAAGGT
TTGGCCATTGCCTAAAAATGGGCGAACTACAAATTATATTTATAAAATAAATAGTTTTTAATACAGTTTTAGGCC
AGAAAGGATAATGCAAAAGCAGTTCGATGGCTCCATGCTTCCGCTACAGTTACAAGTCCACCTTATCGGGCG
CCCACTGTGAGTACGACTCATGCACGCACACATGCGCAGCTCTCTGAGTCACTCGCGCAAACGAAAATGACA
CGTATGGGAGTTTCGCTCTCGCTCTGCGCCAGTGCCTGCATGTGTATGCGAGTTTATCAATGCTGTTGAACT
GCGGTTAACAAAAACAACATCGCACACGTATTTGGCTCATTGAAAACATTCTCACACACACTGGCACACGCA
TTTCAATTAAAGCCTTAAAAGGCCAGAAAAAAACCGCAGTTGAAGCACCCAAA

BA-2 GA

CCATGCATATGCTTATCACAACGGAAGAACATCTTATTGCTGCTCTTTAAGGTTTTTTAAGGAATGAAGGT
TTGGCCATTGCCTAAAAATGGGCGAACTACAAATTATATTTATAAAATAAATAGTTTTTAATACAGTTTTAGGCC
AGAAAGGATAATGCAAAAGCAGTTCGATGGCTCCATGCTTCCGCTACAGTTACAAGTCCACCTTATCGGGCG
CCCACTGTGAGTACGACTCATGCACGCACACATGCGCAGGGTGCGCCGATACTCGCGCAAACGAAAATGACA
CGTATGGGAGTTTCGATCGAGGCTTGCGCCAGTGCCTGCATGTGTATGCGAGTTTATCAATGCTGTTGAACT
GCGGTTAACAAAAACAACATCGCACACGTATTTGGCTCATTGAAAACATTCTCACACACACTGGCACACGCA
TTTCAATTAAAGCCTTAAAAGGCCAGAAAAAAACCGCAGTTGAAGCACCCAAA

BA-2 AP-1

CCATGCATATGCTTATCACAACGGAAAGAACATCTTATTGCTGCTCTTTAAGGTTTTTTAAGGAATGAAGGT
TTGGCCATTGCCTAAATGGGCGAACTACAAATTATATTTATAAAATAAATAGTTTTAATACAGTTTTAGGCC
AGAAAGGATAATGCAAAAGCAGTTCGATGGCTCCATGCTTCCGCTACAGTTACAAGTCCACCTTATCGGGCG
CCCCTGTGAGTACAGCAAATGCACGCACACATGCGCAGCTCTCTATCCGACTCGCGCAAACGAAAATGACA
CGTATGGGAGTTTCGCTCTCGCTCTGCGCCAGTGCCTGCATGTGTATGCGAGTTTATCAATGCTGTTGAACT
GCGGTAAACAAAAACAACATCGCACACGTATTAGTCGAATTGAAAACATTCTCACACACACTGGCACACGCA
TTTCAATTAAAGCCTTAAAAGGCCAGAAAAAAACCGCAGTTGAAGCACCCAAA

BA-2 CA

CCATGCATATGCTTATCACAACGGAAAGAACATCTTATTGCTGCTCTTTAAGGTTTTTTAAGGAATGAAGGT
TTGGCCATTGCCTAAATGGGCGAACTACAAATTATATTTATAAAATAAATAGTTTTAATACAGTTTTAGGCC
AGAAAGGATAATGCAAAAGCAGTTCGATGGCTCCATGCTTCCGCTACAGTTACAAGTCCACCTTATCGGGCG
CCCCTATTGGTACGACTCATGCGGTCCCTCGATGCGCAGCTCTCTGAGTCACTCGCGCAAACGAAAATGACA
CGTATGGGAGTTTCGCTCTCGCTCTGCGCCATCTATTACATAGGTATGCGAGTTTATCAATGCTGTTGAACT
GCGGTAAACAAAAACAACATTACGAAGGTATTTGGCTCATTGAAAACATTCTAAGAATCCTGGCGAAGCCG
TTTCAATTAAAGCCTTAAAAGGCCAGAAAAAAACCGCAGTTGAAGCACCCAAA

BA-2 GATA

CCATGCATATGCACCTAACAACGGAAAGAACATCTTATTGCTGCTCTTTAAGGTTTTTTAAGGAATGAAGGT
TTGGCCATTGCCTAAATGGGCGAACTACAAATTATATTTATAAAATAAATAGTTTTAATACAGTTTTAGGCC
AGAAAGGCTTATGCAAAAGCAGTTCGATGGCTCCATGCTTCCGCTACAGTTACAAGTCCACCTTAGTGGGCG
CCCCTGTGAGTACGACTCATGACGCACACATGCGCAGCTCTCTGAGTCACTCGCGCAAACGAAAATGACA
CGTATGGGAGTTTCGCTCTCGCTCTGCGCCAGTGCCTGCATGTGTATGCGAGTTTATCAATGCTGTTGAACT
GCGGTAAACAAAAACAACATCGCACACGTATTTGGCTCATTGAAAACATTCTCACACACACTGGCACACGCA
TTTCAATTAAAGCCTTAAAAGGCCAGAAAAAAACCGCAGTTGAAGCACCCAAA

BA-3 wt

AATTTGCATTGACTGCTGCCACAACATAACTGCACTTTGCACTCACACACACTAAGATATGTTTTTCAACGC
AAAGCATTCGGGGCGGGTTTTTAATATTGCTATTGAGCTACAGGCAGAGTAATCCGAAATATTAAGAATAAT
AATACTACACACCAGACGCGGAGCTCCCGACTCGACGGCAGACAAAGACGCAGTGTGGTAAAATGTGCCAG
ACCAGACGGGGCGCTCCTCTCTCTCACACTCTCTTTCTCTCTCTCGTTTTCCGCTGCGAGTGTAGATATATGG
ATGGAAAATCGAAACTCTGTGGCTGTGGCCAGCCCCCGCGGCCCCCTTTTGCTATCTCGCTTGCTCGCATGG
ACAAAATCAACACAAGTTCACACACATACAGGCACACGGATGTGAACTCACAATGACAACCACTTCGTCACC
AGCAGATAAAAAAGTGCTGGCCTTAAATCAATTGCG

BA-3 GA

AATTTGCATTGACTGCTGCCACAACATAACTGCACTTTGCACTCACACACACTAAGATATGTTTTTCAACGC
AAAGCATTCGGGGCGGGTTTTTAATATTGCTATTGAGCTACAGGCAGAGTAATCCGAAATATTAAGAATAAT
AATACTACACACCAGACGCGGAGCTCCCGACTCGACGGCAGACAAAGACGCAGTGTGGTAAAATGTGCCAG
ACCAGACGGGGCGCTCATAGCCACCACACTATGATTCCCTGATCAGTTTCCGCTGCGAGTGTAGATATATGG
ATGGAAAATCGAAACTCTGTGGCTGTGGCCAGCCCCCGCGGCCCCCTTTTAGGATGCCGCTTGCTCGCATGG
ACAAAATCAACACAAGTTCACACACATACAGGCACACGGATGTGAACTCACAATGACAACCACTTCGTCACC
AGCAGATAAAAAAGTGCTGGCCTTAAATCAATTGCG

BA-3 CA

AATTTGCATTGACTGCTGCCACAACATAACTGCACTTTGTGATTCTGTATTCTAAGATATGTTTTTCAACGC
AAAGCATTCGGGGCGGGTTTTTAATATTGCTATTGAGCTACAGGCAGAGTAATCCGAAATATTAAGAATAAT
AATACTAGGTATCAGACGCGGAGCTCCCGACTCGACGGCAGACAAAGACGCAACCGAGTAAAATGTGCCAG
ACCAGACGGGGCGCTCCTCTCTCTTGATCTCTCTTTCTCTCTCTCGTTTTCCGCTGCGAGTGTAGATATATGG
ATGGAAAATCGAAACTCTTTGGCTTCGGCCAGCCCCCGCGGCCCCCTTTTGCTATCTCGCTTGCTCGCATGG
ACAAAATCAACACAAGTTCACCTATAGCATGTTCAACGGATGTGAACTCACAATGACAACCACTTCGTCACC
AGCAGATAAAAAAGTGCTGGCCTTAAATCAATTGCG

B.bone

CCAACACAAATCCCAAGAAAGAACACAGTGTCGTGGAAAAAGGACTTGGATACCGCCGGCCCGCTGAAGGTC
AACTAATTTTCGTGTTATATTTGGGACAGGGAACCCCTTCTCCTTTTTTCGCTTTTTCGAATCGATTCCCCTAAAC
TGACCGCCAGGTTCAATCAACCTTTGGCCCGAAAAGTACACTTCAGAAGCAGAAGCCAAATAGATTTGTCAA
CCCATTGACACATCGCTTGAGACTTCCGGCAGCTGGCTCGCAGGCTCTTGACCGATTTCGCTTTACGTCATCC
ACAACATCTTAACCTAATCGAATGGGTTCATGGACTCGATCTAATTGTCACTTAAGGCCACTAAAAAT
AGGCTCAAAAACCACAGAGGTTTCGAGTTCAACGCTCAACGAGTTGGCAGCATTTCATGCACTCGCACAAAG
ATTCATATTCCTGCACAGTCAGCGGATGAAGGCGGAGCCACCAGG

synEnh - BA-2

CCAACACAAATCCCAAGAAAGAACACAGTGTCGTGGAAAAAGGACTTGGATACCGCCGGCCCGCTGAAGGTC
AACTAATTTTCGTGTTATATTTGGGACAGGGAACCCCTTCTCCTTTTTTCGCTTTTTCGAATCGATTCCCCTAAAC
TGACCGCCAGGTTCAATCAACCTTTGGCCCGAAAAGTACACTTCAGAAGCAGAAGCCAAATAGATTTGTCAA
CCCCTGTGAGTACGACTCATGCACGCACACATGCGCAGCTCTCTGAGTCACTCGTTTCGCTTTACGTCATCC
ACAACAGGAGTTTCGCTCTCGCTCTGCGCCAGTGCGTGCATGTGTATGTCACTTAAGGCCACTAAAAAT
AGGCTCAAAAACCACACATCGCACACGTATTTGGCTCATTGATTACATTCTCACACACACTGGCACACGCA
TTTCACTATTCCTGCACAGTCAGCGGATGAAGGCGGAGCCACCAGG

synEnh - BA-3

CCAACACAAATCCCAAGAAAGAACACAGTGTCGTGTTTGCCTCACACACACTAAGCGGCCCGCTGAAGGTC
AACTAATTTTCGTGTTATATTTGGGACAGGGAACCCCTTCTCCTTTTTTCGCTTTTTCGAATCGATTCCCCTAAAC
AATACTACACACCAGATCAACCTTTGGCCCGAAAAGTACACTTCAGAAGCAGTGTTGTAAGATTTGTCAA
CCCATACGGGGCGCTCCTCTCTCTCACACTCTCTTTCTCTCTCTCGTTTCCGCGATTTCGCTTTACGTCATCC
ACAACATCTTAACACTCTGTGGCTGTGGCCATGGACTCGATCCCCCTTTTGCTATCTCGCTCCACTAAAAAT
AGGCTCAAAAACCAAGTTTACACACATACAGGCACACGGAGAGTTGGCAGCATTTCATGCACTCGCACAAAG
ATTCATATTCCTGCACAGTCAGCGGATGAAGGCGGAGCCACCAGG

synEnh - BA-4

CCAACCTACTCACACACGCACGCACACGTGTCGTGGAAAAAGGACTTGGATACCGCCGGCCCGCTGAAGGTC
AACTAATTTTCGTGTTATATTTGGGACAGGGAACCCCTTCTCCTTTTTTCGCTTTTTCGAATCGATTCCCCTAAAC
TGACCGCCAGGTTCAATCAACCGCGCACACACACACGCACCTTCAGAAGCAGAAGCTGTACACACACACACA
TGATTTGACACATCGCTTGAGACTTCCGGCAGCTGGCTCGCAGGCTCTTGACCGATGGGGCTAAACAGCTGC
ACAACATCTTAACCTAATCGAATGGGTTCATGGACTCGATCTAATTGTCACTTAAGGCCACTAAAAAT
AGGCTCAAAAACCACAGAGGTTAACTGCTTATCGCAGCCGAGTTGGCAGCATTTCATGCACTCGCACAAAG
ATTCATATTCCTGCACAGTCGTGATGTGTGTTGTGAGCCACCAGG

synEnh -BA-5

CCAACAGGTTTCTCTCTCGCAGACACAGTGTCGTGGAAAACCTGCTCTCTCTCTCTCGAGCGCGCGATCGGT
GTGCGCTTTTCGTGTTATATTTGGGACAGGGAACCCCTTCTCCTTTTTTCGCTTTTTCGAATCGATTCCCCTAAAC
TGTTTGTGTTATCGCGTTGCCTTTGGCCCGAAAAGTACACTTCAGAAGCAGAAGCCAAATAGATTTGTCAA
CCCATTGACACATCGCTTGAGACTTCCGGCAGCTGGCTCGCAGGCTCTTGACCGATTTCGCTTTACGTCGTGTT
GGTGTGCGTGCGCGTGTTCGAATGGGTTCATTTGTGTGTGAGTGTGTGTATGGTTAAGGCCACTAAAAAT
AGGCTCAAAAACCACAGAGGTTTCGAGTTCAACGCTCAACGAGTTGGCAGCATTTCATGCACTCGCACAAAG
ATTCATATTCCTGCACAGTCAGCGGATGAAGGCGGAGCCACCAGG

HsEnh-1

AAGTCAGGTTGGAGGCGGCCTTTAAGCGCCGCGCTCTCTCGGAACCTTAAATATTACCCGGAATCGTCTCTG
TGAATTGGAAAGCCCATTGAAGCCTCAGAGAAGCCGCGCCATTGCGGAGGCACGTAACGTCAAGCAGGGACT
TCCTGCGTCTTCGCCGCGGTGGCCATTTGGCTTGTGCGGACCATCCACCTATGCTGCGTCTGTACCTCTTT
GTCAGGACCGAGAGCGCGCGAGAGGAGTGTTGTGCCATTGCACACAGGCGGATCTGAGGCTTGGCGACGC
CGCTGGTTCGCGACCTGGGACAGGACGAGCACAGGAACCGCCTGGCCTGCGTCCCCGCGCGCATGAAGTCG
TGCGGCGCGCGCTCAAGTGAGCCCTGCGACCTCCAGGCCTCACCCGCCATCCCCACACAGACCCTGAGACC
CCTGACTGTGAGGGGCGCTTCCTCGCGTC

HsEnh-1 GA

AAGTCAGGTTGGAGGCGGCCTTTAAGCGCCGCGGGACTTAGGAACTTAAATATTACCCGGAATCGTCACCG
TGAATTGGAAAGCCCATTGAAGCCTCATCTTAGCCGCGCCATTGCGGAGGCACGTAACGTCAAGCAGGGACT
TCCTGCGTCCCTCGCCGCGGTGGCCATTTTGGCTTGTCGGGACCATCCACCTATGCTGCGTCTGTACCTCTTT
GTCAGGACCGAGAGCGCGCATGCCGAGTGGTTGTGCCCATTCACACAGGCGGATCTGAGGCTTGCGCAGCGC
CGCTGGTCGCGACCTGGGACAGGACGAGCACAGGAACCGCCTGGCCTGCGTCCCCGCGCGCATGAAGTCG
TGGCGGCGGCCGCTCAAGTGAGCCCTGCGACCTCCAGGCCTCACCCGCCATCCCCACACAGACCCTGAGACC
CCTGACTGTGAGGGGCGCTTCCTCGCGTC

HsEnh-1 CA

AAGTCAGGTTGGAGGCGGCCTTTAAGCGCCGCGCTCTCTCGGAACTTAAATATTACCCGGAATCGTCTCTG
TGAATTGGAAAGCCCATTGAAGCCTCAGAGAAGCCGCGCCATTGCGGAGGCACGTAACGTCAAGCAGGGACT
TCCTGCGTCCCTCGCCGCGGTGGCCATTTTGGCTTGTCGGGACCATCCACCTATGCTGCGTCTGTACCTCTTT
GTCAGGACCGAGAGCGCGGAGAGGAGTGGTTGTGCCCATTTGAATACCGGCGGATCTGAGGCTTGCGCAGCGC
CGCTGGTCGCGACCTGGGACAGGACGAGCACAGGAACCGCCTGGCCTGCGTCCCCGCGCGCATGAAGTCG
TGGCGGCGGCCGCTCAAGTGAGCCCTGCGACCTCCAGGCCTCACCCGCCATCCCTCTTTAGGCCCTGAGACC
CCTGACTGTGAGGGGCGCTTCCTCGCGTC

HsEnh-1 CG

AAGTCAGGTTGGAGGCGGCCTTTAAGCGCATTAAATCTCTCGGAACTTAAATATTACCCGGAATCGTCTCTG
TGAATTGGAAAGCCCATTGAAGCCTCAGAGAAGCATGTACATTGCGGAGGCACGTAACGTCAAGCAGGGACT
TCCTGCGTCCCTCGCCGCGGTGGCCATTTTGGCTTGTCGGGACCATCCACCTATGCTGCGTCTGTACCTCTTT
GTCAGGACCGAGAATCTTTGAGAGGAGTGGTTGTGCCCATTCACACAGGCGGATCTGAGGCTTGCGCAGCGC
CGCTGGTCGCGACCTGGGACAGGACGAGCACAGGAACCGCCTGGCCTGCGTCCCCGCGCTAGATATGAAGTCG
TGGCGGCGGCCGCTCAAGTGAGCCCTGCGACCTCCAGGCCTCACCCGCCATCCCCACACAGACCCTGAGACC
CCTGACTGTGAGGGGCGCTTCCTCGCGTC

HsEnh-2

TCCATCACTGCCTTCAAGTATGATAAAAATTGCTAATTTTCATGGGGTATGTCTAAGCTCTTTTCTTTGGTTTC
TCTGAGGTACTTTTTTAAATGTTTATTGGGGCTATCTTAAACTATCTTCCTTTGATTCTTCTGAAGTTGCAT
TTTAAAGTTTGTCTCCAGTGAATCATGCTTTCTTCATTTCTGTTTCAGCTTCTCTTCTGCCGTCTGCTC
CATAAATGGGGCCCTCTGCCAAGTTTCTGTTCCCTTTCTATGTTTTCTCTCTTATACATTGTATATACCTCT
ATGGCTTTAAATGTTGCTACTAAGTTAAAGATTTCTCAATTCATTTATTTGGCAGCATATCTCTCTTGGTTC
CAGTTAAAAATTTAGCTCCCCATCCGTAAGCTTTACATGTGCATCCTATGTAATATTTCTGGCTTCCAGCCT
GTTCTGATCATGTTGCACTTGGAGAATACCTTTTCTTGGTGA

HsEnh-2 GA

TCCATCACTGCCTTCAAGTATGATAAAAATTGCTAATTTTCATGGGGTATGTCTAAGCTCTTTTCTTTGGTTTT
GCAGAGGTACTTTTTTAAATGTTTATTGGGGCTATCTTAAACTATCTTCCTTTGATTCTTCTGAAGTTGCAT
TTTAAAGTTTGTCTCCAGTGAATCATGCTTTCTTCATTTCTGTTTCAGCTTCTCTTCTGCCGTCTGCTC
CATAAATGGGGCCCTCTGCCAAGTTTCTGTTCCCTTTCTATGTTTTTCATGGTATACATTGTATATACCTCT
ATGGCTTTAAATGTTGCTACTAAGTTAAAGATTTCTCAATTCATTTATTTGGCAGCATACATTTAATGGTTC
CAGTTAAAAATTTAGCTCCCCATCCGTAAGCTTTACATGTGCATCCTATGTAATATTTCTGGCTTCCAGCCT
GTTCTGATCATGTTGCACTTGGAGAATACCTTTTCTTGGTGA

HsEnh-3

GCCGCTCGCAACCCCGACTGTCGCCAGCGCCGCAACACCAGCGCTTCCCGCCTGGCTGAGTGGCCCCGGCTG
CGCGCGGGTTGCCATGGAGACGGTTCCGCCCTCTCGTGCGGACGCACTCAGGCGCGACCTCCGCCCCCTACGC
CGCCATGAGCGGAAAACGGGGAATGTGAGGCTGACGGCGCCATGTTTGAATTGGTTCGACGCGCTCCTGCAA
GACCTGGAAGAAAGAAAAGTAACGATTCTTCTCGGCCAGAGAGAGAAGCATGACCGGTTTTGCATACCCTGT
CCCCAGAAAAAGCACTCTAAGAGTAGGCGCGCTACTCCTGGGCGAGGAACTGCGGAAGTGAGGCATTGGTG
CCAGGTTCAAAGATGGCGCTGAGCAGCCAAGCGCAGAAGCGAAGAGAGGGCGGCAGCCTGCGGCCGTGGCCG
GCCCCGAGAGTCTGGGCCTGGGAGCGCAGATCTGGCTGAGCAGCTGATTCTTCCAGCAGATCCGAGA

HsEnh-3 GA

GCCGCTCGCAACCCCGACTGTCGCCAGCGCCGCAACACCAGCGCTTCCCGCCTGGCTGAGTGGCCCCGGCTG
CGCGCGGGTTGCCATGGAGACGGTTCCGCCGCATGGTGCGGACGCACTCAGGCGCGACCTCCGCCCCCTACGC
CGCCATGAGCGGAAAACGGGGAATGTGAGGCTGACGGCGCCATGTTTGAATTGGTCGCAGCGCCTCCTGCAA
GACCTGGAAGAAAGAAAAGTAACGATTCTTCTCGGCCATAAACTAGAGCATGACCGGTTTTGCATACCCTGT
CCCCAGAAAAAGCACTCTAAGAGTAGGCGCGCTACTCCTGGGCGAGGAAACTGCGGAAGTGAGGCATTGGTG
CCAGGTTCAAAGATGGCGCTGAGCAGCCAAGCGCAGAAGCGAGAATTTGGCGGCAGCCTGCGGCCGTGGCCG
GCCCCGCGAGGTCTGGGCCTGGGAGCGCAGATCTGGCTGAGCAGCTGATTCTTCCAGCAGATCCGAGA

HsEnh-3 CG

GCCGCTCGCAACCCCGACTGTCGCCAGCGCCGCAACACCAGCGCTTCCCGCCTGGCTGAGTGGCCCCGGCTT
CAGGTGGGTTGCCATGGAGACGGTTCCGCCCTCTCGTGCGGACGCACTCAGGTCAGACCTCCGCCCCCTACGC
CGCCATGAGCGGAAAACGGGGAATGTGAGGCTGACGGCGCCATGTTTGAATTGGTCGCAGCGCCTCCTGCAA
GACCTGGAAGAAAGAAAAGTAACGATTCTTCTCGGCCAGAGAGAGAAGCATGACCGGTTTTGCATACCCTGT
CCCCAGAAAAAGCACTCTAAGAGTAGGTCCTATACTCCTGGGCGAGGAAACTGCGGAAGTGAGGCATTGGTG
CCAGGTTCAAAGATGGCGCTGAGCAGCCAAGCGCAGAAGCGAAGAGAGGGCGGCAGCCTGCGGCCGTGGCCG
GCCCCGCGAGGTCTGGGCCTGGGAGCGCAGATCTGGCTGAGCAGCTGATTCTTCCAGCAGATCCGAGA

References

- Kheradpour P, Stark A, Roy S, Kellis M. 2007. Reliable prediction of regulator targets using 12 *Drosophila* genomes. *Genome Res* **17**: 1919–1931.
- Ui K, Nishihara S, Sakuma M, Togashi S, Ueda R, Miyata Y, Miyake T. 1994. Newly established cell lines from *Drosophila* larval CNS express neural specific characteristics. *In vitro cellular & developmental biology Animal* **30**: 209–216.
- Yanez-Cuna JO, Dinh HQ, Kvon EZ, Shlyueva D, Stark A. 2012. Uncovering cis-regulatory sequence requirements for context-specific transcription factor binding. *Genome Res* **22**: 2018–2030.