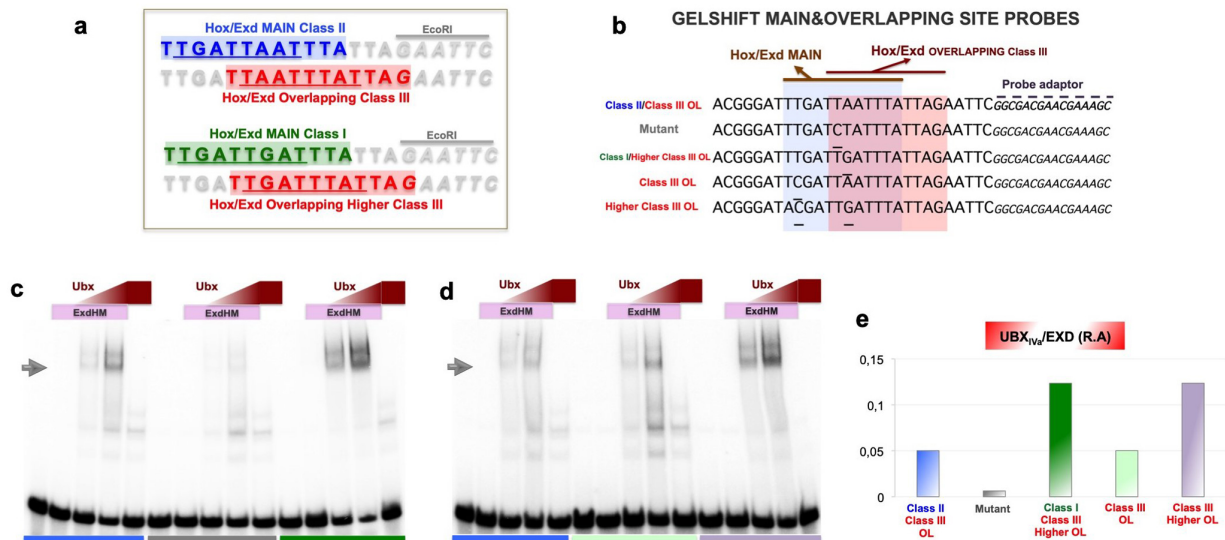


Supplementary Information

***In vivo* Hox binding specificity revealed
by systematic changes to a single cis
regulatory module**

Carlos Sánchez-Higuera et al.



Supplementary Figure 1: Biochemical and *in silico* predicted site affinity to main and overlapping class 3 sites or to the overlapping site

(a) Sequence comparison of the oligo containing the *main* and *overlapping* wild type sites with that carrying the class 1 mutated site (green) that secondarily generates a higher affinity class 3 site. (b) oligo probes used to test in EMSA the affinity to probes containing variants for both sites or only containing the *overlapping* site after mutation of the main site. (c) Ubx binding to both the *main* class 2 site and *overlapping* class 3 containing oligo (blue lanes); to a mutant for both sites (grey); to a *main* class 1 site that transformed the *overlapping* site into a High class 3 (green). (d) Ubx binding to both the *main* class 2 site and *overlapping* class 3 containing oligo (blue); to a mutant *main* with a class 3 *overlapping* site (light green); or to a mutant *main* with a high affinity class 3 *overlapping* site (light purple). (e) *In silico* predicted affinity of Ubx to sites shown in (c-d).

Comparison of blue lanes with light green labelled lanes in d shows Ubx binding is maintained or even increased in the absence of the *main* site, indicating Ubx binding to the oligos containing two sites is mostly occurring through the *overlapping* class 3 site. This conclusion is reinforced by the strong binding of Ubx observed in the light purple labelled lanes where only the *overlapping* site remains. These EMSA results indicate that the trunk expression observed in *S1+55cl1* embryos is mediated through the *overlapping* site. Source data are provided as a Source Data file.

a

Dm-Dfd
Al-Hox4

Sequence alignment table for Dm-Dfd Al-Hox4. Columns include accession numbers (P07548, B8RCB0), gene names (DFD_DROME, B8RCB0_BRALA), positions (1-279), and scores (58, 20, 118, 20, 178, 50, 238, 65, 297, 110, 356, 165, 414, 226, 476, 268, 536, 278, 586, 278).

Deformed A-Hox4 YPWM YPWM Consensus YPWM W motif



b

Dm-Labial
Al-Hox4

Sequence alignment table for Dm-Labial Al-Hox4. Columns include accession numbers (B8CA9, P10105), gene names (B8CA9_BRALA, LAB_DROME), positions (1-597), and scores (0, 60, 120, 0, 180, 0, 240, 24, 300, 84, 328, 144, 358, 195, 416, 195, 476, 234, 536, 596, 288, 635).

Labial A-Hox4 YKWM YDWM Consensus YKWM W motif



c

Dm-Ubx
Al-Hox7

Sequence alignment table for Dm-Ubx Al-Hox7. Columns include accession numbers (P83949, B8CB5), gene names (UBX_DROME, B8CB5_BRALA), positions (1-212), and scores (60, 12, 113, 64, 173, 93, 233, 111, 288, 151, 348, 211, 389, 236).

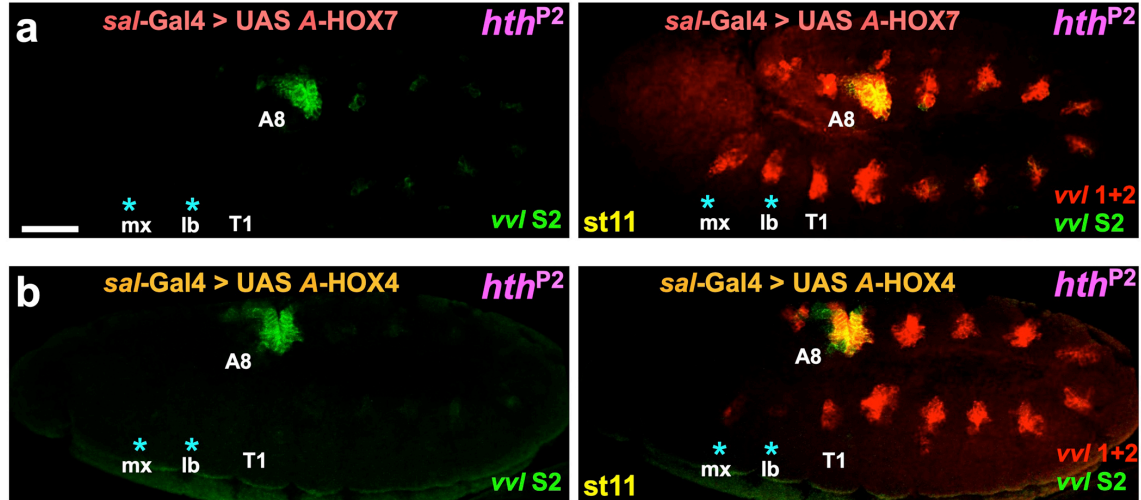
Ubx A-Hox7 YPWM YPWM Consensus YPWM W motif



Supplementary Figure 2: Sequence alignment of *D. melanogaster* and *A. lanceolatus* orthologous proteins

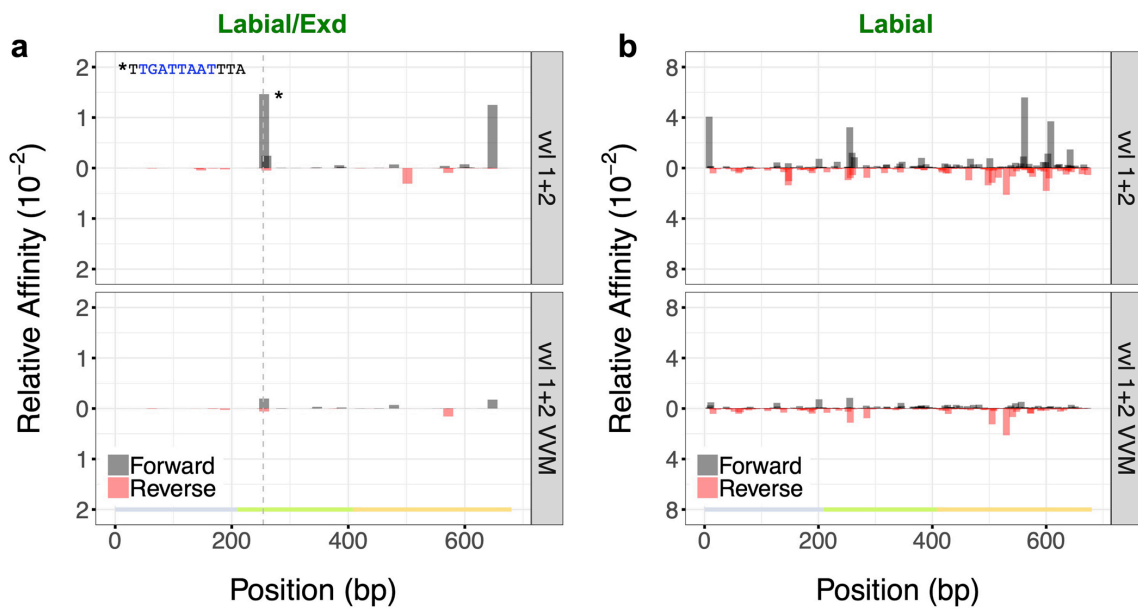
Full sequence Uniprot alignments of (a) Dfd and AI-Hox4, (b) Lab and AI-Hox1 and (c) Ubx and AI-Hox7. In all cases most conservation maps to the homeodomain especially on the DNA-contacting third helix and on the cofactor interacting hexapeptide motif. Specific alignments of the hexapeptide motif (also known as W motif) and the homeodomain performed with Jalview are shown under the full sequence alignment

Note that the most similar cofactor interacting motives (labelled in orange) between UbxIVa and AI-Hox7 do not align when using either Uniprot or Clustal w software with default settings. UbxIVa contains 4 described W motifs ³⁵.



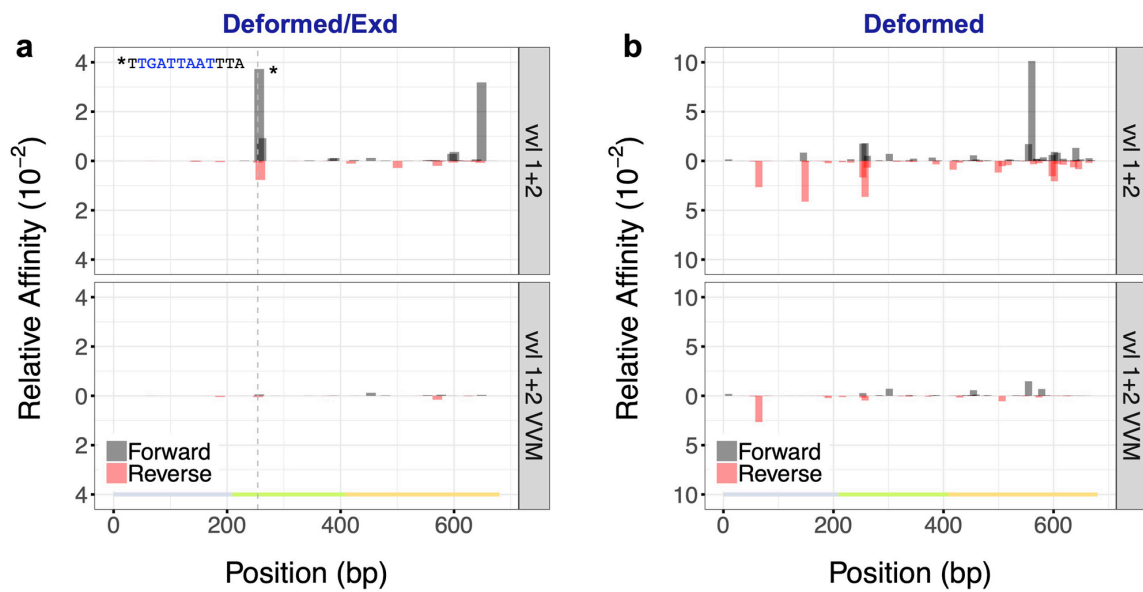
Supplementary Figure 3: *Amphioxus* Hox monomeric function in *Drosophila*

Expression of *S2-GFP* (a-b) or *vvl1+2-mCherry* (red) and *S2-GFP* (a-b, right panels) in *hth^{P2}* mutant embryos that express in the maxilla and labium (asterisks) the *UAS-AHox7* (a) or *UAS-AHox4* (b) with *sal-Gal4*. *AHox-7* can rescue *vvl1+2* but not *S2* expression, while *AHox-4* cannot rescue any of them. We did not test the rescue capacity of *AHox1* in *hth^{P2}* embryos as *AHox1* is unable to rescue any of these constructs in a wild type embryo (see Fig. 4I). Scalebar 50 μ m.



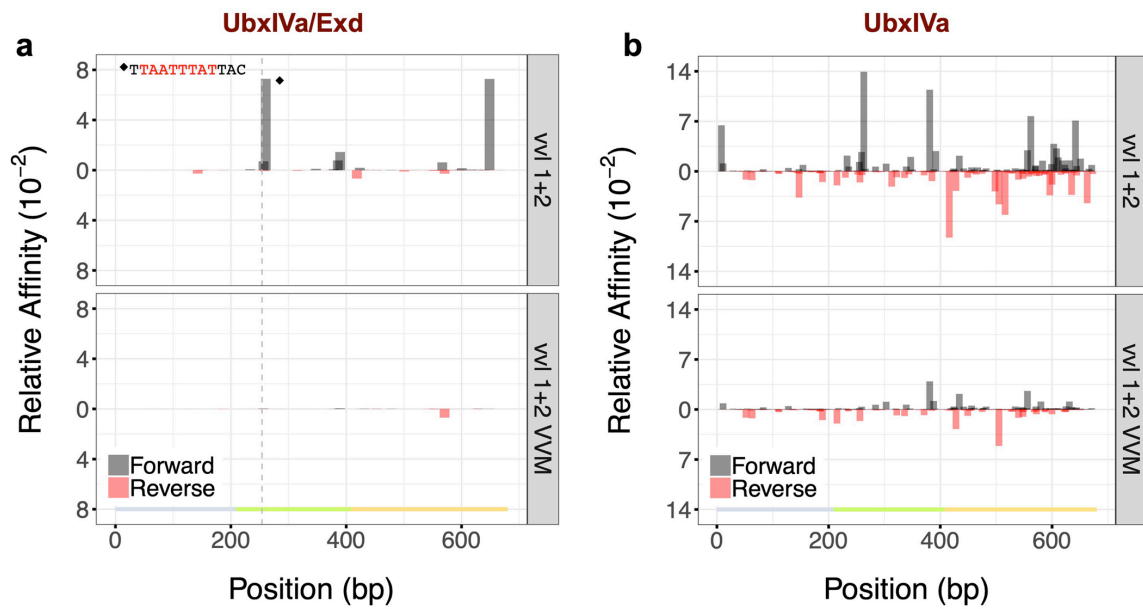
Supplementary Figure 4: Predicted site affinity of Lab to *vv1+2* and *vv1+2vvm*

NRLB predicted relative affinity of Lab-Exd (a) or of monomer Lab (b) to *vv1+2* or to the remaining Hox binding sites in the *vv1+2vvm*. Asterisk and discontinuous vertical line in (a) label the position of the *main* site.



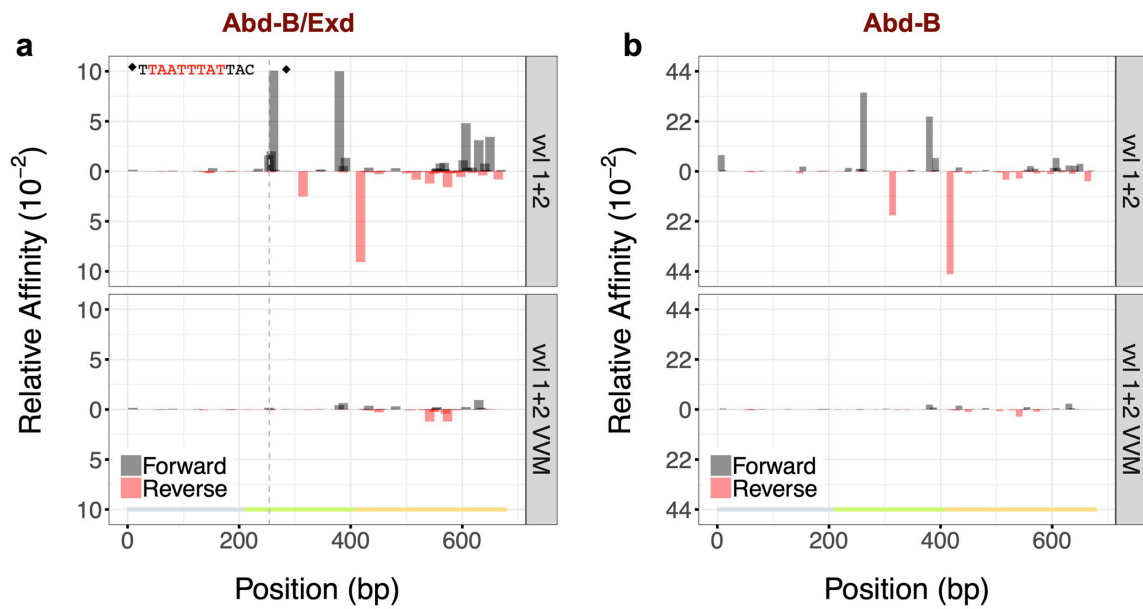
Supplementary Figure 5: Predicted site affinity of Dfd to *vv1+2* and *vv1+2vvm*

NRLB predicted relative affinity of Dfd-Exd (a) or of monomer Dfd (b) to *vv1+2* or to the remaining Hox binding sites in the *vv1+2vvm*. Asterisk and discontinuous vertical line in (a) label the position of the *main* site.



Supplementary Figure 6: Predicted site affinity of Ubx IVa to *vv1+2* and *vv1+2vvm*

NRLB predicted relative affinity of Ubx IVa-Exd (a) or of monomer Ubx IVa (b) to *vv1+2* or to the remaining Hox binding sites in the *vv1+2vvm*. Diamond and discontinuous vertical line in (a) label the position of the *overlapping* site.



Supplementary Figure 7: Predicted site affinity of Abd-B to *vl1+2* and *vl1+2vvm*

NRLB predicted relative affinity of AbdB-Exd (a) or of monomer AbdB (b) to *vl1+2* or to the remaining Hox binding sites in the *vl1+2vvm*. Diamond and discontinuous vertical line in (a) label the position of the *overlapping* site.

<i>vv1</i> 1+2	GGTTAATGATGGCCACACAGGCGACAACCTGGCGACTGAGATCTTGTCCCGCAATTTCCCG ##	60
<i>vv1</i> 1+2 <i>vvm</i>	GGTTggTGATGGCCACACAGGCGACAACCTGGCGACTGAGATCTTGTCCCGCAATTTCCCG	60
<i>vv1</i> 1+2	ATCATTGCTCAGATACGATACGGATTCTGCGAAGTCACCGGGCAAAGGCATCCCACTGA 	120
<i>vv1</i> 1+2 <i>vvm</i>	ATCATTGCTCAGATACGATACGGATTCTGCGAAGTCACCGGGCAAAGGCATCCCACTGA	120
<i>vv1</i> 1+2	GTTTGTGGCTTCTTGTCAAACAAATCATTACTGCGGATACTCCTCGATATTCCCTCAAG 	180
<i>vv1</i> 1+2 <i>vvm</i>	GTTTGTGGCTTCTTGTCAAACAAAgCAgggACTGCGGATACTCCTCGATATTCCCTCAAG	180
<i>vv1</i> 1+2	ATACTATCCATTCTACTGTGTAGGTGTGAGCTGCAATTTCCCTGGAAAAATTACGTCCA 	240
<i>vv1</i> 1+2 <i>vvm</i>	ATACTATCCATTCTACTGTGTAGGTGTGAGCTGCAATTTCCCTGGAAAAAggACGTCCA	240
<i>vv1</i> 1+2	GCACGGGATTTGATTAATTTATTACCGCTTGTGCAAGGGAAAGTGATCTTCGGGTTCTTA 	300
<i>vv1</i> 1+2 <i>vvm</i>	GCACGGGATTTGATcAATTTcTTcCCGCTTGTGCAAGGGAAAGTGATCTTCGGGTTCTTA	300
<i>vv1</i> 1+2	ACGGTTGCGGATCGTAAAAACATTCGGCAGACACAATTTGTTGAATTATCTGCGGGCTGC 	360
<i>vv1</i> 1+2 <i>vvm</i>	ACGGTTGCGGAgCgGgAAAAACATTCGGCAGACACAATTTGTTGAAggATCTGCGGGCTGC	360
<i>vv1</i> 1+2	TGTTGTGGGCACTTTTTATGAGTTATTTATGTGCGACTGTGGCGCAACAGGATCATAA 	420
<i>vv1</i> 1+2 <i>vvm</i>	TGTTGTGGGCACTTTgTTATGgGTTATgTATGTGCGACTGTGGCGCAACAGGATCATgg	420
<i>vv1</i> 1+2	AATATGTAGTTATGGGTAAATCTGTGAAAAATAAATGTAAGCGTAATCTGGAAAAATTGA 	480
<i>vv1</i> 1+2 <i>vvm</i>	AATATGTAGTTATGGGTAAATCTGTGAAAAATAAATGTAAGCGTggTCTGGAAAAATTGA	480
<i>vv1</i> 1+2	GTGGCCTAGATTACCACATTAGTCATTTCAAAGAAATTAAGAAACTAAAGTAGAAGT 	540
<i>vv1</i> 1+2 <i>vvm</i>	GTGGCCTAGATTACCACAggAGTCATTTCAAAGAAAggAAAGAAACTAAAGTAGAAGT	540
<i>vv1</i> 1+2	ATAAAAAATTTAAATTGTAATGATTTAAATATATCAAATAATGCAAATTTAAAAAATTA 	600
<i>vv1</i> 1+2 <i>vvm</i>	ATAAAAAATTTAAATTGTggTGAgggAAATATATCAAATggTGCAAATTTAAAAAaggA	600
<i>vv1</i> 1+2	ATTTGATTACTGCTTTATTTAAATTTTATTTTCATATTAATATGATTTATCGTGCATACG 	660
<i>vv1</i> 1+2 <i>vvm</i>	ATTTGAggACTGCTTTAggAAAATTTTATTTTCATAggAATATGAgggATCGTGCATACG	660
<i>vv1</i> 1+2	GAAATTAACCTGGATTATGG 680 ##	680
<i>vv1</i> 1+2 <i>vvm</i>	GAAAggAAACTGGAggATGG 680	680

S1: [1 → 209]

S2: [210 → 418]


S3: [419 → 680]


vv1 1+2 genomic location: (Chr3L: 6,764,714-6,765,393)


Supplementary Figure 8: Aligned *vv1*+2 and *vv1*+2*vvm* sequences


The *vv1*+2 sequence aligned to the *vv1*+2*vvm* sequence indicating the mutated sites. The S1 sequence is shown in grey, S2 in green and S3 in brown. Genomic location coordinates for the *vv1*+2 enhancer are included.


REPORTER GENE	Intercalary Segment (ic)	Maxillary Segment (mx)	Labial Segment (lb)	Trunk Segments (T2-A7)	Eighth Abdominal Segment (A8)	MAIN&OL Hox/Exd sites	Global Hox Monomer Input per fragment
vv/ 1+2	☒	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	Main cl2 OL cl3	high
vv/ S1+S2	☒	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	Main cl2 OL cl3	high
vv/ S2	☒	✓✓✓	☒	✓✓✓✓	✓✓✓✓	Main cl2 OL cl3	high
vv/ S1+55	☒	✓✓	✓✓✓	☒	☒	Main cl2 OL cl3	weak
vv/ S1+55 cl1	✓✓	✓✓✓	✓✓✓✓	✓	☒	Main cl1 Higher OL cl3	weak
vv/ S1+55 cl3	☒	✓	✓✓✓	✓	✓	Main cl3	weak
vv/ S1+55 cl1OF	✓✓✓✓	✓✓✓	✓✓✓	☒	☒	Main cl1OF	weak
vv/ S1+55 cl2OF	☒	✓✓✓✓	✓✓✓✓	☒	☒	Main cl2OF	weak
vv/ S1+55 cl3OF	☒	✓✓✓	✓✓✓	✓✓	✓✓✓	Main cl3OF	weak
vv/ S1+55 mut	☒	☒	☒	☒	☒	none	weak
vv/ 1+2 ^{Main&OL mut}	☒	✓	✓	✓✓✓	✓✓✓✓	none	high
vv/ S2 ^{Main&OL mut}	☒	☒	☒	☒	✓✓✓✓	none	high
vv/ 1+2 vvm	☒	☒	☒	☒	☒	none	weak
vv/ 1+2 vvm-R	☒	✓✓✓	✓✓✓	✓✓✓✓	✓✓✓✓	Main cl2 OL cl3	weak
vv/ S1	☒	☒	☒	☒	☒	none	weak
vv/ S3	☒	☒	☒	☒	☒	none	high


 Highest optimal expression


 High expression


 Medium expression


 Low Expression

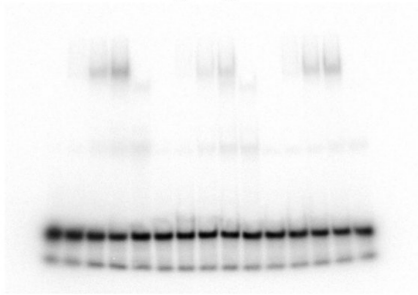

 Occasional not significant expression

Supplementary Table 1: Summary of the segmental expression in different reporter genes

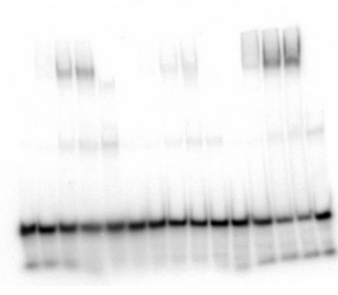
At a glance summary of the expression driven by the different reporters. The levels of expression in each segment are indicated using a subjective five degree scale: occasional not significant expression; low expression; medium expression; high expression; highest optimal expression.

The presence or absence of the *main* and *overlapping* sites and the binding class to which they belong, plus the abundance of monomer sites in each reporter are indicated in the last two columns.

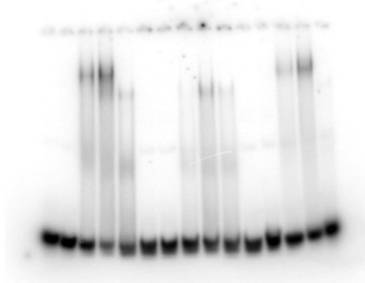
EMSA gel Figure 3i



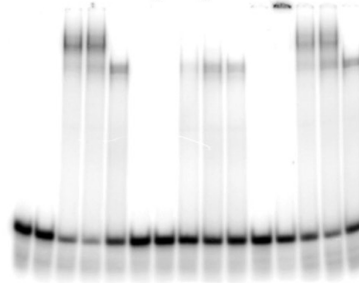
EMSA gel Figure 3j



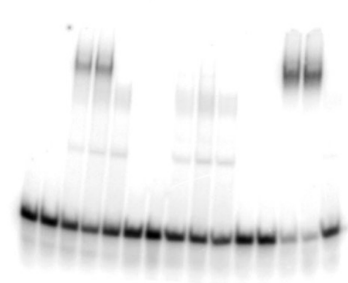
EMSA gel Figure 3l



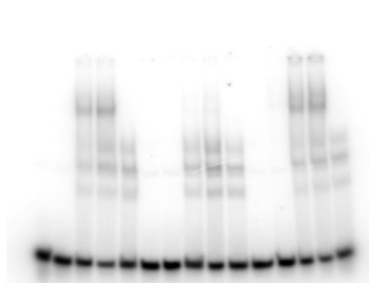
EMSA gel Figure 3m



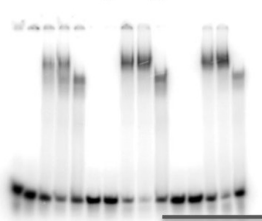
EMSA gel Figure 3o



EMSA gel Figure 3p

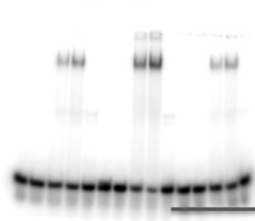


EMSA gel Figure 3r



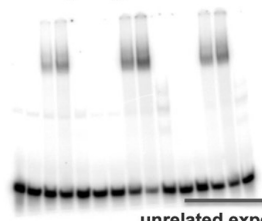
unrelated experiment

EMSA gel Figure 3t



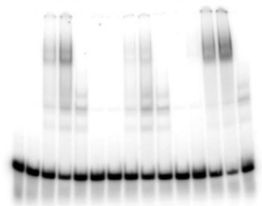
unrelated experiment

EMSA gel Figure 3v

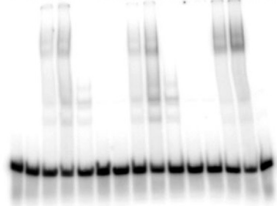


unrelated experiment

EMSA gel Supplementary Figure1c



EMSA gel Supplementary Figure1d



Unprocessed scans of EMSA experiments

Primer List

vv/ S1+55

Forward vv/ S1 EcoRI:

5'-AAAGAATTCGGTTAATGATGGCCACACAGGCGAC-3'

Rvs vv/ S1+55 EcoRI:

5'-AAAGAATTCTAATAAATTAATCAAATCCCGTGCTGGACGTAA-3'

vv/ S1+55 CI1

Forward vv/ S1 EcoRI:

5'-AAAGAATTCGGTTAATGATGGCCACACAGGCGAC-3'

Reverse vv/ S1+55 CI1 EcoRI:

5'-AAAGAATTCTAATAAATCAATCAAATCCCGTGCTGGACGTAA-3'

vv/ S1+55 CI3

Forward vv/ S1 EcoRI:

5'-AAAGAATTCGGTTAATGATGGCCACACAGGCGAC-3'

Reverse vv/ S1+55 CI3 EcoRI:

5'-AAAGAATTCTAATAAATAAATCAAATCCCGTGCTGGACGTAA-3'

vv/ S1+55 CI1 OF

Forward vv/ S1 EcoRI:

5'-AAAGAATTCGGTTAATGATGGCCACACAGGCGAC-3'

Reverse vv/ S1+55 CI1 OF EcoRI:

5'-AAAGAATTCTAAGTAATCAATCATATCCCGTGCTGGACGTAA-3'

vv/ S1+55 CI2 OF

Forward vv/ S1 EcoRI:

5'-AAAGAATTCGGTTAATGATGGCCACACAGGCGAC-3'

Reverse S1+55 CI2 OF:

5'- AAAGAATTCTAAGTCATTAATCATATCCCGTGCTGGACGTAA-3'

vv/ S1+55 CI3 OF

Forward vv/ S1 EcoRI:

5'-AAAGAATTCGGTTAATGATGGCCACACAGGCGAC-3'

Reverse vv/ S1+55 Cl3 OF EcoRI:

5'-AAAGAATTCTAAGTCATAAATCATATCCCGTGCTGGACGTAA-3'

vv/ S1+55 mutant

Forward vv/ S1 EcoRI:

5'-AAAGAATTCGGTTAATGATGGCCACACAGGCGAC-3'

Reverse vv/ S1+55 mutant EcoRI:

5'-AAAGAATTCTAATAAATTGATCAAATCCCGTGCTGGACGTAA-3'

vv/ S2

Forward vv/ S2 EcoRI :

5':AAAGAATTCGCTGCAATTTTCCCTGGAAAAATTACGTCCAGCACGGGATTTGATTAATTTATTA-3'

Rvs S2 EcoRI:

5'-AAAGAATTCATGATCCTGTTGGCGCCACAGTCGCACAT-3'

vv/ S2 MAIN&OL mutant

Forward vv/ S2 Main&OL mut EcoRI:

5'-AAAGAATTCGCTGCAATTTTCCCTGGAAAAATTACGTCCAGCACGGGATTTGATCAATTTATTA-3'

Rvs S2 EcoRI:

5'-AAAGAATTCATGATCCTGTTGGCGCCACAGTCGCACAT-3'

vv/ S1+S2

Forward vv/ S1 EcoRI:

5'-AAAGAATTCGGTTAATGATGGCCACACAGGCGAC-3'

Rvs S2 EcoRI:

5'-AAAGAATTCATGATCCTGTTGGCGCCACAGTCGCACAT-3'

vv/ 1+2 MAIN&OL mutant*

External Forward vv/ S1 EcoRI:

5'-AAAGAATTCGGTTAATGATGGCCACACAGGCGAC-3'

External Reverse vv/ S3 EcoRI:

5'-AAAGAATTCATATAATCCAGTTTAATTTCCGTATGCACGATAAA-3'

Forward vv/ S2 Main&OL mutant:

5'- GATCAATTTCTTCCCGCTTGTCTCGAAGGGAAAGTG-3'

Reverse vv/ S2 Main&OL mutant:

5'-CCTTCGACAAGCGGGAAGAAATTGATCAAATCCCGTGCTGG-3'

*** PCR mutagenesis in two steps (First Round PCRs and Second Round of Overlapping PCRs).**

vv/ 1+2 vvm-R*

External Forward vv/ S1 vvm EcoRI:

5'-AAAGAATTCGGTTGGTGATGGCCACACAGGCGACAA-3'

External Reverse vv/ S3 vvm EcoRI:

5'-AAAGAATTCATCCTCCAGTTTCCTTTCCGTATGCAC-3'

Fwd vv/ S2 Main&OL restored:

5'-AGCACGGGATTTGATTAATTTATTACCGCTTGTC-3'

Rvs vv/ S2 Main&OL restored:

5'-GACAAGCGGTAATAAATTAATCAAATCCCGTGCT-3'

*** PCR mutagenesis in two steps (First Round PCRs and Second Round of Overlapping PCRs)**