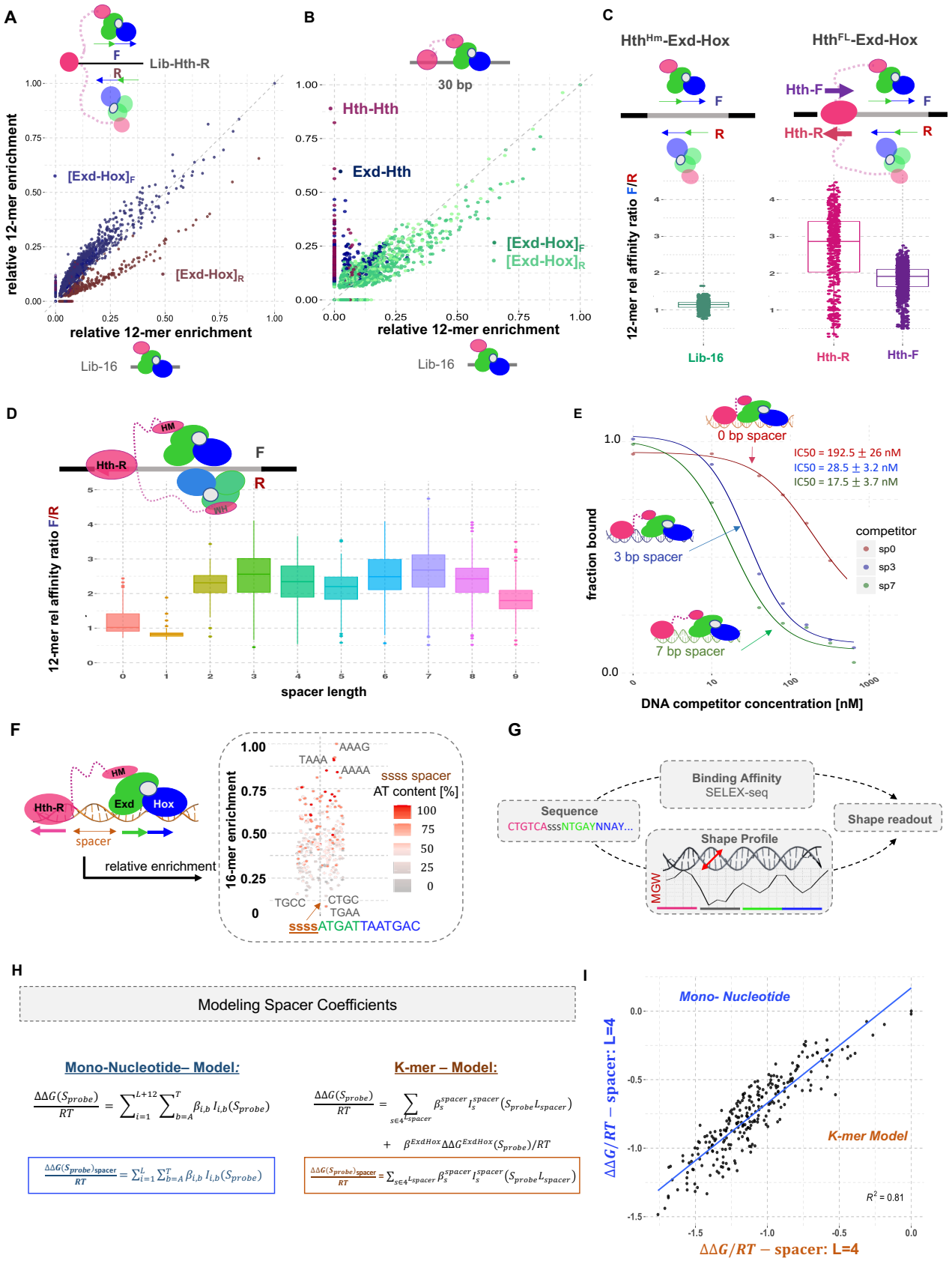

Figure S1. Binding Preferences and Model Comparison. (related to Figure 1)

(A) Comparison of Exd-Hox binding preferences in the presence or absence of a Hth site. Normalized relative 12-mer enrichments for the ternary Hth^{Fl}-Exd-Hox (Lib-Hth-R) are compared to those for Hth^{HM}-Exd-Hox (Lib-16). 12-mers are colored based on the most likely Exd-Hox binding orientation with respect to the to the upstream fixed library flank: $[Exd\text{-}Hox]_F$ with a higher enrichment score for the forward 12-mer (blue); and $[Exd\text{-}Hox]_R$ with a higher score for the reverse orientation (red). The fixed Hth site introduces a preference for the $[Exd\text{-}Hox]_F$ orientation. **(B)** 12-mer relative enrichments are compared between Lib-30 with trimeric Hth^{Fl}-Exd-Dfd and Lib-16 with Hth^{HM}-Exd-Dfd bound. Three main binding modes are recognized: i) Hth-dimer sequences (purple), ii) canonical Exd-Hth sites (dark blue) and ii) Exd-Hox sites (green; shade implies the orientation of the binding site with respect to the library sequencing flanks). **(C)** Effect size of orientation preferences introduced by the fixed Hth site in either Lib-Hth-F or Lib-Hth-R compared to Lib-16 Hth^{Hm}-Exd-Dfd. Binding to the forward or reverse $[Exd\text{-}Hox]_{(F/R)}$ motif occurs in equal ratios in the symmetric Lib-16 (Hth^{HM}-Exd-Hox), whereas both Hth-F and Hth-R libraries have relative enrichment ratios of $[Exd\text{-}Hox]_F$ over $[Exd\text{-}Hox]_R$ greater than one (average of ~ 3 for Lib-Hth-R, pink; of ~2 for Lib-Hth-F, purple). **(D)** Relative 12-mer enrichments by offset to fixed Hth site in Lib-HthR. Preference for $[Exd\text{-}Hox]_F$ varies with spacer length. **(E)** Competition assay validating the spacer preference of the trimeric Hth^{FL}-Exd-Hox complex: three different spacer sequences – 0, 3, and 7bp – in between the Hth and Exd-Hox binding sites (CTGTCA-$(N)_L$-ATGATTAATGAC) were used to compete with a radio-labeled Hth-Exd-Hox probe (CTGTCA-AAA-ATGATTAATGAC). Agreeing with the model (**Figure 2A**), spacers of 3 and 7 bp competed out the labeled probe at lower concentrations compared to a suboptimal spacer of 0 bp (IC50 values of $17.5 \pm 3.7; 28.5 \pm 3.2,$ and 192.5 ± 28 nM for spacers of 7, 3 and 0 bp respectively). **(F)** Sequence preferences of the DNA spacer between Hth and Exd-Hox. Relative enrichment for 16-mers with a fixed Exd-Hox site at positions 5-16 are shown and colored based on A/T content within the spacer (position 1-4). More enriched sequences tend to have a higher spacer A/T content. **(G)** Schematic for detecting shape-readout of DNA spacer sequence: For a specific spacer length, mononucleotide models are fit to the SELEX round 2 count data and the affinity score and the minor grove width (MGW) is computed for each sequence. Intrinsic, average DNA MGW profiles are correlated with TF binding selectivity to test for shape readout. **(H)** Spacer affinity scores were generated by either fitting a mono-nucleotide model (blue) or a k-mer model (orange) to the round-2 SELEX-seq count data. Either sum of coefficients (mono-nucleotide model), or the individual spacer coefficients were used to compute spacer affinity scores. **(I)** The 4-bp spacer $\Delta\Delta G/\mathrm{RT}$ coefficients were compared for either the mono-nucleotide model (blue) or the k-mer model (orange) ($R^2 = 0.81$).

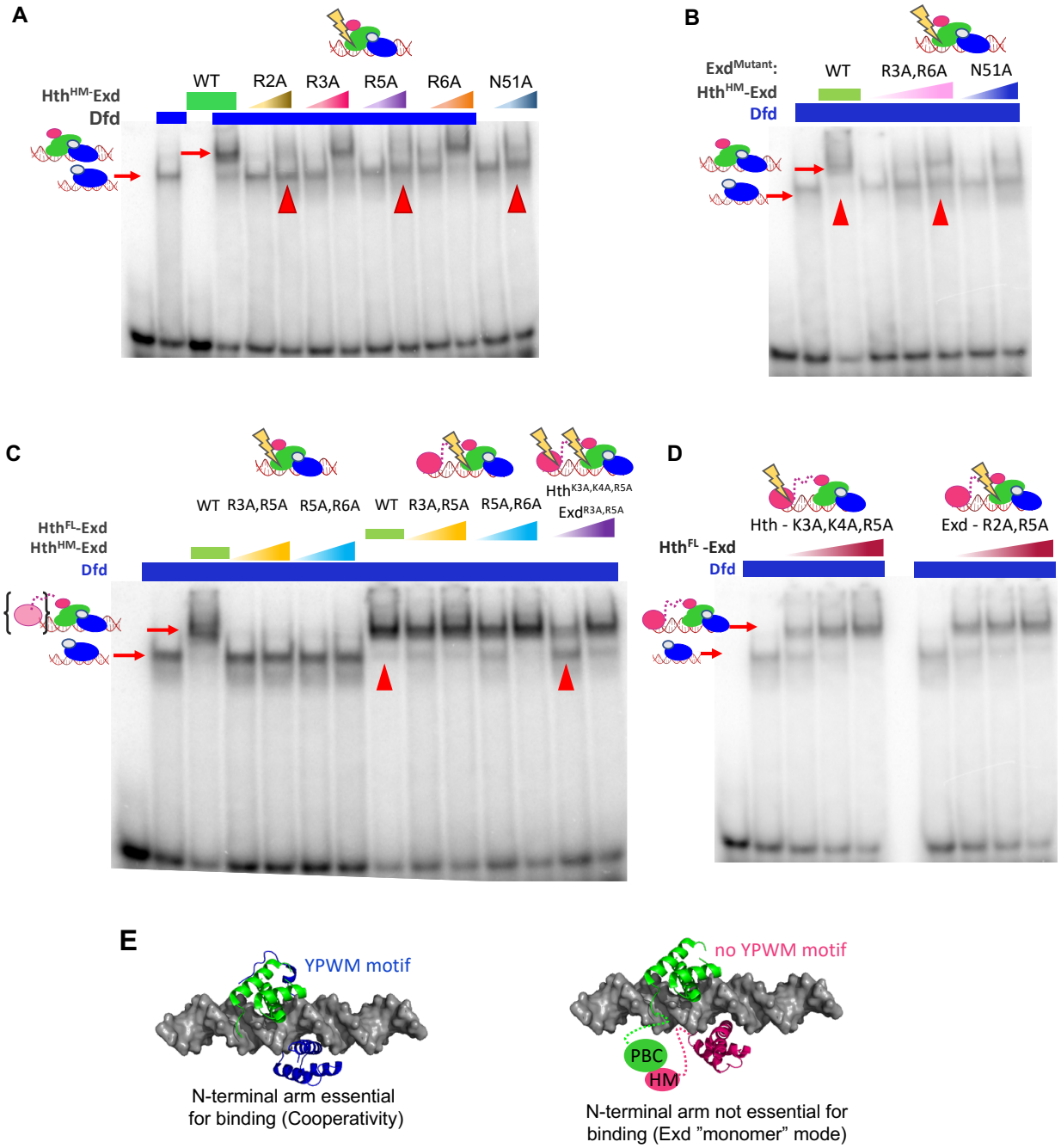

Figure S2. Binding behavior of Exd and Hth mutant proteins. (related to Figure 2 and 3) **(A)** Electromobility shift assay (EMSA) for single amino-acid point mutations of Exd in complex with Hth^{HM} and the Hox protein Dfd. Arrows indicate binding loss. The N51A hydrogen bond disrupting mutation (blue) is shown as a reference for severe binding loss. Lane 1: DNA only; lane 2: Dfd only, lane 3: Hth^{HM} only, lane 4: Hth^{HM}-Exd^{WT}-Dfd, lane 5-6:Hth^{HM}-Exd^{R2A} -Dfd, lane 7-8:Hth^{HM}-Exd^{R3A}-Dfd, lane 9-10:Hth^{HM}-Exd^{R5A}-Dfd, lane 11-12:Hth^{HM}-Exd^{R6A}-Dfd, lane 13-14: Hth^{HM}-Exd^{N51A}-Dfd. **(B)** Electromobility shift assay for Hth^{HM}-Exd-Dfd with Exd double mutant protein (R3A,R6A; containing the two arginines that individually do not cause a loss of binding

for the Hth^{HM}-Exd-Hox complex. Wild type Exd and the hydrogen-bond disrupting Exd^{N51A} proteins are used for comparison: Lane 1: DNA only; lane 2: Dfd only, lane 3: Hth^{HM}-Exd^{WT}-Dfd, lane 4-6: Hth^{HM}-$Exd^{R3A,R6A}$-Dfd, lane 7-8: Hth^{HM}-Exd^{N51A}-Dfd. **(C)** Electromobility shift assay for combinations of Exd N-terminal arginine mutants with either Hth^{HM} or Hth^{Fl} isoform or with Hth^{-shape} mutant protein. Lane 1: DNA only; lane 2: Dfd only, lane 3: Hth^{HM}-Exd-Dfd, lane 4-5: Hth^{HM}-$Exd^{(R3A,R5A)}$-Dfd, lane 6-7: Hth^{HM}-$Exd^{(R5A,R6A)}$-Dfd, lane 8: Hth^{FL}-Exd-Dfd; lane 9-10: Hth^{FL}-$Exd^{(R3A,R5A)}$-Dfd; lane 11-12: Hth^{FL}-$Exd^{(R5A, R6A)}$-Dfd, lane 13-14: $Hth^{K3A,K4A,R5A}$-$Exd^{R2A,R5A}$-Dfd. **(D)** Electromobility shift assay for Hth^{-shape} ($Hth^{K3A,K4A,R5A}$) and Exd^{-shape} ($Exd^{R2A,R5A}$) mutant proteins as used for the mutant Lib-Hth-(F/R) SELEX experiments. Lane 1: DNA only; lanes 2 and 7: Dfd-only; lane 3-5: Hth^{-shape}-Exd-Dfd; lane 8-10: Hth^{FL}-Exd^{-shape}-Dfd. **(E)** Visualizing binding mode differences by structural superimposition of either Exd-Hox (PBD-ID: 2R5Y; Exd=green, Hox=blue) or Exd and MEIS1 (PDB-ID: 4XRM; pink) onto B-DNA. In Exd-Hox the YPWM motif directly interacts with Exd's HD, potentially intensifying Exd's shape readout through cooperative binding behavior. Hth uses its HM-domain that is connected via a flexible linker to Exd's PBC domain (indicated here as colored circles) to connect to Exd. It is likely that the binding is DNA mediated thus allowing Exd to bind in its native, monomeric mode.

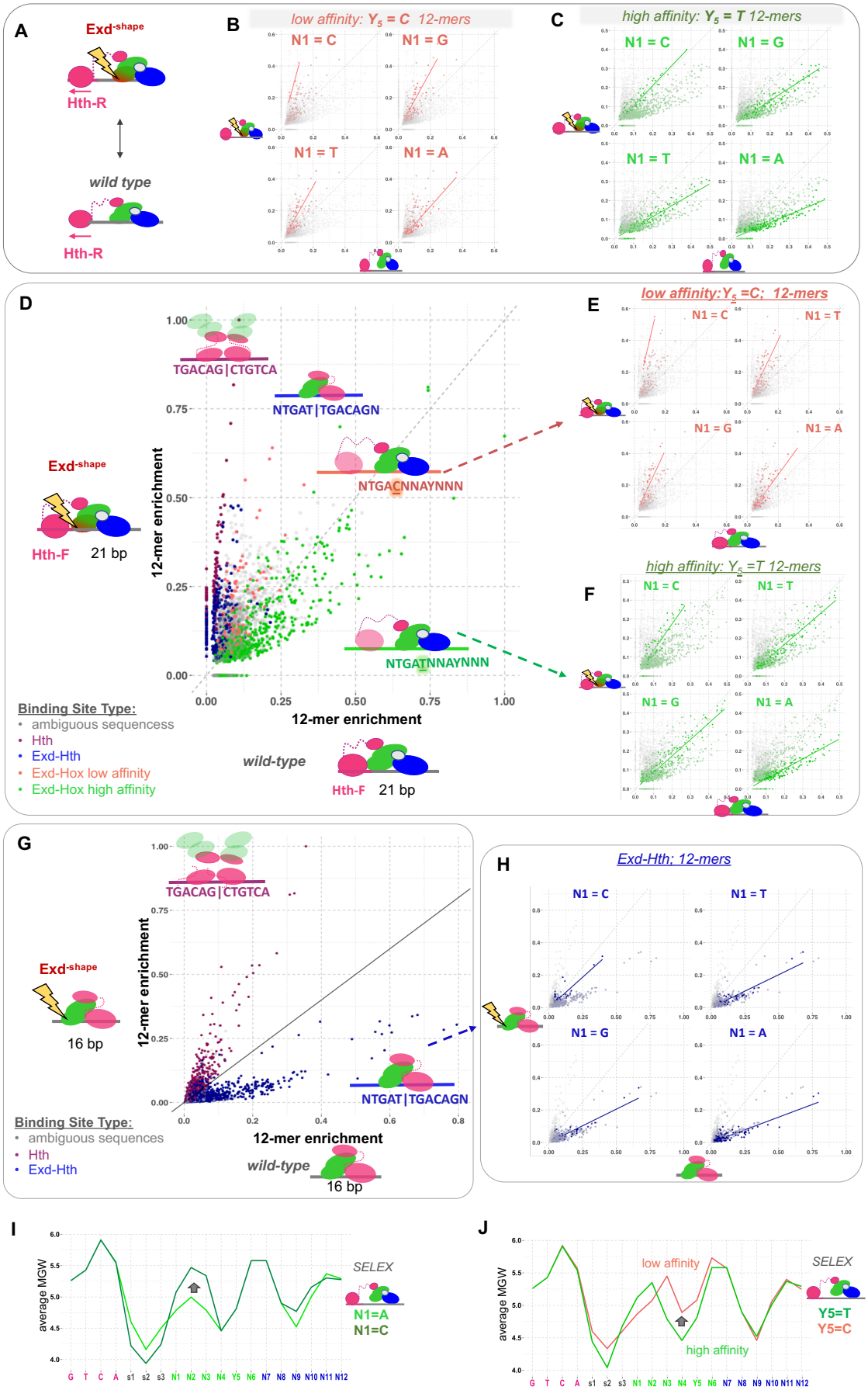
A
Exd-shape
Hth-R
wild type
Hth-R
B
low affinity: Y5 = C 12-mers
N1 = C
N1 = G
N1 = T
N1 = A
C
high affinity: Y5 = T 12-mers
N1 = C
N1 = G
N1 = T
N1 = A
D
TGACAG|CTGTCA
NTGAT|TGACAGN
NTGACNNAYNNN
NTGATNNAYNNN
12-mer enrichment
12-mer enrichment
Exd-shape
Hth-F 21 bp
Binding Site Type:
ambiguous sequencess
Hth
Exd-Hth
Exd-Hox low affinity
Exd-Hox high affinity
wild-type
Hth-F 21 bp
E
low affinity:Y5 =C; 12-mers
N1 = C
N1 = T
N1 = G
N1 = A
F
high affinity: Y5 =T 12-mers
N1 = C
N1 = T
N1 = G
N1 = A
G
TGACAG|CTGTCA
12-mer enrichment
Exd-shape
16 bp
NTGAT|TGACAGN
12-mer enrichment
Binding Site Type:
ambiguous sequences
Hth
Exd-Hth
wild-type
16 bp
H
Exd-Hth; 12-mers
N1 = C
N1 = T
N1 = G
N1 = A
I
average MGW
SELEX
N1=A
N1=C
G T C A s1 s2 s3 N1 N2 N3 N4 Y5 N6 N7 N8 N9 N10 N11 N12
J
average MGW
low affinity
high affinity
SELEX
Y5=T
Y5=C
G T C A s1 s2 s3 N1 N2 N3 N4 Y5 N6 N7 N8 N9 N10 N11 N12

Figure S3. Exd^{-shape} differentiates complex compositions. (related to Figure 4)
(A-C) Comparison of 12-mer relative sequence enrichments between Hth^{FL}-Exd-Hox and Hth^{FL}-Exd^{-shape}-Hox (Lib-Hth-R) continued from Figure 4A. (B) and (C) show blow-ups of Y5=C (top) or Y5=T (bottom) 12-mer sequences. Each individual box shows subsequences that vary in the N1 base identity (N1=A,C,G, or T) and the degree to which they are lost in Exd^{-shape} (lines represent linear fit). **(D)** Comparison of 12-mer relative sequence enrichments between Hth^{FL}-Exd-Hox and Hth^{FL}-Exd^{-shape}-Hox (Lib-Hth-F). 12-mers are colored by the PSAM sequence score most likely explaining their enrichment: Hth-dimers (purple), canonical Exd-Hth^{FL} binding (dark blue), Hth^{FL}-Exd-Hox bound to the lower affinity sites (Y5=C: NTGA**C**NNAYNNN; coral red), and Hth^{FL}-Exd-Hox bound to the high affinity sites (Y5=T: NTGA**T**NNAYNNN; green). **(E)** and **(F)** Blow-up of Y5=C (top) or Y5=T (bottom) 12-mer sequences from plot (D). Each individual box shows subsequences that vary in the N1 base identity (N1=A,C,G, or T) and the degree to which they are lost in Exd^{-shape} (lines represent linear fit). **(G)** Comparison of 12-mer relative sequence enrichments between Hth^{FL}-Exd and Hth^{FL}-Exd^{-shape} (Lib-16). 12-mers are colored by the PSAM sequence score most likely explaining their enrichment: Hth-dimers (purple), canonical Exd-Hth^{FL} binding (dark blue). **(H)** Blow-up of 12-mer sequences from plot (G). Each individual box shows subsequences that vary in the N1 base identity (N1=A,C,G, or T) and the degree to which they are lost in Exd^{-shape} (lines represent linear fit). **(I)** Proposed mechanisms for the sequence-dependent, differential binding loss of Hth^{FL}-Exd^{-shape}-Hox. Average MGW profiles are shown for the top 10 highest scoring sequences when the N1 base is either fixed to A (light green) or C (dark green) before fitting the binding model. N1=A type sequences have an optimal, narrow MGW along the entire N-terminal end of Exd, thus more likely to be affected by a mutation removing shape readout, whereas N1=C type sequences have a much wider MG thus their binding mode is less impacted. **(J)** As in **(I)** but for position Y5 (coral red: Y5=C and green Y5=T). Widening of the MGW at positions N3-6 and the resulting shift in readout preferences is likely to explain the differences in Exd^{-shape} binding to the two types of sequences.

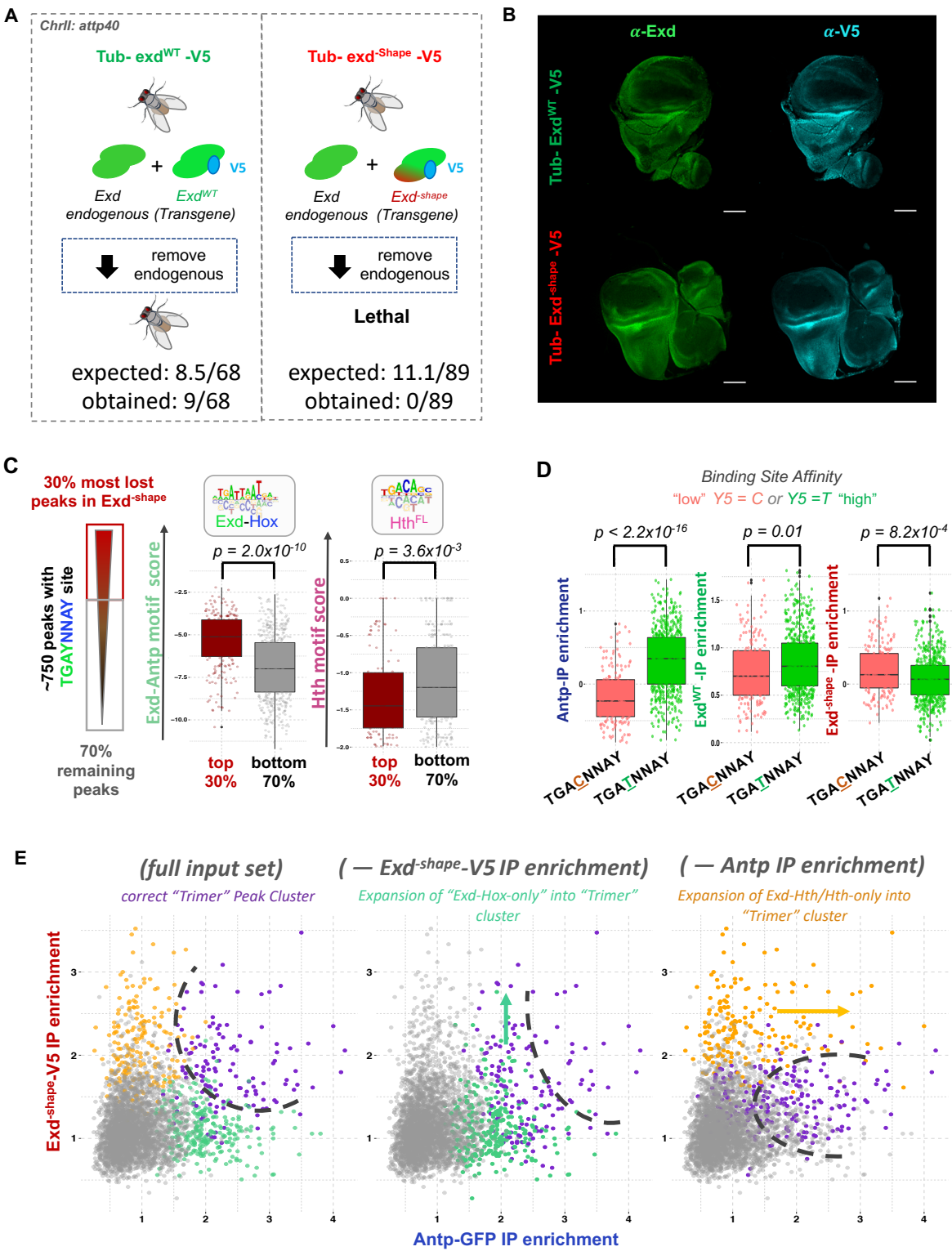

Figure S4. Using Exd^{-shape} to probe complex composition and binding mechanisms in vivo. (related to Figure 4 and 5)

(A) Transgenes carrying either Exd^{WT} or Exd^{-shape} tagged with V5 are inserted into the attp40 landing site on chromosome II. Removing endogenous Exd transcripts is rescued by Exd^{WT}-V5 (9 males of 68 progeny, expected 8.5) but not Exd^{-shape}-V5 (0 males of 89 progeny, expected 11, p = 0.004). **(B)** Anti-Exd (green) and anti-V5 (cyan) stain of homozygous Exd^{WT} or Exd^{-shape} V5-tagged 3rd instar imaginal wing discs (in the background of endogenous Exd). V5 signal follows total Exd stain and is nuclear in both genotypes. Scale bars indicate 100 μm. **(C)** Hth^{FL} binding stabilizes binding loss of Exd-Hox dimers: The 752 peaks containing a match to the consensus Exd-Hox site (TGAYNNAY) were split according to Exd^{-shape} binding loss (ratio Exd^{WT}/Exd^{-shape}; 30:70 - red:grey) and the peak sequences (100 bp around summit) were scored with either Exd-Antp (turquois) or Hth (pink) motif models. 30:70 peak sets were tested for differences in Exd-Antp BS peak score and the presence of a secondary Hth motif. **(D)** The 752 peaks with a match to an Exd-Hox TGAYNNAY site were split into two groups Y_5=C (low affinity; coral red) and Y_5=T (high affinity; green), and the *in vivo* sequence selectivity for Antp, Exd^{WT} and Exd^{-shape} is compared between the two groups. Exd^{-shape} prefers the low affinity Y_5=C over the high affinity Y_5=T sites. **(E)** Scatterplot of Exd^{-shape}-V5 IP enrichment against Hox (Antp-GFP) IP enrichment for all ~3700 Exd^{WT}-V5 peak summits. Coloring is based on peak clustering using i) the full feature set (three IP signals (Exd^{WT}-V5, Exd^{-shape}-V5, Antp-GFP) and three binding site scores (Exd-Antp, Hth-only, Exd-Hth; *in vitro* binding models); ii) 5 out of 6 features, leaving out Exd^{-shape}-V5 (middle panel); iii) 5 out of 6 features, leaving out Antp IP signal (right panel). Only the full model correctly differentiates the "trimeric" cluster that shows stabilized Exd^{-shape} signal at high Antp IP signal (purple). Both reduced feature sets produce impure clustering; with leaving out Exd^{-shape}-V5 resulting in blending of "Exd-Hox-only" and "Trimer" sites and leaving out Antp in blending of "Exd-Hth/Hth-only" and "Trimer" sites (see dashed lines and arrows).

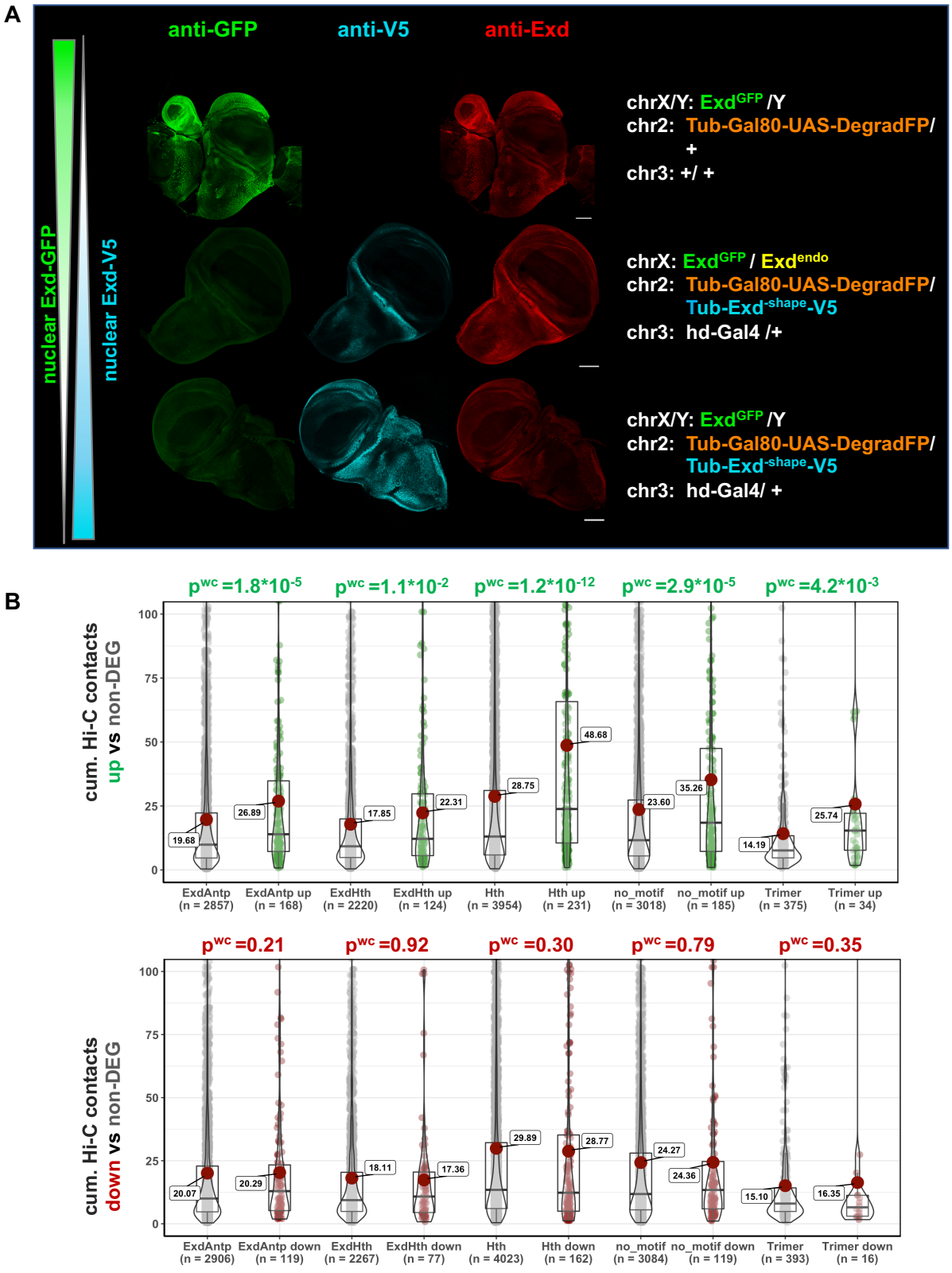

Figure S5. Exd^{-shape} as a tool to study complex-specific function in vivo. (related to Figure 6)

(A) Testing the deGradFP system: Images show a confocal plane of third instar imaginal wing discs, with larva raised at 18℃ and shifted to 29℃ 24h prior to dissection. Three different genotypes are shown: male wing disc containing one copy of Exd^{GFP} on the X chromosome and the Gal80 controlled deGradFP System on the second chromosome (top); female wing disc containing one Exd^{GFP} and one copy of endogenous Exd on the X chromosomes, the Gal80-deGradFP system and the transgenic Exd^{-shape}-V5 on the 2nd chromosome, as well as the headcase-Gal4 driver (hd-Gal4) on the third chromosome (middle); as middle, except male disc with the endogenous Exd copy absent (bottom). In the absence of hd-Gal4, the deGradFP system is not active and Exd^{GFP} is nuclear (top). Upon activation of the deGradFP system the nuclear Exd^{GFP} is greatly reduced and Exd^{-shape}-V5 competes with endogenous Exd for nuclear localization (middle). Upon removal of endogenous Exd (by only considering male flies), and continuing to remove Exd^{GFP}, transgenic Exd^{-shape}-V5 is the only source of Exd protein and nuclear levels increase. Scale bars indicate 100 μm. **(B)** Promoters of up, but not downregulated genes in Exd^{-shape} vs Exd^{WT} display a significantly higher Hi-C contact frequency (Wilcox test (wc)) for each of the five major peak classifications – Exd-Hox, Exd-Hth, Hth-only, motifless ("no-motif"), and Trimer binding sites – compared to promoters whose transcripts are not differentially expressed (non-DEG; FDR threshold set to 5%).

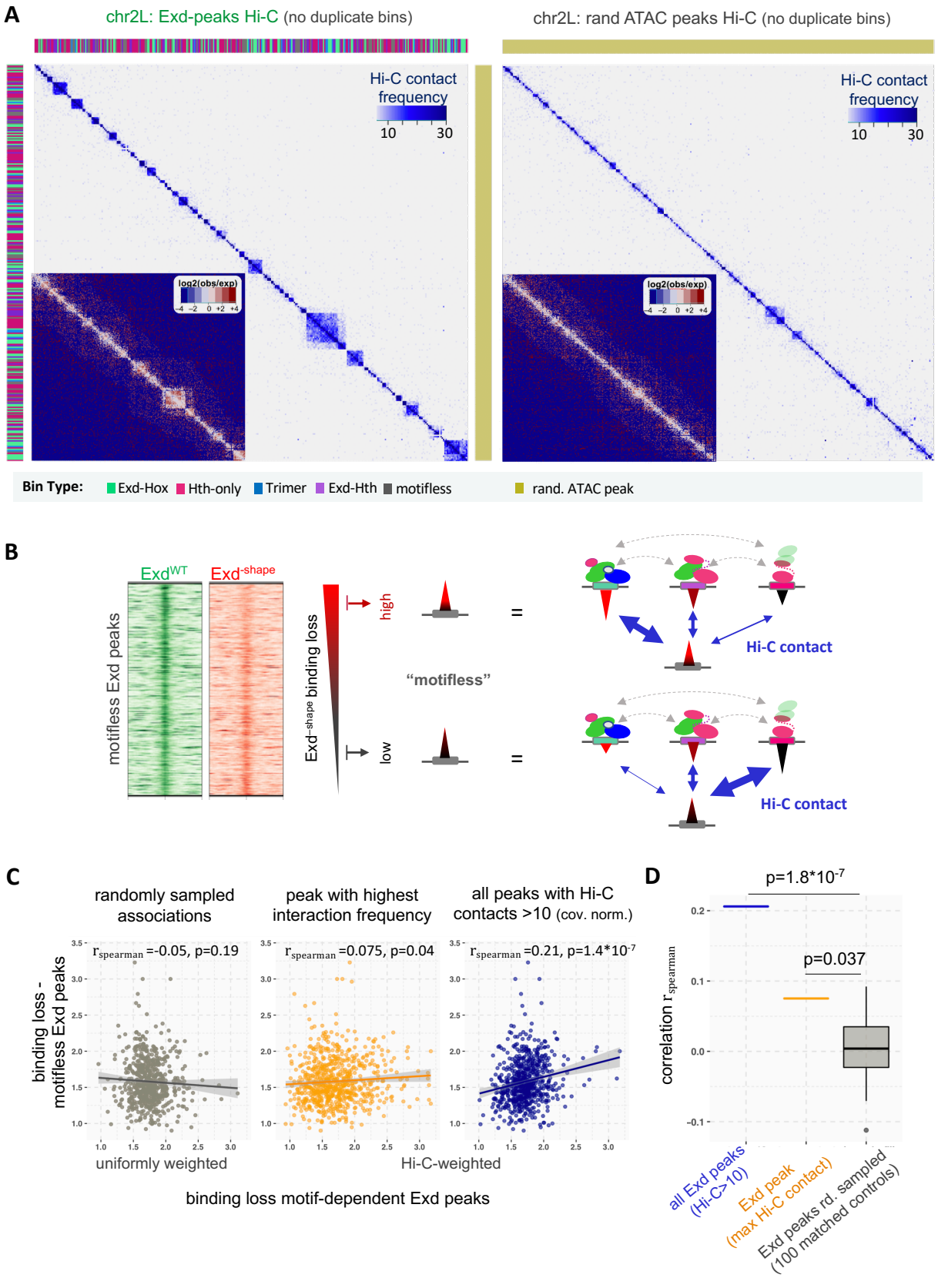

Figure S6. Exd binding sites physically interact. (related to Figure 5 and 6)

(A) Hi-C contact maps of wild-type (incl. tub>exd^{WT}-V5 transgene) third instar larval wing discs isolating either the set of Exd peaks (left) or a size-matched random set of ATAC-seq peaks (right). Clusters of sites corresponding to distinct genomic contact domain structures are seen for the Exd peakset, but not for a size-matched random sample of accessible genomic regions (p-value = $2.6*10^{-32}$). The insets show the logarithm base two of the observed/expected Hi-C count ratio, computed following the method of (Rao et al., 2014), for the same peak sets. Only interactions between unique genomic bins are shown. In the left panel, whenever a genomic bin contains more than one Exd peak, one of these is chosen to represent the "bin-type" color shown along the margin of the matrix. **(B)** The degree of IP signal loss at motifless binding sites in response to Exd^{-shape} mutation is variable, which might be caused by 3D contacts with motif-dependent Exd binding sites. Solid dark blue double arrows indicate potential contacts between motifless and motif-containing sites; dashed grey arrows reflect the overall connectivity among Exd peaks. **(C)** Exd^{-shape} binding loss at motifless Exd peaks may be inherited through direct crosslinking to motif-dependent Exd binding sites. Each plot shows the motifless peak Exd^{WT}/Exd^{-shape} IP ratio on the y-axis and the Hi-C weighted Exd^{WT}/Exd^{-shape} IP ratio at motif-dependent sites that are associated with motifless sites on the x-axis. Three different approaches are shown to compute the motif-dependent signal loss associated with each motifless peak: (i) motif-dependent peaks on the same chromosome were randomly sampled using uniform weighting (the number of sampled peaks considered per motifless site is matched to the threshold used in (iii) see **STAR Methods**; grey); (ii) only the motif-dependent peak that has the highest Hi-C contact for each motifless peak was considered (orange); or (iii) all motif-dependent peaks that have a Hi-C contact >10 (vanilla coverage normalized) were included using the log10 Hi-C contacts as weights (blue). Lines represent linear fits. Spearman correlation estimates and p-values are shown above plot. **(D)** Spearman correlations for Exd^{WT}/Exd^{-shape} IP ratio between motifless and motif-dependent peaks using either approach ii) (orange) or iii) (blue) described in **(C)** are compared to the distribution of 100 correlations generated using approach i) in **(C).** The probability of obtaining the observed correlations given the distribution of random peak associations are indicated above the graph.

Table S1. Statistics for Sequencing Data. (related to Figures 4-7)

Sample	sequencing reads	% aligned
CHIP-seq IP replicate 1 anti-V5 tub>exd^{WT}-V5; wing	24064224	95.4
CHIP-seq INPUT replicate 1 tub>exd^{WT}-V5; wing	26790115	96.5
CHIP-seq IP replicate 1 anti-V5 tub>exd^{-shape}-V5; wing	24665534	94.8
CHIP-seq INPUT replicate 1 tub>exd^{-shape}-V5; wing	28304229	96.4
CHIP-seq IP replicate 2 anti-V5 tub>exd^{WT}-V5; wing	17954539	96.0
CHIP-seq INPUT replicate 2 tub>exd^{WT}-V5; wing	22858962	97.0
CHIP-seq IP replicate 2 anti-V5 tub>exd^{-shape}-V5; wing	20530272	95.3
CHIP-seq INPUT replicate 2 tub>exd^{-shape}-V5; wing	23430975	97.1
CHIP-seq IP against Antp-GFP; wing	29481911	96.2
CHIP-seq INPUT Antp-GFP; wing	32714133	96.7
CHIP-seq IP against Hth (tub>exd^{WT}-V5); wing	24381125	92.5
CHIP-seq INPUT Hth (tub>exd^{WT}-V5); wing	25177504	96.3
ATAC seq yw wild-type; wing	73805409	95.9
In Situ Hi-C in tub>exd^{WT}-V5; wing	36184416	93.1
RNA-seq in tub>exd^{WT}-V5; deGradFP; wing	43661540	96.9
RNA-seq in tub>exd^{-shape}-V5; deGradFP; rep. 1; wing	69717436	96.7
RNA-seq in tub>exd^{-shape}-V5; deGradFP; rep. 2; wing	49007286	96.3
RNA-seq in tub>exd^{WT}-V5; replicate 1; wing	66930460	96.7
RNA-seq in tub>exd^{WT}-V5; replicate 2; wing	50993994	96.7
RNA-seq in tub>exd^{WT}-V5; replicate 3; wing	70598809	96.7
RNA-seq in tub>exd^{WT}-V5; replicate 1; CNS	63836614	96.4
RNA-seq in tub>exd^{WT}-V5; replicate 2; CNS	69668130	96.4
RNA-seq in tub>exd^{WT}-V5; replicate 3; CNS	58194992	96.6