Supplemental Data

Analysis of Homeodomain Specificities Allows the Family-wide Prediction of Preferred Recognition Sites

Marcus B. Noyes, Ryan G. Christensen, Atsuya Wakabayashi, Gary D. Stormo, Michael H. Brodsky, and Scot A. Wolfe

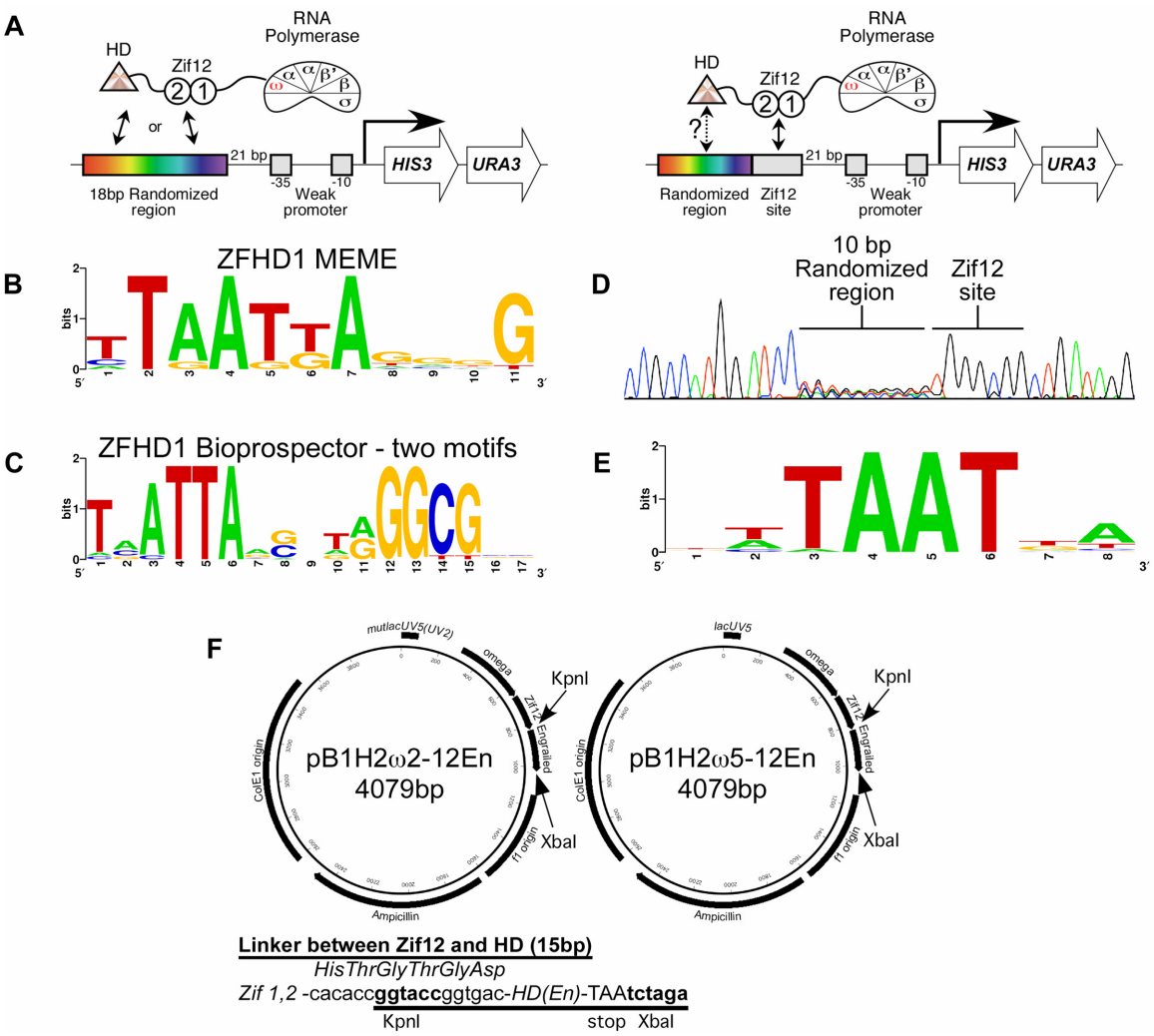
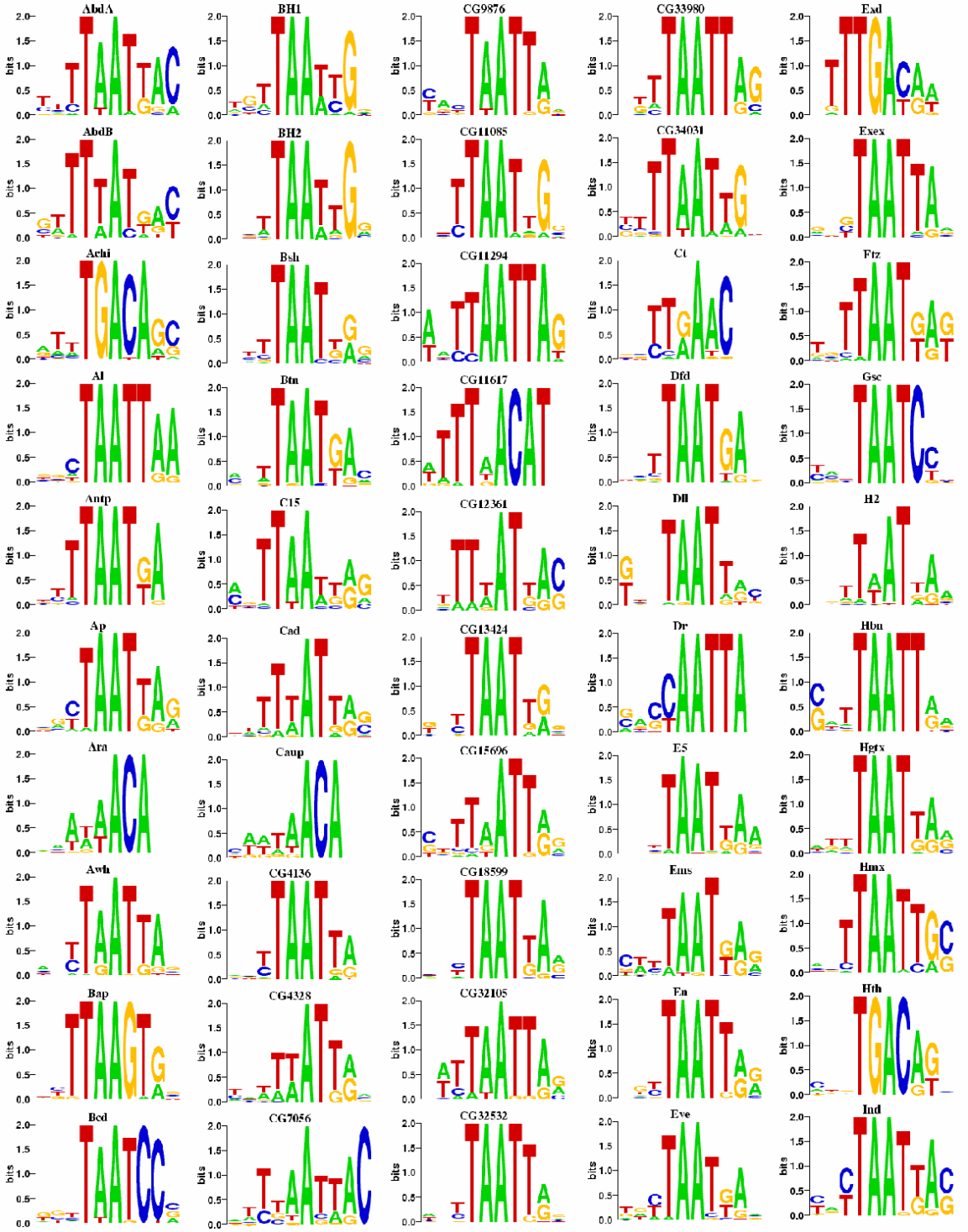


Figure S1. The ZFHD1 framework for determining DNA-binding specificity. A. (left) Cartoon depicting the interaction of the omega-ZFHD1 construct with a reporter with a 18 bp randomized window and (right) one that has a fixed binding site for the Zif12 DNA-binding domain in the context of the ZF10 library. B and C. The logos resulting from the ZFHD1 selection using the 28 bp library as determined by MEME (B. searching for one motif) and by BioProspector (Liu et al., 2001) (C. searching for 2 separate motifs). The discovered motifs in the BioProspector analysis are consistent with the specificity of Zif12 and Oct1

(Pomerantz et al., 1995). D. Sequence chromatogram of the ZF10 library prior to selection with a transcription factor where the 10 bp randomized region and the Zif12 binding sites are indicated. E. The logo resulting from the ZFHD1 selection on the ZF10 library. $5x10^7$ bacteria containing the ZF-10 library and the omega-Zif12-Oct1 expression vector were selected on minimal medium lacking histidine and containing 10 mM 3-AT. Approximately 2000 colonies survived the selection, which represented a > 100-fold increase over the number of surviving clones in the omega-Zif12 negative control. MEME analysis of 22 unique sequenced clones recovered a motif consistent with the specificity of Oct1 from 22 of 22 sequences (Pomerantz et al., 1995; Verrijzer et al., 1992). F. The maps of the bait plasmids used to characterize the homeodomains as omega-Zif12 fusions with key features annotated. These plasmids allow a homeodomain to be characterized at two different promoter strengths: either expressed under control of the *lacUV5* or a mutant *lacUV5* (*lacUV5mut*) promoter. Homeodomains were introduced by simply subcloning between the unique KpnI and XbaI sites in each plasmid. The linker between Zif12 and the TF is indicated below the plasmids.


AbdA
BH1
CG9876
CG33980
Exd
AbdB
BH2
CG11085
CG34031
Exex
Achi
Bsh
CG11294
Ct
Frz
Al
Btn
CG11617
Dfd
Gsc
Antp
C15
CG12361
Dll
H2
Ap
Cad
CG13424
Dr
Hbn
Ara
Caup
CG15696
E5
Hgtx
Awh
CG4136
CG18599
Ems
Hmx
Bap
CG4328
CG32105
En
Hth
Bcd
CG7056
CG32532
Eve
Ind
bits

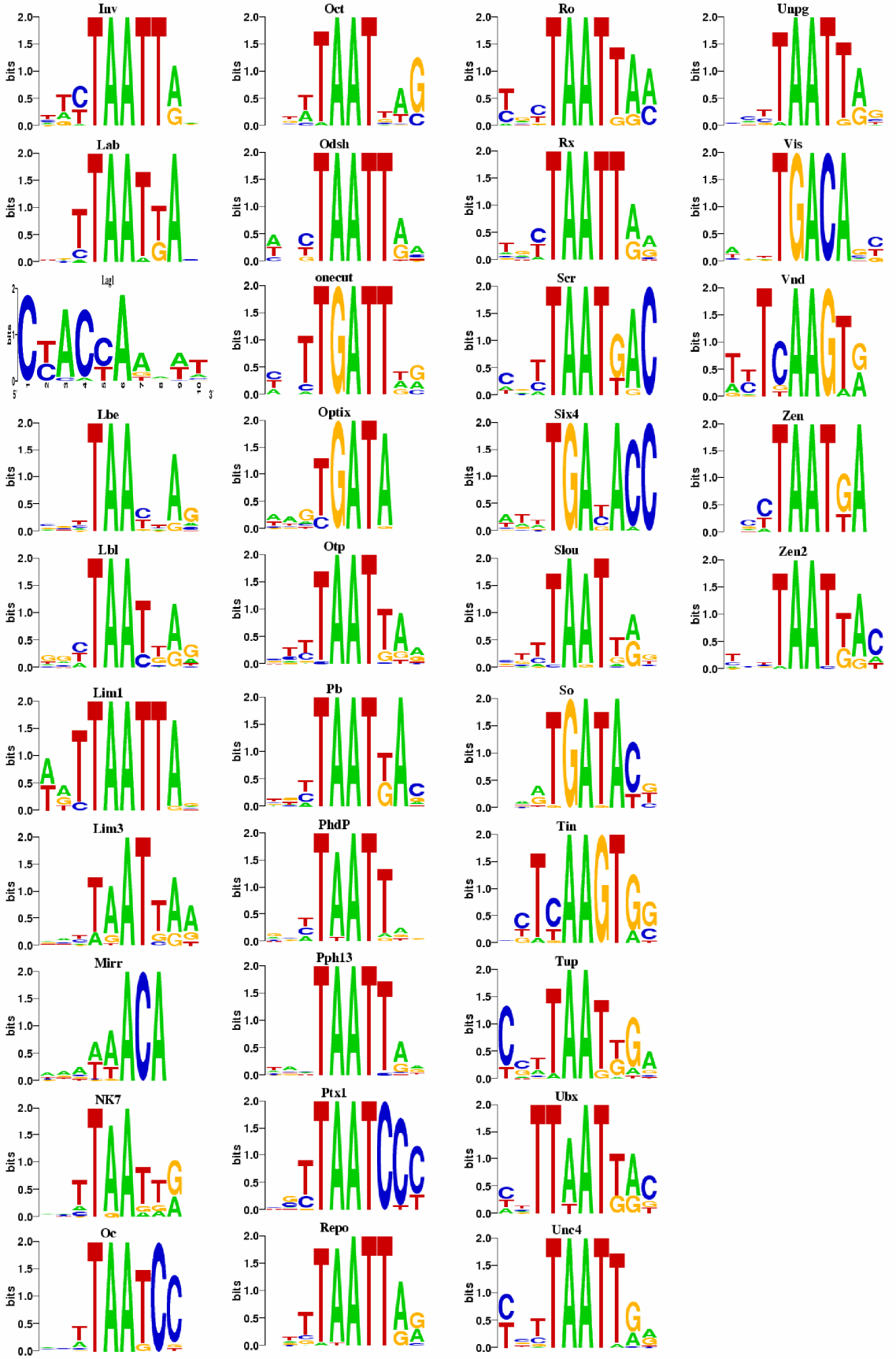

Inv
Oct
Ro
Unpg
Lab
Odsh
Rx
Vis
Lagl
onecut
Scr
Vnd
Lbe
Optix
Six4
Zen
Lbl
Otp
Slou
Zen2
Lim1
Pb
So
Lim3
PhdP
Tin
Mirr
Pph13
Tup
NK7
Ptx1
Ubx
Oc
Repo
Unc4
bits

Figure S2. Sequence logos for 84 homeodomains determined in this study. Binding site alignments for each of the 84 homeodomains were extracted from the master alignment of sites. We generated Sequence logos (Schneider and Stephens, 1990) for each of these alignments using WebLogo version 2.8 (Crooks et al., 2004). WebLogo converts alignments to count matrices. Gaps are treated as missing data and ignored, thus not all column count totals are equal. There are no gaps in the core portion of the master alignment, however (columns 4 through 8, see Experimental Procedures).

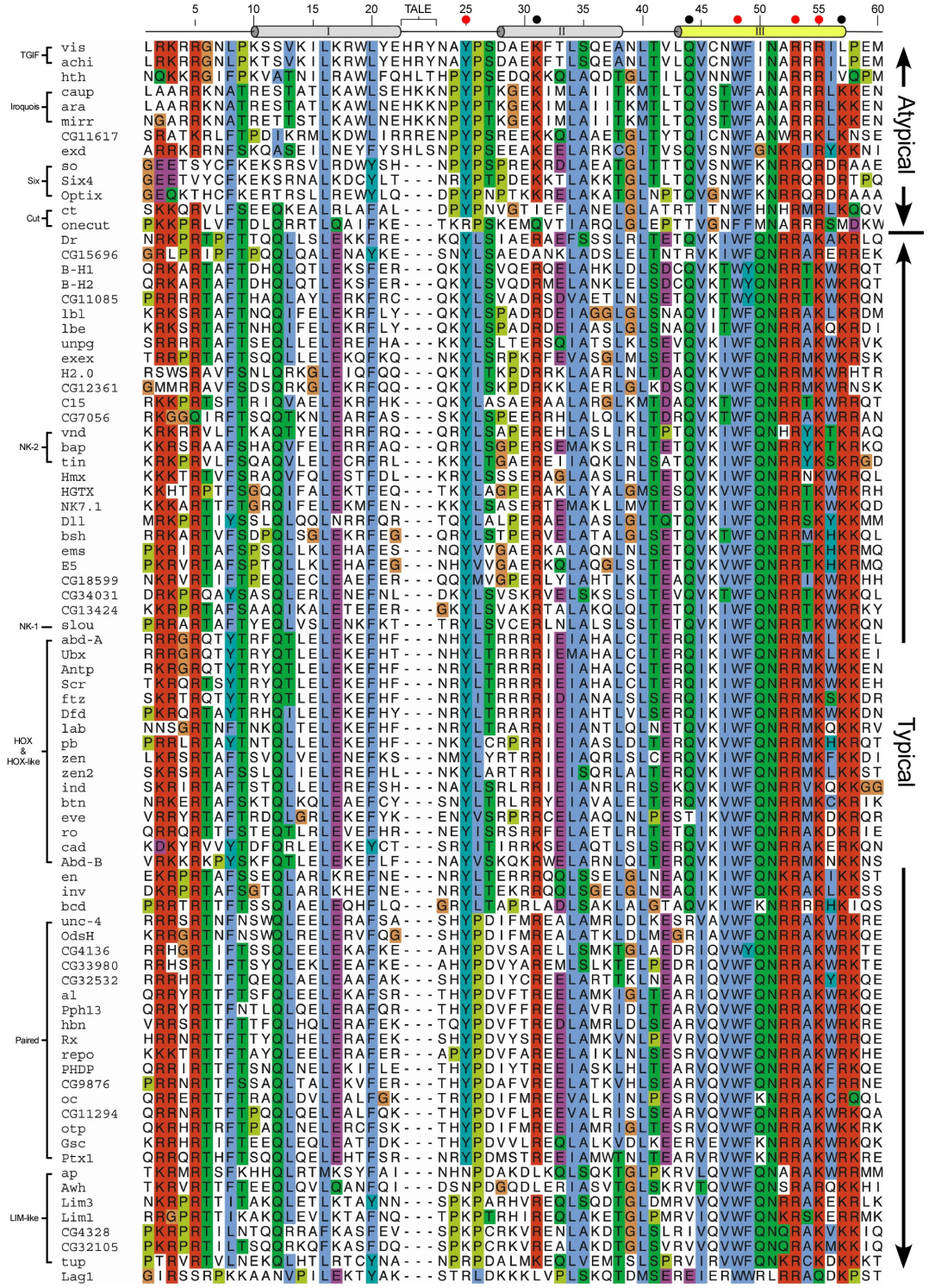


Figure S3. ClustalW alignment of the homeodomain motifs from the 84 homeodomains examined in this study. The position of each amino acid in the canonical homeodomain numbering scheme is indicated at the top of the alignment, along with cylinders that indicate the position of the three alpha helices in the structure. The homeodomains can be divided into two broad superclasses: the Atypical (top) and Typical (bottom), which have distinguishing sequence features and recognition motifs. One of the most prominent sequence differences is the presence of a three amino acid loop extension (TALE) between helix 1 and 2

within the majority of the atypical homeodomains(Banerjee-Basu and Baxevanis, 2001; Mukherjee and Burglin, 2007). Where clearly defined, the classes that the homeodomains fall within are indicated to the left of their names (Banerjee-Basu and Baxevanis, 2001; Harvey, 1996). Colored residues at each position within the alignment indicate a high frequency of occurrence of residues with similar properties, with the exception of Gly and Pro, which are highlighted throughout these sequences. Filled circles above the sequences indicate positions that make phosphate contacts in the structures of both typical and atypical homeodomain-DNA complexes (Fraenkel et al., 1998; Joshi et al., 2007; Wolberger et al., 1991). Red circles indicate positions where an amino acid type capable of making a phosphate contact is conserved in all Asn51 containing homeodomains. At these seven common phosphate-contacting positions 95% (79 of 83) have an appropriate residue for this interaction at 6 of the 7 positions. In the case of position 31, the absence of a Lys or Arg can be compensated by one at position 46, as either position can contact a common phosphate (Grant et al., 2000). Not surprisingly, two of the outliers with regards to conserved phosphate contacts are Cut and Onecut, which display unique specificities.

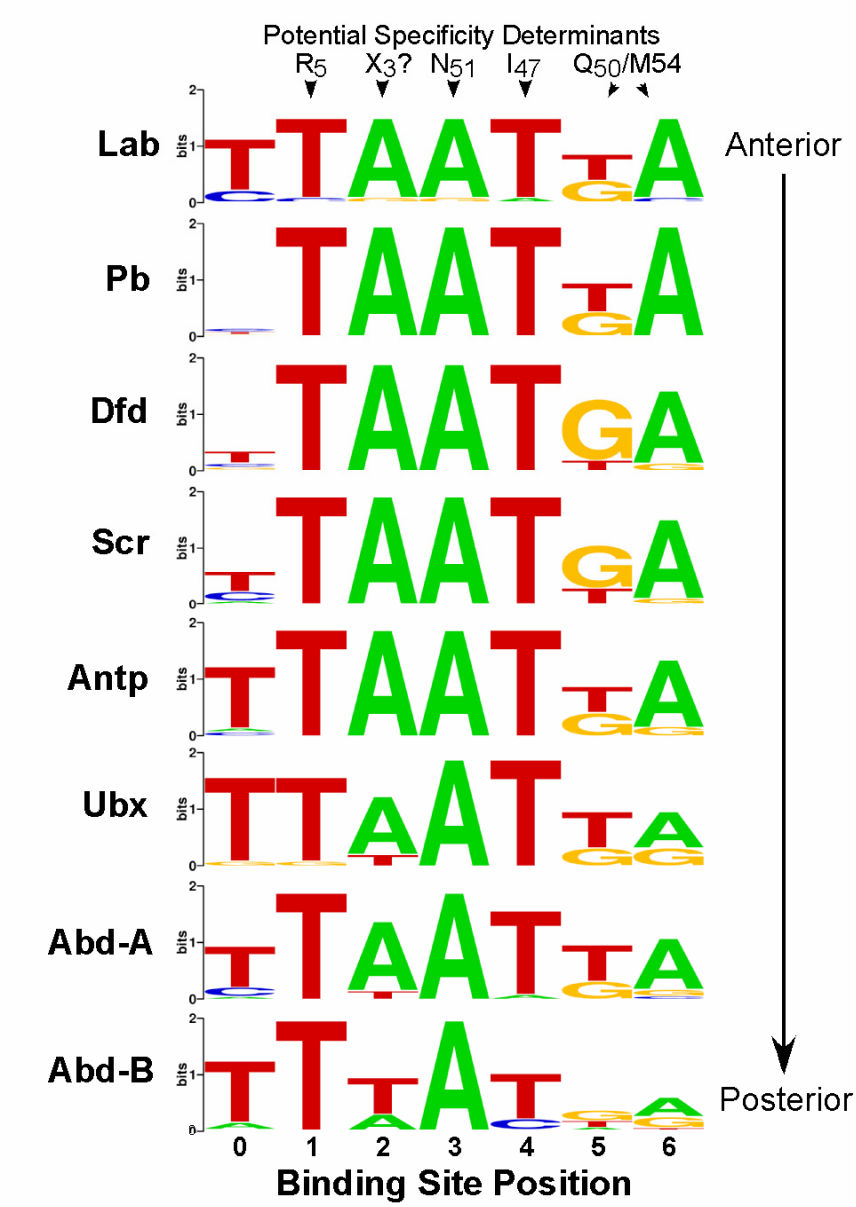
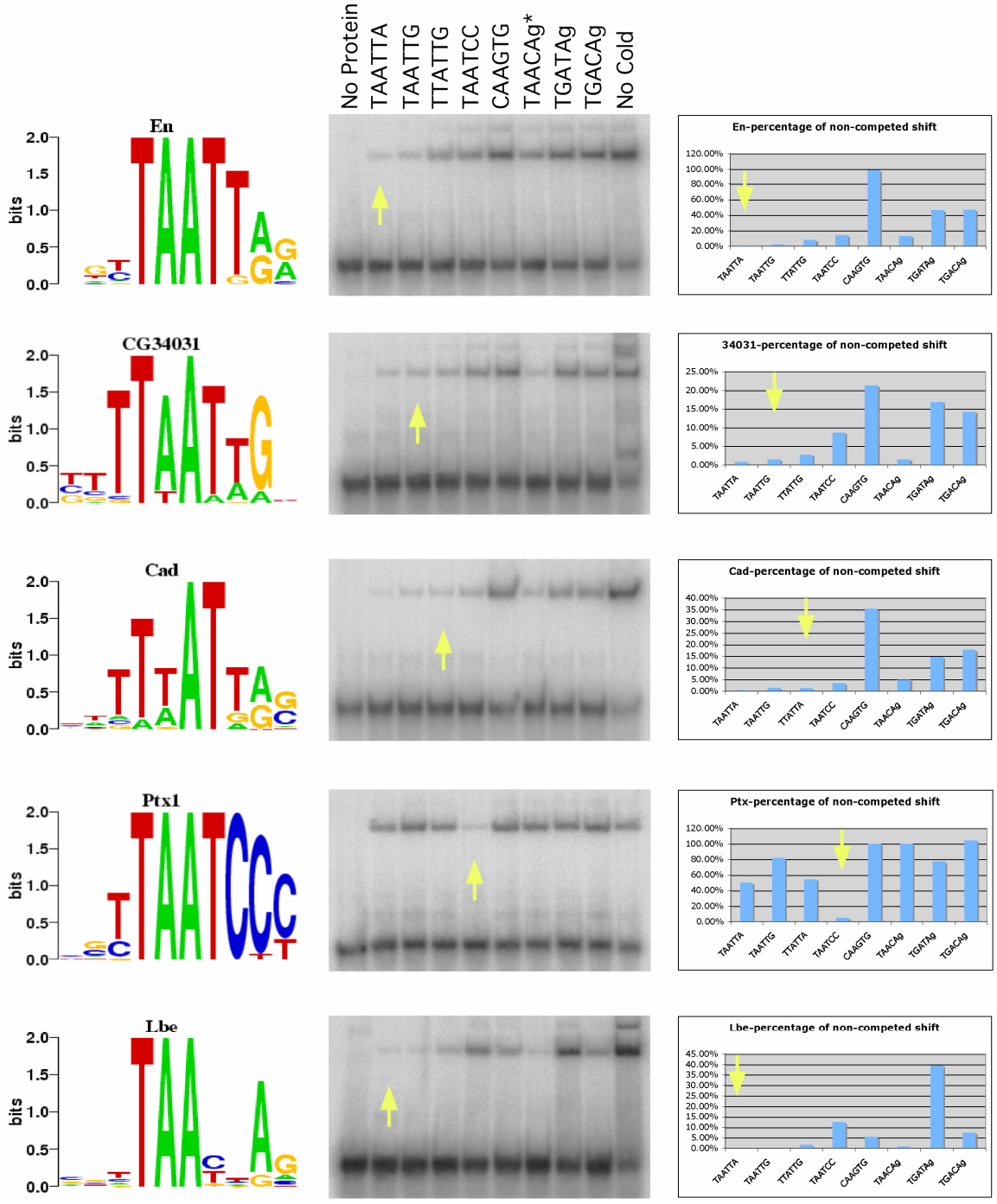


Figure S4. Binding site logos for the eight *D. melanogaster* HOX factors displayed in order of anterior to posterior expression (top to bottom). Potential specificity determinants for each position of the binding site are indicated above the logos. There are subtle differences between the specificity of each HOX factor that are highlighted when they are arranged according to their regional expression pattern within the embryo. For example, at binding site position 2, the specificity shifts from a strong preference for Ade in more anterior factors toward an increasing tolerance for Thy in the most posterior factors, with Abd-B displaying a significant preference for Thy over Ade as was previously noted in the study by Beachy and colleagues(Ekker et al., 1994). This change in base preference was ascribed to residues at position 3, 6 and 7 in the N-terminal arm of Abd-B(Ekker et al., 1994). A second trend is the increasing preference for Thy at position 0 in the more posterior factors. This observation is consistent with a greater preference for Thy in Ubx and Abd-B compared to Dfd in *in vitro* binding site selections(Ekker et al., 1994; Ekker et al., 1992). Another notable difference previously observed for Dfd and Ubx was a stronger preference of Dfd for Gua over Thy at binding site position 5, whereas Ubx did not display an obvious preference for Gua over Thy at this position(Ekker et al., 1992).


No Protein
TAATTA
TAATTG
TTATTG
TAATCC
CAAGTG
TAACAg*
TGATAg
TGACAg
No Cold
En
CG34031
Cad
Ptx1
Lbe
bits
En-percentage of non-competed shift
34031-percentage of non-competed shift
Cad-percentage of non-competed shift
Ptx-percentage of non-competed shift
Lbe-percentage of non-competed shift

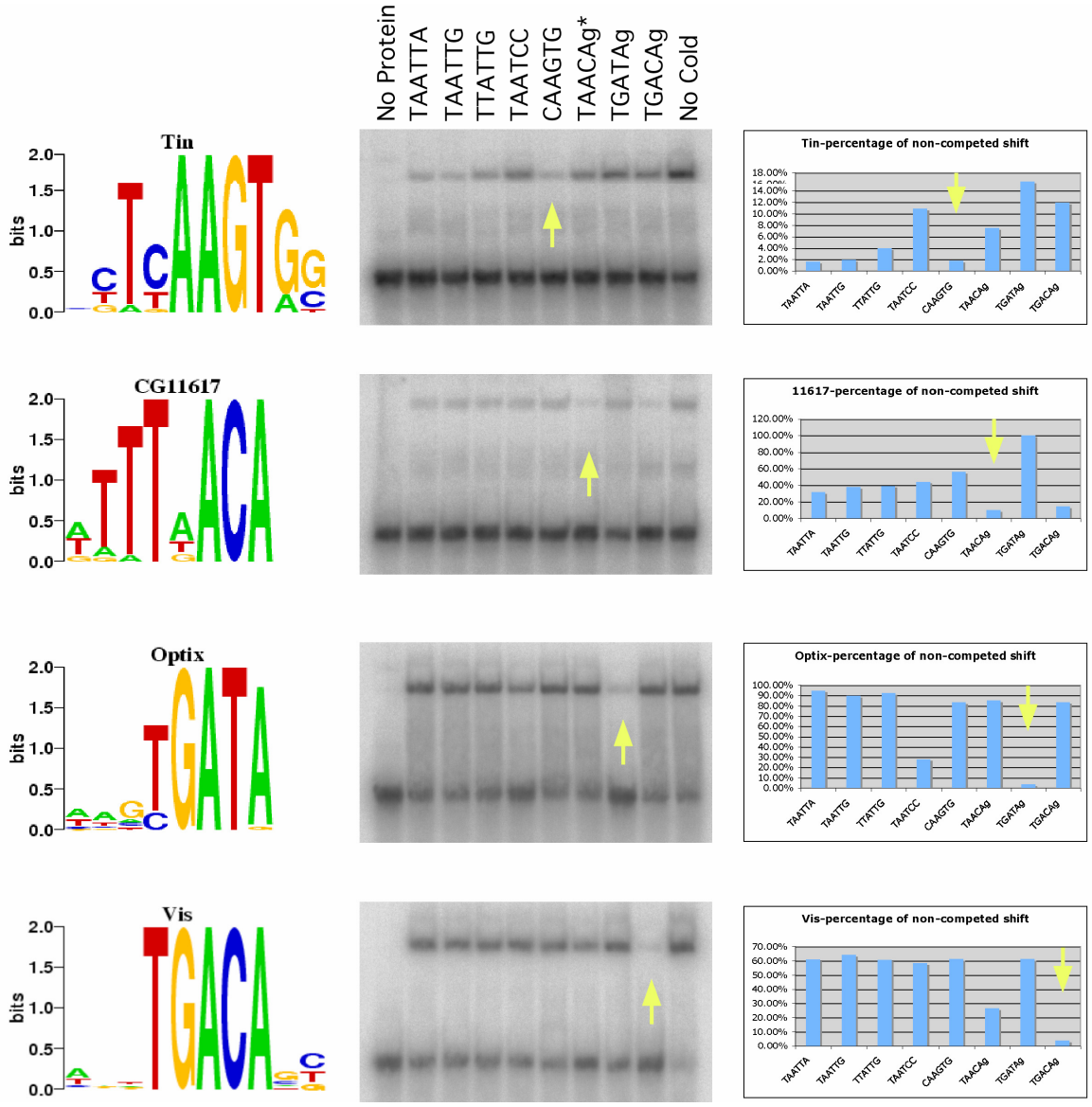


Figure S5. Competitive gel shift assays for nine different fly homeodomains representing nine different specificity groups. The sequence logo for the factor used in each assay is shown on the left of each row. Radiolabeled DNA containing the consensus binding site for each factor was challenged with excess cold competitor DNA containing sequences from the various specificity groups. Each lane represents a different competition, where the cold competitor and protein are at constant concentrations, with the exceptions of the far left (no protein) and far right (no competitor) lanes. The sequence of the homeodomain binding site used in each competition is indicated at the top of the page, where they are at identical positions in each assay. A yellow arrow indicates the competitor that is identical in sequence to the radiolabeled probe. The quantification of the change in percent shift as a function of the concentration of cold competitor is shown on the right where these values are normalized to the percent shift in the no competitor lane. A yellow arrow indicates the competitor that is identical in sequence to the radiolabeled probe. The neighboring 5' constant region of each binding site is CAG with the exception of the indicated site (*), which contains an additional T (CAG**T**). The reverse complement of this element with a portion of the binding site creates an element "TAACTG" which may compete somewhat effectively with some of the typical homeodomains.

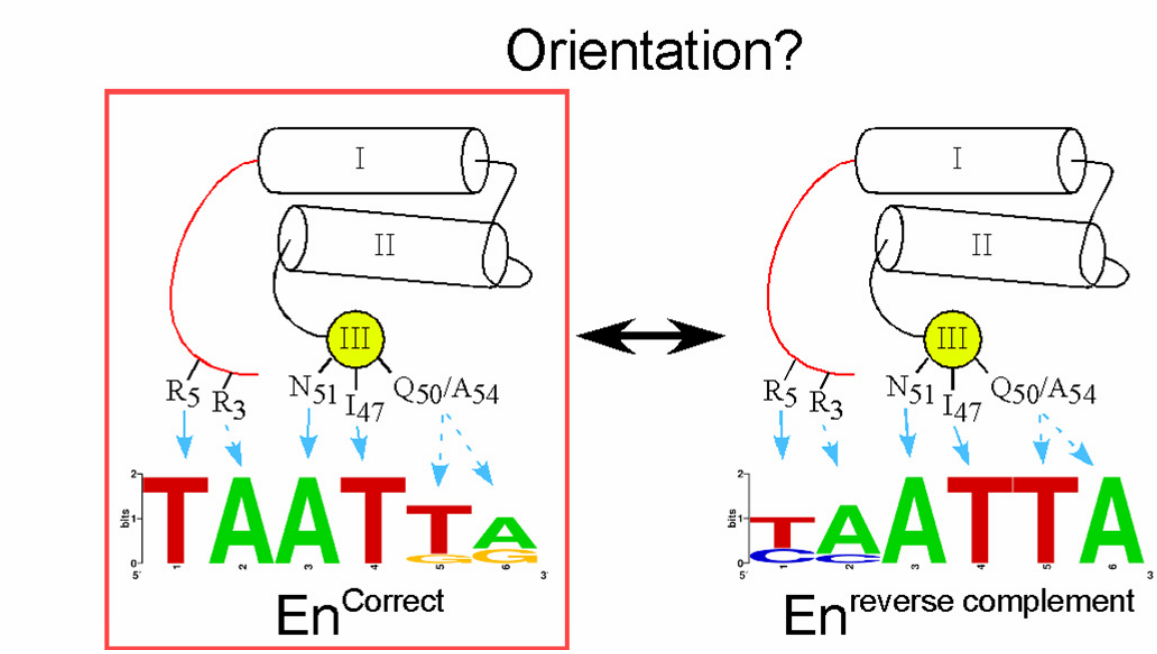


Figure S6. Determining the orientation of the homeodomain on a selected site. To correlate differences in homeodomain specificity with differences in their amino acids, it is also critical to determine the orientation of the homeodomain with respect to the motif. For example, if the consensus sequence for a factor defined from a selection is TCATTA, the N-terminal arm might either specify the TC at the 5' end of this sequence or the TA at the 5' end of the complimentary sequence TAATGA. In this study, base positions predicted to interact with the N-terminal arm are at the 5' positions of the motif (Figure 1). Previous structural and mutagenesis studies of well-characterized homeodomain proteins such as En, Ubx, and Bcd have defined the orientation of the homeodomain on its binding site (Ades and Sauer, 1994; Ekker et al., 1992; Fraenkel et al., 1998; Hanes and Brent, 1989; Percival-Smith et al., 1990; Treisman et al., 1989). For factors, such as En, that can bind a pseudo-C2 symmetric binding sequence (*e.g.* TAATTA), it is not immediately obvious from the binding site logo alone how the sequence motif should be orientated. Fortunately, the short tether between the omega-Zif12 module used for the B1H selection and the homeodomain biases the location of binding sites recovered from the randomized library. Homeodomain binding sites that neighbor the fixed Zif12 binding site are preferentially recovered on the DNA strand complementary to the Zif12 binding site. In contrast, more distant homeodomain sites are generally recovered on the same strand as the Zif12 binding site. This positional bias is most easily observed in the selected binding sites of a factor such as Bcd, which binds a clearly asymmetric binding sequence (TAATCC, Supplementary Table 2). This site bias can be used to define the orientation of the homeodomain on the consensus sequence, which facilitates the comparison of recognition sequences for each homeodomain in a common register.

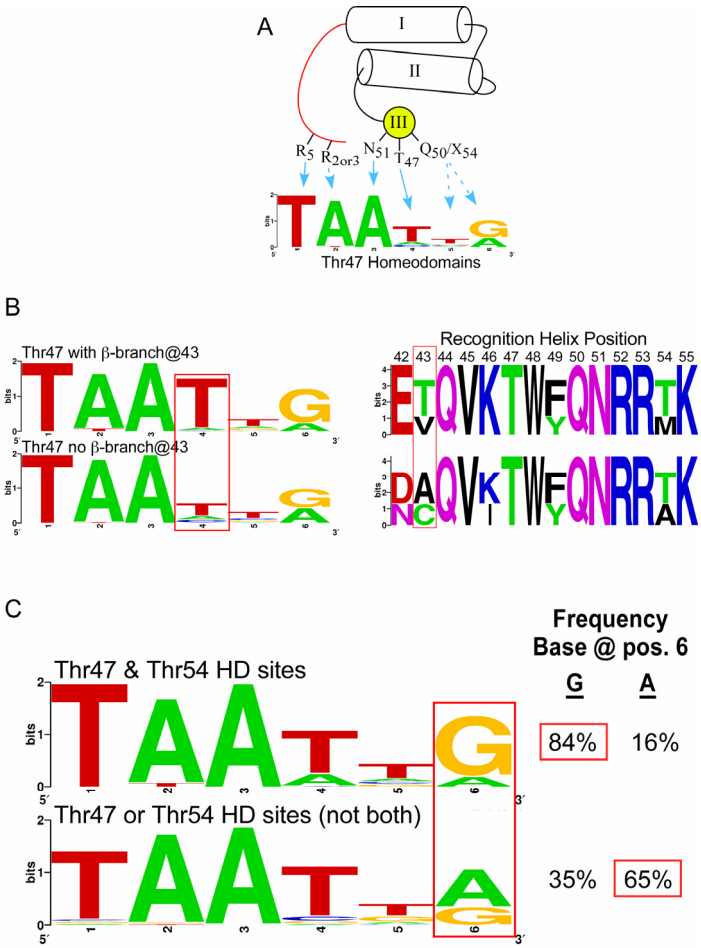


Figure S7. Secondary effects influence recognition for Thr47 containing homeodomains. A) Schematic of the potential interactions between the Thr47 containing homeodomains and their binding sites. When the eight typical homeodomains that contain Thr at position 47 are considered together, they display a clear reduction in specificity at binding site positions 4 and 5 of their as assessed by the information content at these positions. B) This set of Thr47-containing factors can be divided into two sub-sets with regards to the degree of preference for Thy at position 4 of their binding site. Factors that contain a β-branched amino acid at position 43 (Bsh, CG11085 & CG34031) display a strong preference for Thy at position 4, while the remaining five factors (Lbe, Lbl, B-H1, B-H2 & C15) display only a weak preference The difference between these two motifs is statistically significant (p-value = 2.48e-4). Because amino acids at positions 43 and 47 are on the same face of the recognition helix and are in van der Waals contact in some crystal structures (Piper et al., 1999), it is plausible that the type of residue at position 43 could bias the preferred conformation of Thr47 and thereby influence its binding site preference. The Sequence logos of the recognition helices for each subgroup (not corrected for small sample size) are shown to the right of their motifs. C) The combination of Thr47 and Thr54 leads to a preference for Gua over Ade at binding site position 6 for the factors within the Bar group. Five of these factors (BH-1, BH-2, CG11085, CG34031 & C15) contain Thr at positions 47 and 54 of the recognition helix and a display a strong preference for Gua at binding site position 6, a preference that is not observed in any of the other homeodomain groups. Ten other homeodomains (Bsh, Lbl, Lbe, NK7.1, CG7056, Slou, CG13424, Hgtx, E5 & Ems) that contain Thr at either but not both positions, display a preference for an alternate base, typically A, at position 6. This difference is statistically significant (p-value = 7.05e-18). Mechanistically, how the presence of Thr at positions 47 and 54 would create a preference for Gua at position 6 of the binding site is not obvious as these residues are separated by 2 helical turns and appear more appropriately positioned to influence specificity at binding site positions 4 and 5. This combination could indirectly affect specificity at binding site position 6 through an influence on the sequence preference of Gln50.

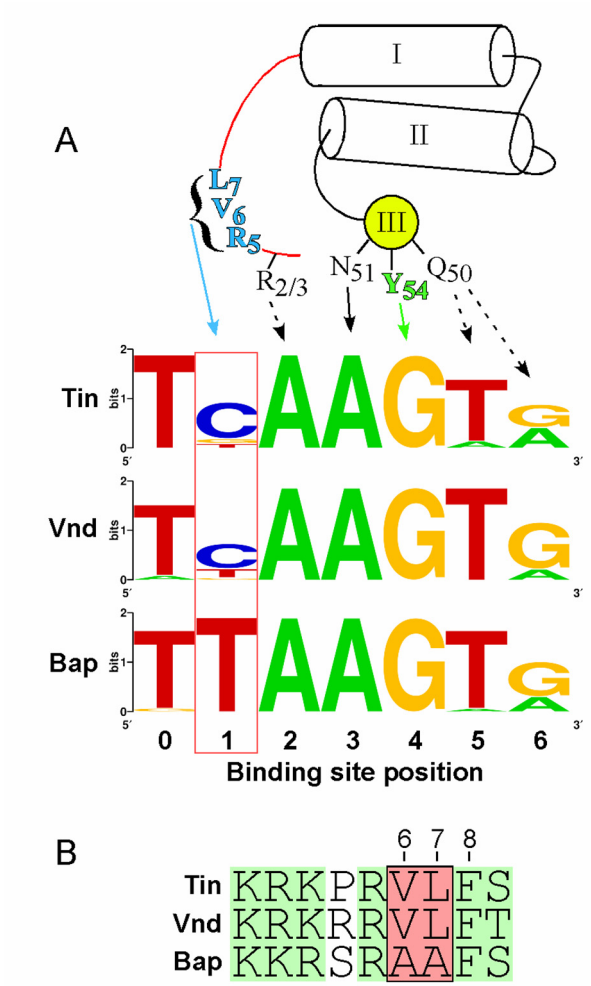


Figure S8. Residues at positions 6 and 7 of the N-terminal arm can influence the specificity at position 1 of the binding site. A) The three NK-2 class family members that share Tyr at position 54 display very similar overall specificities with the exception of binding site position. Bap (NK3) prefers a **T**AAGTG sequence, while Tin (NK4) and Vnd (NK2) prefer **C**AAGTG. A vertebrate homolog of Bap, Nkx3.2 displays an identical consensus sequence TAAGTG (Kim et al., 2003). The modest preference for Cyt at binding site position 1 has been observed for homologs of Tin and Vnd in other species and has been particularly well characterized in bovine TTF-1, which recognizes CAAGTG(Damante et al., 1996). Damante and colleagues narrowed the residues responsible for the Cyt preference at position 1 to the residues at positions 6 through 8 (VLF) in the N-terminal arm(Damante et al., 1996). They demonstrated that when these residues are substituted for the QTY in the N-terminal arm of Antp, its specificity was altered from a strong preference for Thy at binding site position 1 to a tolerance for Cyt or Thy. B) Consistent with this analysis, both Tin and Vnd have VLF motif at these positions in the N-terminal arm, while Bap contains an AAF motif. Our results allow us to further narrow the key residues responsible for the alteration in the specificity observed in the NK-2 family at binding site position 1 to residues at positions 6 and 7 in the N-terminal arm, since F is shared between all three NK-2-type factors. Only one other *D. melanogaster* homeodomain, Cut, has the VL motif at positions 6 and 7 of the N-terminal arm and it tolerates either Thy or Cyt at position 1 in the binding site, although there is a modest preference for Thy (Supplementary Figure 3). This Thy preference in Cut may be due to the presence of a contact to binding site position 2 by Arg55, which can influence the preference for Thy at position 1 in the binding site as described in the text (Figure 3B). A structural explanation for this influence is unavailable, but it is possible that the packing of the VL motif against the homeodomain fold positions a backbone carbonyl to favorably interact with the N2 position of the complementary Gua in the minor groove.

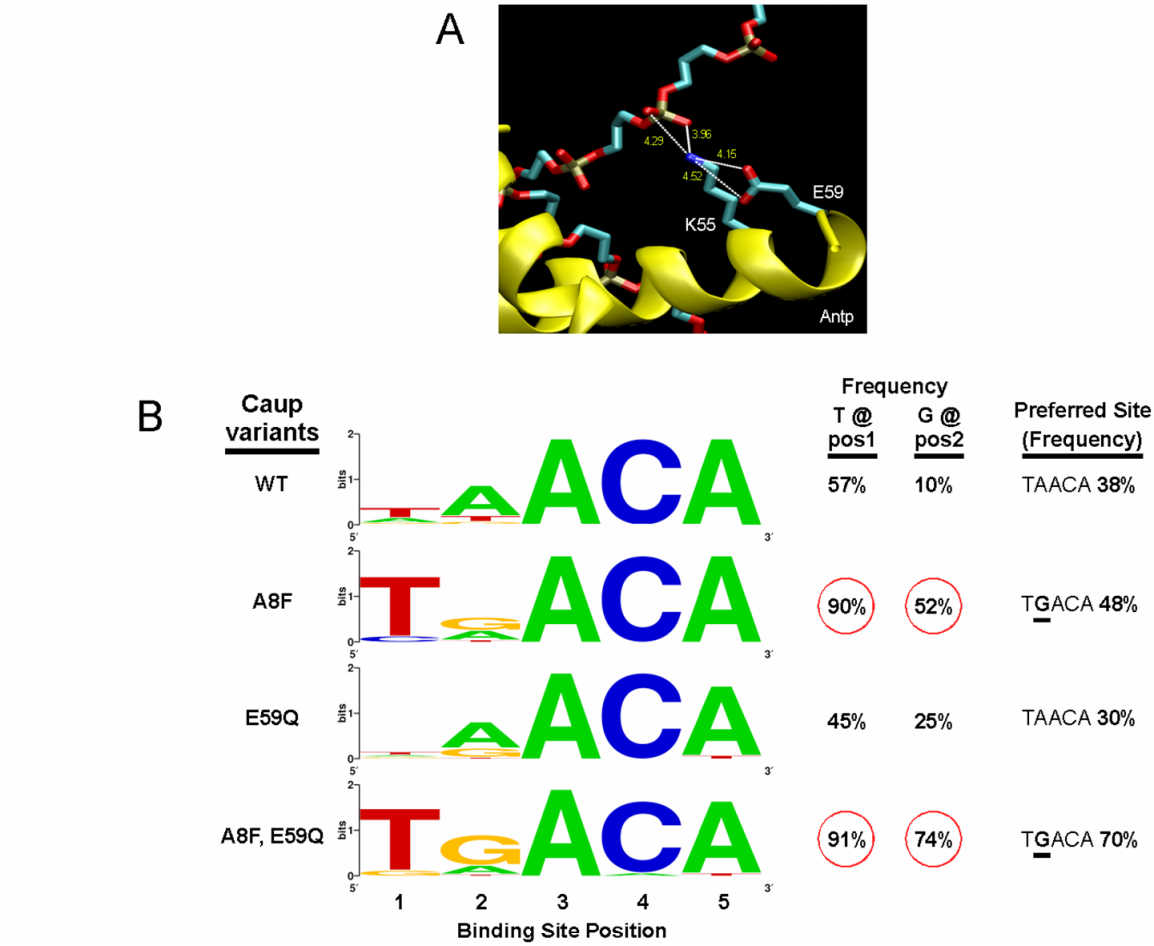


Figure S9. Glu59 may influence the specificity imparted by Arg55 in Caup. The presence of glutamate at position 59 of the homeodomain may also be influencing the degree of preference for Gua at position 2 of the binding site by capturing Arg55. A) A "cationic trap" is present in the Antp structure where Lys55 is captured against the phosphate backbone by Glu59 (Fraenkel and Pabo, 1998). Dotted lines indicate the distances between the lysine amino group and the negatively charged oxygens on the phosphodiester backbone and the acidic side-chain. The Iroquois family has a conserved glutamate at this position that could potentially capture Arg55 against the phosphodiester backbone (Supplementary Figure 11). B) Mutation of Glu59 to Gln in the presence or absence of the A8F mutation results in an increase in the frequency of recovery of Gua at position 2 in the binding site. However, this change in specificity is modest when compared to the effect of the A8F mutation on specificity.

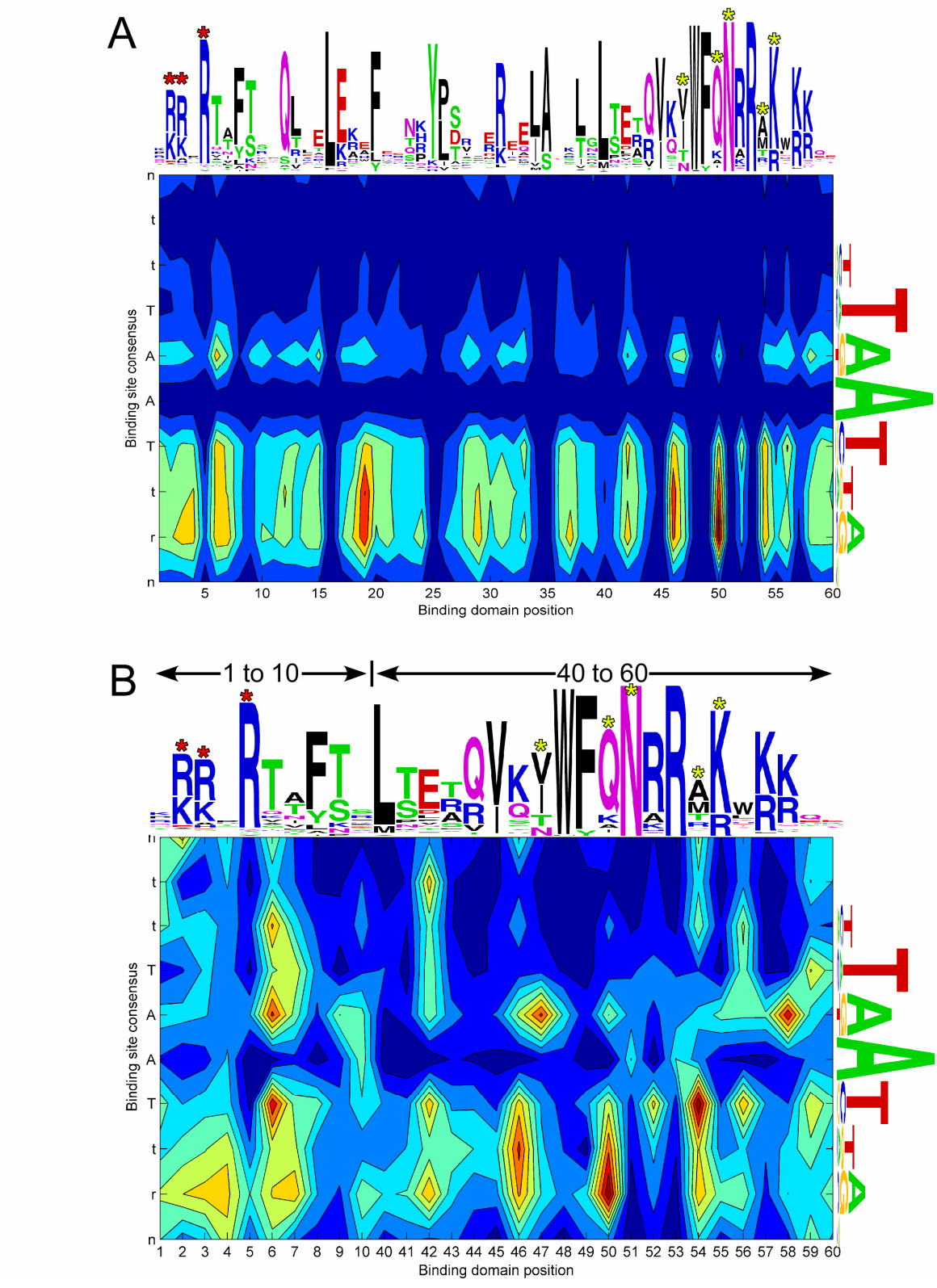


Figure S10. A) Heat plot of Mutual Information (MI) analysis between the amino acid sequences of the homeodomain and the master alignment of their selected binding sites. A two-dimensional heat plot is

shown where background levels of covariance are indicated in dark blue and positions with strong covariance are indicated in red. The x-axis indicates the position within the homeodomain (1 through 60), where the TALE insertions in the atypical family have been removed to achieve a common sequence framework. The amino acid diversity at each position is indicated by a Sequence logo above the plot. Key DNA-recognition positions are indicated by asterisks above the amino acid sequence. The y-axis indicates the position within the binding site (5' top, 3' bottom, with positions 1-4 capitalized). The overall motif of all of the binding sites is represented as a Sequence logo to the right of the plot. Overall the degree of covariation is higher at the 3' end of the binding site than for the 5' end of the site. Positions with little diversity in sequence composition have lower MI values, which is highlighted by the trough in the plot at the position corresponding to the Ade at position 3, which is highly conserved among the population of binding sites. Strong peaks are observed within the plot for expected base-amino acid contacts, such as between position 50 of the homeodomain and positions 5 and 6 of the binding site. B) Heat plot of the joint-ranked Mutual Information (jrMI) analysis. The background levels of covariance are indicated in dark blue and positions with strong covariation are indicated in red. Positions 1 through 10 and 40 through 60 in the homeodomain (x-axis) are shown to highlight the regions that are most likely involved in recognition.

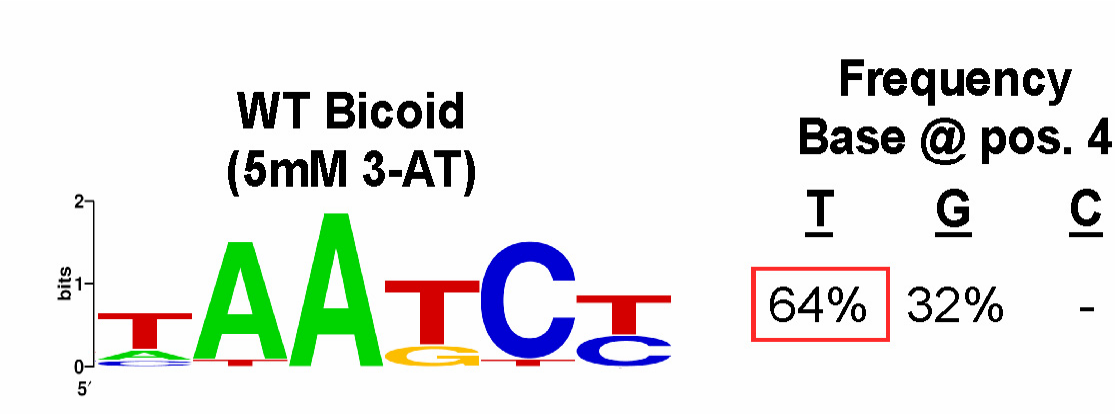


Figure S11. Binding site selection data for Bicoid collected from a selection performed at 5mM 3-AT instead of the typical stringency of 10 mM. The sequence preferences are preserved in this binding site selection except that Cyt and Thy are equally represented at position 6. There is increased tolerance for other bases at many positions, such as position 4 where Thy and Gua are both acceptable, but Cyt is still excluded from the binding sites at this position.

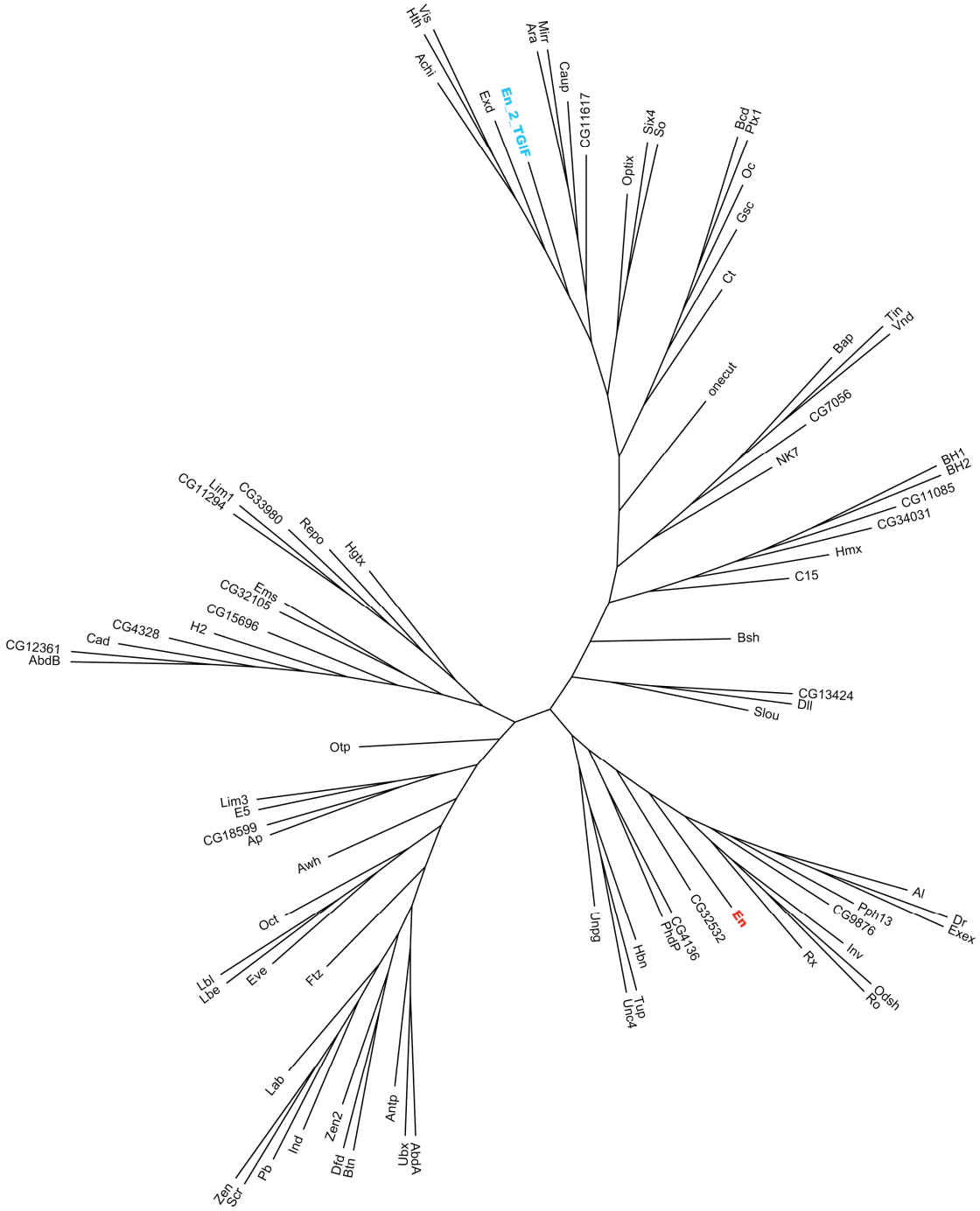


Figure S12. Clustergram of engineered En homeodomain that has TGIF-like specificity (**En_2_TGIF**; R3K, I47N, Q50A, A54R, K55R) with the entire set of fly homeodomains. The specificity of this factor clusters with the other members of the TGIF-Exd specificity group (Vis, Achi, Hth & Exd), which is significantly removed from its original position in the clustergram (**En**) with the En-group.

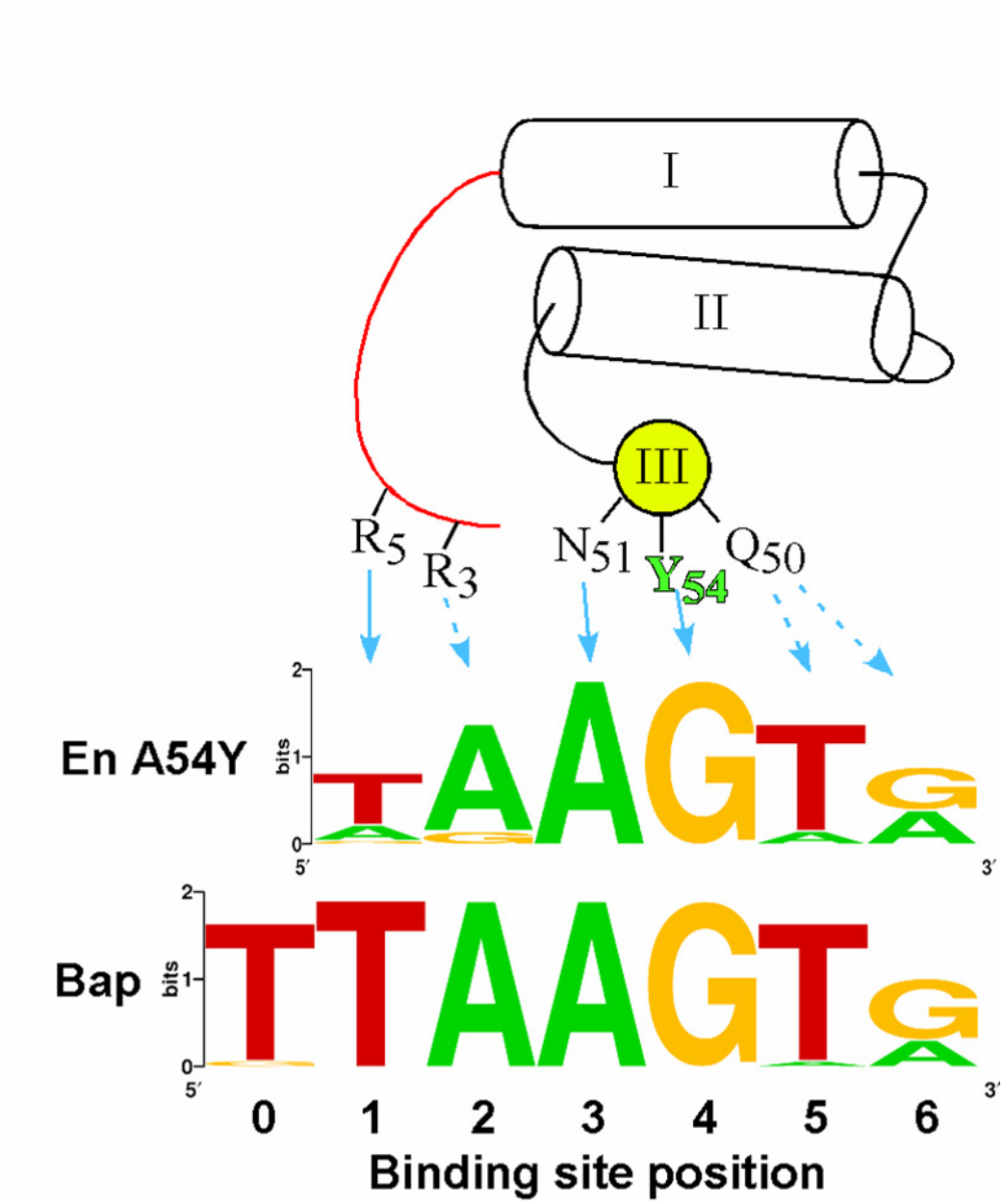
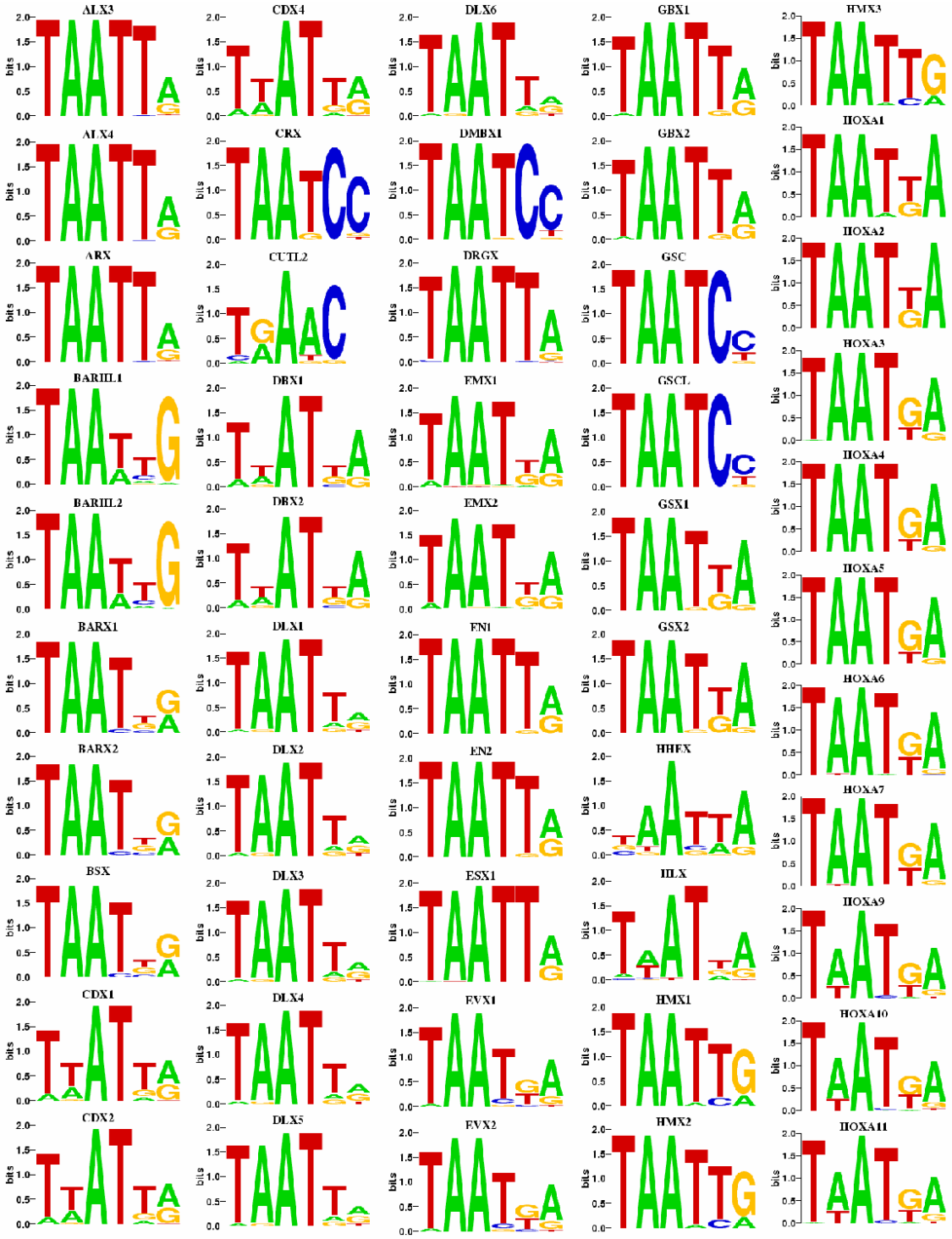
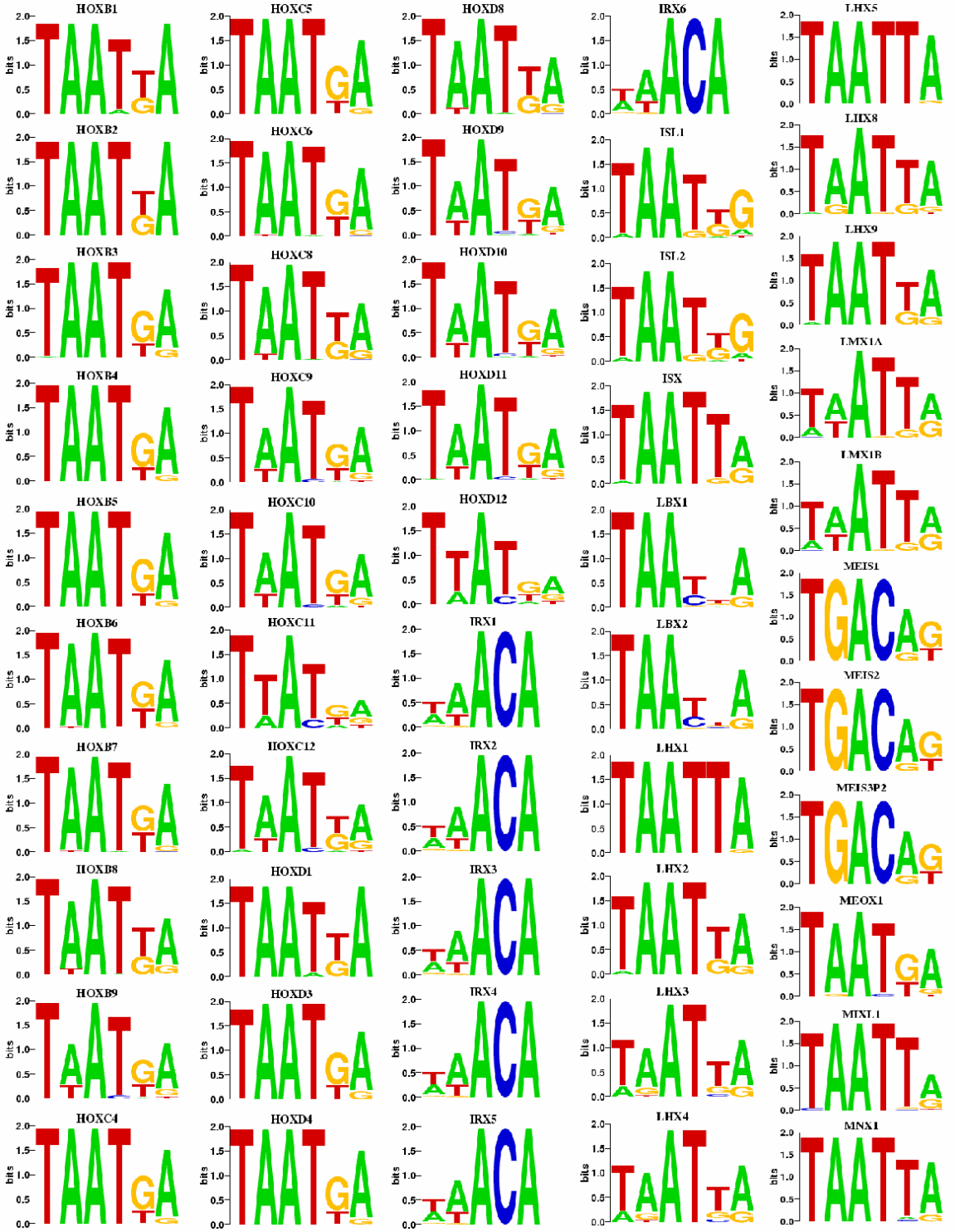


Figure S13. Conversion of Engrailed into a factor with an NK-2 type specificity. A single mutation was introduced into Engrailed at position 54 (A54Y) to attempt to convert this homeodomain into a specificity that is similar to the NK-2 class. This mutation has been introduced previously into Antp and Gsc with a resulting conversion of sequence preference at position 4 of the binding site from Thy to Gua (Damante et al., 1996; Pellizzari et al., 1997). We observe a similar phenomenon, where this single mutation is sufficient to convert the specificity of Engrailed at position 4 from Thy to Gua. For reference the specificity of Bap, a NK-2 family member, is shown below the specificity of the A54Y mutant.


ALX3
CDX4
DLX6
GBX1
HMX3
ALX4
CRX
DMBX1
GBX2
HOXA1
ARX
CUTL2
DRGX
GSC
HOXA2
BARHL1
DBX1
EMX1
GSCL
HOXA3
BARHL2
DBX2
EMX2
GSX1
HOXA4
BARX1
DLX1
FIN1
GSX2
HOXA5
BARX2
DLX2
EN2
HHEX
HOXA6
BSX
DLX3
ESX1
HLX
HOXA7
CDX1
DLX4
EVX1
HMX1
HOXA9
CDX2
DLX5
EVX2
HMX2
HOXA10
HOXA11
bits


HOXB1
HOXC5
HOXD8
IRX6
LHX5
HOXB2
HOXC6
HOXD9
ISL1
LHX8
HOXB3
HOXC8
HOXD10
ISL2
LHX9
HOXB4
HOXC9
HOXD11
ISX
LMX1A
HOXB5
HOXC10
HOXD12
LBX1
LMX1B
HOXB6
HOXC11
IRX1
LBX2
MEIS1
HOXB7
HOXC12
IRX2
LHX1
MEIS2
HOXB8
HOXD1
IRX3
LHX2
MEIS3P2
HOXB9
HOXD3
IRX4
LHX3
MEOX1
HOXC4
HOXD4
IRX5
LHX4
MNX1
bits

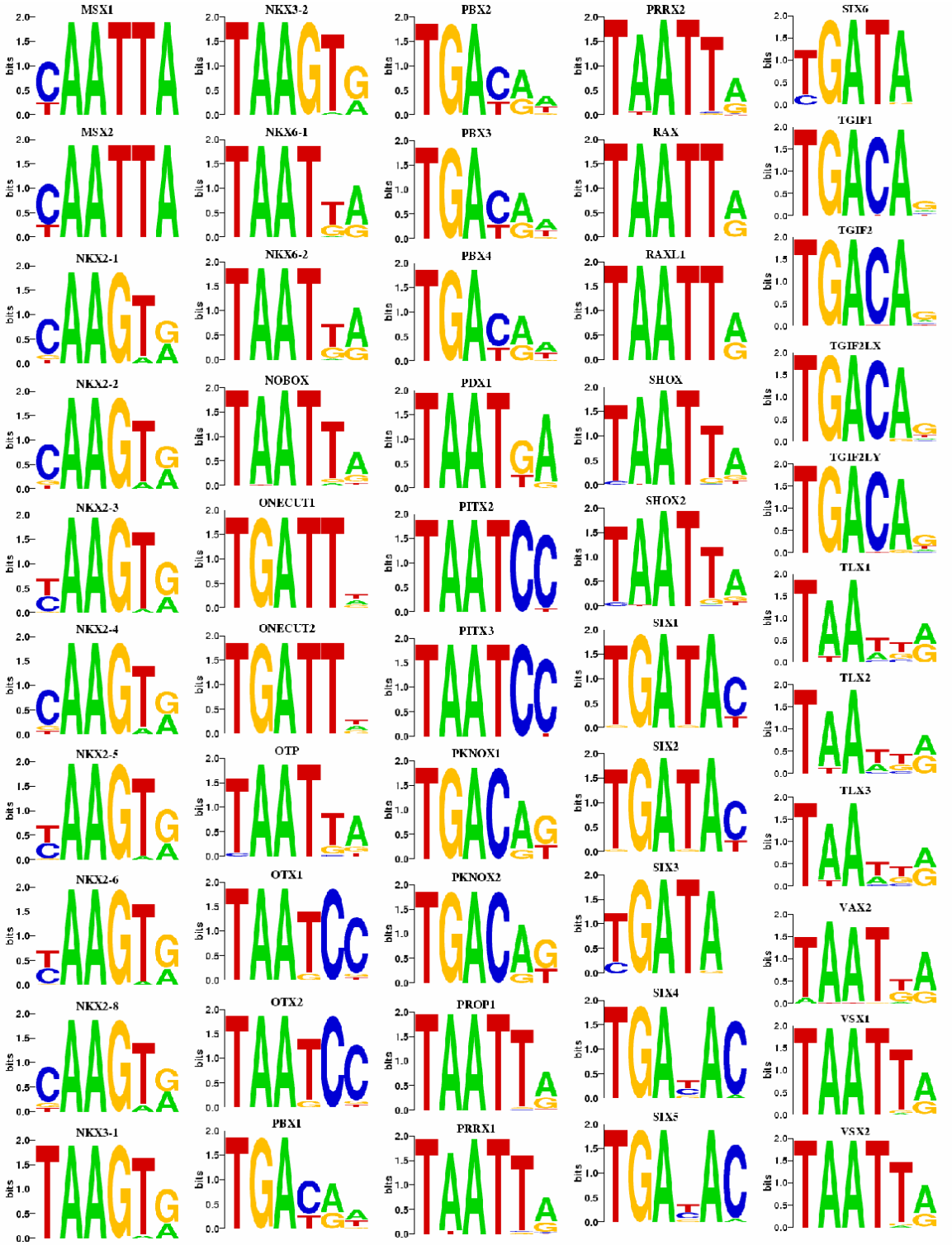
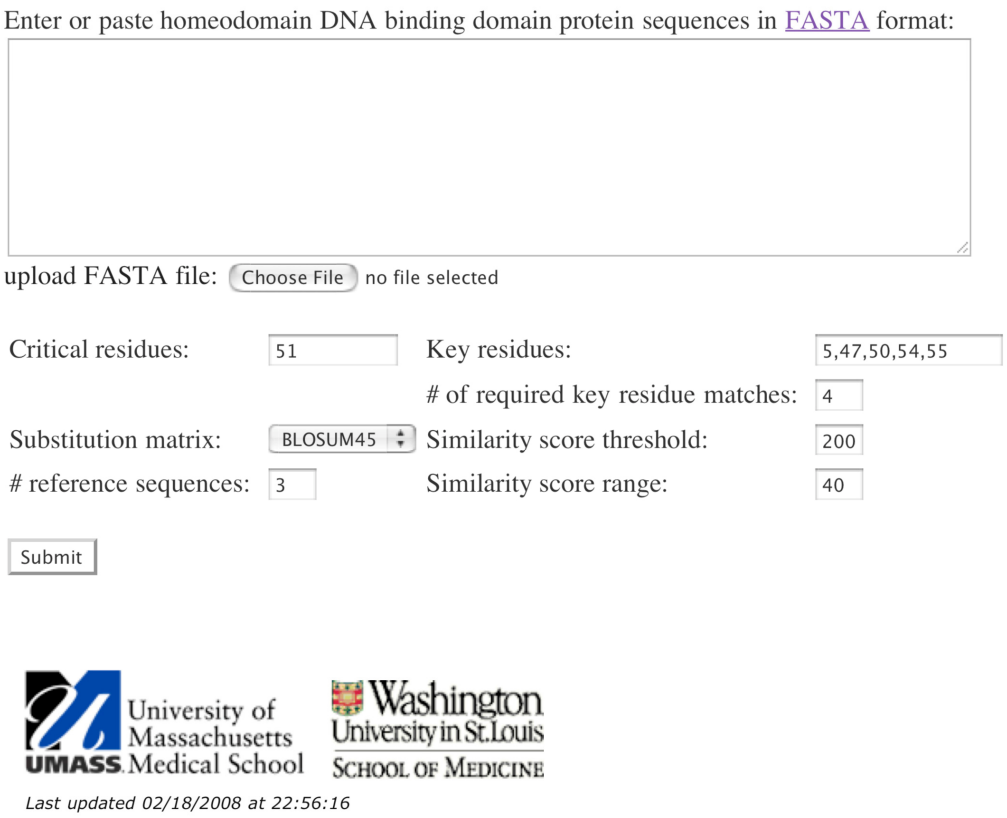


Figure S14. Sequence logos of the 153 predicted human homeodomains.

A



B

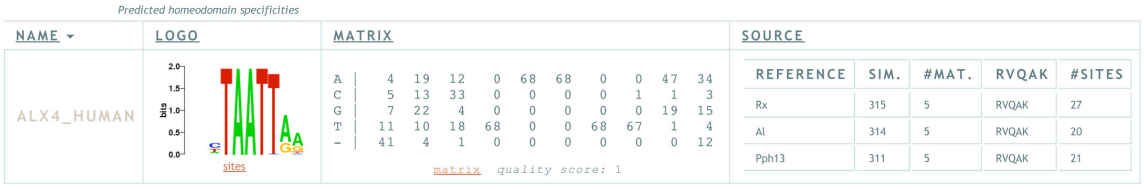


Figure S15. A. Splash page of interactive web-based homeodomain specificity prediction tool at http://ural.wustl.edu/flyhd/. Given a set of homeodomain protein sequences entered as a FASTA file, the program returns a set of predicted sites, a count matrix and a sequence logo for each query protein, as well as the set of reference proteins used to make each prediction. If the homology constraints are not met by any of the reference proteins, no prediction is made. Users can change the default parameters to define alternate sets of key and critical residues and different homology score constraints. B. Example of the information returned from the website upon entering the sequence of the Alx4 homeodomain. The sites used to construct the LOGO can be accessed by clicking the "sites" link. The quality score (1 to 4) under the MATRIX indicates the confidence level of the prediction where 1 is the highest confidence. Under SOURCE, the REFERENCE indicates the fly homeodomains used to construct the predicted LOGO where SIM.= similarity score, #MAT= number of matches at the key recognition positions, RVQAK are the amino acids at the key recognition positions in the query, with the corresponding residues of the utilized fly factors listed below, #SITES = number of unique sequences used from each fly factor.

Table S2. Orientation of the homeodomain on each recovered sequence is apparent from the position of its binding site

Selected sequences for Bicoid, Deformed and En (from the randomized region). All sequences are listed in the same orientation relative to the promoter of the reporter genes (toward the 3' side). Yellow highlight indicates the position of a binding site in the positive strand, magenta indicates the position of a binding site in the negative strand, and green indicates a palidromic site (orientation unknown). "Yellow" sites trend toward the 3' end of the sequence whereas "magenta" sites trend toward the 5' end of the sequence.

Bicoid selected sequences

>bcd1	TCTTAATCCC	>bcd13	TGTTAATCCC
TGTTAATCCG	>bcd7	TGTTAATCC	>bcd20
>bcd2	GCTTAATCCG	>bcd14	CGCTTAATCC
ATGGATTAGA	>bcd8	TGGGATTATA	>bcd21
>bcd3	GGGTTAATCC	>bcd15	TTACTAATCC
CGTTAATCTC	>bcd9	GCGTAATCCA	>bcd22
>bcd4	GAGATAATCC	>bcd16	GTCCTAATCC
GGTTTAATCC	>bcd10	GGCTTAAGCC	>bcd23
>bcd5	AGCTTATCC	>bcd17	GGTTAATCCG
TCTATAATCC	>bcd11	GGTTATCCG	>bcd24
>bcd6	CGGGTAATCC	>bcd18	ATGGATTAGA

Deformed selected sequences

>dfd1	>dfd7	>dfd13	>dfd19
CTTCATTAAG	GATAATTAAT	TCGTAATGA	TACCTAATGA
>dfd2	>dfd8	>dfd14	>dfd20
GGTCATTAAT	CCTAATTAAG	TGCTTAATGG	TGGATAATGA
>dfd3	>dfd9	>dfd15	>dfd21
TATCATTAAA	CCCCATTAAT	ATCGTAATTA	CGACTAATGA
>dfd4	>dfd10	>dfd16	>dfd22
GGTCATTAAT	TTTTTAATGA	CTCATTACT	TATCATTAAC
>dfd5	>dfd11	>dfd17	>dfd23
GTCATTAACA	AGCTATTAAA	CTTCATTAAG	CCGTTAATGA
>dfd6	>dfd12	>dfd18	>dfd24
CCTAATTAAG	GCACTAATGA	AGTCATTAGG	CAATTAATGA

Engrailed selected sequences

>En1A1	>En1A11	>En1B10	>En1*B2
CGCAATTAGA	TTAATTAGGT	TCAATTAAGG	ATTAATTATC
>En1A2	>En1A12	>En1B11	>En1*B3
GACTTAATGA	CCAATTATGT	GCTAATTAAT	GTAATTAGG
>En1A3	>En1B2	>En1*A1	>En1*B4
CTACTAATTG	TTAATTAGTA	GTGTTAATGA	AGCAATTAAG
>En1A4	>En1B3	>En1*A4	>En1*B5
CCAATTATTT	TTTTTAATTA	GCATAATTA	TGTAATTAGA
>En1A5	>En1B4	>En1*A5	>En1*B7
CCAATTACC	ATAATTAGTG	ACAATTATAA	TTTAATTAAG
>En1A6	>En1B5	>En1*A6	>En1*B8
CATAATTAAA	TCCAATTAAG	ATAATTAAAA	TAGTTAATTA
>En1A7	>En1B6	>En1*A8	>En1*B9
CACAATTAAC	CCAATTAGAT	ATCATTAACC	ATCAATTAAG
>En1A8	>En1B7	>En1*A9	>En1*B10
AGTAATTACC	TAATTAACA	GATAATTATC	CTCATTAAGA
>En1A9	>En1B8	>En1*A10	>En1*B11
CTAATTAGAG	GGTAATTAAA	TATCAACCCC	CTCATTAGTG
>En1A10	>En1B9	>En1*A11	
TTAATTAGAC	GGTAATTAAC	CCTCATTAAA	

Table S3. Master alignment of selected binding sites for Drosophila homeodomain family

Master alignment of selected binding sites. Sequences that were identified for each factor to contain an overrepresented motif based on CONSENSUS analysis were aligned at the common Ade at binding site position 3 that is shared by all Asn51 containing homeodomains.

```
>Vis:VisG1b:-:5:1
aaaTGACAc-
>Vis:VisG2b:-:6:2
actTGACA--
>Vis:VisG4b:-:5:4
aggTGACAt-
>Vis:VisG5b:+:4:5
-aaTGACAgc
>Vis:VisG6b:+:4:6
-gtTGACAcc
>Vis:VisG7b:-:6:7
agtTGACA--
>Vis:VisG8b:+:2:8
cgaTGACAg-
>Vis:VisG10b:-:4:10
tccTGACAgt
>Vis:VisG1:-:5:13
aaaTGACA--
>Vis:VisG2:-:5:14
ctcTGACAg-
>Vis:VisG3:-:5:15
ttaTGACAc-
>Vis:VisG4:-:6:16
cgtTGACA--
>Vis:VisG5:-:5:17
aagTGACAa-
>Vis:VisG6:+:4:18
-gaTGACAag
>Vis:VisG7:-:6:19
tttTGACA--
>Vis:VisG8:-:6:20
aagTGACA--
>Vis:VisG9:+:3:21
gttTGACAcg
>Vis:VisG10:+:4:22
-ttTGACAgt
>Vis:VisG11:+:4:23
-agTGACAgc
>Vis:VisG12:+:3:24
taaTGACAgc
>Vis:VisH2:-:5:26
agtTGACAt-
>Vis:VisH10:-:4:33
tccTGACAgt
>Achi:AchiG1:-:1:1
-CTTGACA--
>Achi:AchiG2:+:3:2
aTGTGACA--
>Achi:AchiG3:-:4:3
-CATGACAgc
>Achi:AchiG4:+:4:4
aTCTGACA--
>Achi:AchiG5:+:3:5
gCATGACAg-
>Achi:AchiG6:+:3:6
gTTTGACAt-
>Achi:AchiG8:-:4:8
-TATGACAgg
>Achi:AchiG9:-:3:9
tTTTGACAac
>Achi:AchiG10:+:4:10
aTCTGACA--
>Achi:AchiH1:-:2:11
aTTTGATAt-
>Achi:AchiH2:-:4:12
-TCTGACAgc
>Achi:AchiH3:+:3:13
aAATGACA--
>Achi:AchiH4:-:4:14
-ACTGACAga
>Achi:AchiH6:-:3:16
tTTTGACAgc
>Achi:AchiH11:-:3:19
cTTTGACAgc
>Achi:achi2F1:+:3:20
gAATGACAa-
>Achi:achi2F2:-:4:21
-TTTGACAac
>Achi:achi2F3:-:2:22
gATTGACAg-
>Achi:achi2F4:-:4:23
-TTTGACAgc
>Achi:achi2F6:+:4:25
aGATGACA--
>Achi:achi2F8:-:3:26
-TTTGACAgg
>Achi:achi2F11:-
:4:29
-TTTGACAgc
>Achi:achi2F12:-
:3:30
gTTTGACAgg
>Hth:hthC2:+:5:1
ttgTGACAg-
>Hth:hthC3:-:4:2
-ttTGACAta
>Hth:hthC4:-:2:3
cctTGACGt-
>Hth:hthC5:-:3:4
ctgTGACAgg
>Hth:hthC6:+:4:5
ctgTGACGt-
>Hth:hthC8:-:1:6
-gaTGACA--
>Hth:hthC9:-:4:7
-gcTGACAgt
>Hth:hthC10:+:6:8
accTGACG--
>Hth:hthC11:+:5:9
cgaTGACAg-
>Hth:hthC12:-:1:10
gttTGACA--
>Hth:hthD1:-:3:11
cagTGACAgg
>Hth:hthD2:-:4:12
-ttTGACAgc
>Hth:hthD3:+:5:13
attTGACAt-
>Hth:hthD4:+:5:14
aggTGACAg-
>Hth:hthD6:-:4:16
-aaTGACAgc
>Hth:hthD9:-:1:19
caaTGACA--
>Hth:hthD11:+:6:21
tcgTGACA--
>Six4:5Six4E1:+:1:1
ATTTGATAC-
>Six4:5Six4E2:+:1:2
CTTTGAGAC-
>Six4:5Six4E3:+:1:3
ATTTGATAC-
>Six4:5Six4E4:+:2:4
ATCTGACACc
>Six4:5Six4E5:+:1:5
AAATGAGAC-
>Six4:5Six4E6:+:2:6
TCTTGATACc
>Six4:5Six4E8:+:1:8
AGATGATAC-
>Six4:5Six4E9:+:1:9
TTTTGAGAA-
>Six4:5Six4E10:+:2:1
0
AGTTGACACc
>Six4:5Six4E11:+:2:1
1
GTATGATACc
>Six4:5Six4E12:+:2:1
2
GTCTGACACc
>Six4:5Six4F1:+:2:13
ACATGACACc
>Six4:5Six4F2:+:1:14
AAGTGATAC-
>Six4:5Six4F3:+:2:15
AATTGACACc
>Six4:5Six4F4:+:2:16
CTGTGACACc
>Six4:5Six4F5:+:2:17
TAATGATACc
>Six4:5Six4F6:+:2:18
TGATGACACc
>Six4:5Six4F7:+:2:19
TTTTGATACc
>Six4:5Six4F10:+:1:2
1
ATTTGAGAC-
>Six4:5Six4F11:+:1:2
2
TTGTGATAC-
>Optix:11617C4:+:1:3
AAGTGATA--
>Optix:11617C6:+:1:5
ATGTGATA--
>Optix:11617C7:+:1:6
CAGTGATA--
>Optix:11617C8:+:1:7
TTGCGATA--
>Optix:11617C9:-:1:8
TAGCGATG--
>Optix:11617C11:+:1:
10
TAGTGATA--
>Optix:11617D8:+:1:1
7
AAATGATA--
>Optix:11617D9:+:1:1
8
TAGTGATA--
>Optix:11617D10:+:1:
19
AAGTGATA--
>Optix:11617D11:+:1:
20
TAGTGATA--
>Optix:11617D12:+:1:
21
AGATGATA--
>Optix:OptixE8:+:1:2
9
AAGTGATA--
>Optix:Optix2A2:+:1:
34
CCATGATA--
>Optix:Optix2A3:+:1:
35
ATGTGATA--
>Optix:Optix2A4:+:1:
36
AGCTGATA--
>Optix:Optix2A5:+:1:
37
ATCTGATA--
>Optix:Optix2A6:+:1:
38
AGCTGATA--
>Optix:Optix2A10:+:1
:41
CCGCGATA--
>Optix:Optix2A11:+:1
:42
GATTGATA--
>Optix:Optix2A12:+:1
:43
CCTTGATA--
>Optix:Optix2B1:+:1:
44
TCGCGATA--
>Optix:Optix2B3:+:1:
46
TTGCGATA--
>Optix:Optix2B4:+:1:
47
GATTGATA--
>Optix:Optix2B5:+:1:
48
AAGTGATA--
>Optix:Optix2B6:+:1:
49
AAGTGATA--
>Optix:Optix2B7:+:1:
50
TTATGATA--
>Optix:Optix2B8:+:1:
51
TTCTGATA--
>So:SoE1:-:2:1
cTGTGATAc-
>So:SoE2:+:4:2
cGATGATA--
>So:SoE3:-:3:3
tATTGATAct
>So:SoE5:-:3:5
gAATGATAct
>So:SoE6:+:4:6
aAATGATA--
>So:SoE7:-:2:7
tTGTGATAc-
>So:SoE8:-:2:8
aTCTGATAc-
>So:SoE9:+:4:9
tAGTGATA--
>So:SoE11:-:2:10
tTTTGATAc-
>So:SoE12:+:4:11
gGATGATA--
>So:SoF2:+:4:12
tGATGATA--
>So:SoF4:-:2:14
aAATGATAc-
>So:SoF5:-:2:15
gAGGGATAt-
>So:SoF7:-:2:17
aAATGATAc-
>So:SoF9:+:4:19
gCATGAGA--
>So:SoF11:-:2:21
aAGTGATAc-
>So:So2E1:+:4:22
tCGTGATA--
>So:So2E2:+:4:23
tATTGATA--
>So:So2E3:-:3:24
cAGTGATAtg
>So:So2E4:+:4:25
gCATGATA--
>So:So2E5:-:3:26
```

gTGTGATAcc
>So:So2E6:+:4:27
gGATGATA--
>So:So2E7:+:4:28
cAGTGATA--
>So:So2E9:-:3:30
tTGTGATAtg
>So:So2E10:+:4:31
aCATGATA--
>So:So2E11:+:4:32
tGATGATA--
>So:So2E12:+:4:33
cAATGATA--
>Exd:ExdE1:+:1:1
GTTTGACAt-
>Exd:ExdE3:+:1:3
CTTTGACAt-
>Exd:ExdE5:+:2:5
TTTTGACA--
>Exd:ExdE6:+:2:6
ATTTGACA--
>Exd:ExdE7:+:1:7
CTTTGATGa-
>Exd:ExdE8:+:1:8
GATTGATGa-
>Exd:ExdE9:+:2:9
AGTTGACA--
>Exd:ExdE10:+:1:10
TTTTGACGa-
>Exd:ExdE12:+:2:12
TTTTGACA--
>Exd:ExdF1:+:1:13
AGTTGACAt-
>Exd:ExdF2:+:1:14
TTTTGATGg-
>Exd:ExdF3:+:1:15
CTTTGATGa-
>Exd:ExdF4:+:2:16
ATTTGACA--
>Exd:ExdF5:+:2:17
GTTTGACA--
>Exd:ExdF6:+:1:18
GTTTGATGa-
>Exd:ExdF7:+:2:19
TTTTGACA--
>Exd:ExdF9:+:2:20
CTTTGACA--
>onecut:5onectG1:+:4
:1
agCTGATTA-
>onecut:5onectG2:-
:3:2
-cTTGATTTc
>onecut:5onectG3:-
:2:3
ccTTGATTGc
>onecut:5onectG4:-
:2:4
taTTGATTAa
>onecut:5onectG8:+:4
:8
ctATGATTA-
>onecut:5onectG10:-
:2:9
taTTGATTAa
>onecut:5onectG11:-
:3:10
-aTTGATTTg
>onecut:5onectG12:-
:3:11
-tCTGATTGa
>onecut:5onectH1:+:4
:12
agTTGATTG-
>onecut:5onectH2:-
:3:13
-aTTGATTTg
>onecut:5onectH4:-
:3:15
-cTTGATTTg
>onecut:5onectH5:+:5
:16
tcTTGATTT-
>onecut:5onectH7:-
:3:18
-gTTGATTTg
>onecut:5onectH10:-
:3:21
-tTTGATTGg
>onecut:5onectH11:-
:2:22
cgTTGATTAg
>Vnd:VndE2:+:1:1
TCTCAAGTG-
>Vnd:VndE3:+:1:2
TCTCAAGTA-
>Vnd:VndE5:+:1:3
ATTGAAGTA-
>Vnd:VndE7:+:1:4
GGTCAAGTA-
>Vnd:VndE8:+:1:5
TCTCAAGTA-
>Vnd:VndE9:+:1:6
TCTCAAGTG-
>Vnd:VndE10:+:1:7
TTTTAAGTA-
>Vnd:VndE11:+:1:8
TCTCAAGTG-
>Vnd:VndE12:+:1:9
ATTGAAGTA-
>Vnd:VndF1:+:1:10
GTTCAAGAG-
>Vnd:VndF2:+:1:11
TTTCAAGTG-
>Vnd:VndF4:+:1:13
TCTTAAGTA-
>Vnd:VndF5:+:1:14
TATCAAGAG-
>Vnd:VndF6:+:1:15
TCTCAAGTA-
>Vnd:VndF7:+:1:16
ACTCAAGTG-
>Vnd:VndF8:+:1:17
TTTCAAGTG-
>Vnd:VndF9:+:1:18
TTTCAAGTA-
>Vnd:VndF10:+:1:19
TTTCAAGTG-
>Vnd:VndF12:+:1:21
ATTCAAGTG-
>Tin:TinA2:-:3:1
gCTCAAGTA-
>Tin:TinA3:+:3:2
-GTCAAGTAc
>Tin:TinA4:+:2:3
cGTCAAGTGg
>Tin:TinA5:-:3:4
tCTCAAGTG-
>Tin:TinA6:+:2:5
aTTCAAGTGg
>Tin:TinA7:+:3:6
-GTCAAGTGc
>Tin:TinA8:+:3:7
-CTTAAGTAc
>Tin:TinA9:+:3:8
-CTTAAGTGg
>Tin:TinA10:+:3:9
-TTTAAGTGg
>Tin:TinA11:+:2:10
cCACAAGTGg
>Tin:TinB1:+:3:11
-CTCAAGTGg
>Tin:TinB3:+:3:12
-CTTAAGTGt
>Tin:TinB4:+:2:13
-TTCAAGTGg
>Tin:TinB5:+:3:14
-CTCAAGTGg
>Tin:TinB6:+:3:15
-CTCAAGTGc
>Tin:TinB9:+:3:18
-TTGAAGTGg
>Bap:BapA2:-:1:1
GCTTAAGTGG
>Bap:BapA3:-:1:2
TCTTAAGTGG
>Bap:BapA12:-:1:10
AGTTAAGTGG
>Bap:BapB1:-:1:11
GTTTAAGTGG
>Bap:BapB5:-:1:14
TGTTAAGTGG
>Bap:BapB6:-:1:15
TCTTAAGTGG
>Bap:BapB7:-:1:16
ACTTAAGTAC
>Bap:BapB9:-:1:18
TGTTAAGTGG
>Bap:BapB11:-:1:20
CGTTAAGTGG
>Bap:Bap2C1:+:1:21
CTTTAAGTGT
>Bap:Bap2C2:-:1:22
ACTTAAGTAC
>Bap:Bap2C3:-:1:23
TTTTAAGTGT
>Bap:Bap2C4:-:1:24
CTGTAAGTGT
>Bap:Bap2C6:-:1:26
GCTTAAGTGC
>Bap:Bap2C7:-:1:27
ACTTAAGAAC
>Bap:Bap2C8:-:1:28
TCTTAAGTAC
>Bap:Bap2C9:-:1:29
ATTTAAGTGA
>Bap:Bap2C10:-:1:30
TATTAAGTAC
>Bap:Bap2C12:-:1:31
TTTTAAGTGA
>Bap:Bap2D2:-:1:32
CTTTAAGTAC
>Bap:Bap2D3:-:1:33
GGTTAAGTGG
>Bap:Bap2D10:-:1:38
ATTTAAGTGA
>Bap:Bap2D11:-:1:39
ACTTAAGTAC
>CG7056:CG7065A1:+:3
:1
tTTTGATAA-
>CG7056:CG7065A4:+:3
:2
cTTTAATAA-
>CG7056:CG7065A5:+:3
:3
cTCTAATTA-
>CG7056:CG7065A6:+:3
:4
gCTCAACAA-
>CG7056:CG7065A7:+:3
:5
aATGAATTG-
>CG7056:CG7065A8:+:3
:6
gGTTAATTA-
>CG7056:CG7065A9:+:3
:7
aCTGAAGTA-
>CG7056:CG7065A10:+:
3:8
aTTTAATGA-
>CG7056:CG7065B1:+:3
:10
aGTCAATTA-
>CG7056:CG7065B5:+:3
:13
tTTCGATTA-
>CG7056:CG7065B6:+:3
:14
aTTGTACTA-
>CG7056:CG7065B7:+:3
:15
aACCAATAA-
>CG7056:CG7065B8:+:3
:16
aTTGTACTA-
>CG7056:CG7065B9:+:3
:17
tTTTAATAA-
>CG7056:CG7065B11:+:
2:19
cATTAATAAc
>Ubx:UbxC02:+:3:1
aATTAATTA-
>Ubx:UbxC03:+:3:2
aTTTTATTA-
>Ubx:UbxC04:+:3:3
cTTTAATTA-
>Ubx:UbxC05:+:3:4
cTTTAATTA-
>Ubx:UbxC06:+:3:5
tATTAATTA-
>Ubx:UbxC07:-:3:6
-TTTTATTGc
>Ubx:UbxC08:-:3:7
-CTTAATTGc
>Ubx:UbxC11:+:3:8
tATTAATGA-
>Ubx:UbxC12:+:3:9
aTTTAATGG-
>Ubx:UbxD01:+:2:10
tCTTAATGA-
>Ubx:UbxD02:+:3:11
cGTTAATTA-
>Ubx:UbxD04:+:3:12
cCTTAATTA-
>Ubx:UbxD05:-:3:13
-TTTAATTGt
>Ubx:UbxD06:+:3:14
cGTTAATTA-
>Ubx:UbxD07:-:3:15
-TTTAATGGg
>Ubx:UbxD08:+:3:16
cCTTAATGA-
>Ubx:UbxD09:+:3:17
cCTTAATTA-
>Ubx:UbxD10:+:3:18
tTTTTATGA-
>Ubx:UbxD11:-:3:19
-GTTAATTAc
>Ubx:UbxD12:-:3:20
-TTTAATTGc
>C15:C15A1:-:3:1
aTTTAAAGA-
>C15:C15A2:+:3:2
-GTTAATTGg
>C15:C15A3:+:3:3
-GTTAATGGg
>C15:C15A4:+:3:4
-CTTAAACGa
>C15:C15A5:+:3:5
-CTTAACGAg
>C15:C15A6:-:1:6
-GTTTAACA-
>C15:C15A7:+:1:7
cGTTAAACA-
>C15:C15A8:-:3:8
cGTTAACGA-
>C15:C15A9:-:3:9
tCTTAATTG-
>C15:C15A10:-:3:10
cGTTAATGA-
>C15:C15A11:+:3:11
-CTTAAATGa
>C15:C15A12:-:3:12
aTTTAATTG-
>C15:C15B1:+:3:13
-GTTAATTGg
>C15:C15B2:+:3:14
-AATAATGGg
>C15:C15B4:+:3:15

```
-TTTAATTAg
>C15:C15B6:-:2:16
aTCTTATTAc
>C15:C15B8:+:3:18
-ATTAATTAg
>C15:C15B9:+:3:19
-TTTAAATAg
>C15:C15B11:+:3:21
-GTTAATTGc
>Mirr:MirrE1:+:1:1
TTGTAACA--
>Mirr:MirrE2:+:1:2
CAGTAACA--
>Mirr:MirrE3:+:1:3
TAAATACA--
>Mirr:MirrE4:+:1:4
AGCAAACA--
>Mirr:MirrE5:+:1:5
CTATAACA--
>Mirr:MirrE6:+:1:6
TAATAACA--
>Mirr:MirrE7:+:1:7
TCGTGACA--
>Mirr:MirrE8:+:1:8
AAGTTACA--
>Mirr:MirrE9:+:1:9
AACTTACA--
>Mirr:MirrE10:+:1:10
AAAAAACA--
>Mirr:MirrE11:+:1:11
AGAAAACA--
>Mirr:MirrE12:+:1:12
AAAAAACA--
>Mirr:MirrF2:+:1:14
TACTTACA--
>Mirr:MirrF3:+:1:15
AGAAAACA--
>Mirr:MirrF4:+:1:16
GAAAAACA--
>Mirr:MirrF5:+:1:17
TGAAAACA--
>Mirr:MirrF6:+:1:18
GCCTGACA--
>Mirr:MirrF7:+:1:19
AATTTACA--
>Mirr:MirrF8:+:1:20
TCAAAACA--
>Mirr:MirrF9:+:1:21
AGAAAACA--
>Mirr:MirrF10:+:1:22
ACAAAACA--
>Mirr:MirrF11:+:1:23
AAATAACA--
>Mirr:Mirr2C2:+:1:24
CAGAAACA--
>Mirr:Mirr2C3:+:1:25
AACAAACA--
>Mirr:Mirr2C4:+:1:26
GTACTACA--
>Mirr:Mirr2C5:+:1:27
TTCGAACA--
>Mirr:Mirr2C7:+:1:28
ATTTAACA--
>Mirr:Mirr2C8:+:1:29
TTGTAACA--
>Mirr:Mirr2C9:+:1:30
AGAAAACA--
>Mirr:Mirr2C11:+:1:3
2
TGAAAACA--
>Mirr:Mirr2C12:+:1:3
3
GATATACA--
>Mirr:Mirr2D1:+:1:34
CTCTTACA--
>Mirr:Mirr2D2:+:1:35
AAATTACA--
>Mirr:Mirr2D3:+:1:36
ATATAACA--
>Mirr:Mirr2D4:+:1:37
GTTAAACA--
>Mirr:Mirr2D5:+:1:38
CATAAACA--
>Mirr:Mirr2D7:+:1:40
TGAAAACA--
>Mirr:Mirr2D8:+:1:41
AGTTTACA--
>Mirr:Mirr2D9:+:1:42
AGAAAACA--
>Mirr:Mirr2D10:+:1:4
3
CAATAACA--
>Mirr:Mirr2D11:+:1:4
4
GAATAACA--
>Ara:AraG1:+:1:1
GTATTACA--
>Ara:AraG2:-:1:2
TAAGTACA--
>Ara:AraG4:+:1:3
AGTTTACA--
>Ara:AraG5:+:1:4
TAATTACA--
>Ara:AraG7:+:1:6
GGAAAACA--
>Ara:AraG8:+:1:7
GTATTACA--
>Ara:AraG11:+:1:9
GATATACA--
>Ara:AraG12:+:1:10
TTAGAACA--
>Ara:AraH3:+:1:11
GAATAACA--
>Ara:AraH5:+:1:12
ACATAACA--
>Ara:AraH6:+:1:13
TGTAAACA--
>Ara:AraH7:-:1:14
GCTGAACA--
>Ara:AraH8:+:1:15
CAAAAACA--
>Ara:AraH9:+:1:16
ACATAACA--
>Ara:AraH10:+:1:17
CAAAAACA--
>Ara:AraH11:+:1:18
CTTTTACA--
>Ara:Ara2G1:+:1:19
CCCAAACA--
>Ara:Ara2G2:+:1:20
AATTAACA--
>Ara:Ara2G3:+:1:21
AAGTTACA--
>Ara:Ara2G4:+:1:22
AGAGAACA--
>Ara:Ara2G5:+:1:23
TATAAACA--
>Ara:Ara2G6:+:1:24
GAAAAACA--
>Ara:Ara2G7:+:1:25
TAATAACA--
>Ara:Ara2G8:+:1:26
GTAGAACA--
>Ara:Ara2G10:+:1:27
TGAAAACA--
>Ara:Ara2G12:+:1:29
AATTAACA--
>Ara:Ara2H1:+:1:30
AAAAAACA--
>Ara:Ara2H2:+:1:31
CCAAAACA--
>Ara:Ara2H4:+:1:33
AAATTACA--
>Ara:Ara2H6:+:1:34
GAAAAACA--
>Ara:Ara2H8:+:1:35
ACATTACA--
>Ara:Ara2H9:+:1:36
ACAAAACA--
>Ara:Ara2H10:+:1:37
AGAAAACA--
>Ara:Ara2H11:+:1:38
GAATTACA--
>Caup:CaupA2:+:1:1
TATTAACA--
>Caup:CaupA3:+:1:2
CTTTTACA--
>Caup:CaupA4:+:1:3
CAAAAACA--
>Caup:CaupA5:+:1:4
GTTGTACA--
>Caup:CaupA6:+:1:5
CTATTACA--
>Caup:CaupA7:+:1:6
CATTAACA--
>Caup:CaupA8:+:1:7
AGTTAACA--
>Caup:CaupA9:+:1:8
AAAGAACA--
>Caup:CaupA10:+:1:9
CTTTAACA--
>Caup:CaupA11:+:1:10
AAATGACA--
>Caup:CaupA12:+:1:11
CTAAAACA--
>Caup:CaupB1:+:1:12
TAATAACA--
>Caup:CaupB2:+:1:13
TTGTGACA--
>Caup:CaupB3:+:1:14
CGGAAACA--
>Caup:CaupB4:+:1:15
GGTTAACA--
>Caup:CaupB5:+:1:16
TAATAACA--
>Caup:CaupB9:+:1:19
CAGCAACA--
>Caup:CaupB10:+:1:20
CTGTAACA--
>Caup:CaupB11:+:1:21
AAAAAACA--
>CG11617:LagG2:+:3:1
TTTTTACA--
>CG11617:LagG3:+:3:2
ATTTAACA--
>CG11617:LagG5:+:3:3
TTTTAACA--
>CG11617:LagG6:+:3:4
TTTTAACA--
>CG11617:LagG8:+:3:5
GTTTAACA--
>CG11617:LagG9:+:3:6
AATTTACA--
>CG11617:LagG10:+:3:
7
ATTTAACA--
>CG11617:LagG11:+:3:
8
TTTTGACA--
>CG11617:LagG12:-
:2:9
TTTTGACAt-
>CG11617:LagH1:+:3:1
0
ATTTAACA--
>CG11617:LagH2:+:3:1
1
GTTTTACA--
>CG11617:LagH3:+:3:1
2
ATTTAACA--
>CG11617:LagH4:+:3:1
3
TTTTTACA--
>CG11617:LagH5:+:1:1
4
GTTTAACA--
>CG11617:LagH6:+:3:1
5
AAATAACA--
>CG11617:LagH8:+:3:1
6
AGTTGACA--
>CG11617:LagH11:+:3:
18
TTTTAACA--
>Ct:5CtE1:+:1:1
TGCTAAAC--
>Ct:5CtE2:+:1:2
GTCTGAAC--
>Ct:5CtE3:+:1:3
GGTTAAAC--
>Ct:5CtE4:+:1:4
CGTTAATC--
>Ct:5CtE5:+:1:5
CATTGAAC--
>Ct:5CtE6:+:1:6
GATTAAAC--
>Ct:5CtE7:+:1:7
TCTTGAAC--
>Ct:5CtE9:+:1:9
AGTTAAAC--
>Ct:5CtE10:+:1:10
TGCTAATC--
>Ct:5CtE11:+:1:11
GTTCAAAG--
>Ct:5CtE12:+:1:12
GCTCGAGC--
>Ct:5CtF1:+:1:13
TCCTGAAC--
>Ct:5CtF3:+:1:15
AATTGAAC--
>Ct:5CtF5:+:1:17
TTCTGAAC--
>Ct:5CtF6:+:1:18
CCTTGAAC--
>Ct:5CtF7:+:1:19
ACTTAAAC--
>Ct:5CtF8:+:1:20
GGTTGAAC--
>Ct:5CtF9:+:1:21
GCTAAAAC--
>Ct:5CtF10:+:1:22
TCTTGAAC--
>Ct:5CtF11:+:1:23
TCCTGAAC--
>CG15696:CG15696G1:-
:3:1
cCTTAATTA-
>CG15696:CG15696G2:+
:3:2
-TTTAATGAg
>CG15696:CG15696G3:+
:3:3
-GTCAATTGg
>CG15696:CG15696G5:-
:3:4
gTTTTATTA-
>CG15696:CG15696G6:+
:2:5
cTCTAATTGg
>CG15696:CG15696G7:+
:2:6
tTGTAATTGg
>CG15696:CG15696G8:+
:3:7
-CTAGATTGa
>CG15696:CG15696G9:-
:1:8
-TCTAATTAg
>CG15696:CG15696G10:
+:3:9
-TATTATTAa
>CG15696:CG15696G11:
+:3:10
-CTCAATTGg
>CG15696:CG15696G12:
-:1:11
-GTTAATTAg
>CG15696:CG15696H2:+
:3:13
-GTTGATTGa
>CG15696:CG15696H3:+
:2:14
```

```
-CTTGATGGc
>CG15696:CG15696H4:+
:3:15
-CTTGATTGg
>CG15696:CG15696H5:+
:3:16
-TTTAATTGa
>CG15696:CG15696H7:+
:3:17
-ATTAATTAg
>CG15696:CG15696H8:+
:3:18
-GTTAATTGa
>CG15696:CG15696H9:+
:3:19
-GTTAATTAa
>CG15696:CG15696H10:
+:3:20
-CTTAATTAg
>CG15696:CG15696H11:
+:3:21
-GTTAATTGg
>CG15696:156962G1:+:
3:22
-TGTAATTAc
>CG15696:156962G2:-
:3:23
gGGTAATTG-
>CG15696:156962G3:+:
3:24
-TTTGATTAt
>CG15696:156962G4:+:
3:25
-TATTATTAg
>CG15696:156962G5:-
:1:26
-CTTAATTAc
>CG15696:156962G6:-
:3:27
cGCTAATTA-
>CG15696:156962G7:-
:1:28
-TTCAATTAc
>CG15696:156962G8:+:
3:29
-TTTTATTGc
>CG15696:156962G9:+:
2:30
-ATTGATTGa
>CG15696:156962G10:+
:3:31
-ATTAATTAg
>CG15696:156962G11:+
:3:32
-TTTTATTGg
>CG15696:156962G12:+
:2:33
-TTAAATTGg
>CG4328:5CG4328E1:+:
2:1
tGCTTATTGc
>CG4328:5CG4328E2:-
:3:2
aATATATTA-
>CG4328:5CG4328E3:+:
3:3
aCATTATGAa
>CG4328:5CG4328E4:-
:3:4
tCGTTATTG-
>CG4328:5CG4328E5:+:
2:5
tAGATATTGc
>CG4328:5CG4328E6:+:
4:6:mod
cctCAATTAT
>CG4328:5CG4328E7:+:
3:7
-ATTTATTGa
>CG4328:5CG4328E8:-
:3:8
cTATAATTA-
>CG4328:5CG4328E9:+:
3:9
-AATAATGAc
>CG4328:5CG4328E10:+
:1:10
tATTAATGG-
>CG4328:5CG4328E11:+
:3:11
-TTATATTAc
>CG4328:5CG4328E12:+
:3:12
-CTTAATTAg
>CG4328:5CG4328F1:+:
3:13
-TCTAATTGt
>CG4328:5CG4328F2:+:
2:14
gTATAATTGg
>CG4328:5CG4328F4:+:
3:15
-TGTAATTGg
>CG4328:5CG4328F5:+:
3:16
-AATAATTAa
>CG4328:5CG4328F6:+:
3:17
-ATATATTGc
>CG4328:5CG4328F7:+:
3:18
-CTATATTGc
>CG4328:5CG4328F9:+:
2:19
cGATAATTAa
>CG4328:5CG4328F10:+
:2:20
tCATTATTGa
>CG4328:5CG4328F11:+
:3:21
-AATAATTAa
>CG4328:5CG43282A1:+
:3:22
-AATTATGAg
>CG4328:5CG43282A2:+
:3:23
-TATTATTGa
>CG4328:5CG43282A3:-
:3:24
tATTTATTA-
>CG4328:5CG43282A5:+
:3:25
-ATTTATTAa
>CG4328:5CG43282A6:-
:2:26
cGTATATGAa
>CG4328:5CG43282A7:+
:2:27
aATTTATTGc
>CG4328:5CG43282A8:+
:2:28
tGATAATTGa
>CG4328:5CG43282A10:
-:2:29
tTAATATGAg
>CG4328:5CG43282A11:
+:2:30
cTTTAATTGc
>CG12361:CG12361A1:+
:3:1
tGTTTATGA-
>CG12361:CG12361A2:+
:3:2
tATAAATTA-
>CG12361:CG12361A4:+
:3:4
aATTGATGA-
>CG12361:CG12361A5:-
:1:5
aTTAAATGA-
>CG12361:CG12361A6:+
:3:6
cCTTAATGA-
>CG12361:CG12361A7:-
:2:7
gGTTTATGGg
>CG12361:CG12361A8:-
:2:8
aTTTTATCAc
>CG12361:CG12361A9:-
:3:9
-TTTAATGAc
>CG12361:CG12361A10:
-:3:10
-CTTAATTGc
>CG12361:CG12361A11:
-:2:11
tTTTTATCAc
>CG12361:CG12361A12:
+:3:12
gTATTATTA-
>CG12361:CG12361B1:+
:3:13
gTATTATTA-
>CG12361:CG12361B2:-
:2:14
cTTTGATTAc
>CG12361:CG12361B3:-
:2:15
cGTTTATTAg
>CG12361:CG12361B4:-
:2:16
aTTTTATTGg
>CG12361:CG12361B5:-
:2:17
tTTTAATTAc
>Cad:CadE1:-:3:1
cACAAATTA-
>Cad:CadE2:+:2:2
tGTTGATTAg
>Cad:CadE3:+:2:3
aAGTTATTAc
>Cad:CadE4:-:3:4
aATTAATAG-
>Cad:CadE5:-:3:5
cTTAAATGA-
>Cad:CadE6:-:4:6
gATTTATTT-
>Cad:CadE8:+:2:7
aTTTTATAAg
>Cad:CadE9:-:3:8
cAGTAATTA-
>Cad:CadE11:+:3:10
-CTTAATTGc
>Cad:CadE12:+:2:11
-TTTTATTAg
>Cad:CadF1:-:3:12
cCGTAATTA-
>Cad:CadF2:+:2:13
cATTTATTGg
>Cad:CadF3:+:2:14
cTTTAATGGc
>Cad:CadF4:-:3:15
gTTTAATAA-
>Cad:CadF5:-:3:16
tATTTATTA-
>Cad:CadF6:-:3:17
aTTTTATTA-
>Cad:CadF7:+:2:18
tGTTTATTGc
>Cad:CadF8:+:3:19
-TTTTATTGa
>Cad:CadF9:+:2:20
gTTTTATGAt
>Cad:CadF11:+:2:22
aCTTTATTAc
>Cad:Cad2E1:+:2:23
cCTTTATTGg
>Cad:Cad2E2:-:3:24
aCATTATTA-
>Cad:Cad2E3:+:3:25
-ATTTATTAg
>Cad:Cad2E4:-:3:26
tATTTATTA-
>Cad:Cad2E7:-:3:28
cTCTAATTG-
>Cad:Cad2E8:+:2:29
aGTTTATTGg
>Cad:Cad2E9:+:2:30
tAATGATTGc
>Cad:Cad2E10:-:3:31
cTTAAATTA-
>Cad:Cad2F1:-:3:34
tTCTAATTA-
>Cad:Cad2F2:+:2:35
tGTTTATGAg
>Cad:Cad2F3:-:3:36
gATTTATAA-
>Cad:Cad2F4:-:3:37
cCATAATTA-
>Cad:Cad2F5:+:2:38
tTTTTATTGc
>Cad:Cad2F6:-:3:39
tTTTTATTG-
>Cad:Cad2F7:+:2:40
aATTTATGGg
>Cad:Cad2F9:-:3:41
tTTTTATGA-
>Cad:Cad2F10:+:2:42
aTTATATGGg
>Cad:Cad2F11:+:2:43
tTTTTATTGc
>H2:H20C1:-:4:1:mod
tttTTATATA
>H2:H20C2:+:4:2
tcTTAATGA-
>H2:H20C3:+:3:3
acATTATTGa
>H2:H20C4:+:3:4
ggTTAATGAt
>H2:H20C5:+:3:5
taGATATTAc
>H2:H20C7:+:3:6
ccTTTATGGg
>H2:H20C8:+:3:7
cgTTGATTAa
>H2:H20C9:-:4:8:mod
agtCAATAAA
>H2:H20C10:+:3:9
gaTTAATTAt
>H2:H20C11:-
:4:10:mod
ggaTAATTAA
>H2:H20C12:+:3:11
ggGTAATTAg
>H2:H20D1:+:3:12
ctATTATTAa
>H2:H20D2:+:3:13
tgAATATTGa
>H2:H20D3:+:3:14
gcTTAATTGa
>H2:H20D4:-:4:15:mod
agaTAATAAT
>H2:H20D5:-:4:16:mod
ctgTAATAAA
>H2:H20D6:-:4:17
accTCATAAT
>H2:H20D7:+:3:18
aaTTTATTAa
>H2:H20D8:-:4:19:mod
atcTAATTAA
>H2:H20D9:-:4:20:mod
tgcTAATAAA
>H2:H20D10:+:3:21
cgTTTATTAa
>H2:H20D11:+:3:22
atTTTATGAg
>H2:5H202H2:+:3:23
caTTTATTGg
>H2:5H202H3:+:3:24
tgTTTATGAa
>H2:5H202H4:-
:4:25:mod
cgtTAATAAA
```

```
>H2:5H202H5:-:3:26
-cTTTTTGGc
>H2:5H202H6:+:4:27
atTTTATTA-
>H2:5H202H7:+:3:28
ctTTAATGAg
>H2:5H202H8:-:1:29
ttATTATGG-
>H2:5H202H9:-
:4:30:mod
catTAATAAA
>H2:5H202H10:+:3:31
agGTAATTAc
>H2:5H202H11:+:3:32
gtTTAATGAc
>AbdB:abdb2:+:2:1
GGTTTATAG-
>AbdB:abdb3:+:1:2
GTTTTATTGt
>AbdB:abdb4:+:2:3
TTTTTATGG-
>AbdB:abdb5:+:2:4
GATTAATGG-
>AbdB:abdb6:+:2:5
GCTTTATGT-
>AbdB:abdb7:+:1:6
GGTTTACAAc
>AbdB:abdb8:+:2:7
GTTTAATGT-
>AbdB:abdb10:+:2:8
GATTTATGT-
>AbdB:abdb11:+:2:9
ATATTATGA-
>AbdB:abdb12:+:2:10
CATTTATTA-
>AbdB:abdb13:+:1:11
TTTTTATAAc
>AbdB:abdb14:+:2:12
TATTAATTA-
>AbdB:abdb15:+:1:13
GTTTTATGA-
>AbdB:abdb17:+:2:14
GTTTTATGG-
>AbdB:abdb18:+:2:15
CTTTAACGA-
>AbdB:abdb19:+:2:16
CTTTTATTA-
>AbdB:abdb20:+:2:17
CTTTTACGA-
>AbdB:abdb21:+:2:18
GATTTATTA-
>AbdB:abdb22:+:2:19
GATTTATTA-
>AbdB:abdb23:+:2:20
CTTTAATTA-
>AbdB:abdb24:+:2:21
GTTTTATGA-
>Lim3:Lim3C1:-:3:1
cCCTGATTA-
>Lim3:Lim3C2:-:1:2
-ATTTATTAa
>Lim3:Lim3C3:-:3:3
gCTTAATCA-
>Lim3:Lim3C4:+:2:4
gAATAATTAt
>Lim3:Lim3C5:+:2:5
tAATGATTAt
>Lim3:Lim3C7:-:3:6
aACAAATTA-
>Lim3:Lim3C8:+:3:7
-TCTGATGAa
>Lim3:Lim3C10:+:2:8
tAAAAATTAa
>Lim3:Lim3C11:+:3:9
-CTAAATGAa
>Lim3:Lim3C12:+:3:10
-ATAAATTAg
>Lim3:Lim3D1:+:3:11
-TTTAATGAa
>Lim3:Lim3D3:+:2:13
aTCTAATGAg
>Lim3:Lim3D4:+:1:14
-GGTAATTG-
>Lim3:Lim3D5:+:3:15
-CTTAATTGa
>Lim3:Lim3D6:+:3:16
-TGTAATTGa
>Lim3:Lim3D7:+:2:17
tCTTAATTGa
>Lim3:Lim3D8:-:1:18
-ACTAATTAa
>Lim3:Lim3D9:-:1:19
-ATTAATCAa
>Lim3:Lim3D10:+:1:20
gGCTAATTA-
>Lim3:Lim3D11:+:2:21
aTCTAATTAg
>Awh:AwhC1:-:2:1
atTTGATTAc
>Awh:AwhC2:-:2:2
ggCTGATTGg
>Awh:AwhC3:-:3:3
-tTTAATGAa
>Awh:AwhC4:+:1:4
--CTGATTAc
>Awh:AwhC5:+:4:5
atTTGATTA-
>Awh:AwhC7:-:2:7
aaTTAAGTAg
>Awh:AwhC8:+:1:8
--CTAATTAc
>Awh:AwhC9:+:1:9
--ATAATTAt
>Awh:AwhC10:-:2:10
taTTAATGAa
>Awh:AwhC12:+:4:12
acTTGATTA-
>Awh:AwhD1:-:2:13
cgCTAATGAg
>Awh:AwhD2:+:4:14
gtCTAATTA-
>Awh:AwhD3:-:2:15
acCTAATTAc
>Awh:AwhD5:+:4:16
caCTAATTA-
>Awh:AwhD6:-:2:17
ctTTAATTAc
>Awh:AwhD7:+:4:18
ttTTAATTA-
>Awh:AwhD8:-:1:19
cgCTAATTG-
>Awh:AwhD9:+:4:20
agTTAATTA-
>Awh:AwhD10:+:4:21
taCTAATGA-
>Awh:AwhD11:+:4:22
acCTAATGA-
>Awh:Awh2A1:+:4:23
ccTTAATTT-
>Awh:Awh2A2:-:2:24
atTTGATTAg
>Awh:Awh2A3:+:1:25
--TTAATTAt
>Awh:Awh2A4:+:4:26
ttTTGATTA-
>Awh:Awh2A8:+:4:27
ccTAAATGA-
>Awh:Awh2A9:-:3:28
-tTTAATTAg
>Awh:Awh2A10:-:2:29
tgCTAATTGg
>Awh:Awh2A11:+:2:30
-cATAATTA-
>Awh:Awh2A12:+:1:31
--TTAATTAg
>Awh:Awh2B1:+:4:32
agTTAATTA-
>Awh:Awh2B2:+:4:33
agATAATTA-
>Awh:Awh2B3:-:2:34
-cTTAATTAt
>Awh:Awh2B4:+:4:35
aaCTAATTA-
>Awh:Awh2B5:+:4:36
caCTAATTT-
>Awh:Awh2B6:+:4:37
tcCTAATTA-
>Awh:Awh2B7:-:3:38
-tTTAATTAg
>Awh:Awh2B8:+:4:39
taTTAATTA-
>Awh:Awh2B9:+:4:40
ccCTAATGA-
>Awh:Awh2B10:+:4:41
agTTAATGA-
>Awh:Awh2B11:+:4:42
ctCTAATTG-
>Dll:DllC1:-:3:1
--tTAATTGT
>Dll:DllC2:-:5:2:mod
TGATAATgga
>Dll:DllC3:-:2:3
-tgTAATAGC
>Dll:DllC4:-:2:4
-gtTAATTTC
>Dll:DllC5:+:3:5
-taTAATTTT
>Dll:DllC6:-:2:6
-ccTAATTTG
>Dll:DllC7:-:2:7
-ttTAATGGC
>Dll:DllC8:-:2:8
-acAAATTGT
>Dll:DllC9:-:3:9
--aTAATTAC
>Dll:DllC10:-:4:10
---TAATTAC
>Dll:DllC11:-:4:11
---TAATTAC
>Dll:DllC12:-:2:12
-tcTAATGAT
>Dll:DllD1:-:2:13
-ctTAATAAC
>Dll:DllD2:+:3:14
-caTAATTTT
>Dll:DllD3:+:2:15
--cTGATAGG
>Dll:DllD4:+:4:16
ggtTAATTAC
>Dll:DllD5:-:2:17
-tcTAATTGC
>Dll:DllD6:-:2:18
-taTAATTGC
>Dll:DllD7:-:2:19
-cgTAATTGG
>Dll:DllD8:-:4:20
---TAATTAT
>Dll:DllD9:+:3:21
-agTAATTAC
>Dll:DllD10:+:2:22
--cTAATTAC
>Dll:DllD11:-:4:23
---TAATTAG
>CG4136:5CG4136G1:+:
3:1
agCTAATTA-
>CG4136:5CG4136G2:-
:3:2
-gTTAATTTa
>CG4136:5CG4136G3:-
:3:3
-cTTAATAGa
>CG4136:5CG4136G4:-
:3:4
-gCTAATTAt
>CG4136:5CG4136G5:-
:3:5
-cTTAATTAg
>CG4136:5CG4136G6:+:
3:6
gcCTAATTG-
>CG4136:5CG4136G7:+:
4:7
aaTTAATGA-
>CG4136:5CG4136G8:+:
4:8
cgATAATTA-
>CG4136:5CG4136G9:-
:3:9
-tTTAATAGg
>CG4136:5CG4136G10:-
:3:10
-cCTAATTAc
>CG4136:5CG4136G11:-
:3:11
-cCTAATGAg
>CG4136:5CG4136G12:-
:3:12
-gTTAATGAg
>CG4136:5CG4136H1:+:
4:13
taCTAATTG-
>CG4136:5CG4136H2:+:
4:14
aaCTAATTA-
>CG4136:5CG4136H3:+:
2:15
-cTTAATTAc
>CG4136:5CG4136H4:-
:4:16
--TTAATTAt
>CG4136:5CG4136H5:+:
3:17
gcTTAATTG-
>CG4136:5CG4136H6:+:
4:18
ctCTAATTA-
>CG4136:5CG4136H7:+:
4:19
gtTTAATTA-
>CG4136:5CG4136H9:-
:3:21
-tGTAATTGt
>CG4136:5CG4136H10:-
:3:22
-tTTAATTAa
>CG4136:5CG4136H11:+
:2:23
-cTTAATTAt
>Al:AlE5:+:3:2
--tTAATTAA
>Al:AlE2:-:3:1
-gcTAATTAA
>Al:AlE6:-:4:3
tgcTAATTAA
>Al:AlE7:-:4:4
gcaTAATTAA
>Al:AlE8:-:4:5
gacTAATTAA
>Al:AlE9:+:3:6
--tTAATTAA
>Al:AlE10:-:4:7
ttcTAATTAA
>Al:AlE11:-:3:8
-taTAATTAA
>Al:AlE12:-:4:9
tgcTAATTAA
>Al:AlF2:-:4:11
gacTAATTAA
>Al:AlF3:-:4:12
cgcTAATTGA
>Al:AlF4:-:4:13
tcaTAATTAA
>Al:AlF5:-:4:14
cgcTAATTGG
>Al:AlF6:-:4:15
accTAATTAA
>Al:AlF7:-:1:16
---TAATTAA
>Al:AlF8:+:3:17
--gTAATTAG
>Al:AlF9:-:4:18
gtcTAATTAA
>Al:AlF10:-:4:19
```

```
ggcTAATTAA
>Al:AlF11:-:4:20
cccTAATTGA
>Al:AlF12:-:3:21
-gtTAATTAA
>CG11294:CG11294A1:+
:2:1
-cTTAATTAt
>CG11294:CG11294A3:+
:2:2
-tTTAATTAg
>CG11294:CG11294A9:+
:2:4
-gTCAATTAg
>CG11294:CG11294A11:
+:4:6
aaCTAATTA-
>CG11294:CG11294A12:
-:1:7
-tCTAATTA-
>CG11294:CG11294B1:-
:2:8
taTTAATTAg
>CG11294:CG11294B2:-
:2:9
acCTAATTAg
>CG11294:CG11294B3:-
:4:10
--TTAATTAg
>CG11294:CG11294B4:-
:3:11
-aTTAATTAg
>CG11294:CG11294B5:+
:2:12
-aTCAATTAa
>CG11294:CG11294B6:-
:3:13
-tTTAATTAg
>CG11294:CG11294B7:+
:2:14
-aTTAATTAt
>CG11294:CG11294B8:-
:3:15
-gTTAATTAg
>CG11294:CG11294B9:-
:3:16
-aTTAATTAg
>CG11294:CG11294B10:
-:3:17
-aTTAATTAg
>Lim1:LimC10:+:2:9
-gTTAATTAg
>Lim1:LimC11:+:2:10
-gTTAATTGa
>Lim1:LimC12:+:1:11
--TTAATTAg
>Lim1:LimD1:+:3:12
taCTAATTA-
>Lim1:LimD2:+:2:13
-aTTAATTAg
>Lim1:LimD3:-:4:14
--TTAATTAc
>Lim1:LimD4:-:4:15
--TTAATTAt
>Lim1:LimD10:-:3:20
-aTTAATTAg
>Lim1:LimD11:-:4:21
--TTAATTAa
>Lim1:Lim1C2a:-:4:23
--CTAATTAt
>Lim1:Lim1C9a:-:4:28
--TTAATTAg
>Lim1:Lim1C10a:+:2:2
9
-gTTAATTAc
>Lim1:Lim1D2a:-:4:32
--TTAATTAg
>Lim1:Lim1D4a:+:3:33
aaTTAATTAt
>Lim1:Lim1D6a:-:4:34
--TTAATTAa
>Lim1:Lim1D9a:-:3:35
-tTTAATTAg
>Lim1:Lim1D10a:-
:4:36
--TTAATTAg
>Lim1:Lim1D11a:+:2:3
7
-tTTAATTAc
>Hbn:HbnA1:+:1:1:mod
-aaTAATTAA
>Hbn:HbnA4:+:2:2
-aGTAATTAc
>Hbn:HbnA5:+:2:3:mod
-tgTAATTGA
>Hbn:HbnA6:+:2:4:mod
-caTAATTGA
>Hbn:HbnA7:+:3:5
gaTTAATTAa
>Hbn:HbnA8:-:4:6:mod
ctcTAATTGA
>Hbn:HbnA9:-:4:7:mod
cctTAATTAA
>Hbn:HbnA10:-:3:8
-cTTAATTTg
>Hbn:HbnB2:+:4:9
gaCTAATTA-
>Hbn:HbnB3:-:3:10
-tTTAATTGt
>Hbn:HbnB4:-:3:11
-aTTAATTGt
>Hbn:HbnB5:-:3:12
-tTTAATTAa
>Hbn:HbnB6:+:2:13
-cTTAATTAt
>Hbn:HbnB7:-:3:14
-tTTAATTGg
>Hbn:HbnB9:-:3:15
-tTTAATTAa
>Hbn:HbnB10:-:3:16
-gTTAATTAg
>Hbn:HbnB11:-:2:17
-aTTAATTGg
>Repo:RepoE2:-:3:1
cTTTAATTA-
>Repo:RepoE3:-:3:2
gGGTAATTA-
>Repo:RepoE4:-:2:3
aGCTAATTA-
>Repo:RepoE5:-:3:4
tGTTAATTA-
>Repo:RepoE7:+:2:6
tCTTAATTGa
>Repo:RepoE8:-:3:7
tTTTAATTA-
>Repo:RepoE9:+:2:8
cATTAATTAg
>Repo:Repo2A2:+:3:12
-TTAAATTGc
>Repo:Repo2A3:+:2:13
tAGTAATTGg
>Repo:Repo2A4:-:3:14
cTATAATTA-
>Repo:Repo2A5:+:1:15
aTGTAATTA-
>Repo:Repo2A6:+:3:16
-TTTAATTAg
>Repo:Repo2A7:+:2:17
tATTTATTGa
>Repo:Repo2A8:-:3:18
gCGTAATTA-
>Repo:Repo2A9:-:3:19
aGCTAATTA-
>Repo:Repo2A11:-
:3:20
aTTTAATTG-
>Repo:Repo2A12:-
:3:21
cATTAATTA-
>Repo:Repo2B1:+:3:22
-TTTAATTGa
>Repo:Repo2B2:+:3:23
-TTTAATTAg
>Repo:Repo2B3:-:3:24
gGCTAATTA-
>Repo:Repo2B4:-:3:25
cACTAATTA-
>Repo:Repo2B5:-:3:26
aCTTAATTA-
>Repo:Repo2B6:-:3:27
gTTTAATTA-
>Repo:Repo2B7:-:3:28
cATTAATTA-
>Repo:Repo2B8:+:3:29
-TTTAATTAa
>Repo:Repo2B9:+:2:30
tTTTAATTGg
>Repo:Repo2B10:-
:3:31
aGTTAATTA-
>Repo:Repo2B11:-
:3:32
aTTTAATTA-
>CG32105:CG32105G2:+
:2:1
-TTTAATTAG
>CG32105:CG32105G3:+
:2:2
-ACTAATTAA
>CG32105:CG32105G4:+
:1:3
gTCTAATTGC
>CG32105:CG32105G5:+
:2:4
-AATAATTAG
>CG32105:CG32105G6:+
:1:5
cATTAATTGC
>CG32105:CG32105G7:+
:2:6
-TCTAATTGG
>CG32105:CG32105G8:+
:1:7
gTTATATTAA
>CG32105:CG32105G9:+
:2:8
-TTTAATGAC
>CG32105:CG32105G10:
+:1:9
cTATAATTGA
>CG32105:CG32105G12:
+:2:11
-TTTAATTAA
>CG32105:CG32105H1:+
:2:12
-TTTAATTAA
>CG32105:CG32105H2:+
:1:13
tATTAATTAC
>CG32105:CG32105H3:-
:1:14
-ATTAATTAG
>CG32105:CG32105H4:+
:1:15
tCCAAATTAG
>CG32105:CG32105H5:+
:2:16
-ATTAATTAG
>CG32105:CG32105H7:+
:2:18
-ATTAATTAG
>CG32105:CG32105H9:+
:1:19
aATTTAGTAG
>CG32105:CG32105H10:
+:2:20
-ATTAATTAG
>CG32105:CG32105H11:
-:1:21
-TTTAATTAA
>CG33980:CG33980A1:-
:1:1
-TTTAATTAC
>CG33980:CG33980A2:+
:1:2
-CATAATTAG
>CG33980:CG33980A4:+
:1:4
-GTTAATTAG
>CG33980:CG33980A5:+
:1:5
-GTTAATTAG
>CG33980:CG33980A6:+
:1:6
-GTTAATTGC
>CG33980:CG33980A7:+
:1:7
-GTTAATTGG
>CG33980:CG33980A8:+
:1:8
-TTTAATTAA
>CG33980:CG33980A12:
+:1:12
-GTTAATTGG
>CG33980:CG33980B2:-
:1:13
-TCTAATTAG
>CG33980:CG33980B4:+
:1:15
-TATAATTAG
>CG33980:CG33980B6:+
:1:17
-TTTAATTAG
>CG33980:CG33980B7:+
:1:18
-TCTAATTAG
>CG33980:CG33980B10:
+:1:20
-ATTAATTAG
>Exex:ExexE1:-:3:1
-tCTAATTAa
>Exex:ExexE2:+:2:2:m
od
-ctTAATTGG
>Exex:ExexE3:+:3:3
ggGTAATTAa
>Exex:ExexE4:+:4:4
gaGTAATTA-
>Exex:ExexE5:+:2:5
-cTTAATTAt
>Exex:ExexE6:+:4:6
caGTAATTA-
>Exex:ExexE7:+:4:7
cgTTAATTA-
>Exex:ExexE8:-:4:8
--GTAATAAg
>Exex:ExexE9:-:3:9
-aGTAATTAg
>Exex:ExexE10:+:4:10
gtGTAATTA-
>Exex:ExexE11:+:4:11
ggCTAATTA-
>Exex:ExexE12:+:4:12
gaGTAATTA-
>Exex:ExexF1:+:2:13
-gGTAATTAg
>Exex:ExexF2:-
:4:14:mod
ttcTAATTGA
>Exex:ExexF3:-
:4:15:mod
agtTAATTAC
>Exex:ExexF4:+:2:16
-tGTAATTAa
>Exex:ExexF5:-
:4:17:mod
actTAATCAC
>Exex:ExexF6:+:2:18
-tCTAATTAa
>Exex:ExexF7:+:2:19
-aCTAATTAg
>Exex:ExexF8:+:4:20
ggCTAATTA-
>Exex:ExexF9:+:2:21
```

```
-ttTAATTGC
>Exex:ExexF10:+:4:22
atGTAATTA-
>Exex:ExexF11:+:3:23
cgCTAATTAa
>Rx:RxF2:+:3:1
-GCTAATTAc
>Rx:RxF3:-:3:2
tACTAATTA-
>Rx:RxF4:-:3:3
tACTAATTA-
>Rx:RxF5:+:2:4
gGTTAATTAg
>Rx:RxF7:-:1:6:mod
-gtTAATTGG
>Rx:RxF8:-:1:7:mod
-ttTAATTGA
>Rx:RxF12:-:1:10
-GCTAATTAa
>Rx:Rx2C1:+:1:11
tCATAATTA-
>Rx:Rx2C2:-:1:12
-ACTAATTAg
>Rx:Rx2C3:+:3:13
-GTTAATTAa
>Rx:Rx2C5:-:3:15
aCTTAATTA-
>Rx:Rx2C6:+:2:16
cACTAATTGg
>Rx:Rx2C7:-:1:17
-CCTAATTAt
>Rx:Rx2C8:-:3:18
tGTTAATTA-
>Rx:Rx2C9:-:1:19:mod
-atTAATTGA
>Rx:Rx2C10:-
:1:20:mod
-caTAATTGA
>Rx:Rx2C11:-
:1:21:mod
-gcTAATTGA
>Rx:Rx2D1:-:1:23:mod
-ttTAATTGC
>Rx:Rx2D2:-:1:24
-GCTAATTAg
>Rx:Rx2D3:-:1:25
-AATAATTAg
>Rx:Rx2D4:-:1:26:mod
-ctTAATTGG
>Rx:Rx2D5:-:1:27
-GCTAATTAa
>Rx:Rx2D6:-:1:28:mod
-ttTAATTGA
>Rx:Rx2D8:-:1:29
-GCTAATTAa
>Rx:Rx2D9:+:3:30
-GCTAATTAa
>Rx:Rx2D10:-:1:31
-GCTAATTAa
>Rx:Rx2D11:-
:1:32:mod
-atTAATTGG
>Ro:RoG1:+:3:1
-GGTAATTAc
>Ro:RoG2:-:1:2:mod
-gcTAATTGC
>Ro:RoG3:-:3:3
tACTAATGA-
>Ro:RoG4:+:3:4
-TTTAATTAc
>Ro:RoG5:-:1:5:mod
-ccTAATTGA
>Ro:RoG6:-:3:6
tAATAATTA-
>Ro:RoG7:-:2:7
cTTTAATTA-
>Ro:RoG8:-:3:8
tACTAATGA-
>Ro:RoG9:-:3:9
cACTAATTA-
>Ro:RoG10:-:3:10
tGCTAATTA-
>Ro:RoG11:-:3:11
aGTTAATTA-
>Ro:RoG12:-:3:12
tGTTAATTA-
>Ro:RoH1:-:1:13:mod
-ttTAATTGA
>Ro:RoH2:-:1:14
-ACTAATTAa
>Ro:RoH3:-:2:15
tCGTAATGA-
>Ro:RoH4:-:3:16
cGCTAATTA-
>Ro:RoH5:-:1:17
-GCTAATTAa
>Ro:RoH6:-:1:18
-GCTAATTAa
>Ro:RoH7:-:1:19
-GCTAATTAa
>Ro:RoH8:-:3:20
tCTTAATTA-
>Ro:RoH9:-:1:21
-GCTAATTAa
>Ro:RoH10:-:3:22
cGGTAATTA-
>Ro:RoH11:-:1:23:mod
-ctTAATTGC
>Pph13:Pph13C1:+:3:1
tAATAATTA-
>Pph13:Pph13C2:+:3:2
tGATAATTG-
>Pph13:Pph13C3:+:1:3
-TCTAATTTa
>Pph13:Pph13C4:+:3:4
cCATAATTA-
>Pph13:Pph13C5:-:3:5
-ACTAATTCg
>Pph13:Pph13C6:+:3:6
aTATAATTA-
>Pph13:Pph13C7:+:3:7
aTTTAATTA-
>Pph13:Pph13C8:-:3:8
-ACTAATCAg
>Pph13:Pph13C9:+:1:9
:mod
-caTAATTGT
>Pph13:Pph13C10:+:1:
10:mod
-agTAATTGG
>Pph13:Pph13C11:+:3:
11
tCCTAATTA-
>Pph13:Pph13C12:+:1:
12:mod
-ggTAATTGG
>Pph13:Pph13D1:+:1:1
3
-AATAATTAg
>Pph13:Pph13D2:+:1:1
4:mod
-gcTAATTGT
>Pph13:Pph13D3:+:3:1
5
gCCTAATTA-
>Pph13:Pph13D4:+:1:1
6
-AGTAATTAc
>Pph13:Pph13D5:+:1:1
7:mod
-ttTAATTGA
>Pph13:Pph13D6:+:1:1
8
-ACTAATTAt
>Pph13:Pph13D7:-
:3:19
-ATTAATTAa
>Pph13:Pph13D9:+:1:2
1
-ACTAATTAa
>Pph13:Pph13D11:+:1:
22
-ATTAATTAa
>Inv:InvC1:+:2:1
gGTTAATTAt
>Inv:InvC2:-:1:2
-ACTAATTAa
>Inv:InvC3:+:2:3
tGCTAATTAt
>Inv:InvC5:-:3:4
cACTAATTA-
>Inv:InvC7:+:3:6
-TATAATTAc
>Inv:InvC8:-:1:7
-TCTAATTAa
>Inv:InvC11:+:1:10
cTCTAATTA-
>Inv:InvC12:+:3:11
-TTTAATTAg
>Inv:InvD1:+:2:12
aTTTAATTGg
>Inv:InvD2:+:2:13
cTCTAATTGa
>Inv:InvD3:+:2:14
tTTTAATTGa
>Inv:InvD4:+:2:15
tTCTAATTAg
>Inv:InvD5:+:2:16
tACTAATTAc
>Inv:InvD7:+:2:18
aACTAATTGg
>Inv:InvD8:+:2:19
tTTTAATTGa
>Inv:InvD11:+:3:22
-GTTAATTAc
>CG9876:CG9876A1:-
:2:1
tGCTAATTGt
>CG9876:CG9876A2:+:1
:2:mod
-cgTAATGAG
>CG9876:CG9876A3:+:3
:3
gACTAATTA-
>CG9876:CG9876A4:+:3
:4
tCGTAATTG-
>CG9876:CG9876A5:+:1
:5
-GTTAATTGa
>CG9876:CG9876A6:+:1
:6
-AGTTATTAa
>CG9876:CG9876A7:-
:2:7
cACTAATTGg
>CG9876:CG9876A8:+:3
:8
tATTAATTA-
>CG9876:CG9876A10:+:
1:10
-ACTTATTAa
>CG9876:CG9876B2:-
:2:11
cACTAATTGg
>CG9876:CG9876B3:+:2
:12
cAATAATTAg
>CG9876:CG9876B4:+:1
:13
-ACTAATTAg
>CG9876:CG9876B5:+:1
:14
-TTTAATTAa
>CG9876:CG9876B6:+:1
:15
-TTTAATTAc
>CG9876:CG9876B7:-
:3:16
-ATTAATTAa
>CG9876:CG9876B8:+:1
:17
-TCTAATTAt
>CG9876:CG9876B9:+:1
:18
-GCTAATTAa
>CG9876:CG9876B10:+:
1:19
-ACTAATTAa
>CG9876:CG9876B11:+:
1:20
-ATTAATTAt
>CG9876:CG9876B12:-
:3:21
-GATAATTGc
>En:eng1:-:2:1
cGCTAATTAg
>En:eng2:+:3:2
aTTTAATTA-
>En:eng3:-:2:3
cACTAATGAg
>En:eng4:+:3:4
gGGTAATTA-
>En:eng5:-:2:5
tGATAATTGc
>En:eng7:+:3:6
tGCTAATTA-
>En:eng8:-:2:7
gGTTAATTGa
>En:eng9:+:3:8
cGTTAATTA-
>En:eng10:+:3:9
aGGTAATTA-
>En:eng11:+:3:10
gGCTAATTA-
>En:eng12:-:1:11
tTTTAATTG-
>En:eng13:-:2:12
tTTTAATTGg
>En:eng14:+:3:13
aACTAATTA-
>En:eng15:+:3:14
tCTTAATTG-
>En:eng16:+:3:15
cGATAATTG-
>En:eng17:+:3:16
gACTAATTA-
>En:eng18:-:2:17
tTTTAATTGg
>En:eng19:-:3:18
-CTTAATTGa
>En:eng20:+:3:19
cGTTAATGA-
>En:eng21:+:1:20
-GCTAATTAa
>En:eng22:-:2:21
gTTTAATTGg
>En:eng23:-:2:22
-CTTAATTGa
>En:eng24:+:3:23
gGCTAATTA-
>CG32532:CG32432G1:+
:3:1
-gCTAATTAc
>CG32532:CG32432G2:+
:2:2
tgATAATTGg
>CG32532:CG32432G3:-
:1:3
--TTAATTAt
>CG32532:CG32432G4:+
:3:4
-tTTAATTTg
>CG32532:CG32432G5:-
:4:5
agGTAATTA-
>CG32532:CG32432G6:-
:2:6
-tCTAATTAt
>CG32532:CG32432G7:-
:3:7
gcTTAATGA-
>CG32532:CG32432G8:+
:3:8
```

```
-gGTAATTAc
>CG32532:CG32432G9:+
:2:9
taCTAATTGc
>CG32532:CG32432G10:
+:2:10
atCTAATTGg
>CG32532:CG32432G11:
+:3:11
-aTTAATTGa
>CG32532:CG32432G12:
+:3:12
-cTTAATTAg
>CG32532:CG32432H1:+
:2:13
cgTTAATTGg
>CG32532:CG32432H2:+
:3:14
-cCTAATTGg
>CG32532:CG32432H3:+
:3:15
-tTTAATTAa
>CG32532:CG32432H4:+
:3:16
-aTTAATTGg
>CG32532:CG32432H5:-
:2:17
-tTTAATTAt
>CG32532:CG32432H6:-
:2:18
-gTTAATTAg
>CG32532:CG32432H7:+
:2:19
tcCTAATTGa
>CG32532:CG32432H8:+
:3:20
-aTTAATTGg
>CG32532:CG32432H9:+
:3:21
-cTTAATTGa
>CG32532:CG32432H10:
+:3:22
-gTTAATTAa
>CG32532:CG32432H11:
-:4:23
acATAATGA-
>Unpg:UnpgC1:-:3:1
-ACTAATGAc
>Unpg:UnpgC2:+:3:2
aCGTAATTA-
>Unpg:UnpgC3:+:2:3
cGCTAATTAg
>Unpg:UnpgC4:+:3:4
aGATAATTG-
>Unpg:UnpgC5:+:1:5
-CGTAATTAg
>Unpg:UnpgC6:-:2:6
gCGTAATTAg
>Unpg:UnpgC7:-:3:7
-CTTAATTGc
>Unpg:UnpgC8:-:2:8
-TTTAATGAg
>Unpg:UnpgC9:-:3:9
-TTTAATTGa
>Unpg:UnpgC10:-:2:10
cCTTAATTGc
>Unpg:UnpgC11:+:1:11
-ACTAATTAt
>Unpg:UnpgD1:+:3:12
gCCTAATTA-
>Unpg:UnpgD2:-:2:13
cGCTAATTAt
>Unpg:UnpgD3:-:2:14
-CTTAATTGa
>Unpg:UnpgD4:-:3:15
-CTTAATTGg
>Unpg:UnpgD5:-:3:16
-ATTAATTGg
>Unpg:UnpgD6:-:2:17
tCCAAATTAg
>Unpg:UnpgD8:-:3:18
-ATTAATTAg
>Unpg:UnpgD9:+:1:19
-TGTAATTAt
>Unpg:UnpgD10:-:2:20
cCTTAATTAg
>Unpg:UnpgD11:-:2:21
tATTAATTAg
>PhdP:PhdPA1:+:2:1
-aTTAATTtg
>PhdP:PhdPA2:+:3:2
ggTTAATTac
>PhdP:PhdPA3:-:3:3
-gCTAATAgg
>PhdP:PhdPA4:+:2:4:m
od
-cctAATGAG
>PhdP:PhdPA8:+:4:6
gcATAATTt-
>PhdP:PhdPA9:+:4:7
taATAATTg-
>PhdP:PhdPA10:+:2:8
-cCTAATTgg
>PhdP:PhdPA11:+:4:9
aaATAATTa-
>PhdP:PhdPA12:+:4:10
gaTTAATTa-
>PhdP:PhdPB1:-:4:11
-gATAATTta
>PhdP:PhdPB2:-:4:12
-tTTTATTga
>PhdP:PhdPB3:+:5:13
cgCTAATTt-
>PhdP:PhdPB4:-:4:14
-cTTAATTcc
>PhdP:PhdPB5:+:1:15
--CTAATTat
>PhdP:PhdPB6:+:2:16
-cTTAATTat
>PhdP:PhdPB7:+:2:17
-tTTAATTaa
>PhdP:PhdPB10:+:4:18
acCTAATTa-
>CG7056:CG7056G1:+:3
:20
tATTAATTA-
>CG7056:CG7056G2:+:3
:21
tTTGAAGTA-
>CG7056:CG7056G3:+:3
:22
cCCTAATTG-
>CG7056:CG7056G4:+:2
:23
aATCAATTAc
>CG7056:CG7056G5:+:3
:24
tTTGAATTG-
>CG7056:CG7056G6:+:3
:25
aTTTAATTA-
>CG7056:CG7056G7:-
:3:26
-TCTAATTAc
>CG7056:CG7056G8:+:3
:27
tTTTAATAA-
>CG7056:CG7056G10:+:
2:28
aTCTTATTAc
>CG7056:CG7056G11:+:
3:29
gACCAATTG-
>CG7056:CG7056G12:+:
3:30
gACGAAGTA-
>Oc:OcA2:+:4:2
gcTTAAGCC-
>Oc:OcA3:+:3:3
cgATAATCCc
>Oc:OcA4:-:3:4
-tTTAAGCCc
>Oc:OcA5:-:3:5
-aATAATCCt
>Oc:OcA6:+:4:6
atATAATCC-
>Oc:OcA9:-:4:9
--CTAATCCg
>Oc:OcA10:-:2:10
ccTTAATCCt
>Oc:OcA11:-:3:11
-gTTAATCTg
>Oc:OcB1:-:3:12
-aTTAATCCa
>Oc:OcB2:+:4:13
tcATAATCC-
>Oc:OcB3:-:3:14
-cGTAATCCt
>Oc:OcB4:-:3:15
-cTTAATCGc
>Oc:OcB5:-:3:16
-cTTAATCCa
>Oc:OcB6:-:3:17
-tATAATCCc
>Oc:OcB7:+:4:18
agTTAATCC-
>Oc:OcB8:-:3:19
-aCTAATCCa
>Oc:OcB9:-:2:20
-aTTAATCCt
>Oc:OcB10:-:3:21
-cTTAATCCg
>Oc:OcB11:-:3:22
-gTTAATCCg
>Bcd:bcd1:-:2:1
tGTTAATCCg
>Bcd:bcd2:+:3:2
-TCTAATCCa
>Bcd:bcd3:-:2:3
cGTTAATCTc
>Bcd:bcd4:-:3:4
gTTTAATCC-
>Bcd:bcd5:-:3:5
cTATAATCC-
>Bcd:bcd6:-:2:6
tCTTAATCCc
>Bcd:bcd7:-:2:7
gCTTAATCCg
>Bcd:bcd8:-:3:8
gGTTAATCC-
>Bcd:bcd9:-:3:9
aGATAATCC-
>Bcd:bcd10:-:2:10
aGCTTATCC-
>Bcd:bcd11:-:3:11
gGGTAATCC-
>Bcd:bcd13:-:2:12
tGTTAATCC-
>Bcd:bcd14:+:3:13
-TATAATCCc
>Bcd:bcd15:-:2:14
gCGTAATCCa
>Bcd:bcd16:+:1:15
gCTTAAGCC-
>Bcd:bcd17:-:1:16
-GGTTATCCg
>Bcd:bcd18:-:2:17
tGTTAATCCc
>Bcd:bcd20:-:3:18
gCTTAATCC-
>Bcd:bcd21:-:3:19
tACTAATCC-
>Bcd:bcd22:-:3:20
tCCTAATCC-
>Bcd:bcd23:-:2:21
gGTTAATCCg
>Bcd:bcd24:+:3:22
-TCTAATCCa
>Ptx1:PtxG1:+:4:1
atCTAATCC-
>Ptx1:PtxG2:+:2:2
-tTTAATCCc
>Ptx1:PtxG4:-:4:3
--CTAATCCt
>Ptx1:PtxG6:-:3:4
-aCTAATCCt
>Ptx1:PtxG7:+:2:5
-cTTAATCCt
>Ptx1:PtxG8:-:3:6
-gTTAATCCc
>Ptx1:PtxG9:-:3:7
-gTTAATCCc
>Ptx1:PtxG10:+:4:8
aaTTAATCC-
>Ptx1:PtxG11:+:4:9
ggCTAATCC-
>Ptx1:PtxG12:+:3:10
tcGTAATCCc
>Ptx1:PtxH1:+:2:11
-gCTAATCCt
>Ptx1:PtxH2:+:4:12
tcTTAATCC-
>Ptx1:PtxH4:+:3:14
cgTTAATCTc
>Ptx1:PtxH5:+:3:15
ccTTAATCCc
>Ptx1:PtxH6:+:3:16
cgTTAATCCc
>Ptx1:PtxH7:+:3:17
ggTTAATCCc
>Ptx1:PtxH8:+:3:18
tcTTAATCCc
>Ptx1:PtxH9:+:3:19
cgTTAATCCc
>Ptx1:PtxH10:+:3:20
cgTTAATCCc
>Ptx1:PtxH11:+:3:21
acTTAATCCc
>Gsc:GscA1:+:2:1
tCGTAATCGg
>Gsc:GscA2:-:3:2
aAGTAATCC-
>Gsc:GscA3:+:2:3
gACTAATCTt
>Gsc:GscA4:+:3:4
-AATAATCCt
>Gsc:GscA5:+:2:5
tACTAATCTt
>Gsc:GscA6:+:3:6
-GATAATCTg
>Gsc:GscA7:+:3:7
-AGTAATCCt
>Gsc:GscA8:+:3:8
-CTTAATCGc
>Gsc:GscA9:-:3:9
tCGTAATCC-
>Gsc:GscA10:-:3:10
cGTTAATCT-
>Gsc:GscA11:+:3:11
-CATAATCCt
>Gsc:GscA12:-:3:12
cAATAATCC-
>Gsc:GscB1:-:3:13
cGTTAATCT-
>Gsc:GscB3:+:2:14
tGTTAATCCc
>Gsc:GscB4:+:3:15
-CTTAATCTc
>Gsc:GscB5:+:3:16
-CTTAATCCg
>Gsc:GscB6:+:3:17
-TTTAATCCg
>Gsc:GscB7:-:3:18
cCCTAATCC-
>Gsc:GscB8:+:2:19
tGTTAATCCc
>Gsc:GscB9:+:2:20
tCCTAATCCa
>Gsc:GscB10:+:3:21
-ACTAATCCa
>Gsc:GscB11:+:1:22
cATTAATCC-
>Oct:oct1:+:3:1
cGATAATGA-
```

```
>Oct:oct2:+:3:2
gTTTAATGA-
>Oct:oct3:+:3:3
tTTTAATAA-
>Oct:oct4:+:3:4
aCATAATTT-
>Oct:oct5:+:2:5
cTATAATTAg
>Oct:oct6:+:3:6
aGATAATTT-
>Oct:oct7:+:3:7
aTATAATAA-
>Oct:oct8:+:3:8
cATTAATCA-
>Oct:oct10:-:3:9
-TTTAATCAc
>Oct:oct11:-:3:10
-CCTAATGAg
>Oct:oct13:+:3:11
aGCTAATTA-
>Oct:oct14:+:3:12
cATTAATTT-
>Oct:oct15:+:3:13
gGTAAATGA-
>Oct:oct16:-:3:14
-TTTAATGAg
>Oct:oct17:+:3:15
cCCTAATTA-
>Oct:oct18:-:3:16
-GTTAATGAg
>Oct:oct19:+:3:17
tCATAATCA-
>Oct:oct20:+:3:18
gTTTAATTG-
>Oct:oct21:+:3:19
aGATAATTC-
>Oct:oct22:+:3:20
gTTTAATTT-
>Oct:oct23:+:3:21
cTTTAATTT-
>Oct:oct24:+:3:22
tGATAATTA-
>Bsh:BshE1:+:1:1
-CCTAATGGG
>Bsh:BshE2:+:1:2
-TTTAATCGA
>Bsh:BshE3:+:1:3
-CTTAATGAC
>Bsh:BshE4:+:1:4
-TGTAATTGG
>Bsh:BshE6:+:1:5
-ATTAATTGC
>Bsh:BshE8:+:1:7
-CTTAATGAT
>Bsh:BshE11:+:1:10
-TTTAATCGA
>Bsh:BshE12:+:1:11
-CTTAATCGA
>Bsh:BshF1:+:1:12
-TCTAATGAG
>Bsh:BshF4:+:1:14
-TCTAATGAG
>Bsh:BshF6:+:1:15
-GTTAATTGC
>Bsh:BshF7:+:1:16
-ATTAATTAG
>Bsh:BshF8:+:1:17
-ATTAATTAG
>Bsh:BshF9:+:1:18
-CTTAACGAG
>Bsh:BshF10:+:1:19
-TATAATTGG
>Bsh:BshF11:+:1:20
-TGTAATTGG
>Tup:TupE1:-:3:1
-GGTAATTGa
>Tup:TupE2:-:2:2
cGCTAATTAg
>Tup:TupE3:-:2:3:mod
CACTAATGt-
>Tup:TupE4:+:3:4
tATTAATGG-
>Tup:TupE5:-:3:5
-CTTAATAGa
>Tup:TupE6:-:3:6
-TATAATGGt
>Tup:TupE7:-:3:7
-GATAATTAa
>Tup:TupE9:-:3:9
-GTTAAGTGg
>Tup:TupE10:-:3:10
-CATAATTGa
>Tup:TupE11:+:3:11
cCTTAATGG-
>Tup:TupE12:+:3:12
cCTTAATGG-
>Tup:TupF2:-:3:14
-CTTAATTGc
>Tup:TupF4:-:3:15
-CTAAATGGa
>Tup:TupF6:-:3:17
-CTTAATGGa
>Tup:TupF7:+:3:18
cGATAAGTG-
>Tup:TupF8:-:3:19
-CTTAATTGa
>NK7:NK7E1:+:3:1
aTATAATGA-
>NK7:NK7E2:+:3:2
aATTAAGTG-
>NK7:NK7E3:+:3:3
aATTAAGTG-
>NK7:NK7E4:-:3:4
-TTTAAATAc
>NK7:NK7E6:+:3:5
aAATAATTA-
>NK7:NK7E7:+:2:6
gAGTAAATGa
>NK7:NK7E8:+:3:7
aACTAATTG-
>NK7:NK7E10:-:3:8
-GTTAAGTGg
>NK7:NK7F1:+:3:10
aGCTAATTG-
>NK7:NK7F4:+:3:13
cGCTAATGA-
>NK7:NK7F6:-:1:15
tTTTAATTG-
>NK7:NK7F7:+:3:16
cCTTAATAG-
>NK7:NK7F8:-:3:17
-CTTAATTGc
>NK7:NK7F9:-:3:18
-CTTAATTGg
>NK7:NK7F10:+:3:19
gCTTAATTA-
>NK7:NK7F11:+:3:20
gACTAATTA-
>NK7:NK72A2:-:2:21
gTTTAATGAt
>NK7:NK72A3:+:3:22
aCTTAAGTG-
>NK7:NK72A5:-:3:24
-TTTAATGGa
>NK7:NK72A6:-:3:25
-CTTAATAGc
>NK7:NK72A7:-:3:26
-TATAATTGt
>NK7:NK72A8:-:2:27
tTTTAATTAa
>NK7:NK72A9:+:3:28
aCATAATAG-
>NK7:NK72A10:+:3:29
aGATAATGA-
>NK7:NK72A11:-:2:30
tATTGATAGc
>NK7:NK72A12:+:3:31
cATTAAATG-
>NK7:NK72B2:+:3:32
gGTTAAATA-
>NK7:NK72B4:+:3:34
aATTAAGTG-
>NK7:NK72B5:-:2:35
tATTAATAGt
>NK7:NK72B6:-:3:36
-ATTGATGGt
>NK7:NK72B7:+:3:37
tATTAATGA-
>NK7:NK72B8:-:3:38
-CTTAATTAc
>NK7:NK72B9:+:3:39
cACTAATTA-
>NK7:NK72B10:-:3:40
-TTTAATAGg
>NK7:NK72B11:-:2:41
cCTTAATTGg
>CG13424:CG13424E1:+
:2:1
gCCTAATTGa
>CG13424:CG13424E2:+
:3:2
-TTTAATTAg
>CG13424:CG13424E3:-
:3:3
gTTTAATTA-
>CG13424:CG13424E4:+
:3:4
-TATAATTGt
>CG13424:CG13424E9:+
:2:8
cATTAATGGg
>CG13424:CG13424E11:
+:2:10
tATTAATTAg
>CG13424:CG13424E12:
-:3:11
tGTTAATGA-
>CG13424:CG13424F2:-
:3:12
aACTAATGG-
>CG13424:CG13424F3:+
:3:13
-GGTAATTGc
>CG13424:CG13424F4:+
:3:14
-TTTAATAGg
>CG13424:CG13424F5:-
:3:15
gTCTAATGA-
>CG13424:CG13424F7:+
:2:17
gCATAATTGg
>CG13424:CG13424F8:+
:2:18
gTTTAATTGa
>CG13424:CG13424C1:+
:2:23
gGTTAATTAg
>CG13424:CG13424C5:-
:3:26
gATTAATTA-
>CG13424:CG13424C6:-
:3:27
tACTAATTG-
>CG13424:CG13424C7:+
:3:28
-CTTAATAGc
>CG13424:CG13424C8:-
:3:29
aGTTAATAG-
>CG13424:CG13424C12:
+:2:33
tCCTAATTGg
>CG13424:CG13424D1:+
:2:34
tACTAATTGc
>CG13424:CG13424D2:+
:2:35
tCCTAATTAt
>Hgtx:HgtxC1:+:3:1
-ATTAATAGg
>Hgtx:HgtxC2:-:1:2
-AGTAATTGa
>Hgtx:HgtxC3:-:2:3
aGCTAATTA-
>Hgtx:HgtxC4:+:3:4
-GATAATTGg
>Hgtx:HgtxC5:-:3:5
tTGTAATTA-
>Hgtx:HgtxC6:+:1:6
tTTTAATGA-
>Hgtx:HgtxC7:-:3:7
aTGTAATTA-
>Hgtx:HgtxC8:+:3:8
-GTTAATTGa
>Hgtx:HgtxC9:+:3:9
-CATAATGAg
>Hgtx:HgtxD1:+:3:11
-GATAATTGc
>Hgtx:HgtxD2:+:3:12
-TCTAATTAc
>Hgtx:HgtxD3:+:3:13
-AATAATGAa
>Hgtx:HgtxD4:+:2:14
aATTAATTAa
>Hgtx:HgtxD5:-:3:15
tTTTAATTA-
>Hgtx:HgtxD6:-:3:16
cTTTAATTA-
>Hgtx:HgtxD7:+:2:17
-TTTAATGAg
>Hgtx:HgtxD8:-:3:18
cTTTAATTA-
>Hgtx:HgtxD9:-:3:19
tTTTAATGA-
>Hgtx:HgtxD10:-:3:20
aGTTAATGA-
>Hgtx:HgtxD11:-:3:21
gTTTAATTA-
>Ems:emsA1:+:2:1
-TATAATTAa
>Ems:emsA2:+:1:2
cATAAATGG-
>Ems:emsA3:-:3:3
aTTTTATGA-
>Ems:emsA5:+:2:5
tATTAGTGAa
>Ems:emsA6:+:3:6
-AGTAATGAc
>Ems:emsA7:+:3:7
-ATTAATGGa
>Ems:emsA8:+:3:8
-GTTAATCAc
>Ems:emsA9:-:3:9
cATAAATGA-
>Ems:emsA10:+:3:10
-TTTAATTAg
>Ems:emsA11:+:2:11
cAATAATTGg
>Ems:emsA12:-:3:12
cATAAATTA-
>Ems:emsB2:+:3:14
-TATAATTGg
>Ems:emsB3:+:3:15
-TTTAATGGa
>Ems:emsB5:-:3:16
tTCTAATGA-
>Ems:emsB6:-:3:17
gCCTAATGA-
>Ems:emsB7:-:3:18
cACTAATTA-
>Ems:emsB8:-:3:19
gTCTAATGA-
>Ems:emsB9:+:2:20
cTCTAATTGg
>Ems:emsB10:+:2:21
cTCTAATTAg
>Ems:emsB11:+:2:22
cGCTAATTAg
>Otp:OTPG1:+:1:1
-TTCAATTAt
>Otp:OTPG2:-:2:2
tTCTAATTTg
>Otp:OTPG3:-:3:3
```

```
-ATTAATTGc
>Otp:OTPG4:+:2:4
gTTTAATCA-
>Otp:OTPG5:+:3:5
tTGTAATTA-
>Otp:OTPG6:-:2:6
aATTAATGGt
>Otp:OTPG8:+:4:7
cGCTAATTT-
>Otp:OTPG9:-:2:8
gTTTAATGAg
>Otp:OTPG10:-:1:9
cTTTAATGA-
>Otp:OTPG11:+:1:10
-TTTAATTAa
>Otp:OTPG12:+:1:11
-CCTAATTAg
>Otp:OTPH1:+:2:12
cCTTAATTAa
>Otp:OTPH2:-:2:13
cTGTAATTAg
>Otp:OTPH3:+:2:14
gCATAATTA-
>Otp:OTPH4:-:3:15
-GTTAATGAa
>Otp:OTPH5:-:2:16
cCTTAATTGa
>Otp:OTPH6:-:3:17
-TTTAATTAa
>Otp:OTPH7:-:3:18
-TCTAATTAa
>Otp:OTPH10:+:3:19
aACTAATTA-
>Otp:OTPH11:-:2:20
gTCTAATTAa
>Ftz:FtzG1:+:3:1
tCATAATTG-
>Ftz:FtzG3:-:3:3
-GTTAATGGg
>Ftz:FtzG4:+:3:4
tCTTAATTA-
>Ftz:FtzG6:+:3:6
tACTAATGA-
>Ftz:FtzG7:+:3:7
tATAAATGA-
>Ftz:FtzG8:+:3:8
tTTTAATTG-
>Ftz:FtzG9:+:3:9
gCCTAATGA-
>Ftz:FtzG10:+:3:10
aGTTAATTA-
>Ftz:FtzG11:-:3:11
-GTTAATGAt
>Ftz:FtzG12:+:3:12
gCTTAATTA-
>Ftz:FtzH1:+:3:13
gCTTAATGG-
>Ftz:FtzH3:+:3:15
cGTTAATTA-
>Ftz:FtzH5:+:3:16
aGTTAATGA-
>Ftz:FtzH7:+:3:18
tGTTAATTA-
>Ftz:FtzH8:+:3:19
gGTTAATTA-
>Ftz:FtzH9:+:3:20
tTTTAATGA-
>Ftz:FtzH10:+:3:21
tATTAATTA-
>Ftz:FtzH11:+:3:22
tGTTAATGA-
>Antp:AntpA02:-:3:1
cCTTAATTA-
>Antp:AntpA03:-:3:2
aTTTAATTA-
>Antp:AntpA04:-:3:3
gCTTAATGA-
>Antp:AntpA05:-:3:4
tGTTAATGA-
>Antp:AntpA07:-:3:6
gTTTAATGA-
>Antp:AntpA08:-:3:7
gCTTAATGA-
>Antp:AntpA10:-:3:8
tACTAATTA-
>Antp:AntpA11:-:3:9
cCTTAATGG-
>Antp:AntpB01:-:3:11
tCATAATTA-
>Antp:AntpB02:-:3:12
tGTTAATTA-
>Antp:AntpB04:-:3:13
aTTTAATTA-
>Antp:AntpB06:-:3:14
gTTTAATGA-
>Antp:AntpB07:-:3:15
tTTTAATGA-
>Antp:AntpB09:-:3:17
gTTTAATTA-
>Antp:AntpB11:-:3:18
tTTTAATGA-
>Antp:AntpB12:-:3:19
cTTTAATGA-
>Zen2:Zen2A2:+:2:2
TTATAATGA-
>Zen2:Zen2A4:+:2:4
CCGTAATTA-
>Zen2:Zen2A6:+:2:6
TACTAATTG-
>Zen2:Zen2A7:+:2:7
GAGTAATGA-
>Zen2:Zen2A8:+:2:8
TGATAATGA-
>Zen2:Zen2A9:+:2:9
CCGTAATTA-
>Zen2:Zen2A10:+:2:10
TAGTAATTA-
>Zen2:Zen2B3:+:2:14
TACTAATTA-
>Zen2:Zen2B4:-:2:15
TACTAATGAc
>Zen2:Zen2B9:+:2:20
CCTTAATTA-
>Zen2:Zen2B11:+:2:22
AGTTAATGA-
>Zen2:Zen2E1:-:2:23
ACTTAATTAt
>Zen2:Zen2E3:+:2:25
ATTTAATTA-
>Zen2:Zen2E4:-:2:26
TCATAATTGa
>Zen2:Zen2E5:-:2:27
TTTTAATGAc
>Zen2:Zen2E6:+:2:28
GCTAAATTA-
>Zen2:Zen2E7:+:2:29
CACTAACGA-
>Zen2:Zen2E9:-:2:31
CCTTAATTGc
>Zen2:Zen2E10:+:2:32
CCTTAATTA-
>Zen2:Zen2E11:-:2:33
TTTTAATTGc
>Zen2:Zen2F1:+:2:35
CCGTAATGA-
>Zen2:Zen2F3:+:2:37
TACTAATGA-
>Zen2:Zen2F5:+:2:39
TGCTAATTA-
>Zen2:Zen2F6:+:2:40
AGCTAATTA-
>Zen2:Zen2F7:+:1:41
TGTTAATGA-
>Zen2:Zen2F11:+:2:45
TGTTAATTA-
>Slou:SlouE2:+:3:2
-CTTAATGGc
>Slou:SlouE3:+:2:3
tTTTAATGAg
>Slou:SlouE4:-:3:4
gAGTAATGA-
>Slou:SlouE5:-:2:5
cGTCAATTAc
>Slou:SlouE6:+:3:6
-GCTAATTGt
>Slou:SlouE7:-:1:7
-GTTAATTAt
>Slou:SlouE8:+:3:8
-TTTAATCGg
>Slou:SlouE9:+:3:9
-TTTAATAAt
>Slou:SlouE10:+:1:10
aATTAATTG-
>Slou:SlouE11:-:3:11
gTCTAATGA-
>Slou:SlouE12:-:3:12
tTCTAATGA-
>Slou:SlouF1:-:3:13
gGCTAATTA-
>Slou:SlouF2:-:3:14
cTCTAATTG-
>Slou:SlouF3:-:2:15
cTATAATTAg
>Slou:SlouF4:-:1:16
-GTTAATTAg
>Slou:SlouF5:+:3:17
-CTTAATTGg
>Slou:SlouF6:+:3:18
-TATAATTGt
>Slou:SlouF7:+:3:19
-CTTAATTGa
>Slou:SlouF8:+:1:20
cTTTAGTAG-
>Slou:SlouF9:+:3:21
-TTTAATTGg
>Slou:SlouF10:+:3:22
-TTTAATTAc
>Slou:SlouF11:+:3:23
-GATAATTGg
>Btn:BtnE1:+:3:1
-TTTAATGGc
>Btn:BtnE2:-:3:2
cAGTAATGA-
>Btn:BtnE3:-:3:3
aCTTGATTA-
>Btn:BtnE4:+:3:4
-ATTAATGTa
>Btn:BtnE5:+:3:5
-CTTAATGGg
>Btn:BtnE6:-:3:6
cTGTAATTA-
>Btn:BtnE7:+:3:7
-GGTAATGAc
>Btn:BtnE8:+:3:8
-CTTAACGAc
>Btn:BtnE9:-:3:9
gCCTAATTA-
>Btn:BtnE10:+:3:10
-TATAATTGc
>Btn:BtnE11:-:1:11
-GATAATTAa
>Btn:BtnE12:-:3:12
tCGTAATGA-
>Btn:BtnF1:+:3:13
-TATAATGAt
>Btn:BtnF2:-:1:14
-CTTAATTAa
>Btn:BtnF3:+:3:15
-ATTAATGAc
>Btn:BtnF4:-:3:16
aTATAATGA-
>Btn:BtnF5:+:3:17
-GATAATTAg
>Btn:BtnF6:+:3:18
-ATTAATTAc
>Btn:BtnF7:+:3:19
-CTTAATGAc
>Btn:BtnF8:-:3:20
aCTTAATGA-
>Btn:BtnF9:-:3:21
cCTTAATGA-
>Btn:BtnF10:-:3:22
cGTTAATGA-
>Btn:BtnF11:-:3:23
aGTTAATGA-
>Dfd:dfd1:+:3:1
-CTTAATGAa
>Dfd:dfd2:+:3:2
-ATTAATGAc
>Dfd:dfd3:+:3:3
-TTTAATGAt
>Dfd:dfd4:+:3:4
-ATTAATGAc
>Dfd:dfd5:+:2:5
tGTTAATGAc
>Dfd:dfd6:+:3:6
-CTTAATTAg
>Dfd:dfd7:+:3:7
-ATTAATTAt
>Dfd:dfd8:+:3:8
-CTTAATTAg
>Dfd:dfd9:+:3:9
-ATTAATGGg
>Dfd:dfd10:-:3:10
tTTTAATGA-
>Dfd:dfd11:+:3:11
-TTTAATAGc
>Dfd:dfd12:-:3:12
cACTAATGA-
>Dfd:dfd13:-:2:13
tCGTAATGA-
>Dfd:dfd14:-:3:14
gCTTAATGG-
>Dfd:dfd15:-:3:15
tCGTAATTA-
>Dfd:dfd16:+:2:16
-AGTAATGAg
>Dfd:dfd17:+:3:17
-CTTAATGAa
>Dfd:dfd18:+:3:18
-CCTAATGAc
>Dfd:dfd19:-:3:19
aCCTAATGA-
>Dfd:dfd20:-:3:20
gGATAATGA-
>Dfd:dfd21:-:3:21
gACTAATGA-
>Dfd:dfd22:+:3:22
-GTTAATGAt
>Dfd:dfd23:-:3:23
cGTTAATGA-
>Dfd:dfd24:-:3:24
aATTAATGA-
>Scr:ScrE04:+:1:1
CCTTAATGA-
>Scr:ScrE05:+:2:2
ACATAATGA-
>Scr:ScrE08:+:2:4
TACTAATTA-
>Scr:ScrE12:+:2:7
CCTTAATGA-
>Scr:ScrF02:+:1:9
CGATAATGA-
>Scr:ScrF04:+:2:10
ACTTAATGA-
>Scr:ScrF06:+:2:11
CGCTAATGA-
>Scr:ScrF11:+:2:15
CATTAATGA-
>Scr:ScrF12:+:2:16
CGTTAATGA-
>Scr:ScrG1:-:2:17
CTTTAATTGc
>Scr:ScrG2:+:2:18
AATTAATGA-
>Scr:ScrG3:+:2:19
GTTTAATGA-
>Scr:ScrG5:+:2:20
CACTAATTA-
>Scr:ScrG6:-:2:21
CGTTAATTGc
>Scr:ScrG7:+:2:22
TGTTAATTA-
>Scr:ScrG8:+:1:23
```

CACTAATTA-
>Scr:ScrG9:+:2:24
CGTTAATTA-
>Scr:ScrG12:+:2:26
CGCTAATGA-
>Scr:ScrH1:+:1:27
TGTTAATGA-
>Scr:ScrH2:+:2:28
ACATAATGA-
>Scr:ScrH3:+:2:29
CGTTAATGA-
>Scr:ScrH4:+:2:30
TATTAATGA-
>Scr:ScrH9:+:2:32
TACTAATGA-
>Scr:ScrH10:+:2:33
ATTTAATGA-
>Scr:ScrH11:+:2:34
CACTAATGA-
>Zen:ZenC1:+:2:1
CCTTAATTA-
>Zen:ZenC3:+:1:2
GACTAATTA-
>Zen:ZenC4:+:2:3
GGCTAATTA-
>Zen:ZenC5:+:2:4
TAATAATGA-
>Zen:ZenC8:+:2:7
ACCTAATGA-
>Zen:ZenC9:+:2:8
GTTTAATGA-
>Zen:ZenC10:+:2:9
CCCTAATGA-
>Zen:ZenC11:+:2:10
CGCTAATGA-
>Zen:ZenD2:+:2:12
CCCTAATGA-
>Zen:ZenD3:+:2:13
GCTTAATGA-
>Zen:ZenD4:+:2:14
ACCTAATGA-
>Zen:ZenD5:+:2:15
TGCTAATTA-
>Zen:ZenD6:+:2:16
CCTTAATGA-
>Zen:ZenD7:+:2:17
TTTTAATTA-
>Zen:ZenD9:+:2:18
ACATAATGA-
>Zen:ZenD11:+:1:19
TGCTAATGA-
>Pb:Pb1:+:3:1
-TCTAATGAc
>Pb:Pb2:+:3:2
-GTTAATTAc
>Pb:Pb3:+:3:3
-TTTAATTAc
>Pb:Pb4:+:3:4
-GTTAATGAc
>Pb:pBG2:-:3:5
tCCTAATTA-
>Pb:pBG3:+:2:6
-TTTAATGAg
>Pb:pBG4:+:3:7
-TATAATTAc
>Pb:pBG5:-:3:8
tGTTAATTA-
>Pb:pBG6:+:3:9
-ATTAATTAc
>Pb:pBG7:-:3:10
aCATAATGA-
>Pb:pBG8:-:3:11
gACTAATGA-
>Pb:pBG9:-:3:12
cGTTAATGA-
>Pb:pBG11:+:3:13
-CTTAATGAg
>Pb:pBG12:-:1:14
-GCTAATTAa
>Pb:pBH1:+:3:15
-TTTAATTAc
>Pb:pBH3:+:3:16
-TATAATTAc
>Pb:pBH4:-:1:17
-GATAATTAt
>Pb:pBH6:-:3:19
tGCTAATGA-
>Pb:pBH7:-:1:20
-GCTAATTAa
>Pb:pBH8:-:3:21
gGTTAATGA-
>Pb:pBH9:+:1:22
cATTAATGA-
>Pb:pBH10:+:3:23
-TCTAATGAg
>Pb:pBH11:+:3:24
-TCTAATTAc
>Pb:pBH12:-:3:25
tGCTAATTA-
>Lab:LabE1:+:1:1
cgTTAATGA-
>Lab:LabE2:-:4:2
ccTTAATTA-
>Lab:LabE3:-:4:3
ggCTAATTA-
>Lab:LabE4:+:3:4
-aTTAATTAt
>Lab:LabE6:-:4:6
atTTAATTA-
>Lab:LabE7:+:3:7
-aTTAATTAg
>Lab:LabE8:-:4:8
agTTAATTA-
>Lab:LabE9:-:4:9
taCTAATTA-
>Lab:LabE10:+:4:10
--TTAAAGAa
>Lab:LabF1:+:3:12
-cTTAATGAc
>Lab:LabF5:-:4:16
tgTTAATTA-
>Lab:LabF9:+:3:20
-aTTAATGAc
>Lab:LabF11:-:4:22
gtCTAATGA-
>Lab:LabH6:-:4:25
tgTTAATTA-
>Lab:LabH7:-:4:26
tcTTAATTA-
>Lab:LabH12:-:4:31
acATAATGA-
>AbdA:AbdAG02:-:3:1
-tTTAATTAc
>AbdA:AbdAG03:+:4:2
cgTTAATGA-
>AbdA:AbdAG04:+:4:3
ttTTAATTA-
>AbdA:AbdAG06:+:1:5
--TTAATTAc
>AbdA:AbdAG07:+:4:6
tcTTTATTA-
>AbdA:AbdAG10:+:3:9
caCTAATTA-
>AbdA:AbdAG11:+:2:10
-cATAATTA-
>AbdA:AbdAG12:+:4:11
ttTTAATTA-
>AbdA:AbdAH03:-:3:14
-tTTAATGAc
>AbdA:AbdAH04:+:4:15
ttTTTATGA-
>AbdA:AbdAH05:+:4:16
taCTAATTC-
>AbdA:AbdAH06:-:3:17
-tTTAATTGc
>AbdA:AbdAH07:-:4:18
--TTAAAGAa
>AbdA:AbdAH08:+:4:19
atTTAATTA-
>AbdA:AbdAH09:+:4:20
cgCTAATGA-
>AbdA:AbdAH10:+:4:21
gtTTAATGA-
>AbdA:AbdAH11:-:3:22
-cTTAATTGc
>AbdA:AbdAH12:+:4:23
tcTTAATTAc
>Ap:ApA2:-:2:1
gGTTAATGAt
>Ap:ApA3:+:3:2
gACTAATTG-
>Ap:ApA4:+:3:3
gGTTAATTA-
>Ap:ApA5:+:3:4
aAATAATGA-
>Ap:ApA6:-:2:5
cGCTAATTAg
>Ap:ApA7:+:3:6
tGCTAATTG-
>Ap:ApA8:+:1:7
-GCTAATTAa
>Ap:ApA9:-:2:8
tCATAATTGg
>Ap:ApA10:-:2:9
cCTTAATTAg
>Ap:ApA11:+:1:10
-ACAAATTAa
>Ap:ApB3:+:3:12
cCTTAATGA-
>Ap:ApB4:-:2:13
aCTTAATTAg
>Ap:ApB5:-:3:14
-TTTAATGAg
>Ap:ApB6:+:1:15
-ACTAATTAa
>Ap:ApB7:+:3:16
cGTTAATGA-
>Ap:ApB8:+:3:17
cGCTAATTA-
>Ap:ApB9:+:3:18
gACTAATTA-
>Ap:ApB10:+:2:19
cGCTAATGA-
>Ap:ApB11:-:2:20
tGCTAATTAg
>Ind:IndC1:+:2:1
GATTAATTA-
>Ind:IndC2:+:1:2
CGCTAATGA-
>Ind:IndC3:+:2:3
ACCTAATGA-
>Ind:IndC4:-:1:4
TATTAAGTG-
>Ind:IndC5:-:1:5
CTCTAATTA-
>Ind:IndC7:+:2:7
CGTTAATGA-
>Ind:IndC8:+:2:8
GATTAATGA-
>Ind:IndC9:+:2:9
CGTTAATGA-
>Ind:IndC10:-:2:10
TACTAATTAc
>Ind:IndC11:-:2:11
TATTAATTAc
>Ind:IndC12:-:2:12
TGTTAATGAg
>Ind:IndD1:-:2:13
CCTTAATTAg
>Ind:IndD3:-:2:14
TTCTAATTAc
>Ind:IndD4:+:2:15
CACTAATGA-
>Ind:IndD5:+:2:16
TGCTAATTA-
>Ind:IndD6:+:2:17
TCCTAATTA-
>Ind:IndD7:+:2:18
CACTAATGA-
>Ind:IndD8:-:2:19
TATTAATTGc
>Ind:IndD9:+:2:20
CACTAATTA-
>Ind:IndD10:+:2:21
CCCTAATTA-
>Ind:IndD11:-:2:22
TGCTAATTAg
>CG18599:CG18599C1:-:3:1
-tTTAATTGa
>CG18599:CG18599C2:+:4:2
tgCTAATGA-
>CG18599:CG18599C3:+:4:3
tgCTAATTA-
>CG18599:CG18599C4:+:4:4
aaTTAATTA-
>CG18599:CG18599C5:+:4:5
ctTTAATTA-
>CG18599:CG18599C6:+:1:6
--TTAATTAa
>CG18599:CG18599C7:+:3:7
aaTTAATTA-
>CG18599:CG18599C8:+:4:8
ccCTAATTG-
>CG18599:CG18599C9:+:4:9
gaTTAATGA-
>CG18599:CG18599C10:+:4:10
atCTAATTA-
>CG18599:CG18599C11:+:4:11
caCTAATGA-
>CG18599:CG18599C12:+:4:12
cgCTAATTA-
>CG18599:CG18599D2:-:2:13
ttATAATGAg
>CG18599:CG18599D9:+:4:16
ccGTAATTA-
>CG18599:CG18599D10:+:1:17
--TTAATCAc
>CG18599:CG18599D11:+:4:18
acATAATGA-
>CG18599:CG18599F1:+:4:19
cgTTAATTA-
>CG18599:CG18599F2:+:4:20
taATAATGA-
>CG18599:CG18599F3:+:4:21
cgCTAATTA-
>CG18599:CG18599F4:+:4:22
cgTTAATTA-
>CG18599:CG18599F5:+:3:23
caCTAATTA-
>CG18599:CG18599F6:+:1:24
--CTAATTAg
>CG18599:CG18599F7:+:3:25
gcTTAATGA-
>CG18599:CG18599F11:+:4:27
ttATAATGA-
>CG18599:CG18599F12:+:4:28
gcCTAATTA-
>Lbe:IbeG2:-:3:1
-CATAATCAt

```
>Lbe:IbeG3:+:3:2
aTCTAAGTA-
>Lbe:IbeG4:+:3:3
gGTTAACCA-
>Lbe:IbeG5:+:2:4
tCCTAATCAc
>Lbe:IbeG6:+:2:5
cGTTAAATGa
>Lbe:IbeG7:-:3:6
-GGTAATTAc
>Lbe:IbeG8:+:3:7
cTATAAGTA-
>Lbe:IbeG9:-:3:8
-TGTAACAAg
>Lbe:IbeG10:-:3:9
-GTTAACCAg
>Lbe:IbeG11:+:3:10
gCCTAATTA-
>Lbe:IbeG12:+:4:11
gATTAACTA-
>Lbe:IbeH1:+:3:12
cACTAACAA-
>Lbe:IbeH2:-:3:13
-ACTAACGAg
>Lbe:IbeH3:+:3:14
aCATAATCA-
>Lbe:IbeH4:+:3:15
cTTTAACAA-
>Lbe:IbeH5:+:1:16
-GCTAATTAa
>Lbe:IbeH6:-:3:17
-GTTAATTGg
>Lbe:IbeH7:-:3:18
-GTTAATTAg
>Lbe:IbeH8:-:1:19
-GATAACAA-
>Lbe:IbeH9:-:3:20
-GTTAATTGg
>Lbe:IbeH10:-:3:21
-TTTAACGAg
>Lbe:IbeH11:-:3:22
-CTTAACGAg
>Lbl:LblC1:-:3:1
-GATAATTAt
>Lbl:LblC2:-:3:2
-TATAATTAc
>Lbl:LblC3:+:3:3
gGATAATTG-
>Lbl:LblC4:+:3:4
cACTAATCA-
>Lbl:LblC5:+:1:5
-GCTAATTAt
>Lbl:LblC6:+:3:6
tGCTAATTG-
>Lbl:LblC7:+:1:7
-GTTAATTAa
>Lbl:LblC8:+:3:8
aACTAACGA-
>Lbl:LblC9:-:3:9
-CTTAATCAg
>Lbl:LblC10:-:2:10
-GATAATTGg
>Lbl:LblC11:-:2:11
-GTTAACAAg
>Lbl:LblC12:-:3:12
-CTTAACGAg
>Lbl:LblD1:-:3:13
-TTTAATTGg
>Lbl:LblD2:-:3:14
-CTTAACGAg
>Lbl:LblD3:-:3:15
-CTTAATTGg
>Lbl:LblD4:-:3:16
-GATAATTGg
>Lbl:LblD5:+:3:17
gACTAATGA-
>Lbl:LblD6:+:3:18
tGCTAATCA-
>Lbl:LblD7:+:1:19
-GCTAATTAa
>Lbl:LblD8:+:1:20
-GCTAATTAa
>Lbl:LblD9:+:3:21
gACTAATGA-
>Lbl:LblD10:-:3:22
-CTTAACGAg
>Lbl:LblD11:+:3:23
gACTAATGA-
>Eve:Eve-G2:+:2:2
aATTAAGTAa
>Eve:Eve-G3:+:2:3
tAATAATCGa
>Eve:Eve-G4:-:3:4
tTCTAATCA-
>Eve:Eve-G5:-:3:5
aGTAAATTA-
>Eve:Eve-G6:-:3:6
tTCTAACGA-
>Eve:Eve-G7:+:3:7
-TATAATGAt
>Eve:Eve-G8:+:2:8
tACTAACGAc
>Eve:Eve-G9:-:3:9
gGCTAATTG-
>Eve:Eve-G10:+:2:10
gTCTAATTGa
>Eve:Eve-G11:+:3:11
-GTTAATGTg
>Eve:Eve-G12:+:3:12
-CATAATGAg
>Eve:Eve-H1:-:1:13
-GTTAATTAa
>Eve:Eve-H2:+:3:14
-GTTAATGGg
>Eve:Eve-H3:+:3:15
-TTTAATGAc
>Eve:Eve-H4:-:3:16
gTTTAATGA-
>Eve:Eve-H5:-:3:17
gGCTAATTA-
>Eve:Eve-H6:-:3:18
tGCTAATTA-
>Eve:Eve-H7:-:3:19
cACTAATTA-
>Eve:Eve-H8:+:2:20
tACTAATTAc
>Eve:Eve-H9:-:3:21
tGCTAATGA-
>Eve:Eve-H10:+:3:22
-ATTAATGAg
>Eve:Eve-H11:+:3:23
-GTTAATGAc
>E5:E5C1:-:3:1
-CCTAATTGa
>E5:E5C2:-:2:2
aTTTAATTAa
>E5:E5C3:-:2:3
tGTAAATTAg
>E5:E5C4:-:1:4
gCTTAATGG-
>E5:E5C5:+:1:5
-ATAAATTAa
>E5:E5C6:+:3:6
tGTTAATAA-
>E5:E5C7:+:3:7
cCGTAATTA-
>E5:E5C8:+:2:8
gCTTAAGTA-
>E5:E5C9:-:3:9
-GCTAATTGa
>E5:E5C10:+:3:10
cTGTAATGA-
>E5:E5C11:+:3:11
tTATAATTA-
>E5:E5C12:-:2:12
tTCTAATTAa
>E5:E5D2:+:3:14
gTATAATGA-
>E5:E5D3:-:1:15
cACTATTGA-
>E5:E5D4:+:3:16
tGTTAATTG-
>E5:E5D5:+:3:17
aGCTAATGA-
>E5:E5D7:+:3:19
gGCTAATTA-
>E5:E5D8:+:3:20
tATTAATTA-
>E5:E5D9:+:3:21
aTCTAATTA-
>E5:E5D10:+:3:22
cACTAATGA-
>E5:E5D11:+:3:23
gACTAATGA-
>E5:E52C1:+:3:24
gACTAATAG-
>E5:E52C2:-:2:25
gGTTAATGGg
>E5:E52C3:+:3:26
gGCTAAATA-
>E5:E52C4:-:2:27
aCGTAATGAa
>E5:E52C5:-:3:28
-CTTAATAGt
>E5:E52C6:+:3:29
aTTAAATGA-
>E5:E52C7:+:3:30
aCTTAATAA-
>E5:E52C8:+:3:31
cAATAATGA-
>E5:E52C10:-:3:33
-ACTAATGAa
>E5:E52C11:+:3:34
gGGTAATTA-
>E5:E52C12:+:3:35
gTTTAATGA-
>E5:E52D1:+:3:36
aTGTAATGA-
>E5:E52D2:+:3:37
tTTTAATTA-
>E5:E52D3:+:3:38
gCATAATTA-
>E5:E52D4:+:3:39
cGTTAATAG-
>E5:E52D5:+:3:40
aACTAATTA-
>E5:E52D6:-:2:41
tCTTAATTGa
>E5:E52D7:+:3:42
cATTAATGA-
>E5:E52D8:+:3:43
cAATAATTA-
>E5:E52D9:+:3:44
aTCTAATGA-
>E5:E52D10:+:3:45
cGGTAATTA-
>E5:E52D11:-:3:46
-GCTAATTAc
>BH1:B1HG8:+:1:8
-GCTAATTGA
>BH1:B1HG9:+:1:9
-GTTAATTGA
>BH1:B1HG10:+:2:10
tGTTAAACGG
>BH1:B1HH9:-:2:20
-CTTAATTGC
>BH1:BH12E2:+:2:22
tTCTAAACGG
>BH1:BH12E3:-:2:23
-GTTAATTGG
>BH1:BH12E4:+:2:24
gGCTAATTGA
>BH1:BH12E5:+:2:25
aGTTAATAGG
>BH1:BH12E6:-:2:26
-GTTAATTGT
>BH1:BH12E7:-:2:27
-GTTAATTGA
>BH1:BH12E8:-:2:28
-CATAATTGC
>BH1:BH12E9:+:2:29
tGTTAAACGG
>BH1:BH12E10:+:2:30
tCTTAAACGG
>BH1:BH12E11:+:2:31
tCTTAAACGG
>BH1:BH12E12:+:2:32
aGATAATTGC
>BH1:BH12F1:+:2:33
gGATAATTGA
>BH1:BH12F2:+:2:34
aCTTAAACGT
>BH1:BH12F3:+:1:35
-GATAATTAA
>BH1:BH12F4:+:2:36
cTTTAAACGG
>BH1:BH12F6:-:2:38
-ATTAAATGT
>BH1:BH12F11:+:2:42
tTCTAATTGA
>BH2:BH2E1:+:1:1
-GATAAACGG
>BH2:BH2E2:+:1:2
-TTTAATTGC
>BH2:BH2E4:+:1:3
-CTTAATAGT
>BH2:BH2E5:+:1:4
-ACTAAATGG
>BH2:BH2E6:+:1:5
-TTTAATAGG
>BH2:BH2E7:+:1:6
-CTTAATGGC
>BH2:BH2E9:+:1:8
-CTTAAAAGG
>BH2:BH2E10:+:1:9
-ATTAATTGG
>BH2:BH2E11:+:1:10
-GATAAATGA
>BH2:BH2E12:+:1:11
-TCTAATGGG
>BH2:BH2F1:+:1:12
-GATAATTGG
>BH2:BH2F2:+:1:13
-CTTAAATGA
>BH2:BH2F3:+:1:14
-TGTAATTGG
>BH2:BH2F4:+:1:15
-CATAATTGG
>BH2:BH2F5:+:1:16
-GATAATTGG
>BH2:BH2F6:+:1:17
-TATAATTGC
>BH2:BH2F7:+:1:18
-GTTAATTGA
>BH2:BH2F8:+:1:19
-GTTAATTGC
>BH2:BH2F9:+:1:20
-GTTAATTGA
>BH2:BH2F10:+:1:21
-CTTAATTGA
>BH2:BH2F11:+:1:22
-CTTAATTGG
>CG11085:CG11085G1:+
:1:1
-CTTAATGGG
>CG11085:CG11085G3:+
:1:3
-TTTAATTGG
>CG11085:CG11085G4:+
:1:4
-TTTAATTGC
>CG11085:CG11085G5:+
:1:5
-TTTAATAGG
>CG11085:CG11085G6:+
:1:6
-CTTAATGGG
>CG11085:CG11085G7:+
:1:7
-TTTAATTAC
>CG11085:CG11085G8:+
:1:8
-TCTAATAGC
```

```
>CG11085:CG11085G9:+
:1:9
-CTTAATCGG
>CG11085:CG11085G11:
+:1:11
-CCTAATTGC
>CG11085:CG11085H2:+
:1:14
-GCTAATTGA
>CG11085:CG11085H4:+
:1:16
-GTTAATTGG
>CG11085:CG11085H6:+
:1:18
-ATTAAATGT
>CG11085:CG11085H11:
+:1:23
-ATTAATTGA
>CG34031:CG34031E1:+
:3:1
-TGTAATTGt
>CG34031:CG34031E2:+
:2:2
cTTTAATTGc
>CG34031:CG34031E3:+
:3:3
-GTTAATTAg
>CG34031:CG34031E4:-
:3:4
tTTTTATTG-
>CG34031:CG34031E5:+
:3:5
-ATTAAATGt
>CG34031:CG34031E6:-
:3:6
tTTTAATAG-
>CG34031:CG34031E7:-
:3:7
cTTTTATAG-
>CG34031:CG34031E8:+
:2:8
cTTTAATAGt
>CG34031:CG34031F3:+
:3:15
-GTTAATTGc
>CG34031:CG34031F4:+
:3:16
-CTTAATGGt
>CG34031:CG34031A1:+
:2:23
tATTAATTAg
>CG34031:CG34031A2:+
:3:24
-CTTAATAGa
>CG34031:CG34031A3:+
:2:25
gTTTAATTGg
>CG34031:CG34031A4:+
:2:26
gTTTAATTGc
>CG34031:CG34031A5:+
:3:27
-TTTAATTGc
>CG34031:CG34031A6:+
:3:28
-TCTAATTGt
>CG34031:CG34031A7:+
:2:29
gTTTAATTGc
>CG34031:CG34031A8:+
:3:30
-TTTTATAGt
>CG34031:CG34031A10:
-:3:32
tGTTAATTG-
>CG34031:CG34031A11:
+:3:33
-CTTAATTGt
>CG34031:CG34031A12:
+:3:34
-TTTAATAGg
>CG34031:CG34031B2:+
:3:35
-CTTAATTGc
>CG34031:CG34031B3:+
:3:36
-GTTAATTGa
>CG34031:CG34031B5:+
:3:37
-TTTAAATAg
>CG34031:CG34031B6:+
:3:38
-TTTAATAGa
>Hmx:HMXA2:+:4:1
tgTTAATTG-
>Hmx:HMXA3:+:4:2
agATAATTG-
>Hmx:HMXA4:+:4:3
cgCTAATTG-
>Hmx:HMXA5:-:4:4
--TTAAATGg
>Hmx:HMXA6:+:4:5
taTTAATTG-
>Hmx:HMXA7:+:4:6
ccTTAATTG-
>Hmx:HMXA8:+:4:7
tgTTAATTG-
>Hmx:HMXA12:+:1:10
--TTAATTG-
>Hmx:HMXE1:+:4:11
cgTTAATTA-
>Hmx:HMXE2:+:4:12
acTTAATCG-
>Hmx:HMXE3:+:4:13
ctCTAATTG-
>Hmx:HMXE4:+:4:14
acTTAATCG-
>Hmx:HMXE5:+:4:15
taTTAATCG-
>Hmx:HMXE6:+:4:16
acTTAATTA-
>Hmx:HMXE7:+:4:17
agTTAATTA-
>Hmx:HMXE8:+:4:18
caCTAATTA-
>Hmx:HMXE9:+:4:19
gtTTAATTG-
>Hmx:HMXE10:+:4:20
atTTAATTG-
>Hmx:HMXF4:-:3:25
-cTTAATTGc
>Hmx:HMXF8:-:3:28
-cTTAATTGc
>Unc4:Unc4C1:-:1:1
-TCTAATTAg
>Unc4:Unc4C2:-:3:2
cTGTAATTA-
>Unc4:Unc4C3:+:3:3
-CTTAATTCg
>Unc4:Unc4C4:-:1:4
-GCTAATTAt
>Unc4:Unc4C5:-:1:5
-CTTAATTAt
>Unc4:Unc4C6:+:3:6
-CTTAATAGa
>Unc4:Unc4C7:+:2:7
tTTTAATTGa
>Unc4:Unc4C8:-:3:8
tCTTAATTG-
>Unc4:Unc4C9:+:3:9
-CCTAATTGa
>Unc4:Unc4C10:+:3:10
-TGTAATTGa
>Unc4:Unc4C11:+:3:11
-GTTAATTGc
>Unc4:Unc4C12:+:3:12
-CCTAATTGa
>Unc4:Unc4D1:-:3:13
cACTAATTA-
>Unc4:Unc4D3:+:3:15
-AATAATTGg
>Unc4:Unc4D4:+:3:16
-CTTAATTGa
>Unc4:Unc4D5:+:3:17
-GTTAATTGa
>Unc4:Unc4D6:+:3:18
-CTTAATTGg
>Unc4:Unc4D7:+:3:19
-TTTAATTGg
>Unc4:Unc4D8:+:3:20
-GTTAATTGg
>Unc4:Unc4D10:+:3:22
-GTTAATTGa
>Unc4:Unc4D11:+:3:23
-CTTAATTAg
>Odsh:OdshC1:-:4:1
ttGTAATTA-
>Odsh:OdshC2:+:2:2
-tTTAATTTc
>Odsh:OdshC3:+:2:3
tgCTAATTAt
>Odsh:OdshC4:-:3:4
taCTAATTAa
>Odsh:OdshC5:-:4:5
aaCTAATTG-
>Odsh:OdshC7:-:3:6
agCTAATTA-
>Odsh:OdshC8:-:2:7
-cCTAATTAc
>Odsh:OdshC9:+:2:8
ccTTAATTGc
>Odsh:OdshC10:+:3:9
-gGTAATTAc
>Odsh:OdshC11:-:4:10
atTTAATTA-
>Odsh:OdshC12:+:3:11
-cCTAATTGt
>Odsh:OdshD1:+:3:12
-aTTAATTGt
>Odsh:OdshD2:+:2:13
atTTAATTGg
>Odsh:OdshD3:-:3:14
tgCTAATTAa
>Odsh:OdshD4:-:4:15
acTTAATTA-
>Odsh:OdshD5:-:2:16
-cCTAATTAa
>Odsh:OdshD6:-:2:17
-tCTAATTAa
>Odsh:OdshD7:-:4:18
cgCTAATTA-
>Odsh:OdshD8:+:3:19
-aCTAATTGa
>Odsh:OdshD9:+:3:20
-tGTAATTGa
>Odsh:OdshD10:-:2:21
-cTTAATTAa
>Odsh:OdshD11:+:3:22
-gGTAATTGa
>Dr:DrA2:+:1:1
CCTCAATTA-
>Dr:DrA3:+:1:2
AAGCAATTA-
>Dr:DrA4:+:1:3
GGCCAATTA-
>Dr:DrA5:+:1:4
AACTAATTA-
>Dr:DrA6:+:1:5
CTCCAATTA-
>Dr:DrA7:+:1:6
CACCAATTA-
>Dr:DrA8:+:1:7
GAGCAATTA-
>Dr:DrA9:+:1:8
GGGTAATTA-
>Dr:DrA10:+:1:9
GACTAATTA-
>Dr:DrA11:+:1:10
CGCTAATTA-
>Dr:DrB2:+:1:11
CTCCAATTA-
>Dr:DrB3:+:1:12
GGCCAATTA-
>Dr:DrB4:+:1:13
AAACAATTA-
>Dr:DrB5:+:1:14
AACCAATTA-
>Dr:DrB6:+:1:15
GAGCAATTA-
>Dr:DrB7:+:1:16
GACCAATTA-
>Dr:DrB8:+:1:17
CTCCAATTA-
>Dr:DrB9:+:1:18
GAGTAATTA-
>Dr:DrB10:+:1:19
CAGCAATTA-
>Dr:DrB11:+:1:20
CCCCAATTA-
>Dr:DrB12:+:1:21
GTCCAATTA-
```

Table S5. Crossvalidation analysis of fly homeodomains.

Query	ALLR	Distance	p-value	e-value	Pred. Cons.	Actual Cons.	Combined Cons.
Repo	9.9809	0.0512	7.90E-009	7.90E-009	TAATTa	TAATTA	TAATTa
Rx	9.9038	0.1058	9.29E-009	9.29E-009	TAATTa	TAATTR	TAATTa
Hbn	9.8023	-0.2713	1.15E-008	1.15E-008	TAATTa	TAATTR	TAATTa
Al	9.7524	0.4103	1.28E-008	1.28E-008	TAATTR	TAATTA	TAATTa
Ptx1	9.6421	0.6916	1.61E-008	1.61E-008	TAATCC	TAATCC	TAATCC
Oc	9.6284	0.3647	1.65E-008	1.65E-008	TAATCC	TAATCC	TAATCC
Pph13	9.5617	0.1223	1.90E-008	1.90E-008	TAATTA	TAATTa	TAATTA
Gsc	9.5553	0.6155	1.93E-008	1.93E-008	TAATCC	TAATCc	TAATCC
Dfd	9.4429	0.2754	2.44E-008	2.44E-008	TAATgA	TAATGA	TAATGA
CG32532	9.4385	0.074	2.46E-008	2.46E-008	TAATTa	TAATTR	TAATTa
CG11294	9.3819	0.3812	2.77E-008	2.77E-008	TAATTA	TAATTA	TAATTA
Odsh	9.3406	0.2868	3.02E-008	3.02E-008	TAATTR	TAATTR	TAATTR
Scr	9.2996	0.3014	3.29E-008	3.29E-008	TAATKA	TAATGA	TAATKA
Ind	9.2047	0.2277	4.01E-008	4.01E-008	TAATKA	TAATKA	TAATKA
Zen2	9.1465	0.6703	4.53E-008	4.53E-008	TAATGA	TAATKA	TAATKA
Lab	9.1042	0.7574	4.95E-008	4.95E-008	TAATGA	TAATKA	TAATgA
Inv	9.0763	0.3488	5.25E-008	5.25E-008	TAATTR	TAATTa	TAATTR
En	9.0763	0.3488	5.25E-008	5.25E-008	TAATTa	TAATTR	TAATTR
Unc4	9.0702	0.8337	5.32E-008	5.32E-008	TAATTa	TAATTg	TAATTR
CG9876	9.067	0.3904	5.35E-008	5.35E-008	TAATTa	TAATTa	TAATTa
Otp	9.0266	0.456	5.83E-008	5.83E-008	TAATTA	TAATTA	TAATTA
Ftz	8.9761	0.0222	6.48E-008	6.48E-008	TAATKA	TAATKA	TAATKA
Zen	8.9689	0.699	6.57E-008	6.57E-008	TAATKA	TAATgA	TAATKA
Antp	8.8538	0.4076	8.37E-008	8.37E-008	TAATKA	TAATKA	TAATKA
PhdP	8.8293	0.1977	8.81E-008	8.81E-008	TAATTA	TAATTn	TAATTa
AbdA	8.7462	0.1147	1.05E-007	1.05E-007	TAATKA	TAATtA	TAATKA
Dr	8.6782	2.0874	1.21E-007	1.21E-007	TAATTA	TAATTG	TAATTR
Btn	8.6048	0.3618	1.41E-007	1.41E-007	TAATKA	TAATgA	TAATKA
Vis	8.5718	1.0373	1.51E-007	1.51E-007	TGACAg	TGACAn	TGACA
Achi	8.5718	1.0373	1.51E-007	1.51E-007	TGACAn	TGACAg	TGACA
Unpg	8.5501	1.4249	1.58E-007	1.58E-007	TAATTA	TAATTa	TAATTA
Ro	8.39	1.8609	2.21E-007	2.21E-007	YAATTA	TAATTA	tAATTA
Eve	8.3891	0.4253	2.21E-007	2.21E-007	TAATKA	TAATKA	TAATKA
CG33980	8.3842	0.9261	2.24E-007	2.24E-007	TAATTa	TAATTA	TAATTA
CG4136	8.3842	0.9261	2.24E-007	2.24E-007	TAATTA	TAATTa	TAATTA
Ubx	8.3467	0.5666	2.42E-007	2.42E-007	TAATKA	TAATta	TAATKA
Hth	8.3339	1.5492	2.48E-007	2.48E-007	TGACAg	TGACAg	TGACA
CG18599	8.325	1.2562	2.53E-007	2.53E-007	TAATKA	TAATtA	TAATKA
Pb	8.2429	1.6834	3.01E-007	3.01E-007	TAATKA	TAATKA	TAATKA
Awh	8.2022	1.0905	3.27E-007	3.27E-007	TAATtA	TAATTA	TAATTA
Ap	8.2022	1.0905	3.27E-007	3.27E-007	TAATTA	TAATtA	TAATTA
So	8.1879	2.4025	3.37E-007	3.37E-007	TGATAC	TGATAC	TGATAC
Tin	8.0367	1.177	4.63E-007	4.63E-007	YAAGTR	cAAGTG	YAAGTg
Bap	7.8641	2.6081	6.64E-007	6.64E-007	CAAGTg	TAAGTg	YAAGTg
Dll	7.8095	1.8343	7.45E-007	7.45E-007	YAATTA	YAATTA	YAATTA
BH2	7.7818	1.3614	7.89E-007	7.89E-007	TAAWtG	TAATtG	TAAttG

BH1	7.7818	1.3614	7.89E-007	7.89E-007	TAATtG	TAAWtG	TAAttG
Optix	7.7705	2.0452	8.08E-007	8.08E-007	TGATAC	TGATAn	TGATAC
CG34031	7.7609	1.7932	8.24E-007	8.24E-007	TAATtG	TAATtG	TAATtG
CG13424	7.7235	0.9239	8.92E-007	8.92E-007	TAATtR	TAATtR	TAATtR
Vnd	7.6846	1.9895	9.67E-007	9.67E-007	tAAGTG	CAAGTR	YAAGTg
Slou	7.4995	1.1247	1.43E-006	1.43E-006	TAATtR	TAATtR	TAATtR
Six4	7.4962	3.3833	1.44E-006	1.44E-006	TGATAC	TGAnAC	TGATAC
Ems	7.4919	0.9041	1.45E-006	1.45E-006	TAATKA	TAATKa	TAATKA
E5	7.4919	0.9041	1.45E-006	1.45E-006	TAATKa	TAATKA	TAATKA
CG11085	7.4246	1.0618	1.67E-006	1.67E-006	TAAttG	TAATtG	TAAttG
Hgtx	7.4223	1.6526	1.68E-006	1.68E-006	TAATtR	TAATtA	TAATtR
NK7	7.376	1.4885	1.85E-006	1.85E-006	TAATtR	TAATtR	TAATtR
Bsh	7.3414	1.302	1.98E-006	1.98E-006	TAATtR	TAATKR	TAATtR
Exex	7.2781	4.8019	2.27E-006	2.27E-006	TAATKA	TAATTA	TAAT
Lim1	7.1013	3.0796	3.28E-006	3.28E-006	TAATtA	TAATTA	TAATTA
Mirr	6.8517	0.3406	5.53E-006	5.53E-006	taACAn	WaACAn	WaACA
Ara	6.7625	0.493	6.67E-006	6.67E-006	WAACAn	WaACAn	WaACA
Caup	6.6398	0.3492	8.62E-006	8.62E-006	WaACAn	tAACAn	WaACA
Lbe	6.5811	1.563	9.75E-006	9.75E-006	TAATtA	TAAYnA	TAAtnA
Lbl	6.5811	1.563	9.75E-006	9.75E-006	TAAYnA	TAATtA	TAAtnA
Lim3	6.5461	2.1844	1.05E-005	1.05E-005	TWATTR	TAATtA	TaATTa
C15	6.4381	2.6673	1.32E-005	1.32E-005	TAATtG	TAAttR	TAAttG
CG32105	6.4272	2.9105	1.35E-005	1.35E-005	TWATTR	TAATTA	TWATTR
CG4328	6.4272	2.9105	1.35E-005	1.35E-005	TAATTA	TWATTR	TWATTR
Hmx	6.1889	4.7795	2.22E-005	2.22E-005	TAAKTR	TAATTG	TAAtTg
H2	5.7098	1.9904	6.04E-005	6.04E-005	TWATKA	TWATnA	TWATnA
AbdB	5.647	4.8563	6.89E-005	6.89E-005	TAATKA	TTATga	TAATKA
CG12361	5.2354	5.7449	0.000163	0.000163	TAATTA	TMATWA	TAATtA
Cad	5.0652	6.9451	0.0002328	0.0002328	TAATgA	TtATtR	TAATKA

Table S6. Fly factors used to predict the human homeodomains and confidence scores.

Confidence Score	Human HD Query	fly HD Ref	Similarity Score	#Key Residue matches	Ref. Key Residues (5,47,50,54 &55)	#Sites Contributed by Ref
1	ALX3				RVQAK	
		Pph13	316	5	RVQAK	21
		Hbn	311	5	RVQAK	17
		Al	310	5	RVQAK	20
1	ALX4				RVQAK	
		Rx	315	5	RVQAK	27
		Al	314	5	RVQAK	20
		Pph13	311	5	RVQAK	21
	ARGFX				RVRFK	
	No predictions made					
1	ARX				RVQAK	
		Al	349	5	RVQAK	20
		Pph13	348	5	RVQAK	21
		Hbn	333	5	RVQAK	17
1	BARHL1				RTQTK	
		BH2	320	5	RTQTK	21
		BH1	310	5	RTQTK	21
1	BARHL2				RTQTK	
		BH2	314	5	RTQTK	21
		BH1	304	5	RTQTK	21
3	BARX1				RTQMK	
		Bsh	254	5	RTQMK	16
3	BARX2				RTQMK	
		Bsh	242	5	RTQMK	16
2	BSX				RTQMK	
		Bsh	339	5	RTQMK	16
2	CDX1				RIQAK	
		Cad	303	5	RIQAK	38
2	CDX2				RIQAK	
		Cad	308	5	RIQAK	38
2	CDX4				RIQAK	
		Cad	284	5	RIQAK	38
2	CRX				RVKAK	

		Oc	327	5	RVKAK	19
	CUTL1				RNHSR	
	No predictions made					
4	CUTL2				RNHSR	
		Ct	203	4	RNHMR	20
2	DBX1				RIQMK	
		CG12361	334	5	RIQMK	16
2	DBX2				RIQMK	
		CG12361	306	5	RIQMK	16
2	DLX1				RIQSK	
		Dll	332	5	RIQSK	23
2	DLX2				RIQSK	
		Dll	316	5	RIQSK	23
2	DLX3				RIQSK	
		Dll	304	5	RIQSK	23
2	DLX4				RIQSK	
		Dll	320	5	RIQSK	23
2	DLX5				RIQSK	
		Dll	313	5	RIQSK	23
2	DLX6				RIQSK	
		Dll	321	5	RIQSK	23
3	DMBX1				RVKAK	
		Gsc	260	5	RVKAK	22
		Ptx1	252	5	RVKAK	20
		Oc	243	5	RVKAK	19
1	DRGX				RVQAK	
		Al	312	5	RVQAK	20
		CG11294	308	5	RVQAK	15
		Pph13	295	5	RVQAK	21
1	EMX1				RVQTK	
		E5	311	5	RVQTK	43
		Ems	300	5	RVQTK	20
1	EMX2				RVQTK	
		E5	323	5	RVQTK	43
		Ems	307	5	RVQTK	20

3	EN1				RIQAK	
		En	283	5	RIQAK	23
		Inv	272	5	RIQAK	16
3	EN2				RIQAK	
		En	287	5	RIQAK	23
		Inv	276	5	RIQAK	16
3	ESX1				RVQAK	
		Hbn	265	5	RVQAK	17
		Repo	264	5	RVQAK	28
		Rx	260	5	RVQAK	27
2	EVX1				RVQMK	
		Eve	354	5	RVQMK	22
2	EVX2				RVQMK	
		Eve	358	5	RVQMK	22
2	GBX1				RIQAK	
		Unpg	338	5	RIQAK	21
2	GBX2				RIQAK	
		Unpg	337	5	RIQAK	21
2	GSC				RVKAK	
		Gsc	300	5	RVKAK	22
2	GSCL				RVKAK	
		Gsc	282	5	RVKAK	22
2	GSX1				RIQVK	
		Ind	300	5	RIQVK	21
2	GSX2				RIQVK	
		Ind	299	5	RIQVK	21
	HESX1				RIQAK	
	No predictions made					
3	HHEX				QTQAK	
		CG7056	267	5	QTQAK	26
2	HLX				RVQMK	
		H2	308	5	RVQMK	32
	HMBOX1				RNAKE	
	No predictions made					
2	HMX1				RIQNK	
		Hmx	346	5	RIQNK	20

3	HMX2				RTQNK	
		Hmx	321	4	RIQNK	20
2	HMX3				RIQNK	
		Hmx	345	5	RIQNK	20
	HNF1A				RNAKE	
	No predictions made					
	HNF1B				RNAKE	
	No predictions made					
2	HOXA1				RIQMK	
		Lab	315	5	RIQMK	16
3	HOXA10				RIQMK	
		AbdB	259	5	RIQMK	21
		Antp	248	5	RIQMK	16
		Scr	248	5	RIQMK	25
		AbdA	247	5	RIQMK	18
4	HOXA11				RIQMK	
		AbdB	251	5	RIQMK	21
		Scr	218	5	RIQMK	25
		Ftz	213	5	RIQMK	18
	HOXA13				RIQVK	
	No predictions made					
2	HOXA2				RVQMK	
		Pb	365	5	RVQMK	24
3	HOXA3				RIQMK	
		Scr	279	5	RIQMK	25
		Dfd	278	5	RIQMK	24
		Ftz	273	5	RIQMK	18
1	HOXA4				RIQMK	
		Dfd	355	5	RIQMK	24
		Scr	351	5	RIQMK	25
		Antp	334	5	RIQMK	16
1	HOXA5				RIQMK	
		Scr	361	5	RIQMK	25
		Antp	360	5	RIQMK	16
		Dfd	342	5	RIQMK	24
1	HOXA6				RIQMK	
		Antp	375	5	RIQMK	16
		Scr	351	5	RIQMK	25
		AbdA	341	5	RIQMK	18

1	HOXA7				RIQMK	
		Antp	390	5	RIQMK	16
		Scr	372	5	RIQMK	25
		AbdA	356	5	RIQMK	18
3	HOXA9				RIQMK	
		AbdB	272	5	RIQMK	21
		Scr	260	5	RIQMK	25
		Antp	254	5	RIQMK	16
2	HOXB1				RIQMK	
		Lab	303	5	RIQMK	16
	HOXB13				RIQVK	
	No predictions made					
2	HOXB2				RVQMK	
		Pb	365	5	RVQMK	24
3	HOXB3				RIQMK	
		Dfd	280	5	RIQMK	24
		Scr	276	5	RIQMK	25
		Ftz	275	5	RIQMK	18
1	HOXB4				RIQMK	
		Dfd	355	5	RIQMK	24
		Scr	351	5	RIQMK	25
		Antp	334	5	RIQMK	16
1	HOXB5				RIQMK	
		Scr	361	5	RIQMK	25
		Antp	360	5	RIQMK	16
		Dfd	342	5	RIQMK	24
1	HOXB6				RIQMK	
		Antp	374	5	RIQMK	16
		Scr	353	5	RIQMK	25
		AbdA	345	5	RIQMK	18
1	HOXB7				RIQMK	
		Antp	385	5	RIQMK	16
		Scr	353	5	RIQMK	25
		AbdA	345	5	RIQMK	18
1	HOXB8				RIQMK	
		Antp	342	5	RIQMK	16
		AbdA	315	5	RIQMK	18
		Ubx	310	5	RIQMK	20
3	HOXB9				RIQMK	
		AbdB	267	5	RIQMK	21

		Scr	258	5	RIQMK	25
		Antp	255	5	RIQMK	16
3	HOXC10				RIQMK	
		AbdB	254	5	RIQMK	21
		Dfd	250	5	RIQMK	24
		Antp	248	5	RIQMK	16
3	HOXC11				RIQMK	
		AbdB	265	5	RIQMK	21
4	HOXC12				RIQMK	
		AbdB	232	5	RIQMK	21
		Zen2	207	5	RIQMK	26
		Ftz	202	5	RIQMK	18
	HOXC13				RIQVK	
	No predictions made					
1	HOXC4				RIQMK	
		Dfd	359	5	RIQMK	24
		Scr	349	5	RIQMK	25
		Antp	332	5	RIQMK	16
1	HOXC5				RIQMK	
		Scr	346	5	RIQMK	25
		Antp	340	5	RIQMK	16
		Dfd	322	5	RIQMK	24
1	HOXC6				RIQMK	
		Antp	370	5	RIQMK	16
		Scr	343	5	RIQMK	25
		AbdA	341	5	RIQMK	18
1	HOXC8				RIQMK	
		Antp	334	5	RIQMK	16
		AbdA	307	5	RIQMK	18
		Ubx	302	5	RIQMK	20
3	HOXC9				RIQMK	
		AbdB	277	5	RIQMK	21
		Scr	268	5	RIQMK	25
		Antp	262	5	RIQMK	16
2	HOXD1				RIQMK	
		Lab	306	5	RIQMK	16
3	HOXD10				RIQMK	
		AbdB	257	5	RIQMK	21
		AbdA	247	5	RIQMK	18
		Scr	246	5	RIQMK	25

4	HOXD11				RIQMK	
		AbdB	254	5	RIQMK	21
		Ftz	221	5	RIQMK	18
		Scr	221	5	RIQMK	25
3	HOXD12				RIQMK	
		AbdB	229	5	RIQMK	21
	HOXD13				RIQVK	
	No predictions made					
3	HOXD3				RIQMK	
		Scr	277	5	RIQMK	25
		Dfd	276	5	RIQMK	24
		Ftz	273	5	RIQMK	18
1	HOXD4				RIQMK	
		Dfd	357	5	RIQMK	24
		Scr	353	5	RIQMK	25
		Antp	336	5	RIQMK	16
1	HOXD8				RIQMK	
		Antp	339	5	RIQMK	16
		AbdA	322	5	RIQMK	18
		Ubx	307	5	RIQMK	20
3	HOXD9				RIQMK	
		AbdB	281	5	RIQMK	21
		Scr	265	5	RIQMK	25
		AbdA	256	5	RIQMK	18
3	IRX1				KTARR	
		Mirr	309	4	RTARR	41
		Ara	300	4	RTARR	34
		Caup	300	4	RTARR	19
1	IRX2				RTARR	
		Ara	330	5	RTARR	34
		Caup	330	5	RTARR	19
		Mirr	329	5	RTARR	41
3	IRX3				KTARR	
		Mirr	302	4	RTARR	41
		Ara	301	4	RTARR	34
		Caup	301	4	RTARR	19
1	IRX4				RTARR	
		Mirr	352	5	RTARR	41
		Ara	335	5	RTARR	34
		Caup	335	5	RTARR	19
1	IRX5				RTARR	

		Ara	330	5	RTARR	34
		Caup	330	5	RTARR	19
		Mirr	329	5	RTARR	41
1	IRX6				RTARR	
		Mirr	344	5	RTARR	41
		Ara	340	5	RTARR	34
		Caup	340	5	RTARR	19
2	ISL1				RVQCK	
		Tup	353	5	RVQCK	16
2	ISL2				RVQCK	
		Tup	355	5	RVQCK	16
4	ISX				RIQAK	
		Unpg	204	5	RIQAK	21
1	LBX1				RTQAK	
		Lbe	332	5	RTQAK	22
		Lbl	327	5	RTQAK	23
3	LBX2				RTQAK	
		Lbl	287	5	RTQAK	23
		Lbe	280	5	RTQAK	22
2	LHX1				RVQSK	
		Lim1	330	5	RVQSK	18
2	LHX2				RVQAK	
		Ap	345	5	RVQAK	19
2	LHX3				RVQAK	
		Lim3	339	5	RVQAK	20
2	LHX4				RVQAK	
		Lim3	337	5	RVQAK	20
2	LHX5				RVQSK	
		Lim1	330	5	RVQSK	18
3	LHX8				RVQAR	
		Awh	253	5	RVQAR	40
2	LHX9				RVQAK	
		Ap	345	5	RVQAK	19
1	LMX1A				RVQAK	
		CG4328	317	5	RVQAK	30
		CG32105	308	5	RVQAK	19

1	LMX1B				RVQAK	
		CG4328	317	5	RVQAK	30
		CG32105	308	5	RVQAK	19
2	MEIS1				RNIRR	
		Hth	366	5	RNIRR	17
2	MEIS2				RNIRR	
		Hth	371	5	RNIRR	17
	MEIS3				DNIRR	
	No predictions made					
2	MEIS3P2				RNIRR	
		Hth	335	5	RNIRR	17
2	MEOX1				RVQMK	
		Btn	295	5	RVQMK	23
3	MIXL1				RVQAK	
		CG11294	234	5	RVQAK	15
		Pph13	231	5	RVQAK	21
		CG32532	229	5	RVQAK	23
	MKX				RNARR	
	No predictions made					
2	MNX1				RIQMK	
		Exex	339	5	RIQMK	23
2	MSX1				RIQAK	
		Dr	333	5	RIQAK	21
2	MSX2				RIQAK	
		Dr	335	5	RIQAK	21
	NANOG				RTQMR	
	No predictions made					
	NANOGP1				RTQMR	
	No predictions made					
	NANOGP8				RTQMR	
	No predictions made					
2	NKX2-1				RIQYK	
		Vnd	326	5	RIQYK	19
2	NKX2-2				RIQYK	

		Vnd	363	5	RIQYK	19
3	NKX2-3				RIQYK	
		Vnd	293	5	RIQYK	19
		Bap	276	5	RIQYK	23
		Tin	255	5	RIQYK	16
2	NKX2-4				RIQYK	
		Vnd	326	5	RIQYK	19
3	NKX2-5				RIQYK	
		Vnd	289	5	RIQYK	19
		Bap	268	5	RIQYK	23
		Tin	250	5	RIQYK	16
3	NKX2-6				RIQYK	
		Vnd	283	5	RIQYK	19
		Bap	257	5	RIQYK	23
		Tin	252	5	RIQYK	16
2	NKX2-8				RIQYK	
		Vnd	337	5	RIQYK	19
2	NKX3-1				RIQYK	
		Bap	306	5	RIQYK	23
2	NKX3-2				RIQYK	
		Bap	333	5	RIQYK	23
2	NKX6-1				RVQTK	
		Hgtx	368	5	RVQTK	20
2	NKX6-2				RVQTK	
		Hgtx	371	5	RVQTK	20
4	NOBOX				RVQAK	
		Al	235	5	RVQAK	20
		CG4136	228	5	RVQAK	22
		PhdP	222	5	RVQAK	17
2	ONECUT1				RNMRR	
		onecut	316	5	RNMRR	15
2	ONECUT2				RNMRR	
		onecut	297	5	RNMRR	15
2	OTP				RVQAK	
		Otp	372	5	RVQAK	20
2	OTX1				RVKAK	

		Oc	360	5	RVKAK	19
2	OTX2				RVKAK	
		Oc	364	5	RVKAK	19
2	PBX1				RNGIR	
		Exd	350	5	RNGIR	17
2	PBX2				RNGIR	
		Exd	351	5	RNGIR	17
2	PBX3				RNGIR	
		Exd	346	5	RNGIR	17
2	PBX4				RNGIR	
		Exd	329	5	RNGIR	17
3	PDX1				RIQMK	
		Scr	279	5	RIQMK	25
		Dfd	271	5	RIQMK	24
		Antp	271	5	RIQMK	16
2	PITX2				RVKAK	
		Ptx1	374	5	RVKAK	20
2	PITX3				RVKAK	
		Ptx1	374	5	RVKAK	20
2	PKNOX1				RNIRR	
		Hth	285	5	RNIRR	17
2	PKNOX2				RNIRR	
		Hth	301	5	RNIRR	17
3	PROP1				RVQAK	
		CG32532	281	5	RVQAK	23
		Al	271	5	RVQAK	20
		Pph13	251	5	RVQAK	21
1	PRRX1				RVQAK	
		CG9876	329	5	RVQAK	20
		Pph13	301	5	RVQAK	21
1	PRRX2				RVQAK	
		CG9876	335	5	RVQAK	20
		Pph13	302	5	RVQAK	21
	Predicted				RNIHK	
	No predictions made					

2	RAX				RVQAK	
		Rx	388	5	RVQAK	27
2	RAXL1				RVQAK	
		Rx	374	5	RVQAK	27
	RHOXF1				RVKAR	
	No predictions made					
	RHOXF2				VIEAK	
	No predictions made					
	SATB1				RKQYY	
	No predictions made					
	SATB2				RKQYH	
	No predictions made					
3	SHOX				RVQAK	
		PhdP	280	5	RVQAK	17
		CG11294	276	5	RVQAK	15
		Otp	275	5	RVQAK	20
3	SHOX2				RVQAK	
		PhdP	280	5	RVQAK	17
		CG11294	276	5	RVQAK	15
		Otp	275	5	RVQAK	20
2	SIX1				SNKQR	
		So	361	5	SNKQR	27
2	SIX2				SNKQR	
		So	365	5	SNKQR	27
2	SIX3				TNKQR	
		Optix	372	5	TNKQR	27
2	SIX4				VNKQR	
		Six4	306	5	VNKQR	20
2	SIX5				VNKQR	
		Six4	307	5	VNKQR	20
2	SIX6				TNKQR	
		Optix	367	5	TNKQR	27
1	TGIF1				RNIRR	
		Vis	299	5	RNIRR	22
		Achi	298	5	RNIRR	23
1	TGIF2				RNIRR	

		Vis	311	5	RNIRR	22
		Achi	310	5	RNIRR	23
4	TGIF2LX				KNIRR	
		Vis	239	4	RNIRR	22
		Achi	238	4	RNIRR	23
		Hth	211	4	RNIRR	17
4	TGIF2LY				KNIRR	
		Vis	234	4	RNIRR	22
		Achi	233	4	RNIRR	23
		Hth	206	4	RNIRR	17
2	TLX1				RTQTK	
		C15	345	5	RTQTK	19
2	TLX2				RTQTK	
		C15	328	5	RTQTK	19
2	TLX3				RTQTK	
		C15	347	5	RTQTK	19
	VAX1				RSNFK	
	No predictions made					
2	VAX2				RVQTK	
		Ems	247	5	RVQTK	20
		E5	235	5	RVQTK	43
	VENTX				RTQMK	
	No predictions made					
1	VSX1				RVQAK	
		CG33980	328	5	RVQAK	13
		CG4136	296	5	RVQAK	22
		Rx	291	5	RVQAK	27
1	VSX2				RVQAK	
		CG33980	341	5	RVQAK	13
		CG4136	306	5	RVQAK	22
		Rx	303	5	RVQAK	27

Table S7. Prediction of TRANSFAC homeodomains.

Predicted	Actual	ALLR	Dist	E value	P value	Predicted Cons.	Actual Cons.	Combined Cons.
T02970 mouse Chx10	M00437.cons.L8.c2.topMx1	9.2938	4.1572	4.44E-008	4.44E-008	TAATTa	GCTAATTA	TAATTA
T04139 human Chx10	M00437.cons.L8.c2.topMx1	9.2938	4.1572	4.44E-008	4.44E-008	TAATTa	GCTAATTA	TAATTA
T04142 chick Chx10	M00437.cons.L8.c2.topMx1	9.2938	4.1572	4.44E-008	4.44E-008	TAATTa	GCTAATTA	TAATTA
T08863 mouse S8	M00099.cons.L9.c2.topMx1	8.8487	1.2748	1.27E-007	1.27E-007	TAATTa	TAATTRRnt	TAATTR
T03978 human Cart-1	M00416.cons.L14.c2.topMx1	7.836	7.8714	1.64E-006	1.64E-006	TAATTA	SnTAATtRnATTAn	TAATTA
T03981 clawed frog Cart-1	M00416.cons.L14.c2.topMx1	7.836	7.8714	1.64E-006	1.64E-006	TAATTA	SnTAATtRnATTAn	TAATTA
T03980 rat Cart-1	M00416.cons.L14.c2.topMx1	7.836	7.8714	1.64E-006	1.64E-006	TAATTA	SnTAATtRnATTAn	TAATTA
T08295 mouse Nkx2-2	M00485.cons.L9.c2.topMx1	7.2991	5.2662	3.25E-006	3.25E-006	CAAGTR	TAAGTRnTT	YAAGTR
T04337 human Nkx2-2	M00485.cons.L9.c2.topMx1	7.2991	5.2662	3.25E-006	3.25E-006	CAAGTR	TAAGTRnTT	YAAGTR
T04272 chick Nkx2-2	M00485.cons.L9.c2.topMx1	7.2991	5.2662	3.25E-006	3.25E-006	CAAGTR	TAAGTRnTT	YAAGTR
T04265 golden Syrian hamster Nkx2-2	M00485.cons.L9.c2.topMx1	7.2991	5.2662	3.25E-006	3.25E-006	CAAGTR	TAAGTRnTT	YAAGTR
T03489 cattle Crx	M00623.cons.L15.c2.topMx1	6.4759	5.537	3.04E-005	3.04E-005	TAATCC	YnnnTAAtCnnMnnn	TAATC
T03458 rat Crx	M00623.cons.L15.c2.topMx1	6.4759	5.537	3.04E-005	3.04E-005	TAATCC	YnnnTAAtCnnMnnn	TAATC
T03461 mouse Crx	M00623.cons.L15.c2.topMx1	6.4759	5.537	3.04E-005	3.04E-005	TAATCC	YnnnTAAtCnnMnnn	TAATC
T02792 human Crx	M00623.cons.L15.c2.topMx1	6.4759	5.537	3.04E-005	3.04E-005	TAATCC	YnnnTAAtCnnMnnn	TAATC
T00857 human Nkx2-1	M00432.cons.L8.c2.topMx1	6.264	4.6036	2.52E-005	2.52E-005	CAAGTR	ASTCAAGT	CAAGT
T00856 rat Nkx2-1	M00432.cons.L8.c2.topMx1	6.264	4.6036	2.52E-005	2.52E-005	CAAGTR	ASTCAAGT	CAAGT
T02098 dog Nkx2-1	M00432.cons.L8.c2.topMx1	6.264	4.6036	2.52E-005	2.52E-005	CAAGTR	ASTCAAGT	CAAGT
T00859 mouse Nkx2-1	M00432.cons.L8.c2.topMx1	6.264	4.6036	2.52E-005	2.52E-005	CAAGTR	ASTCAAGT	CAAGT
T00856 rat Nkx2-1	M00794.cons.L9.c2.topMx1	5.911	5.2132	5.95E-005	5.94E-005	CAAGTR	cTcAAGnGY	cAAGtg

T04271 chick Nkx2-1	M00794.cons.L9.c2.topMx1	5.911	5.2132	5.95E-005	5.94E-005	CAAGTR	cTcAAGnGY	cAAGtg
T00859 mouse Nkx2-1	M00794.cons.L9.c2.topMx1	5.911	5.2132	5.95E-005	5.94E-005	CAAGTR	cTcAAGnGY	cAAGtg
T00857 human Nkx2-1	M00794.cons.L9.c2.topMx1	5.911	5.2132	5.95E-005	5.94E-005	CAAGTR	cTcAAGnGY	cAAGtg
T02098 dog Nkx2-1	M00794.cons.L9.c2.topMx1	5.911	5.2132	5.95E-005	5.94E-005	CAAGTR	cTcAAGnGY	cAAGtg
T05157 human RX	M00623.cons.L15.c2.topMx1	5.8299	7.2643	0.0001174	0.0001174	TAATTR	YnnnTAAtCnnMnnn	TAAT
T00863 fruit fly Ubx	M00018.cons.L4.c2.topMx1	5.7005	3.4202	4.11E-005	4.11E-005	TAATKA	TAAT	TAAT
T03848 red flour beetle Ubx	M00018.cons.L4.c2.topMx1	5.7005	3.4202	4.11E-005	4.11E-005	TAATKA	TAAT	TAAT
T01481 human Pbx1a	M00124.cons.L9.c2.topMx1	5.1275	11.4862	0.0003066	0.0003065	TGAcaa	TTGATTGAT	TGA
T01481 human Pbx1a	M00096.cons.L9.c2.topMx1	5.0303	11.8509	0.0003757	0.0003757	TGAcaa	AAGCTTGAT	TGA
T01481 human Pbx1a	M00998.cons.L8.c2.topMx1	4.0228	8.6473	0.002752	0.002749	TGAcaa	tGATTGAT	TGA
T01992 fruit fly Abd-A	M01083.cons.L10.c2.topMx1	2.4675	12.5216	0.08927	0.0854	TAATKA	AARTaAWWWW	TAAT
T04367 human NCX	M00484.cons.L10.c2.topMx1	1.9307	7.2197	0.2746	0.2401	TAAttR	nngtAAntng	TAA
T04368 mouse Ncx	M00484.cons.L10.c2.topMx1	1.9307	7.2197	0.2746	0.2401	TAAttR	nngtAAntng	TAA

Table S8. Examples of functional (*in vivo*) homeodomain binding sites consistent with their monomeric specificity

HD	Target gene	References	Notes	enhancer element name*
Cad	*ftz*	Dearolf, C.R., Topol, J., and Parker, C.S. (1989). *Nature* 341, 340-343.	Activation of the ftz zebra stripe element is facilitated by cad binding sites where direct mutation of these sites abrogates activity.	ftz zebra stripe
Tin	*mef2*	Gajewski, K., Kim, Y., Lee, Y.M., Olson, E.N., and Schulz, R.A. (1997). *EMBO J* 16, 515-522.	A pair of tinman sites is required for enhancer function in myocyte precursors as demonstrated by loss of *lacZ* expression when the sites are mutated in report assay.	Mef2_IIA237
Tin	*beta3Tub60D*	Kremser, T., Gajewski, K., Schulz, R.A., and Renkawitz-Pohl, R. (1999). *Dev biol* 216, 327-339	Identified 3 tinman sites, 2 are required for reporter transcription in dorsal vessel cells. Mutation of tin sites in enhancer disrupts expression.	betaTub60D_ b3-lac333
Bap	*beta3Tub60D*	Zaffran, S., and Frasch, M. (2002). *Mech. Dev.* 114, 85-93.	Identifies single pair of overlapping bap sites that are responsible for tissue specific gene expression	betaTub60D_ beta3-14/vm1
Tin	*Sur*	Akasaka, T., Klinedinst, S., Ocorr, K., Bustamante, E.L., Kim, S.K., and Bodmer, R. (2006). *Proc Natl Acad Sci U S A* 103, 11999-12004. & Hendren, J.D., Shah, A.P., Arguelles, A.M., and Cripps, R.M. (2007). *Mech. Dev.* 124, 416-426.	In this gene a tin-responsive enhancer was discovered using bioinformatics approaches looking for tin binding sites based on the consensus sequence.	En3
Tin	*svp*	Ryan, K.M., Hendren, J.D., Helander, L.A., and Cripps, R.M. (2007). *Dev biol* 302, 694-702.	In this gene a tin-responsive enhancer was discovered using bioinformatics approaches looking for tin binding sites based on the consensus sequence where the identified sites were conserved over multiple genomes.	SCE
Tin	*eve*	Knirr, S., and Frasch, M. (2001). *Dev biol* 238, 13-26.	Mutation of 2 tin sites identified based on its consensus recognition element inactivates enhancer	EME b3'
Bcd & Cad	*kni*	Rivera-Pomar, R., Lu, X., Perrimon, N., Taubert, H., and Jackle, H. (1995). *Nature* 376, 253-256.	Demonstrates that sets of binding sites for cad and bcd are sufficient for anterior patterned expression in absence of other factor binding sites. Composition of sites not critical, just number and quality.	kni_64
Ap	*Ser*	Yan, S.J., Gu, Y., Li, W.X., and Fleming, R.J. (2004). *Development* 131, 285-298.	14 Ap sites identified by DNaseI footprinting when mutated abrogate activitiy.	Ser_minimal_ wing_enhancer
Bcd	*eve*	Arnosti, D.N., Barolo, S., Levine, M., and Small, S. (1996). *Development* 122, 205-214.	Mutating individual bcd sites reduces activity. Adding novel sites to new positions restores activity demonstrating that the bcd site position is not critical.	eve_stripe2
Ubx	*sal*	Galant, R., Walsh, C.M., and Carroll, S.B. (2002). *Development* 129, 3115-3126.	Ubx represses *sal* expression in the haltere. The development of the haltere is not dependent on Exd or Hth, so the interaction of Ubx with the sal 328 element is thought to be independent of these TFs (although potentially dependent on other unknown TFs). Mutation of individual Ubx sites results in a loss of repression of the reporter gene in the haltere	sal 328
Dfd	*rpr*	Lohmann, I., McGinnis, N., Bodmer, M., and McGinnis, W. (2002). *Cell* 110, 457-466.	Dfd regulates *rpr* in maxillary segment boundary. Loss of 4 Dfd sites severely decreases reporter expression. Gel shift analysis suggests that Exd does not bind cooperatively with Dfd on this element.	rpr_4S3

HD= homeodomain; *Enhancer names were taken directly from the literature or extracted from REDfly

Table S9. Similarity of predicted and determined recognition motifs for six human homeodomains

Factor Comparison	ALLR score	ALLR Distance	E value	P value	predicted consensus	actual consensus	combined consensus
BarHL1	7.8848	1.5896	8.48E-007	8.48E-007	TAAttG	nTAAtTGn	TAAtTG
Nkx3-2	9.1699	0.6191	5.04E-008	5.04E-008	TAAGTg	YTAAGTG	TAAGTG
PitX2	10.1543	0.114	5.50E-009	5.50E-009	TAATCC	TAATCC	TAATCC
Six3	7.6043	0.962	1.27E-006	1.27E-006	TGATA	nnSTGATA	TGATA
TGIF2	8.5538	1.3843	1.83E-007	1.83E-007	TGACAg	ntTGACA	TGACA
Vsx1	9.5817	0.2637	2.13E-008	2.13E-008	TAATTa	tTAATTA	TAATTa

Supplemental Discussion

A more detailed discussion of the specificity determinants that can influence each binding site position.

BS Position 1: 89% of the aligned recognition sequences have Thy at this position. Consistent with this preference, the majority of homeodomains (94%) have Arg5 in the N-terminal arm, which specifies Thy (Ades and Sauer, 1995). In the Six family, Thy is preferred in the absence of Arg5, which is potentially due to the favorable stacking of the T-methyl group over Arg55 as it interacts with Gua at position 2 of the binding site (Figure 3B). In the NK-2 class, Val6 and Leu7 promote a preference for Cyt over Thy (Damante et al., 1996).

BS Position 2: Preferences for Ade, Gua or Thy are observed among the different homeodomains. 83% of the aligned sequences have Ade at this position. Consistent with this preference, most typical homeodomains contain Arg2 or Arg3 within the N-terminal arm, which structural and biochemical studies have implicated in biasing specify toward Ade (Ades and Sauer, 1995; Hovde et al., 2001). However, both Lab and HGTX, which lack either arginine, strongly prefer Ade, which is consistent with reports of an additional unidentified specificity determinant (Ades and Sauer, 1995). Members of the Abd-B group display a preference for Thy over Ade, but the responsible specificity determinants are unclear. In atypical homeodomains Arg55 specifies Gua; however, several typical homeodomains contain Arg55, yet display a preference for Ade (e.g. Awh and Bcd), indicating that additional residues contribute to specification of Gua.

BS Position 3: Asn51 specifies Ade at this position.

BS Position 4: Any base can be specified at this position. Thy is the most common base (80%) and is strongly correlated with the presence of Ile or Val at position 47. The three exceptions that contain Val at position 47 - Bap, Tin and Vnd – are NK-2 class homeodomains, which prefer Gua due to a direct contact by Tyr54 to the complementary strand Cyt (Gruschus et al., 1997). Cyt specificity is strongly correlated with the presence of Arg54, which interacts with the complementary Gua. Bcd is one notable exception; it contains both Ile47 and Arg54, yet displays strong preference for Thy, which suggests that the base preference of Ile47 may supercede Arg54 (explored in the Bcd mutagenesis). Decreased specificity at position 4 is correlated with either Thr47 or Asn47, although the presence of a β-branched amino acid at position 43 modifies the degree of preference (Supplementary Figure 7B). Finally, Cut displays a unique preference for Ade at position 4; His50, which only occurs in this homeodomain, may be responsible.

BS Position 5: For many specificity groups correlations exist between the residues at positions 47, 50 and 54 and certain base preferences. Specificity groups that have Gln50 and either Ile47 or Val47 display preferences that correlate with the residue at position 54. When Met54 is present (Antp group) a tolerance is observed for either Gua or Thy, whereas when Ala54 (En group) or Tyr54 (NK-2 group) are present, a preference for Thy is observed. When Lys50 occurs with Ile47 or Val47 and Ala54 (Bcd group) a strong preference for Cyt is observed (Ades and Sauer, 1994; Hanes and Brent, 1989; Percival-Smith et al., 1990; Treisman et al., 1989). However, Lys50 in the presence of Asn47 and Gln54 (Six family) specifies Ade instead of Cyt. An influence of position 54 on the specificity mediated by Lys50 has also been observed in other contexts (Pellizzari et al., 1997). Finally, when Arg54 is present in the absence of Ile47 or Val47

(Iroquois and TGIF groups) a preference for Ade is observed, which is likely due to preferential stacking of the complementary T-methyl group over the guanidinium group of Arg54 as it interacts with Gua at position 4 (Figure 3B). Typical homeodomains containing Thr47 display relaxed specificity at position 5.

BS Position 6: Like position 5, the determinants of specificity at this position appear influenced by the residues at positions 47, 50 and 54. Ade is modestly preferred over Gua when Gln50 and Met54 (Antp group) or Ala54 (En group) are present. Conversely, Gua is preferred over Ade when both Thr47 and Thr54 are present with Gln50 (Bar and NK-1 groups, Supplementary Figure 7C) or less exclusively when Tyr54 is present with Ile47 and Gln50 (NK-2 group). Finally, the presence of Lys50 provides a preference for Cyt at position 6 (Bcd and Six groups). The majority of atypical homeodomains with the exception of the Six group display no strong preference at this position.

Supplemental Experimental Procedures

Identification and boundary definition of the independent homeodomains.
These 84 homeodomains represent all of the fly homeodomains in the SMART database that are not associated with another major type of DNA-binding domain: 18 unique homeodomains are associated with PAX, POU or ZnF-C2H2 domains or an additional homeodomain based on the SMART annotation (Letunic et al., 2006; Schultz et al., 1998). There are homeodomains in our set that are associated with another DNA-binding domain, such as those in the Cut family, which were retained because of interesting sequence composition. The sequences of the homeodomains used in the B1H selection and the raw selected binding sites are found in Supplementary Table 1. The region of the homeodomain that was fused to omega was defined by the length of the core homeodomain identified by SMART: 60 to 63 amino acids depending on the presence of a TALE insertion. 10 additional amino acids in the protein sequence were included beyond the C-terminus of the homeodomain (if present). The amino acids were removed from the terminus if a hydrophobic residue occurred (YFIVLWPM), since terminal hydrophobic residues can induce protein degredation in E coli (Parsell et al., 1990).

Omega-Zif12 fusions for the selection of homeodomain specificity
Homeodomains were expressed as omega fusions in combination with fingers one and two of Zif268 (Zif12) under control of the *lacUV5mut* promoter plasmid (pB1H2ω2-12HD; Supplementary Figure 1). Each homeodomain (with two additional N-terminal residues) was cloned between the KpnI site and the Xba1 site downstream with a stop codon introduced just prior to the Xba1 site. Two additional amino acids were added after the Kpn1 site prior to the start of the homeodomain. The first amino acid was always glycine. In the majority of the homeodomains, the second amino acid was the -1 amino acid of the specific homeodomain being assayed, however for a subset we used the -1 residue of Oct1, which is arginine, for purely historical reasons. The KpnI site and the inserted residues created a 5 amino acid linker between the 2nd His of Zif268 finger 2 and the beginning of the HD (Zif12-TGTGN-HD; Supplementary Figure 1). All expression constructs were sequence verified.

Construction of the ZF10 Library
The ZF10 library was created by annealing oligonucleotides with complimentary ends to the library oligonucleotide to make duplex DNA with a gap spanning the randomized region and appropriate overhangs for cloning into the pH3U3 plasmid between the NotI and EcoRI sites upstream of the promoter controlling the HIS3 and URA3 genes.
ZF10 library oligonucleotide:
5'-GGCCGCCATGGATCCNNNNNNNNNNTGGGCGGCTGATAGGCGCGCCG-3', 5 prime complimentary oligonucleotide:
5'-GGATCCATGGC-3'
3 prime complimentary oligonucleotide:
5'-AATTCGGCGCGCCTATCAGCCGCCCA-3'
The oligonucleotides were 5' phosphorylated as a mixture by combining 200 pmol of each oligonucleotide in 100µl of 1xT4 polynucleotide kinase buffer (NEB) with 1mM ATP and 20 Units T4 polynucleotide kinase (NEB). This reaction was incubated at 37°C for 30 minutes and then boiled for 5 minutes before annealing by a gradual reduction in temperature to 4°C. 1 µl of the phosporylated, annealed

oligonucleotides (2 μM) was ligated into 1 μg of gel purified pH3U3 plasmid backbone that had been digested with NotI and EcoRI in a 30 μl reaction containing 1xT4 DNA Ligase buffer and 1 μl of T4 DNA Ligase (400 units, NEB) overnight at 16°C. Following completion, the ligation reaction was ethanol precipitated, and the DNA pellet was resuspended in 2 μl of H_20, and transformed into 80 μl of electrocompetent XLI-Blue cells (Stratagene). The transformed cells were recovered in 50 ml SOC for 1 hour at 37°C. Following the recovery, a 200 μl sample was titrated by 10-fold serial dilution on 2xYT plates containing 25 μg/ml Kanamycin to determine the total number of transformants. The number of transformants in a dilution normalized to the fraction of library culture should reflect the constructed library size. Kanamycin (25 μg/ml) was then added to the remaining culture and the cells were expanded for an additional hour at 37°C. After expansion, the cells were plated on 10 large 2xYT plates (150mm round) containing 25 μg/ml kanamycin and grown overnight at 37 °C. 200μl of the culture was again titrated by 10-fold dilutions on 2xYT plates containing kanamycin to determine the degree of expansion that occurred during the additional hour of growth. After these large plates had grown overnight, cells were harvested from these plates by resuspending the colonies in 10 ml 2xYT per plate. The resuspensions were pooled and cells pelleted by centrifugation for 15 minutes at 3000 rpm. The plasmid DNA was recovered from this pellet by scaling 20-fold the procedure for DNA isolation using the QIAGEN plasmid Miniprep Kit.

Counterselection of the ZF10 library

Counterselections were performed on the ZF10 library to remove self-activating sequences. 1 μg of raw ZF10 library material was transformed into 80 μl of the *rpoZ* positive version of the selection strain (US0 *ΔhisB, ΔpyrF*). These cells were recovered in SOC for 1 hour at 37°C while rotating. The cells were then pelleted by centrifugation for 15 minutes at 3000 rpm and the resulting pellet was resuspended in 5 ml of YM medium (a type of minimal medium). The cells were acclimated to the YM medium for 1 hour at 37°C while rotating. Following recovery the cells were pelleted, washed 2 times with YM medium (by pelleting and resuspension) and then resuspended in a final volume of 1 ml YM medium. 20 μl of this final resuspension was titrated by 10-fold serial dilution on rich media plates (2xYT + 25 μg/ml Kanamycin) to determine the total transformants. The titration plates were grown overnight at 37°C while the remaining 980 μl cell resuspension was stored at 4 °C. The following day cells were counted from the rich media titrations and approximately $5x10^6$ transformants were plated on ten 150mm round YM plates containing 2.5mM 5-FOA. The plates were wrapped with parafilm and incubated at 37°C for 24 to 36 hours. The cells were then harvested from the plates and the plasmid DNA recovered as described above for the raw library.

General selection procedure

Approximately 2 μg of the bait plasmid (pB1H2ω2-12HD) and 50 ng of the library plasmid were electroporated into 80 μl of the selection strain (US0 Δ*hisB*, Δ*pyrF*, Δ*rpoZ*; *Noyes and Wolfe, unpublished results*). The cells and the two plasmids were mixed on ice and moved to a pre-chilled 1 mm electroporation cuvette. The cell and plasmid suspension was electroporated at 4°C and immediately resuspended in 10 ml pre-warmed SOC. The cells were then recovered while rotating at 37 °C for 1 hour. Next, the cells were pelleted by centrifugation at 3000 rpm for 15 minutes and resuspended in 5 ml NM medium that was supplemented with 200 mM uracil, and 0.1% histidine (Meng and Wolfe, 2006). These cells were acclimated to the NM medium while rotating for 1 hour at 37°C. Cells were pelleted, washed 4 times in NM medium (no supplement) by sequential pelleting and resuspension and then resuspended to a final volume of 1 ml NM medium. 20 ml of this final resuspension was titrated by 10-fold serial dilutions on rich media plates (2xYT + 25 μg/ml Kanamycin, and 100 μg/ml Carbenicillin) to determine the total number of transformants. The titration plates were grown overnight at 37 °C while the remaining 980 μl cell culture was stored at 4 °C. The following day cell counts were determined from the rich media titrations and based on the total number of transformants, between $1x10^7$ and $1x10^8$ cells were placed on one 5 mM 3-AT and one 10mM 3-AT NM selective plate (150mm diameter rounds). Cells were spread on the plates with sterile glass beads and allowed to air dry under a flame. The plates were then wrapped individually with parafilm and grown at 37°C for 36 to 48 hours. Typically five to fifteen binding site selections were performed in parallel with positive and negative controls included to allow a qualitative assessment of the success of each experiment. Surviving colonies on each plate were counted and the fraction of surviving clones was determined based on the number of cells that were plated. A 10-fold increase in the fraction of surviving clones compared to the negative control (omega without a tethered TF)

was highly correlated with a successful selection. Selections with ratios lower than a 10-fold increase were successful in many instances, but in this range there was more variability in the number of sequenced clones that contained binding sites.

Alternate selection conditions

The vast majority of the selections were successful in the initial attempt, however a handful did not yield an obvious enrichment in the number of selected clones as defined by a low fold increase (or no increase) in the number of surviving clones on selective media relative to the background when normalized to the number of cells plated. In most cases this was resolved by expressing the omega-Zif12-homeodomain at higher levels using a stronger promoter (*lacUV5*). In the case of one homeodomain, Eve, where our initial selections failed, we found that the removal of a small string of hydrophobic residues from the C-terminus of the protein just after the end of the homeodomain resulted in a significant improvement in activity. Hydrophobic residues at the C-terminus of a protein can lead to lower levels of functional expression in bacteria (Parsell et al., 1990).

Colony PCR and Sequencing

The binding sites from successful selections were recovered by PCR amplification of the corresponding pH3U3 Library window from individual surviving colonies picked from each selection plate. These PCR products were sequenced to generate the desired data for computational analysis. PCR reactions were done in a 96-well plate format where 25 μl of the PCR mix (1mM HU100 primer, 1mM OK181 primer, 300mM Denville dNTP mix, 1x NEB ThermoPol Buffer and 1 unit of NEB Taq polymerase) was added to each well of the plate. For each of the 96 wells, a single colony was picked from a selection plate with an autoclaved toothpick to inoculate the 25 μl PCR mix. For each set of PCRs from a given selection plate a negative control reaction (no inoculation) was run in parallel in one well to insure that the appearance of an amplified product was not due a contaminating DNA source. Once each well had been inoculated, the plate was covered with aluminum film and placed in the thermocycler. The PCR reaction initiated with a single, 2 minute denaturation step at 94^{o}C. This was followed by 35 reaction cycles consisting of one minute denaturation at 94^{o}C, a 1.5 minute annealing step at 56^{o}C, and a 2 minute extension at 68^{o}C. After these 35 cycles were complete, there was a final extension for 5 minutes at 68^{o}C. The plate was held at 4^{o}C from that point until being removed from the block. To confirm successful PCR reactions, 5 μl of each 25 μl PCR reaction was run out on a 1.5% agarose gel and the mobility of product in each well was compared to a DNA ladder standard (NEB). Successful plates were sequencing (Agencourt) using HU100 as the sequencing primer.

HU100
5'-GAAATATGTATCCGCTCATGAC-3'
OK181
5'-CCAGAGCATGTATCATATGGTCCAGAAACCC-3'

Construction of the master alignment of sites:

The master alignment contains 1860 binding sites for 83 of the 84 Drosophila homeodomain proteins as well as Oct1 (Lag1 was excluded because it lacks Asn51, which makes the alignment of its sites to all others within the dataset problematic). CONSENSUS selected substrings from 1868 of the 2211 input sequences and took the reverse complement of 657 of these substrings, wherein alignments were generated for each factor independently(Hertz and Stormo, 1999). In every case, the length of the motif was selected by varying the motif length parameter (-L) and selecting the alignment with the smallest e-value. These 84 separate alignments formed the basis for the construction of the master alignment. The orientation of the alignments for individual factors produced by CONSENSUS was somewhat arbitrary; consequently, we manually reversed the orientation of 35 sets of sequences (about 588 sites). As described previously, the high information content Ade (recognized by Asn51) was used as an anchor to help align the sets of sites (notice that all but 4 sites contain Ade at position 6 in the master alignment below). Information about the probable orientation of each individual site gleaned from the observed site biases (described above) led us to manually 'flip' the orientation of some individual sites (47), overriding some orientation decisions made by CONSENSUS. For the master alignment of all 84 sets of sites we used the entire sequence of each aligned site, not just the 1868 substrings returned by CONSENUS (8 problematic sites were removed; Supplementary Table 3). At this point, the alignment contained 15 columns as the registers of the aligned subsites in each sequence varied, so the 5' and 3' flanking columns 1, 2, 13, 14, and 15 were removed to

generate a master alignment with 10 columns because from 57 to 99 percent of these columns were comprised of gaps as the library sequence elements are only 10 bp in length.

Count matrix for the entire master alignment:

	1	2	3	4	5	6	7	8	9	10
A	297	435	269	78	1511	1856	48	326	983	260
C	310	396	419	66	1	0	219	143	109	254
G	242	446	161	19	213	2	84	328	512	345
T	358	542	1006	1697	135	2	1509	1063	60	119
-	653	41	5	0	0	0	0	0	196	882

All Sequence logos (Schneider and Stephens, 1990) for these factors were generated using WebLogo (Crooks et al., 2004). We note that the number of selected binding sites that comprise a particular logo are modest (typically 20 to 40) and consequently, the significance of bases that are absent or occur infrequently in a motif cannot be fully assessed.

Clustering of binding site motifs

The master alignment of sites was used to determine the pairwise global alignments between every set of homeodomain binding sites. The aligned sites for each homeodomain protein were converted to count matrices. Pairwise distances between all matrices were calculated based on average log likelihood ratio (ALLR) similarity scores (Wang and Stormo, 2003). When calculating the ALLR scores, gaps were treated as missing data and ignored. The formula for the ALLR score was modified slightly: instead of using the natural logarithm function (log base e), log base 2 was used. The Neighbor program from the Phylip phylogenetic analysis package (Felsenstein, 2005) was used to cluster the motifs using the neighbor joining method. The input to Neighbor was a pairwise distance table based on the master alignment of sites. The radial logarithmic neighbor joining tree of the motifs in Figure 2 was produced using the TreeIllustrator program (Trooskens et al., 2005). The branch lengths displayed in this image are logarithmically proportional to the actual branch lengths calculated by the Neighbor program. The phylogram of the homeodomain amino acid sequences in Figure 2 was produced using TreeIllustrator with the pairwise distances determined by ClustalW (Thompson et al., 1994).

MI analysis

MI analysis was performed on the dataset using the Master alignment of binding sites as previously described (Gutell et al., 1992). The MI plot was transformed into a joint rank product matrix by transforming each element in the MI matrix by calculating the rank of each element's MI value in that column (the column-wise rank) and the rank of each element's MI value in that row (the row-wise rank). The column-wise rank and row-wise rank for each element were multiplied to yield the joint rank product matrix. The product matrix was transformed to generate a heat plot using the following formula:

$$\frac{\max(Ln(X)) - Ln(X_{ij})}{\max(Ln(X))}$$

where X_{ij} is the joint rank product matrix element ij and max(X) is the maximum value in X (600).

G-test significance analysis

The significance of an apparent difference between motifs for two groups of homeodomains was estimated using a G-test (Sokal and Rohlf, 1995). Aligned binding sites for each group of factors were pooled and one position (column) in the DNA binding motif was analyzed by generating a 2 by 4 contingency table, where rows contained the 2 classes and columns 4 DNA bases. Small pseudo counts (0.01) were added to each value and the G-test statistic was calculated allowing 3 degrees of freedom for each base, unless a base was not observed in both of the two classes, in which case 1 degree of freedom was subtracted.

Competition Gel Shift Assay

Oligonucleotides.

The Oligonucleotides used for this assay were designed to have a single, central homeodomain binding site that represents the consensus recognition sequence of one of 7 core specificity groups (Engrailed, Bar,

Abd-B, Bicoid, NK-2, Six, and TGIF) as well as one outlier (CG11617). Once annealed, the resulting duplex oligonucleotides contain a 5' GG overhang at each end that can be used to radiolabel the DNA. These sequences of the oligonucleotides are listed below with the recognition sequence in **bold**. Where multiple binding sites were examined for a single specificity group, the differences within these sequences are underlined:

Engrailed Top
GGGCAGGCAG***TAATTA***GGACGTCG
Engrailed Bottom
GGCGACGTCC***TAATTA***CTGCCTGC
Bar Top
GGGCAGGCAG***TAATTG***GGACGTCG
Bar Bottom
GGCGACGTCC***CAATTA***CTGCCTGC
Abd-B-A Top
GGGCAGGCAG***TTATTA***GGACGTCG
Abd-B-A Bottom
GGCGACGTCC***TAATAA***CTGCCTGC
Abd-B-G Top
GGGCAGGCAG***TTATTG***GGACGTCG
Abd-B-G Bottom
GGCGACGTCC***CAATAA***CTGCCTGC
Bicoid Top
GGGCAGGCAG***TAATCC***GGACGTCG
Bicoid Bottom
GGCGACGTCC***GGATTA***CTGCCTGC
NK-2 Top
GGGCAGGCAG***CAAGTG***GGACGTCG
NK-2 Bottom
GGCGACGTCC***CACTTG***CTGCCTGC
CG11617-A Top
GGGCAGGCAG***TTAACA***GGACGTCG
CG11617-A Bottom
GGCGACGTCC***TGTTAA***CTGCCTGC
CG11617-C Top
GGGCAGGCAG***TTCACA***GGACGTCG
CG11617-C Bottom
GGCGACGTCC***TGTGAA***CTGCCTGC
CG11617-T Top
GGGCAGGCAG***TTTACA***GGACGTCG
CG11617-T Bottom
GGCGACGTCC***TGTAAA***CTGCCTGC
Six Top
GGGCAGGCAG***TGATA***CGGACGTCG
Six Bottom
GGCGACGTCCG***TATCA***CTGCCTGC
TGIF Top
GGGCAGGCAG***TTGACA***GGACGTCG
TGIF Bottom
GGCGACGTCC***TGTCAA***CTGCCTGC

Expression and Purification of Proteins.
Each homeodomain was expressed as a C-terminal fusion to maltose binding protein (MBP) from pJH196 (a generous gift from Keith Joung, (Hurt et al., 2003)) using an *in vitro* transcription-translation system (Expressway™ Cell-Free E. coli Expression System, Invitrogen). The zinc fingers utilized for the binding site selection in the B1H system were not incorporated into these constructs. Each MBP-HD construct was

expressed in two 100ml reactions from approximately 2mg of plasmid DNA per reaction. These reactions were incubated while rotating for 6.5 hours at 37°C. The reactions for each construct were combined together with one of the binding buffers listed below (900 ml final volume) and the MBP-HD proteins were captured on 100ml of Amylose Resin (New England BioLabs) by incubation at 4 °C for 1.5 hours while rotating. The resin-bound MBP-HD proteins were washed 4 times with 1ml binding buffer. Finally, the protein was eluted from the resin by incubation with 50ml of binding buffer supplemented with 40mM maltose at room temperature, while rotating for 30 minutes. Aliquots of protein were stored at -80°C.

Two different binding buffers were used in these experiments. The majority of the gel shift assays utilized a binding buffer consisting of 10mM Tris-HCl pH 7.5, 0.1mM EDTA, 25mM NaCl, 1mM DTT and 5% glycerol. The more complex binding buffer for Ptx and Caudal consisted of 15mM HEPES pH 7.8, 50mM KCl, 50mM KGlutamate , 50mM KOAc, 5mM $MgCl_2$, 1mM DTT and 5% glycerol. All binding buffers were supplemented with 0.1mg/ml BSA and 0.1% IGEPAL CA-630 for the gel-shift assays.

Gel Shift Competitions.

To perform the competition gel shift assays, oligonucleotides for the consensus binding site corresponding to each homeodomain were annealed and then endlabeled. 30 ml labeling reactions were done using 200ng of annealed oligonucleotide, 40mCi a-^{32}P dCTP, 5 units Klenow (exo-), and a final concentration of 3.33mM dNTPs minus dCTP. These reactions were incubated at 37°C for 30 minutes and then chased with 3.33mM dNTPs (including dCTP) for an additional 30 minutes at 37°C. The labeled oligonucleotides were recovered from free radionucleotides using a G-25 spin column (BioRAD).

The optimal amount of protein for each homeodomain to be used in the DNA-binding reaction for the competition assay was determined by performing shift assays with a titration of protein on its labeled consensus site. Titrations of both protein and DNA were performed to ensure that binding reactions were under K_d conditions with [labeled DNA] << K_d and that the amount of HD-DNA complex formation was not saturated (data not shown). The appropriate concentration of cold consensus binding site needed to effectively compete the majority of the HD-DNA* complex at the optimal protein concentration was determined by titration of competitor (data not shown).

The competition assays were then performed by equilibrating the homeodomain with 80pM of its labeled target site and one of the cold competitor duplexes in a 20ml reaction for at least 2 hours at room temperature. Each protein was challenged in a separate reaction with each of the eight of the specificity group binding site oligonucleotides and one control reaction without competitor. The final concentrations of cold competitor DNA used for each homeodomain is as follows:

CG11617 used 5.625nM

En, Vis, and Tin used 9nM

Lbe, Optix, CG34031, Ptx, and Cad used 90nM

10ml of each reaction was then run on a pre-run (35 minutes at 300V) 7.5% native polyacrylamide gel (0.5xTBE) for 35 minutes at 300V. The gels were dried and then exposed phosphoimaging plates for 8-12 hours. These plates were scanned with a FUJIFILM FLA-5000 and the percentage of protein-DNA complex in each reaction was determined by quantifying the free DNA and bound DNA bands with FUJIFILM's program, Image Gauge 4.22.

Supplemental References

Ades, S.E., and Sauer, R.T. (1994). Differential DNA-binding specificity of the engrailed homeodomain: the role of residue 50. Biochemistry *33*, 9187-9194.

Ades, S.E., and Sauer, R.T. (1995). Specificity of minor-groove and major-groove interactions in a homeodomain-DNA complex. Biochemistry *34*, 14601-14608.

Banerjee-Basu, S., and Baxevanis, A.D. (2001). Molecular evolution of the homeodomain family of transcription factors. Nucl Acids Res *29*, 3258-3269.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome research *14*, 1188-1190.

Damante, G., Pellizzari, L., Esposito, G., Fogolari, F., Viglino, P., Fabbro, D., Tell, G., Formisano, S., and Di Lauro, R. (1996). A molecular code dictates sequence-specific DNA recognition by homeodomains. Embo J *15*, 4992-5000.

Ekker, S.C., Jackson, D.G., von Kessler, D.P., Sun, B.I., Young, K.E., and Beachy, P.A. (1994). The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins. EMBO J *13*, 3551-3560.
Ekker, S.C., von Kessler, D.P., and Beachy, P.A. (1992). Differential DNA sequence recognition is a determinant of specificity in homeotic gene action. EMBO J *11*, 4059-4072.
Felsenstein, J. (2005). PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author Department of Genome Sciences, University of Washington, Seattle.
Fraenkel, E., and Pabo, C.O. (1998). Comparison of X-ray and NMR structures for the Antennapedia homeodomain-DNA complex. Nat Struct Biol *5*, 692-697.
Fraenkel, E., Rould, M.A., Chambers, K.A., and Pabo, C.O. (1998). Engrailed homeodomain-DNA complex at 2.2 A resolution: a detailed view of the interface and comparison with other engrailed structures. Journal of molecular biology *284*, 351-361.
Grant, R.A., Rould, M.A., Klemm, J.D., and Pabo, C.O. (2000). Exploring the role of glutamine 50 in the homeodomain-DNA interface: crystal structure of engrailed (Gln50 --> ala) complex at 2.0 A. Biochemistry *39*, 8187-8192.
Gruschus, J.M., Tsao, D.H., Wang, L.H., Nirenberg, M., and Ferretti, J.A. (1997). Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy: basis of binding specificity. Biochemistry *36*, 5372-5380.
Gutell, R.R., Power, A., Hertz, G.Z., Putz, E.J., and Stormo, G.D. (1992). Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. Nucl Acids Res *20*, 5785-5795.
Hanes, S.D., and Brent, R. (1989). DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. Cell *57*, 1275-1283.
Harvey, R.P. (1996). NK-2 homeobox genes and heart development. Developmental biology *178*, 203-216.
Hertz, G.Z., and Stormo, G.D. (1999). Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics (Oxford, England) *15*, 563-577.
Hovde, S., Abate-Shen, C., and Geiger, J.H. (2001). Crystal structure of the Msx-1 homeodomain/DNA complex. Biochemistry *40*, 12013-12021.
Hurt, J.A., Thibodeau, S.A., Hirsh, A.S., Pabo, C.O., and Joung, J.K. (2003). Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. Proc Natl Acad Sci U S A *100*, 12271-12276.
Joshi, R., Passner, J.M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M.A., Jacob, V., Aggarwal, A.K., Honig, B., and Mann, R.S. (2007). Functional specificity of a Hox protein mediated by the recognition of minor groove structure. Cell *131*, 530-543.
Kim, D.W., Kempf, H., Chen, R.E., and Lassar, A.B. (2003). Characterization of Nkx3.2 DNA binding specificity and its requirement for somitic chondrogenesis. J Biol Chem *278*, 27532-27539.
Letunic, I., Copley, R.R., Pils, B., Pinkert, S., Schultz, J., and Bork, P. (2006). SMART 5: domains in the context of genomes and networks. Nucl Acids Res *34*, D257-260.
Liu, X., Brutlag, D.L., and Liu, J.S. (2001). BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. Pac Symp Biocomput *6*, 127-138.

Meng, X., and Wolfe, S.A. (2006). Identifying DNA sequences recognized by a transcription factor using a bacterial one-hybrid system. Nat protocols *1*, 30-45.
Mukherjee, K., and Burglin, T.R. (2007). Comprehensive Analysis of Animal TALE Homeobox Genes: New Conserved Motifs and Cases of Accelerated Evolution. J Mol Evol *65*, 137-153.
Parsell, D.A., Silber, K.R., and Sauer, R.T. (1990). Carboxy-terminal determinants of intracellular protein degradation. Genes Dev *4*, 277-286.
Pellizzari, L., Tell, G., Fabbro, D., Pucillo, C., and Damante, G. (1997). Functional interference between contacting amino acids of homeodomains. FEBS Lett *407*, 320-324.
Percival-Smith, A., Müller, M., Affolter, M., and Gehring, W.J. (1990). The interaction with DNA of wild-type and mutant fushi tarazu homeodomains. EMBO J *9*, 3967-3974.
Piper, D.E., Batchelor, A.H., Chang, C.P., Cleary, M.L., and Wolberger, C. (1999). Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. Cell *96*, 587-597.
Pomerantz, J.L., Sharp, P.A., and Pabo, C.O. (1995). Structure-based design of transcription factors. Science *267*, 93-96.
Schneider, T.D., and Stephens, R.M. (1990). Sequence logos: a new way to display consensus sequences. Nucl Acids Res *18*, 6097-6100.
Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. Proceedings of the National Academy of Sciences of the United States of America *95*, 5857-5864.
Sokal, R.R., and Rohlf, F.J. (1995). Biometry: the principles and practice of statistics in biological research., 3rd edition edn (New York, W. H. Freeman and Co.).
Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl Acids Res *22*, 4673-4680.
Treisman, J., Gönczy, P., Vashishtha, M., Harris, E., and Desplan, C. (1989). A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell *59*, 553-562.
Trooskens, G., De Beule, D., Decouttere, F., and Van Criekinge, W. (2005). Phylogenetic trees: visualizing, customizing and detecting incongruence. Bioinformatics (Oxford, England) *21*, 3801-3802.
Verrijzer, C.P., Alkema, M.J., van Weperen, W.W., Van Leeuwen, H.C., Strating, M.J., and van der Vliet, P.C. (1992). The DNA binding specificity of the bipartite POU domain and its subdomains. EMBO J *11*, 4993-5003.
Wang, T., and Stormo, G.D. (2003). Combining phylogenetic data with co-regulated genes to identify regulatory motifs. Bioinformatics (Oxford, England) *19*, 2369-2380.
Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D., and Pabo, C.O. (1991). Crystal structure of a MATalpha2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell *67*, 517-536.