Supplementary Figure 1

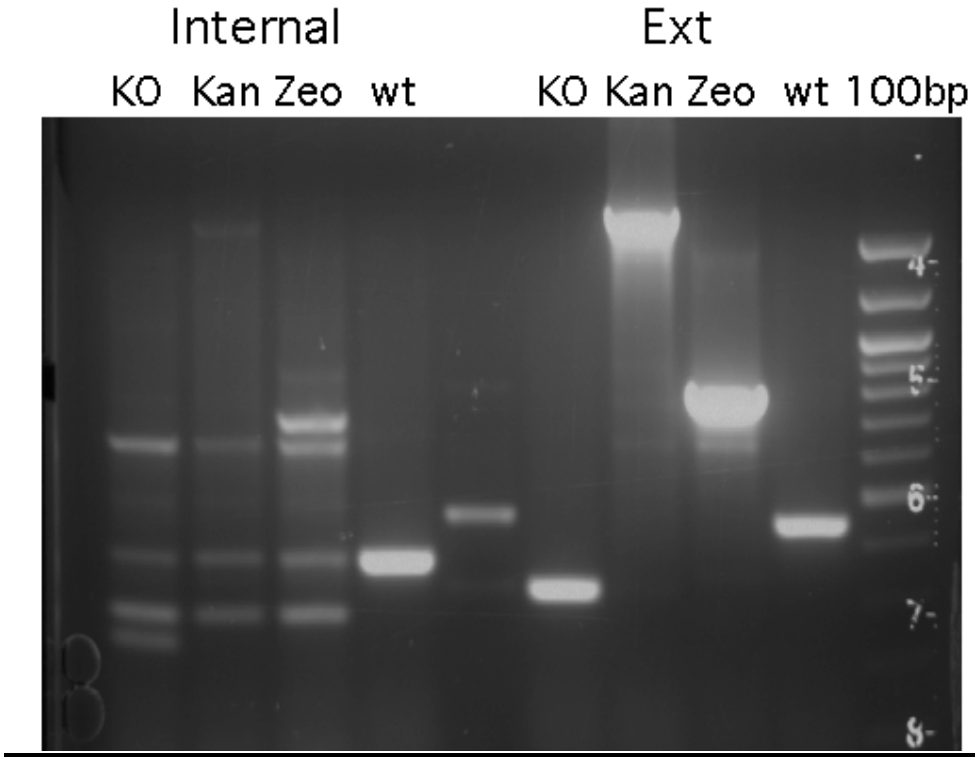
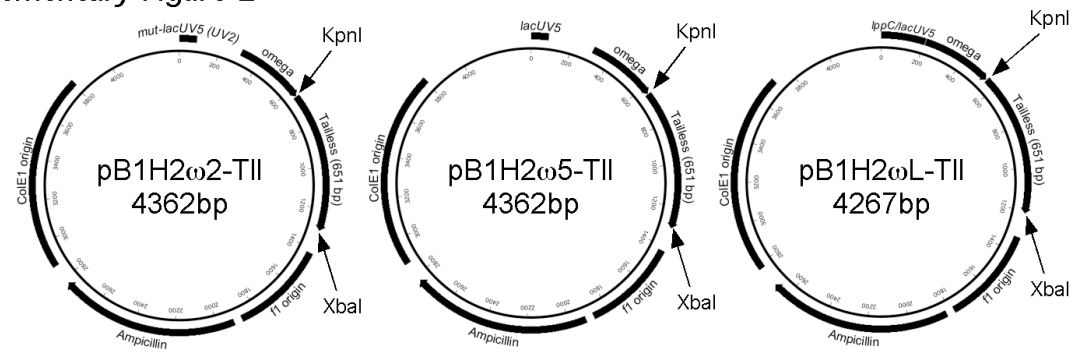


Figure 1._PCR products of rpoZ locus. The PCR products generated from genomic DNA of our *rpoZ* cell types run on a 1.5% agarose gel. Above each lane the insertion at the *rpoZ* locus of the particular cell type is listed. Wt is the wild type *rpoZ* gene, Kan is the kanamycin cassette, Zeo is the zeocin cassette, and KO has the antibiotic cassette removed resulting in a complete removal of the *rpoZ* locus. On the left side of the gel are the PCR products from the *rpoZ 5'* primer and an internal omega primer (omega int 3') that should only amplify the wt gene. On the right side of the gel are PCR products using the two external primers described in the text, *rpoZ 5'* and *rpoZ 3'*. These give an indication of the relative size of the insertion/deletion at the *rpoZ* locus.

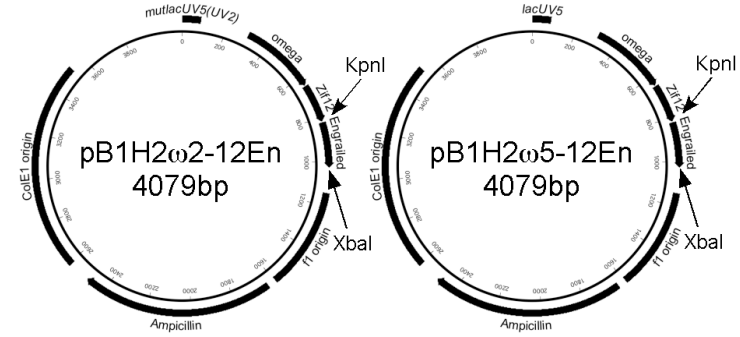
Supplementary Figure 2



Linker between omega subunit and TF (69bp)

gcggccgcggactacaaggatgacgacgacaagttccggaccggttccaagacacccccccat**ggtacc**-*TF(TII)*-TAA**tctaga**

NotI KpnI stop XbaI



Linker between Zif12 and HD (15bp)

HisThrGlyThrGlyAsp

Zif 1,2 -cacacc**ggtacc**ggtgac-*HD(En)*-TAA**tctaga**

KpnI stop XbaI

Figure 2. Maps of the bait plasmids used in the omega-B1H system. The maps of the bait plasmids used to characterize the TF reported here are displayed with key features annotated. The plasmids allow a TF to be characterized at 3 different promoter strengths as direct fusions to omega or as fusions mediated by Zif268 fingers 1 and 2 (Zif12) by simply subcloning between the unique Kpn1 and Xba1 sites in each plasmid. The linkers between the omega subunit or Zif12 and the TF are indicated below the plasmid sets.

Supplementary Figure 3

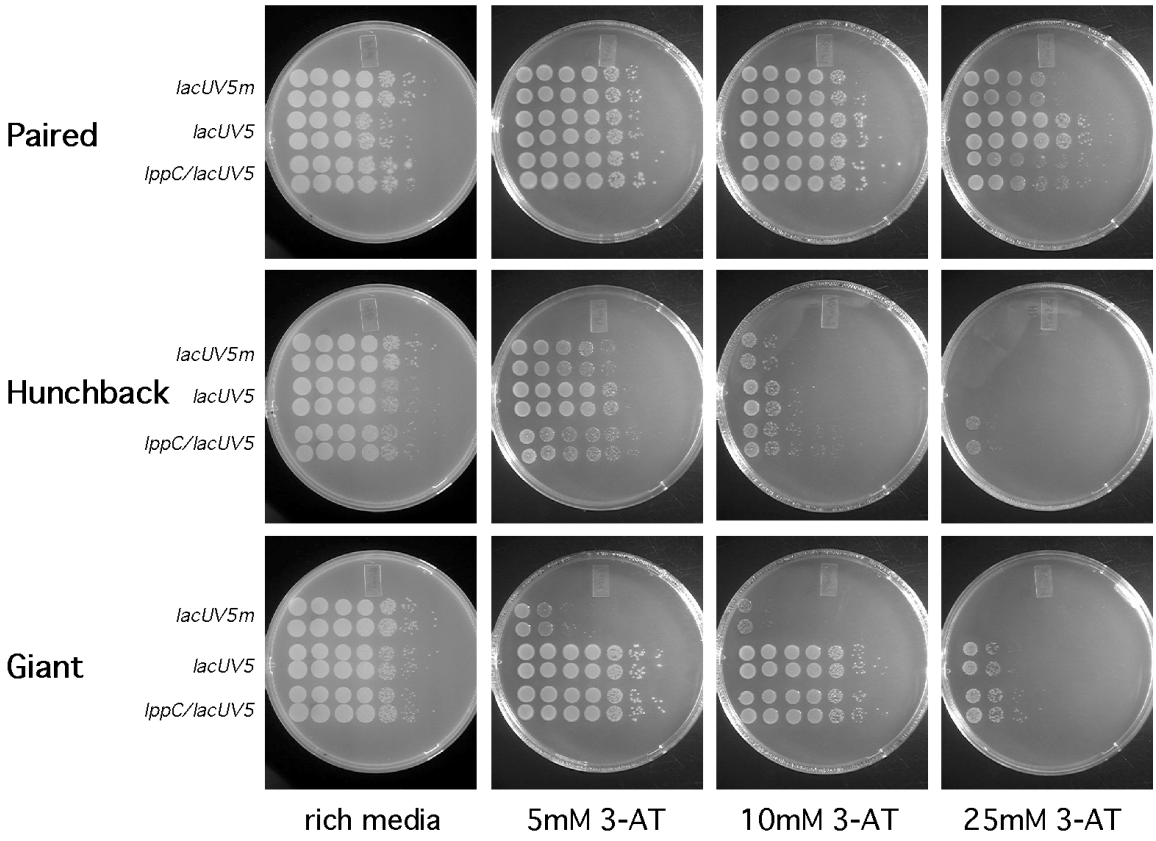


Figure 3. Comparison of the activity of three different factors at three different promoter strengths. The activity of three different TFs (Paired, Hunchback and Giant) that represent three different families of DNA-binding domains (a Paired motif, a Cys_2His_2 zinc finger protein, and basic leucine zipper motif, respectively) were characterized on consensus binding sites for each factor at three different promoter strengths. The optimal window of activity varied depending on the factor. Paired displays good activity when expressed from all three promoters (nearly 100% at 10 mM 3-AT), with the *lacUV5* promoter slightly superior to the other two. Hunchback displayed weaker activity than Paired at all three promoter strengths, with the strongest activity again with the *lacUV5* promoter. Giant displayed almost no activity when expressed under the *lacUV5m* promoter, which was optimal for Zif268, but strong activity when expressed from the other two promoters, with the dual promoter perhaps slightly superior. The reporter pH3U3 plasmids used in each case had the consensus binding site for each factor positioned at the most highly represented position relative to the -35 box based on B1H binding site selections. The activity of the omega-TF hybrids with these reporters was examined at 3 different concentrations of 3-AT where the rich media plate serves as a control for the number of cells plated in each set. Each combination of expression plasmid and reporter plasmid was plated in duplicate, where each spot represents a 10-fold serial dilution of cells from left to right.

Supplementary Figure 4

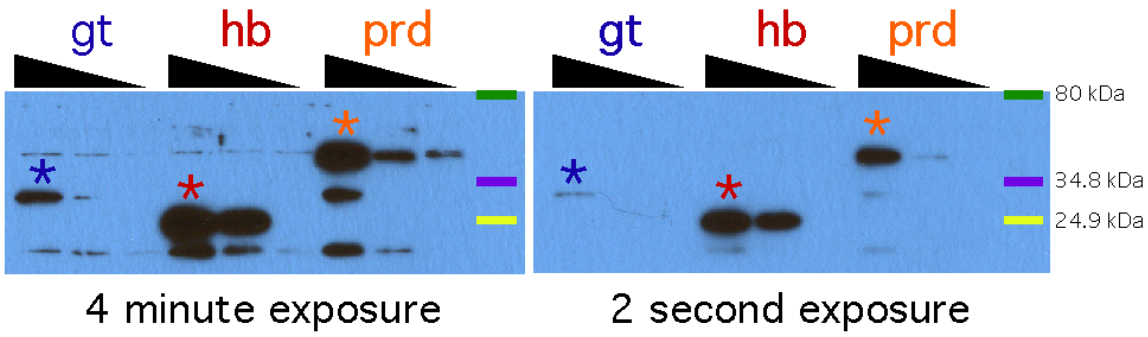


Figure 4. Western blots of 3 ω-TF's at 3 different promoter strengths. Three factors, Prd, Hb, and Gt, each at 3 different promoter strengths were grown under inducing conditions (10 μM IPTG). For each lane on the gel a normalized amount of each bacterial culture was created based on its OD_{600} and these cells were solubilized in SDS-loading buffer. A identical fraction of this sample was run on an SDS gel and the flag tag in each construct was visualized by Western blot. Each factor is shown from left to right with decreasing promoter strength (*lppC*, *lacUV5*, and *lacUV5m*). The approximate molecular weights of the omega-TF constructs are: Prd = 41.58kDa, Hb = 25.74kDa, Gt = 29.7kDa.

Supplementary Figure 5

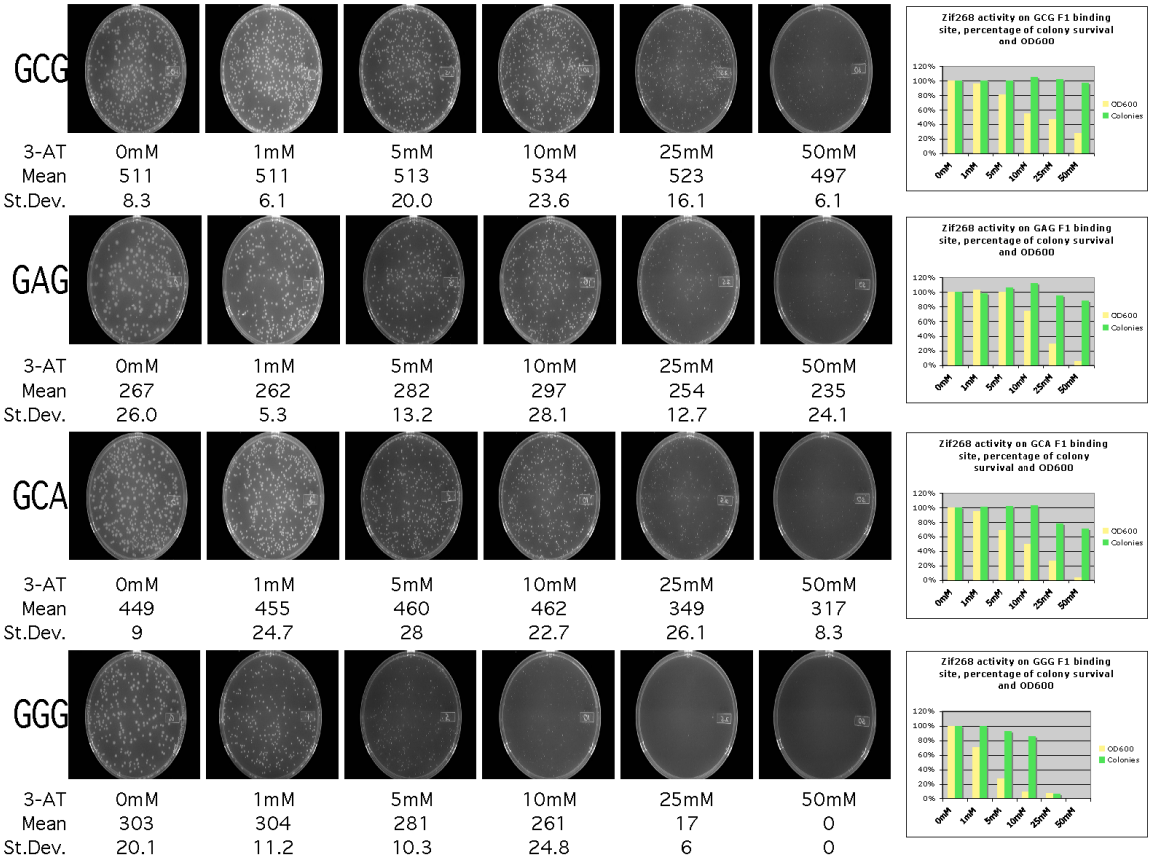


Figure 5. Comparison of ω-Zif268 activity with 4 different binding sites. Cells harboring the pB1H2ω2-Zif268 plasmid and one of 4 different pH3U3 plasmids were challenged at several 3-AT concentrations. The pH3U3 plasmids contained one of 4 Zif268 binding sites (GCGTGG**GCG**) placed 10 bp upstream of the -35 box with modifications in the finger 1 binding sequence. The finger 1 sequence is listed to the left of each row and their previously determined relative K_ds are GAG=4.9-fold over GCG, GCA=4.7-fold over GCG, and GGG=20.3-fold over GCG. Cells were plated in triplicate and a representative plate is displayed here with the mean and standard deviation colony count is listed below. To the right is a histogram displaying the percentage of colonies that survive at each 3-AT concentration in comparison to non-selective conditions (0mM 3-AT) in green. In yellow is the percentage OD_{600} that was measured from the cells pooled at each stringency.

Supplementary Figure 6

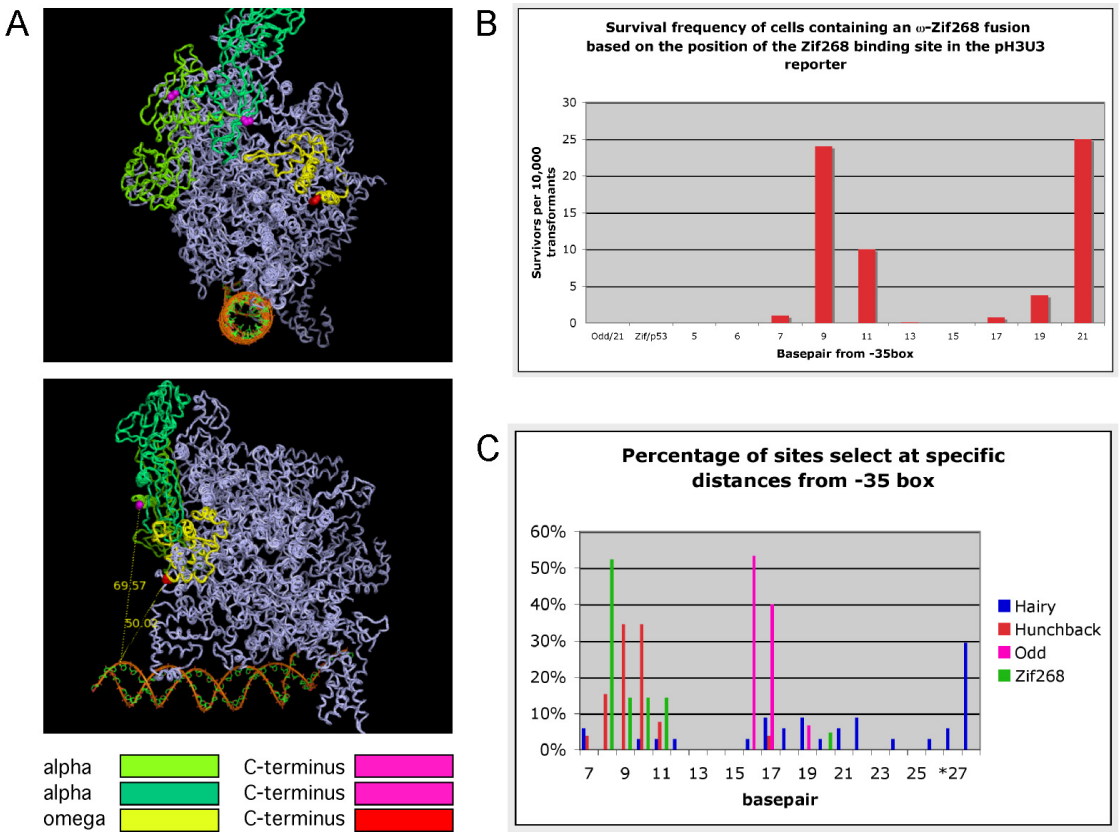


Figure 6. The omega subunit and the positions utilized by omega-TF hybrids. A. The structure of RNA Polymerase (RNAP) bound to DNA (1). Top. An upstream view of the RNAP shows the relative positions of the C-termini of both alpha subunits (green and avocado) and the omega subunit (yellow). The position of the omega subunit is rotated about the DNA by about 30 degress relative to the nearest alpha subunit. Bottom. Measurements from the nearest alpha subunit and omega subunit C-termini to the DNA roughly one turn upstream of the -35 box reveals that the omega subunit is approximately 20Å closer to the DNA. B. The survival frequency of cells under stringent conditions harboring the pB1H2ω2-Zif268 plasmid and a reporter pH3U3 plasmid with the Zif268 consensus binding site placed at various distances from the -35 box reveals two clear peaks of activity at approximately 10 and 21 bp upstream. Ostermeier and colleagues observed a peak of activity for a site positioned 14 bp away from -35 box of the alpha-B1H system, which could be the result of differences in the positions of the polymerase subunits relative to the DNA (2) C. A comparison of the selected binding site positions of four factors characterized with the omega-B1H system. Some factors have a clear preference for a specific distance from the -35 box (Zif268, Hb, and Odd) and these distances differ from one factor to another (Hb sequences are centered 9 or 10 bp from the -35 box while Odd sequences are centered almost exclusively 16 or 17 bp away). There are other factors such as Hairy that utilize almost the entire library window. (*sequences that came through 27 and 28 basepair from the -35 box are utilizing a fixed region of the pH3U3 plasmid and could be a reason for the peak here. These sequences were not used for the construction of the Hairy logo).

Supplementary Figure 7

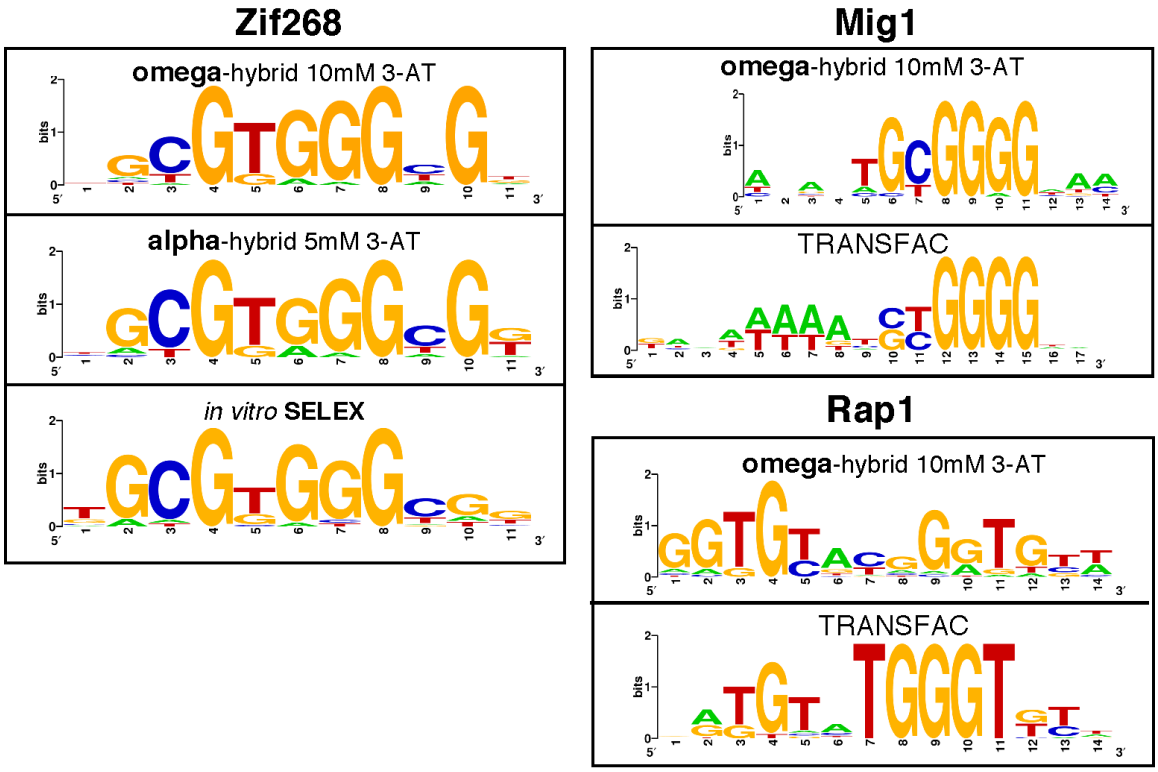
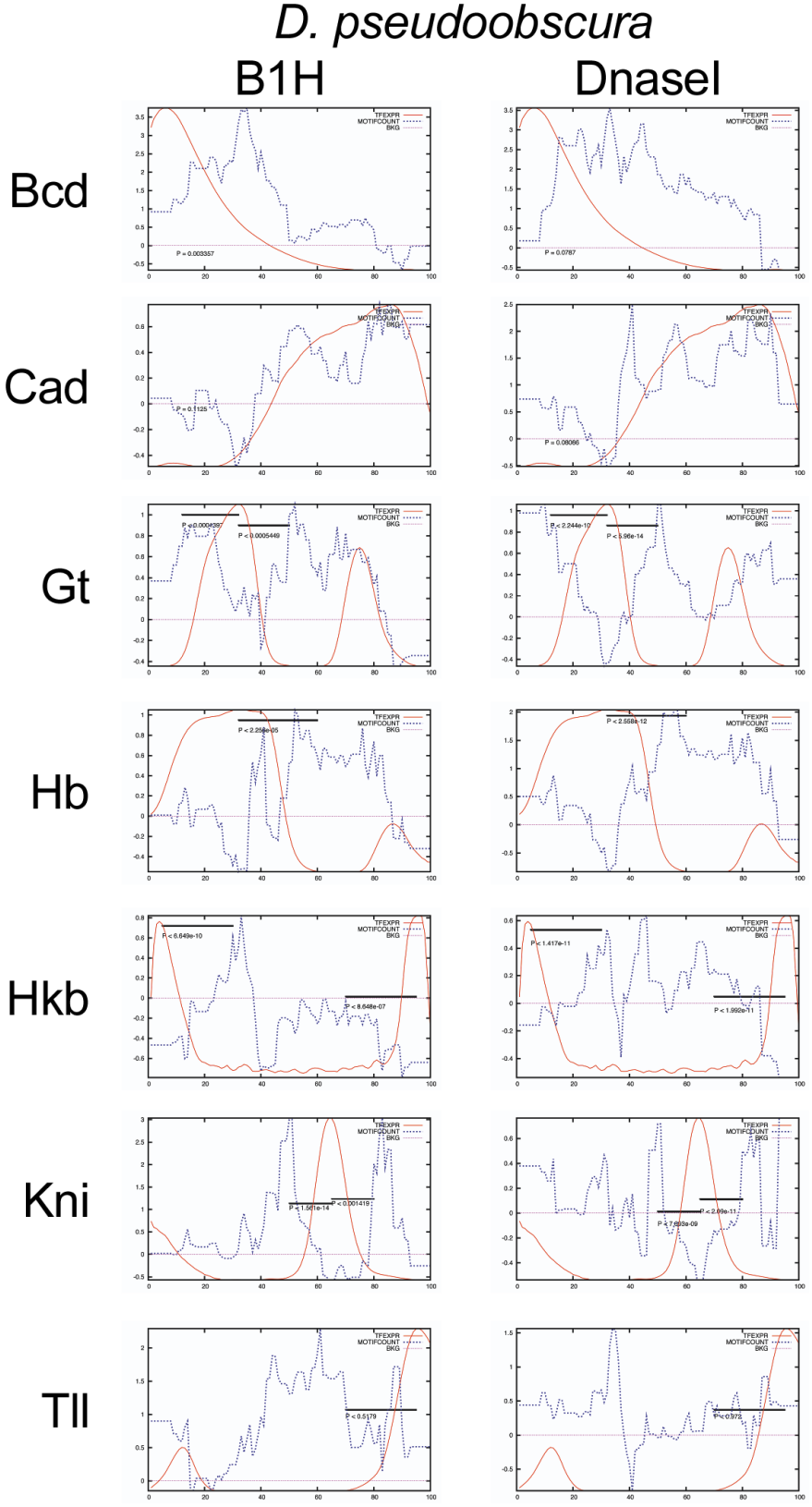


Figure 7. Comparison of Zif268, Mig1 and Rap1 logos. The DNA-binding domains of all three factors were expressed (*lacUV5m*) as C-terminal fusions to the omega subunit. Complementary sequences within the 28 bp randomized library to each TF were isolated under selective conditions (10 mM 3-AT). Each factor yielded a significant increase (>10-fold) in the number of surviving colonies over background when compared to a negative control (omega without a fusion partner). Approximately 20 reporters were sequenced from each binding site selection and binding sites were determined by identifying overrepresented motifs within these sequences using MEME (3). The logos identified for each factor as an omega-TF fusion are compared to logos generated for these factors by other methods. The TRANSFAC logos were generated from the available PWMs at TRANSFAC (4). The Zif268 logo is compared to the logos from characterization as an alpha-Zif268 hybrid (5) and from *in vitro* SELEX data (6).

Supplementary Figure 8.



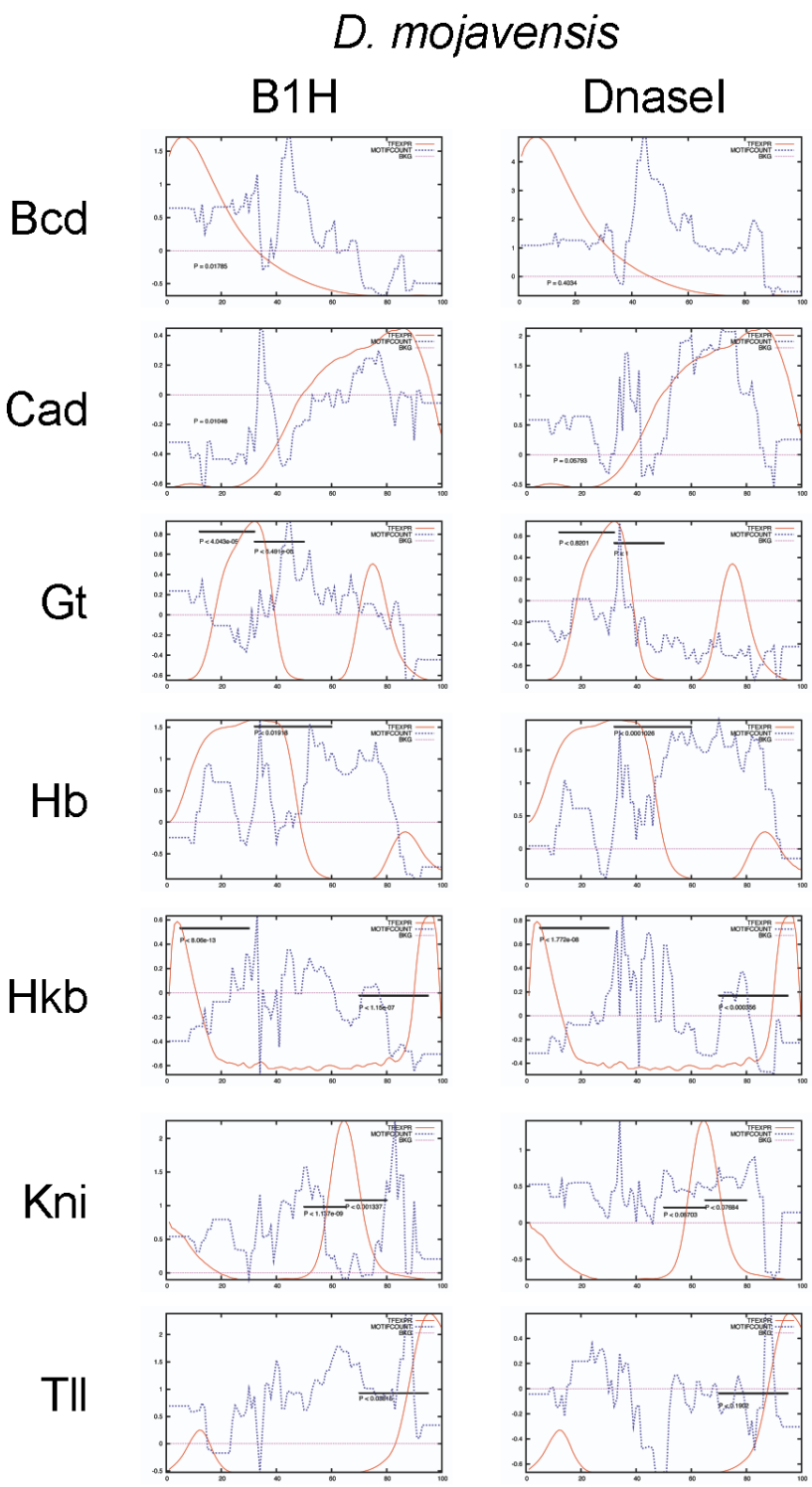
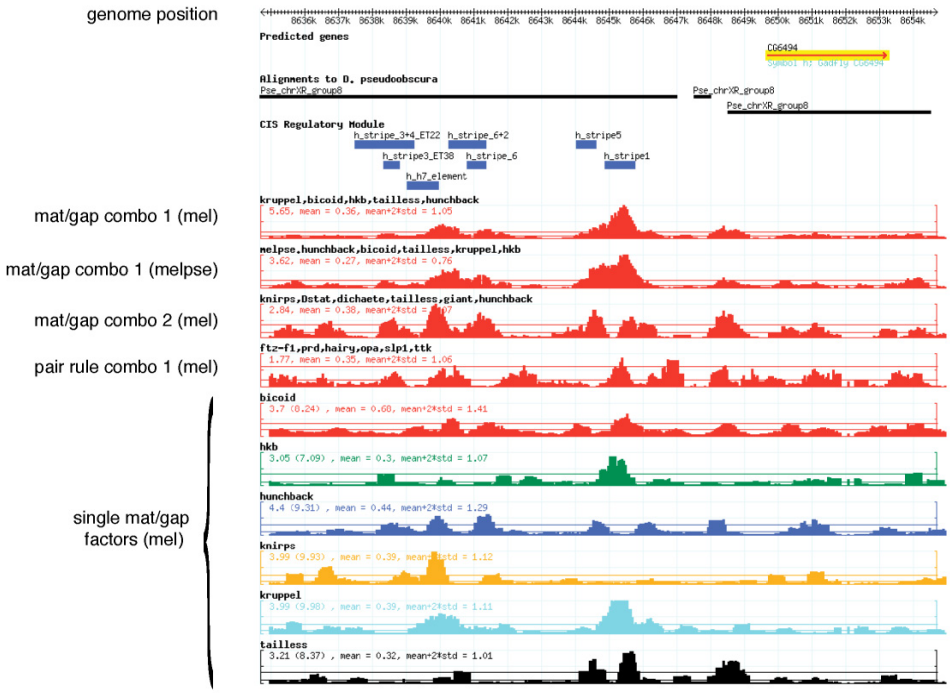


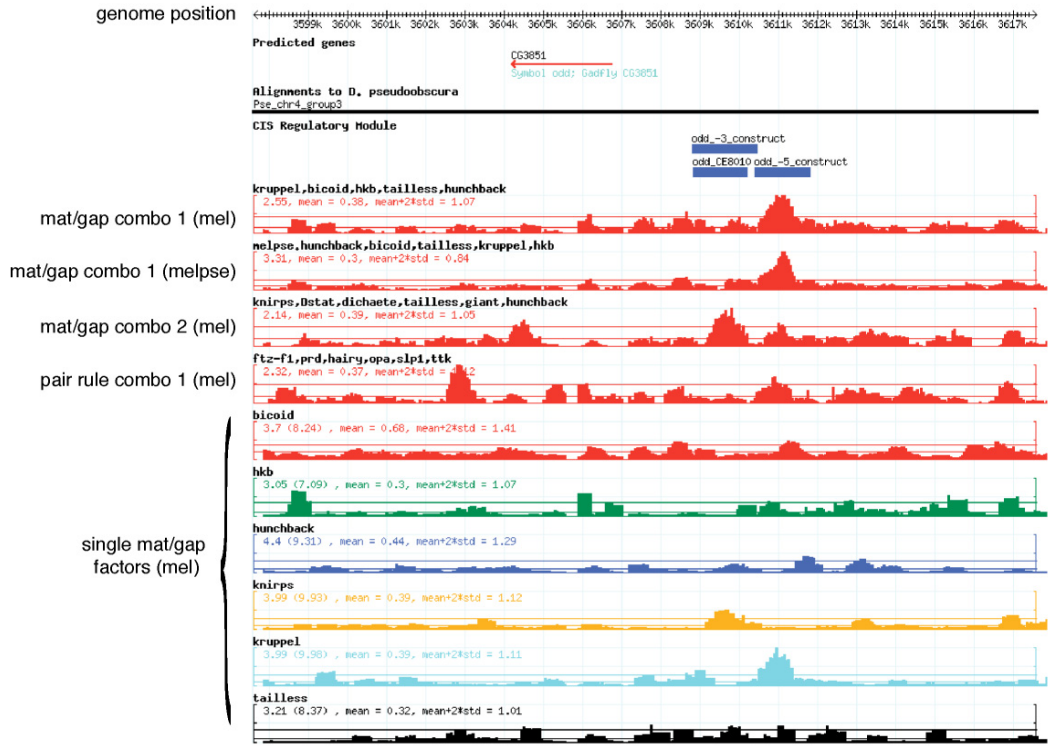
Figure 8. Motifcount plots for the DNase and B1H PMWs over syntenic regions for the 48 CRMs within the *D. pseudoobscura* and *D. mojavensis* genomes. Overall these plot look similar to the *D. melanogaster* plots in Figure 5, although some of the feature are more poorly defined. The p-values are significant for a number of factors and in most cases superior for the B1H PWMs. Otherwise, these plots have the same features as described in Figure 5.

Supplementary Figure 9.

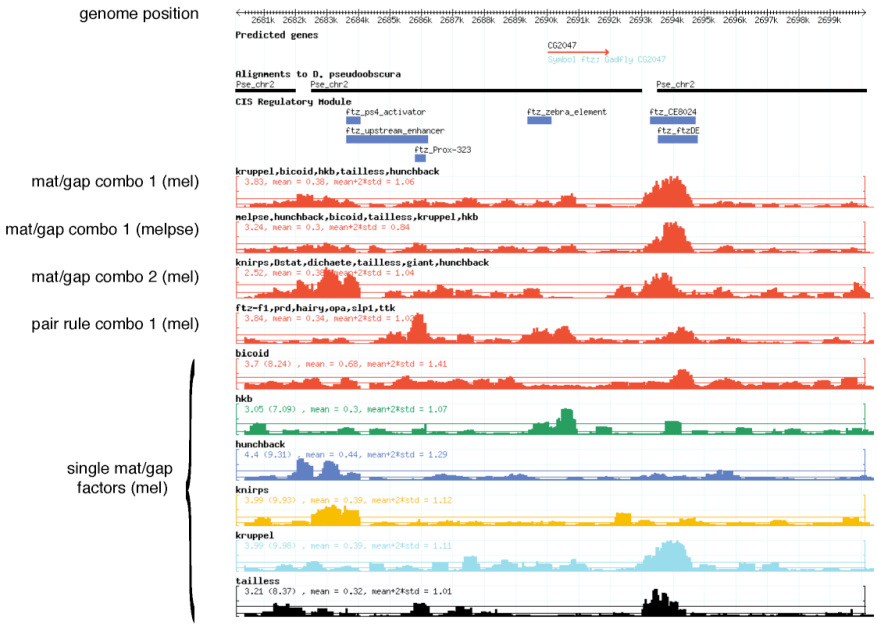
A



B



C



D

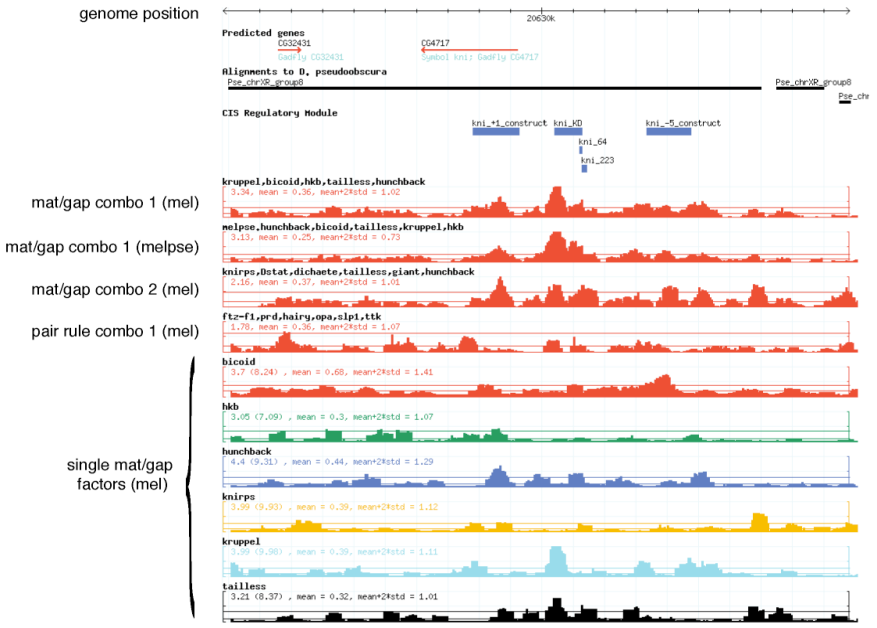


Figure 10. Hits two through five from the anterior TF search. Loci surrounding hits neighboring A) *hairy*, B) *odd*, C) *ftz* and D) *knirps*. Strong peaks of binding site overrepresentation are observed in all four of these loci and they overlap with previously identified CRMs that control patterned gene expression in the early embryo. For the three pair-rule genes, these peaks overlap with CRMs that control expression of the most anterior stripe (stripe 1) in their patterned expression. Otherwise the tracks displayed in these images are identical to those in Figure 6.

Supplementary Tables 1 & 2.

Due to their length, they have been submitted as separate files to the editor.

Supplementary Table 3.

Evaluation of CRM scoring function using 48 defined CRMs and 4800 random regions.

	Number of CRMs of 48 sampled above threshold		
Scoring Function	P≤0.005	P≤0.01	P≤0.05
Stubb, all PWMs together	13	14	24
Sum-of-PWMs, dict	9	15	23
Sum-of-PWMs, dict, z,	13	17	25
Sum-of-PWMs, dict, z + thresholding	**16**	**19**	**29**
Sum-of-PWMs, dict, z + thresholding, Mel-Pse	17	20	33
Sum-of-PWMs, dict, z + thresholding, Mel-Moj (43 CRMs)	16	20	31
Above values extrapolated to 48 CRMs*	18	22	35
Sum-of-PWMs, dict, z + thresholding, Mel-Pse-Moj (43 CRMs)	14	21	28
Above values extrapolated to 48 CRMs*	16	23	31

*Syntenic regions for only 43 CRMs are available in *D. mojavensis*

Supplementary Table 4.

Top 20 hits for D. melanogaster search with anterior TFs.

Rank	Location	Dmel	Dpse	DmelDpse	Motifs	gene	CRM	expression
1	3L 8645450	5.65	1.6	3.62	bcd, hkb, tll, kr	h	h_stripe 1	stripe 1
2	2R 5498250	5.3	3.59	4.44	bcd, hkb, tll, kr	eve	eve_stripe1	stripe1
3	3R 2694150	3.83	1.79	2.78	bcd, hkb, tll, kr	ftz	ftz_ftzDE	stripe1+5
4	2L 12682200	3.66	2.14	2.9	hkb, kr, hb	pdm2		stripes
5	2R 697950	3.43	n.a.	1.71	hkb, hb	nip		
6	X 9535700	3.41	1.94	2.67	bcd, tll, kr	btd	btd_Ss-Bg	head
7	3L 20630550	3.34	2.93	3.12	tll, kr	kni	kni_KD	ant+post
8	X 20534450	3.28	2.98	3.11	hkb, kr	run		stripes
9	3R 22133100	3.26	0.96	2.04	hkb, tll, kr, hb	CG33970		
10	2L 3834050	3.22	2.09	2.59	bcd, hkb, kr	slp2	slp2_-3	ant
11	X 7500350	3.2	1.89	2.54	tll, kr, hb	cut		CNS
12	3L 18806250	3.09	n.a.	1.54	bcd, tll, kr	CG32027		
13	3L 14858400	3.07	0.77	1.87	hkb, tll, kr	CG13467		
14	2R 999850	3.05	0.53	1.67	bcd, hkb, kr, hb	CG30440		
15	3R 12150	3.03	n.a.	1.26	hkb, hb	Dsk		
16	3R 4100	3.03	n.a.	1.51	hkb, hb	CG12581		
17	X 9284900	3.02	0.82	1.92	tll, kr, hb	CG15316		
18	2L 6295750	3.02	0.32	1.48	tll, kr, hb	Ddr		
19	X 10580150	3.01	0	1.5	hkb, tll, kr	CG32676		
20	3R 17474150	3.01	1.03	1.9	kr, hb	E2F		weak stripes

Supplementary Methods:

Construction of the *ΔrpoZ* Selection Strain

The omega selection strain was created by knocking out the *rpoZ* gene from the selection strain used with the alpha-based B1H system (US0 Δ*hisB*, Δ*pyrF*). This gene inactivation was accomplished using the method detailed by Datsenko and Wanner. To avoid using the Kanamycin or Chloramphenicol resistance markers for gene knock-outs that was developed by Datsenko and colleagues (7), which are already utilized in the B1H system, we created a Zeocin resistant version of the pKD4 plasmid. This template has the Zeocin resistance cassette flanked by FRT sites with the P1 and P2 priming sequences 5' and 3', respectively, to the FRT flanked cassette. The following primers were used to amplify both the Zeocin and the Kanamycin recombination cassette using the P1 and P2 sites:

rpoZ 5' (genomic homology italics, P1 site is bold)
5'*ATGCCCAGTCATTTCTTCACCTGTGGAGCTTTTTAAGTATGGCACGCGTAAC TGTTCA***TGTGTAGGCTGGAGCTGCTTCG**-3'

rpoZ 3' (genomic homology italics, P2 site is bold)
5'*ACAAGGGCGACCCGCTTTGTGATTAACGACGACCTTCAGCAATAGCGGTAA CGGC***CATATGAATATCCTCCTTAGTTCCT**-3'

The resulting PCR products were introduced as recombination donors in conjunction with lambda red recombinase to inactivate the *rpoZ* locus. The resistance marker was removed following isolation of a confirmed knockout strain by FLP recombinase. PCR primers were designed to prime approximately 100 basepairs upstream and downstream from the *rpoZ* locus to confirm insertion of the resistance marker and its removal from this locus.

rpoZ 5' PCR primer
5'-GCAGCGTCATGACGCTTTAA-3'
rpoZ 3' PCR primer
5'-GATTTGGTCTTCCGGCAGG-3'

Following amplification of this locus, the PCR products were run on a 1.5% agarose gel to confirm the change in mobility of the products that should be associated with insertion of antibiotic resistance marker and their removal (Supplementary figure 1). These PCR products were sequenced to verify *rpoZ* replacement and deletion. The strain containing the Zeocin insertion into the *rpoZ* gene (US0 Δ*hisB*, Δ*pyrF, ΔrpoZ:zeo*) was used in all of the B1H selections because it provides an additional selectable marker for identification and maintenance of the selection strain that is orthogonal to all of the plasmid-based markers and it demonstrated the same activity when compared to the complete *rpoZ* knockout (data not shown).

Omega Constructs (Figure 2b)

Direct Omega fusions

The coding sequences for the wt omega subunit contains a BamHI and KpnI site, both of which are potential cloning candidates in our expression vectors. Therefore, the following primers were used to PCR amplify the omega subunit

from *E. coli* genomic DNA in two fragments that would remove these target sequences.

Omega 5' (introduces a NcoI site at 5' end of gene)
5'-GCGGAATTCCATGGCACGCGTAACTGTTCAG-3'
Omega int 3' (introduces two silent mutations, removing restriction sites)
5'-TTCTTCCGGTACGAGCGGGTCCTTTCCGCC-3'
Omega int 5' (compliment to Omega int 5')
5'-GGCGGAAAGGACCCGCTCGTACCGGAAGAA-3'
Omega 3' (removes R91 and stop codon, introduces NotI site)
5'-TGCGCGGCCGCACGACCTTCAGCAATAGCGGT-3'

The first fragment was amplified using Omega 5' and Omega int 3'. This fragment introduced an NcoI site at the 5' end of the gene which contains the ATG start site that will be used for expression of the omega-TF hybrid. The second fragment was amplified with Omega int 5' and Omega 3' which removed the stop codon and the last residue (R91) of the omega protein while introducing a NotI site. The two internal primers introduced two silent mutations that removed the BamHI and KpnI target sites. The final PCR was done using an equal molar mix of the first two fragments as templates for the external Omega 5' and Omega 3' primers. The resulting PCR product was recovered and digested with NcoI and NotI. The digestion product was ligated into the previously described pB1H2 expression plasmid (5). This resulted in a plasmid that could express omega-TF hybrids with exactly the same promoter and linker that had been used in the alpha system. This plasmid for expressing a transcription factor as a fusion to the C-terminus of omega was named pB1H2ωL where the L signifies the *lppC lacUV5* dual promoter that is driving expression of the omega-TF hybrid (8). The pB1H2ω2 and pB1H2ω5 expression plasmids were created by replacing a fragment of the pB1H2ωL plasmid that contains the promoter (EcoRI to NcoI) with a fragment that contains either the *lacUV5* promoter (pB1H2ω5) or a mutant version of the *lacUV5* promoter (*lacUV5m*) that has two mutations in the -10 box (pB1H2ω2). Once the three primary expression plasmids (pB1H2ω2, pB1H2ω5, and pB1H2ωL) had been constructed, a TF could be introduced into any of these plasmids for expression by designing a KpnI site in frame into the 5' primer and a stop codon and XbaI site into the 3' primer used to amplify the TF from a DNA template (genomic DNA or cDNA). The PCR product could then be digested with KpnI and XbaI and ligated into the pB1H2ω backbone DNA with the desired promoter strength (Supplementary figure 2). This universal cloning strategy allowed a TF to be easily moved between the plasmids by sub-cloning to examine the impact of changing the promoter strength. The linker between the omega-subunit and the TF was identical to the linker previously described in pB1H2, which contains a flag-tag allowing the expression of a TF to be verified by Western blot.

Omega-Zif12 fusions (Homeodomain constructs)

Homeodomains were expressed as omega fusions in combination with fingers one and two of Zif268 (Zif12) under control of the *lacUV5m* promoter plasmid (pB1H2ω2-12HD; Supplementary Figure 2). These expression constructs and the

validation of this system is described elsewhere (Noyes, Christensen, Wakabayashi, Stormo, Brodsky and Wolfe *manuscript in preparation*)

Library Construction

28 Basepair Library

The 28 basepair library was constructed in pH3U3 as previously described by Meng, Brodsky and Wolfe except that it was introduced between the NotI and EcoRI sites using the following oligonucleotides:

28 bp library oligonucleotide:

5'GGCGCGAATTCGNNNNNNNNNNNNNNNNNNNNNNNNNNGCGGCC GCAAGGTAGCTGATTCCGTTCTCGC-3'

Library extension primer:

5'-GCGAGAACGGAATCAGCTACCTT-3'

This library places the randomized region 7 bp away from the 5' edge of the -35 box of the weak promoter that controls expression of the HIS3/URA3 reporter genes. The 28 bp raw library contains ~2 x 10^8 independent clones based on titration of the initial transformants following electroporation of the ligated library after a 1 hour recovery in SOC medium.

ZF10 Library

The ZF10 library contains a binding site for Zif12 neighboring a 10 bp randomized region. The construction of this library and its counter-selection is described elsewhere (Noyes, Christensen, Wakabayashi, Stormo, Brodsky and Wolfe *manuscript in preparation*)

Counterselection of the 28 bp library

Counterselections were performed on the 28 bp library to remove self-activating sequences. 250 ng of the raw library material was transformed into 80 μl of the *rpoZ* positive version of the selection strain (US0 Δ*hisB*, Δ*pyrF*). 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 1ml 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 27 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. A second round of counterselection was performed to further reduce the background of self-activating clones, which remained higher than desired following the first counterseleciton. This additional counterselection was performed as described above except that approximately 2.5x10^8 transformed

cells were split over 20 5-FOA/YM plates. The number of self-activating sequences in the library was reduced by counter-selection to ~1 in 10^6 at a typical selection stringency (10 mM 3-AT).

Selection Overview (Figure 2)

Two different libraries were constructed to assess the specificity of each TF: a 28 bp randomized library for general use in characterizing factors, and a 10 bp library containing a neighboring binding site for fingers 1 and 2 of Zif268 (ZF10), which is for characterizing factors with low affinity or specificity. The 28 bp library, which was constructed for the characterization of direct omega-TF fusions has a relatively high background of false positive clones when compared to the previously described 18 pair library. The majority of this background results from a population of self-activating sequences in the randomized region that are coupled to a deactivated *URA3* reporter gene that allows these clones to persist even under stringent counterselection. This subpopulation is sensitive to the absence of uracil during the positive selection of binding sites, and consequently uracil was omitted from binding site selections using this library in the description of general procedure that follows. The false positive rate within the ZF10 library following counterselection was quite low and did not change significantly when uracil was omitted, consequently all selections using the ZF10 library were performed in the presence of uracil (200 μM uracil was added to each NM selection plate), which provides slightly greater growth rates under selective conditions.

General selection procedure (Figure 2c)

Approximately 2 μg of the bait plasmid (pB1H2ω2/5/L or pB1H2ω2/5-12HD) and 50 ng of the library plasmid were electroporated into 80 μl of the selection strain. The cells and the two plasmids were mixed on ice and moved to a pre-chilled 1 mm 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 μM uracil, and 0.1% histidine. 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 μl of this final resuspention 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.

Failed Selections

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 was a reliable indicator of a failed selection. This outcome was predominantly the result of improper expression of the TF (too low or too high). Factors that had too high an expression level displayed toxicity on rich media plates (reduced colony number relative to selections performed with other factors in parallel and/or extreme variability in colony size), which allowed us to gauge whether the expression level of a TF from a failed selection should be increased or decreased. If the expression plasmid for a particular factor appeared toxic, the TF was moved to a version of the pB1H2ω plasmid that contains a weaker promoter (*lacUV5m* or *lacUV5*) and the selections were repeated. If the expression plasmid for a particular factor did not appear to be toxic cells based on the rich media titrations but the selection failed, the TF was moved to a version of the pB1H2ω plasmid with a stronger promoter (*lacUV5* or *lppC*) and reselected. This approach proved successful in almost every example. In principle the concentration of the inducer IPTG could also be increased or decreased to further modify the expression levels, but we found adjusting the overall promoter strengths on the expression plasmid to be the most reliable path to success. 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 (9). Croc is the single example of a TF that did not generate a binding site motif using the omega-based B1H system. This factor produced colony numbers following selection that were between 20 and 100-fold over the background when expressed form either a *lacUV5* and *lppC* promoter at either 5 or 10mM 3-AT, however computational analysis of the selected binding sites was unable to identify a significantly overrepresented motif within this population. It is not clear why Croc was able to achieve a high fold over the background without selecting a specific binding site; one possible explanation is recognition of a binding site present in a fixed region of the pH3U3 plasmid outside the library window, which would return sequences that were not related to sequence-specific binding.

Medium for Selection procedure

SOC medium

Contains a final concentration of 1% filter sterilized glucose in autoclaved SOB (Becton Dickinson, cat. no. 244310).

NM medium

Prepare the following solution: 1xM9 Salts, 4 mg/ml glucose, 200 μM adenine-HCl, 1x Amino acid mixture (below), 1 mM MgSO4, 10 μg/ml thiamine, 10 μM ZnSO4, 100 μM CaCl2, 25ug/ml kanamycin, 100 μg/ml carbenicillin, and 10 μM IPTG. Filter sterilize through a 0.22 μm filter.

NM selective plates

Selective plates contain: 1.5% autoclaved agar, 1xM9 Salts, 4 mg/ml glucose, 200 μM adenine-HCl, 1x Amino acid mixture (below), 1 mM MgSO4, 10 μg/ml thiamine, 10 μM ZnSO4, 100 μM CaCl2, 25 μg/ml kanamycin, 100 μg/ml carbenicillin, 10 μM IPTG and the desired concentration of 3-AT.

YM medium

Prepare the following solution: 1× M9 salts, 4 mg/ml glucose, 0.01% yeast extract, 10 μg/ml thiamine, 200 μM uracil, 0.1% histidine, 10 μM $ZnSO_4$, 100 μM $CaCl_2$ and 1 mM $MgSO_4$. Filter-sterilize through a 0.22-μm filter.

YM 5-FOA plates

Selective plates contain: 1.5% autoclaved agar, 1× M9 salts, 4 mg/ml glucose, 0.01% yeast extract, 10 μg/ml thiamine, 200 μM uracil, 0.1% histidine, 10 μM $ZnSO_4$, 100 μM $CaCl_2$, 1 mM $MgSO_4$, 25 μg/ml kanamycin and the desired concentration of 5-FOA.

Amino acid mixture (33.3x)

Contains 17 of the 20 amino acids, omitting His, Met, and Cys.

Prepare the following six solutions (all percentages are wt/vol):

Solution I (200x): dissolve 0.99g Phe (0.99%), 1.1g Lys (1.1%) and 2.5g Arg (2.5%) in water to a final volume of 100ml.

Solution II (200x): dissolve 0.2 g Gly (0.2%), 0.7 g Val (0.7%), 0.84 g Ala (0.84%) and 0.41 g Trp (0.41%) in water to a final volume of 100ml.

Solution III (200x): dissolve 0.71g Thr (0.71%), 8.4 g Ser (8.4%), 4.6 g Pro (4.6%) and 0.96 g Asn (0.96%) in water to a final volume of 100ml.

Solution IV (200x): add 1.04g Asp (1.04%) and 18.7g potassium-Glu (18.7%) to water, bring to a final volume of 100ml.

Solution V (200x): add 14.6g Gln (14.6%) and 0.36g Tyr (0.36%) to roughly 90ml of water. Add solution V to solution IV. Add NaOH pellets slowly until all amino acids go into solution. Bring final volume to 200ml.

Solution VI (200x): dissolve 0.79 g Ile (0.79%) and 0.77 g Leu (0.77%) in water to a final volume of 100ml.

Mix solutions I to VI together and filter sterilize through a 0.22 μm filter and store at 4C. This results in a 33.3x amino acid mixture.

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 25µl of the PCR mix (1µM HU100 primer, 1µM OK181 primer, 300µM 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 25ul 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, 5µl of each 25µl 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'

Motif discovery and alignment

The chromatogram from each sequence read was inspected for quality and accuracy, and if judged interpretable, the sequence between the NotI site and the EcoRI site (the library window) was confirmed by visual inspection of the peaks. The group of all unique library sequences recovered for a TF was then analyzed using MEME motif discovery tool (http://meme.sdsc.edu/meme/intro.html). Overrepresented motifs were recovered using the following settings in MEME. Selections employing the 28 bp library: zero or one motif per sequence could be discovered, a motif could be discovered on either strand of the DNA and the search width for a motif was set from 3 to 28 bases. Selections employing the ZF10 library were identical except that the search width for a motif was set from 3 to 10 bases. In all instance where a binding site motif was successfully identified, it was the top hit recovered from the MEME search and had an expectation value that was $<e^{-5}$. The aligned sequences that compose each motif recovered by MEME were used to generate a Sequence logo using the WebLogo tool (http://weblogo.berkeley.edu/logo.cgi).

Omega-TF activity assays

Target sites in pH3U3 for Zif268, Paired, Giant and Hunchback

The following binding sites were designed for the promoter strength activity assays based on the preferred consensus sequence and the preferred position of that site relative to the promoter that were determined for each factor from an omega-B1H binding site selection. For the Zif268 mutant sites, mutant target sequences were designed based on previously determined binding constants for variants of the finger 1 binding site.

Giant
5'-GGCCGCAGACCGGAG**ATTACGTAAC**TATAAGACACG-3'
Hunchback
5'-GGCCGCCTACCGGAGCGATA**CACAAAAAAACA**TGCG-3'
Paired
5'-GGCCGCGAGTCTCACATAC**ATCCGTCACGCT**ACCCG-3'
Zif268-F1 wt (GCG)
5'-GGCCGCTGCGTGG**GCG**GGACG-3'
Zif268-F1 GAG
5'-GGCCGCTGCGTGG**GAG**GGACG-3'
Zif268-F1 GGG
5'-GGCCGCTGCGTGG**GGG**GGACG-3'
Zif268-F1 GCA
5'-GGCCGCTGCGTGG**GCA**GGACG-3'

These oligonucleotides were annealed to a complementary oligonucleotide that contains a 5' AATT overhang and were truncated by four bp on the 3' end to leave a GGCC overhang on the complementary strand. These overhangs are complimentary to the overhangs created by a NotI/EcoRI digestion of pH3U3. Each complementary pair of oligonucleotides were annealed and ligated between the Not1 and EcoR1 site of the pH3U3 plasmid. Cloned binding sites were verified by sequencing.

Zif268 activity on mutant finger 1 binding sites

100 ng of pB1H2ω2-Zif268 and 50 ng of a reporter pH3U3 plasmid containing one of four different Zif268 finger 1 binding sites (GCG, GCA, GAG, or GGG) were transformed into 80 μl of the selection strain by electroporation. The cells were recovered in 1 ml of SOC for 1 hour at 37°C. After recovery, the cells were plated on rich media plates (2xYT + 25 μg/ml Kanamycin, and 100 μg/ml Carbenicillin) and grown overnight at 37°C. A single colony was selected from each plate and a 5 ml 2xYT culture with Kanamycin and Carbenicillin was grown overnight at 37°C. The following day, a second 5 ml culture in identical medium was inoculated with 50 μl of the overnight, saturated culture and grown to an OD600 of approximately 0.2 to 0.4. The cells in each culture were pelleted by centrifugation at 3000 rpm for 15 minutes. The pellet was then resuspended in 5 ml NM media supplemented with 200 μM uracil. The cells were acclimated to NM media for 1 hour at 37°C while rotating. This cells in this culture were pelleted by centrifugation at 3000 rpm for 15 minutes. The cells were then washed 4 times by sequential pelleting and resuspension in NM media

supplemented with 200 μM uracil and finally resuspended in a volume of 1ml. 20 μl 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 cells. 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 a fraction of the culture was diluted to roughly 500 cells per 750 μl. Next, 750 μl of each diluted culture was spread on small NM selective plates with various concentrations of 3-AT. Each 3-AT concentration challenge was performed in triplicate for each binding site. Plates were wrapped in parafilm and grown at 37 °C for 36 hours. The number of surviving colonies was counted on each plate. Next, the colonies on each plate were resuspended in 5 ml of 2xYT and the cell suspension from the three duplicate plates for each 3-AT concentration were pooled. These cells were pelleted by centrifugation at 3000 rpm for 15 minutes and then the pellet was resuspended in 100 ml of 2xYT and an OD600 reading was measured in triplicate to determine the approximate number of cells recovered form each group of selection plates. For easy interpretation, the mean colony counts from each condition for each binding site were normalized by the colony count from non-selective conditions (NM medium with no 3-AT and 0.1% Histidine) to provide a fractional colony count. This fractional colony count for each binding site at each stringency describes the percentage of surviving colonies relative to the number observed at non-selective conditions (maximum survival rate). The measured OD600's were normalized by colony count to provide quantitative basis for evaluating the colony size that was recovered for each binding site at each stringency. To remove the complication of the number of colonies contributing to the OD600 at a stringency, the mean OD600 measurement at each stringency was multiplied by the fraction of the non-selective colony count divided by the colony count at a that specific stringency. This product represents OD600's at equivalent cell densities. Once the OD600's were normalized by colony count, the percentage of the OD600 from non-selective conditions (0mM 3-AT with 0.1% Histidine) was calculated for each binding site, 3-AT combination.

Paired, Giant and Hunchback Activity Assay

100 ng of pB1H2ω2, 5, and L versions of each TF was cotransformed with 50 ng of a reporter pH3U3 plasmid containing the optimal binding site for the corresponding TF into 80 μl of the selection strain by electroporation. The cells were recovered in 1ml of SOC for 1 hour at 37°C. After recovery, the cells were plated on rich media plates (2xYT + 25 μg/ml Kanamycin, and 100 μg/ml Carbenicillin) and grown overnight at 37°C. The following day, a single colony was selected from each plate and a 5ml 2xYT culture with Kanamycin and Carbenicillin was grown overnight at 37°C. A second 5 ml culture in identical medium was inoculated with 50 μl of this overnight, saturated culture and grown to an OD600 of approximately 0.2 to 0.4. The cells in this culture were pelleted by centrifugation at 3000 rpm for 15 minutes. The pellet was resuspended in 5

ml NM medium supplemented with 200 μM uracil, and 0.1% histidine. The cells were acclimated to NM media for 1 hour at 37°C while rotating. This culture was pelleted by centrifiugation at 3000 rpm for 15 minutes. The cells were then washed 4 times by sequential pelleting and resuspension in NM medium supplemented with 200 μM uracil and then resuspended in a final volume of 1ml. 20 μl of this final resuspention was titrated by 10-fold serial dilution on rich media plates (2xYT + 25 μg/ml Kanamycin and 100 μg/ml Carbenicillin) to determine the total number of cells. 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. The cell concentration in each sample was normalized based on the titration results to achieve uniform cell densities at each cell dilution. The normalized cell cultures were then titrated by 10-fold serial dilution on rich media and NM selective plates supplemented with 200μM uracil and various 3-AT concentrations. The NM selective plates were wrapped with parafilm and grown at 37°C for 36 hours whereas the rich media plates were grown overnight at 37°C.

Western blot assay

100 ng of pB1H2ω2, 5, and L versions of each TF assayed was cotransformed with 50 ng of a reporter pH3U3 plasmid containing the optimal binding site for the corresponding TF into 80 μl of the selection strain by electroporation. The cells were recovered in 1ml of SOC for 1 hour at 37°C. After recovery, the cells were plated on rich media plates (2xYT + 25 μg/ml Kanamycin, and 100 μg/ml Carbenicillin) and grown overnight at 37°C. The following day, a single colony was selected from each plate and a 5ml 2xYT culture with Kanamycin and Carbenicillin was grown overnight at 37°C. A second 5 ml culture in identical medium with the addition of 10μM IPTG was inoculated with 50 μl of this overnight, saturated culture and grown to an OD600 of approximately 0.4 to 0.6. 1.5ml of each culture was pelleted by centrifugation at 18,000g for 2 minutes. The pellets were resuspended in a volume of 1xSDS loading buffer that was normalized by the OD600 of each sample in comparison to the sample with the lowest OD600, which was resuspended in 50μl. The samples were promptly boiled for 10 minutes. After boiling, the samples were then centrifuged briefly to pellet debris and diluted 1:10 in 1xSDS loading buffer. 10μl of each of the diluted samples were loaded on to a 14% SDS-polyacrylamide gel along with 5μl of Kaleidoscope Prestained Standard (BioRad). The samples were run on the gel for approximately 2.5 hours. The gel was then transferred to 0.45 micron polyvinylidene difluoride (PVDF) membrane and the membrane was blocked overnight with gentle rocking in binding buffer (5% nonfat dry milk resupended in wash buffer). After blocking the membrane was washed four times in wash buffer (20mM Tris, 150mM NaCl, and 0.05% Tween 20). The membrane was then labeled for 2 hours with 2.5μl of anti-FLAG per 10ml of binding buffer. The membrane was again washed four times with wash buffer. The secondary labeling was done for 2 hours with 1ml anti-murine-hc-HRP per 15ml binding buffer. The membrane was washed four times with wash buffer. Finally, the

HRP was reacted using the Millipore Immobolin reagents for HRP substrates. The membrane was bathed with this Peroxide Solution/Luminol Reagent, 1:1 mix for 1 minute before exposing film. Film exposures were taken from 2 seconds to 5 minutes.

Genome-wide search tool

The genome-wide search tool retrieves the top 10-100 scoring 500 bp windows using any single motif or motif combination. Once a particular window is selected, all windows within 2 Kbp on either side are excluded from consideration to avoid redundancies in the final report. An optional filter forces the tool to report only those windows where at least some (user-specified) number of motifs have individual z-scores above a user-specified threshold. The position of each window in the genome is reported along with the single species (*D. melanogaster* and *D. pseudoobscura*) and two-species scores. The results can be sorted either on the single species or on the two-species score.

References:

1. Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O. and Darst, S.A. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science*, **296**, 1285-1290.
2. Durai, S., Bosley, A., Abulencia, A.B., Chandrasegaran, S. and Ostermeier, M. (2006) A bacterial one-hybrid selection system for interrogating zinc finger-DNA interactions. *Comb Chem High Throughput Screen*, **9**, 301-311.
3. Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*, **2**, 28-36.
4. Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V. *et al.* (2003) TRANSFAC(R): transcriptional regulation, from patterns to profiles. *Nucl. Acids Res.*, **31**, 374-378.
5. Meng, X., Brodsky, M.H. and Wolfe, S.A. (2005) A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nat. Biotechnol.*, **23**, 988-994.
6. Wolfe, S.A., Greisman, H.A., Ramm, E.I. and Pabo, C.O. (1999) Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J. Mol. Biol.*, **285**, 1917-1934.
7. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A*, **97**, 6640-6645.
8. Hu, J.C., Kornacker, M.G. and Hochschild, A. (2000) *Escherichia coli* one- and two-hybrid systems for the analysis and identification of protein-protein interactions. *Methods*, **20**, 80-94.
9. Parsell, D.A., Silber, K.R. and Sauer, R.T. (1990) Carboxy-terminal determinants of intracellular protein degradation. *Genes Dev*, **4**, 277-286.