Supplemental Discussion

The splice isoforms of *broad* and *tramtrack*

br (broad) is an ecdysone-regulated gene that is essential for metamorphosis during *Drosophila* development (Thummel 2001). Four splice isoforms (Z1, Z2, Z3 and Z4) from this gene play both distinct and overlapping roles in gene regulation in a large number of tissues during metamorphosis (Bayer et al. 1996; Spokony and Restifo 2007). Mutations that individually disrupt the function of three of the four isoforms present phenotypes that represent distinct complementation groups (Bayer et al. 1996; Bayer et al. 1997; Spokony and Restifo 2007). All four of these isoforms encode unique pairs of zinc fingers (**Supplementary Figure 4A, B**) and consequently have the potential to recognize distinct sequences. Analysis of their DNA-binding specificities reveals that each isoform displays a unique recognition sequence (**Supplementary Figure 4C**). Interestingly, isoforms Z1 (PE) and Z4 (PC) display the greatest similarity in their recognition sequences, and Z4 is the only isoform (besides Z1) that can partially complement the loss of Z1 during *Drosophila* metamorphosis (Bayer et al. 1997).

ttk (tramtrack) is a transcription factor that plays fundamental roles in early *Drosophila* development (Brown and Wu 1993; Harrison and Travers 1990; Read et al. 1992). Ttk has two alternate spliced isoforms (-PF, 69kD and -PA, 88kD, **Supplementary Figure 5**) with different fingers sets and distinct DNA-binding specificities (Read and Manley 1992) that appear to play independent roles in development (Xiong and Montell 1993). The DNA-binding specificity of the 69kD isoform is well defined (Bergman et al. 2005; Noyes et al. 2008; Read and Manley 1992), however the preferred recognition motif for the 88kD isoform (AGGG(C/T)GG) is based on the footprinting analysis of only three bound sequence (Badenhorst et al. 1996; Read and Manley 1992). Our analysis of these isoforms provides greater clarity for the preferred recognition sequences for the 88kD isoform and confirms that the 69kD and 88kD isoforms have

fundamental differences in their recognition preferences (**Supplementary Figure 5**).

Current frameworks for deconvoluting zinc-finger DNA recognition preferences

Common specificity determinants for DNA-recognition by zinc fingers have largely been extracted from the biochemical and structural characterization of a small number of naturally-occurring ZFPs (Badis et al. 2009; Badis et al. 2008; Noyes et al. 2008; Wolfe et al. 2000; Zhu et al. 2009) and the selection and characterization of artificial ZFPs that recognize novel target sequences (Bae et al. 2003; Bulyk et al. 2001; Dreier et al. 2001; Dreier et al. 2005; Dreier et al. 2000; Greisman and Pabo 1997; Gupta et al. 2012; Isalan et al. 1998; Isalan et al. 2001; Lam et al. 2011; Liu et al. 2002; Maeder et al. 2008; Sander et al. 2011; Segal et al. 1999; Wolfe et al. 1999). These data have provided a foundation for predictive recognition models to estimate the DNA-binding specificity of naturally-occurring ZFPs (Benos et al. 2002; Cho et al. 2008; Kaplan et al. 2005; Liu and Stormo 2008; Persikov et al. 2009; Workman et al. 2005). The growing archive of DNA-binding specificities for naturally-occurring ZFPs (Badis et al. 2009; Badis et al. 2008; Jolma et al. 2010; Noyes et al. 2008; Zhu et al. 2009) and the available structural information on a number of canonical-binding variants that recognize different target sequences (Elrod-Erickson and Pabo 1999; Kim and Berg 1996; Stoll et al. 2007; Wolfe et al. 2001) have also facilitated structural-based modeling approaches to predict DNA-binding specificity (Havranek et al. 2004; Siggers and Honig 2007; Yanover and Bradley 2011). However, systematic incorporation of each ZFP motif into a predictive recognition model typically requires the association of each individual finger with a DNA subsite within the target sequence. Thus, our desire to extract associations between fingers and subsites within our dataset.

Analysis of the recognition preferences at position 2

Serine is the most common residue present at position 2 in our ZFPs, and consistent with previous studies (Kim and Berg 1995), this residue displays no

particular preference for recognition at this position. Aspartate also occurs frequently, almost exclusively in the context of RXD motifs (40 of 44 occurrences), which is consistent with prior studies shows a strong preference for Gua or Thy as the neighboring 3' base (Isalan et al. 1997; Swirnoff and Milbrandt 1995). Contrastingly, positively charged residues at position 2 (Arg/Lys) occur primarily with Ade or Cyt at the neighboring base position, which is a trend also observed in the selection of artificial fingers where residues on neighboring recognition helices have been simultaneously randomized (Gupta et al. 2012; Isalan et al. 1998). The consistency of our data with previous analyses of ZFP specificity suggests that other novel trends that are observed within this dataset (e.g. correlation of Phe/Tyr at position 2 with Gua or Cyt at the neighboring base position) will provide a valuable framework for predicting recognition by uncharacterized naturally-occurring ZFPs.

Supplemental Figure 1. Genome-wide distribution of canonically linked Cys_2-His_2 ZFPs. A) Distribution of the linker lengths (< 20 amino acids) joining neighboring fingers within the zinc finger containing genes. B) Sequence logo (information content) of the 730 five amino acid linkers connecting zinc fingers within the *Drosophila* genome. There is a strong bias toward sequences that correspond to TGE(K/R)P-type composition. C) Pie chart representing the 282 multi-finger proteins in *Drosophila*. Forty-six of these ZFPs have all of the fingers linked by canonical (five amino acid) linkers. Sixty-six of these ZFPs have one or more of the fingers linked by a canonical linker. D) Schematic of the zinc finger domains in CG4360 and how these are deconstructed into 3 clusters based the linker size that connects neighboring fingers. Individual clusters are then incorporated into the B1H system for specificity analysis (Noyes et al. 2008).

Supplemental Figure 2. Ninety-four B1H characterized Cys_2-His_2 ZFP recognition motifs displayed as information content Sequence Logos.

Supplemental Figure 3. Success rate for attempted Cys_2-His_2 zinc finger *clusters* in the B1H system based on: A), B) Canonical linkage of neighboring fingers within the characterized cluster and C), D) Number of constituent fingers within the characterized cluster, where a cluster can represent a subset of fingers from a gene.

Supplemental Figure 4. BR family analysis. A) Alignment of Broad isoforms that contain two zinc fingers. B) Information content Sequence Logo of the aligned two-finger modules in A. C) DNA-binding specificity of the four splice isoforms (br-PA, br-PC, br-PE and br-PL) encoded by the *br* locus.

Supplemental Figure 5. TTK family analysis. A) Alignment of two finger containing Tramtrack isoforms (-PA 88kD, and -PF 69kD). B) Preferred recognition motifs for ttk-PA and –PF.

Supplemental Figure 6. Lola family analysis. A) Alignment of two-finger units from Lola isoforms that contain two or more fingers. All of these fingers have a characteristic CCHC zinc coordination residues in finger 1. The sequences of these recognition helices are generally diverse, although there are pairs of fingers that are identical (e.g. -PT and -PU). B) Information content Sequence Logo of the aligned two-finger modules in A. Note: In part A, only fingers 2 and 3 of lola-PN (the only three-finger lola isoform) are included in the alignment, as these are the fingers responsible for DNA binding based on the recovered recognition motif (*i.e*. fingers 2 and 3 of lola-PN have the same amino acid residue content as lola-PY).

Supplemental Figure 7. TOMTOM comparative motif analysis (Gupta et al. 2007) identifies strong similarity between the Shn and NF-KB recognition sequences. Comparison of the Shn recognition motif with motifs present in the

JASPAR database (Bryne et al. 2008) identifies significant similarity with NF-κB recognition motifs.

Supplemental Figure 8. Assigning subsites of recognition for individual fingers. A) A schematic depicting canonical DNA recognition by a Cys_2-His_2 zinc finger. The numbered spheres on the α-helix represent the residues that are anticipated to contact DNA in the canonical recognition mode. These residues are numbered relative to the start of the α-helix and make contact (arrows) with their respective color-coded DNA bases (boxes). Each finger (in an N-terminal to C-terminal orientation) binds its DNA subsite (labeled 5' to 3') in an anti-parallel arrangement. The labels in the boxes 5, M, 3 and N refer to the 5', middle, 3' and neighboring subsite bases relative to the three base pair core subsite recognized by the finger. B) A recognition "code" for defining finger register within a recognition sequence. Based on previously defined correlations between specificity determinants and preferred bases at each position of the DNA subsite (Wolfe et al. 2000) we can assign fingers to subsites within the recognition motifs for many ZFPs. C) CG9895 provides an example of a ZFPs where the register of the fingers on the target site can be readily assigned based on previously determined amino acid - base correlations. The amino acids in the recognition helix of CG9895 are shown above the recognition motif in an anti-parallel orientation where the number of amino acids in the linker connecting neighboring fingers is indicated. Amino acid - base correlations are denoted by arrows.

Supplemental Figure 9. Specificity assignments for ZFPs based on the characterization of finger subsets. Sens contains 4 canonically linked zinc fingers, however the orientation and position of the fingers on the recognition motif is not inherently obvious. We determined the specificity of fingers 1 to 3, which provides information both on the register and orientation of the fingers relative to this submotif. Based on this information the position of the Sens fingers on the DNA can be confidently assigned.

Supplemental Figure 10. Specificity assignments for ZFPs based on the characterization of a finger swap construct. D19B (fingers 10-12) contains 3 canonically linked zinc fingers, however the orientation of the fingers on the recognition motif is not inherently obvious. We determined the specificity of D19B-F10-11 (F1-2 in this construct), fused to finger 3 of CG4360 whose specificity is clearly defined from our previous analysis. This yields information both on the register and orientation of the fingers on the motif enabling assignment of the D19B fingers specificities.

Supplemental Figure 11. Binding site motifs for finger-subsets. A) B1H recognition motifs for thirty-three *Drosophila* zinc finger subsets or finger sets generated by splicing together fingers from different zinc finger arrays. The header for each motif indicates the gene of origin of the fingers and the finger numbers are indicated as "F#", where the fingers are numbered in order of occurrence in the gene. Spliced finger sets are indicated by two gene names

and the fingers from each gene (e.g. Blimp-1-F1_CG4360F2-3) or in the case of two isoforms that have been combined (for the Broad or Lola families), by the isoform fingers that were combined. For example “Br-PAPC” indicates Br-PA-F1 fused to Br-PC-F2. B) Finger subsets analyzed for the human ZFP BCL6.

Supplemental Figure 12. Frequency Sequence logos for larger triplet bins of recognition helices (in Figure 4) displaying a number of unique members. The position of the determinants in the recognition helix are numbered relative to the start of the helix. The average recognition motif over the core triplet for members in each bin is labeled as in Figure 1B.

Supplemental Figure 13. Amino acid - base correlations**.** Frequency logo displaying the average base preference for each amino acid at potential recognition position on the Recognition Helix (RH) assuming canonical recognition. The total number of recognition helices and the number of unique recognition helices that contain the amino acid at that position are indicated above each logo, where a unique set considers residues at positions -1,1, 2, 3 and 6). Base position nomenclature is defined in Supplemental Figure 9.

Supplemental Figure 14. Tyrosine at position 3 specifies A. Recognition motif of an artificial zinc finger protein containing Tyr at position 3 of finger 3 displays a strong preference for A at the middle position of the finger triplet (circled).

Supplemental Figure 15. B1H-determined specificities of the artificial ZFAs assembled entirely from artificial finger sets for use in the ZFN activity assays. In the case of 5p_irs1b-like_n, arginine has been introduced at position 6 of the second finger to encourage a preference for guanine at the corresponding position in the finger subsite (indicated with red box).

Supplemental Figure 16. Activity and toxicity of the *nhlh2* ZFNs. A) Dose response curve of the *nhlh2* ZFNs in zebrafish embryos. Embryos were sorted into groups based on morphology and survival at 24 hpf where “Monsters” indicates embryos with morphological defects (Meng et al. 2008). Lesion frequency was assessed by loss of sensitivity of a PCR product spanning the genomic target site to HaeIII. M denotes 100bp DNA ladder (NEB). Lesion frequency was determined by ImageJ analysis of the uncleaved and cleaved DNA bands. C) Lesions in shotgun cloned sequences from the HaeIII resistant band.

Supplemental Figure 17. Activity and toxicity of the *irs1* ZFNs. A) Dose response curve of the *irs1* ZFNs in zebrafish embryos. Embryos were sorted into groups based on morphology and survival at 24 hpf where “Monsters” indicates embryos with morphological defects (Meng et al. 2008). Lesion frequency was assessed by T7EI sensitivity of a PCR product spanning the genomic target site (Kim et al. 2009). M denotes 100bp DNA ladder (NEB). Lesion frequency was determined by ImageJ analysis of the uncleaved and cleaved DNA bands. C)

Lesions in shotgun cloned sequences from *lacZ* assay for frame-shifted products.

Supplemental Figure 18. Activity and toxicity of the *nr3c1* ZFNs. A) Dose response curve of the *nr3c1* ZFNs in zebrafish embryos. Embryos were sorted into groups based on morphology and survival at 24 hpf where "Monsters" indicates embryos with morphological defects (Meng et al. 2008). Lesion frequency was assessed by T7EI sensitivity of a PCR product spanning the genomic target site (Kim et al. 2009). M denotes 100bp DNA ladder (NEB). Lesion frequency was determined by ImageJ analysis of the uncleaved and cleaved DNA bands. C) Lesions in shotgun cloned sequences from *lacZ* assay for frame-shifted products.

Supplemental Figure 19. Activity and toxicity of the *irs1b*-like ZFNs. A) Dose response curve of the *irs1b*-like ZFNs in zebrafish embryos. Embryos were sorted into groups based on morphology and survival at 24 hpf where "Monsters" indicates embryos with morphological defects (Meng et al. 2008). Lesion frequency was assessed by loss of sensitivity of a PCR product spanning the genomic target site to PflMI. M denotes 100bp DNA ladder (NEB). Lesion frequency was determined by ImageJ analysis of the uncleaved and cleaved DNA bands. C) Lesions in shotgun cloned sequences from the PflMI resistant band.

Supplemental Figure 20. Relationship between number of characterized fingers and the width of recognition motif recovered in B1H analysis. Each motifs edge was defined by the final position with information content >0.5 bits. Filled circles in the plot represent individual motif-finger combinations, where the members in each finger number bin are colored differently. There is a clear increase in motif width from 2 to 3 finger units, after which the distribution plateaus.

Supplemental Dataset 1. Position Frequency Matrices of the core triplet for the assigned fingers within the ZFP recognition motifs. The recognition residues at positions -1, 2, 3 and 6 for each finger, and its position and strand in the larger motif are indicated in the header of each entry.

Supplemental Table 1. Three hundred and twenty-seven Cys2-His2 ZFPs in the *Drosophila* genome.

Supplemental Table 2. Successfully characterized Cys2-His2 ZFPs from *Drosophila melanogaster*.

Supplemental Table 3. Alternately spliced genes that display isoform-dependent changes in zinc finger composition or number.

Supplemental Table 4. Single finger – DNA subsite combinations derived from characterized 83 Cys2-His2 ZFPs from *D. melanogaster*.

Supplemental Table 5. Recognition helices for all *Drosophila* single finger – DNA subsite combinations represented in Figure 4.

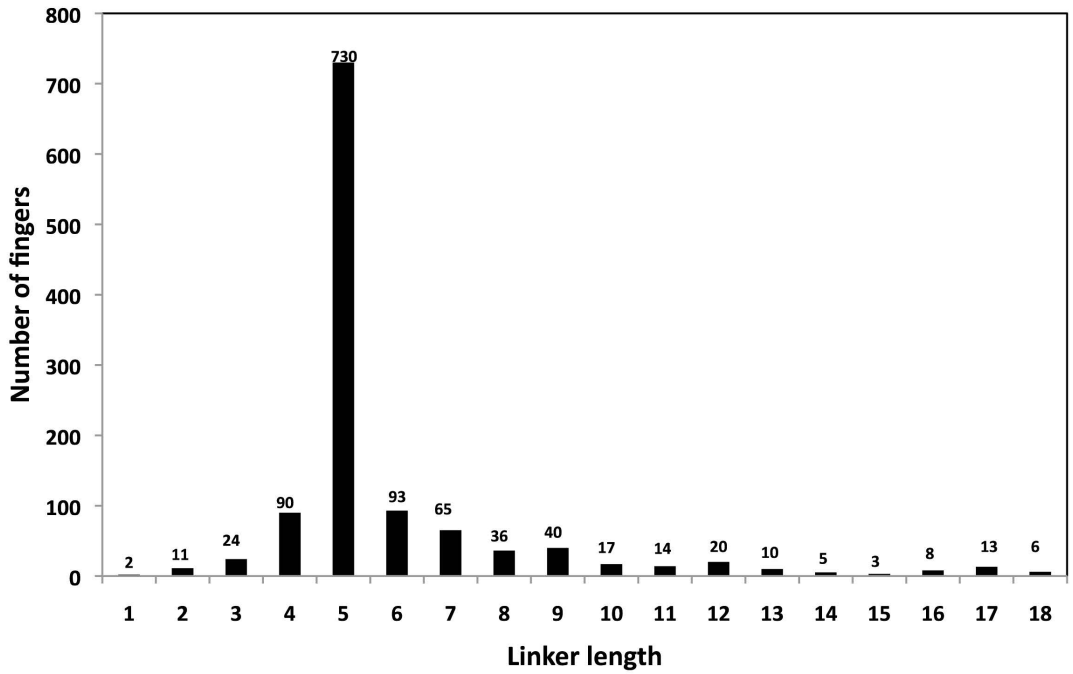
Supplemental Table 6. ZFA amino acid sequences and ZFN target sites.

Supplemental Table 7. Ratios of ZFNs injected into zebrafish embryos relative to robustness in B1H selection system.

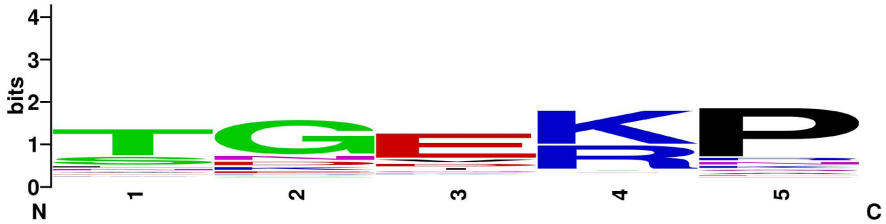
SUPPLEMENTAL FIGURES:

Supplemental Figure 1

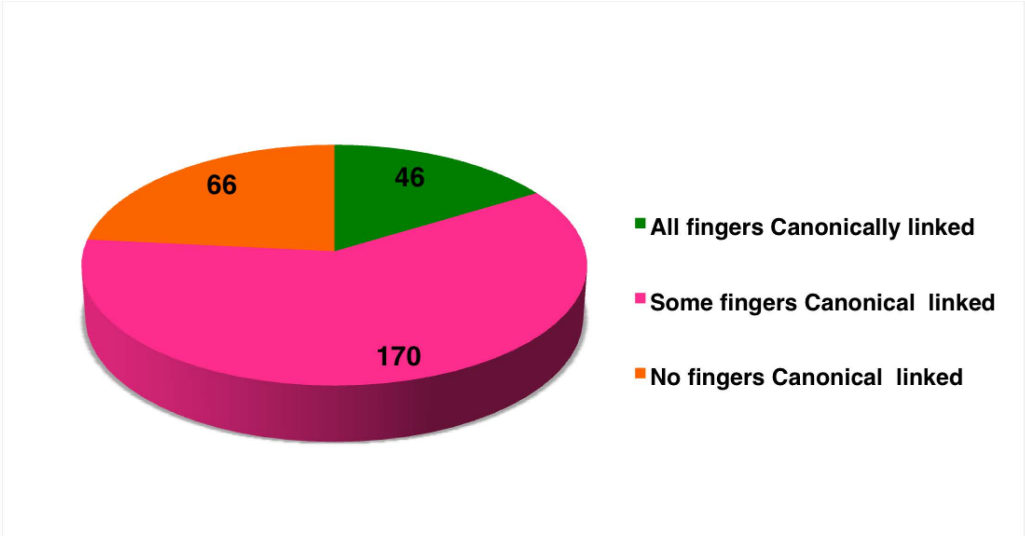
A



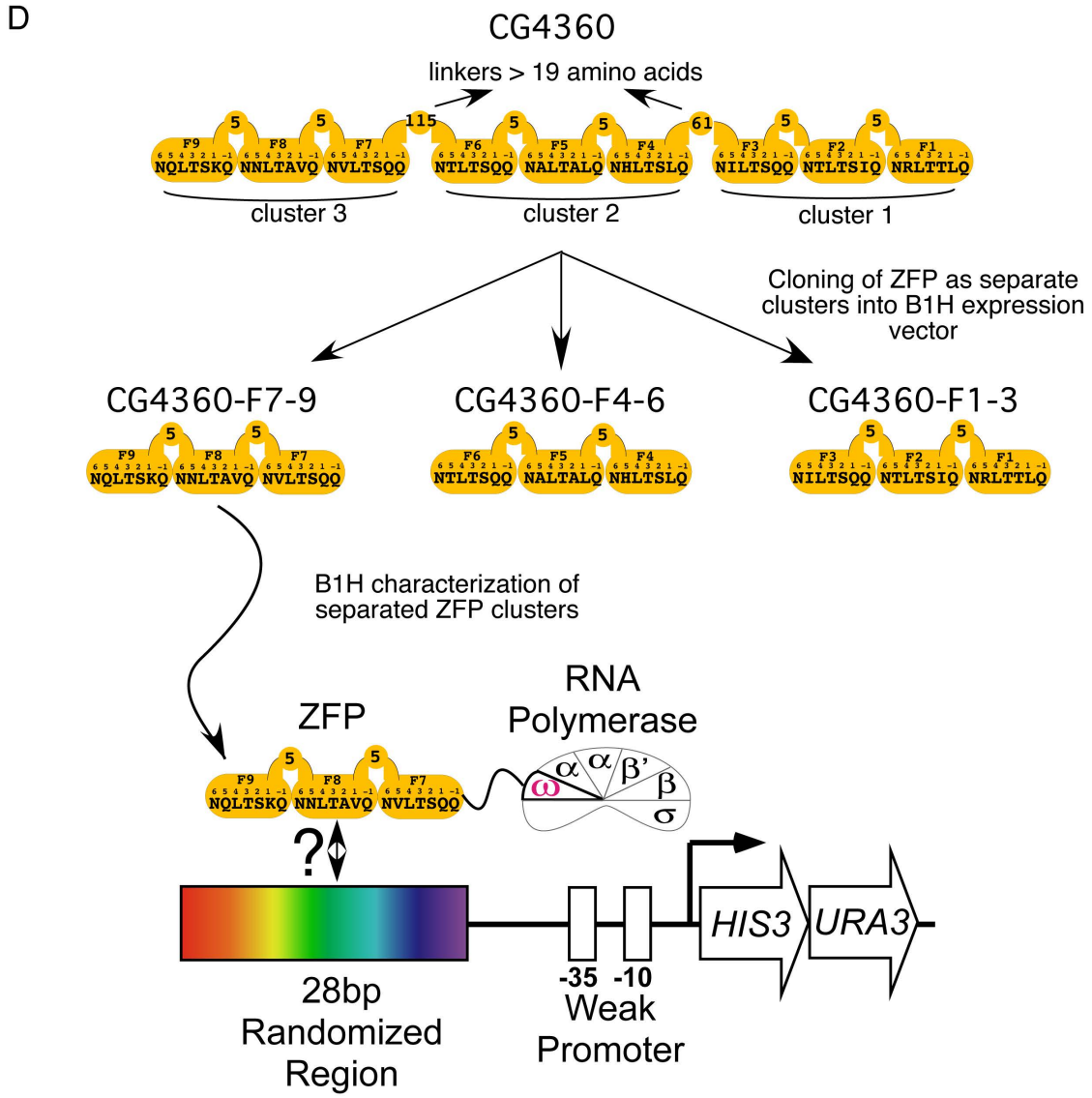
B



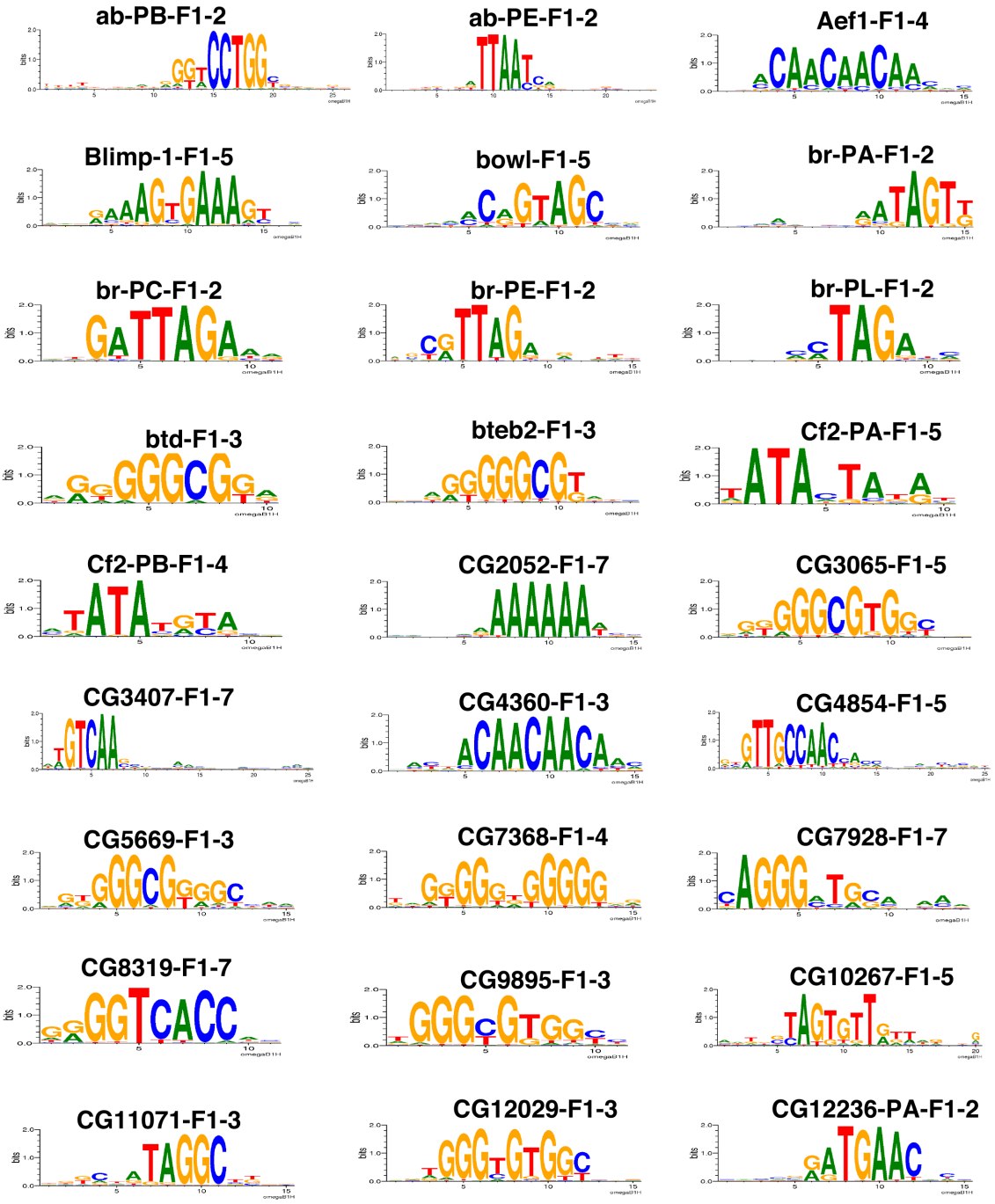
C

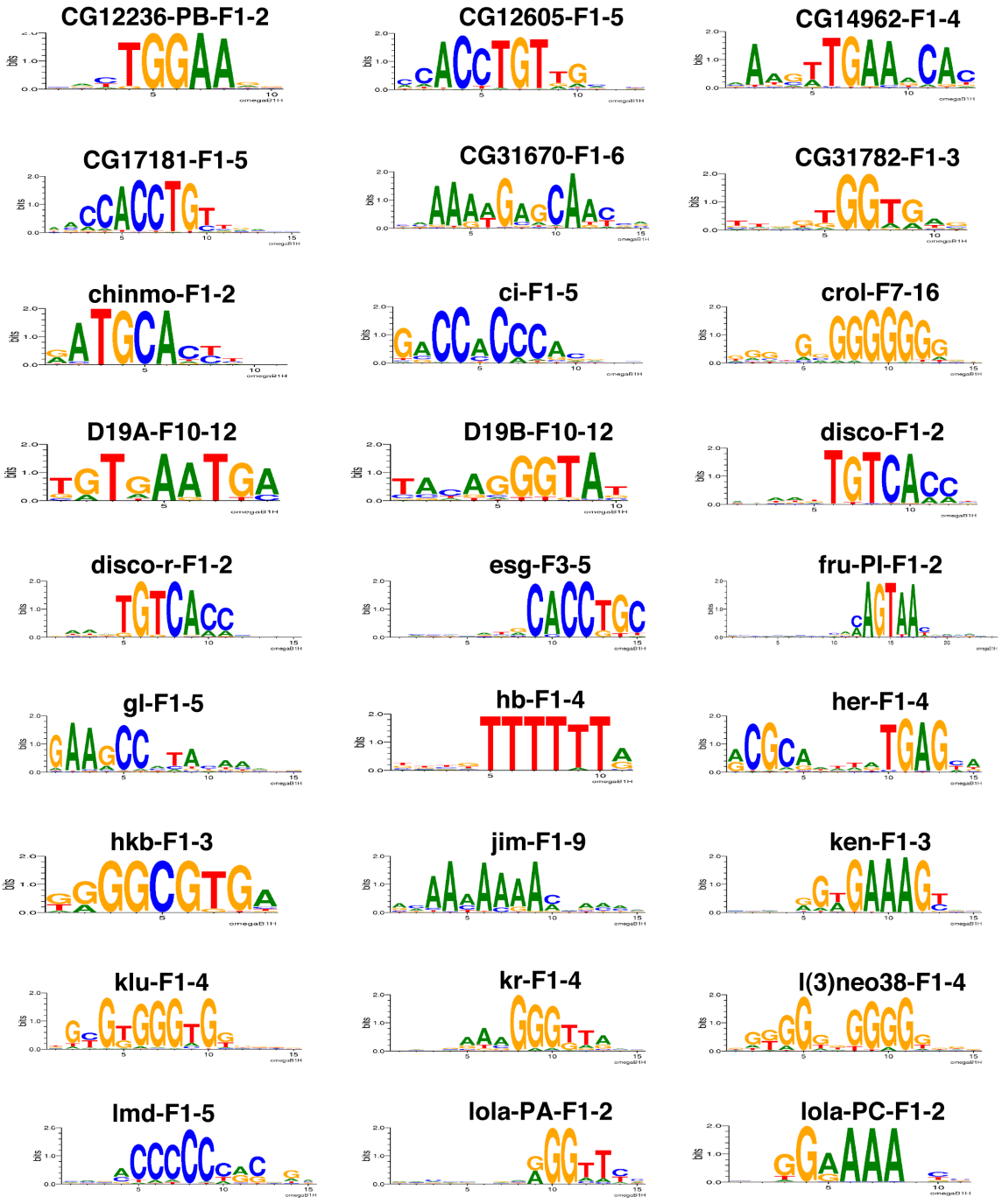


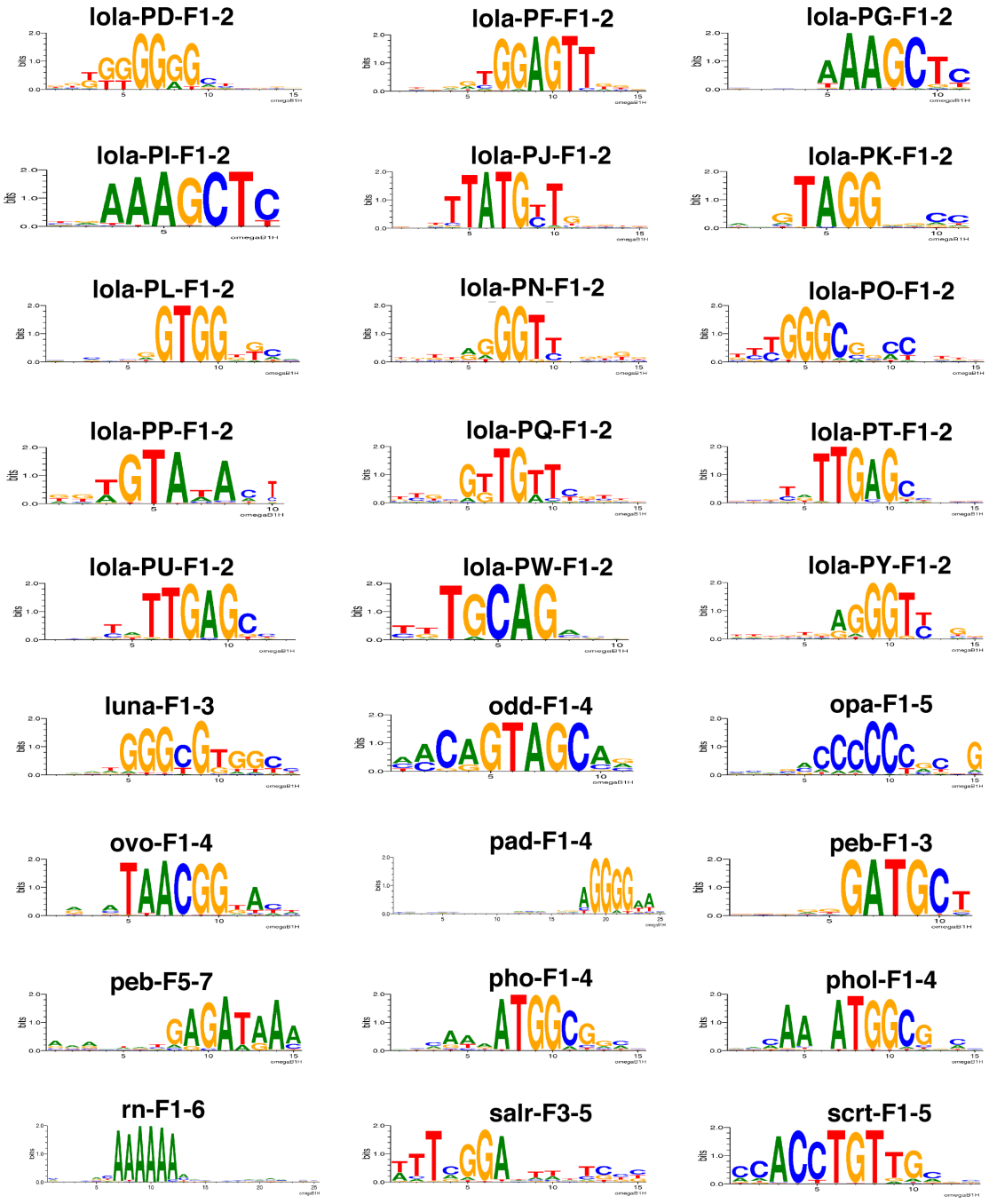
D

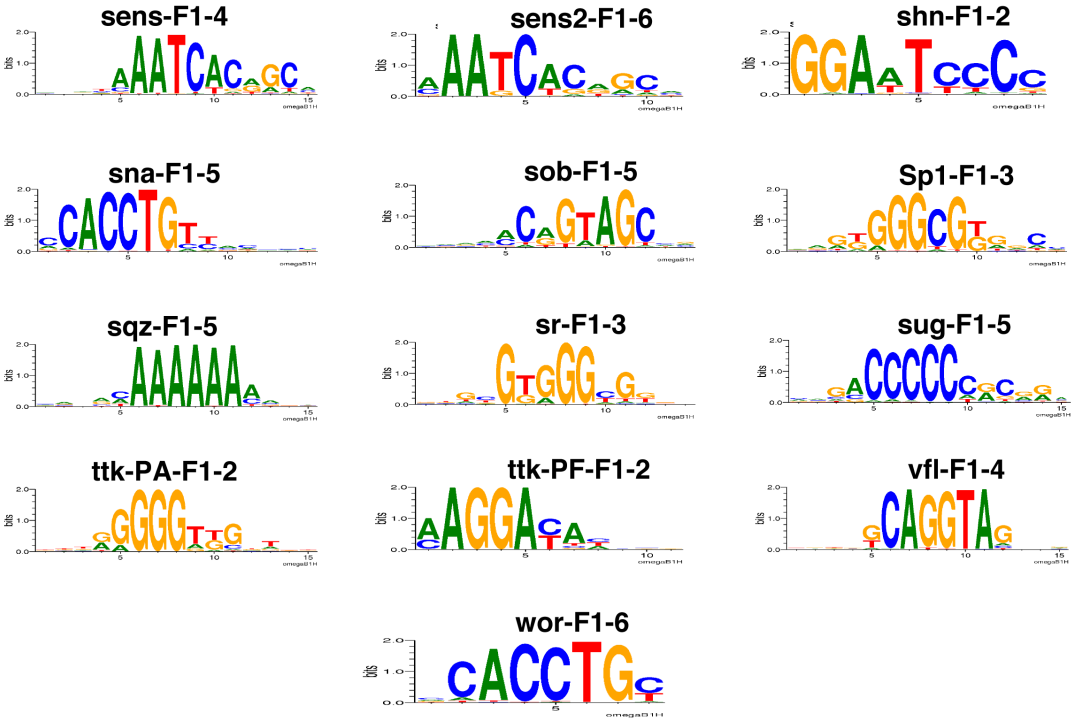

CG4360
linkers > 19 amino acids
115
61
F9 F8 F7
NQLTSKQ NNLTAVQ NVLTSQQ
F6 F5 F4
NTLTSQQ NALTALQ NHLTSLQ
F3 F2 F1
NILTSQQ NTLTSIQ NRLTTLQ
cluster 3
cluster 2
cluster 1
Cloning of ZFP as separate clusters into B1H expression vector
CG4360-F7-9
CG4360-F4-6
CG4360-F1-3
B1H characterization of separated ZFP clusters
ZFP
RNA Polymerase
α α β' ω β σ
?
28bp Randomized Region
-35 -10
Weak Promoter
HIS3
URA3

Supplemental Figure 2

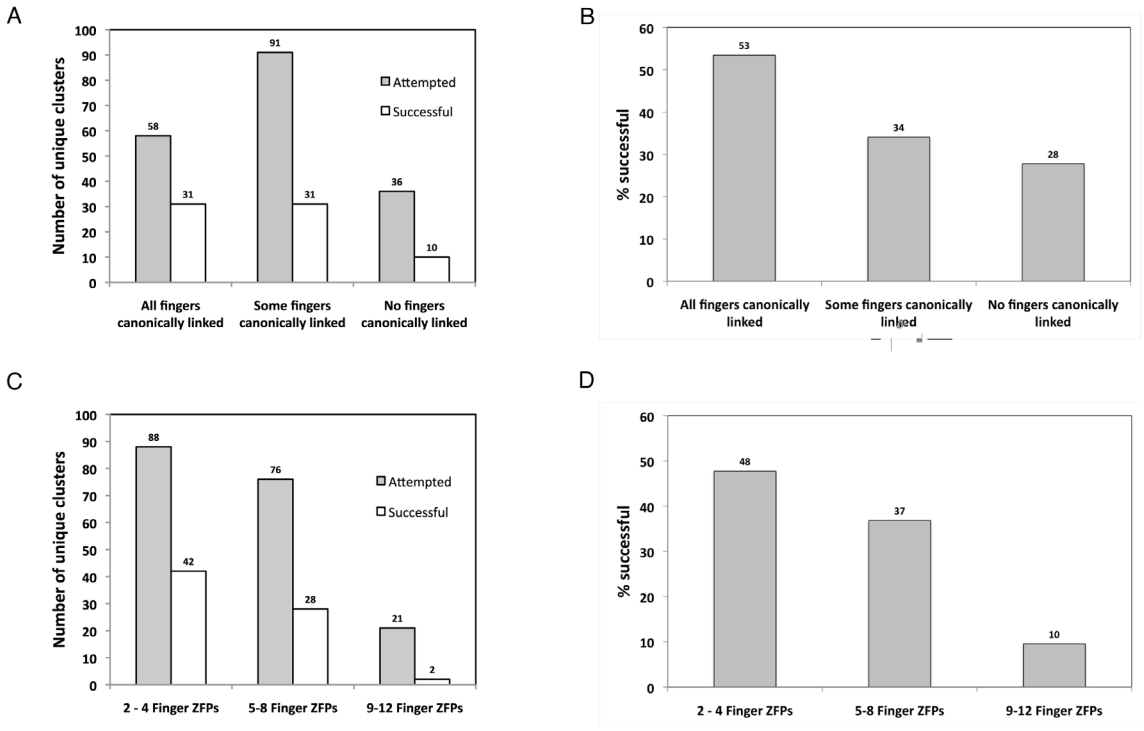



CG12236-PB-F1-2
CG12605-F1-5
CG14962-F1-4
CG17181-F1-5
CG31670-F1-6
CG31782-F1-3
chinmo-F1-2
ci-F1-5
crol-F7-16
D19A-F10-12
D19B-F10-12
disco-F1-2
disco-r-F1-2
esg-F3-5
fru-PI-F1-2
gl-F1-5
hb-F1-4
her-F1-4
hkb-F1-3
jim-F1-9
ken-F1-3
klu-F1-4
kr-F1-4
l(3)neo38-F1-4
lmd-F1-5
lola-PA-F1-2
lola-PC-F1-2

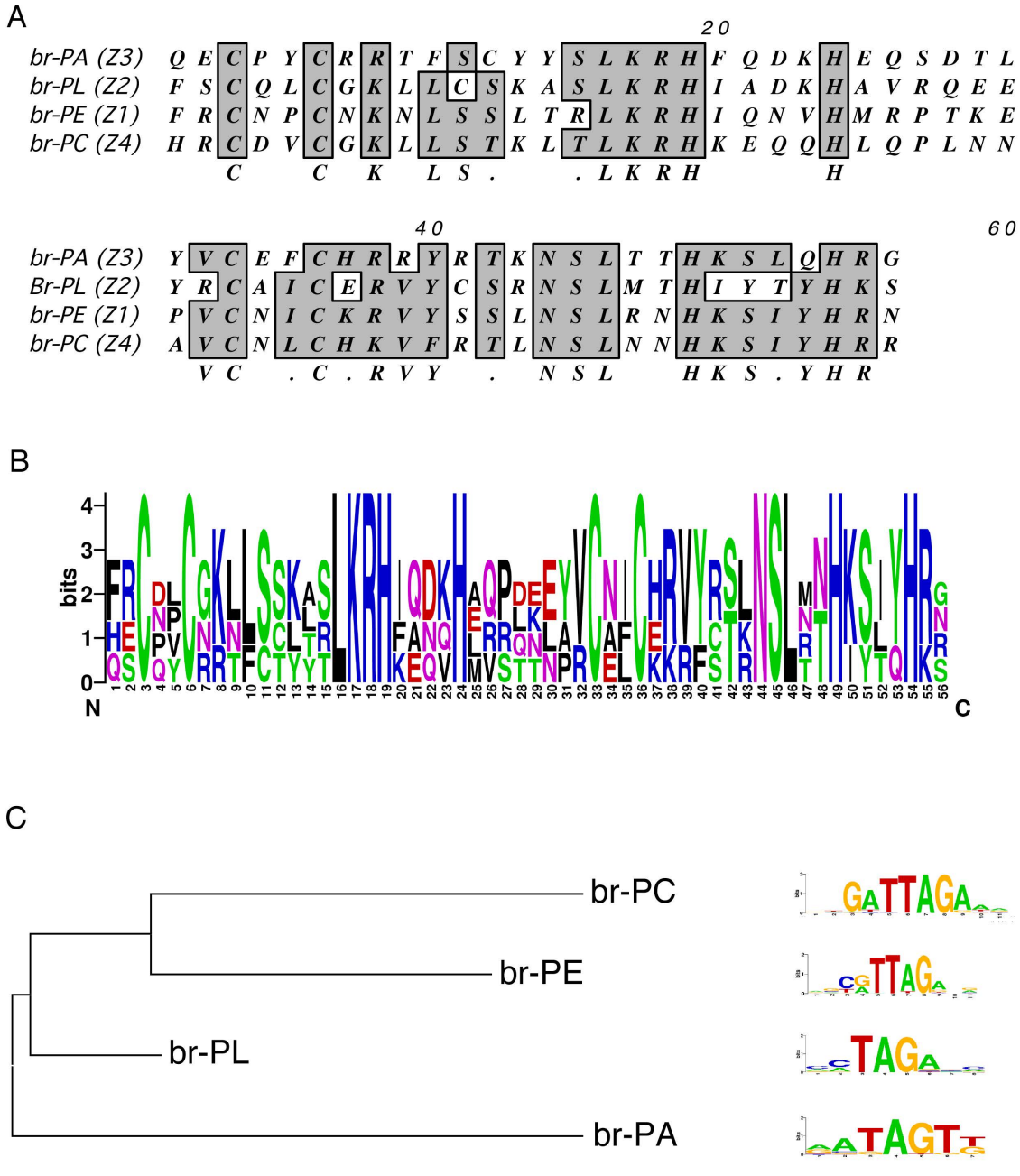

lola-PD-F1-2
lola-PF-F1-2
lola-PG-F1-2
lola-PI-F1-2
lola-PJ-F1-2
lola-PK-F1-2
lola-PL-F1-2
lola-PN-F1-2
lola-PO-F1-2
lola-PP-F1-2
lola-PQ-F1-2
lola-PT-F1-2
lola-PU-F1-2
lola-PW-F1-2
lola-PY-F1-2
luna-F1-3
odd-F1-4
opa-F1-5
ovo-F1-4
pad-F1-4
peb-F1-3
peb-F5-7
pho-F1-4
phol-F1-4
rn-F1-6
salr-F3-5
scrt-F1-5


sens-F1-4
sens2-F1-6
shn-F1-2
sna-F1-5
sob-F1-5
Sp1-F1-3
sqz-F1-5
sr-F1-3
sug-F1-5
ttk-PA-F1-2
ttk-PF-F1-2
vfl-F1-4
wor-F1-6

Supplemental Figure 3. Linkage of Fingers in each B1H Characterized Zinc Finger Cluster.

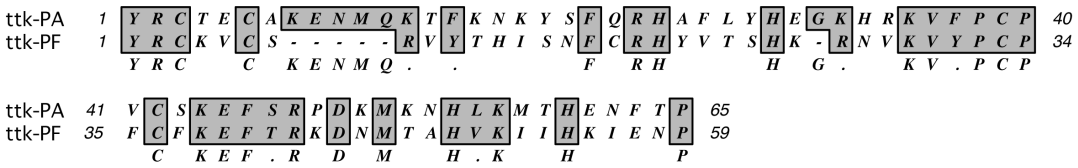


Supplemental Figure 4

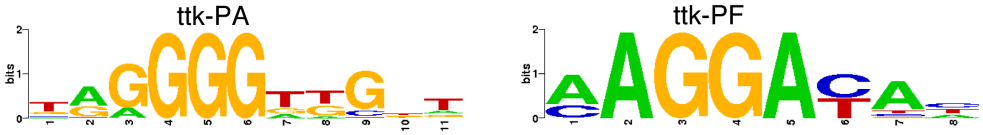


Supplemental Figure 5

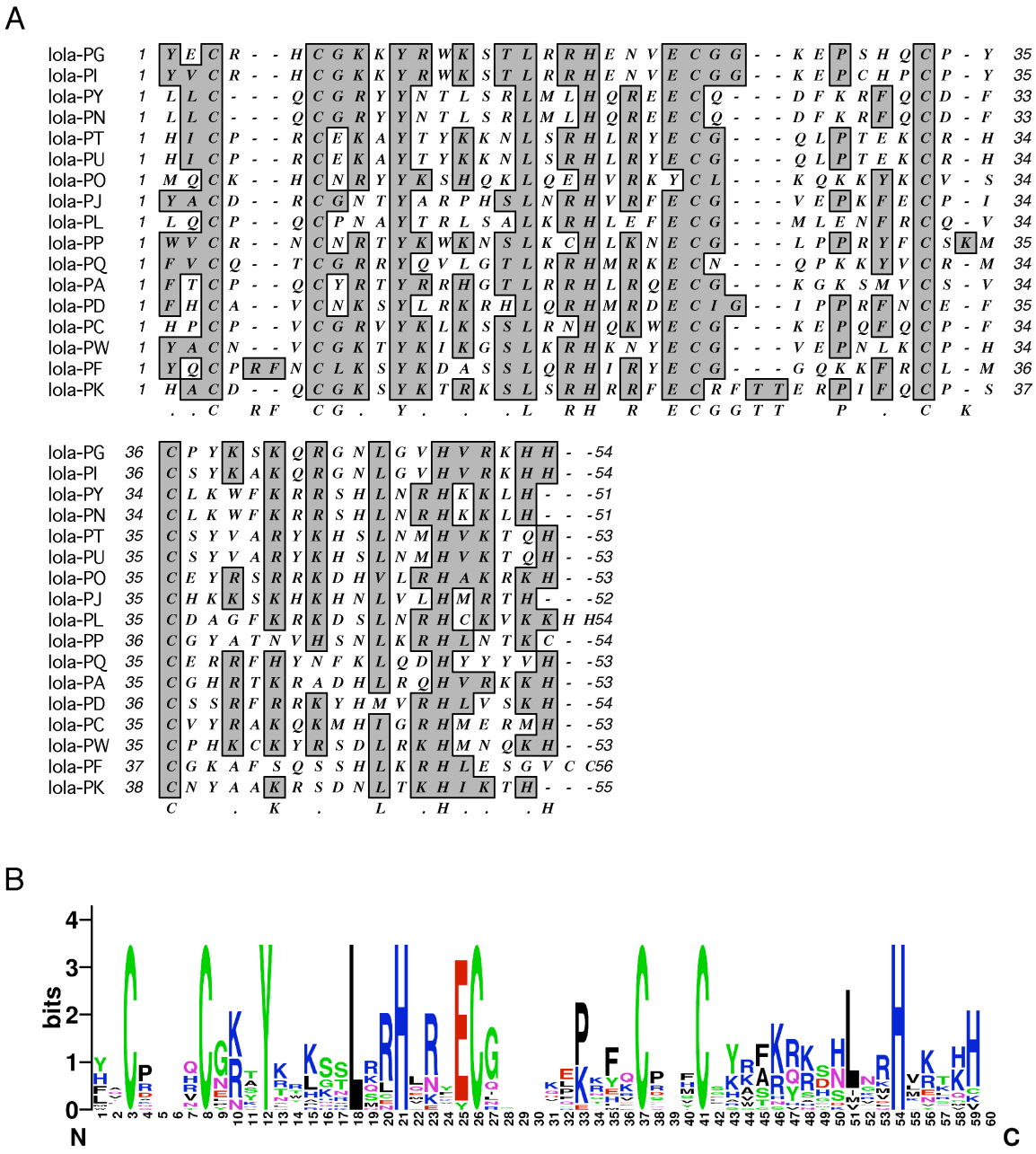
A



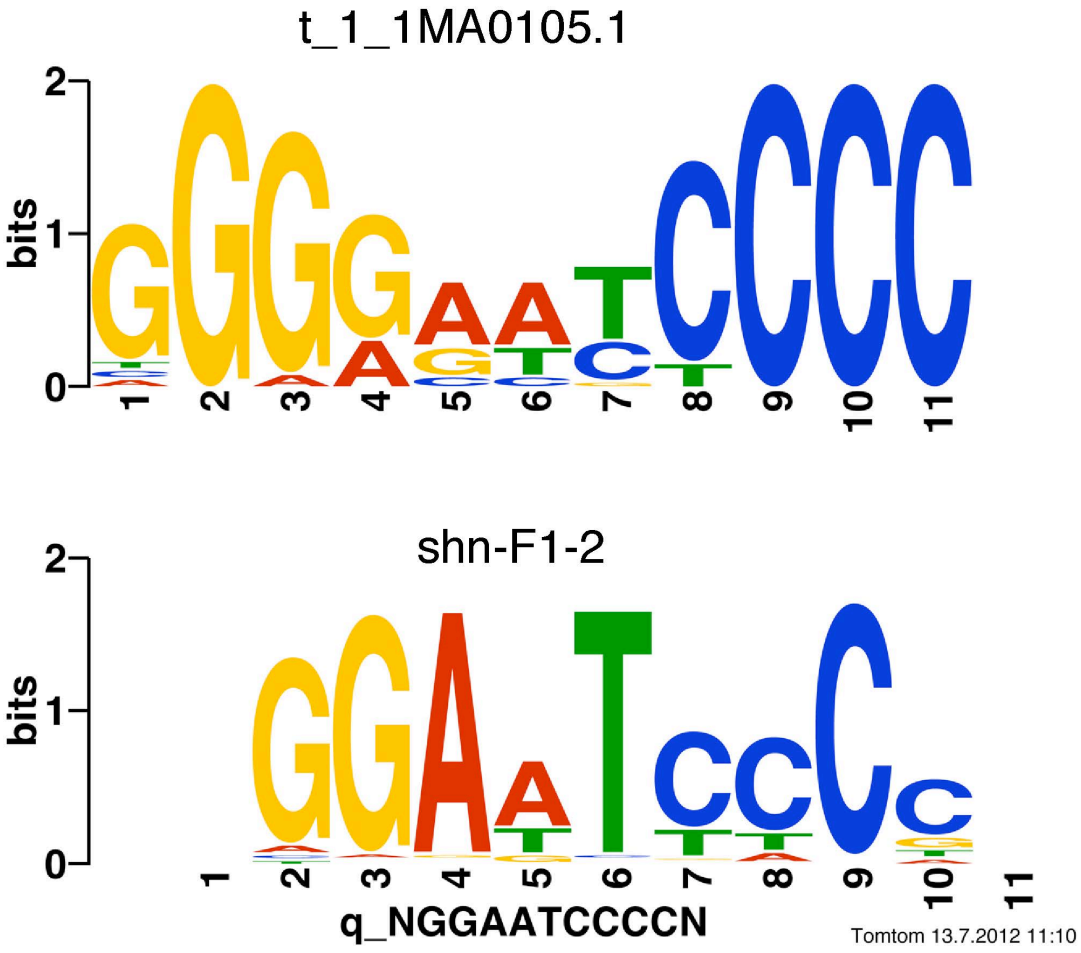
B



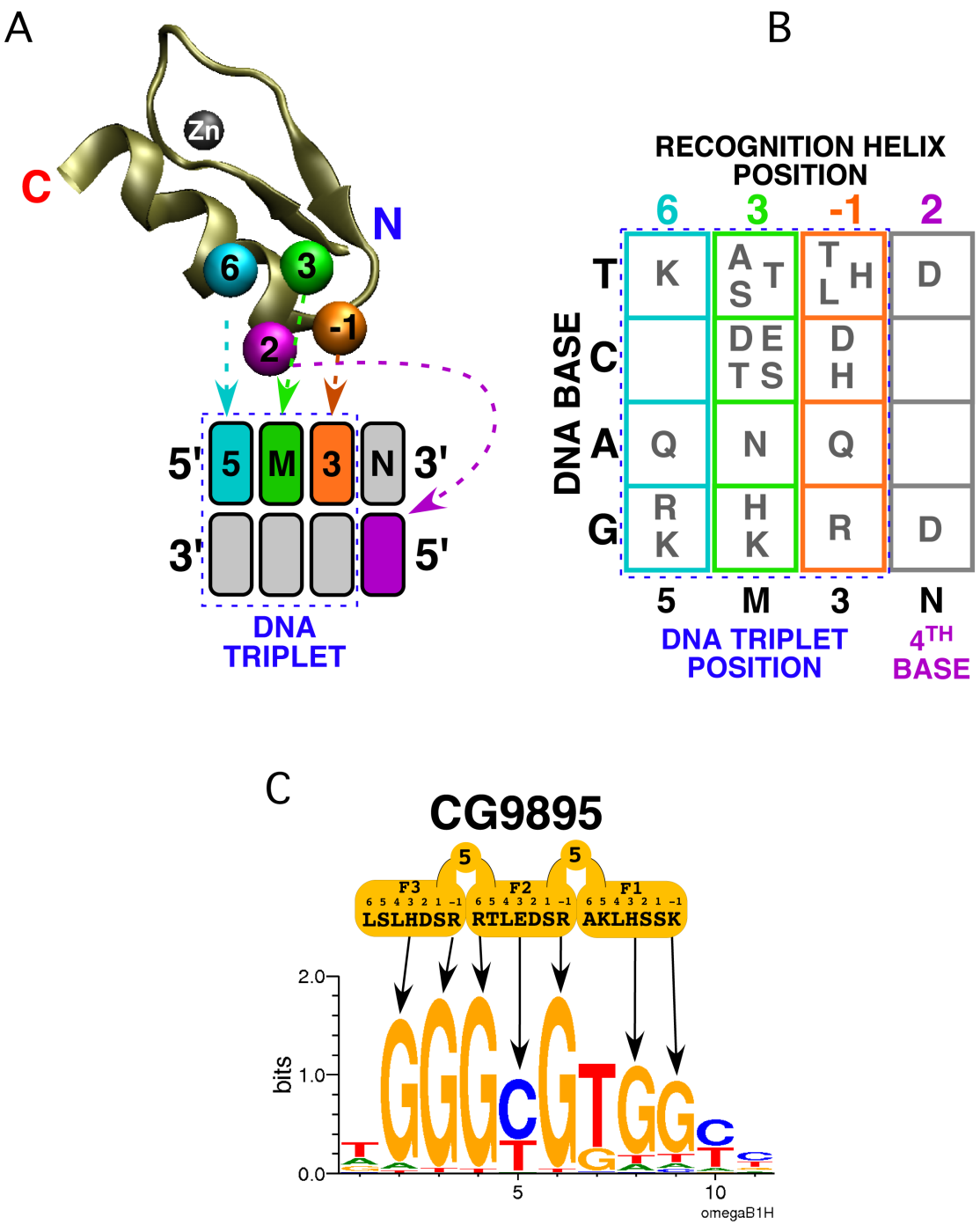
Supplemental Figure 6



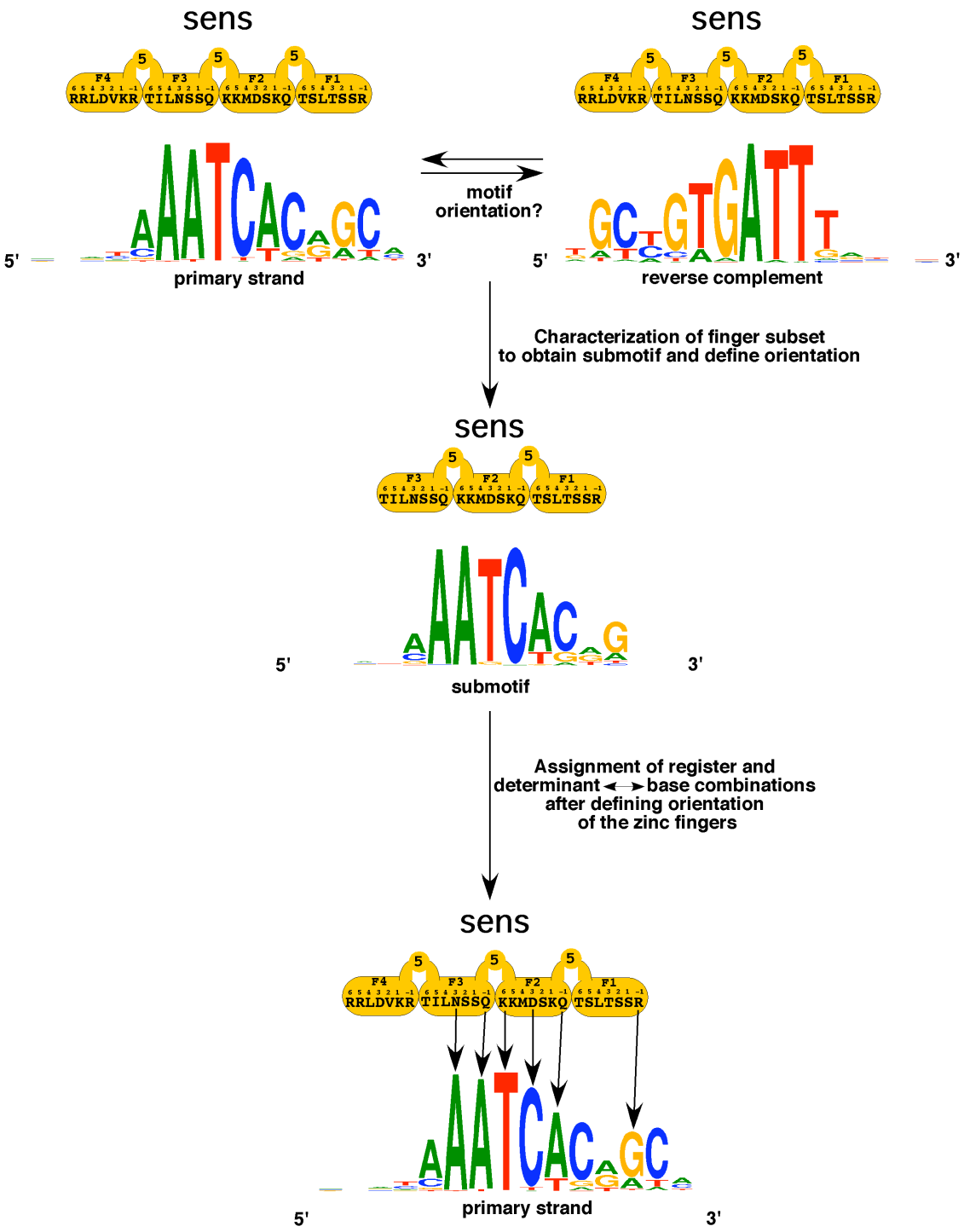
Supplemental Figure 7



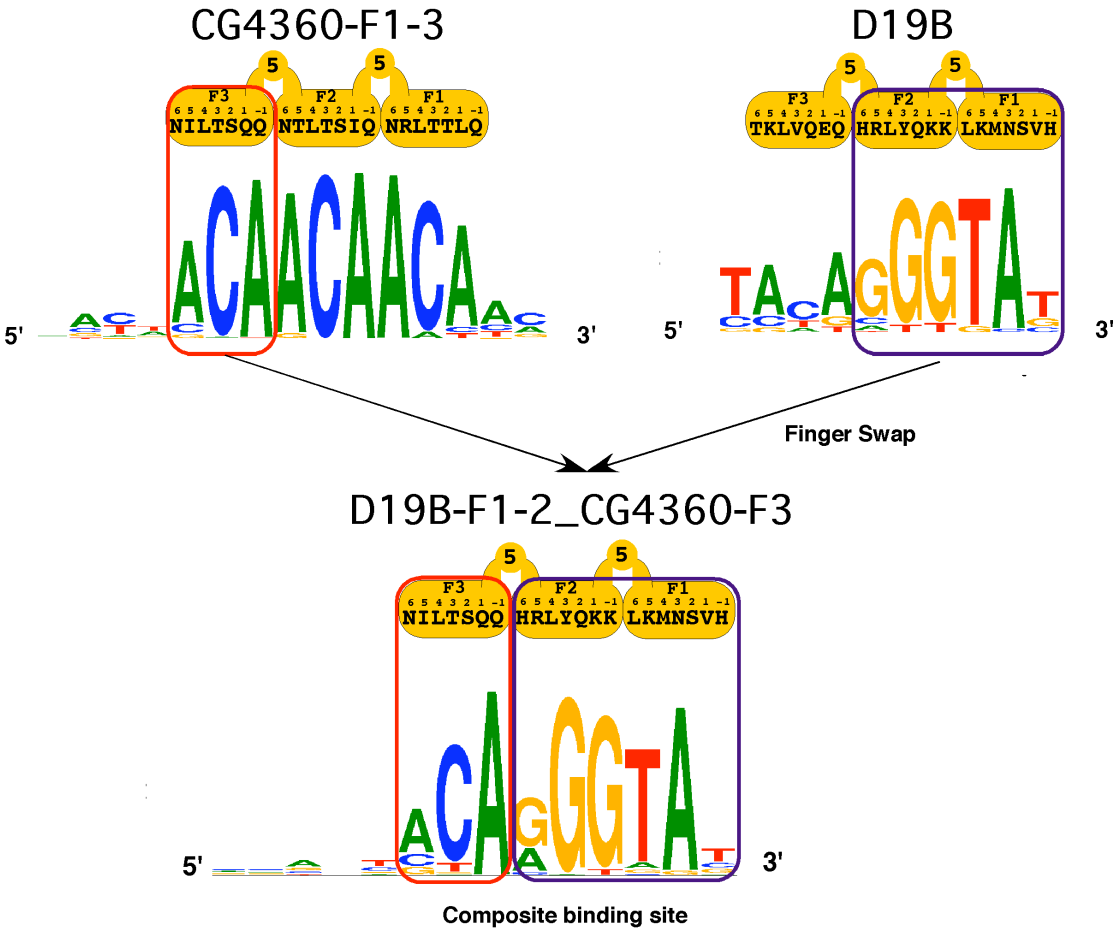
Supplemental Figure 8



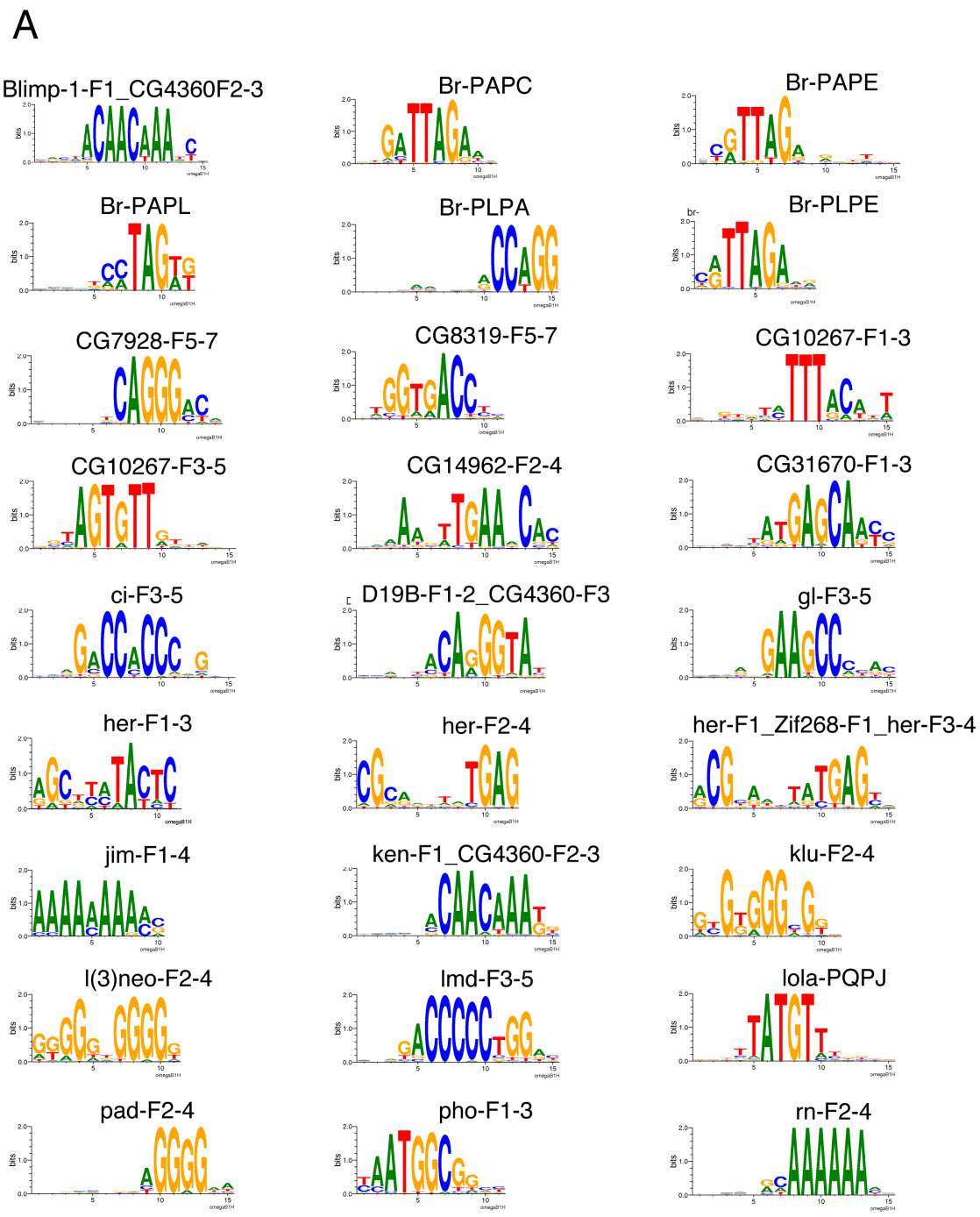
Supplemental Figure 9

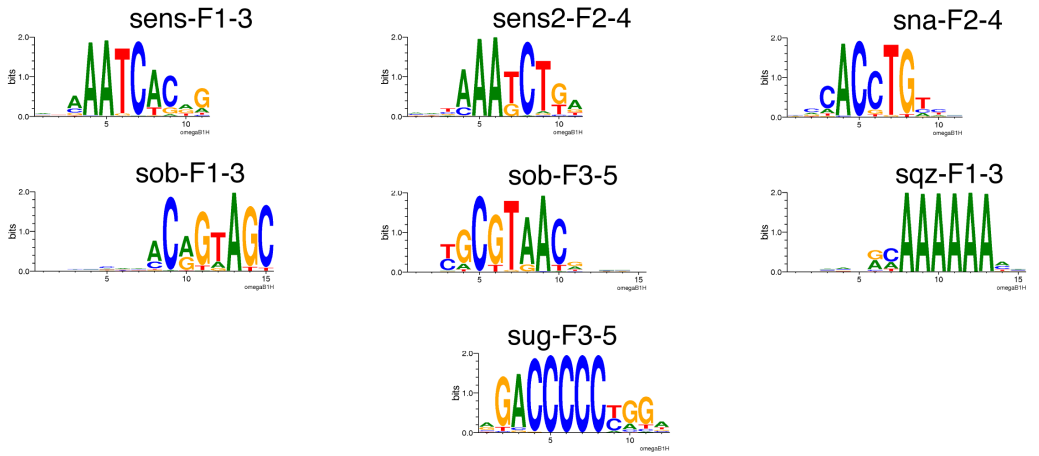


Supplemental Figure 10

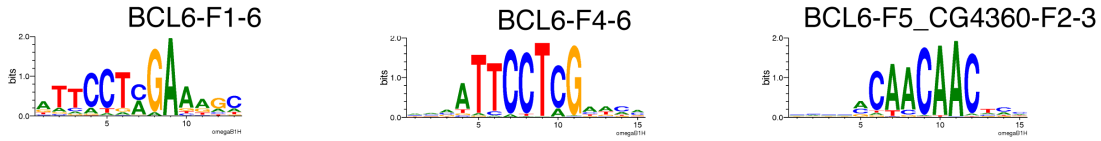


Supplemental Figure 11



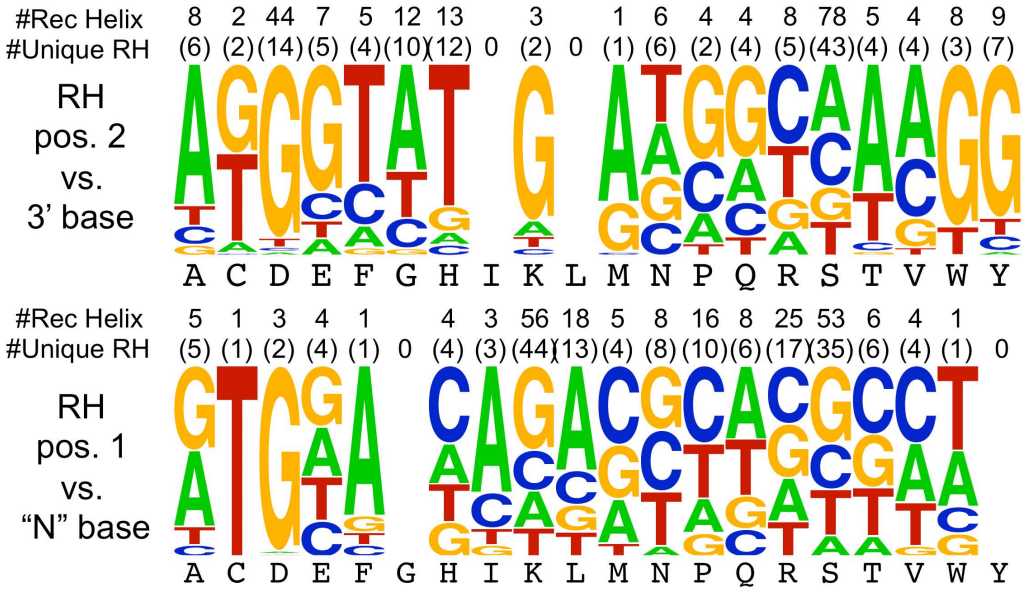

sens-F1-3
sens2-F2-4
sna-F2-4
sob-F1-3
sob-F3-5
sqz-F1-3
sug-F3-5

B

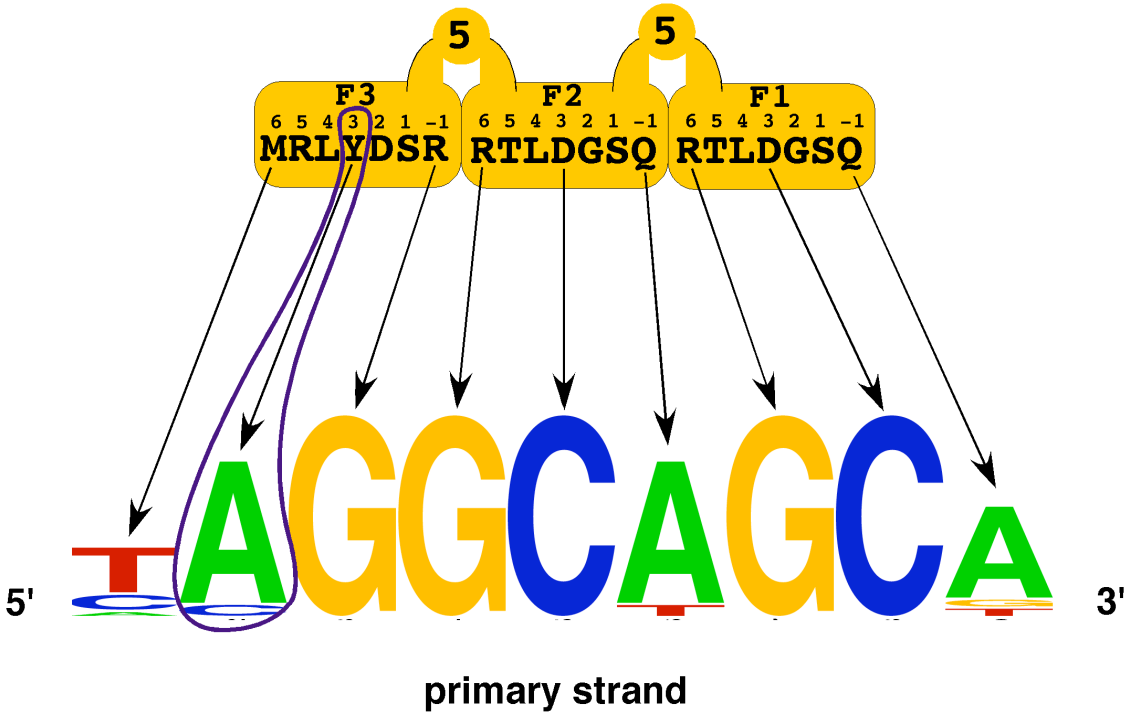

BCL6-F1-6
BCL6-F4-6
BCL6-F5_CG4360-F2-3

Supplemental Figure 12

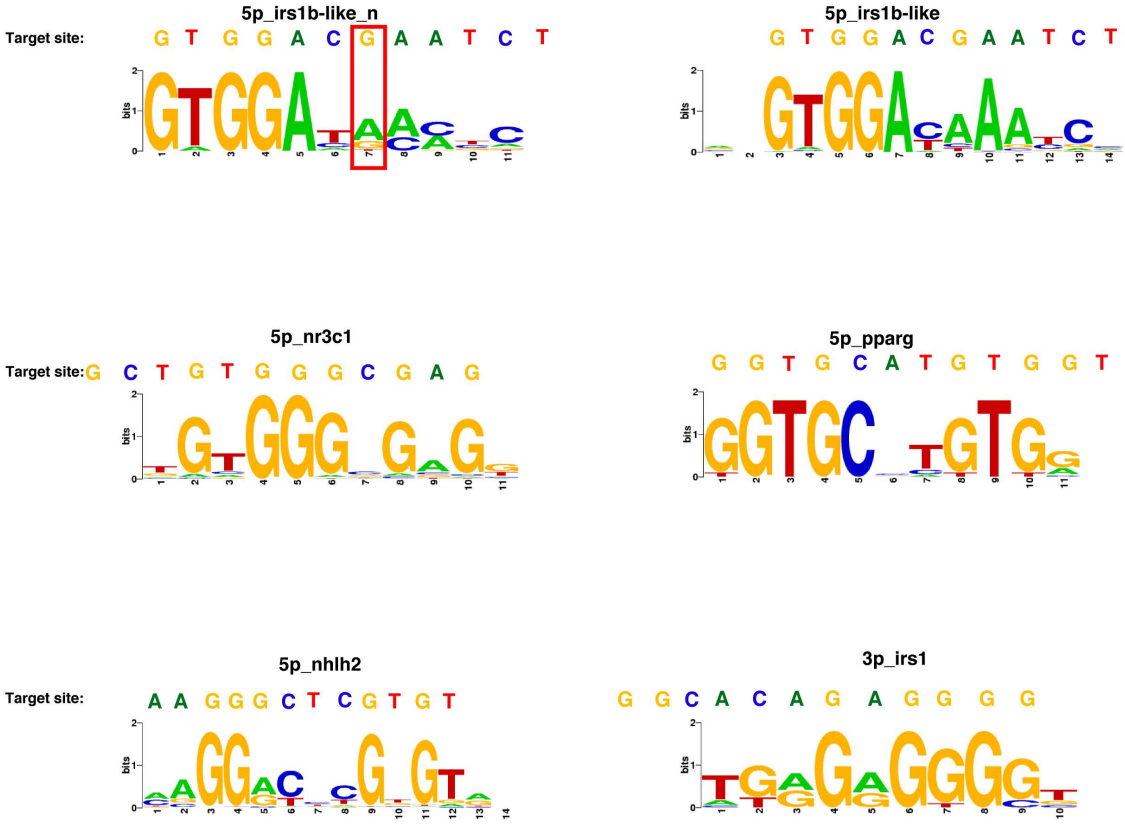
Supplemental Figure 13



Supplemental Figure 14

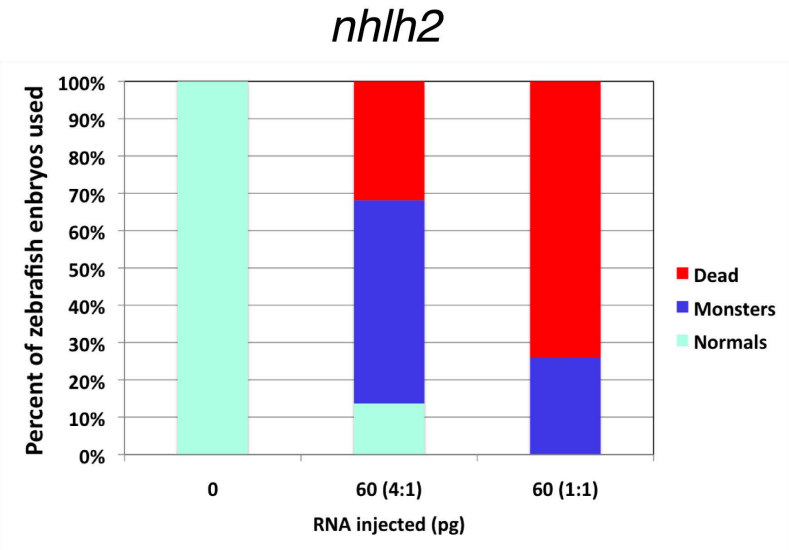


Supplemental Figure 15

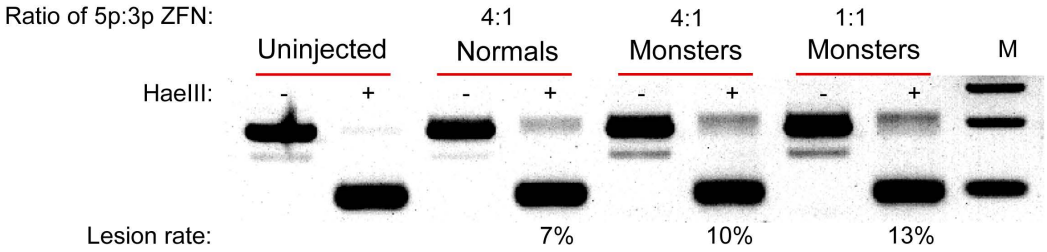


Supplemental Figure 16

A



B

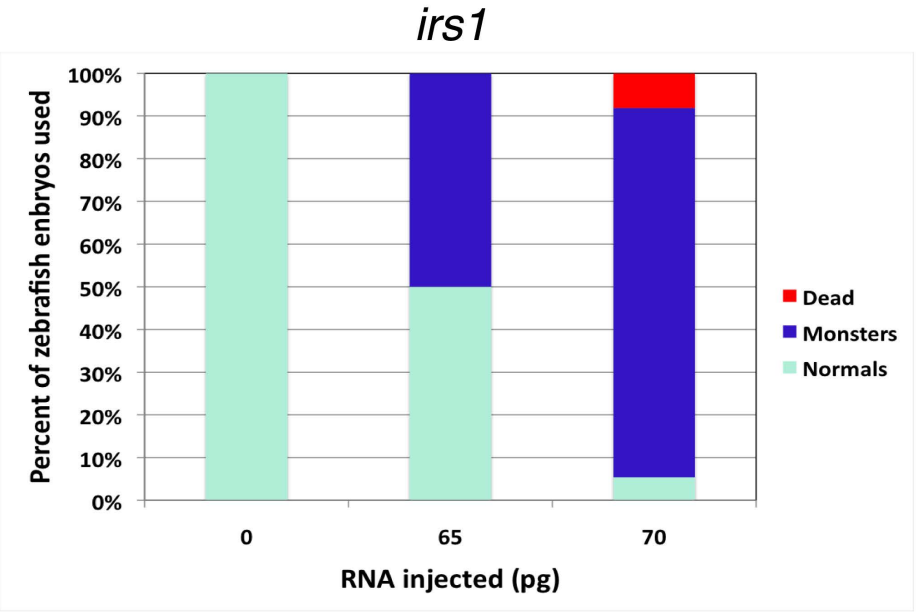


C

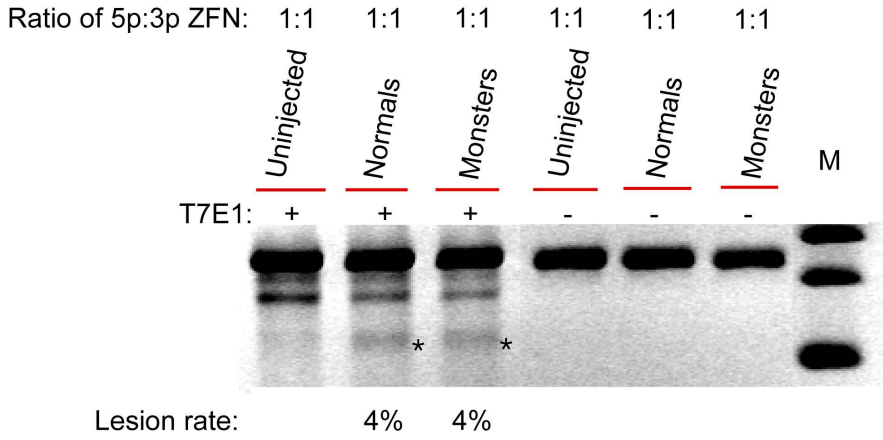
```
gtgtggcACACGAGCCCTTagaggccGACGATGGCAAGa - Wild-type target sequence
gtgtggcACACG---------------ACGATGGCAAGa - 15bp deletion
gtgtggcAC-----------------ACGATGGCAAGa - 18bp deletion
gtgtggcACACGAGCCC-------ATGACGATGGCAAGa -  9bp deletion & insertion
gtgtggcACACGAGCCCTTAGAGGACACTTGTCCCGACGATGGCAAGa - 12bp insertion
```

Supplemental Figure 17

A



B

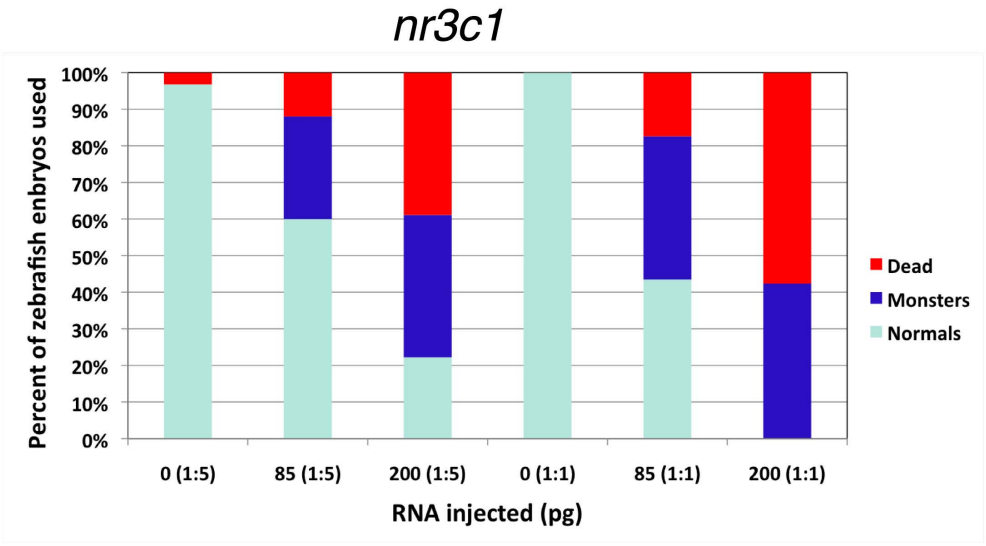


C

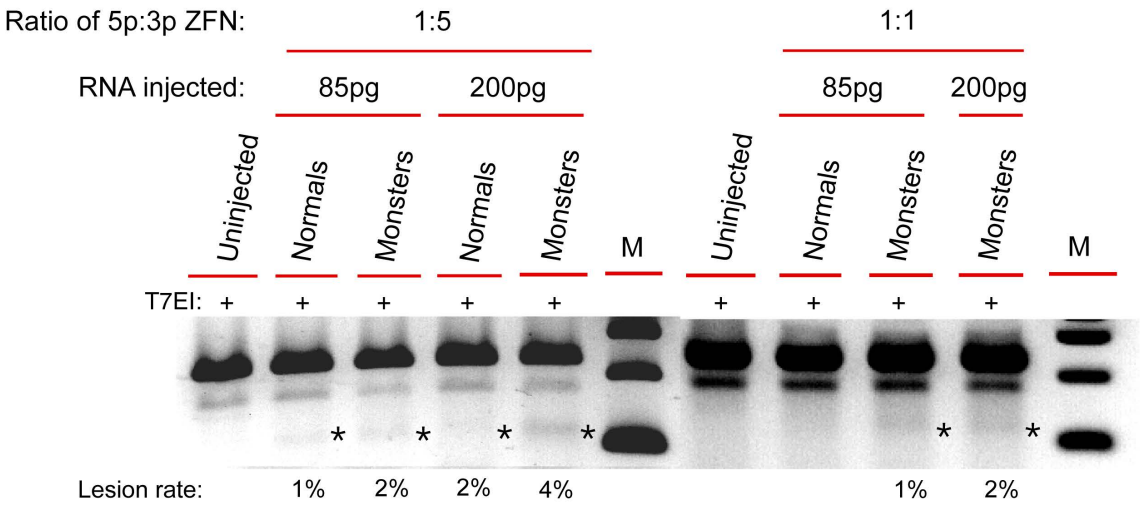
```
gcagcCCgagcaatgCCCGACCTCAATcgcaacGGCACAGAGGGGgatccaga - Wild-type target sequence
gcagcCCgagcaatgCCCGACCTCAATcg-----GCACAGAGGGGgatccaga -  5bp deletion
gcagcCCgagcaatgCCCGACCTCAAT--------CACAGAGGGGgatccaga -  8bp deletion
gcagccc-----------------------------------gatccaga - 38bp deletion
gcagcCCgagcaatgCCCG-CGGTCCTACACACTCTATGCCTAC-gatccaga - 26bp deletion
                                                        & 24bp insertion
```

Supplemental Figure 18

A



B

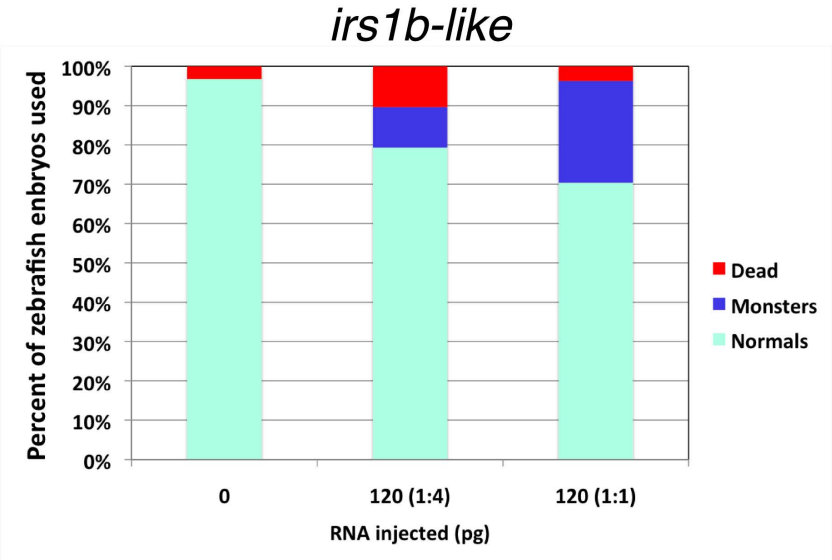


C

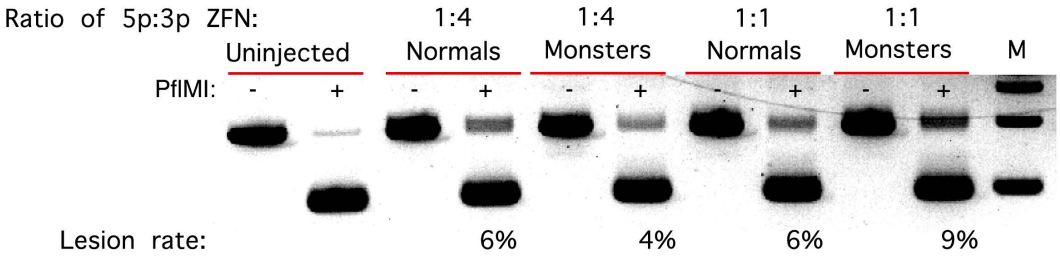
```
gCTCGCCCACAGCtgtcgtCGGTCTGCAGAAa - Wild-type target sequence
gCTCGCCCACAG------cgtCGGTCTGCAGAAa -  6bp deletion
gCTCGCCCACAGCtg----------CAGAAa - 13bp deletion
gCTCGCTCAC----------TCGGTCTGCAGAAa - 10bp deletion
gCTCGCCCACAG----------CTGCAGAAa - 13bp deletion
gCTCGCCCACAGctgtcgtc--GGTCTGCAGAAa -  2bp deletion
gCTCGCCCACAGAAAGATCAGAAGTCTGCAGAAa -  9bp deletion & 11bp insertion
gCTCGCCCACAGCtgCAGAAAGAtcgtCGGTCTGCAGAAa - 8bp insertion
gCTCGCCCACAGCtgATCTTTCTGATCTTTCGGTCTGCAGAAa - 4bp deletion & 15bp insertion
```

Supplemental Figure 19

A



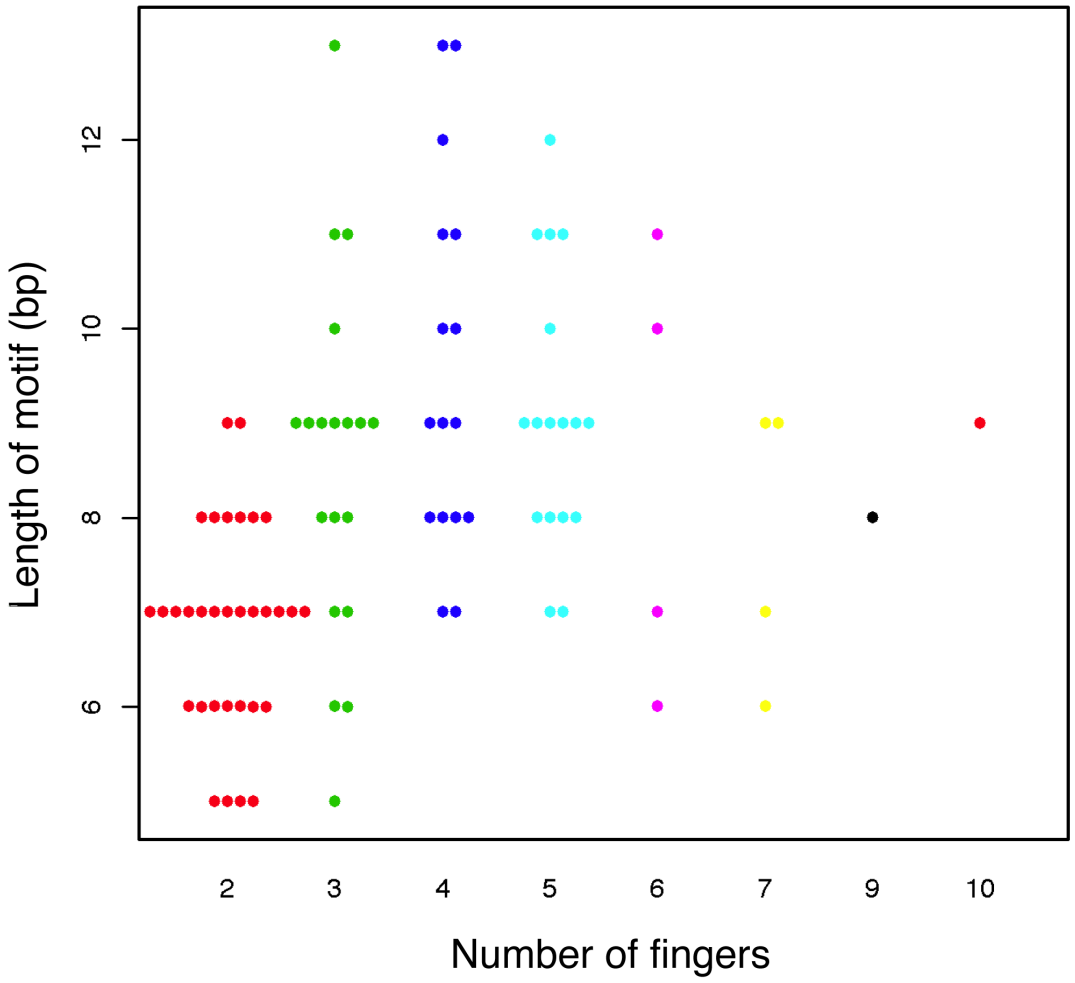
B



C

```
tccacggcAGATTCGTCCACcttctGGCACAGATGGCtacatgat - Wild-type target sequence
tccacggcAGATTCGTCCACcGAATGAATATGATG-----atgat - 20bp deletion & insertion
tccacggcAGATTCGTCCAC----------AGATGGCtacatgat - 10bp deletion
tccacggcAGATTCGTCCAC----------------tacatgat - 17bp deletion
tccacggcAGATTCGTCC---ttctGGCACAGATGGCtacatgat -  3bp deletion
tccacggcAGATTC----------tGGCACAGATGGCtacatgat - 10bp deletion
tc--------------------tctGGCACAGATGGCtacatgat - 20bp deletion
tccacggcAGAGTCGTCCAC----------AGATGGCtacatgat - 10bp deletion
tccacggcAGATTCGTCCACGGC----------------atgat - 17bp deletion & insertion
tccacggcAGATTCGTCCACc---tGGCACAGATGGCtacatgat -  3bp deletion
```

Supplemental Figure 20



Supplementary Table 1.

Please see attached.

Supplementary Table 2. Successfully characterized Cys_2-His_2 ZFPs from Drosophila melanogaster.

Please see attached.

Supplementary Table 3. Alternately spliced genes that display isoform-dependent changes in zinc finger composition or number.

gene name	isoforms with different recognition potential	isoform differences
Lola	15	different composite finger sets
Br	4	different composite finger sets
Crol	4	different composite finger sets
Fru	3	different composite finger sets
CG12236	2	different composite finger sets
Ttk	2	different composite finger sets
ab	2	different composite finger sets
Cf2	2	one additional internal finger in array
CG17829	2	one additional internal finger in array
CG9817	2	one additional internal finger in array
CG12054	2	additional N-terminal finger
Rgr	2	additional N-terminal finger
Rn	2	additional N-terminal finger
CG1529	2	two additional N-terminal fingers
zfh1	2	one additional N-terminal fingers
CG10274	2	three additional N-terminal fingers
MTF-1	2	four additional N-terminal fingers
CG6813	2	additional C-terminal finger
Gl	2	additional C-terminal finger
Hang	2	additional C-terminal finger
CG12071	2	two additional C-terminal fingers
CG6791	2	two additional C-terminal fingers
mid	2	two additional C-terminal fingers
CG2678	2	three additional C-terminal fingers
CG4360	2	three additional C-terminal fingers
CG14667	2	three additional C-terminal fingers
CG11456	2	four additional C-terminal fingers
CG31388	2	seven additional C-terminal fingers

Supplementary Table 4. Single finger – DNA subsite combinations derived from characterized 83 Cys_2-His_2 ZFPs from *D. melanogaster* .

Please see attached.

Supplementary Table 5. Recognition helices for all *Drosophila* single finger – DNA subsite combinations represented in Figure 4.

Please see attached.

Supplementary Table 6. ZFA amino acid sequences and ZFN target sites.

Fingers	ZFA Sequence	Target Sequence	ZFN Target Site
	5p_cpe		
	GTKP	GGACAAAAGGAA	TTCCTTTTGTCCtgtcagcCAATCTGCATGG
F1	YKCPECGKSFS**QKCNLVR**HQRTHTGEKP		
F2	YACDICKKRFS**STSNLKT**HLRLHSGQKP		
F3	YACDLCPQKFT**QFVHLKL**HKRLHTGEKP		
F4	FACDICGRKFA**QRGHLTR**HTKIHLRGS		
	3p_cpe		
	GTKP	CAATCTGCATGG	TTCCTTTTGTCCtgtcagcCAATCTGCATGG
F1	YKCPE--CGKSFS**RSDHLTT**HQRTHTGEKP		
F2	YACPVESCDRRFS**QSGDLTR**HIRIHTGQKP		
F3	FQCQY--CGKRFH**QKSDMKK**HTYIHTGEKP		
F4	HKCTV--CLKAFS**QSSNLIT**HMRKHLRGS		
	5p_irs1b-like		
	GTKP	GTGGACGAATCT	AGATTCGTCCACcttctGGCACAGATGGC
F1	YKCNYCGKRFH**QKSDMKK**HTYIHTGEKP		
F2	HKCQVCGKAFS**QSSNLII**HSRKHTGQKP		
F3	FQCRICMRNFS**LKGNLTR**HIRTHTGEKP		
F4	FACDICGRKFA**RSDALTR**HTKIHLRGS		
	3p_irs1b-like		
	GTKP	GGCACAGATGGC	AGATTCGTCCACcttctGGCACAGATGGC
F1	YKCNV--CGKSFV**ESSKLKR**HQLVHTGEKP		
F2	FECTFEGCGKRFS**LDFNLRT**HVRIHTGQKP		
F3	FQCPVESCDRRFS**QRGTLAE**HIRIHTGQKP		
F4	FQCRI--CMKAFS**CARNLTR**HIRTHLRGS		
	5p_nr3c1		
	GTKP	GCTGTGGGCGAG	CTCGCCCACAGCtgtcgtCGGTCTGCAGAA
F1	YKCPE--CGKSFS**RSDNLTR**HQRTHTGEKP		
F2	YACPVESCDRRFS**EKSHLTR**HIRIHTGQKP		
F3	FQCRI--CMRNFS**RSDALTR**HIRTHTGEKP		
F4	FACDI--CGRKFA**HRQSLTR**HTKIHLRGS		
	3p_nr3c1n		
	GTKP	CGGTCTGCAGAA	CTCGCCCACAGCtgtcgtCGGTCTGCAGAA
F1	YKCPE--CGKSFS**QKCNLVR**HQRTHTGEKP		
F2	YACPVESCDRRFS**IKGSLKR**HIRIHTGQKP		
F3	FQCRI--CMRNFS**YRSDLRK**HIRTHTGEKP		
F4	FACDI--CGRKFA**RSDHLSD**HTKIHLRGS		
	5p_irs1		
	GTKP	ATTGAGGTCGGG	CCCGACCTCAATcgcaacGGCACAGAGGGG
F1	YKCPE--CGKSFS**RSDHLTR**HQRTHTGEKP		
F2	YACPVESCDRRFS**DRSALAR**HIRIHTGQKP		
F3	FQCRI--CMRNFS**RSDNLTR**HIRTHTGEKP		
F4	FACKI--CSRSFG**YKHVLQN**HERTHLRGS		
	3p_irs1		
	GTKP	GGCACAGAGGGG	CCCGACCTCAATcgcaacGGCACAGAGGGG
F1	YKCPE--CGKSFS**RSDHLTR**HQRTHTGEKP		
F2	YACPVESCDRRFS**RSDNLTR**HIRIHTGQKP		
F3	FQCPVESCDRRFS**QRGTLKQ**HIRIHTGQKP		
F4	FQCRI--CMKAFS**DKGHLTR**HIRTHLRGS		
	5p_nhlh2		
	GTKP	AAGGGCTCGTGT	ACACGAGCCCTTagaggccGACGATGGCAAG
F1	YKCPE--CGKSFS**LRHHLVG**HQRTHTGEKP		
F2	YACPVESCDRRFS**RSDTLKG**HIRIHTGQKP		
F3	FQCRI--CMKAFS**DRCHLTR**HIRTHTGEKP		
F4	FACDI--CGRKFA**RSDNLTQ**HTKIHLRGS		
	3p_nhlh2		
	GTKP	GACGATGGCAAG	ACACGAGCCCTTagaggccGACGATGGCAAG
F1	YKCPE--CGKSFS**RSDNLTQ**HQRTHTGEKP		
F2	YACNV--CGKSFV**ESSKLKR**HQLVHTGEKP		
F3	FECTFEGCGKRFS**LDFNLRT**HVRIHTGEKP		
F4	FACDI--CGRKFA**LKGNLTR**HTKIHLRGS		
	5p_pparg		
	GTKP	GGTGCATGTGGT	ACCACATGCACCcggagTCTGCAGACCTG
F1	YKCPE--CGKSFS**CAHHLTR**HQRTHTGEKP		
F2	YACPVESCDRRFS**LRHHLVG**HIRIHTGQK		
F3	PFQCGI-CMRNFS**QSGDLTR**HIRTHTGEKP		
F4	FACDI--CGRKFA**LSHHLTR**HTKIHLRGS		
	3p_ppargn		
	GTKP	TCTGCAGACCTG	ACCACATGCACCcggagTCTGCAGACCTG
F1	YKCPI--CGKAFS**RPWLLQG**HIRTHTGEKP		
F2	FQCKAPGCTKRYT**DPSSLRK**HVKTVHTGQKP		
F3	FQCRI--CMRNFS**QSGDLTR**HIRTHTGEKP		
F4	FACQY--CGKRFH**QKSDMKK**HTYIHLRGS		

Supplementary Table 7. Ratios of ZFNs injected into zebrafish embryos relative to robustness in B1H selection system.

Factor Name:	**5p_irs1b-like**	**3p_irs1b-like**
[3-AT]:	5mM	5mM
Fold-over-background	14	4
ZFN Ratio:	1	4
Factor Name:	**5p_nhlh2**	**3p_nhlh2**
[3-AT]:	5mM	5mM
Fold-over-background	4	22
ZFN Ratio:	4	1
Factor Name:	**5p_nr3c1**	**3p_nr3c1nn**
[3-AT]:	10mM	5mM
Fold-over-background	331	30
ZFN Ratio:	1	5
Factor Name:	**5p_irs1**	**3p_irs1**
[3-AT]:	10mM	5mM
Fold-over-background	16	812
ZFN Ratio:	1	1

Supplemental References:

Badenhorst, P., Harrison, S., and Travers, A. 1996. End of the line? Tramtrack and cell fate determination in Drosophila. *Genes to cells : devoted to molecular & cellular mechanisms* **1:** 707-716.

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X. et al. 2009. Diversity and complexity in DNA recognition by transcription factors. *Science* **324:** 1720-1723.

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L. et al. 2008. A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Molecular cell* **32:** 878-887.

Bae, K.H., Kwon, Y.D., Shin, H.C., Hwang, M.S., Ryu, E.H., Park, K.S., Yang, H.Y., Lee, D.K., Lee, Y., Park, J. et al. 2003. Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nature biotechnology* **21:** 275-280.

Bayer, C.A., Holley, B., and Fristrom, J.W. 1996. A switch in broad-complex zinc-finger isoform expression is regulated posttranscriptionally during the metamorphosis of Drosophila imaginal discs. *Dev Biol* **177:** 1-14.

Bayer, C.A., von Kalm, L., and Fristrom, J.W. 1997. Relationships between protein isoforms and genetic functions demonstrate functional redundancy at the Broad-Complex during Drosophila metamorphosis. *Dev Biol* **187:** 267-282.

Benos, P.V., Lapedes, A.S., and Stormo, G.D. 2002. Probabilistic code for DNA recognition by proteins of the EGR family. *J Mol Biol* **323:** 701-727.

Bergman, C.M., Carlson, J.W., and Celniker, S.E. 2005. Drosophila DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, Drosophila melanogaster. *Bioinformatics* **21:** 1747-1749.

Brown, J.L. and Wu, C. 1993. Repression of Drosophila pair-rule segmentation genes by ectopic expression of tramtrack. *Development* **117:** 45-58.

Bryne, J.C., Valen, E., Tang, M.H., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B., and Sandelin, A. 2008. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res* **36:** D102-106.

Bulyk, M.L., Huang, X., Choo, Y., and Church, G.M. 2001. Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc Natl Acad Sci U S A* **98:** 7158-7163.

Cho, S.Y., Chung, M., Park, M., Park, S., and Lee, Y.S. 2008. ZIFIBI: Prediction of DNA binding sites for zinc finger proteins. *Biochemical and biophysical research communications* **369:** 845-848.

Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D., and Barbas, C.F., 3rd. 2001. Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial

transcription factors. *The Journal of biological chemistry* **276:** 29466-29478.

Dreier, B., Fuller, R.P., Segal, D.J., Lund, C.V., Blancafort, P., Huber, A., Koksch, B., and Barbas, C.F., 3rd. 2005. Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *The Journal of biological chemistry* **280:** 35588-35597.

Dreier, B., Segal, D.J., and Barbas, C.F., 3rd. 2000. Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J Mol Biol* **303:** 489-502.

Elrod-Erickson, M. and Pabo, C.O. 1999. Binding studies with mutants of Zif268. Contribution of individual side chains to binding affinity and specificity in the Zif268 zinc finger-DNA complex. *The Journal of biological chemistry* **274:** 19281-19285.

Greisman, H.A. and Pabo, C.O. 1997. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275:** 657-661.

Gupta, A., Christensen, R.G., Rayla, A.L., Lakshmanan, A., Stormo, G.D., and Wolfe, S.A. 2012. An optimized two-finger archive for ZFN-mediated gene targeting. *Nat Methods*.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. 2007. Quantifying similarity between motifs. *Genome Biol* **8:** R24.

Harrison, S.D. and Travers, A.A. 1990. The tramtrack gene encodes a Drosophila finger protein that interacts with the ftz transcriptional regulatory region and shows a novel embryonic expression pattern. *The EMBO journal* **9:** 207-216.

Havranek, J.J., Duarte, C.M., and Baker, D. 2004. A simple physical model for the prediction and design of protein-DNA interactions. *J Mol Biol* **344:** 59-70.

Isalan, M., Choo, Y., and Klug, A. 1997. Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc Natl Acad Sci U S A* **94:** 5617-5621.

Isalan, M., Klug, A., and Choo, Y. 1998. Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry* **37:** 12026-12033.

Isalan, M., Klug, A., and Choo, Y. 2001. A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nature biotechnology* **19:** 656-660.

Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J.M., Yan, J., Sillanpaa, M.J. et al. 2010. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome research* **20:** 861-873.

Kaplan, T., Friedman, N., and Margalit, H. 2005. Ab initio prediction of transcription factor targets using structural knowledge. *PLoS Comput Biol* **1:** e1.

Kim, C.A. and Berg, J.M. 1995. Serine at position 2 in the DNA recognition helix of a Cys2-His2 zinc finger peptide is not, in general, responsible for base recognition. *J Mol Biol* **252:** 1-5.

Kim, C.A. and Berg, J.M. 1996. A 2.2 A resolution crystal structure of a designed zinc finger protein bound to DNA. *Nature structural biology* **3:** 940-945.

Kim, H.J., Lee, H.J., Kim, H., Cho, S.W., and Kim, J.S. 2009. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome research* **19:** 1279-1288.

Lam, K.N., van Bakel, H., Cote, A.G., van der Ven, A., and Hughes, T.R. 2011. Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. *Nucleic Acids Res* **39:** 4680-4690.

Liu, J. and Stormo, G.D. 2008. Context-dependent DNA recognition code for C2H2 zinc-finger transcription factors. *Bioinformatics* **24:** 1850-1857.

Liu, Q., Xia, Z., Zhong, X., and Case, C.C. 2002. Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *The Journal of biological chemistry* **277:** 3850-3856.

Maeder, M.L., Thibodeau-Beganny, S., Osiak, A., Wright, D.A., Anthony, R.M., Eichtinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A. et al. 2008. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Molecular cell* **31:** 294-301.

Meng, X., Noyes, M.B., Zhu, L.J., Lawson, N.D., and Wolfe, S.A. 2008. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nature biotechnology* **26:** 695-701.

Noyes, M.B., Meng, X., Wakabayashi, A., Sinha, S., Brodsky, M.H., and Wolfe, S.A. 2008. A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system. *Nucleic Acids Res* **36:** 2547-2560.

Persikov, A.V., Osada, R., and Singh, M. 2009. Predicting DNA recognition by Cys2His2 zinc finger proteins. *Bioinformatics* **25:** 22-29.

Read, D., Levine, M., and Manley, J.L. 1992. Ectopic expression of the Drosophila tramtrack gene results in multiple embryonic defects, including repression of even-skipped and fushi tarazu. *Mechanisms of development* **38:** 183-195.

Read, D. and Manley, J.L. 1992. Alternatively spliced transcripts of the Drosophila tramtrack gene encode zinc finger proteins with distinct DNA binding specificities. *The EMBO journal* **11:** 1035-1044.

Sander, J.D., Dahlborg, E.J., Goodwin, M.J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S.J., Blackburn, J.S., Thibodeau-Beganny, S., Qi, Y. et al. 2011. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods* **8:** 67-69.

Segal, D.J., Dreier, B., Beerli, R.R., and Barbas, C.F., 3rd. 1999. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A* **96:** 2758-2763.

Siggers, T.W. and Honig, B. 2007. Structure-based prediction of C2H2 zinc-finger binding specificity: sensitivity to docking geometry. *Nucleic Acids Res* **35:** 1085-1097.

Spokony, R.F. and Restifo, L.L. 2007. Anciently duplicated Broad Complex exons have distinct temporal functions during tissue morphogenesis. *Development genes and evolution* **217:** 499-513.

Stoll, R., Lee, B.M., Debler, E.W., Laity, J.H., Wilson, I.A., Dyson, H.J., and Wright, P.E. 2007. Structure of the Wilms tumor suppressor protein zinc finger domain bound to DNA. *J Mol Biol* **372:** 1227-1245.

Swirnoff, A.H. and Milbrandt, J. 1995. DNA-binding specificity of NGFI-A and related zinc finger transcription factors. *Molecular and cellular biology* **15:** 2275-2287.

Thummel, C.S. 2001. Molecular mechanisms of developmental timing in C. elegans and Drosophila. *Dev Cell* **1:** 453-465.

Wolfe, S.A., Grant, R.A., Elrod-Erickson, M., and Pabo, C.O. 2001. Beyond the "recognition code": structures of two Cys2His2 zinc finger/TATA box complexes. *Structure* **9:** 717-723.

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.

Wolfe, S.A., Nekludova, L., and Pabo, C.O. 2000. DNA recognition by Cys2His2 zinc finger proteins. *Annual review of biophysics and biomolecular structure* **29:** 183-212.

Workman, C.T., Yin, Y., Corcoran, D.L., Ideker, T., Stormo, G.D., and Benos, P.V. 2005. enoLOGOS: a versatile web tool for energy normalized sequence logos. *Nucleic Acids Res* **33:** W389-392.

Xiong, W.C. and Montell, C. 1993. tramtrack is a transcriptional repressor required for cell fate determination in the Drosophila eye. *Genes Dev* **7:** 1085-1096.

Yanover, C. and Bradley, P. 2011. Extensive protein and DNA backbone sampling improves structure-based specificity prediction for C2H2 zinc fingers. *Nucleic Acids Res* **39:** 4564-4576.

Zhu, C., Byers, K.J., McCord, R.P., Shi, Z., Berger, M.F., Newburger, D.E., Saulrieta, K., Smith, Z., Shah, M.V., Radhakrishnan, M. et al. 2009. High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome research* **19:** 556-566.