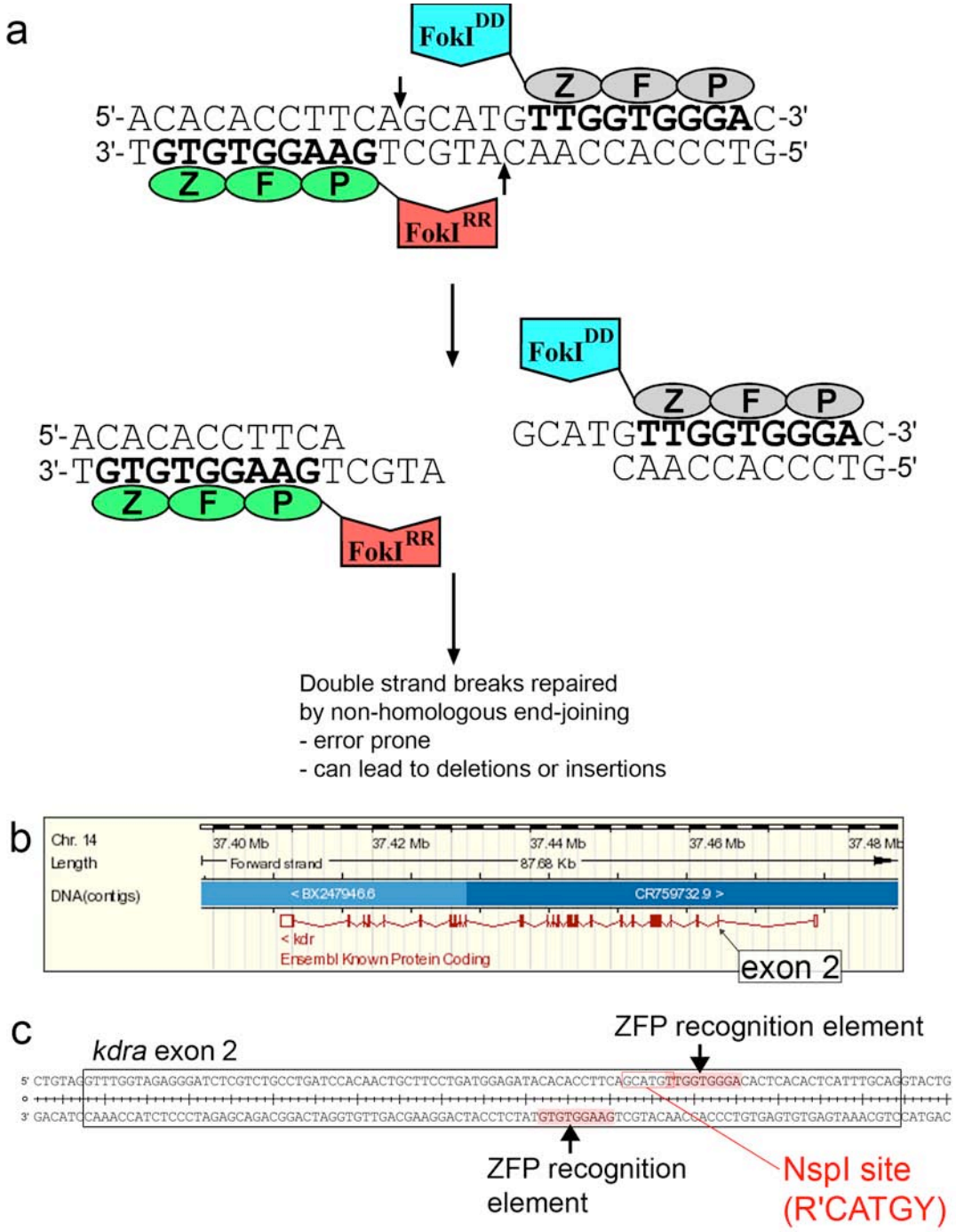


Supplementary Figure 1. Targeting the zebrafish *kdra* gene with zinc finger nucleases (ZFNs). **a**. Schematic representation of ZFN induced cleavage. Upon binding, the heterodimeric nucleases will cleave the target DNA (indicated by arrow) leading to a double strand break (DSB). The DSB can then be repaired by non-homologous end-joining (NHEJ). Since this is an error prone repair pathway, NHEJ often results in deletions or insertions at the site of the repaired DSB. **b**. Image from ENSEMBL Genome browser showing the *kdra* locus. Position of exon 2 is indicated. **c**. Sequence of *kdra* exon 2. Target zinc finger recognition sites and the NspI restriction site are indicated.

Supplementary Figure 2.

Algorithm for the identification of favorable ZFN recognition sequences. A simple PERL script was constructed that allows an enduser to input a series of Exon sequences for a gene of interest (fasta format) and set thresholds for the number of purines and guanines required for the identification of suitable pairs of 9bp ZFP recognition elements. This program will output a “hit” list in sequential order that can be evaluated by the enduser with regards to position and quality of the target sites for the most favorable set of ZFPs to create. This program differs from other webtools, ZiFiT (bindr.gdcb.iastate.edu/ZiFiT/) and Zinc Finger Tools (zincfingertools.org), in that it only considers the number of purines in the site, which we believe is a hallmark of a favorable ZFP recognition site. The other zinc finger tools consider sites that are compatible with preexisting zinc finger modules that can be assembled to generate ZFPs with a desired specificity. The file “kdra_exons_fasta.txt” can be used as test input for this program. The PERL program creates output file “tempout.txt” in the current working directory.

```
#!/usr/bin/perl -w
#
use strict;

# test file use 'kdra_exons_fasta.txt';

###
#
# input statements
#
###

print "\nProgram creates output file : tempout.txt : in current directory with output\n\n";

print "Enter Exon file Name (Must be in fasta format plain text w/line breaks): ";

my($ZFlist) = '';

$ZFlist= <STDIN>;

chomp $ZFlist;

print "Enter the minimum number of purines per composite site: ";

my($purinecount) = '';

$purinecount= <STDIN>;

chomp $purinecount;

$purinecount = $purinecount - 1;
```

```
print "Enter the minimum number of guanines per composite site: ";

my($guaninecount) = '';

$guaninecount= <STDIN>;

chomp $guaninecount;

$guaninecount = $guaninecount - 1;

###
# open output file
#

open (OUTFILE, ">tempout.txt");

###

my($revcom)='', my($x)=0, my($ii)=0,
my($r1)=0,my($r2)=0,my($y1)=0,my($y2)=0,my($i)=0;
my($fastahead)='', my($fastaseq)='', my($temp)='', my($temp5)='', my($temp3)='';
my($invprimerlist) = '';
my($header) = '';
my($headlength)=0;

open(ZFL, $ZFlist);

while ( $fastahead= <ZFL> ) {

###
# Extract header

chomp $fastahead;

$headlength = length ($fastahead) -1;

$header = substr ($fastahead, 1, $headlength);

#
###
# Extract exon sequence
#
$fastaseq= <ZFL>;

$x = 0, $i=0,  $temp = '', $ii = '18 to 21';

$fastaseq =~ s/\s//g;

$x = length $fastaseq;

##
```

```
# Analyze 24 bp blocks (6bp gap) for composition of ZFP sites

while ($i < ($x -24)) {

$r1=0, $r2=0, $y1=0, $y2=0;

$temp = substr ($fastaseq, $i, 24);

$temp5 = substr ($fastaseq, $i, 9);

$temp3 = substr ($fastaseq, $i + 15, 9);

while ( $temp3 =~ /a/ig) {$r2++}
while ( $temp3 =~ /g/ig) {$r1++}
while ( $temp5 =~ /c/ig) {$r1++}
while ( $temp5 =~ /t/ig) {$r2++}

$y1 = $r1 + $r2;

if ($y1 > $purinecount && $r1 > $guaninecount) {

$revcom = reverse $temp5;

$revcom =~ tr/ACGT/TGCA/;

print OUTFILE $header,  "\n", "Site position ", $i, "\n", "6bp gap", "\nComposite site: ",
$temp, "\n5' ZFP site:    ", $revcom, "\n3' ZFP site:    ", $temp3, "\nPurines in site:  ",$y1,
"\nGuanines in site:  ",$r1, "\n\n";

}

$i++;

}

##
# Analyze 23 bp blocks (5bp gap) for composition of ZFP sites

$i = 0;

$x = length $fastaseq;

while ($i < ($x - 23)) {

$r1=0, $r2=0, $y1=0, $y2=0;

$temp = substr ($fastaseq, $i, 23);

$temp5 = substr ($fastaseq, $i, 9);

$temp3 = substr ($fastaseq, $i + 14, 9);
```

```
while ( $temp3 =~ /a/ig) {$r2++}
while ( $temp3 =~ /g/ig) {$r1++}
while ( $temp5 =~ /c/ig) {$r1++}
while ( $temp5 =~ /t/ig) {$r2++}

$y1 = $r1 + $r2;

if ($y1 > $purinecount && $r1 > $guaninecount) {

$revcom = reverse $temp5;

$revcom =~ tr/ACGT/TGCA/;

print OUTFILE $header,  "\n", "Site position ", $i, "\n", "5bp gap", "\nComposite site: ",
$temp, "\n5' ZFP site:    ", $revcom, "\n3' ZFP site:    ", $temp3, "\nPurines in site:  ",$y1,
"\nGuanines in site:  ",$r1, "\n\n";

}

$i++;

}

}

print "\nOutput placed in tempout.txt\n";

close ZFL;

close OUTFILE;

exit;
```

Supplementary Figure 3.

Sequence analysis of clones in selected single finger library pools. Sequences of a small number of fingers (recognition helix positions -1 through 6) from each Stage 1 finger selection. Selected positions in each finger are indicated in bold. The pool of clones isolated from the finger 3 selection against the TTG subsite (Finger 3-TTG) within the 9 bp recognition element **TTG**GTGGGA does not display as tight a consensus as the other sets of selected fingers. It appears that this is a more challenging site for zinc finger recognition based on our efforts to improve this finger through further selection beyond that described herein (Meng, X and Wolfe, SA *unpublished results*).

Fingers selected against 3 bp subsites in target GAAGGTGTG

Finger 1-GTG subsite

Only a single position is randomized in the finger so the raw library was recombined prior to the second selection stage.

Finger 2-GGT subsite

CRHHLTR
LKHHLTR (3 unique codon sets)
QKHHLTR
LRHHLTR (3 unique codon sets)

Finger 3-GAA subsite

QGANLTR
QGANLTR
QLCNLTR
QHGNLXR
QACNLTR
QSSNLTR
QSHNLTR
QNGNLTR

Fingers selected against 3 bp subsites in target TTGGTGGGA

Finger 1-GGA subsite

CSGHLTR
IQGHLTR
QKGHLTR
IKGHLTR
QQGHLTR
CKGHLTR (2 duplicates)
TKGHLTR

Finger 2-GTG subsite

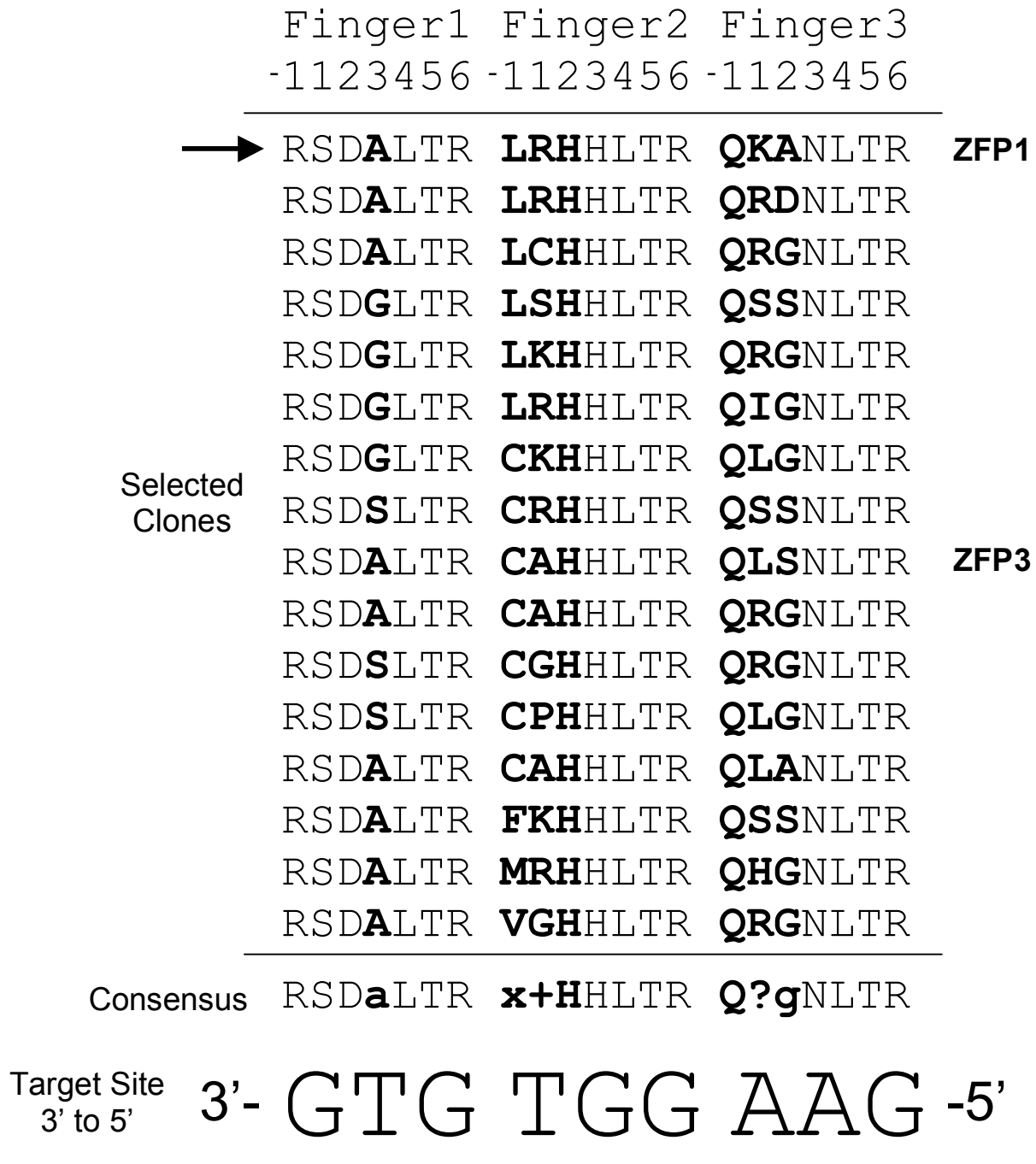
Only a single position is randomized in the finger so the raw library was recombined prior to the second selection stage.

Finger 3-TTG subsite

RSD**G**L**WL**
RSD**Q**L**VV**
RSD**R**L**FT**
RSD**Y**L**LF**
RSD**P**L**KL**
RSD**W**L**FV**
RSD**G**L**SA**

Supplementary Figure 4.

Sequences of the recognition helices from ZFPs selected to recognize the GAAGGTGTG recognition element. Bold residues indicate selected positions. An arrow indicates the clone used within the ZFN1 to target exon 2 in *kdra*. Names beside the sequences indicate clones characterized by binding site selections (Supplementary Fig. 6).



Supplementary Figure 5.

Sequences of the recognition helices from ZFPs selected to recognize the TTGGTGGGA recognition element. Bold residues indicate selected positions. An arrow indicates the clone used within the ZFN2 to target exon 2 in *kdra*. A clone (ZFP5) with Glu at position 3 in finger 2 was characterized, since this is the consensus residue at this position, but it preferred C over T in the binding site selections (Supplementary Fig. 6) and so was not utilized.

	Finger1 -1123456	Finger2 -1123456	Finger3 -1123456	
→	**QKG**HLTR	RSD**A**LTR	RSD**S**L**LG**	**ZFP2**
	QKGHLTR	RSD**E**LTR	RSD**M**L**LG**	
	QQGHLTR	RSD**S**LTR	RSD**E**L**SE**	
	QKGHLTR	RSD**E**LTR	RSD**M**L**LG**	
	QLGHLTR	RSD**A**LTR	RSD**M**L**GS**	
	QRGHLTR	RSD**A**LTR	RSD**L**L**LL**	**ZFP4**
	QKGHLTR	RSD**F**LTR	QSS**D**L**QR**	
Selected Clones	**CWG**HLTR	RSD**G**LTR	RSD**V**L**QW**	
	CKGHLTR	RSD**E**LTR	RSD**A**L**GL**	**ZFP5**
	CKGHLTR	RSD**E**LTR	RSD**C**L**YT**	
	CKGHLTR	RSD**E**LTR	RSD**A**L**RL**	
	SKGHLTR	RSD**A**LTR	RSD**R**L**SF**	
	SKGHLTR	RSD**E**LTR	RSD**S**L**QG**	
	SKGHLTR	RSD**V**LTR	RSD**C**L**LG**	
	HKGHLTR	RSD**E**LTR	RSD**S**L**FL**	
	HKGHLTR	RSD**S**LTR	QSS**D**L**QR**	
	AKGHLTR	RSD**E**LTR	RSD**A**L**LS**	
	NKGHLTR	RSD**G**LTR	RSD**G**L**KG**	
	AKGHLTR	RSD**E**LTR	RSD**A**L**LS**	
Consensus	**?kG**HLTR	RSD**e**LTR	RSD**?**L**??**	

Target Site 3' to 5' 3'- AGG GTG GTT -5'

Supplementary Figure 6.

Binding site analysis for ZFPs indicated in Supplementary Figures 5 & 6. Raw sequences of the selected binding sites are listed first followed by the overrepresented motif present in the MEME alignment. The Sequence logo for clones not used in each ZFN is present below each alignment.

ZFP1 – ZFP used in ZFN1 (RSD**A**LTR **LRH**HLTR **QKA**NLTR) DNA-binding site selection results (27 clones)

Raw sequences:

GTCCACACAGGCCGACAACACACCACCA
GCCCCTGACAGTTTGATCCACACCCGCA
GCTCCCCACCACTACGACCACACCCTCG
CACTCGCCCACGTTTGAGAAGGTGCGGc
GTCCCACGCGCAAACCACACCATCACCC
CTTGTGATCCCGGGACCCCACACCCTCA
ACGCCAAAACGCCCCACACACCCTCACC
CCCTACACACCTCCCCGCACCCTCACCA
ACCGCGACCACGTTAGGTGGTGTGGAAC
GCCATTCCGACTATGCGGGGGTGTGGTG
GACTTACCACGGGACGAAGGTGCGCCTA
ACGGAGCCAGCACACCACACCCTCACTG
ATCTCTGGTCGAAACGAAGGTGTGGAAC
TGGGTGATTTGCGCGCCCACACCACCGC
TTGAGGCAGATATGCACACACCCTCATA
AGGCCCGAGGGTGTGTTCGCCCGCGGCT
CAACCCGACTGAGTCGATACCACACCCT
ACGCTAGGAACACTTAGGAGGGTGTGGC
ACGGCATCCACTAACTCCCACACCCCCAC
CTGCCACACCTTCATAGCACTGTCCCGA
CGCAGCATTTATCCCCACACCACCCCCA
CAACCCGACTGAGTCGATACCACACCCTC
CGCGTACCCTGCTTTACTCCACACCACCC
GCAGCAGAGGGTGTGTCAGCTCCCAAAT
TACCACACCCCCGATAAACACACCCTCCC
CGGACTTCAGTCCTCCCACACCCTCCCA
CAAACACCGAAAACCCACCACACCCCCA

MEME identified the DNA-binding motif from 26 clones (E-value=7.6e-056)(Note: the motif is reverse complementary to the target site (GAAGGTGTG))

TTCAGTCCTC**CCACACCCTC**CCA
TGAGTCGATA**CCACACCCTC**
G**CCACACCCTC**CTAAGTGTTC
AGCCAGCACA**CCACACCCTC**ACTG
TCCCGGGACC**CCACACCCTC**A
ACCACTACGA**CCACACCCTC**G

CACGCGCAAACCACACCATCACCC
CCCCCGATAAACACACCCTCCC
TTGGGAGCTGACACACCCTCTGCTGC
CCGCGGGCGAACACACCCTCGGGCCT
GCAGATATGCACACACCCTCATA
AAAACGCCCCACACACCCTCACC
CTGCCACACCTTCATAGCACTGT
GTTCCACACCTTCGTTTCGACCA
CGAAAACCCACCACACCCCCA
CCACTAACTCCCACACCCCCAC
CACCACACCCCCGCATAGTCGG
CTGCTTTACTCCACACCACCCCCC
GCATTTATCCCCACACCACCCCCA
ATTTGCGCGCCCACACCACCGC
GTTCCACACCACCTAACGTGGTC
CAGGCCGACAACACACCACCA
CCCTACACACCTCCCCGCACCCTC
ACAGTTTGATCCACACCCGCA
GCCGCACCTTCTCAAACGTGG
TAGGCGCACCTTCGTCCCGTGGT

For Sequence Logo see Figure 1c.

ZFP2 – ZFP used in ZFN2 (**QKG**HLTR RSD**A**LTR RSD**S**L**LG**) clone DNA-binding site selection result (22 clones)

Raw sequences:
CGAACTGAGGTAGTAAAACCCACCACCC
ACCTTGCGCGCACGACCGGTGGGAAGGT
GCATACCGGAACGTTAACAGGGGGAAGA
CTCAACACCGGTCTCAAGGGTGGGAAAC
TCCCCATGGTGGGAGAGCACAGACTCAG
GATACATAAATTAAACACGGAGGGATGG
CAACCCTATGACACCCTCCCACCACACT
AATGTTCTCGCGATCTTGGTGGGACTAG
GCGTTTCACTGTGGCCCGTCCCACCATT
ACTAAGCACAGCGCCCACCTATCGACCG
GCACCAGGCGTAGCATACGGTGGGACTC
TAGGCCCCGGTGGGACCCGTACTACCTA
GAGGGCCGCAGGAACTCCCACCGAGACA
GTAATCGATCGCTCACCTGGTGGGTTGG
CCACCACGGATAACACAAACCCACCATA
ACACCCCCGCTGGAGTGCAACCAATCAG
CGATATCCGTCTATGACATCCCACCCGA
CCATTTGGCCGTCACCTTGGTGGGATAC
CACAGAGGTAATGGAGCGGTGGGACGCG
GCAACCCTTGGAAAAACATCCCACCGGC

GGATCAACCACGCAATCACGGAGGGATG
ATTGAGCACCAGGATAGGGTGGGACCCA

MEME identified the DNA-binding motif from 22 clones (E-value=2.1e-025)

GCCGTCACCTTGGTGGGATAC
AATGGTGGGACGGGCCACAG
CTCGCGATCTTGGTGGGACTAG
AGTGTGGTGGGAGGGTGTCATA
TCCCCATGGTGGGAGAGCACAGAC
GCCGGTGGGATGTTTTTCCA
GGTAATGGAGCGGTGGGACGCG
TGTCTCGGTGGGAGTTCCTGCGG
TAGGCCCCGGTGGGACCCGTACTAC
GCGTAGCATACGGTGGGACTC
CGCGCACGACCGGTGGGAAGGT
CACCAGGATAGGGTGGGACCCA
TCGGGTGGGATGTCATAGAC
CCGGTCTCAAGGGTGGGAAAC
TATGGTGGGTTTGTGTTATC
ATCGCTCACCTGGTGGGTTGG
GGGTGGTGGGTTTTACTACCT
CACGCAATCACGGAGGGATG
AAATTAAACACGGAGGGATGG
CGGTCGATAGGTGGGCGCTGTGCTTA
CTGATTGGTTGCACTCCAGCGGG
GGAACGTTAACAGGGGGAAGA

For Sequence Logo see Figure 1c.

ZFP3 (RSD**A**LTR **CAH**HLTR **QLS**NLTR)clone DNA-binding site selection result (21 clones)

Raw sequences:
ACGCAAGAGGGTGTGTCAACGGTAAGTG
AGTTATCCAAAGTAACACACCTTCAGCA
GTGGCCAGAAAATGAGAAGGTGCGGAGA
AAGCGACGCCAAACCTGAAGGTGTGTTA
CTACCGCCTCTCCGAACACACCTTCGCT
ACGCCAAAACGCCCCACACACCCTCACC
TTACCCAAACTACCCCCACACCTTCCCC
GTTTGACCATTGCAACCACACCTTCAAC
GGGTACCAGCACGTCAAACACACCCTCA
GCAACGGAAGGTGTGAGGAATACCCACG
AGTAGTGAAGGGGTGGATGACAAAACTG
AGGCCCGAGGGTGTGTTCGCCCGCGGCT

ACACCTTATCTTGTATGAAGGTGCGGGT
ACGCTAGGAACACTTAgGAGGGTGTGGc
ATGTCCCGGGCAAAGTGAAGGTGTGGGC
CCCCACACCAGGACAAGCCACACCCTCA
CGTCACCATTACACAACACACCTTCACC
AAACTGTGACATCTACCACACCTTCATC
ATAGTTACTCGTCTACACACCTTCATAC
GACACAGTAACGGAAGAGGGTGTGATGC
CTGCCACACCTTCATAGCACTGTCCCGA

MEME identified the DNA-binding motif from 21 clones (E-value=1.7e-055)(Note: the motif is reverse complementary to the target site (GAAGGTGTG))

TTACTCGTCTACACACCTTCATAC
CCATTACACAACACACCTTCACC
TAACACACCTTCAGGTTTGGCGT
ATCCAAAGTAACACACCTTCAGCA
CTGCCACACCTTCATAGCACTGTC
GTGACATCTACCACACCTTCATC
GCCCACACCTTCACTTTGCCCGG
ACCATTGCAACCACACCTTCAAC
AGCACGTCAAACACACCCTCA
AAAACGCCCCACACACCCTCACC
CCAGGACAAGCCACACCCTCA
CAAACTACCCCCACACCTTCCCC
GCCTCTCCGAACACACCTTCGCT
CTTACCGTTGACACACCCTCTTGCGT
ACCCGCACCTTCATACAAGATAA
GGCCACACCCTCCTAAGTGTTCC
CCGCGGGCGAACACACCCTCGGGCCT
GTTTTGTCATCCACCCCTTCACTACT
TGGGTATTCCTCACACCTTCCGTTGC
GCATCACACCCTCTTCCGTTACTG
TCTCCGCACCTTCTCATTTTCTGG

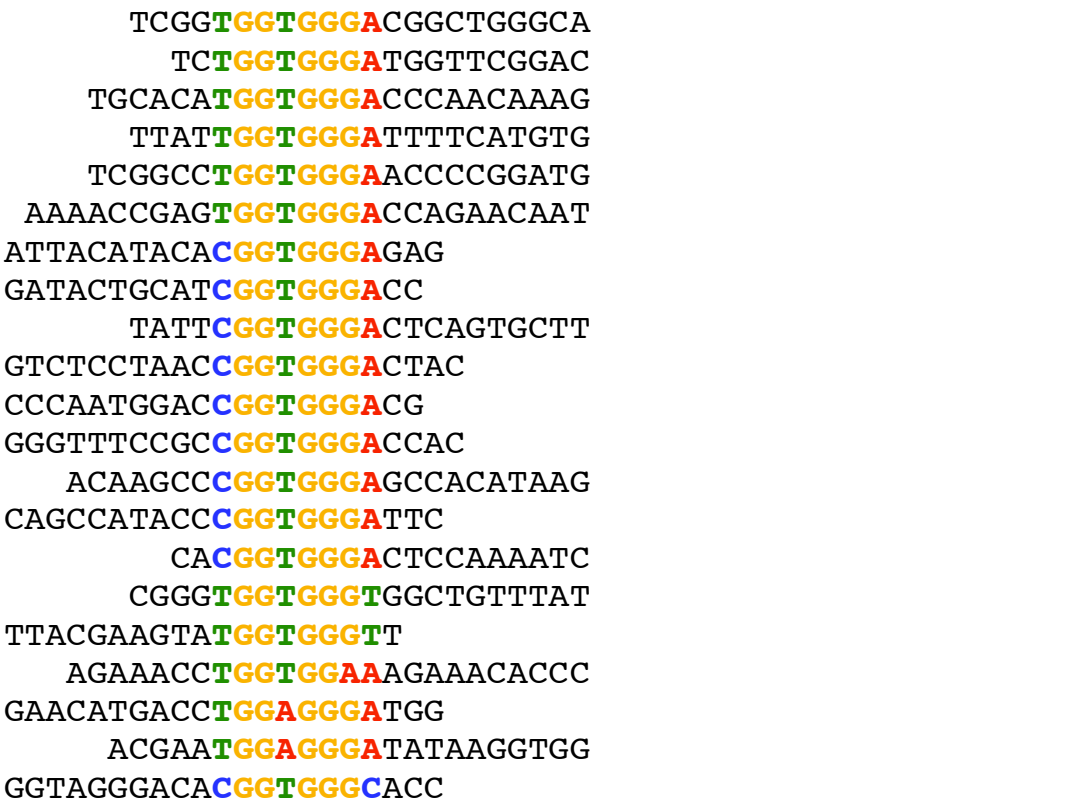
ZFP4 (**QRG**HLTR RSD**A**LTR RSD**LLLL**) DNA-binding site selection result (23 clones)

Raw sequences:

CGTTAGTGCTTACGAAGTATGGTGGGTT

AAAACCGAGTGGTGGGACCAGAACAATG
TAGTCCATGATTTTGGAGTCCCACCGTG
ACAAGATCAGCCATACCCGGTGGGATTC
ACAAGCCCGGTGGGAGCCACATAAGACT
CCAAGTGGGTTTCCGCCGGTGGGACCAC
TCGGCCTGGTGGGAACCCCGGATGCACT
ACCAAACACATGAAAATCCCACCAATAA
AGCAGTGGCCCAATGGACCGGTGGGACG
TGCACATGGTGGGACCCAACAAAGATCA
GGCATTTATGGTGGATCTCCAGGACAAG
GGTCACATGTCCGAACCATCCCACCAGA
CACCCTTGTCTCCTAACCGGTGGGACTAC
CACGCGTTGGTAGGGACACGGTGGGCACC
AGAAACCTGGTGGAAAGAAACACCCTG
GCTACAAGCACTGAGTCCCACCGAATA
ACGAATGGAGGGATATAAGGTGGACTTG
ACCGGTAGGATACTGCATCGGTGGGACC
CGCACTTGCCCAGCCGTCCCACCACCGA
CCACGGTATTACATACACGGTGGGAGAG
CCAGAGTTCAGTGAAACCCATGGTGGGAC
CACATCATAAACAGCCACCCACCACCCG
GCCGAAGGAACATGACCTGGAGGGATGG

MEME identified the DNA-binding motif from 23 clones (E-value=1.7e-037)



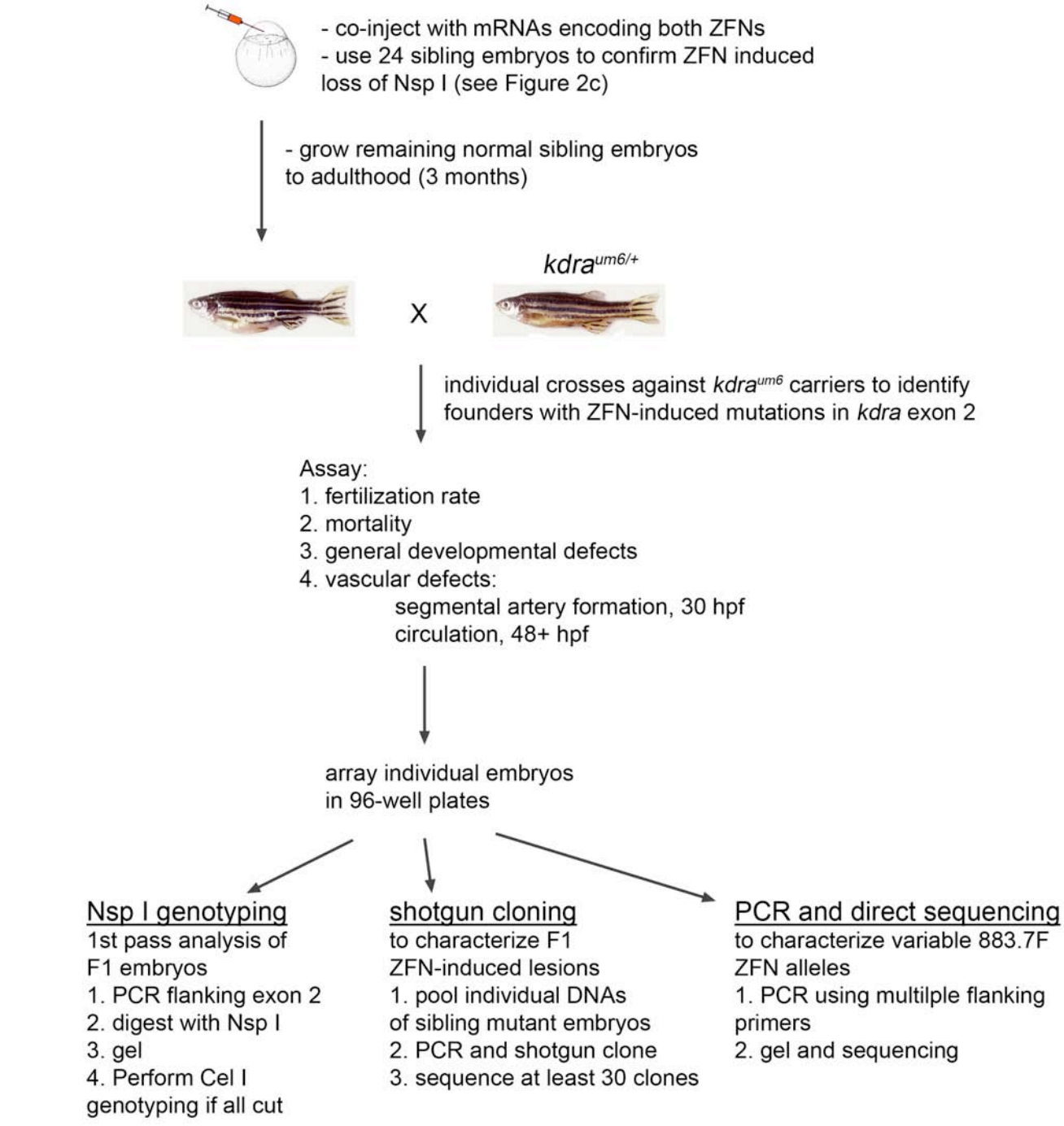
GGCATTTATGGTGGATCTCCAGGACA

ZFP5 (**CKG**HLTR RSD**E**LTR RSD**A**L**GL**) DNA-binding site selection results (6 clones)

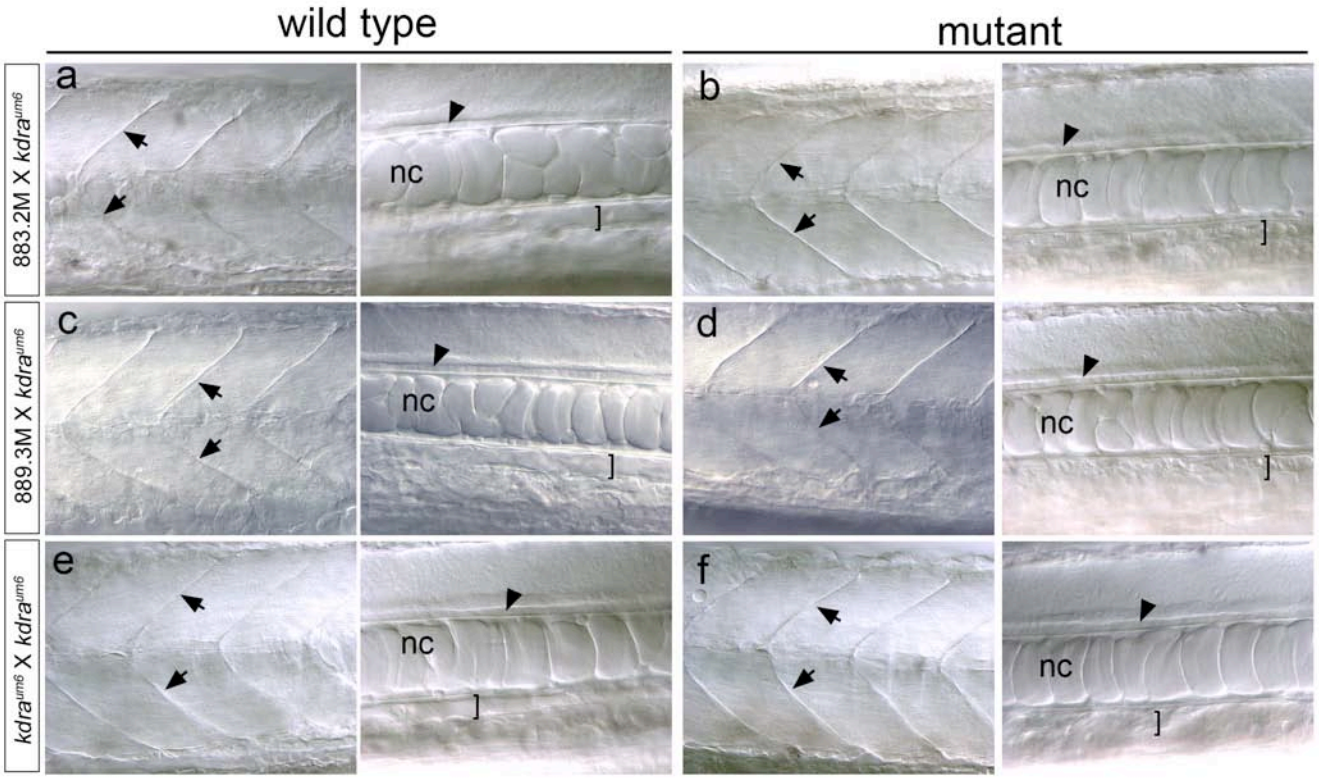
Raw sequences:
AGACTACGAAGCGCACCAGGCGGGATCC
ATAGCGAGAATTGACCCCTGGCGGGATC
TAACAGctGcccGTGcCgtggCGGGATCC
TATACCACGTTCCCAACCCGGCGGGATT
GTTCACAACCCACAGCCAAGGCGGGATCC
ACCTTGCAATGGAATTAAGGTGGGATCC

MEME identified the DNA-binding motif from 6 clones (E-value=2e-006)

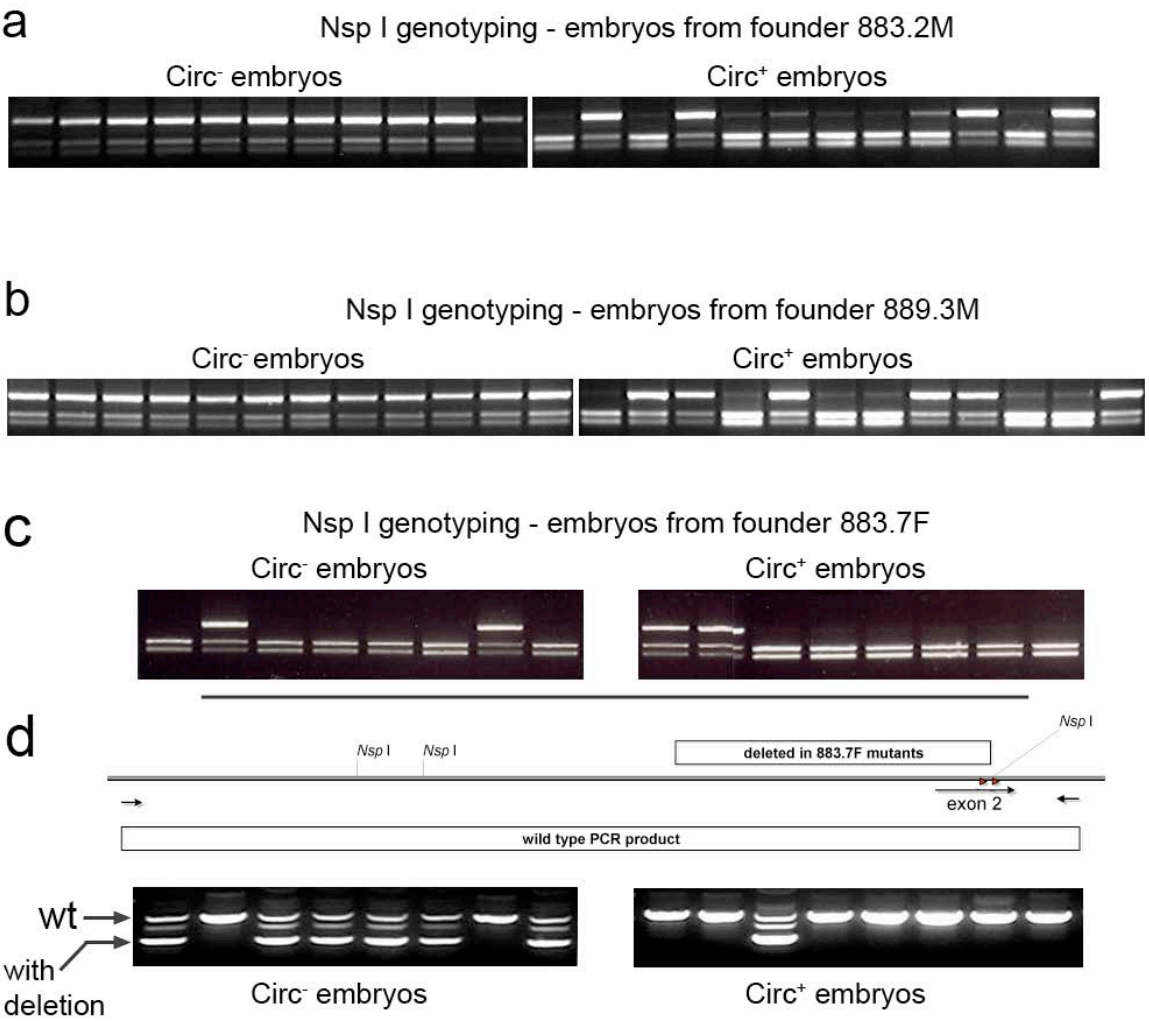
AAGCGCACCAGGCGGGATCC
CCACAGCCAAGGCGGGATCC
ATTGACCCCTGGCGGGATC
CCCGTGCCGTGGCGGGATCC
ATGGAATTAAGGTGGGATCC
TTCCCAACCCGGCGGGATT



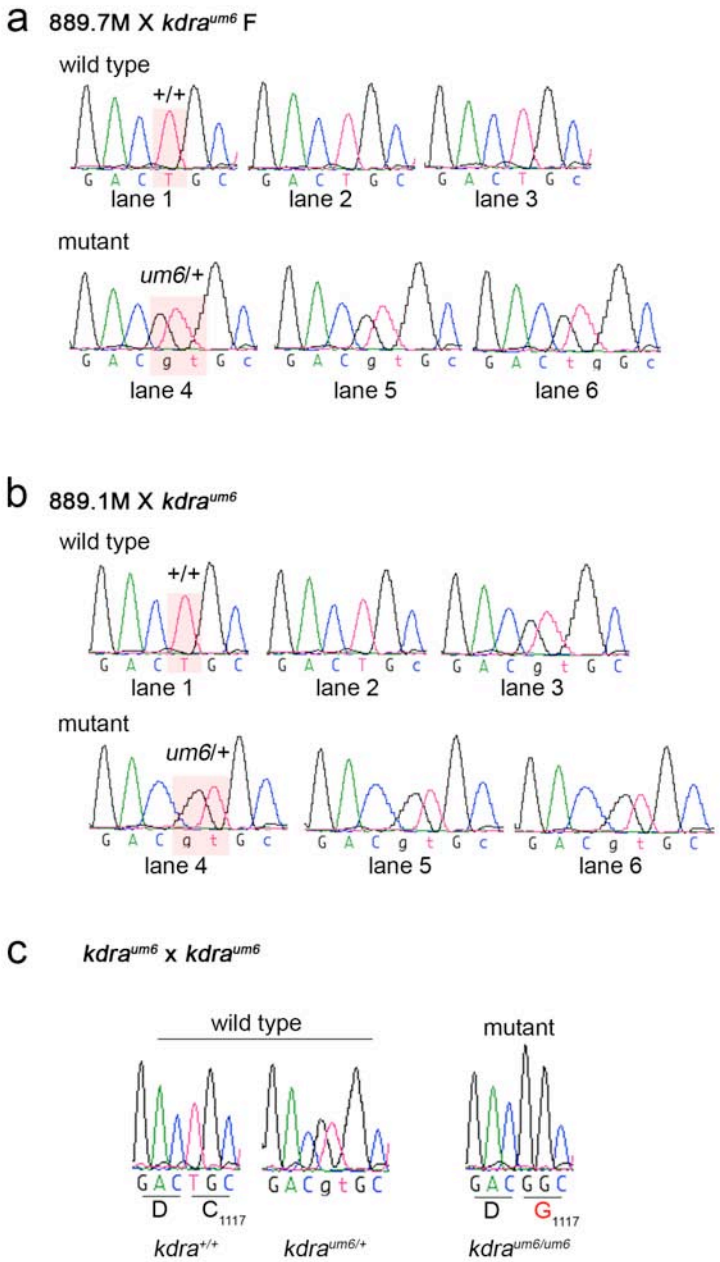
Supplementary Figure 7. Experimental approach for the analysis of F1 founders bearing ZFN-induced alleles at the *kdra* locus.



Supplementary Figure 8. Nomarski images of trunk morphology in wild type and mutant embryos derived from ZFN founders. In all panels, arrows indicate somite boundaries, arrowhead denotes floor plate, nc - notochord and bracket indicates lumen of the dorsal aorta. Left and right paired panels are from the same embryos in each case. **a.** Wild type embryo from 883.2M. **b.** Mutant embryo from 883.2M. **c.** Wild type embryo from 889.3M. **d.** Mutant embryo from 889.3M. **e.** Wild type embryo from incross of *kdra*^{um6} heterozygotes. **f.** Mutant embryo from from incross of *kdra*^{um6} heterozygotes. General trunk morphology appears normal in all cases, including mutant embryos. Dorsal aorta lumenization is not apparent in mutant embryos.



Supplementary Figure 9. Genotype analysis of embryos from F1 founders bearing ZFN-induced *kdra* alleles. In all panels, Circ^{-} indicates phenotypically mutant embryos that lack circulation at 48 hours post fertilization (hpf); Circ+ embryos display normal circulation at 48 hpf. **a**. NspI genotyping (see Figure 2) of embryos derived from a cross between founder 883.2M and a *kdra*^{um6} heterozygous female. All Circ^{-} embryos bear an NspI-resistant allele indicative of a ZFN-induced lesion at this site. **b**. NspI genotyping of embryos derived from a cross between founder 889.3M and a *kdra*^{um6} heterozygous female. As above, all Circ^{-} embryos bear an NspI-resistant allele. **c**. NspI genotyping of embryos derived from a cross between founder 883.7F and a *kdra*^{um6} heterozygous male. In this case, only a fraction of the Circ^{-} embryos display a NspI-resistant allele, suggesting the existence of multiple ZFN-induced alleles in the germ cells of the 883.7F founder. **d**. PCR genotyping of the same embryos in c. using flanking primers. Mutant embryos that displayed apparently wild type NspI genotypes show a smaller fragment suggesting a 300 to 400 bp deletion. The original primers used for NspI genotype analysis in c. would fail to amplify the mutant allele in embryos bearing this deletion. Thus, founder 883.7F bears two distinct alleles: a 382 bp deletion and a 12 bp deletion.



Supplementary Figure 10. SNP analysis for genotyping the *kdra*^{um6} allele. All chromatograms are from direct sequencing of PCR products amplified from wild type or mutant embryos from the indicated crosses. **a**. Genotypes from embryos of a cross between 889.7M and a *kdra*^{um6} heterozygous female. Embryos are the same as shown in Figure 4 (Lane numbers are indicated). Wild type and heterozygous sequences are indicated by pink shaded box. **b**. Genotypes from embryos of a cross between 889.1M and a *kdra*^{um6} heterozygous female. Embryos are the same as shown in Figure 4 (Lane numbers are indicated). **c**. SNP genotyping of embryos derived from an in-cross of *kdra*^{um6} heterozygote carriers.

Supplementary Table 1.
Solexa sequence reads and statistical analysis of ZFN-induced lesions at on- and off-target sites.

	Untreated	10 pg "normal"	10 pg "monster"	20 pg "monster"
All Sequence Reads				
Correct @ off-target site	8230	8281	10957	10649
Correct @ on target site	34	21	20	28
Indel* @ off-target site	0	2	6	9
Indel @ on-target site	0	4	2	4
Statistical Analysis on-target vs. off-target				
p-value	NA	9.53E-10	0.00011	4.64E-08
estimated odds ratio	NA	772.7	180.2	166.9
95%conf lower bound odds ratio	NA	103.5	16.9	35.9
95%conf upper bound odds ratio	NA	8192	1070.5	635.4
Only Heterodimeric ZFN sequence reads				
Correct @ off-target site	1278	1193	1061	1634
Correct @ on target site	34	21	20	28
Indel @ off-target site	0	2	5	8
Indel @ on-target site	0	4	2	4
Statistical Analysis on-target vs. off-target				
p-value	NA	2.01E-06	0.0077	4.90E-05
estimated odds ratio	NA	110.7	21	28.9
95%conf lower bound odds ratio	NA	14.9	1.9	6
95%conf upper bound odds ratio	NA	1291.7	137.7	115.7
Experimental comparisons	Overall	Pairwise comparison vs. untreated sequences		
p value heterodimeric ZFN sequence reads	0.02452	0.23	0.019	0.011
p value all sequence reads	0.02523	0.49	0.041	0.006

* - "Indel" = insertions and deletions in spacer region

		Fertilization		at 24 hpf		GFP+				%	germ cell
ID#	# emb	#	rate	dead	mon.	SeA+	SeA-	circ+	circ-	mutants	mosaicism
883.1M	300	187	62%	7	0	94	0	180	0	0%	0%
883.2M	108	64	59%	0	0	15	2	31	11	26%	52%
883.3M	~100	0	0%	-	-	-	-	-	-	-	-
883.4M	148	86	58%	0	0	42	0	86	0	0%	0%
883.5M	304	277	91%	63	0	107	0	214	0	0%	0%
883.4F	214	173	81%	10	6	77	0	157	0	0%	0%
883.5F	0	-	-	-	-	-	-	-	-	-	-
883.6F	77	50	65%	0	0	21	0	50	0	0%	0%
883.7F	55	53	96%	2	0	21	2	43	8	16%	31%
883.8F	77	76	99%	0	0	39	0	73	3	4%	8%
889.1F	45	29	64%	0	0	-	-	29	0	0%	0%
889.2F	50	28	56%	0	0	-	-	28	0	0%	0%
889.1M	305	198	65%	5	0	-	-	181	12	6%	12%
889.2M	202	175	87%	1	0	-	-	174	0	0%	0%
889.3M	117	101	86%	0	0	-	-	74	27	27%	53%
889.4M	130	50	38%	0	0	-	-	48	2	4%	8%
889.5M	222	209	94%	0	0	-	-	209	0	0%	0%
889.6M	108	10	9%	2	7	-	-	0	1	-	-
889.7M	29	22	76%	0	0	-	-	19	3	14%	27%
889.8M	151	126	83%	0	0	-	-	126	0	0%	0%
kdra^{um6} *inx*	326	216	66%	2	5	117	35	155	56	27%	53%

Supplementary Table 2. Identification and characterization of ZFN founders. Results are from embryo clutches from initial out-cross of indicated ZFN founder to *kdra*^{um6} heterozygous carriers. Putative ZFN allele bearing founders are indicated in yellow. ID# is a family number followed by an individual identifier. Thus, 883 and 889 are separate families derived from separate injections. M – male founder, F – female founder. GFP+ refers to crosses in which the *kdra*^{um6} heterozygote also harbored *Tg(fli1:egfp)*^{y1}. SeA(+/-) - scored for presence or absence of segmental arteries at 30 hours post fertilization (hpf). Circ(+/-) – scored for presence or absence of circulation at 48 hpf. Males 883.3 and 889.6 failed to fertilize eggs on subsequent crosses against gravid females (data not shown). Circ- embryos from female 883.8 also displayed other defects, including split axis. No defects were apparent on subsequent crosses.

	First cross					Second cross*				
	Fertilization		circulation			Fertilization		circulation		
Founder	#	rate	+	-	% mutants	#	rate	+	-	%mutants
883.2M	64	59%	31	11	26%	339	86%	42	20	32%
883.7F	53	96%	43	8	16%	-	-	-	-	
889.1M	198	64%	181	12	6%	389	91%	95	9	9%
889.3M	101	86%	74	27	27%	316	92%	67	26	28%
889.4M	50	38%	48	2	4%	336	92%	70	3	4%
889.7M	22	75%	19	3	14%	323	93%	54	16	23%
um6 inx	216	66.26%	155	56	27%	150	94%	122	28	19%

Supplementary Table 3. Fertilization rates and proportion of mutant embryos in successive ZFN founder crosses. In both cases, indicated founders were individually crossed against *kdra*^{um6} heterozygotes. Data from first cross are derived from Supplemental Table 1. Circulation was scored at 48 hours post fertilization. Female 883.7 failed to produce eggs in her second mating. * - only a random sampling of embryos was scored for circulatory defects in the second cross.

Supplementary Table 4. Genetic lesions and frequency of occurrence in mutant embryos from ZFN mutation bearing founders.

	Phenotypic scoring[1]				
Founder	**circ+**	**circ-**	**% mutants**	**Lesion**	**Frequency[2]**
883.2 M	31	11	26%	4 bp deletion	21/34
883.7 F	43	8	16%	variable[3]	4/39
889.1 M	181	12	6%	4 bp insertion	18/34
889.3 M	74	27	27%	24 bp deletion	17/39
889.7 M	19	3	14%	8 bp insertion	19/33

1. "Phenotypic scoring" – observation of circulation at 48 hours post fertilization. "Mutant" embryos display no circulation. Scoring data is same as in Supplementary Table 1 for indicated founders. **2**. "Frequency" is the number of **mutant** exon 2 sequences *over* **total number** of sequences obtained from shotgun cloning of PCR fragments from *pooled sibling* ***mutant*** embryos. Since mutant embryos were derived from a cross between a ZFN founder and a *kdra*$^{um6/+}$ carrier, we expect approximately 50% of *kdra* exon 2 alleles will be wild type and 50% mutant. Deviation from this proportion, or the appearance of more than one lesion would suggest the existence of multiple alleles. **3**. Two distinct mutant alleles were identified in embryos from 883.7F founders: a 382 bp deletion and a 12 bp deletion (see Supplementary Fig. 10c).

Supplemental Table 5. Off-site PCR primer information

Chr (chromosome) and pos (position) refer to coordinates in Zv7

off-target heterodimer primers

Chr	pos	ZFN site	5' primer (ZFN proximal)	3' primer
2	40209400	CACACACTCCACGCGCGGTGGGA	GCGCctgaagAAAACACAAGCTCTAATTGCCTCACA	GCGCctgaagCAGTCATTATGTTTGCATTCACAAGG
24	17282892	CACACCGTCATTAGCCGCTGGGA	GCGCctgaagTGCTCGCTGAATGCGTCTCCACA	GCGCctgaagCTTTAAGGTCCACTGTACAATAAAGG
24	17509645	TCCAACCAGAAATCGAGGGTGTG	GCGCctgaagAGCATTTACATGTTTTTGAGAAAATCCA	GCGCctgaagTGTTACCAATGTTTAATATTTAATTCAGG
5	28435365	TCCCACATCTGTTAGAGGGTGTG	GCGCctgaagTGACCACAAAAAAATCACATACTCCC	GCGCctgaagCTAAATTAATGAATATTGAGCAATTGGG
13	26355883	TCCTACCACTATGTGAGGGTGTG	GCGCctgaagAATTCTTGTCTTCTTTCTCTCCTCCT	GCGCctgaagAAAGAAAAGCAGAAATCATTCACCTAAT
17	48663741	TCCCACCCCTCATGGATAGTGTG	GCGCctgaagTGCTGAGCTGGCATCACAATCCC	GCGCctgaagCAGGTGAAGAGCTTTACAATTACAGA
6	27321737	CACACCCTCTTCGCCTTGTGGGA	GCGCctgaagCAGCAAGCGTGAGCCCTCCACA	GCGCctgaagCCTGCAGCTCTCACATGGGCTG
16	7576157	CACACGTTCTGTGTTAAGGTGGGA	GCGCctgaagTGACAATCTTACACAAACACTCACAC	GCGCctgaagAAAATCTGACCCTCACAAGTACAGTC
16	24055188	CACACCATCCTACCTTTGGTGGGT	GCGCctgaagTTTGTGTATGTATGCATCTAAACACAC	GCGCctgaagAAATGCAAGCCACATAATGTACTGTAC
18	30932339	CACACATTCTGCTTTCTGGTGGGA	GCGCctgaagATTTTCCACCTACTCTTCTCCACAC	GCGCctgaagGATGTGGTTCACTGAACATTTACATTC
10	13078601	CAGACCATCATCATGGTGGTGGGA	GCGCctgaagCATCTACACAGACACACTGCAGAC	GCGCctgaagGAAAACAAATGATGCTGAAATAAGTTAC
5	7717307	TCCCACCCAGGAAGTGATGGTGAG	GCGCctgaagTAAAGAGAGCGCTCATCTCTCCCA	GCGCctgaagCCACCTTCGACATCTCAATAGAGTC
7	31397854	TCCCACCCGAGTCCTGCAGGTGTG	GCGCctgaagCATTACACAAACACAGCTCAATCCCA	GCGCctgaagCGAGCGTAGAGCCCACATGTAAG
2	23206696	TCCCACCTTTGGTTTGATGGTGTC	GCGCctgaagTTCGTCTCAATCAGAGTCAATCCCA	GCGCctgaagTTCATTAAAATGTTGCCTCACTTTCAAA
13	25900641	TCCCATCAATGTTGTGAAGGTGTG	GCGCctgaagCTGTAAAGACATAATAACAGTATCCCA	GCGCctgaagCATTCTGTGAGCAGAAATCCCCCA
12	15778558	TCCCACCAAACACTAGACGTTGTG	GCGCctgaagGTTTTATTTTATTTTGAAGAAACGTCCCA	GCGCctgaagAGTCCGACGTCAATATGTCGACCA
15	30156993	TCCCACCAAACACTAGACGTTGTG	GCGCctgaagTTTGGTCATGTGATATCAACATCCCA	GCGCctgaagAGTCCGACGTCAATATGTCAACCAG

off-target homodimer primers

Chr	pos	ZFN site	5' primer (ZFN proximal)	3' primer
14	30956533	CACACGATCCTGATGACGGTGTG	GCGCctgaagAGGCAAGTTTGTGCCCTGCCACA	GCGCctgaagTCATAAGTTGCATTGATGATGCCCC
7	42818862	CAGACCTTCCAGTTGACGGTGTG	GCGCctgaagTGTTGACTCTTTCCACAACTCCAGA	GCGCctgaagGGTGGCTCTTGCTCTCTGTCAGT
15	7510964	GACACCCTCATCAAGAAGGTGTG	GCGCctgaagCATCGACATCAAGGTCAACAATGACA	GCGCctgaagCAAATTCAACTCTCTAACCTGATTAATT
19	30922679	CACACTGTCCTGGTGAGGGTGTG	GCGCctgaagGATTGGGGGAAAAACATTTACACACA	GCGCctgaagTTATGAAGACACAGCTAATGCATATAG
20	41407585	CACACTCTCACCACTGAGGGTGTG	GCGCctgaagACATCGCCTGTCCTCACTCACAC	GCGCctgaagGCAGGGCCAGTGGAAACTCAGG
18	40217861	TGCCACCATTTGGAGTTGGTGGGA	GCGCctgaagCTGCCAACATTTGTCAACCTGCCA	GCGCctgaagGTTGCCAACCTTGGTATAGTATCCA
19	38699383	CACACGGTCACAACAGACGGTGTG	GCGCctgaagAGAGACGGACATATTAGCCCACAC	GCGCctgaagGGGACCCAACTGTGAATCGGACC
21	17855081	CACACCATCAAGGGAGAGGGTTTG	GCGCctgaagAGCCTTGGGGATCCAAAAACACAC	GCGCctgaagCAGTAGTGTAATGATGTTTGTGTTGTT
5	51241235	TTCCACCAGCTTCTGTTGGTGGGA	GCGCctgaagTCCAATTCGTTCTCCTGCGTTCCA	GCGCctgaagGCCTGTTTGCACCTCCTCAAAGG
4	18062432	CACACCCTCTCACTGGTGCTGTG	GCGCctgaagCGCTGTCTCCCGAGTACCCACA	GCGCctgaagCCCCTGTCTCCACACACAAACTG
4	23336953	CACACCTTCACTGTGAGGATGAG	GCGCctgaagATGTCTCTTCACAGCAAGCCCACA	GCGCctgaagAATACACAAGGTGCATTTATGATCTAAG
4	29941926	CACACACTCATTCGTACGGTGTG	GCGCctgaagCAGATGCTTGCCTAGATTGAACACA	GCGCctgaagTCTTTCCCCTCGTCTATCGATGTG
5	21630841	CACAGCCTCAGACAGATGATGTG	GCGCctgaagCAGAAACCACAAGCTGTTTATACACA	GCGCctgaagAGACACAGAGGAATAAACCAAAACTG
5	24806109	CACACCCACAAATTCAGGGTGTG	GCGCctgaagCTGAGTTGTCCCTAGGGGTCACA	GCGCctgaagTCATACACTGCTTGATCTGCCAAGG
2	5408316	CACACCGTCTCTTTGAAGGAGTT	GCGCctgaagAAATGCTAGAGATCCTGCACACACA	GCGCctgaagTATGATGTTGTTCTTGACTTCATCCTT
2	29953473	TACCACCACGCACCTTGGTGGGT	GCGCctgaagTGGCCAGATGGCAGTTTTTGTACC	GCGCctgaagAACACGTTCCGTGCAGGTGACGC
2	44624948	CACAACCTCAGAAAGAAGGAGTG	GCGCctgaagTGAGAACATTTGGAGGTTCACCACA	GCGCctgaagAAGTTTTGGCAAGCACATTATCCCTG
1	20116979	TCCCACCATCACACTATGTGGGA	GCGCctgaagTGCTTTGACAGCAGTTACACTCCC	GCGCctgaagGATTAGGGAGTTAAACAGAGGGAAAT
1	55323	ATCAACATCTACAGGACGGTGTG	GCGCctgaagGGCGCTGTTGGTCCAGTACATCA	GCGCctgaagGCGAGGCTGGGGTCAGGCCC
3	5485267	CTCACCTTCTGGCAGACGCTGTG	GCGCctgaagTTATGTATCAGGGATGAAAACTACTCA	GCGCctgaagTTTCAATACAAAAGTGGAAAGATGGCT
3	11650710	CTCTCCGTCTCGTGGATGGTGTG	GCGCctgaagTGCAGGGTGGTGCTGTCGCTCT	GCGCctgaagCAGATTCACTGTGCTGCTCACTCG
3	13074803	CACAGTATCCGCTTGACGGTGTG	GCGCctgaagGAGACACTCCCAATATGAGTATCACA	GCGCctgaagAATATCAATCTAAAATCGATAAAAGTCCT
3	28470201	CACAACATCACAGAGAAGCTGTG	GCGCctgaagCCCTCACAGAGTCTGGCATCACA	GCGCctgaagCTCACACAAAAACACATTAAAAGTATCA
3	43597602	CACACAATATGGTGGAAGGTGTG	GCGCctgaagTATCTACTTTGTCTTTGAGGAGCACA	GCGCctgaagGTGTCACCCCATCTTCTTCATCTTC

Supplementary Methods.

Additional Zinc Finger Library information

Putative high confidence specificity determinants:

DNA Subsite (5' -> 3')	Recognition Helix Positions -1,1,2,3,5,6	Structure w/ recognition motif (PDB ID)
GNN	XXXX**TR**	1AAY
N**G**N	XXX**H**XX	1AAY
NN**G(g/t)**	**RSD**XXX	1AAY
N**A**N	XXX**N**XX	1MEY& 1G2D

Positions within the recognition helix could be engineered with reasonable confidence when a purine base occurs at a position within the DNA subsite (bold DNA bases and corresponding amino acids, where the number indicates the position of the residue relative to the start of the alpha helix). For example when G occurs at the 5' end of a subsite the residues at positions 5 and 6 of the recognition helix can be fixed as T and R respectively. This particular recognition motif is present in finger 1 of the Zif268 structure. For each recognition motif listed in the table, a ZFP-DNA structure is listed as a Protein Data Bank ID that contains the recognition motif. The lower case residue at the 3' end of the NN**G** subsite indicates the necessity to consider binding site overlap between neighboring fingers. Position 4 of the recognition helix is not considered as it plays a structural role and thus is always encoded as leucine, which is the most common residue found in zinc finger populations at this position.

Creating zinc finger libraries for finger 1 subsite GGAc (TTGGTG**GGAc**) and GTGt (GAAGGT**GTGt**)

Finger 1 modules specific for the GAAc subsite were selected in the context of a 3-finger framework to recognize the target site 5'-GCTGCG**GGA**c-3'. Based on recognition elements within the structure of Zif268, three of the recognition positions (3, 5 & 6) were fixed. The anchor fingers at positions 2 and 3 were derived from the corresponding fingers of Zif268 with the exception that the recognition helix in finger 2 was replaced with RSDELTR (position -1,1,2,3,4,5,6) to provide specificity for a GCG subsite and the recognition helix in finger 3 was replaced with QSSDLQR (position -1,1,2,3,4,5,6) to provide specificity for a GCT subsite

Finger 1 modules specific for the GTGt subsite were primarily designed based on the structure of Zif268, where five of the six positions within the recognition helix (every position but 3) were fixed. Because only a single position was randomized, this initial library was directly used in the PCR-based assembly of the final ZFP library for the second stage of selection.

Each constructed library exceeded the complexity of the encoded sequences: F1-GAA 1.8 $x10^5$ members, F1-GTG $2x10^4$ members.

Oligonucleotides:
F1-GGAc library:
C D R R F S X X X H L T R H I R I H T
CCTGTGACCGCCGCTTCAGCNNKNNKNNKCACCTCACCCGCCACATCCGGATCCACAC
F1-GTG library:
C D R R F S R S D X L T R H I R I H T
CCTGTGACCGCCGCTTCAGCCGTAGCGACNNKCTCACCCGCCACATCCGGATCCACAC
Complementary oligonucleotides:
5′: 5′-GCTGAAGCGGCGGTCA-3′; 3′:5′-
GCCGGTGTGGATCCGGATGTGGCGGGTGAGGTG-3′.

Creating zinc finger libraries for finger 2 subsite GTGg (TTG**GTG**GGAc) and GGTg (GAA**GGT**GTGt)

Finger 2 modules specific for the GTGg subsite were designed and utilized as described for the GTGt Finger 1 subsite above

Finger 2 modules specific for the GGTg subsite were selected in the context of a 3-finger framework to recognize the target site 5'- GCT**GGT**GCG -3'. Based on recognition elements within the structure of Zif268, three of the recognition positions (3, 5 & 6) were fixed. The anchor finger at positions 3 was derived from the corresponding finger of Zif268 with the exception that the recognition helix was replaced with QSSDLQR (position -1,1,2,3,4,5,6) to provide specificity for a GCT subsite.

Each constructed library exceeded the complexity of the encoded sequences: F2-GGT $5.8x10^5$ members and F2-GTG $1.3x10^4$ members.

Oligonucleotides:
F2-GGT library:
C M R N F S X X X H L T R H I R T H T
CTGCATGAGGAACTTCAGCNNKNNKNNKCACCTCACCCGCCACATCCGCACACACAC
F2-GTG library:
C M R N F S R S D X L T R H I R T H T
CTGCATGAGGAACTTCAGCCGCAGCGACNNKCTCACCCGCCACATCCGCACACACAC
Complementary oligonucleotides:
5′: GCTGAAGTTCCTCAT;3′: GCCGGTGTGTGTGCGGATGTGGCGGGTGAG

Creating zinc finger libraries for finger 3 subsite TGGg (**TTG**GTGGGAc) and GAAg (**GAA**GGTGTGt)

Finger 3 modules specific for the TTGg subsite were selected in the context of a 3-finger framework to recognize the target site 5'- **TTG**GCGGCG -3'. Based on recognition elements within the structure of Zif268, three of the recognition

positions (-1, 1 & 2) were fixed. The anchor fingers at positions 2 and 3 were derived from the corresponding fingers of Zif268 with the exception that the recognition helix in finger 2 was replaced with RSDDLTR (position -1,1,2,3,4,5,6) to provide specificity for a GCG subsite.

Finger 3 units specific for the GAAg subsite were selected in the context of a 3-finger framework to recognize the target site 5'- **GAA**GCGGCG -3'. Based on recognition elements within the structure of Zif268, three of the recognition positions (3, 5 & 6) were fixed. The anchor fingers at position 2 was derived from the corresponding finger of Zif268 with the exception that the recognition helix was replaced with RSDDLTR (position -1,1,2,3,4,5,6) to provide specificity for a GCG subsite.

Each constructed library exceeded the complexity of the encoded sequences within the recognition helix : F3-GAA $5.7x10^5$ members and F3-TTG $2.3x10^5$ members.

Oligonucleotides:
F3-GAA library:
C G R K F A X X X N L T R H T K I H
CTGCGGCAGGAAGTTCGCGNNKNNKNNKAACCTCACCCGCCACACCAAGATCCACA
F3-TTG library:
C G R K F A R S D X L X X H T K I H
CTGCGGCAGGAAGTTCGCGCGCAGCGACNNKCTCNNKNNKCACACCAAGATCCACA
Complementary oligonucleotides:
5': CGCGAACTTCCTGCC;3': CCCGTGTGGATCTTGGTGTG

pH3U3 reporter vectors

Binding sites for each ZFP were cloned into the pH3U3 reporter vector between the BspEI and EcoRI such that the 3' end of the binding site was located 10 bp upstream of -35 box of the promoter.
Example:
Top oligonucleotide: CCGGT**GCGGCGGGA**CTGTG
Bottom oligonucleotide: AATTCACA**GTCCCGCCGC**A

ZFP library assembly primers

Primers used for single finger extraction:
Finger 1
F1forward CAAGAGCAGGAAGCCGCTG
F1reverse CACTGGAAGGGCTTCTGGCCTGTGTGAATCCGGATATG
Finger 2
F2forward
TATCCGGATTCACACAGGCCAGAAGCCCTTCCAGTGTCGCATCTGC
F2reverse ATGTCGCATGCAAAAGGCTTCTCGCCTGTGTGGGTGCGGATGTG
Finger 3

F3forward: CGAGAAGCCTTTTGCATGCGACA
F3reverse: GTAGGATCCACCTGTATGGATTTTGGTGTG

Off-target site identification

Off-target sites with the highest propensity to be recognized by the exon 2 ZFNs were identified by searching the genome (Zv7 repeat-masked) for matches to the consensus sequence for each ZFP based on their determined specificity ('GAXGGTGTG' and 'XXGGTGGGA', where X allows any base) with the appropriate spacing (5 or 6 bp) and orientation using a Perl algorithm. Heterodimer off-target sites were derived from sites that match 14 of 15 bp (*kdra* exon 2 is the only 15 of 15 bp match). Homodimeric sites were derived from sites that match either the 'GAXGGTGTG' composite site at 14 or 15 out of 16 bp or the 'XXGGTGGGA' composite site at 14 of 14 bp. A list of these sites is provided in Supplementary Table 5.

Off-target site lesion adaptor seqeunces

Barcodes were appended to oligonucleotides adaptor sequences originally designed for Solexa sequencing (© 2006 Illumina, Inc, All rights reserved). The sequences for each adaptor pair are as follows (underlined nucleotides indicate barcode; “p” indicates 5’ phosphate):

Adapter strand1-AA
p-**AA**AGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Adapter strand2-AA
ACACTCTTTCCCTACACGACGCTCTTCCGATCT**TT**T

Adapter strand1-AC
p-**AC**AGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Adapter strand2-AC
ACACTCTTTCCCTACACGACGCTCTTCCGATCT**GT**T

Adapter strand1-AG
p-**AG**AGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Adapter strand2-AG
ACACTCTTTCCCTACACGACGCTCTTCCGATCT**CT**T

Adapter strand1-AT
p-**AT**AGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Adapter strand2-AT
ACACTCTTTCCCTACACGACGCTCTTCCGATCT**AT**T