Supplementary Methods:

Binding site analysis using Illumina Sequencing

Binding site selections were performed using the 28 bp randomized library as described previously (11,36). The colonies surviving on selective media plate containing 5mM 3-AT, 10uM IPTG were counted and then washed off the plate. The plasmid DNA from the pooled colonies was isolated and the binding site was PCR amplified using Phusion (NEB) enzyme starting with 50ng plasmid DNA as template. The PCR reaction conditions were as follows: 98°C 3 min; 25 cycles (98°C, 20 sec.; 60°C, 20 sec.; 72°C, 30sec); 72°C 5min. The primer sequences employed were:

Forward:

CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGTGAACGCTCTCCTGA

GTAGG

Reverse: CTGCTCTGTCATAGCTGTTTCC

One of the Solexa adapter sequences (adapter-P2) was incorporated into the forward primer so that only single adapter ligations were required (see below). The PCR product (~1ug) was digested with 40 units of *Eco*RIHF (NEB) enzyme at 37°C and gel purified. The purified DNA was treated with Klenow Exo$^-$ (NEB) in the presence of 0.1mM dNTPs for 45min at 37°C. The DNA was spin purified using QIAquick PCR purification kit (Qiagen) and barcoded adapters (Supplementary Table 2) were ligated using 20 units of T4 DNA ligase (NEB) at room temperature for 2hrs. **Important: the barcoded adapters should <u>not</u> be 5' phosphorylated so that they will ligate to only the *Eco*RI digested end of**

the DNA molecule. Following ligations, the DNA was PCR amplified (Phusion polymerase (NEB)) using in-house Illumina primers, where 20% of the ligation reaction was used as the starting template. The reaction conditions are as follows: 98°C 3 min; 6 cycles (98°C, 20 sec.; 60°C, 20 sec.; 72°C, 30sec); 72°C 5min. The primer sequences employed were:

In-house Illumina P1 (Invitrogen):

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCT

In-house Illumina P2 (Invitrogen):

CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT

Following PCR of the barcoded-adapter-ligated DNA, 20% of the reaction was run on an agarose gel containing ethidium bromide to estimate DNA quantity. An equal amount of PCR-amplified DNA (~80ng) from different barcoded-adapter ligated samples was pooled. The pooled DNA was run on gel and the adapter-ligated fraction of the DNA was gel purified. This purified DNA (~25ng) was amplified by PCR using the Illumina Genomic DNA primers (1.1 and 2.1) and Phusion polymerase (NEB) using the follows conditions: 98°C 3 min; 9 cycles (98°C, 20 sec.; 60°C, 20 sec.; 72°C, 30sec); 72°C 5min. The PCR sample was gel purified and sequenced at 4pM concentration. The sequences were binned according to their barcode. For each ZFP, unique sequences were isolated and ranked according to their frequency of occurrence. The top most frequent sequences (equivalent to the number of surviving colonies on the selective media plates) were analyzed by MEME to discover the recognition motif (35). The

aligned sequences were then used to generate a Sequence logo using WebLogo (37).

Illumina sample preparation for lesion analysis at "Off-Target Sites I"

For the **"Off-Target Sites I"** group and the on-target site (target site run 1)**,** the Illumina sample was prepared as described previously (11). Genomic DNA from the following 7 samples were used: uninjected, oZFPs DD/RR 10pg normal Replicate 1, oZFPs DD/RR 10pg deformed Replicate 1, oZFPs DD/RR 20pg deformed Replicate 1, nZFPs DD/RR 10pg normal, nZFPs DD/RR 20pg normal, nZFPs DD/RR 20pg deformed. The number of embryos used for isolating the genomic DNA is given in Supplementary Figure 2. The sequences of the primers used to PCR DNA flanking the off-target sites as well as the number of sequences with wild type sequence and indels recovered at each site are provided in Supplementary Table 1.

Primer design for amplifying off-target sites

The sequences for primers used to amplify the DNA flanking the off-target sites are listed in the Supplementary Table 1. For each primer pair, the proximal primer binds neighboring the off-target site while the distal primer binds ~150bp away. The proximal primer contains an *Acu*I restriction site that following cleavage allows the Illumina adapter to be ligated within a few basepairs of the center of the putative off-target site. The distal primer contains the Illumina adapter (P2) sequence, and consequently Illumina adapter ligation is only necessary at the proximal end.

Solexa Sample preparation for the Off-Target Sites II

Lesion analysis at “Off-Target sites II” group (10g and non-10g) and the on-target site (target site run 2) was performed for the following 14 samples: uninjected, oZFPs DD/RR 10pg normal Replicate 1, oZFPs DD/RR 10pg deformed Replicate 1, oZFPs DD/RR 20pg deformed Replicate 1, oZFPs DD/RR 10pg normal Replicate 2, oZFPs DD/RR 10pg deformed Replicate 2, oZFPs DD/RR 20pg deformed Replicate 2, nZFPs DD/RR 10pg normal, nZFPs DD/RR 20pg normal, nZFPs DD/RR 20pg deformed, oZFPs EL/KK 50pg normal, oZFPs EL/KK 50pg deformed, oZFPs EL/KK 100pg normal, oZFPs EL/KK 100pg deformed. Genomic DNA was isolated from the ZFN-injected or uninjected embryos 24-hpf using DNeasy Blood and Tissue Kit (Qiagen). The number of embryos used for each group is listed in Supplementary Fig 2. Using the isolated genomic DNA as template, the DNA flanking the off-target sites was PCR amplified with the primers listed in Supplementary Table 1. For each of the 14 conditions mentioned above, PCR-amplified DNA for all 96 off-target sites was pooled. The pooled DNA for the off-target sites was digested with 25 units of *Acu*I restriction enzyme for 4hrs at 37°C and gel purified using QIAquick Gel Extraction kit (Qiagen). The purified DNA was treated with 1 unit of T4 DNA polymerase at 12°C for 15min in the presence of 0.1mM dNTPs to polish the 3’ overhangs (from *Acu*I digestion). The reaction was stopped immediately after 15min by adding EDTA (final concentration of 10mM) and heating it to 75°C for 20min. DNA was then spin purified using Qiaquick PCR purification kit (Qiagen) and was treated with Klenow Exo$^-$ for 45min at 37°C in the presence of 0.1mM dATP to add 3’ A overhangs. The DNA was spin purified as described above and the barcoded

adapters were ligated as described above (**Binding site analysis using Illumina Sequencing)**. The 14 off-target samples were ligated to the 2bp-barcoded-adapters and PCR amplified with in-house Illumina primers as described above. 20% of the PCR reaction was run on an agarose gel containing ethidium bromide to estimate DNA quantity. An equal amount of the PCR-amplified DNA (~80ng) from 14 barcoded-adapter ligated samples was pooled (~1ug total).

For the on-target site lesion analysis, the DNA flanking the *kdrl* ZFN site was PCR-amplified from the genomic DNA samples from the uninjected or ZFN treated embryos (14 samples in total) using the following primers:

kdrex2 solexa on-site 5p: CCTGATCCACAACTGCTTCCTGATGGATATCCAC

kdrex2 solexa on-site-P2 3p:

CGGCATACGAGCTCTTCCGATCTATAAAGTGGCCATTGAACGTAGATGCAC

The PCR amplified DNA was digested with *Eco*RV restriction enzyme and gel purified. The purified DNA was treated with Klenow Exo$^-$ as above and spin purified. The 14 on-target samples were ligated to the 2bp-barcoded-adapters, PCR amplified with in-house Illumina primers and pooled as described above. ~10ng of the pooled on-target DNA was added to ~1ug of the pooled off-target DNA and was run on gel. The adapter-ligated DNA was gel purified and used as a template for PCR amplification with Illumina primers (1.1 and 2.1) as described above. The PCR sample was gel purified, combined with the off-target site pool at the appropriate ratio, and then was sequenced at 4pM concentration.

Comparisons of the off-target lesion frequency for different ZFN treatments

To compare the difference in lesion frequency in embryos from any two ZFN treatments (oZFN 10pg normal vs oZFN 10pg deformed; oZFNs vs nZFNs and oZFNs^{DDRR} vs oZFNs^{ELKK}) the number of reads with wild type sequence for all analyzed off-target sites was combined as were all of the indel reads. This provided an overall lesion frequency that could be compared between different ZFN treatments using the Chi-square test.

Comparing the distribution of active and inactive off-target sites for the number of matches to the target sites

Each heterodimeric off-target site containing a 5-or 6-bp spacing was scored based on the number of matches to the target site (considering only ZFPL and ZFPR half-sites) and, based on the criteria given above, off-target sites were divided into active and inactive off-target sites. Kendall correlation tests were performed to determine the significance of the correlation between the number of active sites and the number of inactive sites, and the significance of the correlation between the number of matches to the target site and the ratio of active sites. In addition, a two-tailed Fisher Exact Test was performed to determine whether the relative number of active off-target sites from two different groups of ZFN-treated embryos was significantly different.

Comparing the distribution of active and inactive off-target sites for the number of conserved Guanines

Each heterodimeric off-target site containing a 5-or 6-bp spacing was scored based on the number of positions that had a guanine found in the target sequence (GXXGGXGXG and XXGGXGGGX; 10Gs) and, based on the criteria

given above, off-target sites were divided into active and inactive off-target sites.

Frequency Plots for ZFPL and ZFPR binding sites from Inactive, Active and All off-target sites

The ZFPL and ZFPR half-sites were extracted from the relevant (active, inactive and all) groups of heterodimeric off-target site containing a 5-or 6-bp spacing and aligned using MEME (35). Frequency-logos for each group of half-site sequences were generated using Weblogo (37).

Germline transmission of off-target lesions

To assay germline transmission of off-target lesions, we crossed the *kdrl* founder zebrafish (889.7) obtained in Meng *et. al.* with wild type zebrafish, which yielded 33 surviving embryos. These embryos were genotyped for the kdrl ZFN target site using the NspI digestion assay (11). Out of 33 embryos, 17 were found to be heterozygous for the *kdrl* mutation whereas 16 did not carry the mutant allele. The genomic DNA from the 17 heterozygous and 16 homozygous embryos were pooled as two separate groups. To identify any lesions at the off-target sites, the *kdrl* ZFN active off-target sites were PCR amplified as described above from the two genomic DNA pools. The following off-target sites were analyzed: OT1, OT2, OT3, OT4, OT5, OT6, OT7, OT8, OT10, OT11, OT13, OT14, OT15, OT16, OT18, OT20. The target site was also included in the analysis. The Solexa sample was prepared and sequenced as described above. The sequences for each site were binned for analysis using the unique 9 bp “prefix” described above.

Supplementary Table 1: The table gives the sequences and the genomic information for the 141 off-target sites analyzed in this study. The primer sequences used to amplify each of these sites are provided. Number of reads with wild type (W), insertions (I) and deletions (D) for each condition at each of the site is also provided. Reads that did not contain the suffix sequence were marked as undefined (U). The BH adjust p-value is listed for each site and condition (BH). Sites with significant cleavage under one treatment condition are colored as described for Table 1. The target site is colored cyan and was assayed two separate times for some conditions.

Supplementary Table 2: Adapter sequences used for barcoding the Illumina sequencing sample.

Supplementary Figure 1: Proportion of ZFN-treated embryos with different morphology at 24hpf. The number in parentheses represents the total number of embryos used for analysis. The activity of ZFNs containing two different sets of ZFPs were compared the original ZFNs (oZFNs) (11) and new ZFNs (nZFNs) generated in this study.

Supplementary Figure 2: The RFLP analysis performed for ZFN-treated embryos. Genomic DNA from untreated or ZFN-treated embryos was pooled. The DNA flanking the *kdrl* target site was PCR-amplified (11) and digested with *Nsp*I enzyme which cleaves the unaltered spacer region within the ZFN target site resulting in two bands each ~110 bp in size. Indels within the spacer region at the target site result in the loss of *NspI* site, which is detected as the presence of undigested band of 220 bp in size.

Supplementary Figure 3: Pearson correlation analysis of the lesion log odds ratio for 41 identical sites between three identical treatment groups from different biological replicates. After excluding sites where the odds ratio was 0, there are total of 71 dots in the plot with each dot representing a pair of log odds ratios from replicate 1 and replicate2 that have the same ZFN dose at the same site. The three treatment groups are Original ZFNs DD/RR 10pg Monsters, Original ZFNs DD/RR 10pg Normals and Original ZFNs DD/RR 20pg Monsters. A significant correlation exists between the replicates ($r = 0.78$, p-value= 8.9e-16) where the red dots indicate sites with a significant increase in lesion frequency in either replicate 1 or replicate 2 (BH adjusted p-value <0.05) as compared to the uninjected control.

Supplementary Figure 4: The distribution of the type of indels observed at the off-target sites (A) and the target site (B).

Supplementary Figure 5: Enrichment of Guanine-contacts in the active off-target sites. Base frequency at each position in the ZFPL and ZFPR binding sites are displayed as a logo for (top) the group of 18-active off-target sites, (middle) the group of 41-inactive off-target sites, and (bottom) all 59 sites together. Guanines at seven positions (red boxes) in the active off-target sites were absolutely conserved within the active sequences, but are more variable in the inactive sites.

Supplementary Figure 6: High stringency analysis of importance of positions within each ZFP binding site. Each base in the binding site of the ZFPL and ZFPR was independently mutated to cytosine and its influence on ZFP binding

was assayed using B1H-based activity assay (36) at 10 mM 3-AT to detect more subtle influences on recognition, where a reduction in cell survival (plotted as the -log of surviving colonies) indicates a position important for recognition. All of the conserved guanines - indicated by an asterisk (*) - are critical for activity.

Supplementary Figure 7: Influence of the type of the engineered nuclease domain (DD/RR or EL/KK) on the precision of the original ZFNs. The lesion frequencies for normal embryos treated with 10 pg dose of oZFNs^{DDRR}, 50 pg dose of oZFN^{ELKK} or 100 pg dose of oZFN^{ELKK} were plotted for the on-target site and a subset (6 of 8) of the active off-target sites for oZFNs^{DDRR} that were assayed in this experiment. Asterisks indicate 2 positions in the 50 pg oZFN^{ELKK} where there were insufficient sequencing reads to provide a confident assessment of the lesion frequency (none were observed). Normal embryos treated with 50 pg of oZFNs^{ELKK}, like the 100 pg of oZFNs^{ELKK} embryos, displayed a significantly lower average off-target lesion frequency (0.1%) than normal embryos treated with 10 pg oZFNs^{DDRR} (0.3%, p-value < 0.0001) among the active off-target sites with sufficient number of sequencing reads.

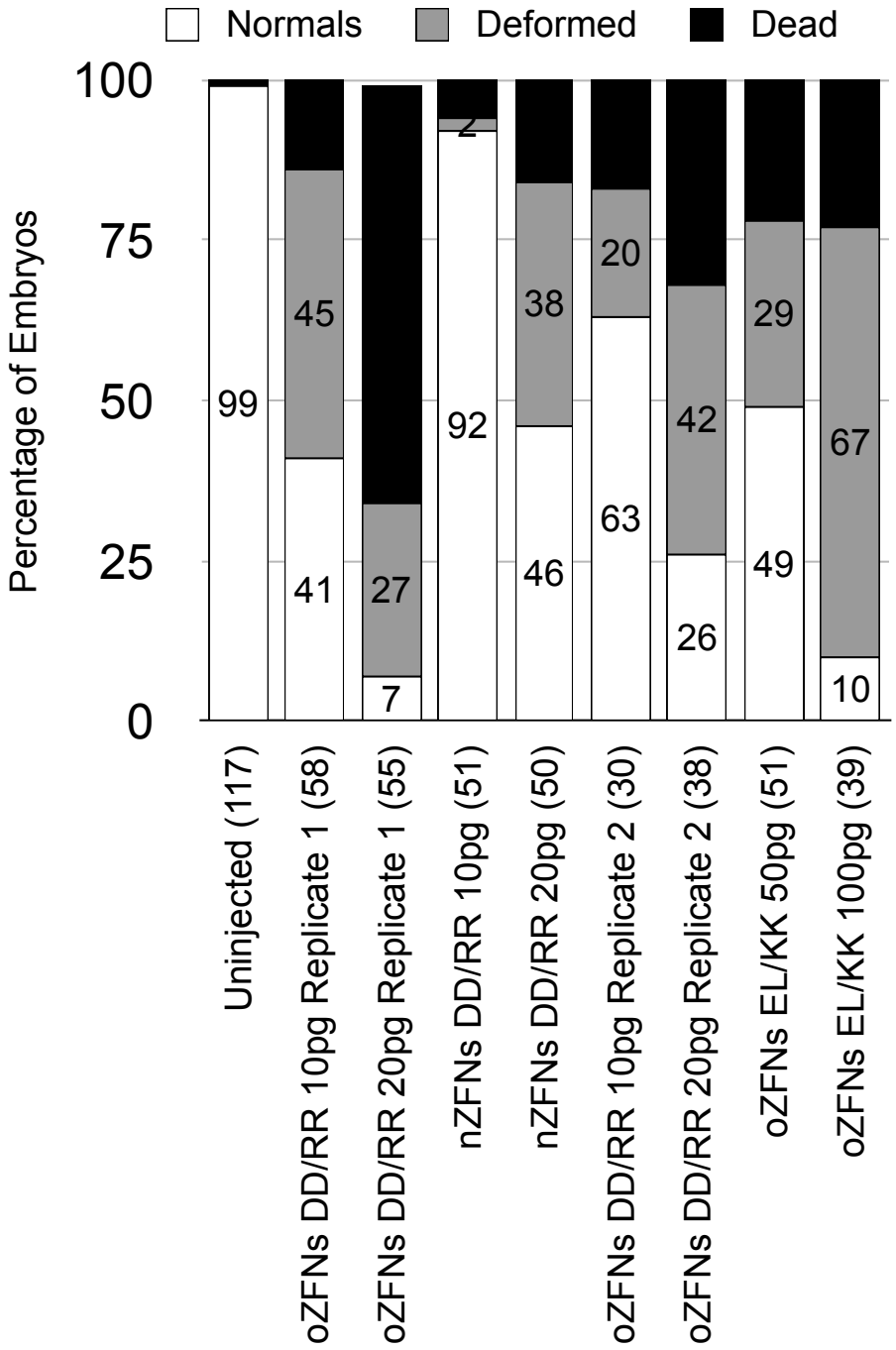
SUPPLEMENTARY TABLE 1

ATTACHED AS AN EXCEL FILE

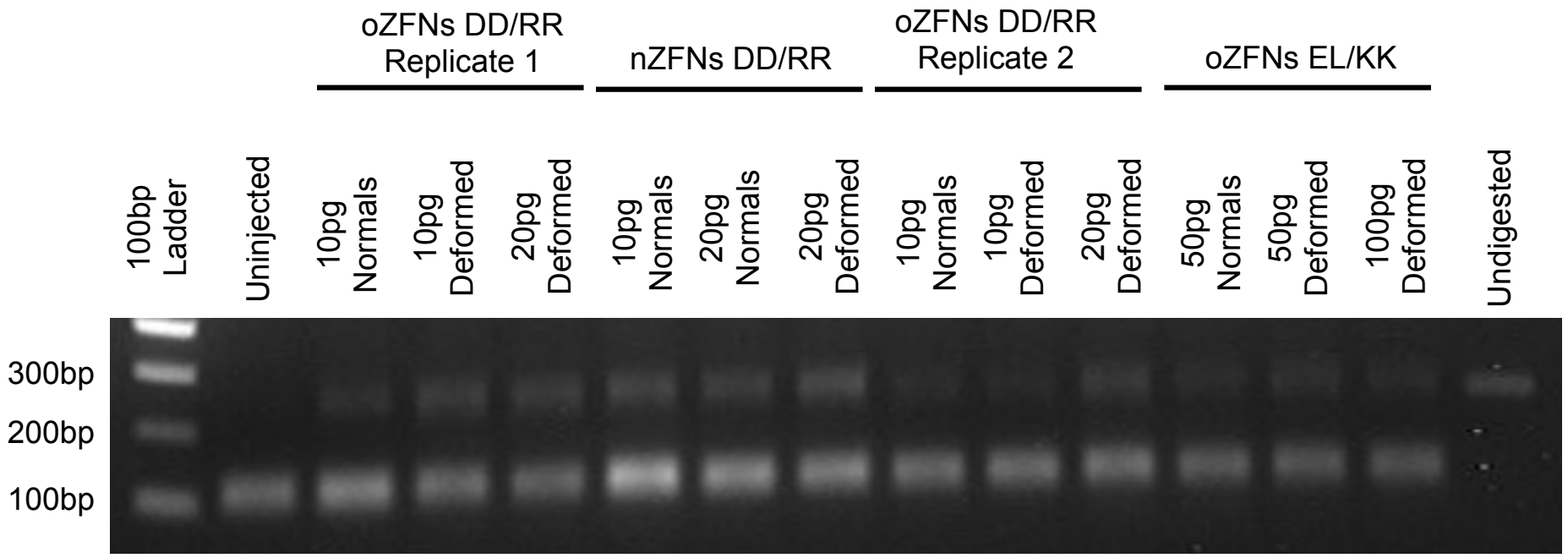
SUPPLEMENTARY TABLE 2

Barcode	Strand 1 Sequence (no phosphorylation)	Strand 2 Sequence (no phosphorylation)
TT	*aa*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*tt*T
TG	c*a*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*tg*T
TC	*ga*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*tc*T
TA	t*a*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*ta*T
GT	*ac*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*gt*T
GG	cc*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*gg*T
GC	*gc*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*gc*T
GA	tc*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*ga*T
CT	*ag*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*ct*T
CG	c*g*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*cg*T
CC	*gg*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*cc*T
CA	t*g*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*ca*T
AT	*at*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*at*T
AG	ct*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*ag*T
AC	*gt*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*ac*T
AA	tt*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT*aa*T

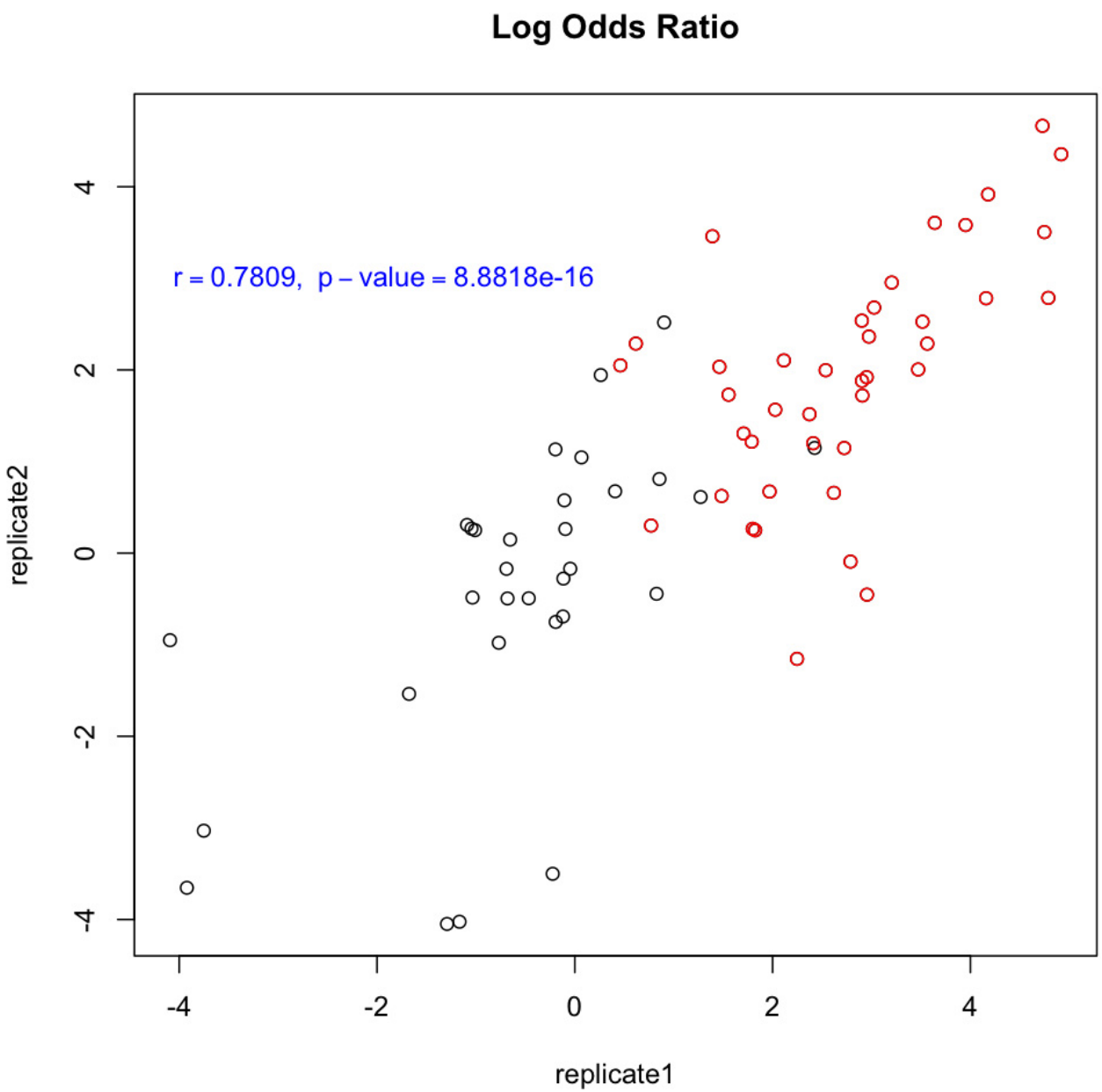
SUPPLEMENTARY FIG 1



SUPPLEMENTARY FIG 2



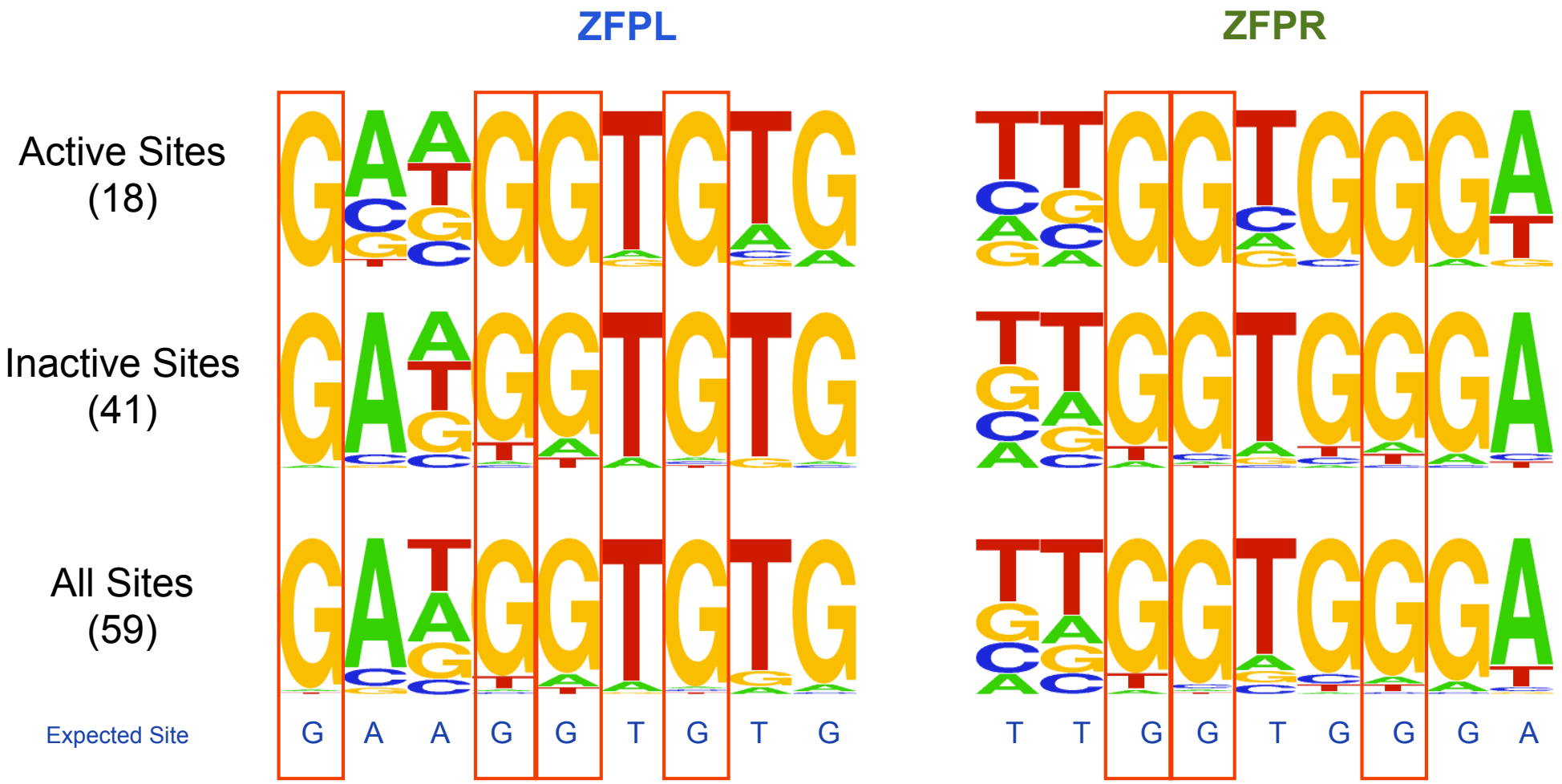
SUPPLEMENTARY FIG 3



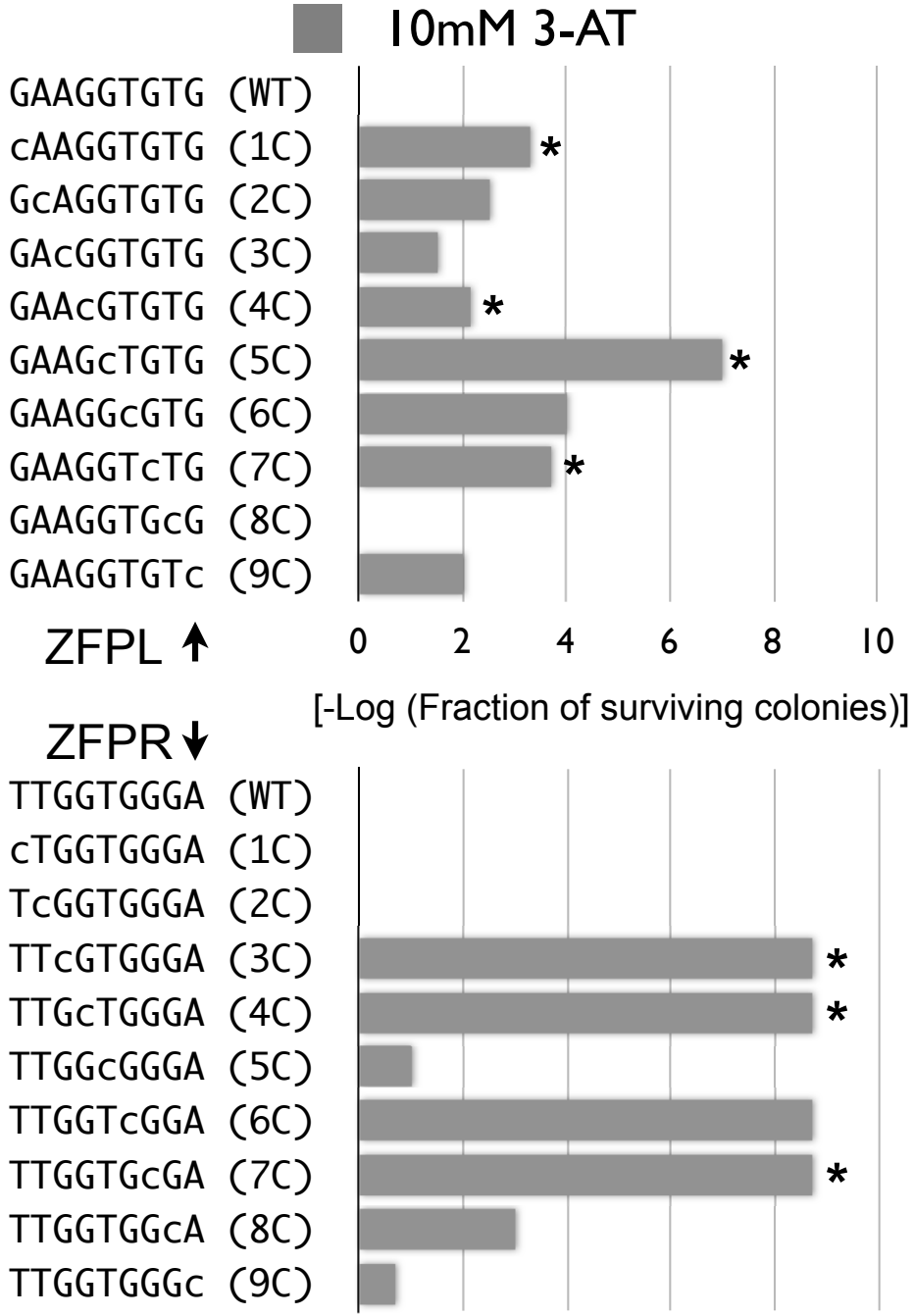
SUPPLEMENTARY FIG 4A

SUPPLEMENTARY FIG 4B

SUPPLEMENTARY FIG 5



SUPPLEMENTARY FIG 6



SUPPLEMENTARY FIG 7