An optimized two-finger archive for ZFN-mediated gene targeting

Ankit Gupta1,2, Ryan G Christensen3, Amy L Rayla1, Abirami Lakshmanan1, Gary D Stormo3 & Scot A Wolfe1,2

The widespread use of zinc-finger nucleases (ZFNs) for genome engineering is hampered by the fact that only a subset of sequences can be efficiently recognized using published finger archives. We describe a set of validated two-finger modules that complement existing finger archives and expand the range of ZFN-accessible sequences threefold. Using this archive, we introduced lesions at 9 of 11 target sites in the zebrafish genome.

ZFNs are artificial restriction enzymes containing a zinc-finger array (ZFA) engineered to recognize different DNA sequences, fused through a flexible linker to the nuclease domain of FokI (ref. 1). These enzymes function as heterodimers to create site-specific double-strand DNA breaks. This technology has been applied in a variety of cell lines and organisms that previously lacked efficient tools for targeted genome editing1. However, widespread use of ZFNs is hindered by the challenge of designing ZFAs with sufficient affinity and specificity for most DNA sequences in a genome.

Highly specific ZFAs can be selected from randomized finger libraries using phage or bacterial selection systems2−5, but this process is labor-intensive. By contrast, modular assembly6−8, wherein precharacterized single zinc finger modules that recognize 3-base-pair (bp) subsites are joined into ZFAs, rapidly yields ZFNs, albeit with lower success rates9 presumably owing to unfavorable 'context-dependent' interactions at the finger-finger interface10. Efforts to generate more reliable tools for ZFA assembly (Supplementary Discussion 1) have focused on randomizing interface sequences for the selection of two-finger modules5,11 or selecting compatible finger pairs2−4. Analogous two-finger modules have been used by Sangamo BioSciences to build highly specific ZFAs and active ZFNs1,12, but the archive is proprietary, limiting its use to ZFNs purchased through Sigma-Aldrich. Recently, the Zinc Finger Consortium (ZFC) described a context-dependent assembly (CoDA) approach, whereby two-finger modules selected from oligomerized pool engineering (OPEN) pools are assembled into three-finger ZFNs13. CoDA-derived ZFNs constructed from prescreened ZFAs had high success rates (~50%), but the assayed ZFNs were almost entirely constructed from two-finger modules that recognize GNNGNN 6-bp sites, where N is any nucleotide, and junction positions are underlined (Supplementary Fig. 1). These ‘N-G’-type junctions are not the limiting factor in effective ZFN design; ZFAs composed of only GNN recognition modules generate the most reliable ZFNs when using a variety of single-finger archives8,9.

Here we describe an archive of two-finger modules that bind all possible 2-bp junctions at the finger-finger interface, for assembly with each other or with predefined one-finger modules to create active, multifinger ZFNs. We isolated these two-finger modules from two orthogonal libraries containing randomized amino acids at the interface recognition positions, via bacterial one-hybrid (B1H) selection against all 16 GANNCG target sites, where adenine at position 2 of the target site is recognized by either asparagine at position +3 of finger 2 (‘Asn+3F2’ library) or histidine at position +3 of finger 2 (‘His+3F2’ library) (Fig. 1a,b). Analysis of clones recovered from selections for 30 of 32 library–target site combinations yielded a partial or full consensus sequence at the randomized positions, implying the recovery of motifs compatible with sequence-specific DNA recognition. Higher stringency yielded a more constrained consensus for seven of these selections (Supplementary Table 1).

To confirm the target specificity of the B1H-selected two-finger modules, we determined the DNA-binding specificity of 87 two-finger modules using our constrained variation B1H (CV-B1H) method14 (Supplementary Fig. 2). For 19 of 32 junctions, including 11 of 24 non-‘N-G’ junctions, we identified two-finger modules that preferred the desired binding site (Supplementary Fig. 3). For an additional 7 junctions we identified two-finger modules with specificity compatible to the desired target site but that preferentially recognize an alternate DNA sequence. Two-finger modules derived from selections with a more refined consensus sequence at higher stringency displayed improved sequence selectivity (Supplementary Fig. 4).

We realized additional improvements in specificity for five junction sequences by rational design of complementary two-finger modules, applying principles of DNA recognition derived from our B1H-based selections and from previous studies. In total, 24 junction sequences can be preferentially recognized by our two-finger modules, with an additional six junctions that can be recognized by two-finger modules with ‘compatible’ specificity (Fig. 1c and Supplementary Fig. 5; see Online Methods for rating of specificity). These two-finger modules recognized 60 GRNNYG 6-bp sites, where R is adenine or guanine (recognized by asparagine or histidine, respectively,

---

1Program in Gene Function and Expression, University of Massachusetts Medical School, Worcester, Massachusetts, USA. 2Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts, USA. 3Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA. Correspondence should be addressed to S.A.W. (scot.wolfe@umassmed.edu).

RECEIVED 7 DECEMBER 2011; ACCEPTED 7 APRIL 2012; PUBLISHED ONLINE 29 APRIL 2012; DOI:10.1038/NMETH.1994

588 | VOL.9 NO.6 | JUNE 2012 | NATURE METHODS
at position +3 of finger) and Y is cytosine or thymine (recognized by threonine at position +3 of finger 1).

Although some of our two-finger modules contain previously observed residues at the finger-finger interface, many contain new combinations. Based on these interface sequences, some CoDA-derived two-finger modules previously described as recognizing ‘N-A’ junctions might prefer other junction sequences. We assessed sequence preferences for five of these CoDA-derived two-finger modules using B1H binding site selections and activity assays. These CoDA-derived modules preferred ‘N-G’ to ‘N-A’ junctions, highlighting the advantage of explicit optimization of the finger-finger interface for generation of highly specific ZFAs. To generate additional modules targeting new hexameric sites, we substituted constant positions in our selected two-finger modules. In many instances, desired alterations in specificity could be obtained through substitutions at specific determinant positions or through substitution of sets of residues (at positions –1, 1 and 2 relative to the start of the α-helix) at the N terminus of finger 1 (Supplementary Fig. 7). These modifications expanded the archive of two-finger modules to encompass 162 unique 6-bp sites including 132 non–‘N-G’ junction-containing sites (Supplementary Table 2).

To demonstrate the utility of these two-finger modules for gene disruption, we combined them with each other or with published one-finger modules to create ZFAs (three pairs of three-finger ZFAs and eight pairs of four-finger ZFAs) targeting 11 sites in the zebrafish genome, where each site contains at least one non–‘N-G’ junction (Supplementary Table 3). We determined the DNA-binding specificities of these assembled ZFAs using our B1H system and a 28-bp randomized library, and found that the incorporated two-finger modules displayed the desired DNA-binding specificity (Supplementary Fig. 8). Most of the resulting ZFAs (9 of 11) were active in a yeast-based chromosomal reporter assay (Supplementary Fig. 9). For some four-finger ZFAs, we found that the presence of a noncanonical linker (TGSQKP) between the second and the third finger could both increase ZFN activity and reduce its toxicity, presumably by moderating its activity at nontarget sequences (Supplementary Fig. 10). Finally, we injected mRNAs encoding these ZFN into zebrafish embryos and evaluated the induction of insertions and deletions (indels) at the target site. Nine of 11 ZFAs induced indels (>1 bp) at the target sites at >0.5% frequency (Table 1 and Supplementary Table 4). Consistent with the yeast assay, lesion frequency in zebrafish was also improved for some ZFAs when incorporating a noncanonical linker (Supplementary Table 5 and Supplementary Discussion 2). We assayed ZFN-injected adult zebrafish for germ-line transmission of mutant alleles for four targets and in all cases identified founders from a small number of screened zebrafish (Supplementary Table 6).

Here we report 87 unique validated two-finger modules that recognize 162 6-bp target sites with high specificity, which can be combined together or with available single-finger modules to rapidly create active ZFAs that target sequences containing

| Gene | 5′ ZFA binding site | 3′ ZFA binding site | Spacer length (bp) | Lesion frequency (%) |
|------|---------------------|---------------------|-------------------|---------------------|
| dab2lp | GACTTGacg | GACATGgac | 6 | 8.0 |
| hey2 | GGTATGtt | gggGAACGT | 6 | 0.6 |
| rock1 | gctGGACG | GACTGgcc | 6 | 0.0 |
| zgc77041 | GACTGGAAGGTG | GACTGGTACG | 6 | 15.7 |
| dck2 | GAGCGGAAATTG | GAAACGGAGAT | 5 | 1.1 |
| mc4r | gctgcagataga | gtagacGACTG | 6 | 12.9 |
| ltp8 | gtcgagggaggcc | GAATCGGAGG | 6 | 7.3 |
| mc3r | GACGGAGACCTG | gtcgctgaagag | 6 | 3.1 |
| apoeb | GGGTCGaggtt | gggGGCCGgtg | 5 | 2.8 |
| lep | aagGAATG Tat | gcatggaaggt | 6 | 0.9 |
| irs2a | aagGGATGgata | cagGAAGGAgta | 6 | 0.4 |

5′ ZFA and 3′ ZFA binding sites are shown with two-finger module 6-bp subsites in uppercase. In some cases 2 two-finger modules abut.

*Targets for which a noncanonical linker (TGSQKP) between the second and the third finger was used to increase ZFN activity; the position of the noncanonical linker is underlined in each half-site where it is present.
non–‘N-G’ junctions in vivo. We determined the number of potential ZFN target sites in protein-coding exons in zebrafish and human genomes and compared this with two other recently described two-finger archives for ZFN construction (CoDA and the two-finger archive from ref. 15; Supplementary Table 7 and Supplementary Discussion 1). Our combined archive allows targeting of ~95% of the protein-coding genes in the zebrafish genome (exons from Zv9 assembly), with an average density of one unique ZFN site every ~140 bp, which is an approximately fivefold higher density than available through the CoDA archive. The archive in ref. 15 has the highest targeting density of the three archives with an average of one unique ZFN site every 10 bp, albeit with a lower overall success rate for assayed ZFNs. To facilitate public use of our archive, we developed a web interface that allows users to search for potential ZFN sites in an input sequence (http://pgfe.umassmed.edu/ZFPmodularsearchV2.html). The website ranks the quality of each ZFN site in the input sequence and provides information for the assembly or direct synthesis of the requisite ZFAs (Supplementary Discussion 3).

Although our archive is to our knowledge the largest set of non–‘N-G’ junction–recognizing two-finger modules described to date, it comprises only 132 of the possible 3,072 non–‘N-G’ junction sites. Additional archives of non–‘N-G’ junction two-finger modules exist; for example, 61 are found in the CoDA archive, but we note that only 3 of 10 tested ZFAs containing these modules were active. Thus, there is a need to expand the set of high-quality two-finger modules covering these junctions to increase the targeting resolution of ZFNs.

As we observed, selections alone may not always be sufficient to obtain highly specific modules for a given target sequence because both affinity and specificity have a role in module activity. The continued development of more accurate predictive models of DNA recognition for zinc fingers is likely to be needed to inform design efforts. Ultimately, these efforts should lead to important advances in nuclease precision and activity not only for engineering model systems but also for creating therapeutic reagents for the treatment of disease.

**METHODS**

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

**ACKNOWLEDGMENTS**

This research was supported by the US National Institutes of Health R01GM068110 (S.A.W.), R24GM078369, R01HL093766 (N. Lawson & S.A.W.) and R01HG00249 (G.D.S.). We thank N. Lawson and members of his laboratory for their insightful advice and zebrafish husbandry training, and J. Zhu for assistance with website construction.

**AUTHOR CONTRIBUTIONS**

S.A.W. conceived the study; A.G. and A.L.R. carried out the selection experiments. R.G.C. and G.D.S. developed the computational platform for motif analysis. A.L. performed the analysis of ZFN sites in multiple genomes. A.G. and S.A.W. wrote the manuscript with input from all authors.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nmeth.1994.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. & Gregory, P.D. Nat. Rev. Genet. 11, 636–646 (2010).
2. Greisman, H.A. & Pabo, C.O. Science 275, 657–661 (1997).
3. Maeder, M.L. et al. Mol. Cell 31, 294–301 (2008).
4. Meng, X., Noyes, M.B., Zhu, J.J., Lawson, N.D. & Wolfe, S.A. Nat. Biotechnol. 26, 695–701 (2008).
5. Isalan, M., Klug, A. & Choo, Y. Nat. Biotechnol. 19, 656–660 (2001).
6. Carroll, D., Morton, J.J., Beumer, K.J. & Segal, D.J. Nat. Protoc. 1, 1329–1341 (2006).
7. Kim, H.J., Lee, H.J., Kim, H., Cho, S.W. & Kim, J.S. Genome Res. 19, 1279–1288 (2009).
8. Zhu, C. et al. Development 138, 4555–4564 (2011).
9. Ramirez, C.L. et al. Nat. Methods 5, 374–375 (2008).
10. Sander, J.D., Zaback, P., Joung, J.K., Voytas, D.F. & Dobbs, D. Nucleic Acids Res. 37, 506–515 (2009).
11. Isalan, M., Klug, A. & Choo, Y. Biochemistry 37, 12026–12033 (1998).
12. Doyon, Y. et al. Nat. Biotechnol. 26, 702–708 (2008).
13. Sander, J.D. et al. Nat. Methods 8, 67–69 (2011).
14. Christensen, R.G. et al. Nucleic Acids Res. 39, e83 (2011).
15. Kim, S., Lee, M.J., Kim, H., Kang, M. & Kim, J.S. Nat. Methods 8, 7 (2011).
ONLINE METHODS

ZFN website scoring function. Our new ZFN site identification tool (http://pgfe.umassmed.edu/ZFPmodularcheckV2.html) uses two-finger (2F) modules from this study and one-finger (1F) modules from our previous archive8 to define favorable combinations of these modules for constructing active ZFNs. These ZFNs are designed to target sequences with 5-, 6- or 7-bp gaps between the monomer recognition sequences, where each ZFN monomer can contain three or four fingers. ZFNs with higher scores are more likely to be active, where the current 2F modules are scored based on their DNA-binding specificity (as determined in the B1H system), where good, fair and poor represent 4, 3 and 2 points, respectively. If the modules use an A-cap (Gln-Arg-Gly at the N terminus of the 2F module) instead of the standard RSD sequence for G-recognition, one point is subtracted from the score. The 1F modules were scored as previously described8. ZFNs containing 2F modules are readily identified in the output from the website by the presence of lowercase triplet sequences in the site breakdown, and by the presence of “2FM-#” in the output module identifier information.

Animal husbandry. Zebrafish were handled according to established protocols16 and in accordance with Institutional Animal Care and Use Committee guidelines of the University of Massachusetts Medical School.

2F-library construction. 2F libraries were constructed in two stages. First, individual F1 and F2 libraries were independently constructed via cassette mutagenesis of annealed randomized oligonucleotides into pBluescript vector containing the appropriate zinc-finger backbone sequence derived from Zif268. The sequences for the randomized oligonucleotides are listed in Supplementary Table 8, where lowercase letters denote the randomized bases. Individual finger library diversity greatly exceeded the theoretical library size: ~1 × 10^8 transformed cells were obtained for the F1 library (>100 times theoretical size) and ~1 × 10^8 transformed cells for the F2 library (30 times theoretical size of the library). Constructed libraries were grown at low density on 2×YT agar plates containing 100 µg/ml carbenicillin at 37 °C for 14 h. Individual F1 and F2 libraries in pBluescript were collected from pooled cells from these surviving colonies.

The 2F library was constructed from the single-finger libraries by PCR assembly, individual F1 and F2 libraries were separately amplified from the pooled pBluescript clones by PCR and then joined via overlapping PCR, where the number of amplification cycles in both steps was minimized by using high concentrations of template DNA. This 2F library was then cloned into the B1H expression vector 1352-omega-UV2 between unique BssHII and Acc65I restriction enzyme sites such that a protein is encoded with the 1352-omega-UV2 between unique BssHII and Acc65I restriction enzyme sites. After electro- transfection of the library, 1 × 10^8 cells (5 times the theoretical size of the library) were plated on ten 2×YT agar plates with 100 µg/ml carbenicillin (150 mm × 15 mm) and grown at 37 °C for 14 h. The 1352-omega-UV2 plasmids containing the 2F library were isolated from pooled surviving colonies and used for selections.

Zinc-finger binding site cloning. The 16 GANNCG zinc-finger binding sites (ggccTAATTACCTGANNCGGacg) were cloned between the EcoRI and NotI sites in the pH3U3-mcs reporter vector. The Homeodomain (Engrailed) binding site TAATTA (underlined) is present 3 bp away and on the strand opposite to the zinc-finger binding site to minimize any interference between the Homeodomain and the zinc fingers. For selecting 2F modules from the Asn+3F2 library that recognize the ‘G-G’ interface, sufficient stringency could not be obtained to narrow the selected clones merely through increased 3-aminotriazole (3-AT) concentration or reduced inducer (isopropyl-b-D-thiogalactoside; IPTG) levels. To reduce the activity of the 2F module–homeodomain fusion construct the homeodomain site was mutated to TAAAGG to increase the dependence on zinc-finger binding.

2F B1H-based selections. Selections for 2F modules were performed as described previously4. The zinc-finger library (20 ng) and the reporter vector (1 µg) containing the zinc finger target site were transformed together via electroporation in the selection strain that lacks endogenous expression of the omega subunit of RNA polymerase (US0AhisBΔpyrFΔarpOZ). We plated 2 × 10^7 cotransformed cells were plated on selective NM minimal medium plates (where stringency was controlled via 3-AT and IPTG concentration) and grown at 37 °C until moderate number of colonies (typically hundreds) were visible. After selection, 2F modules from 6–10 surviving colonies were sequenced to identify functional amino acid sequences for subsequent evaluation. The success of the selection was judged on the diversity of sequences obtained from these selections, with the expectation that successful selections will converge on a small number of functional residues at the critical recognition positions.

Cloning B1H-selected 2F modules into 3F F1-GCG constructs. To determine the binding specificities of 2F modules sequence encoding a ‘GCG’-binding anchor zinc finger (recognition helix: RSDTLAR) fused at the N terminus of the 2F module via overlapping PCR (Supplementary Table 8). After overlapping PCR, the 3F-ZFA sequence was cloned into 1352-omega-UV2 vector between the Acc65I and BamHI sites for expression as a fusion to the omega subunit of RNA polymerase.

CV-B1H method. To determine binding-site specificities of 2F modules, the CV-B1H assay was performed as described before14. After transformations into the selection strain, 1 × 10^6 cells containing the zinc-finger plasmid (1352-omega-UV2-ZFP) and the randomized binding site library plasmid (pH3U3) were plated on selective NM minimal medium plates (100 mm × 15 mm) containing 50 µM IPTG and 1 mM or 2 mM 3-AT and grown at 37 °C for 22–30 h. The surviving colonies were pooled and the binding site plasmid was isolated for identification of the functional DNA sequences. The binding site region was PCR-amplified and Sanger-sequenced to rapidly obtain binding site profiles for each 2F module. For quantitative modeling, the binding site pools for multiple 2F modules were barcoded and sequenced via Illumina sequencing, and then binding specificities were modeled from this data using both W log-odds and GraMSz methods (Supplementary Methods).

Rating of 2F modules. For every 2F module, the frequency of each of the 16 possible 2-bp junctions was determined in the binding sites that were recovered by Illumina sequencing.
The 2F modules for which the frequency of the desired 2-bp junction was the highest among all 16 2-bp junctions were designated as possessing 'preferential specificity'. If the frequency of the desired 2-bp junction was the second highest and represented more than 20% of the dinucleotide population, the 2F module was designated as having 'compatible specificity'. The remaining 2F modules were designated as having 'poor specificity'.

Comparison of CoDA-2F modules and B1H-selected 2F modules. The CoDA 2F modules were created using overlapping PCR where the desired recognition helix sequences were introduced into the Zif268 F2 backbone. The 2F modules were fused to the N-terminal ‘GGG’ binding finger and CV-B1H assay was performed followed by binding site modeling using the W log-odds and GraMS methods as described above. B1H-based activity assay were performed as described previously 17.

Rational design of 2F modules with improved specificity. The archive of selected 2F modules was expanded through rational design. For improved recognition of junction sequences, the specificity determinants at the interface positions were altered based on recognition trends that were observed in our selected modules or prior interface selection studies 11. Changing the residue at position 3 of finger 2 or finger 1, respectively, created modules with altered specificity at the second or fifth position of the 6-bp recognition sequence. Substituting the three N-terminal cap residues in finger 1 (RSD at positions −1, 1 and 2) produced alterations in sequence preference at the sixth position, where substitution of a QRG cap reliably produced an alteration in sequence preference from G to A.

Creating ZFAs. Three-finger (3F) and four-finger (4F) ZFAs for use in ZFNs were assembled from the 2F-module archive described herein and a 1F-module archive that we recently described using overlapping PCR. The primer sequences used for these different assemblies are listed in Supplementary Table 9. If desired, these ZFAs can also be synthesized from the DNA sequence output from our website application.

For amplifying individual 1F and 2F modules, the following PCR conditions were used: 10 ng DNA template, 1 µM each of forward and reverse primer, 200 µM dNTPs and 0.5 unit of Phusion High Fidelity DNA polymerase (New England Biolabs) in 25 µl reaction volume. PCR cycles were 98 °C for 3 min, (98 °C for 15 s, 50 °C for 15 s, 72 °C for 30 s) 6 repeats, (98 °C for 15 s, 56 °C for 15 s, 72 °C for 30 s) 24 repeats, 72 °C for 5 min and then 4 °C.

ZF assembly from the individual 1F- and 2F-module amplions was mediated by overlapping PCR under the following conditions: 1–5 ng DNA for each component, 200 µM dNTPs and 0.5 unit of Phusion High Fidelity DNA polymerase (New England Biolabs) in 25 µl reaction volume. PCR cycles were 98 °C for 3 min, (98 °C for 15 s, 50 °C for 15 s, 72 °C for 30 s) 6 repeats and 72 °C for 5 min. After this initial assembly step, forward and reverse primers (final concentration of 1 µM each) were added to the reaction and PCR amplification proceeded using the following cycles: 98 °C for 3 min, (98 °C for 15 s, 56 °C for 15 s, 72 °C for 30 s), 25 repeats and 72 °C for 5 min. After amplification, the 3F or 4F PCR products were digested with Acc65I and BamHI enzymes and cloned into appropriate vectors.

The QRG cap was introduced into the 2F module using a special QRG(X) primer set that substitutes the RSD cap with the QRG cap in F1. When threonine is present at position 3 of F1 use the QRG(T) primer, when asparagine is present at position 3 of F1 use the QRG(N) primer, and when histidine is present at position 3 of F1 use the QRG(H) primer. For ZFNs recognizing a seven base pair gap use the F3RnTGPGAAGS or 2FM-F3RnTGPGAAGS instead of the F3RnLRGS or F3RnLRGS primers to incorporate the longer linker associated with increased activity (Supplementary Discussion 2). To incorporate the noncanonical linker (TGSKQKP) between F1 and F2 fingers of a 4F construct, the sequences for F2-forward primer and F1-reverse primer were modified to introduce the additional serine in the linker.

3F-ZFA assemblies from F1, F2 and F3 1F modules. The single fingers were amplified individually and then assembled. F1 was amplified using the F1(noF0)Fn and F1Rn primers. F2 was amplified using the F2Fn and F2Rn primers. F3 was amplified using F3Fn and F3RnLRGS primers. The amplified DNA was gel-purified using a Qiagen gel purification kit. For finger assembly, 5 ng of the F1, F2 and F3 amplicons were combined and assembled as described above, where the F1(noF0)Fn and F3RnLRGS primers were added to the PCR for the final amplification.

3F-ZFA assemblies from F1 1F module and 2F module. The F1-module sequence was amplified using the F1(noF0)Fn and F1Rn primers. The 2F-module sequence was amplified using 2FM-F2Fn and 2FM-F3RnLRGS primer. The amplified DNA was gel-purified and the finger amplicons were assembled as described above, where the 2FM-F1(noF0)Fn and F3RnLRGS primers were added to the PCR for the final amplification.

3F-ZFA assemblies from a 2F module and F3 1F module. The 2F-module sequence was amplified using the 2FM-F1(noF0)Fn and 2FM-F2Rn primers and F3-module sequence was amplified using the F3Fn and F3RnLRGS primers. The amplified DNA was gel-purified and the finger amplicons were assembled as described above, where the 2FM-F1(noF0)Fn and F3RnLRGS primers were added to the PCR for the final amplification.

3F-ZFA assemblies from a F1 1F modules and 2F module-QRG cap. F1 was amplified using the F1(noF0)Fn and F1Rn primers. The 2F-module sequence was first amplified with 2FM-F1Fn and 2FM-F2Rn primers, and gel-purified. Then 1–5 ng of gel-purified DNA was used as template for amplification with 2FM-F2-QRG(X)Fn and 2FM-F3RnLRGS primers to substitute the RSD N-terminal cap with the QRG-N-terminal cap. The amplified 2F-module-QRG and F1 were gel-gel-purified and the finger amplicons were assembled as described above, where the F1(noF0)Fn and 2FM-F3RnLRGS primers were added to the PCR for the final amplification.

3F-ZFA assemblies from a 2F module-QRG cap and F3 1F modules. The 2F-module sequence was first amplified with 2FM-F1Fn and 2FM-F2Rn primers, gel-purified and then 1–5 ng of gel-purified DNA was used as template for amplification with 2FM-F1(noF0)-QRG(X)Fn and 2FM-F2Rn primers. F3 was amplified using the F3(noF0)Fn and F3RnLRGS primers. The amplified 2F-module-QRG and F3 module sequences were
gel-purified and the finger amplicons were assembled as described above, where the 2FM-F1(noF0)Fn and F3RnLRGS primers were added to the PCR for the final amplification.

4F-ZFA assemblies from F0, F1, F2 and F3 1F modules. F0 was amplified using the F0Fn and F0Rn primers. F1 was amplified using the F1Fn and F1Rn primers. F2 was amplified using the F2Fn and F2Rn primers. F3 was amplified using F3Fn and F3RnLRGS primers. The amplified DNA was gel-purified, and the finger amplicons were assembled as described above, where the F0Fn and F3RnLRGS primers were added to the PCR for the final amplification.

4F-ZFA assemblies from F0 and F1 1F modules, and 2F module. F0 was amplified using the F0Fn and F0Rn primers. F1 was amplified using the F1Fn and F1Rn primers. The 2F-module sequence was amplified using 2FM-F1Fn and 2FM-F2Rn primers. The amplified DNA was gel-purified and the finger amplicons were assembled as described above, where the F0Fn and 2FM-F3RnLRGS primers were added to the PCR for the final amplification.

4F-ZFA assemblies from F0, 2F-module sequence and F3. F0 was amplified using the F0Fn and F0Rn primers. The 2F-module sequence was amplified using the 2FM-F1Fn and 2FM-F2Rn primers and F3 sequence was amplified using the F3Fn and F3Rn primers. The amplified DNA was gel-purified and the finger amplicons were assembled as described above, where the F0Fn and 2FM-F3RnLRGS primers were added to the PCR for the final amplification.

4F-ZFA assemblies from 2F-module sequence, F2 and F3. The 2F-module sequence was amplified using the 2FM-F0Fn and 2FM-F1Rn primers. F2 was amplified using the F2Fn and F2Rn primers. F3 was amplified using the F3Fn and F3Rn primers. The amplified DNA was gel-purified and the finger amplicons were assembled as described above, where the 2FM-F0Fn and F3RnLRGS primers were added to the PCR for the final amplification.

4F-ZFA assemblies from N-terminal 2F-module sequence, C-terminal 2F-module sequence. The sequence encoding the N-terminal region of the 2F module was amplified with the 2FM-NT-in-Fn and 2FM-F1Rn primers and the C-terminal 2F module was amplified with the 2FM-F2Fn and 2FM-F3RnLRGS primers. The amplified products were gel-purified and the finger amplicons were assembled as described above, where the 2FM-NT-out-Fn and 2FM-CT-out-Rn primers were added to the PCR for the final amplification.

4F-ZFA assemblies from F0, F1 and 2F-module-QRG. F0 was amplified using the F0Fn and F0Rn primers. F1 was amplified using the F1Fn and F1Rn primers. The 2F-module sequence was first amplified with 2FM-F1Fn and 2FM-F2Rn primers, gel-purified and then 1–5 ng of gel-purified DNA was used as template for amplification with 2FM-F2-QRG(X)Fn and 2FM-F3RnLRGS primers to substitute in the encoded product the RSD N-terminal cap with the QRG-N-terminal cap. The amplified 2F-module-QRG, F0 and F1 modules were gel-purified and the finger amplicons were assembled as described above, where the F0Fn and 2FM-F3RnLRGS primers were added to the PCR for the final amplification.

4F-ZFA assemblies from F0, 2F-module-QRG and F3. F0 was amplified using the F0Fn and F0Rn primers. The 2F-module sequence was first amplified with 2FM-F1Fn and 2FM-F2Rn primers, gel-purified and then 1–5 ng of gel-purified DNA was used as template for amplification with 2FM-F1-QRG(X)Fn and 2FM-F2Rn primers. F3 was amplified using the F3Fn and F3Rn primers. The amplified the F0, 2F-module-QRG and F3 modules were gel-purified and amplicons were assembled as described above, where the F0Fn and F3RnLRGS primers were added to the PCR for the final amplification.

4F-ZFA assemblies from 2F-module-QRG, F2 and F3. The 2F-module sequence was first amplified with 2FM-F1Fn and 2FM-F2Rn primers, gel-purified and then 1–5 ng of gel-purified DNA was used as template for amplification with 2FM-F0-QRG(X)Fn and 2FM-F1Rn primers. F2 was amplified using the F2Fn and F2Rn primers. F3 sequence was amplified using the F3Fn and F3Rn primers. The amplified DNA was gel-purified and the finger amplicons were assembled as described above, where the 2FM-F0Fn and F3RnLRGS primers were added to the PCR for the final amplification.

4F-ZFA assemblies from N-terminal 2F module-QRG, C-terminal 2F module. The N-terminal 2F module sequence was amplified with 2FM-F1Fn and 2FM-F2Rn, gel-purified and then 1–5 ng of gel-purified DNA was used as template for amplification with 2FM-F0-QRG(X)Fn and 2FM-F1Rn primers. This amplified 2F-module-QRG was again gel-purified and PCR-amplified with 2FM-NT-in-Fn and 2FM-F1Rn primers. The C-terminal 2F-module sequence was amplified with 2FM-F2Fn and 2FM-F3RnLRGS. The amplified sequences encoding N-terminal and C-terminal 2F-module sequences were gel-purified, and the finger amplicons were assembled as described above, where the 2FM-NT-out-Fn and 2FM-CT-out-Rn primers were added to the PCR for the final amplification.

4F-ZFA assemblies from N-terminal 2F module, C-terminal 2F module-QRG. The N-terminal 2F-module sequence was amplified with 2FM-NT-in-Fn and 2FM-F1Rn. The C-terminal 2F-module sequence was amplified with 2FM-F1Fn and 2FM-F2Rn, gel-purified and then 1–5 ng of gel-purified DNA was used as template for amplification with 2FM-F2-QRG(X)Fn and 2FM-F3RnLRGS primers. The amplified sequences encoding N-terminal and C-terminal 2F module were gel-purified, and the finger amplicons were assembled as described above, where the 2FM-NT-out-Fn and 2FM-CT-out-Rn primers were added to the PCR for the final amplification.

4F-ZFA assemblies from N-terminal 2F module-QRG, C-terminal 2F-module-QRG. The N-terminal 2F-module sequence was amplified with 2FM-F1Fn and 2FM-F2Rn, gel-purified and then 1–5 ng of gel-purified DNA was used as template for amplification with 2FM-F0-QRG(X)Fn and 2FM-F1Rn primers. This amplified 2F-module-QRG sequence was again gel-purified and PCR-amplified with 2FM-NT-in-Fn and 2FM-F1Rn primers. The sequence encoding C-terminal 2F-module was amplified with 2FM-F1Fn and 2FM-F2Rn, gel-purified and then 1–5 ng of gel-purified DNA was used as template for amplification with 2FM-F2-QRG(X)Fn and 2FM-F3RnLRGS primers. The amplified
sequences encoding N-terminal and C-terminal 2F-module were gel-purified, and the finger ampiclonics were assembled as described above, where the 2FM-N T-out-Fn and 2FM-CT-out-Rn primers were added to the PCR for the final amplification.

BIH-binding site selections using the 28-bp library. The selections for 3F and 4F ZFAs were performed as previously described. We plated 1–5 × 10^7 selection strain cells transformed with the 1352-omega-UV2 ZFA expression plasmid and the 28-bp pH3U3 library plasmid on NM minimal medium selective plates lacking uracil and containing 3-AT (2.5 mM, 5 mM or 10 mM) as the competitor and grown at 37 °C for 36–72 h. The number of surviving bacterial colonies on each plate was estimated, then these colonies were pooled, and the population of recovered DNA sequences was determined via Illumina sequencing. Unique sequences were ranked based on the number of recovered reads. From this list an over-represented sequence motif was determined with MEME using as input the number of unique sequences from the top of the list that correspond to the estimated number of colonies on the selection plate (typically >1,000). The aligned sequences were then used to generate sequence logos using Weblogo.

Yeast-based ZFN activity assay. To assess the activity of our ZFNs in an independent system we used a Mel1-based yeast activity assay. The target sites for test ZFNs and the positive control ZFN were cloned in the modified ySSA vector and then integrated into the yeast genome (BY4741 strain) at the HO locus. The ZFA sequences were cloned such that they were at the 5'-end of the wild-type FokI nuclease domain in the pYHis3 and pYNLeu2 vectors between Acc651 and BamH1 sites. pYHis3 and pYNLeu2 vectors encoding the ZFN sequences (or EGFP as a negative control) were transformed in the yeast strain containing the selection plate (typically >1,000). The aligned sequences were then used to generate sequence logos using Weblogo.

ZFN injections and lesion analysis. For gene targeting in zebrafish, ZFA sequences were cloned in pCS2 vectors containing the sequence encoding the DD/RR obligate heterodimer version of the FokI nuclease domain. pCS2-ZFN constructs were linearized with NotI, and mRNA was transcribed using the mMesagemMachine SP6 kit from Ambion. ZFN mRNAs were injected into the blastomere of one-cell-stage zebrafish embryos as previously described. ZFN mRNA–injected embryos with normal appearance (8–30 embryos) and uninjected embryos were collected 24 h after fertilization and incubated in 50 mM NaOH (15 µl/embryo) for 15 min at 95 °C to isolate genomic DNA and then neutralized with 0.5 M Tris-HCl (4 µl/embryo). The DNA solution was centrifuged for 1 min at 13,000 r.p.m. and supernatant was taken for lesion analysis. For initial validation of ZFN activity, the region flanking the ZFN target site was amplified using the Phire Hot Start DNA polymerase (Finnzymes) and restriction fragment length polymorphism analysis or Cel I nuclease assay (Transgenomics) was performed as described previously. For Illumina sequencing, the region flanking the ZFN sites was amplified using the primers and then digested with the appropriate restriction enzyme (Supplementary Table 10). The ends for the digested DNA were polished using Klenow exo− enzyme (New England Biolabs) or T4 DNA polymerase (New England Biolabs) and A-tailed using Klenow exo− enzyme (New England Biolabs). The barcoded adapters (Supplementary Table 10) were ligated to each DNA pool and then PCR-amplified with the Illumina genomic primers 1.1 and 1.2. After sequencing, indels were identified as described previously. Briefly, two tags unique to a ZFN target site were used, a 5' tag and a 3' tag (Supplementary Table 10), and the distance between the tags was used to distinguish wild-type sequence from the indel-containing sequence. Lesion frequency was calculated as follows: lesion frequency = (100 × Nindels)/Ntotal where Nindels represents number of sequences containing indels that are >1 bp in length, and Ntotal represents number of total number of recovered sequences.

Genomic analysis of ZFN target sites. The targeting density and overlap of ZFN sites were determined for three archives (one-finger and two-finger modules from this work (1/2FM), CoDA13 two-finger modules (2FM) and one-finger and two-finger modules from ref. 15(1/2FM)) on the unique protein-coding exons in the zebrafish genome (Zv9 assembly) and the human genome (GRCh37.p5 assembly) from Ensembl release 64. Target sites for each finger archive were determined using custom Perl scripts, where only ZFN sites that map to a single unique gene were counted in this analysis. This analysis provided information on the fraction of genes that can be targeted and the density of the sites per base pair.

Germline transmission analysis. ZFNs were injected at optimal doses in wild-type zebrafish embryos. Injected embryos were grown to maturity and crossed with wild-type zebrafish to identify carriers. PCR products spanning the target loci in F1 generation embryos were screened using Cel1 nuclease nuclease assay for presence of lesions. The compositions of these lesions were characterized through cloning and sequencing PCR products spanning the ZFN target site for each gene (Supplementary Table 10).

16. Westerfield, M. The Zebrafish Book (University of Oregon Press, Eugene, Oregon, USA, 1993).
17. Gupta, A., Meng, X., Zhu, L.J., Lawson, N.D. & Wolfe, S.A. Nucleic Acids Res. 39, 381–392 (2010).
18. Bailey, T.L. & Elkan, C. Proc. Int. Conf. Intell. Syst. Mol. Biol. 2, 28–36 (1994).
19. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. Genome Res. 14, 1188–1190 (2004).
20. Ryan, M.P., Jones, R. & Morse, R.H. Mol. Cell Biol. 18, 1774–1782 (1998).
21. Miller, J.C. Nat. Biotechnol. 25, 778–785 (2007).
22. Szczepk, M. et al. Nat. Biotechnol. 25, 786–793 (2007).