181 zinc finger arrays constructed by CoDA were each evaluated for sequence-specific DNA binding activity using the B2H reporter system. Each of the 181 zinc finger arrays we tested binds to a different nine bp target site (Supplementary Table 3). Fold-activation of a lacZ reporter gene in the B2H system was calculated as previously described.\(^1,2\) Values of B2H activity shown are the mean of three replicates and were plotted from lowest to highest from left to right. Thresholds of fold-activation we used to predict failure (<1.57) or success (>3.00) as ZFNs are indicated.
Supplementary Figure 2  Comparison of zinc finger arrays made by modular assembly and CoDA

Fold-activation values (as measured in the B2H reporter assay) of the most active (A) modular assembly zinc finger arrays and (B) CoDA zinc finger arrays for 26 target sites (Supplementary Table 4) are shown. Fold-activation values represent the mean of three replicates and have been arranged from lowest to highest plotted from left to right in each panel. Thresholds of B2H reporter assay fold-activation values we used to predict failure (<1.57) or success (>3.00) as ZFNs are indicated.
Supplementary Figure 3  Frequencies and sequences of CoDA ZFN-induced mutations in somatic zebrafish cells

For each zebrafish gene targeted by CoDA ZFNs, the wild-type sequence is shown at the top with ZFN half-sites highlighted in yellow. Deletions are indicated by grey highlighted red dashes and insertions by blue highlighted lower case blue letters. The net number of nucleotides inserted or deleted and the number of times each wild-type mutant allele was isolated is shown on the right.

A.  *actn1* (60 sequences total):

| ZFN-L | ZFN-R |
|-------|-------|
| GCCTTCTCCGGGGCAGAAGGT | WT [50x] |
| GCCTTCTCCGGGGCAGAAGGT | ∆2 [3x] |
| GCCTTCTCCGGGGCAGAAGGT | ∆6 [1x] |
| GCCTTCTCCGGGGCAGAAGGT | ∆7 [1x] |
| GCCTTCTCCGGGGCAGAAGGT | ∆13 [1x] |
| GCCTTCTCCGGGGCAGAAGGT | ∆14 [1x] |
| GCCTTCTCCGGGGCAGAAGGT | ∆25 [1x] |
| GCCTTCTCCGGGGCAGAAGGT | ∆25 [1x] |
| GCCTTCTCCGGGGCAGAAGGT | ∆65 [1x] |

mutations in 10 of 60 sequences: 16.6%

B.  *rag2* (52 sequences total):

| ZFN-L | ZFN-R |
|-------|-------|
| ATCTTCTGCTCCAGGGGTGAAGGT | WT [48x] |
| ATCTTCTGCTCCAGGGGTGAAGGT | +1 [3x] |
| ATCTTCTGCTCCAGGGGTGAAGGT | +3 [1x] |

mutations in 4 of 52 sequences: 7.7%

C.  *gad2* (43 sequences total):

| ZFN-L | ZFN-R |
|-------|-------|
| AGCCGCAGCTCTCGGCTGTAGAC | WT [40x] |
| AGCCGCAGCTCTCGGCTGTAGAC | ∆4 [1x] |
| AGCCGCAGCTCTCGGCTGTAGAC | ∆13 [1x] |
| AGCCGCAGCTCTCGGCTGTAGAC | +5 (∆8 and +13) [1x] |

mutations in 3 of 43 sequences: 7.0%

D.  *lmna* (41 sequences total):

| ZFN-L | ZFN-R |
|-------|-------|
| CTCCACAGCTCTGGGGGAGAAGAG | WT [39x] |
| CTCCACAGCTCTGGGGGAGAAGAG | ∆2 [1x] |
| CTCCACAGCTCTGGGGGAGAAGAG | ∆3 [1x] |

mutations in 2 of 49 sequences: 4.9%

E.  *apoeb* (64 sequences total):

| ZFN-L | ZFN-R |
|-------|-------|
| CCCCTCAGCCAGATGGGAGGAG | WT [61x] |
| CCCCTCAGCCAGATGGGAGGAG | ∆4 [1x] |
| CCCCTCAGCCAGATGGGAGGAG | ∆9 [1x] |
| CCCCTCAGCCAGATGGGAGGAG | ∆29 [1x] |

mutations in 3 of 64 sequences: 4.7%
F. **trpm7** (55 sequences total):

```
ZFN-L     ZFN-R
CACACCTGCACACAAGATGCTGCT  WT [53x]
CACCTGCACACAGATGCTGCT   Δ14 [1x]
CACCTGCACACAAGATGCTGCT   +5  [1x]
```

mutations in 2 of 55 sequences: **3.6%**

G. **grip1** (90 sequences total):

```
ZFN-L     ZFN-R
GCCCGCTGCTGGTGGAGGTGGCC   WT [88x]
GCCCGCTGAGGGGTGGCC       Δ3  [2x]
GCCCGCTGAGGGGTGGCC       Δ13 [1x]
```

mutations in 3 of 90 sequences: **3.3%**

H. **pclo** (96 sequences total):

```
ZFN-L     ZFN-R
CCCCTCTCCTCAAAAGCAGATGCA  WT [93x]
CCCCTCTCCTCAAAAGCAGATGCA  Δ7  [1x]
CCCCTCTCCTCAAAAGCAGATGCA  Δ22 [1x]
CCCCTCTCCTCAAAAGCAGATGCA  +6  [1x]
```

mutations in 3 of 96 sequences: **3.1%**

I. **jak3** (71 sequences total):

```
ZFN-L     ZFN-R
GGCCCCACCAAGCCTGCTGGAGGA  WT [70x]
GGCCCCACCAAGCCTGCTGGAGGA  Δ3  [1x]
```

mutations in 1 of 71 sequences: **1.4%**

J. **ago1** (96 sequences total):

```
ZFN-L     ZFN-R
CTCTGCCGCCACCTAGAGGATGGT  WT [95x]
CTCTGCCGCCACCTAGAGGATGGT  Δ5  (Δ10 and +5) [1x]
```

mutations in 1 of 96 sequences: **1.0%**

K. **slitrk1** (96 sequences total):

```
ZFN-L     ZFN-R
GCCCACAGCAATGGCGGAGCCGCC  WT [95x]
GCCCACAGCAATGGCGGAGCCGCC  Δ11 [1x]
```

mutations in 1 of 96 sequences: **1.0%**

L. **bmp2a** (117 sequences total):

```
ZFN-L     ZFN-R
GCCGACTGACACAAGAGAAGTC   WT [116x]
GCCGACTGACACAAGAGAAGTC   Δ5   [1x]
```

mutations in 1 of 117 sequences: **0.9%**
Supplementary Figure 4  Frequencies and sequences of CoDA ZFN-induced mutations in somatic plant cells

For each plant gene targeted by CoDA ZFNs, the wild-type sequence is shown at the top with ZFN half-sites highlighted in yellow. Deletions are indicated by grey highlighted red dashes and insertions by blue highlighted lower case blue letters. The net number of nucleotides inserted or deleted and the number of times each wild-type mutant allele was isolated is shown on the right. Genes shown in (A) and (B) are from soybean and those shown in (C)-(H) are from Arabidopsis.

A.  **Dcl4a** (32 sequences total):

| ZFN-L          | ZFN-R                  | WT  |
|----------------|------------------------|-----|
| TGCTTCATCACAATGAGATGAT | WT [26x]               |
| TGCTTCATCACAATGAGATGAT | Δ1  [1x]               |
| TGCTTCATCACAATGAGATGAT | Δ3  [1x]               |
| TGCTTCATCACAATGAGATGAT | Δ6  [1x]               |
| TGCTTCATCACAATGAGATGAT | Δ6  [1x]               |
| TGCTTCATCACAATGAGATGAT | Δ10 [1x]              |
| TGCTTCATCACAATGAGATGAT | Δ12 [1x]              |

mutations in 6 of 32 sequences: **18.8%**

B.  **Dcl4b** (28 sequences total):

| ZFN-L          | ZFN-R                  | WT  |
|----------------|------------------------|-----|
| TGCTTCATCACAATGAGATGAT | WT [25x]               |
| TGCTTCATCACAATGAGATGAT | Δ1  [1x]               |
| TGCTTCATCACAATGAGATGAT | Δ4  [1x]               |
| TGCTTCATCACAATGAGATGAT | Δ5  [1x]               |

mutations in 3 of 28 sequences: **10.7%**

C.  **MPK8** (83 sequences total):

| ZFN-L          | ZFN-R                  | WT  |
|----------------|------------------------|-----|
| CTCCACAACATCGAGACGAA | WT [76x]              |
| CTCCACAACATCGAGACGAA | Δ3  [3x]               |
| CTCCACAACATCGAGACGAA | Δ6  [1x]               |
| CTCCACAACATCGAGACGAA | Δ6  [1x]               |
| CTCCACAACATCGAGACGAA | Δ19 [2x]              |
| CTCCACAACATCGAGACGAA | +1  [1x]               |

mutations in 7 of 83 sequences: **8.4%**

D.  **MPK11** (90 sequences total):

| ZFN-L          | ZFN-R                  | WT  |
|----------------|------------------------|-----|
| CTCTTCGTCTATCGGCAGAGGC | WT [87x]              |
| CTCTTCGTCTATCGGCAGAGGC | Δ8  [1x]               |
| CTCTTCGTCTATCGGCAGAGGC | +1  [1x]               |
| CTCTTCGTCTATCGGCAGAGGC | +2  [1x]               |

mutations in 3 of 90 sequences: **3.3%**

E.  **MKK9** (95 sequences total):

| ZFN-L          | ZFN-R                  | WT  |
|----------------|------------------------|-----|
| GCCAGCGACGTGGTGGTGGG | WT [92x]             |
| GCCAGCGACGTGGTGGTGGG | Δ2  [1x]               |
| GCCAGCGACGTGGTGGTGGG | Δ3  [2x]               |

Nature Methods: doi: 10.1038/nmeth.1542
mutations in 3 of 95 sequences: 3.2%

F. **MPK15** (73 sequences total):

| ZFN-L | ZFN-R | WT      | Δ3 [1x] | Δ2 [1x] |
|--------|--------|---------|---------|---------|
| TTCTTCATCCAGATGTTGTTGAG | TTCTTCATTCAGTTGTTGAG | WT [71x] |
| TTCTTCATCCAGTTGTTGAG | TTCTTCATCCGTTGTTGAG |

mutations in 2 of 73 sequences: 2.7%

G. **MAPKK18** (75 sequences total):

| ZFN-L | ZFN-R | WT      | Δ13 [1x] | Δ39 [1x] |
|--------|--------|---------|---------|---------|
| CCCTTCCACAACACGGAGAAGCT | CCCTTCCACAACACGGAGAAGCT | WT [73x] |
| ------------------------------------- | ------------------------------------- |
| CAACGGAGAAGCT | CAACGGAGAAGCT |

mutations in 2 of 75 sequences: 2.7%

H. **GA3OX2** (94 sequences total):

| ZFN-L | ZFN-R | WT      | Δ2 [1x] |
|--------|--------|---------|---------|
| AGCTACGCCGTAGCCGGAGACGCC | AGCTACGCCGTAGCCGGAGACGCC | WT [93x] |
| AGCTACGCCGCCCAGACCC | AGCTACGCCGCCCAGACCC |

mutations in 1 of 94 sequences: 1.1%
Supplementary Figure 5  Screenshot of CoDA targeting output from ZiFiT Version 3.3

In the example shown, the output provides the full target sequence, a summary of CoDA zinc finger units (together with their 3 bp subsite targets) required to construct a ZFN pair to the target sequence, and unique identification numbers for the clone encoding the CoDA zinc finger units. For those interested in using DNA synthesis to create CoDA ZF arrays, a link entitled “ZF DNA sequence” is provided. Clicking on this link leads to the generation of a new window providing the full DNA sequences of the CoDA zinc finger arrays.
Supplementary Table 1  F1 CoDA Finger Units

The recognition helix sequences of F1 units selected to bind specific three bp subsites (top row) are listed and highlighted in purple. Each F1 unit was identified from an active three-finger array in which it was positioned adjacent to the F2 unit shown in the grey column on the far left. N.F. indicates where selections were attempted to isolate a unit but “no finger” was obtained. “-” indicates that no attempt has yet been made to identify fingers.

| Target Helix | F1 CoDA Finger Units |
|--------------|----------------------|
| GQG | GQA | GGC | GCT | GAG | GAA | GAC | GAT | DOQ | GCA | GCC | GCT | GTG | GTA | GTLC | GLL | AGG | AAC | AGG | TGC | TGTT | TAG | TCG | TCT | TGC |
| 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L |
| 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L |
| 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L |
| 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L |
| 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L |
| 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L |
| 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L |
| 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L |
| 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L |
| 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L |
| 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L |
| 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L | 3B1L |
Supplementary Table 2  F3 CoDA Finger Units

The recognition helix sequences of F3 units selected to bind specific three bp subsites (top row) are listed and highlighted in purple. Each F3 unit was identified from an active three-finger array in which it was positioned adjacent to the F2 unit shown in the grey column on the far left. N.F. indicates where selections were attempted to isolate a unit but “no finger” was obtained. “-“ indicates that no attempt has yet been made to identify fingers.

| F2 | F3 |
|---|---|
| Target helix | GGG | GGA | GCC | GGT | GAG | GAC | GAT | GGC | GCT | GTG |
| GGG | GGG | GGG | GGG | GGG | GGG | GGG | GGG | GGG | GGG | GGG |
| GGA | GGA | GGA | GGA | GGA | GGA | GGA | GGA | GGA | GGA | GGA |
| GCC | GCC | GCC | GCC | GCC | GCC | GCC | GCC | GCC | GCC | GCC |
| GGT | GGT | GGT | GGT | GGT | GGT | GGT | GGT | GGT | GGT | GGT |
| GAG | GAG | GAG | GAG | GAG | GAG | GAG | GAG | GAG | GAG | GAG |
| GAC | GAC | GAC | GAC | GAC | GAC | GAC | GAC | GAC | GAC | GAC |
| GAT | GAT | GAT | GAT | GAT | GAT | GAT | GAT | GAT | GAT | GAT |
| GGC | GGC | GGC | GGC | GGC | GGC | GGC | GGC | GGC | GGC | GGC |
| GCT | GCT | GCT | GCT | GCT | GCT | GCT | GCT | GCT | GCT | GCT |
| GTG | GTG | GTG | GTG | GTG | GTG | GTG | GTG | GTG | GTG | GTG |
| TGG | TGG | TGG | TGG | TGG | TGG | TGG | TGG | TGG | TGG | TGG |
| TCG | TCG | TCG | TCG | TCG | TCG | TCG | TCG | TCG | TCG | TCG |
| TCC | TCC | TCC | TCC | TCC | TCC | TCC | TCC | TCC | TCC | TCC |
| TTC | TTC | TTC | TTC | TTC | TTC | TTC | TTC | TTC | TTC | TTC |
Supplementary Table 3  B2H Reporter Assay Activity and Target Sites for 181 CoDA Zinc Finger Arrays

B2H reporter assay activities (reported as mean fold-activation values of a lacZ reporter gene performed in triplicate) and the cognate 9 bp target sites for 181 CoDA zinc finger arrays are shown. For most of the zinc finger arrays, IPTG was added to the culture medium at 500 µM to induce zinc finger protein expression in the B2H reporter assay as previously described. For a small number of arrays (highlighted in orange), a lower concentration of IPTG was used as indicated to minimize toxicity associated with zinc finger array expression.

| Target Site | Fold-Activation in B2H reporter assay (IPTG in µM) | Target Site | Fold-Activation in B2H reporter assay (IPTG in µM) | Target Site | Fold-Activation in B2H reporter assay (IPTG in µM) |
|-------------|-----------------------------------------------|-------------|-----------------------------------------------|-------------|-----------------------------------------------|
| GAGAAGUAT   | 6.06                                         | GCAAGGCCC   | 6.98                                         | GCGTAUAGT  | 4.45                                         |
| GAGAAGCIA   | 3.85                                         | GCAACTAAGA  | 1.61                                         | GGGAGAACG  | 12.84                                        |
| GAGAAGAA    | 14.16                                        | GCAAGGTGTC  | 4.36                                         | GGGAGATAAG | 5.92                                         |
| GAGACACCA   | 6.58                                         | GCAAGGTGTC  | 1.08                                         | GGGAGATAAG | 7.36                                         |
| GAGACACCT   | 5.18                                         | GCAAGGTGTC  | 2.65                                         | GGGTAGTGGT | 13.53                                        |
| GAGACTGCT   | 5.63                                         | GCAAGGTGTC  | 2.75                                         | GGGTCGATT  | 7.37                                         |
| GAGACCTGCC  | 5.48                                         | GCAAGGTGTC  | 5.69                                         | GGGTCGATT  | 1.49                                         |
| GAGACTGTTC  | 6.04                                         | GCAAGGTGTC  | 1.34                                         | GGGTGGAGA  | 3.86                                         |
| GAGACGTGCC  | 2.06                                         | GCAAGGTGTC  | 5.91                                         | GGGTGGAGA  | 5.54                                         |
| GAGACGTGCTC | 13.88                                        | GCAAGGTGTC  | 5.20                                         | GGGTGGAGA  | 5.10                                         |
| GAGACGTGCTG | 8.05                                         | GCAAGGTGTC  | 2.81                                         | GGGTGGAGA  | 10.34                                        |
| GAGATGGAG   | 7.28                                         | GCAAGGTCTG  | 1.97                                         | GGGTGGAGA  | 5.92                                         |
| GAGAAGAAG   | 7.39                                         | GCAAGGTCTG  | 5.53                                         | GGGTGGAGA  | 7.20                                         |
| GAGAAGAAGA  | 3.15                                         | GCAAGGTCTG  | 2.67                                         | GGTAAAGCG  | 8.03                                         |
| GAGAAGATCTC | 1.78                                         | GCAAGGTCTG  | 9.34                                         | GGTAAAGCG  | 5.40                                         |
| GAGAAGCAGA  | 3.42                                         | GCAAGGTCTG  | 2.88                                         | GGTAAAGCG  | 1.15                                         |
| GAGAAGCTGC  | 7.34                                         | GCAAGGTCTG  | 10.35                                        | GGTAAAGCG  | 5.50                                         |
| GAGCAGCC    | 6.67                                         | GCAAGGTCTG  | 5.31                                         | GGTCAATAG  | 2.70                                         |
| GAGCAGCCAA  | 1.74                                         | GCAAGGTCTG  | 5.38                                         | GGTCAATAG  | 8.34                                         |
| GAGGCTGCTG  | 6.22                                         | GCAAGGTCTG  | 3.74                                         | GGTCTGAGC  | 11.74                                        |
| GAGGCAGCTC | 6.48                                         | GCAAGGTCTG  | 11.04                                        | GGTCTGAGC  | 1.25                                         |
| GAGGCGAG    | 6.18                                         | GCAAGGTCTG  | 5.91                                         | GGTCTGAGC  | 3.96                                         |
| GAGGCGCTC  | 4.50                                         | GCAAGGTCTG  | 1.62                                         | GGTCTGAGC  | 5.99                                         |
| GAGGCGCTG  | 7.01                                         | GCAAGGTCTG  | 7.21                                         | GGTGGAAGA  | 5.39                                         |
| GAGGCGCTGGA | 11.24                                        | GCAAGGTCTG  | 1.04                                         | GGTGGAAGA  | 11.09                                        |
| GAGGCGCTGG  | 6.27                                         | GCAAGGTCTG  | 9.41                                         | GGTGGAAGA  | 10.26                                        |
| GAGGCGCTGGG | 8.29                                         | GCAAGGTCTG  | 3.10                                         | GGTGGAAGA  | 12.58                                        |
| GAGGCGCTGGGTGCAGGC | 6.27 | GCAAGGTCTG | 5.33 | GGTGGAAGA | 12.58 |
| GAGGCGCTGGGGGTAATAGG | 6.29 | GCAAGGTCTG | 5.33 | GGTGGAAGA | 12.58 |
| GAGGCGCTGGGTAATAGG | 6.29 | GCAAGGTCTG | 5.33 | GGTGGAAGA | 12.58 |
| GAGGCGCTGGGTAATAGG | 6.29 | GCAAGGTCTG | 5.33 | GGTGGAAGA | 12.58 |
| GAGGCGCTACCAATCCGTGGAAG | 4.60 | GCAAGGTCTG | 4.83 | GGTGGAAGA | 5.79 |
| GAGGCGCTACCAATCCGTGGAAG | 11.27 | GCAAGGTCTG | 5.27 | GGTGGAAGA | 4.33 |
| GAGGCGCTACCAATCCGTGGAAG | 7.85 | GCAAGGTCTG | 7.85 | GGTGGAAGA | 9.06 |
| GAGGCGCTACCAATCCGTGGAAG | 4.73 | GCAAGGTCTG | 4.73 | GGTGGAAGA | 1.58 |
| GAGGCGCTACCAATCCGTGGAAG | 13.84 | GCAAGGTCTG | 2.73 | GGTGGAAGA | 3.97 |
| GAGGCGCTACCAATCCGTGGAAG | 7.95 | GCAAGGTCTG | 8.59 | GGTGGAAGA | 0.76 |
| GAGGCGCTACCAATCCGTGGAAG | 2.32 | GCAAGGTCTG | 8.12 | GGTGGAAGA | 3.94 |
| GAGGCGCTACCAATCCGTGGAAG | 3.81 | GCAAGGTCTG | 14.79 | GGTGGAAGA | 0.97 |
| GAGGCGCTACCAATCCGTGGAAG | 4.24 | GCAAGGTCTG | 3.53 | GGTGGAAGA | 12.53 |
| GAGGCGCTACCAATCCGTGGAAG | 5.56 | GCAAGGTCTG | 4.58 | GGTGGAAGA | 1.25 |
| GAGGCGCTACCAATCCGTGGAAG | 3.12 | GCAAGGTCTG | 6.16 | GGTGGAAGA | 6.81 |
| GAGGCGCTACCAATCCGTGGAAG | 9.39 | GCAAGGTCTG | 13.96 | GGTGGAAGA | 1.81 |
| GAGGCGCTACCAATCCGTGGAAG | 2.48 | GCAAGGTCTG | 3.15 | GGTGGAAGA | 3.78 |
| GAGGCGCTACCAATCCGTGGAAG | 4.26 | GCAAGGTCTG | 10.29 | GGTGGAAGA | 11.28 |
| GAGGCGCTACCAATCCGTGGAAG | 6.62 | GCAAGGTCTG | 1.89 | GGTGGAAGA | 5.85 |
| GAGGCGCTACCAATCCGTGGAAG | 12.33 | GCAAGGTCTG | 3.68 | GGTGGAAGA | 8.64 |
| GAGGCGCTACCAATCCGTGGAAG | 1.46 | GCAAGGTCTG | 7.21 | GGTGGAAGA | 5.64 |
| GAGGCGCTACCAATCCGTGGAAG | 4.42 | GCAAGGTCTG | 49.38 | GGTGGAAGA | 6.56 |
| GAGGCGCTACCAATCCGTGGAAG | 2.08 | GCAAGGTCTG | 10.57 | GGTGGAAGA | 3.86 |
| GAGGCGCTACCAATCCGTGGAAG | 2.44 | GCAAGGTCTG | 8.17 | GGTGGAAGA | 7.00 |
| GAGGCGCTACCAATCCGTGGAAG | 15.21 | GCAAGGTCTG | 7.51 | GGTGGAAGA | 15.21 |
| GAGGCGCTACCAATCCGTGGAAG | 8.28 | GCAAGGTCTG | 7.51 | GGTGGAAGA | 15.21 |
Supplementary Table 4  Direct comparison of modularly assembled and CoDA zinc finger arrays in the B2H reporter assay

B2H reporter assay fold-activation values for zinc finger arrays targeted to 26 different DNA sites (left column) are shown. Fold-activation values are the mean of assays performed in triplicate. Zinc finger arrays were made by either modular assembly (using one of three module archives described in references 3-5) or CoDA. (Note that multiple zinc finger arrays can be made in some cases using the module archive of reference 5.) For each of the 26 target sites, the most active and second most active arrays (as judged by B2H fold-activation values) are highlighted in green and grey, respectively. The number of sites and the percentage of sites for which CoDA or a particular module set yielded the most active protein are shown in the third-to-last and second-to-last rows, respectively. The average fold-activation values for all arrays made by CoDA or a particular modular set are shown in the last row of the table.

| Target Site | Modular Assembly ZFNs made from archive of | CoDa |
|-------------|--------------------------------------------|------|
| 5'-GAA-GCG-GTC-3' | Ref. 4 1.75 2.16 2.54 0.81 | 7.05 |
| 5'-GAC-GGC-GCT-3' | Ref. 3 1.05 0.75 | 3.64 |
| 5'-GAG-GCT-GTG-3' | Ref. 5 1.91 8.22 1.08 1.09 2.47 | 3.66 |
| 5'-GAT-GGA-GTT-3' | Ref. 5 1.79 3.34 2.19 2.48 2.70 5.79 | 14.66 |
| 5'-GCA-GCC-GGA-3' | Ref. 5 0.72 1.42 | 5.60 |
| 5'-GCA-GCG-GCA-3' | Ref. 5 0.98 2.40 | 2.30 |
| 5'-GCA-GCT-GTG-3' | Ref. 3 1.06 2.19 | 4.37 |
| 5'-GCA-GTC-GAC-3' | Ref. 5 0.97 0.79 | 5.92 |
| 5'-GCA-GTC-GGC-3' | Ref. 5 0.90 0.99 | 8.20 |
| 5'-GCA-GTC-GCT-3' | Ref. 5 1.20 1.37 | 20.17 |
| 5'-GCA-GTC-GGC-3' | Ref. 5 0.81 1.50 | 3.52 |
| 5'-GCA-GTG-GCC-3' | Ref. 3 0.79 1.29 | 2.10 |
| 5'-GCC-GCA-GGA-3' | Ref. 5 2.22 1.58 1.10 | 5.00 |
| 5'-GCC-GCT-GTG-3' | Ref. 3 2.76 2.37 | 1.88 |
| 5'-GCC-GGC-GTC-3' | Ref. 3 2.37 1.90 1.27 | 7.68 |
| 5'-GCC-GTC-GGC-3' | Ref. 3 0.85 1.07 | 2.81 |
| 5'-GCC-GTC-GCC-3' | Ref. 3 1.24 1.38 | 2.57 |
| 5'-GGC-GCG-GGC-3' | Ref. 5 1.62 2.20 | 3.08 |
| 5'-GGG-GCA-GCA-3' | Ref. 5 1.45 1.37 1.68 | 2.34 |
| 5'-GGT-GAA-GGT-3' | Ref. 5 1.42 2.17 1.24 | 3.15 |
| 5'-GTA-GTT-GCC-3' | Ref. 5 1.15 2.92 | 7.61 |
| 5'-GTG-GCC-GGC-3' | Ref. 5 0.87 0.92 | 3.69 |
| 5'-GTT-GAG-GGT-3' | Ref. 5 0.99 Toxic 8.14 | 12.46 |
| 5'-GTT-GTC-GCC-3' | Ref. 5 0.80 4.45 | 2.83 |
| 5'-GCA-GAA-AGG-3' | Ref. 3 0.94 1.92 1.13 1.62 | 7.23 |
| 5'-TAG-GGA-GGC-3' | Ref. 5 1.94 | 2.40 |

# Best 2 3 % 1 20
% Best 8% 12% 4% 77%
Average Fold-Activation in B2H Reporter Assay 1.38 2.09 2.55 5.60

Nature Methods: doi: 10.1038/nmeth.1542
Supplementary Table 5  Sequences of oligonucleotide primers used in this study

| Primer Name | Primer Sequence                                      |
|-------------|------------------------------------------------------|
| OK1424      | 5'-GAGCGCCCCCTCCAGTGTCG-3’                           |
| OK1427      | 5'-TCGGCATTGGAATGGCTTCTCG-3’                         |
| OK1428      | 5'-GCCATTCCAATGCCGAATATGCA-3’                        |
| OK1429      | 5'-CCCTCCAATGCCGAATATGCA-3’                         |
| OK1430      | 5'-GGGGAGCGCCCCTCCAGTGTCG-3’                         |
| OK1432      | 5'-GTGCAGAGGATCCCTCAGGGTTTCTTAGGTG-3’               |
| OK61        | 5'-GGGTATACGATGACGGAACCTGTC-3’                       |
Supplementary Discussion

Direct comparisons of CoDA and modular assembly zinc finger arrays for 26 target DNA sites

The DNA sites used for this experiment were chosen from among 104 sites we had previously tested to assess the efficacy of modular assembly\(^6\) (Supplementary Table 4) and represent all of these sites for which finger arrays can currently be made using CoDA. Nearly all of these sites (24 out of 26) matched the consensus sequence 5’GNNGNGNN3’, a category of target sites for which modular assembly showed the highest success rates in our earlier report.\(^6\) In addition, it is important to note that although we made and tested only one CoDA finger array for each of the 26 target sites, multiple modularly assembled arrays (two to six arrays) were made and tested for nearly all (25 of the 26) sites (Supplementary Table 4), using three previously published module archives.\(^2\)\(^-\)\(^6\)

Our results demonstrate that CoDA yielded the zinc finger array with the highest B2H assay activity for 20 of the 26 target sites (Supplementary Table 4). Furthermore, the mean B2H fold-activation of all CoDA proteins tested (5.59-fold) is higher than those made using the three different modular assembly sets (1.43-, 2.11-, and 2.53-fold; Supplementary Table 4). To further compare CoDA and modular assembly, we examined fold-activation values in the B2H reporter assay of the most active protein made by each of the two methods for the 26 target DNA sites. Of these proteins, ~38% of the modular assembled arrays activated transcription by 1.57-fold or less in the B2H compared with 0% of the CoDA arrays (Supplementary Figure 2). Furthermore, ~23% of the modularly assembled arrays activated transcription by three-fold or more in the B2H assay compared with ~69% of the CoDA arrays (Supplementary Figure 2).
Comparison of mutation frequencies induced by ZFNs made using CoDA and other engineering platforms

The range of CoDA ZFN-induced mutation frequencies we observed in zebrafish somatic cell experiments (≤1% to 16.7%) are similar but somewhat lower than those from previously published experiments. Somatic mutation rates were reported to be 3% to 20% for ZFNs made by OPEN or B1H selection\textsuperscript{7,8} and 3% to 32% for ZFNs made by the proprietary Sangamo platform.\textsuperscript{9} Nonetheless, our previous experience suggests that the frequencies of somatic mutations we observed are high enough to make it likely that germline founders could be readily identified using these ZFNs (ref. 7 and unpublished data). For Arabidopsis, the frequencies of mutagenesis (as measured by number of mutated alleles) induced by CoDA ZFNs are comparable to those previously observed with ZFNs made by OPEN.\textsuperscript{10} No comparisons to prior experiments could be made for the soybean experiments because, to our knowledge, these are the first examples of ZFN-targeted mutations in endogenous soybean genes.

Predicted Targeting Range of CoDA ZFNs in Random DNA Sequence

The number of 9 bp sequences that can be targeted for each F2 triplet can be calculated as the product of F1 and F3 domains selected to function well in the context of the fixed F2 anchor finger for that triplet (i.e. CoDA targets where F2 target is GGG = 23 F1s x 20 F3s). Therefore, total number of 9 bp sequences that can be targeted with our current CoDA finger units is the sum of the 9bp targets for each of the 18 different F2 triplets which equals 6680 9bp sequences or approximately 2.55% of all possible 9bp DNA sequences. Because each ZFN requires two 9bp arrays and can be designed with three spacer sizes (5, 6 or 7 bp),\textsuperscript{11,12} CoDA targets are expected to occur about every 500bp of random sequence ($\frac{1}{2.55\% \times 2.55\% \times 3}$). We note that the theoretical targeting range in random DNA sequence for ZFNs made using the proprietary Sangamo platform has been reported to be 1 in ~31 bp of random sequence,\textsuperscript{13} a capability currently superior to that of CoDA. Therefore, an important priority for future experiments will be
to identify additional context-sensitive finger units for CoDA to further expand the range of potentially targetable genes and sequences.

**Modified ZiFiT software for identifying potential CoDA ZFN target sites**

Our new ZiFiT V3.3 program can be accessed at: [http://bindr.gdcb.iastate.edu/ZiFiT/](http://bindr.gdcb.iastate.edu/ZiFiT/) or at [http://www.zincfingers.org/software-tools.htm](http://www.zincfingers.org/software-tools.htm). Output from the ZiFiT program (Supplementary Figure 5) provides the sequence of potential CoDA target sites and unique identification numbers for requesting plasmids encoding the finger units required to assemble arrays (individual CoDA zinc finger units can be requested from the Joung lab by any interested academic researcher). Alternatively, ZiFiT can also generate DNA sequences encoding CoDA zinc finger arrays required to target a given site; these <290 bp DNA fragments can be synthesized through a commercial provider and then seamlessly cloned into existing Zinc Finger Consortium ZFN expression vectors.1, 7, 14, 15
References:

1. Maeder, M.L., Thibodeau-Beganny, S., Sander, J.D., Voytas, D.F. & Joung, J.K. Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. *Nat Protoc* **4**, 1471-1501 (2009).

2. Wright, D.A. et al. Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc* **1**, 1637-1652 (2006).

3. Mandell, J.G. & Barbas, C.F., 3rd Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* **34**, W516-523 (2006).

4. Liu, Q., Xia, Z., Zhong, X. & Case, C.C. Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J Biol Chem* **277**, 3850-3856 (2002).

5. Bae, K.H. et al. Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol* **21**, 275-280 (2003).

6. Ramirez, C.L. et al. Unexpected failure rates for modular assembly of engineered zinc-fingers. *Nat Methods* **5**, 374-375 (2008).

7. Foley, J.E. et al. Rapid Mutation of Endogenous Zebrafish Genes Using Zinc Finger Nucleases Made by Oligomerized Pool ENgineering. *PLoS ONE* **4**, e4348 (2009).

8. Meng, X., Noyes, M.B., Zhu, L.J., Lawson, N.D. & Wolfe, S.A. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol* **26**, 695-701 (2008).

9. Doyon, Y. et al. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol* **26**, 702-708 (2008).

10. Zhang, F. et al. High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. *Proc Natl Acad Sci U S A* **107**, 12028-12033 (2010).

11. Smith, J. et al. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res* **28**, 3361-3369 (2000).
12. Handel, E.M., Alwin, S. & Cathomen, T. Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. *Mol Ther* **17**, 104-111 (2009).

13. Shukla, V.K. et al. Precise genome modification in the crop species Zea mays using zinc-finger nucleases. *Nature* **459**, 337-338 (2009).

14. Maeder, M.L. et al. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* **31**, 294-301 (2008).

15. Foley, J.E. et al. Targeted mutagenesis in zebrafish using customized zinc-finger nucleases. *Nat Protoc* **4**, 1855-1867 (2009).