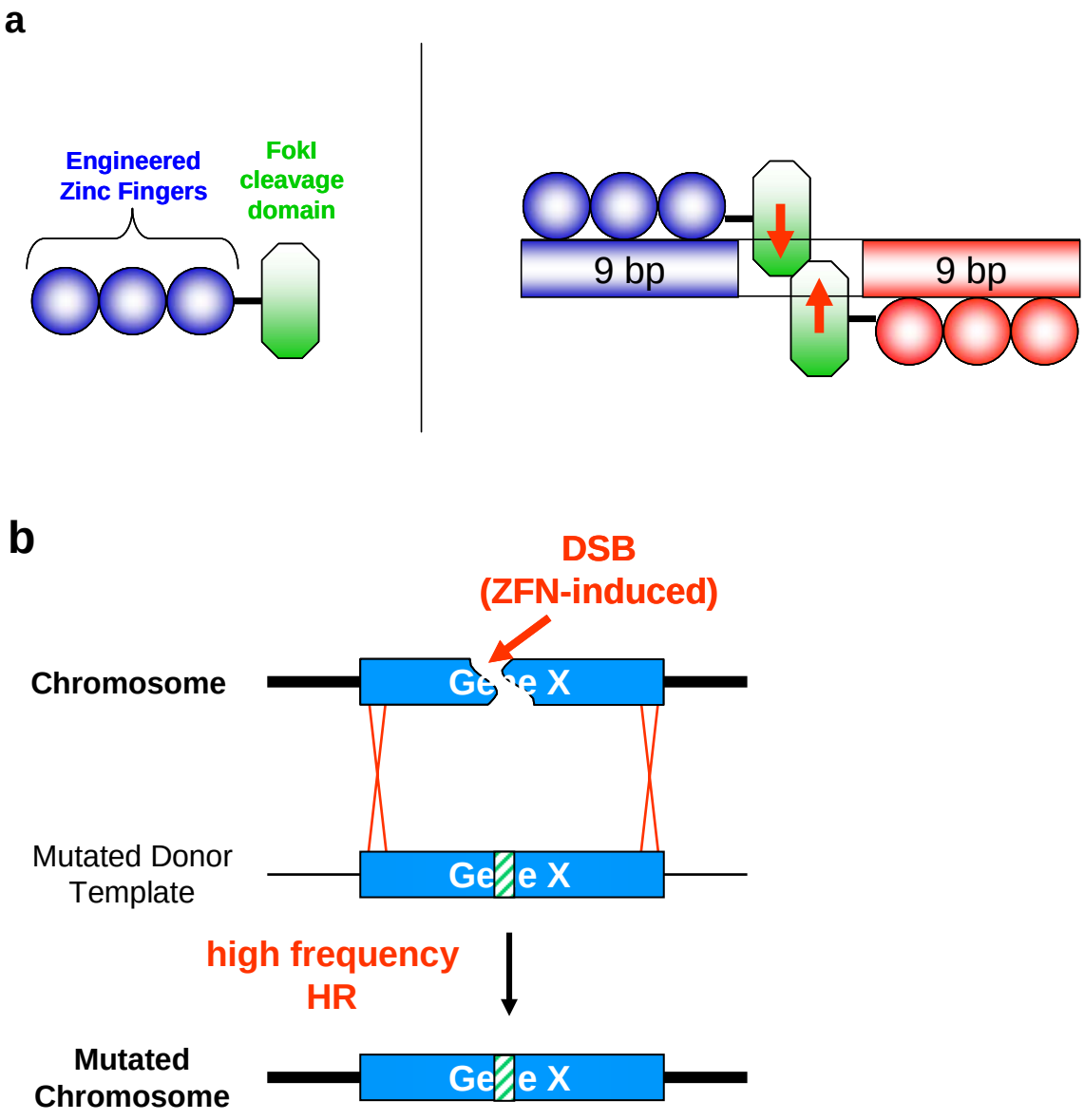
Correspondence

Unexpected failure rates for modular assembly of engineered zinc fingers

Cherie L Ramirez, Jonathan E Foley, David A Wright, Felix Müller-Lerch, Shamim H Rahman, Tatjana I Cornu, Ronnie J Winfrey, Jeffry D Sander, Fengli Fu, Jeffrey A Townsend, Toni Cathomen, Daniel F Voytas & J Keith Joung

Supplementary figures and text:

Supplementary Figure 1 ZFNs induce highly efficient gene targeting events

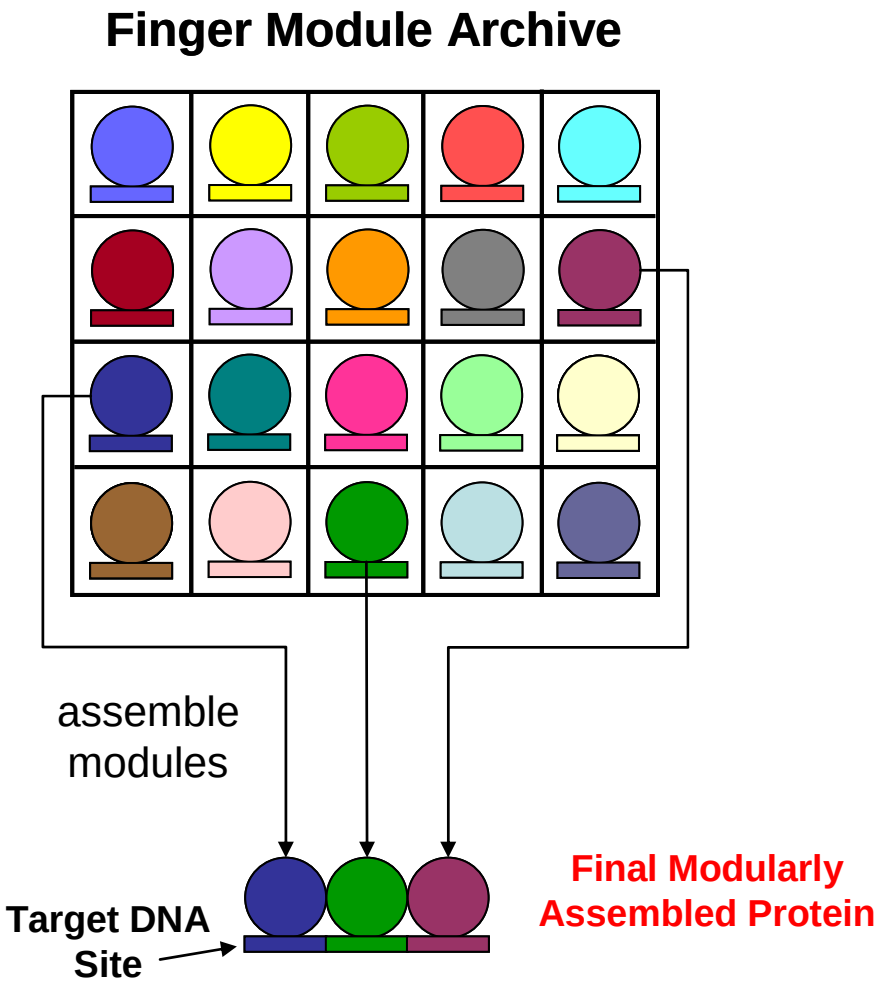


(a) Schematic of a ZFN monomer (left) and a pair of ZFNs cleaving DNA as a dimer (right).

(b) Schematic representing ZFN-enhanced recombination-based genome manipulation. ZFNs introduce site-specific double-strand DNA breaks (DSBs) that can be harnessed to

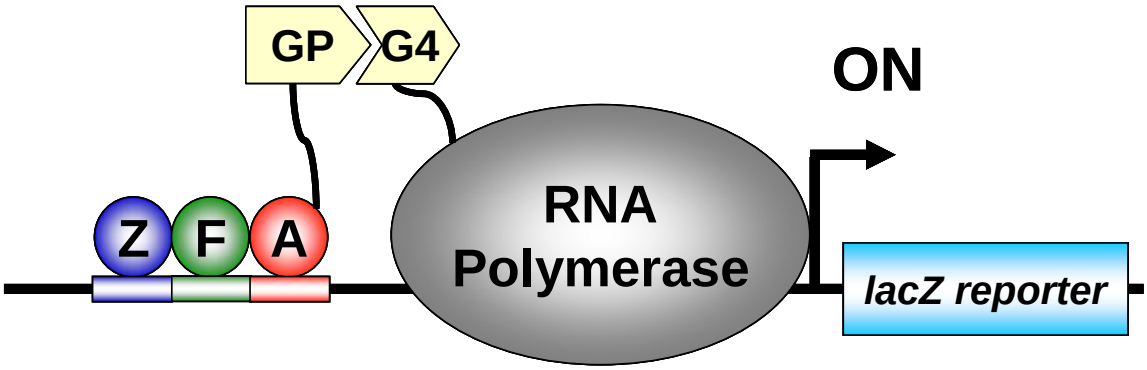
mediate gene targeting via homologous recombination with an exogenously introduced "donor template."

Supplementary Figure 2 Schematic illustrating the "modular assembly" method of engineering multi-finger domains



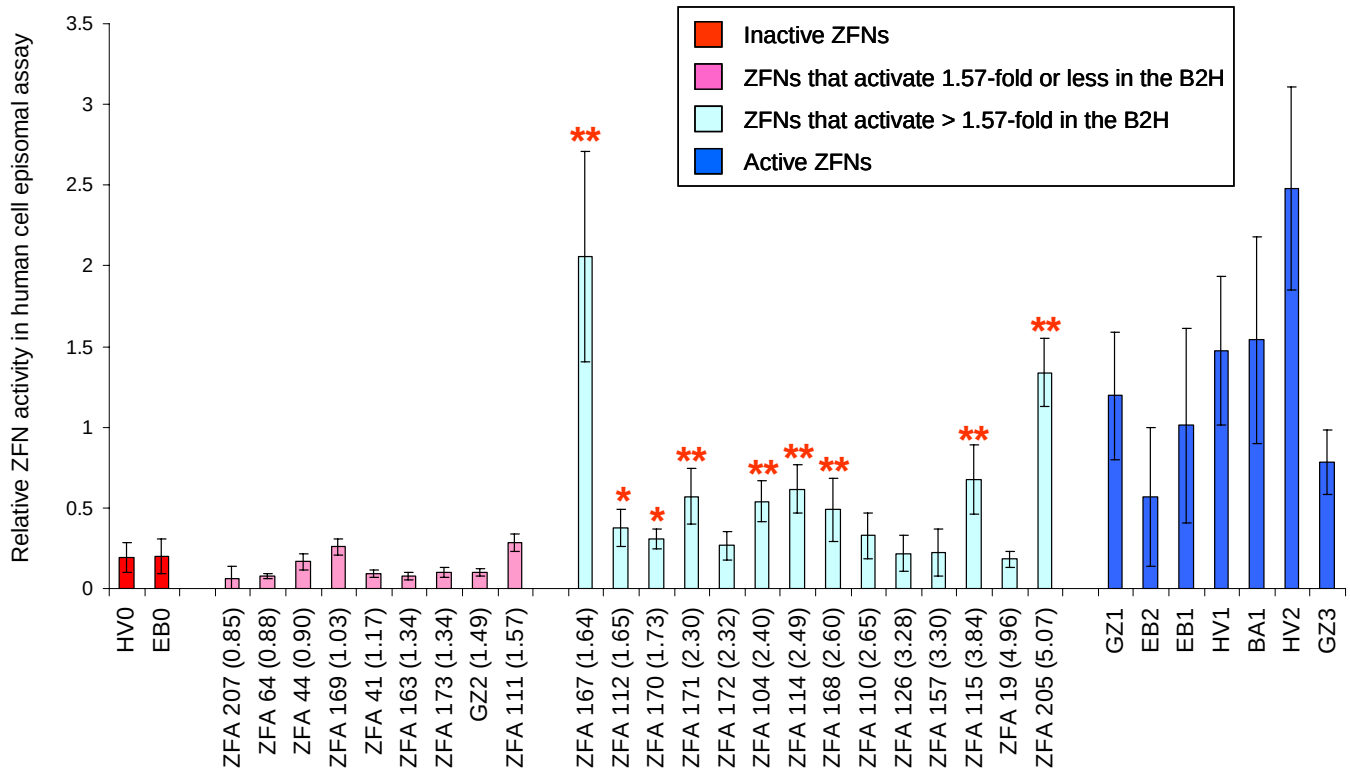
Fingers from archives of pre-characterized modules are joined together to create three-finger domains capable of recognizing a 9 bp target DNA site. Single fingers (colored spheres) recognize their cognate 3 bp "subsites" (colored rectangles).

Supplementary Figure 3 Schematic of the bacterial two-hybrid (B2H) reporter used to assess DNA-binding activities of zinc finger arrays



In the B2H assay, the ability of a zinc finger array to bind a target DNA site is reflected as increased expression of a *lacZ* reporter gene encoding β-galactosidase.[1-3] The B2H requires the expression of two hybrid proteins: one a fusion between the test zinc finger array (ZFA) and a fragment of the yeast Gal11P protein (GP) and the other a hybrid of the RNA polymerase α-subunit and a fragment of the yeast Gal4 protein (G4). If the test ZFA binds to a target DNA site positioned upstream of a promoter, RNA polymerase complexes that have incorporated Rα hybrid proteins are recruited to the promoter and expression of the downstream *lacZ* reporter is activated.

Supplementary Figure 4 The B2H assay identifies zinc-finger arrays that fail to show significant activity as ZFNs in human cells



23 different zinc-finger arrays (pink and light blue bars) which activated transcription in the B2H assay to various levels (fold-activations are indicated in parentheses on the x-axis) were tested as homodimeric ZFNs using a previously described human cell-based episomal recombination (HR) assay.[1] The names of these 23 zinc-finger arrays are shown on the x-axis and details about their binding sites and the modules used to construct them can be found in **Supplementary Tables 1** and **2** online. ZFN activity is reported as a percentage of EGFP-rescue relative to an I-*Sce*I control. In addition, the episomal repair activities of control ZFNs previously shown to lack or possess significant activity as ZFNs on chromosomal targets in human cells[1] are shown for comparison (red and blue bars, respectively). Names of the control ZFNs are shown on the x-axis and are as previously described.[1] Means of at least three independent experiments and standard deviations (error bars) are shown. ZFNs with mean episomal repair activities that are significantly higher

than that of the inactive EB0 ZFN are indicated with one ($p<0.05$) or two ($p<0.01$) red asterisks. The expression levels of all ZFNs shown were verified by Western blot using an antibody to a HA tag present on all proteins (data not shown).

Supplementary Figure 4 Reference:

1. Cornu, T. I. et al. DNA-binding Specificity Is a Major Determinant of the Activity and Toxicity of Zinc-finger Nucleases. *Mol Ther* **16**, 352-8 (2008).

Supplementary Table 1 Small-scale tests of modular assembly using various activity assays

ID#	F1	F2	F3	Target Sequence (F3-F2-F1)	# GXX subsites	Activity*	
						Assay	Results
ZFA 1	73	72	60	5′-GGT-GCT-GCC-3′	3	EMSA	-
ZFA 2	72	72	106	5′-TGG-GCT-GCT-3′	2	EMSA	-
ZFA 3	15	19	43	5′-GCT-GAC-GTG-3′	3	EMSA	-
ZFA 4	14	30	53	5′-AGG-GTC-GTC-3′	2	EMSA	-
ZFA 5	12	23	44	5′-GGA-GCG-GGT-3′	3	EMSA	-
ZFA 6	64	73	63	5′-GAA-GCC-GAT-3′	3	EMSA	-
ZFA 7	58	72	59	5′-GGA-GCT-GGG-3′	3	EMSA	-
ZFA 8	67	60	64	5′-GAT-GGT-GTA-3′	3	EMSA	-
ZFA 9	61	61	63	5′-GAA-GGC-GGC-3′	3	EMSA	-
ZFA 10	70	61	59	5′-GGA-GGC-GCG-3′	3	EMSA	-
ZFA 11	10	23	40	5′-GCA-GCG-GGC-3′	3	EMSA	-
ZFA 12**	15	19	43	5′-GCT-GAC-GTG-3′	3	EMSA	-
ZFA 13**	4	21	39	5′-GAT-GAT-GAT-3′	3	EMSA	-
ZFA 14**	15	31	132	5′-AGG-GTG-GTG-3′	2	EMSA	-
ZFA 15	61	70	71	5′-GCA-GCG-GGC-3′	3	EMSA	+
ZFA 16	62	68	63	5′-GAA-GTT-GAG-3′	3	EMSA	+
ZFA 17	60	70	59	5′-GGA-GCG-GGT-3′	3	EMSA	+
ZFA 18	67	66	65	5′-GAC-GTG-GTA-3′	3	EMSA	+
ZFA 19	66	66	84	5′-AGG-GTG-GTG-3′	2	EMSA	+
ZFA 20	62	62	62	5′-GAG-GAG-GAG-3′	3	EMSA	+
ZFA 21**	67	31	65	5′-GAC-GTG-GTA-3′	3	EMSA	-
						PSSA	-
ZFA 22**	8	24	57	5′-TGG-GCT-GCT-3′	2	EMSA	-
						PSSA	-
ZFA 23	15	31	49	5′-GTG-GTG-GTG-3′	3	EMSA	-
						PSSA	-
ZFA 24**	8	24	43	5′-GCT-GCT-GCT-3′	3	EMSA	-
						PSSA	-
ZFA 25**	15	31	84	5′-AGG-GTG-GTG-3′	2	EMSA	-
						PSSA	-
ZFA 26**	15	31	84	5′-AGG-GTG-GTG-3′	2	PSSA	-
ZFA 27**	13	31	37	5′-GAC-GTG-GTA-3′	3	PSSA	-
ZFA 28***	61	65	65	5′-GAC-GAC-GGC-3′	3	HR	-
						EMSA	-
ZFA 29	100	96	75	5′-AAC-CCT-CGT-3′	0	HR	-
ZFA 30	61	80	61	5′-GGC-ACG-GGC-3′	2	HR	-
ZFA 31	91	104	85	5′-AGT-TAG-CAG-3′	0	HR	-
ZFA 32	83	103	104	5′-TAG-CTT-AGC-3′	0	HR	-
ZFA 33	83	68	74	5′-AAA-GTT-AGC-3′	1	HR	-
ZFA 34	100	104	72	5′-GCT-TAG-CGT-3′	1	HR	-
ZFA 35	17	25	119	5′-AGA-GGA-AGA-3′	1	HR	-
ZFA 36***	60	64	63	5′-GAA-GAT-GGT-3′	3	HR	+
						EMSA	+

Thirty-six three-finger arrays (named "ZFA __" for zinc finger array) were assembled using modules based on the Zinc Finger Consortium Modular Assembly Kit 1.0,[1] and the identities of the three modules used to construct each three-finger array use the Consortium numbering scheme (F1, F2, and F3 are the amino-terminal, middle, and carboxy-terminal fingers, respectively). All target binding sites are written 5' to 3' (note that the F3 module binds to the 5' most triplet subsite while F1 binds to the 3' most triplet subsite). The number of GXX subsites in each target site is also indicated. For nine of the arrays, the amino acid sequence of the zinc finger backbone differs from the modules in the archive. These arrays are noted in the far left column by a double asterisk, and their complete amino acid sequences are available upon request.

Activity of 27 arrays was tested by electrophoretic mobility shift assays (EMSA). A plus sign in the far right column indicates that the finger array caused a shift in mobility of a double-stranded oligonucleotide corresponding to the intended target sequence on polyacrylamide gels. In all cases, binding could be competed away with an excess of the target oligonucleotide. Western blots were performed for all zinc finger arrays to ensure they were expressed.

Plant single-strand annealing (PSSA) assays (see **Supplementary Methods** online) were used to assess function of seven arrays, five of which were also tested by EMSA. The PSSA assay tests the ability of the zinc finger array to function as a ZFN. A minus sign in the table indicates that activity was comparable to negative controls in which plant protoplasts were not transformed with the ZFN construct.

An episomal recombination (HR) assay was used to assess the activity of nine ZFNs in human 293T cells (see **Supplementary Methods** online). A plus sign indicates that ZFN activity in stimulating HR was >40% as compared to the activity of a control I-*Sce*I meganuclease on the

same target locus.

*EMSA, electrophoretic mobility shift assay; a positive value indicates DNA binding that could be competed away by an excess of an oligonucleotide corresponding to the target site. PSSA, plant single-strand annealing assay; a positive value indicates reporter gene function that is at least two-fold over background controls. HR, episomal recombination assay; a positive value indicates at least 40% activity as compared to I-*Sce*I on the same target template. For further information on the activity assays, see **Supplementary Methods** online.

**Finger arrays with backbone sequences that differ from the ZF finger archive.

***Finger arrays previously described.[2]

Reference for Supplementary Table 1:

1. Wright, D. A. et al. Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc* **1**, 1637-52 (2006).

2. Alwin, S. et al. Custom zinc-finger nucleases for use in human cells. *Mol Ther* **12**, 610-7 (2005).

Supplementary Table 2 Modularly assembled zinc finger arrays, cognate target binding sites, and their activities in the B2H assay

ID#	F1	F2	F3	Module Source	Target Sequence (F3-F2-F1)	# GXX subsites	B2H fold activation (mean)	SEM
ZFA 37	75	75	74	B	5'-AAA-AAC-AAC-3'	0	1.152	0.030
ZFA 38	91	78	74	B	5'-AAA-ACA-CAG-3'	0	1.097	0.030
ZFA 39	73	91	74	B	5'-AAA-CAG-GCC-3'	1	1.325	0.095
ZFA 40	100	101	74	B	5'-AAA-CTA-CGT-3'	0	1.347	0.037
ZFA 41	83	68	74	B	5'-AAA-GTT-AGC-3'	1	1.166	0.106
ZFA 42	91	93	75	B	5'-AAC-CCA-CAG-3'	0	0.974	0.075
ZFA 43	68	93	75	B	5'-AAC-CCA-GTT-3'	1	0.980	0.010
ZFA 44	100	96	75	B	5'-AAC-CCT-CGT-3'	0	0.900	0.105
ZFA 45	64	86	77	B	5'-AAT-ATA-GAT-3'	1	1.606	0.037
ZFA 46	90	75	78	B	5'-ACA-AAC-CAC-3'	0	0.850	0.022
ZFA 47	85	75	79	B	5'-ACC-AAC-AGT-3'	0	1.001	0.028
ZFA 48	86	83	79	B	5'-ACC-AGC-ATA-3'	0	1.036	0.028
ZFA 49	83	100	79	B	5'-ACC-CGT-AGC-3'	0	1.235	0.016
ZFA 50	63	59	79	B	5'-ACC-GGA-GAA-3'	2	1.386	0.068
ZFA 51	91	83	80	B	5'-ACG-AGC-CAG-3'	0	0.955	0.027
ZFA 52	72	91	80	B	5'-ACG-CAG-GCT-3'	1	1.236	0.007
ZFA 53	70	95	80	B	5'-ACG-CCG-GCG-3'	1	1.266	0.040
ZFA 54	100	65	80	B	5'-ACG-GAC-CGT-3'	1	1.799	0.060
ZFA 55	71	73	80	B	5'-ACG-GCC-GCA-3'	2	1.295	0.031
ZFA 56	82	59	82	B	5'-AGA-GGA-AGA-3'	1	1.287	0.021
ZFA 57	119	117	119	T	5'-AGA-GGA-AGA-3'	1	0.911	0.012
ZFA 58	119	114	119	T	5'-AGA-GGA-AGA-3'	1	1.928	0.082
ZFA 59	119	129	119	T	5'-AGA-GGA-AGA-3'	1	1.974	0.066
ZFA 60	119	118	119	T	5'-AGA-GGA-AGA-3'	1	2.371	0.047
ZFA 61	64	69	83	B	5'-AGC-GTC-GAT-3'	2	1.206	0.019
ZFA 62	103	88	85	B	5'-AGT-ATT-CTT-3'	0	1.015	0.022
ZFA 64	91	104	85	B	5'-AGT-TAG-CAG-3'	0	0.879	0.092
ZFA 65	85	93	86	B	5'-ATA-CCA-AGT-3'	0	0.989	0.017
ZFA 66	78	81	87	B	5'-ATG-ACT-ACA-3'	0	0.956	0.043
ZFA 67	92	90	87	B	5'-ATG-CAC-CAT-3'	0	1.122	0.068
ZFA 68	78	96	88	B	5'-ATT-CCT-ACA-3'	0	1.064	0.135
ZFA 69	95	72	88	B	5'-ATT-GCT-CCG-3'	1	1.256	0.016
ZFA 70	63	98	90	B	5'-CAC-CGC-GAA-3'	1	1.099	0.011
ZFA 71	60	67	90	B	5'-CAC-GTA-GGT-3'	2	1.476	0.023
ZFA 72	86	98	92	B	5'-CAT-CGC-ATA-3'	0	0.985	0.027
ZFA 74	90	100	93	B	5'-CCA-CGT-CAC-3'	0	1.173	0.034
ZFA 75	73	64	93	B	5'-CCA-GAT-GCC-3'	2	1.085	0.006
ZFA 76	103	85	94	B	5'-CCC-AGT-CTT-3'	0	1.376	0.153
ZFA 77	66	100	94	B	5'-CCC-CGT-GTG-3'	1	1.431	0.069
ZFA 78	80	79	95	B	5'-CCG-ACC-ACG-3'	0	1.090	0.020
ZFA 79	74	91	95	B	5'-CCG-CAG-AAA-3'	0	1.123	0.009
ZFA 80	100	98	95	B	5'-CCG-CGC-CGT-3'	0	1.146	0.074
ZFA 81	86	63	95	B	5'-CCG-GAA-ATA-3'	1	0.907	0.048
ZFA 82	91	80	96	B	5'-CCT-ACG-CAG-3'	0	1.013	0.037
ZFA 83	90	81	96	B	5'-CCT-ACT-CAC-3'	0	1.113	0.042
ZFA 84	70	87	96	B	5'-CCT-ATG-GCG-3'	1	1.239	0.104
ZFA 85	58	67	96	B	5'-CCT-GTA-GGG-3'	2	1.504	0.009

ZFA 86	101	92	98	B	5'-CGC-CAT-CTA-3'	0	1.052	0.052
ZFA 87	88	81	101	B	5'-CTA-ACT-ATT-3'	0	0.884	0.013
ZFA 88	83	59	101	B	5'-CTA-GGA-AGC-3'	1	1.303	0.089
ZFA 89	80	83	103	B	5'-CTT-AGC-ACG-3'	0	1.037	0.086
ZFA 90	94	88	103	B	5'-CTT-ATT-CCC-3'	0	1.031	0.018
ZFA 91	91	60	103	B	5'-CTT-GGT-CAG-3'	1	1.044	0.032
ZFA 92	69	70	63	B	5'-GAA-GCG-GTC-3'	3	2.309	0.129
ZFA 93	14	23	36	S	5'-GAA-GCG-GTC-3'	3	1.620	0.059
ZFA 94	109	130	124	T	5'-GAA-GCG-GTC-3'	3	1.339	0.165
ZFA 95	109	130	122	T	5'-GAA-GCG-GTC-3'	3	2.279	0.123
ZFA 96	87	94	65	B	5'-GAC-CCC-ATG-3'	1	1.074	0.034
ZFA 97	72	61	65	B	5'-GAC-GGC-GCT-3'	3	1.118	0.045
ZFA 98	8	26	37	S	5'-GAC-GGC-GCT-3'	3	0.930	0.039
ZFA 101	4	24	38	S	5'-GAG-GCT-GAT-3'	3	1.327	0.126
ZFA 102	112	139	113	T	5'-GAG-GCT-GAT-3'	3	1.142	0.091
ZFA 103	112	139	136	T	5'-GAG-GCT-GAT-3'	3	1.185	0.087
ZFA 104	112	139	115	T	5'-GAG-GCT-GAT-3'	3	2.399	0.120
ZFA 105	66	72	62	B	5'-GAG-GCT-GTG-3'	3	4.553	0.295
ZFA 106	15	24	38	S	5'-GAG-GCT-GTG-3'	3	3.038	0.153
ZFA 107	138	139	136	T	5'-GAG-GCT-GTG-3'	3	0.959	0.026
ZFA 108	138	139	113	T	5'-GAG-GCT-GTG-3'	3	1.004	0.037
ZFA 109	138	139	115	T	5'-GAG-GCT-GTG-3'	3	3.277	0.153
ZFA 110	68	59	64	B	5'-GAT-GGA-GTT-3'	3	2.647	0.158
ZFA 111	16	25	39	S	5'-GAT-GGA-GTT-3'	3	1.571	0.111
ZFA 112	111	118	112	T	5'-GAT-GGA-GTT-3'	3	1.645	0.071
ZFA 113	111	129	112	T	5'-GAT-GGA-GTT-3'	3	2.013	0.083
ZFA 114	111	117	112	T	5'-GAT-GGA-GTT-3'	3	2.489	0.136
ZFA 115	111	114	112	T	5'-GAT-GGA-GTT-3'	3	3.840	0.308
ZFA 116	58	60	64	B	5'-GAT-GGT-GGG-3'	3	2.390	0.065
ZFA 117	11	28	39	S	5'-GAT-GGT-GGG-3'	3	0.959	0.098
ZFA 118	131	140	112	T	5'-GAT-GGT-GGG-3'	3	1.985	0.067
ZFA 119	133	140	112	T	5'-GAT-GGT-GGG-3'	3	2.444	0.103
ZFA 120	134	140	112	T	5'-GAT-GGT-GGG-3'	3	6.665	0.290
ZFA 121	60	78	71	B	5'-GCA-ACA-GGT-3'	2	1.388	0.014
ZFA 122	100	80	71	B	5'-GCA-ACG-CGT-3'	1	1.274	0.031
ZFA 123	81	90	71	B	5'-GCA-CAC-ACT-3'	1	0.851	0.023
ZFA 124	59	73	71	B	5'-GCA-GCC-GGA-3'	3	1.574	0.099
ZFA 125	9	33	40	S	5'-GCA-GCC-GGA-3'	3	0.953	0.111
ZFA 126	71	70	71	B	5'-GCA-GCG-GCA-3'	3	3.276	0.111
ZFA 127	5	23	40	S	5'-GCA-GCG-GCA-3'	3	0.887	0.068
ZFA 128	58	72	71	B	5'-GCA-GCT-GGG-3'	3	1.186	0.045
ZFA 129	11	24	40	S	5'-GCA-GCT-GGG-3'	3	1.157	0.039
ZFA 130	65	69	71	B	5'-GCA-GTC-GAC-3'	3	0.867	0.046
ZFA 131	2	30	40	S	5'-GCA-GTC-GAC-3'	3	0.897	0.052
ZFA 132	72	69	71	B	5'-GCA-GTC-GCT-3'	3	1.365	0.049
ZFA 133	8	30	40	S	5'-GCA-GTC-GCT-3'	3	1.116	0.092
ZFA 134	61	69	71	B	5'-GCA-GTC-GGC-3'	3	1.077	0.024
ZFA 135	10	30	40	S	5'-GCA-GTC-GGC-3'	3	0.999	0.039
ZFA 136	70	66	71	B	5'-GCA-GTG-GCG-3'	3	1.915	0.050
ZFA 137	7	31	40	S	5'-GCA-GTG-GCG-3'	3	0.819	0.051
ZFA 138	95	79	73	B	5'-GCC-ACC-CCG-3'	1	1.000	0.032
ZFA 139	59	87	73	B	5'-GCC-ATG-GGA-3'	2	0.974	0.043
ZFA 140	103	63	73	B	5'-GCC-GAA-CTT-3'	2	1.137	0.057

ZFA 141	66	71	73	B	5'-GCC-GCA-GTG-3'	3	1.061	0.036
ZFA 142	15	22	41	S	5'-GCC-GCA-GTG-3'	3	0.914	0.023
ZFA 143	59	61	73	B	5'-GCC-GGC-GGA-3'	3	1.182	0.049
ZFA 144	9	26	41	S	5'-GCC-GGC-GGA-3'	3	0.814	0.021
ZFA 145	132	116	130	T	5'-GCG-GAA-AGG-3'	2	1.342	0.134
ZFA 146	132	124	130	T	5'-GCG-GAA-AGG-3'	2	1.736	0.063
ZFA 147	132	122	130	T	5'-GCG-GAA-AGG-3'	2	1.759	0.054
ZFA 148	132	123	130	T	5'-GCG-GAA-AGG-3'	2	2.136	0.064
ZFA 149	102	59	70	B	5'-GCG-GGA-CTG-3'	2	1.522	0.088
ZFA 150	59	66	70	B	5'-GCG-GTG-GGA-3'	3	1.751	0.124
ZFA 151	9	31	42	S	5'-GCG-GTG-GGA-3'	3	2.774	0.116
ZFA 152	114	138	130	T	5'-GCG-GTG-GGA-3'	3	1.029	0.046
ZFA 153	129	138	130	T	5'-GCG-GTG-GGA-3'	3	1.272	0.039
ZFA 154	118	138	130	T	5'-GCG-GTG-GGA-3'	3	1.384	0.059
ZFA 155	117	138	130	T	5'-GCG-GTG-GGA-3'	3	1.953	0.050
ZFA 156	60	66	70	B	5'-GCG-GTG-GGT-3'	3	2.229	0.208
ZFA 157	12	31	42	S	5'-GCG-GTG-GGT-3'	3	3.298	0.266
ZFA 158	70	106	70	B	5'-GCG-TGG-GCG-3'	2	2.713	0.181
ZFA 159	7	34	42	S	5'-GCG-TGG-GCG-3'	2	0.856	0.034
ZFA 160	78	86	72	B	5'-GCT-ATA-ACA-3'	1	0.988	0.024
ZFA 161	62	100	72	B	5'-GCT-CGT-GAG-3'	2	1.703	0.012
ZFA 163	100	104	72	B	5'-GCT-TAG-CGT-3'	1	1.343	0.143
ZFA 164	61	74	59	B	5'-GGA-AAA-GGC-3'	2	1.392	0.084
ZFA 165	92	90	59	B	5'-GGA-CAC-CAT-3'	1	1.182	0.016
ZFA 166	80	101	59	B	5'-GGA-CTA-ACG-3'	1	1.085	0.034
ZFA 167	66	70	59	B	5'-GGA-GCG-GTG-3'	3	1.640	0.043
ZFA 168	15	23	44	S	5'-GGA-GCG-GTG-3'	3	2.599	0.092
ZFA 169	138	130	117	T	5'-GGA-GCG-GTG-3'	3	1.027	0.068
ZFA 170	138	130	118	T	5'-GGA-GCG-GTG-3'	3	1.732	0.041
ZFA 171	138	130	129	T	5'-GGA-GCG-GTG-3'	3	2.296	0.121
ZFA 172	138	130	114	T	5'-GGA-GCG-GTG-3'	3	2.316	0.046
ZFA 173	61	80	61	B	5'-GGC-ACG-GGC-3'	2	1.341	0.109
ZFA 174	63	95	61	B	5'-GGC-CCG-GAA-3'	2	1.171	0.030
ZFA 175	66	73	61	B	5'-GGC-GCC-GTG-3'	3	1.116	0.072
ZFA 176	15	33	45	S	5'-GGC-GCC-GTG-3'	3	0.868	0.017
ZFA 177	73	58	61	B	5'-GGC-GGG-GCC-3'	3	1.203	0.035
ZFA 178	6	27	45	S	5'-GGC-GGG-GCC-3'	3	1.258	0.053
ZFA 180	61	70	58	B	5'-GGG-GCG-GGC-3'	3	2.191	0.105
ZFA 181	10	23	51	S	5'-GGG-GCG-GGC-3'	3	6.162	0.147
ZFA 182	73	72	58	B	5'-GGG-GCT-GCC-3'	3	1.363	0.053
ZFA 183	6	24	51	S	5'-GGG-GCT-GCC-3'	3	1.494	0.054
ZFA 184	108	139	134	T	5'-GGG-GCT-GCC-3'	3	2.022	0.253
ZFA 185	79	88	60	B	5'-GGT-ATT-ACC-3'	1	1.078	0.016
ZFA 186	72	63	60	B	5'-GGT-GAA-GCT-3'	3	1.589	0.094
ZFA 187	8	18	46	S	5'-GGT-GAA-GCT-3'	3	0.944	0.081
ZFA 188	139	122	140	T	5'-GGT-GAA-GCT-3'	3	1.176	0.125
ZFA 189	139	124	140	T	5'-GGT-GAA-GCT-3'	3	1.800	0.124
ZFA 190	139	123	140	T	5'-GGT-GAA-GCT-3'	3	2.189	0.234
ZFA 191	139	116	140	T	5'-GGT-GAA-GCT-3'	3	2.431	0.090
ZFA 192	98	103	67	B	5'-GTA-CTT-CGC-3'	1	1.274	0.034
ZFA 193	61	68	67	B	5'-GTA-GTT-GGC-3'	3	1.056	0.098
ZFA 194	10	32	47	S	5'-GTA-GTT-GGC-3'	3	0.854	0.028
ZFA 195	91	101	69	B	5'-GTC-CTA-CAG-3'	1	1.132	0.035

ZFA 196	81	65	66	B	5'-GTG-GAC-ACT-3'	2	1.387	0.028
ZFA 197	73	61	66	B	5'-GTG-GGC-GCC-3'	3	1.104	0.036
ZFA 198	6	26	49	S	5'-GTG-GGC-GCC-3'	3	0.984	0.055
ZFA 199	66	82	68	B	5'-GTT-AGA-GTG-3'	2	1.683	0.046
ZFA 200	138	119	111	T	5'-GTT-AGA-GTG-3'	2	2.872	0.107
ZFA 201	12	20	50	S	5'-GTT-GAG-GGT-3'	3	0.671	0.015
ZFA 202	140	136	111	T	5'-GTT-GAG-GGT-3'	3	3.477	0.076
ZFA 203	140	113	111	T	5'-GTT-GAG-GGT-3'	3	4.735	0.489
ZFA 204	101	71	68	B	5'-GTT-GCA-CTA-3'	2	1.009	0.043
ZFA 205	73	70	68	B	5'-GTT-GCG-GCC-3'	3	5.065	0.290
ZFA 206	6	23	50	S	5'-GTT-GCG-GCC-3'	3	1.232	0.014
ZFA 207	83	103	104	B	5'-TAG-CTT-AGC-3'	0	0.849	0.100
ZFA 208	70	59	104	B	5'-TAG-GGA-GCG-3'	2	2.185	0.068
ZFA 209	3	20	56	S	5'-TGA-GAG-GAG-3'	2	1.704	0.068
ZFA 210	86	105	105	B	5'-TGA-TGA-ATA-3'	0	1.001	0.030

See legend to **Supplementary Table 1** online for descriptive detail pertaining to each of this table's columns. In addition, the original source of the modules used to construct each array (B=Barbas group,[1] S=Sangamo Biosciences,[2, 3] and T=Toolgen, Inc[4].) is indicated. Because the Zinc Finger Consortium Modular Assembly Kit 1.0 includes more than one module for certain subsites,[3] we were able to construct 168 zinc finger arrays that could potentially recognize the 104 different target sites. Additional details regarding the construction of the B2H zinc finger expression plasmids and the B2H binding site reporter plasmids are provided in **Supplementary Methods** online. Fold-activation of transcription in the B2H assay was determined for each zinc finger array using β-galactosidase assays as previously described.[5] β-galactosidase assays for each zinc finger array were performed a minimum of four times and means and standard errors of the mean are shown. The expression of all zinc finger arrays that failed to activate transcription by more than 1.57-fold in the B2H assay were verified by Western blot using a monoclonal antibody which detects a FLAG epitope present on all arrays (data not shown).

References for Supplementary Table 2:

1. 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).

2. 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).

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

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

5. Cornu, T.I. et al. DNA-binding Specificity Is a Major Determinant of the Activity and Toxicity of Zinc-finger Nucleases. *Mol Ther* (2007).

Supplementary Discussion:

Small-scale surveys suggest a low success rate for modular assembly

Motivated by a desire to use ZFNs for genome modification, our labs constructed the 36 zinc finger arrays listed in **Supplementary Table 1** online. These arrays are highly biased for GXX subsites, and collectively include 22 3-GXX arrays, 8 2-GXX arrays, 2 1-GXX arrays and 4 arrays without GXX subsites. Using three different assays to test for function (see legend to **Supplementary Table 1** online), seven arrays were deemed functional. Six of these sites were composed of three GXX subsites and one was composed of two GXX subsites. Since six of the seven arrays were only tested by EMSA, it is difficult to extrapolate how many of these arrays would function *in vivo* when challenged with the diverse sequence targets present in a genome.

The B2H assay accurately identifies zinc finger arrays that fail to show ZFN function in human cells

In previously published studies from our groups, we observed a general correlation between failure to activate in the B2H assay and failure to show ZFN function in human cells.[1] To further assess this correlation, we tested the activities of 23 ZFNs using a human cell-based episomal homologous recombination (HR) assay (**Supplementary Methods** online). The zinc finger arrays in these 23 ZFNs each activated transcription to different levels in the B2H system ranging from 0.85- to 5.07-fold. As shown in **Supplementary Fig. 4** online, we found that zinc finger arrays which activated transcription by 1.57-fold or less in the B2H system all failed to show significant activity as ZFNs in human cells (nine out of nine ZFNs tested; pink bars in **Supplementary Fig.**

4 online). Conversely, many (although not all) of the zinc finger arrays which activated transcription by >1.57-fold in the B2H system showed significant activity as ZFNs in the episomal repair assay (light blue bars with red asterisks in **Supplementary Fig. 4** online). We conclude that zinc finger arrays that activate transcription by 1.57-fold or less in the B2H system are unlikely to function as ZFNs in human cells and we used this threshold level to interpret the results of our experiments.

Predicted failure rates for identifying zinc finger arrays needed to engineer a ZFN pair

Analysis of our results shows that modular assembly is far less effective for target sites composed of two, one, or no GXX subsites (71%, 88% and 100% failure rates, respectively; **Fig. 1a**) compared with those that contain three GXX subsites (41% failure rate; **Fig. 1a**). Because ZFNs function as dimers, the failure rates for making a ZFN pair will therefore be ~65%, ~92%, ~99%, and 100% for ZFN targets composed of pairs of 3-GXX, 2-GXX, 1-GXX, and 0-GXX 9 bp "half-sites", respectively. These failure rates are calculated by multiplying estimated success rates for each monomer in the ZFN pair and subtracting this percentage from 100%.

Alternative selection-based strategies for engineering multi-finger arrays

A fundamental flaw with the modular assembly method (and a likely cause of its low success rate) is its assumption that zinc finger domains behave as independent modular units. A number of studies have shown that zinc fingers can cross over and interact with adjacent fingers and neighboring finger binding sites.[2-5] Various engineering strategies have been described in the literature that account for these context-dependent effects on

zinc finger DNA-binding. Greisman and Pabo first described a sequential optimization strategy in which combinations of fingers that work well together are identified using serial selections from randomized libraries.[6] Isalan, Klug, and Choo described a "bi-partite" optimization strategy in which "halves" of a three-finger domain are first optimized by randomization and selection and then joined together to create a final protein.[7] Finally, Joung and colleagues described a domain shuffling approach in which pools of fingers are first identified for each subsite in parallel and then shuffled together to create a recombined library for use in a final stringent selection.[8] A limitation of all of these different approaches is that they require specialized expertise in the construction of multiple very large randomized libraries and in the interrogation of these libraries using selection methods.

References for Supplementary Discussion:

1. Cornu, T. I. et al. DNA-binding Specificity Is a Major Determinant of the Activity and Toxicity of Zinc-finger Nucleases. *Mol Ther* (2007).
2. Isalan, M., Choo, Y. & Klug, A. Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc Natl Acad Sci U S A* **94**, 5617-21 (1997).
3. Isalan, M., Klug, A. & Choo, Y. Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry* **37**, 12026-33 (1998).
4. Elrod-Erickson, M., Rould, M. A., Nekludova, L. & Pabo, C. O. Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. *Structure* **4**, 1171-80 (1996).

5. Wolfe, S. A., Grant, R. A., Elrod-Erickson, M. & Pabo, C. O. Beyond the "recognition code": structures of two Cys2His2 zinc finger/TATA box complexes. *Structure (Camb)* **9**, 717-23 (2001).

6. Greisman, H. A. & Pabo, C. O. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**, 657-61 (1997).

7. Isalan, M., Klug, A. & Choo, Y. A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat Biotechnol* **19**, 656-60 (2001).

8. Hurt, J. A., Thibodeau, S. A., Hirsh, A. S., Pabo, C. O. & Joung, J. K. Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc Natl Acad Sci U S A* **100**, 12271-6 (2003).

Supplementary Methods:

Construction of B2H zinc finger expression vectors and reporter strains

104 B2H reporter strains each harboring a single copy target site-reporter plasmid were constructed as previously described.[1] B2H expression vectors encoding different three-finger arrays were constructed essentially as previously described[1] using the Zinc Finger Modular Assembly Kit 1.0 which includes 141 modules made by the Barbas group, Sangamo Biosciences, and Toolgen, Inc.[1] For each target site, we made three-finger arrays using modules from the Barbas, Sangamo, and Toolgen collections but we did not mix and match modules across these different sets. We chose not to make "cross-platform" arrays because: (1) the Barbas website software does not advocate use of their modules with others,[2] (2) the Toolgen group discovered that their human zinc fingers worked best with one another and not as well with other engineered modules,[3] and (3) the Sangamo modules are position-specific and have linkers joining them that differ from the canonical TGEKP linker used by the Barbas and Toolgen modules.[4] 15 of the arrays we constructed were toxic when expressed in the B2H assay and these proteins were not included in our analysis. After setting aside these toxic proteins, we characterized the remaining 168 zinc finger arrays. All expression and reporter plasmids were confirmed to be correct by DNA sequencing prior to use in B2H assays.

B2H assays, verification of protein expression, and re-verification of DNA sequences

The activities of three-finger arrays were tested in the B2H assay by co-transforming a target site reporter strain with a B2H zinc finger array-Gal11P expression vector and a compatible plasmid encoding an *E. coli* RNA polymerase alpha-Gal4 hybrid protein.[1]

Fold-activation was calculated as the ratio of the level of β-galactosidase (*lacZ*) reporter activity in the presence and absence of the zinc finger array as previously described.[5] All β-galactosidase measurements were performed at least four independent times. The expression of each zinc finger array that failed to activate transcription by more than 1.57-fold in the B2H assay was verified by performing Western blot on cell extracts from the β-galactosidase assays using a monoclonal antibody against the FLAG epitope present on all zinc finger array-Gal11P fusion proteins. For zinc finger arrays that exhibited less than 1.57-fold transcriptional activation in the B2H assay, we re-sequenced the zinc finger coding sequences and the reporter plasmid binding sites from the same cells used for β-galactosidase assays to re-confirm the identities of the zinc finger recognition helices and binding sites tested in these cells. We performed this additional re-sequencing control on >80% of our samples (94 of the 115 zinc finger array/binding site combinations) that were negative in the B2H assay and found that all helices and binding sites were correct as expected.

Plant single-strand annealing (PSSA) assays

Zinc finger arrays were fused to the *FokI* nuclease domain and the resulting ZFNs were transiently expressed in tobacco protoplasts. Plasmids encoding these ZFNs were co-transformed with a target plasmid carrying a non-functional β-glucuronidase (*GUS*) gene. The *GUS* gene was rendered non-functional by a direct duplication of part of its coding sequence. Between the direct repeats, a recognition site was inserted for the given ZFN being tested. After expression of the ZFN and cleavage of the *GUS* reporter, repair by single strand annealing can restore GUS function. This can be measured by standard

GUS activity assays. We routinely observe GUS activity 25-fold above background when the assay is performed with a ZFN based on the transcription factor Zif268.

Episomal recombination (HR) assay in human cells

For the episomal HR assay, 293T cells in 24-well plates were transfected by calcium phosphate precipitation with 20 ng of the respective target plasmid, 1 µg of the donor plasmid (pUC.Zgfp/REx), and 100 ng of a CMV-driven endonuclease expression vector encoding the ZFNs, I-SceI (pRK5.LHA-Sce1) or a control vector (pCMV.Luc). Two days after transfection, 50,000 cells were analyzed by flow cytometry (FACSCalibur, BD Bioscience) to determine the percentage of EGFP-positive cells. The number of DsRedExpress (REx)-positive cells was used to normalize for transfection efficiency. All experiments were repeated at least three times. ZFN activity is indicated as the fraction of EGFP-positive cells relative to the number of EGFP-positive cells measured in the presence of I-*Sce*I.

Revisions to the Zinc Finger Consortium website

We have previously described ZiFiT, a web-based software program that enables users to rapidly identify potential zinc finger target sites within genes of interest.[6] ZiFiT identifies target sites for which zinc finger arrays might be made using modular assembly and the Zinc Finger Consortium Modular Assembly Kit 1.0.[1] Given the results of this current study, we have revised the ZiFiT program by adding the following section to its instructions:

Scoring

To assess the effectiveness of modular assembly, we generated a large number of zinc finger arrays and tested their activity using a bacterial cell-based two-hybrid (B2H) assay (Ramirez et al., 2008). In the B2H assay, productive interaction of a zinc finger array with its cognate binding site activates transcription of a linked reporter gene. A total of 168 three finger arrays were assembled that recognize 104 different target DNA sites that varied in their GXX, AXX, CXX, and TXX subsite composition. Transcriptional activation of >1.57-fold over negative controls was determined to be indicative of target site recognition by a given protein. After measuring the activity of the 168 zinc finger arrays, it was found that the subsite composition of binding sites critically affected success rates. Modular assembly was far less effective for target sites composed of one or no GXX subsites (12% and 0% success rates, respectively) compared with those that contain three or two (59% and 29% success rates, respectively). Based on these results, we provide users of ZiFiT with an approximation of the likelihood that a given three finger protein will recognize its target. For example, three finger proteins comprised of three GXX subsites will receive a score of 0.59, reflecting the success rate observed in our survey. Most users will be interested in the likelihood that two arrays will function together as a ZFN. This can be approximated by multiplying the success rate of individual arrays. For example, the likelihood of success that a ZFN with all GXX subsites will recognize its target is 0.59 X 0.59 = 0.35 or 35%. Because multiple modules are often

available for given subsites, investigators may increase their likelihood of success by making multiple different proteins against a single target. One note of caution: the scoring function is a prediction for the activity of a given three finger array in the B2H assay; the protein may behave differently when tested for activity in a eukaryotic cell.

We have also revised the output that ZiFiT currently provides to users to include guidance about the likelihood of success for different target DNA binding sites.

References for Supplementary Methods:

1. Wright, D. A. et al. Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc* **1**, 1637-52 (2006).
2. Mandell, J. G. & Barbas, C. F., 3rd. Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* **34**, W516-23 (2006).
3. Bae, K. H. et al. Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol* **21**, 275-80 (2003).
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**, 3850-6 (2002).
5. Thibodeau, S. A., Fang, R. & Joung, J. K. High-throughput beta-galactosidase assay for bacterial cell-based reporter systems. *Biotechniques* **36**, 410-5 (2004).
6. Sander, J. D., Zaback, P., Joung, J. K., Voytas, D. F. & Dobbs, D. Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Res* **35**, W599-605 (2007).