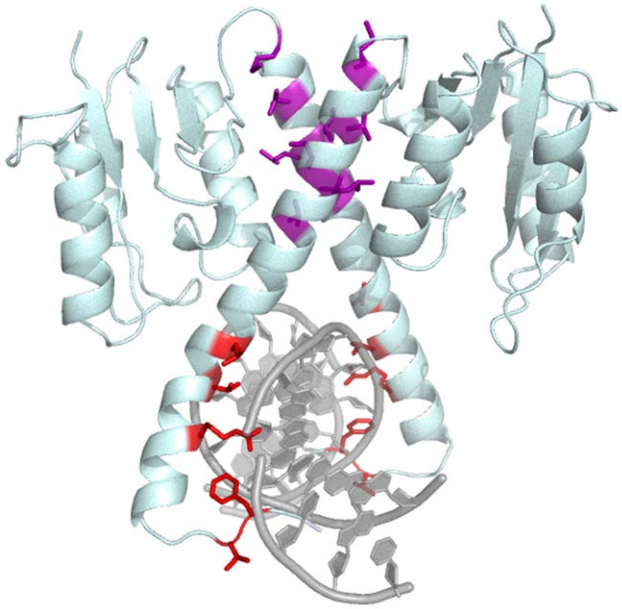
Supporting Information:

A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells

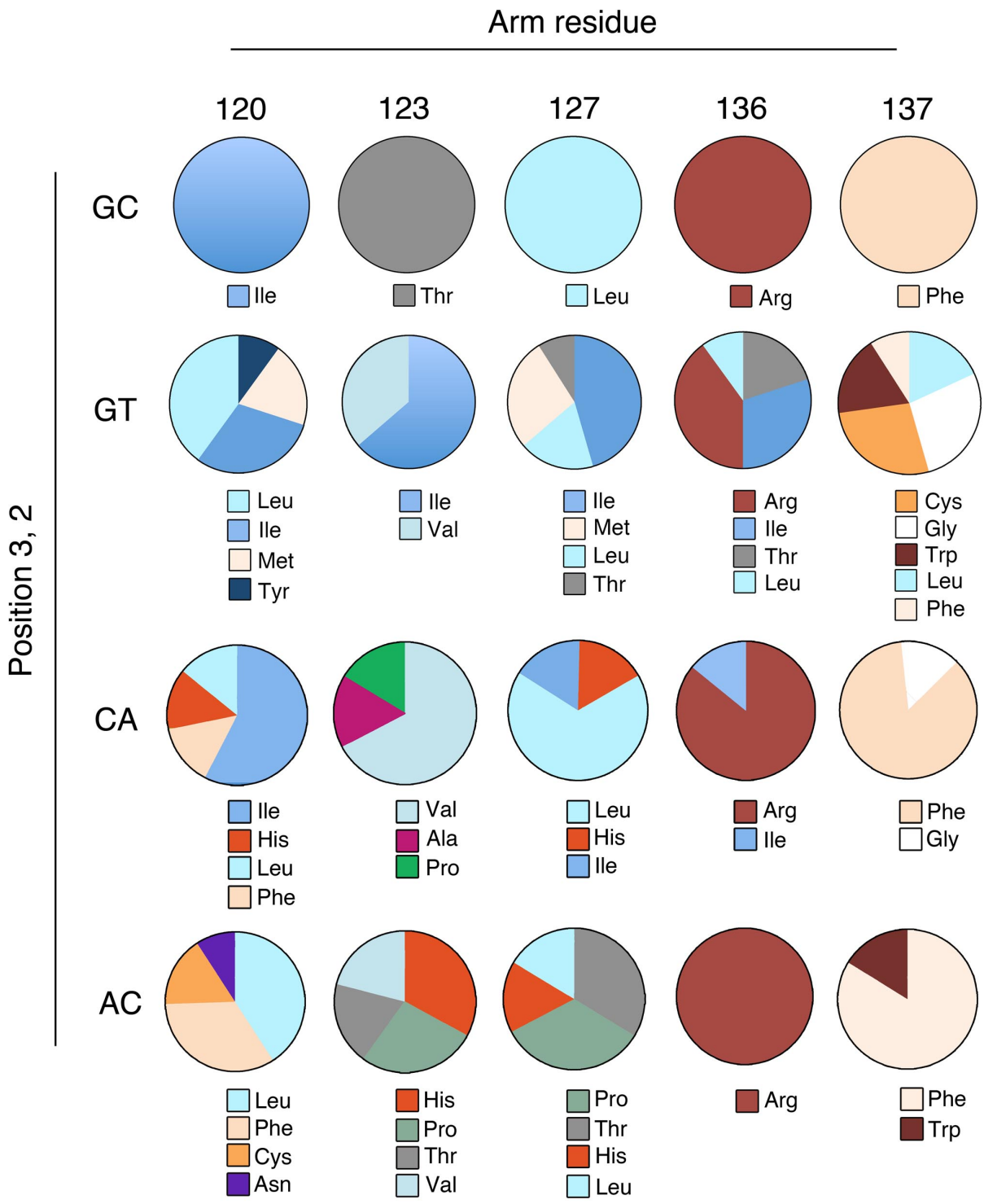
Thomas Gaj, Andrew C. Mercer, Shannon J. Sirk, Heather L. Smith, and Carlos F. Barbas III*

The Skaggs Institute for Chemical Biology and the Departments of Molecular Biology and Chemistry, The Scripps Research Institute, La Jolla, CA 92037 USA

*Corresponding author: carlos@scripps.edu



Supplementary Figure S1. Recombinase DNA-binding residues are located outside the dimer interface. The γδ resolvase in complex with target DNA. Catalytic domain dimer is colored cyan. DNA is colored grey. Arm region residues are shown as red sticks. Residues at the dimer interface are shown as purple sticks (PDB ID: 1GDT). Because the residues essential for specific DNA-binding are located outside the dimer interface, recombinase heterodimers with unique, asymmetric catalytic specificity can be generated.



Supplementary Figure S2. Sequence analysis of selected recombinases. Pie charts showing the percentage of amino acid substitutions at each targeted arm position. After the 4th round of selection, >20 clones were sequenced from each library. Sequence analysis of clones that recombine TT are described elsewhere (1).

Positions 3 and 2

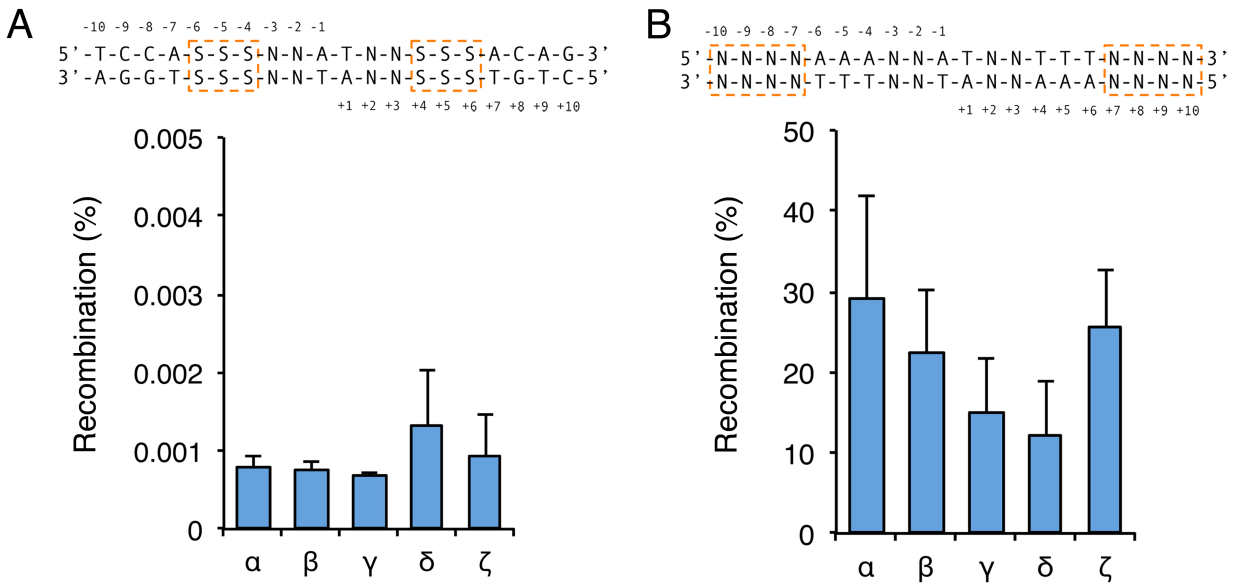
	Clone	Mutations	CC	GG	CG	GC	GT	TT
	Parental		36	25	55	0.01	0.003	0.08
*	GC-1	L127I, I136R, G137F	44	78	57	20	0.09	0.06

	Clone	Mutations	CC	TT	GA	CA	GT	CT
	Parental		36	0.002	0.008	0.004	0.003	0.086
*	GT-1	I120L, T123V, L127I, I136R, G137W	0.009	0.053	0.08	3.1	27	13.4
	GT-2	L127M, I136R, G137C	0.016	0.0003	0.092	0.207	39	6
	GT-3	L127I	0.0209	0.0009	0.41	0.77	39	13.5
	GT-4	D12A*, I120Y, T123I, L127I, I136T, G137L	0.0318	0.034	-	-	15	-
	GT-5	I120M, T123V	0.0068	0.013	-	-	6	-
	GT-6	I120L, T123I, L127I, G137F	0.0125	0.022	-	-	3	-

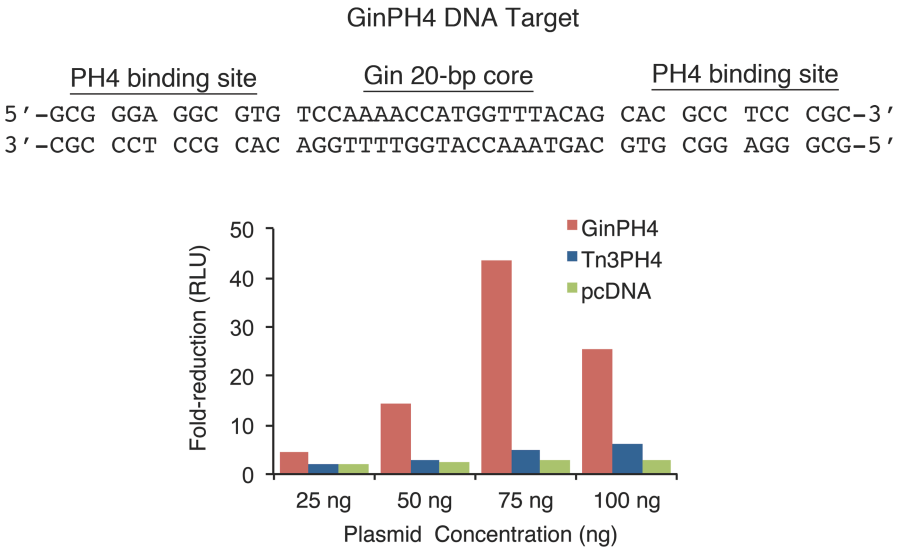
	Clone	Mutations	CC	TT	GA	CA	GT	CT
	Parental		36	0.002	0.01	0.004	0.003	0.09
*	CA-1	T123V, I136R, G137F	0.018	0.003	1.2	13	20	14
	CA-2	T123V	0.02	0.06	-	11	27	-
	CA-3	T123V, N132D	0.02	0.6	-	8.2	11.2	-
	CA-4	L52I*, L118S*, I120F, L127H, I136R, G137F	-	-	-	0.87	-	-
	CA-5	I120L, T123V, I136R, G137F	-	-	-	0.43	-	-
	CA-6	E117Q, L118C*, I120H, T123P, L127I	-	-	-	0.5	-	-

	Clone	Mutations	TT	AC	CA	GT	AG	GC
	Parental		0.002	0.005	0.004	0.003	0.008	0.01
*	AC-1	E117L, L118S, I120L, T123P, L127H, I136R, G137F	0.09	1.7	0.05	6	1	0.008
	AC-2	E117L, I120N, I136R, G137F	0.0002	0.7	-	0.12	-	0.03
	AC-3	R116Q*, E117L, L118S, I120F, T123H, L127T, I136P, G127F	0.0007	0.1	0.0005	0.0014	0.2	-
	AC-4	E117Y, I120F, T123P, L127P, I136R, G137F	0.0005	0.01	0.056	0.001	0.016	-

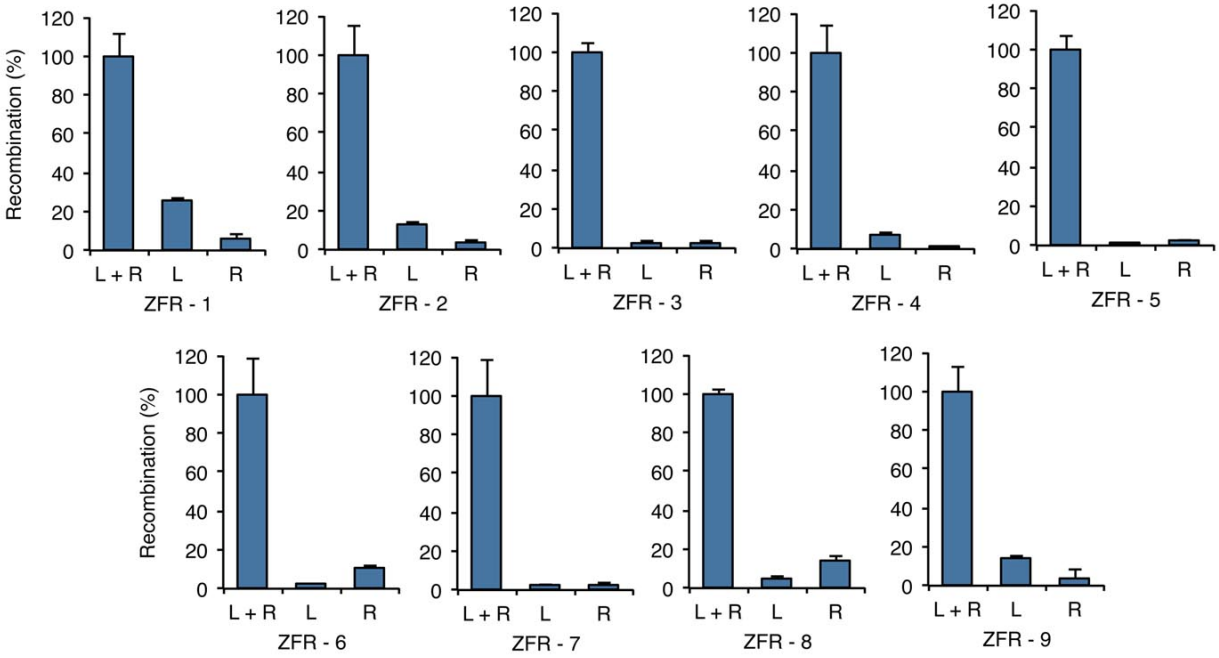
Supplementary Figure S3. Core specificity of isolated catalytic domains. After 4 rounds of selection, the ability of selected catalytic domains to recombine core sequences with substitutions at positions 3 and 2 was evaluated. Assigned DNA targets are underlined. Recombinase mutations are shown. Asterisks indicate catalytic domains selected for further analysis. Wild-type base combination at positions 3 and 2 is CC. Recombination was determined by split gene reassembly (2) and performed in triplicate. Catalytic domains that recombine TT substitutions are described elsewhere (1).



Supplementary Figure S4. Position specificity of selected catalytic domains. Recombination assays between the α, β, γ, δ and ζ catalytic domains and symmetrically substituted target sites. Recombination was measured on a library DNA targets that contained **(a)** >4,000 random strong base (S: G or C) substitutions at positions 6, 5 and 4 and **(b)** $>10^6$ (of a possible 4.29 x 10^9) unique base combinations at positions 10, 9, 8 and 7 (N: A, T, C or G). Recombination was measured by split gene reassembly (2) ($n = 3$).



Supplementary Figure S5. Reduced luciferase expression is dependent on the presence of ZFRs with complementary zinc-finger DNA-binding and recombinase catalytic domains. HEK293T cells were co-transfected with sub-optimal amounts of pGL3-PH4.20G reporter plasmid (25 ng) and increasing amounts of pcDNA-GinPH4 (fully complementary), pcDNA-Tn3PH4 (not fully complementary; mismatched catalytic domain) or empty pcDNA (negative). *Renilla* luciferase expression was used to normalize for transfection efficiency and cell numbers.



Supplementary Figure S6. ZFR homodimer activity. HEK293T cells were co-transfected with 150 ng ZFR-L or 150 ng ZFR-R with 2.5 ng of corresponding pGL3 ZFR reporter plasmid. Recombination was normalized to co-transfection with 150 ng ZFR-L and 150 ng ZFR-R with 2.5 ng pGL3 ZFR reporter plasmid.

Supplementary Figure S7. Clonal analysis of ZFR-modified cells. PCR primer combinations amplified either unmodified genomic target or integrated plasmid in the forward or reverse orientation.

Supplementary Table S1. Primers used in this study.

>GFP-ZFR-20G-XbaI-Fwd
TTAATTAAGAGTCTAGAGGAGGCGTGTCCAAAACCATGGTTTACAGCACGCCTCCAG
ATCTAGGAGGAATTTAAAATGAG

>GFP-ZFR-20G-HindIII-Rev
ACTGACCTAGAGAAGCTTGGAGGCGTGCTGTAAACCATGGTTTTGGACACGCCTCCC
TGCAGTTATTTGTACAGTTCATC

>SV40-ZFR-1-BglII-Fwd
TTAATTAAGAGAGATCTGCTGATGCAGATACAGAAACCAAGGTTTTCTTACTTGCTG
CTGCGCGATCTGCATCTCAATTAGTCAGC

>SV40-ZFR-1-HindIII-Rev
ACTGACCTAGAGAAGCTTGCAGCAGCAAGTAAGAAAACCTTGGTTTCTGTATCTGCA
TCAGCTTTGCAAAAGCCTAGGCCTCCAAA

>CMV-PstI-ZFR-1 Rev
CACCACCACGGATCCGCAGCAGCAAGTAAGAAAACCTTGGTTTCTGTATCTGCATCA
GCAATTTCGATAAGCCAGTAAGCAG

>5' Gin-HBS-Koz
CACCACCACGCGCGCAAGCTTAGATCTGGCCCAGGCGGCCACCATGCTGATTGGCT
ATGTAAGGG

>3' Gin-AgeI-Rev
CACCACCACACCGGTTCCCGATTTAGGTGGGCGAC

> ZFR-Target-1-Fwd
GTTCCTGCCAGGATCCACTAG

> ZFR-Target-1-Rev
GCATGTGTCCAGATGCATAGG

> ZFR-Target-2-Fwd
CACCTTCTCCCAGGATAAGG

> ZFR-Target-2-Rev
GTTGGCCTGTATTCCTCTGG

> ZFR-Target-3-Fwd
AATGAAGTTCCCTTGGCACTTC

> ZFR-Target-3-Rev
CTGAAGGGTTTTAAGTGCAGAAG

> CMV-Mid Prim-1
TGACGTCAATGACGGTAAATGG

ZFR targets are underlined.

Supplementary Table S2. The amino acid sequences of the catalytic domains described in this study.

>Gin α
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGLA
AARNKGRIGGRPPKSG

>Gin β
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGIA
AARNKGRRFGRPPKSG

>Gin γ
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELILERVMAGIA
AARNKGRRWGRPPKSG

>Gin δ
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERVMAGLA
AARNKGRRFGRPPKSG

>Gin ε
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERLSILERPMAGHA
AARNKGRRFGRPPKSG

>Gin ζ
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTSAGRA
AAINKGRIMGRPRKSG

Targeted arm region positions are highlighted red. Random substitutions are highlighted blue. The hyperactivating H106Y mutation is highlighted green.

Supplementary Table S3. The amino acid sequences of the ZFRs described in this study.

>ZFR-1 Left
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGLA
AARNKGRIGGRPPKSGTGEKPYKCPECGKSFSTSGNLVRHQRTHTGEKPYKCPECGKSFSQSGD
LRRHQRTHTGEKPYKCPECGKSFSTSGNLVRHQRTHTGEKPYKCPECGKSFSTSGELVRHQRTH
TGKKTSGQAGQ

>ZFR-1 Right
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGLA
AARNKGRIGGRPPKSGTGEKPYKCPECGKSFSHRTTLTNHQRTHTGEKPYKCPECGKSFSQSGD
LRRHQRTHTGEKPYKCPECGKSFSQSGDLRRHQRTHTGEKPYKCPECGKSFSQSGDLRRHQRTH
TGKKTSGQAGQ

>ZFR-2 Left
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGLA
AARNKGRIGGRPPKSGTGEKPYKCPECGKSFSQSGDLRRHQRTHTGEKPYKCPECGKSFSQRAH
LERHQRTHTGEKPYKCPECGKSFSTSGNLVRHQRTHTGEKPYKCPECGKSFSRSDELVRHQRTH
TGKKTSGQAGQ

>ZFR-2 Right
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGLA
AARNKGRIGGRPPKSGTGEKPYKCPECGKSFSRSDKLVRHQRTHTGEKPYKCPECGKSFSRKDN
LKNHQRTHTGEKPYKCPECGKSFSTSGELVRHQRTHTGEKPYKCPECGKSFSRSDKLVRHQRTH
TGKKTSGQAGQ

>ZFR-3 Left
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGLA
AARNKGRIGGRPPKSGTGEKPYKCPECGKSFSTTGNLTVHQRTHTGEKPYKCPECGKSFSDPGA
LVRHQRTHTGEKPYKCPECGKSFSQSSNLVRHQRTHTGEKPYKCPECGKSFSRSDHLTNHQRTH
TGKKTSGQAGQ

>ZFR-3 Right
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGLA
AARNKGRIGGRPPKSGTGEKPYKCPECGKSFSRKDNLKNHQRTHTGEKPYKCPECGKSFSRSDH
LTNHQRTHTGEKPYKCPECGKSFSDPGNLVRHQRTHTGEKPYKCPECGKSFSRKDNLKNHQRTH
TGKKTSGQAGQ

>ZFR-4 Left
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELILERVMAGIA
AARNKGRRWGRPPKSGTGEKPYKCPECGKSFSQRANLRAHQRTHTGEKPYKCPECGKSFSQSSS

LVRHQRTHTGEKPYKCPECGKSFSTTGNLTVHQRTHTGEKPYKCPECGKSFSQRAHLERHQRTH
TGKKTSGQAGQ

>ZFR-4 Right
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELILERVMAGIA
AARNKGRRWGRPPKSGTGEKPYKCPECGKSFSQRANLRAHQRTHTGEKPYKCPECGKSFSRRDE
LNVHQRTHTGEKPYKCPECGKSFSQLAHLRAHQRTHTGEKPYKCPECGKSFSQRAHLERHQRTH
TGKKTSGQAGQ

>ZFR-5 Left
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELILERVMAGIA
AARNKGRRWGRPPKSGTGEKPYKCPECGKSFSRRDELNVHQRTHTGEKPYKCPECGKSFSRSDH
LTNHQRTHTGEKPYKCPECGKSFSQLAHLRAHQRTHTGEKPYKCPECGKSFSQRAHLERHQRTH
TGKKTSGQAGQ

>ZFR-5 Right
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGLA
AARNKGRIGGRPPKSGTGEKPYKCPECGKSFSTSGSLVRHQRTHTGEKPYKCPECGKSFSRSDK
LVRHQRTHTGEKPYKCPECGKSFSQSGDLRRHQRTHTGEKPYKCPECGKSFSTSGELVRHQRTH
TGKKTSGQAGQ

>ZFR-6 Left
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTSAGRA
AAINKGRIMGRPRKSGTGEKPYKCPECGKSFSQLAHLRAHQRTHTGEKPYKCPECGKSFSQLAH
LRAHQRTHTGEKPYKCPECGKSFSDPGHLVRHQRTHTGEKPYKCPECGKSFSDSGNLRVHQRTH
TGKKTSGQAGQ

>ZFR-6 Right
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELILERVMAGIA
AARNKGRRWGRPPKSGTGEKPYKCPECGKSFSQRAHLERHQRTHTGEKPYKCPECGKSFSTTGN
LTVHQRTHTGEKPYKCPECGKSFSDSGNLRVHQRTHTGEKPYKCPECGKSFSQSSNLVRHQRTH
TGKKTSGQAGQ

>ZFR-7 Left
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGIA
AARNKGRRFGRPPKSGTGEKPYKCPECGKSFSTHLDLIRHQRTHTGEKPYKCPECGKSFSTTGN
LTVHQRTHTGEKPYKCPECGKSFSQSSSLVRHQRTHTGEKPYKCPECGKSFSRSDNLVRHQRTH
TGKKTSGQAGQ

>ZFR-7 Right
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGIA

AARNKGRRFGRPPKSGTGEKPYKCPECGKSFSRSDKLVRHQRTHTGEKPYKCPECGKSFSRRDE
LNVHQRTHTGEKPYKCPECGKSFSQSSSLVRHQRTHTGEKPYKCPECGKSFSRSDHLTNHQRTH
TGKKTSGQAGQ

>ZFR-8 Left
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGLA
AARNKGRIGGRPPKSGTGEKPYKCPECGKSFSQRAHLERHQRTHTGEKPYKCPECGKSFSTSGN
LVRHQRTHTGEKPYKCPECGKSFSRSDELVRHQRTHTGEKPYKCPECGKSFSHKNALQNHQRTH
TGKKTSGQAGQ

>ZFR-8 Right
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGLA
AARNKGRIGGRPPKSGTGEKPYKCPECGKSFSRRDELNVHQRTHTGEKPYKCPECGKSFSQSSN
LVRHQRTHTGEKPYKCPECGKSFSQSSSLVRHQRTHTGEKPYKCPECGKSFSTTGNLTVHQRTH
TGKKTSGQAGQ

>ZFR-9 Left
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELILERVMAGIA
AARNKGRRWGRPPKSGTGEKPYKCPECGKSFSTTGNLTVHQRTHTGEKPYKCPECGKSFSQSSN
LVRHQRTHTGEKPYKCPECGKSFSQRAHLERHQRTHTGEKPYKCPECGKSFSQKSSLIAHQRTH
TGKKTSGQAGQ

>ZFR-9 Right
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLD
RLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTSAGRA
AAINKGRIMGRPRKSGTGEKPYKCPECGKSFSDPGALVRHQRTHTGEKPYKCPECGKSFSQSSS
LVRHQRTHTGEKPYKCPECGKSFSQLAHLRAHQRTHTGEKPYKCPECGKSFSQRANLRAHQRTH
TGKKTSGQAGQ

Arm region mutations are highlighted red. Specificity-determining α-helical zinc-finger residues are highlighted green.

Supplementary Notes

The total number of 20-bp core sequences that can be targeted by each possible combination of Gin catalytic domain heterodimers was calculated as the product of the consensus sequence: 5'-NNNNAAABNWWNVTTTNNNN-3', where N: A, T, C or G, B: T, C or G, V: A, C or G and W: A or T. This value was determined to be 3.77×10^7.

To determine the frequency of a 44-bp ZFR target site occurring randomly in the human genome, we first divided the total number of possible 44-bp ZFR targets sites based on our archive of 45 pre-selected zinc-finger modules (3-5) (6.33×10^{20}) with the total number of unique base combinations possible across this region (3.094×10^{26}). This relative frequency (1.606×10^{-6}) was then multiplied by one haploid human genome (3.1×10^9 bp). The total number of ZFR target sites was thus determined to be 4,980.

We next estimated the overall frequency by dividing one haploid human genome (3.1×10^9 bp) with the total number of possible ZFR targets (4,980). By this calculation, one potential ZFR target site was predicted to occur approximately once every 620,000 bp of random sequence.

We note that by expanding the sequence constraints at the dinucleotide core to include all base combinations except those that are SS (S: C or G), the total number of 'targetable' 20-bp core sequences can be increased to ~1.1×10^8. Using this value, we estimate that one potential ZFR target site is predicted to occur in the human genome approximately once every 160,000 bp of random sequence.

Supplementary References

1. Gaj, T., Mercer, A.C., Gersbach, C.A., Gordley, R.M. and Barbas, C.F., 3rd. (2011) Structure-guided reprogramming of serine recombinase DNA sequence specificity. *Proc. Natl. Acad. Sci. USA*, **108**, 498-503.

2. Gersbach, C.A., Gaj, T., Gordley, R.M. and Barbas, C.F., 3rd. (2010) Directed evolution of recombinase specificity by split gene reassembly. *Nucleic Acids Res.*, **38**, 4198-4206.

3. Segal, D.J., Dreier, B., Beerli, R.R. and Barbas, C.F., 3rd. (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc. Natl. Acad. Sci. USA*, **96**, 2758-2763.

4. Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D. and Barbas, C.F., 3rd. (2001) Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.*, **276**, 29466-29478.

5. Dreier, B., Fuller, R.P., Segal, D.J., Lund, C.V., Blancafort, P., Huber, A., Koksch, B. and Barbas, C.F., 3rd. (2005) Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.*, **280**, 35588-35597.