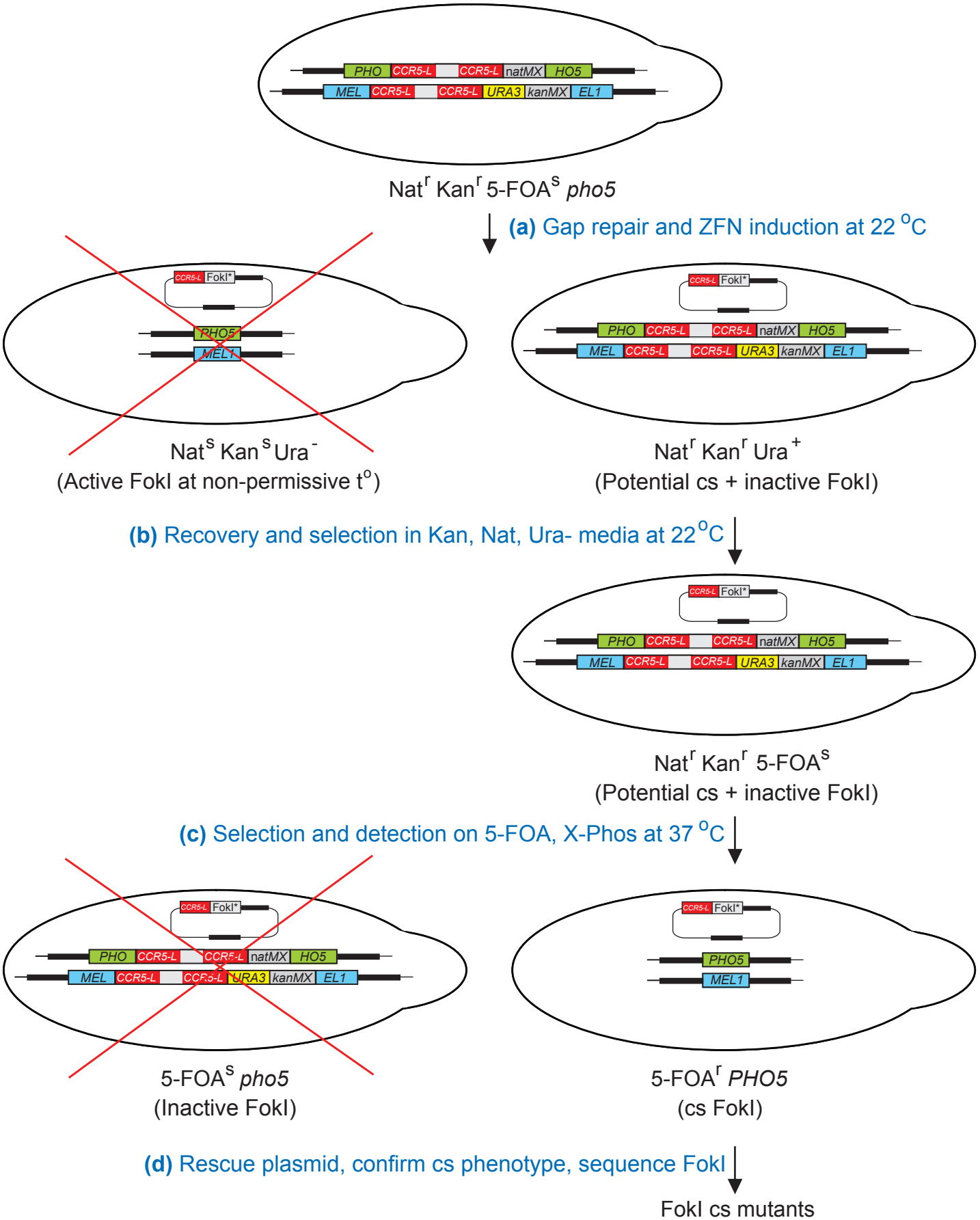
Nature Methods

Enhanced zinc-finger-nuclease activity with improved obligate heterodimeric architectures

Yannick Doyon, Thuy D Vo, Matthew C Mendel, Shon G Greenberg, Jianbin Wang, Danny F Xia, Jeffrey C Miller, Fyodor D Urnov, Philip D Gregory & Michael C Holmes

Supplementary Figure 1	Genetic screening scheme.
Supplementary Figure 2	Improved activity and specificity of the ELD:KKK and ELD:KKR ZFN variants.
Supplementary Figure 3	ZFN variants ELD:KKK and ELD:KKR are more active than the original obligate heterodimeric EL:KK.
Supplementary Figure 4	Preservation of the obligate heterodimer specificity.
Supplementary Figure 5	Characterization of ZFN variants activity at heterodimer off-target sites.
Supplementary Figure 6	Activity of obligate heterodimeric Fok I variants at the CCR5 locus.
Supplementary Figure 7	Anti-homodimerization is further improved by the introduced mutations.
Supplementary Figure 8	The N496D:H537R mutant pair enhance the activity of the DA:RV architecture.
Supplementary Figure 9	The N496D:H537R mutant pair enhance the activity of the D:R architecture.
Supplementary Figure 10	Orthogonal behavior of the ELD:KKR and DAD:RVR architectures.
Supplementary Figure 11	Elimination of transheterodimer off-target activity using combinations of the ELD:KKR and DAD:RVR architectures.
Supplementary Figure 12	Additive effect of the ELD:KKR and “Sharkey” FokI domains on ZFN activity.
Supplementary Figure 13	Additive effect of the DAD:RVR and “Sharkey” FokI domains on ZFN activity.
Supplementary Figure 14	Preservation of the obligate heterodimer specificity in the ELD:KKR and DAD:RVR architectures combined with the “Sharkey” FokI domains.
Supplementary Figure 15	Unprocessed scans of full-length gels and blots for the primary figures.
Supplementary Table 1	Isolated cold-sensitive mutants.
Supplementary Table 2	Primers used for the Cel-1 assays.
Supplementary Note 1	ZFN coding sequences.
Supplementary Note 2	RFLP donor vector sequence.

Supplementary Figure 1 | Genetic screening scheme.

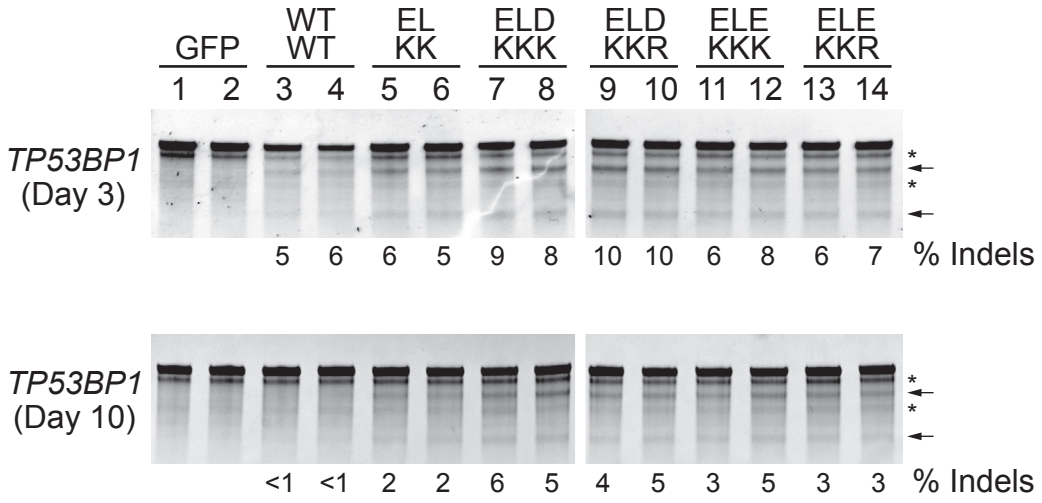


Nature Methods: doi:10.1038/nmeth.1539

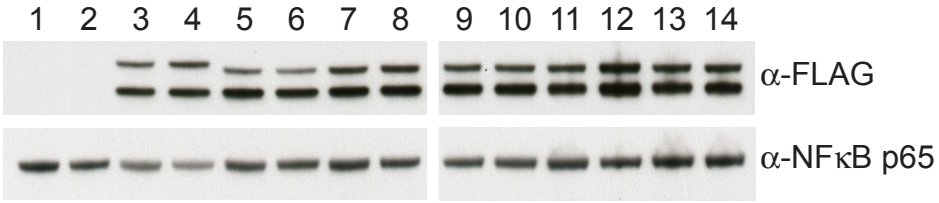
(**a**) Following gap repair, ZFN expression was induced at the non-permissive temperature of 22°C. (**b**) After recovery, the cells were incubated in KAN (G418), NAT and Ura^- media to eliminate all cells containing active ZFNs. This step selected for potential cold-sensitive as well as for inactive mutants. (**c**) Next, the cells were shifted to 37°C (permissive temperature) and plated on media containing 5-FOA and X-Phos (colorimetric Pho5 substrate). Only cells containing a cold-sensitive ZFN should form blue colonies. (**d**) The plasmids from these cells were then isolated and the reporter strain retransformed to confirm the cold-sensitive phenotype. The mutations were identified by direct sequencing of the FokI domain.

Supplementary Figure 2 | Improved activity and specificity of the ELD:KKK and ELD:KKR ZFN variants.

a



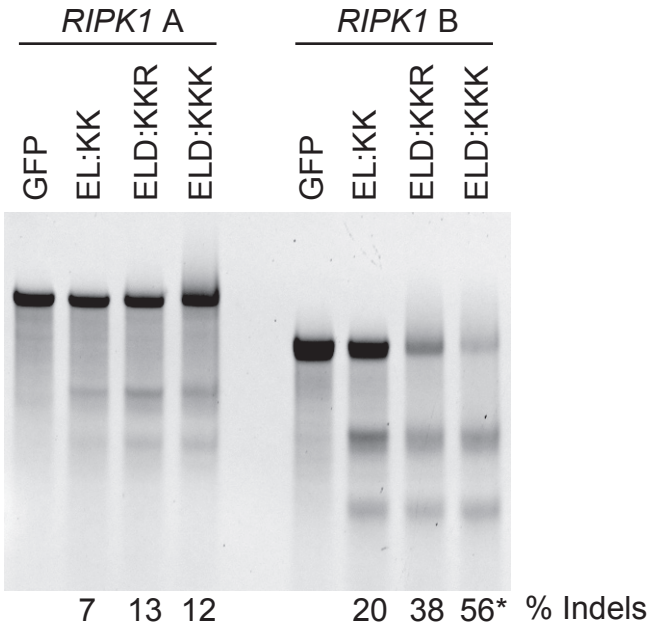
b



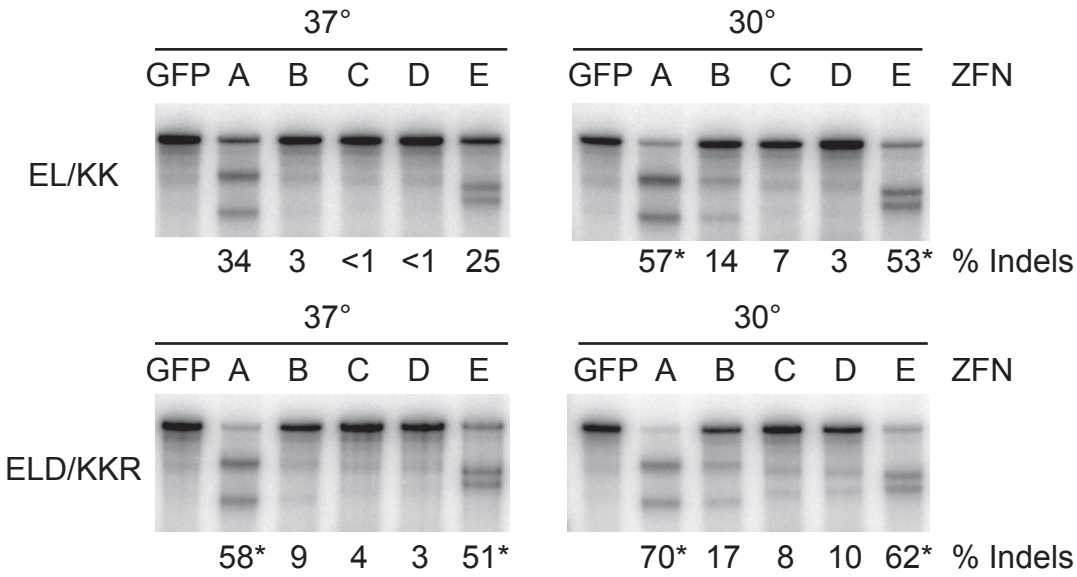
(**a**) Expression plasmids encoding the indicated ZFN variants targeting *TP53BP1* were transfected (400 ng) in duplicate into K562 and the cells were harvested 3 days post-transfection. Non-radioactive Cel-1 assay was used to determine the frequency of ZFN-induced insertions and deletions (indels). An aliquot of the cells was also cultivated for an additional week to determine the stability of the modified cells in extended cultures. (**b**) ZFN expression was monitored by western blot using the FLAG antibody and α-NFκB p65 was used as a loading control. * Indicates non-specific cutting by the Cel-1 nuclease. Arrows indicate the cleavage products of the predicted size.

Supplementary Figure 3 | ZFN variants ELD:KKK and ELD:KKR are more active than the original obligate heterodimeric EL:KK.

a RIPK1

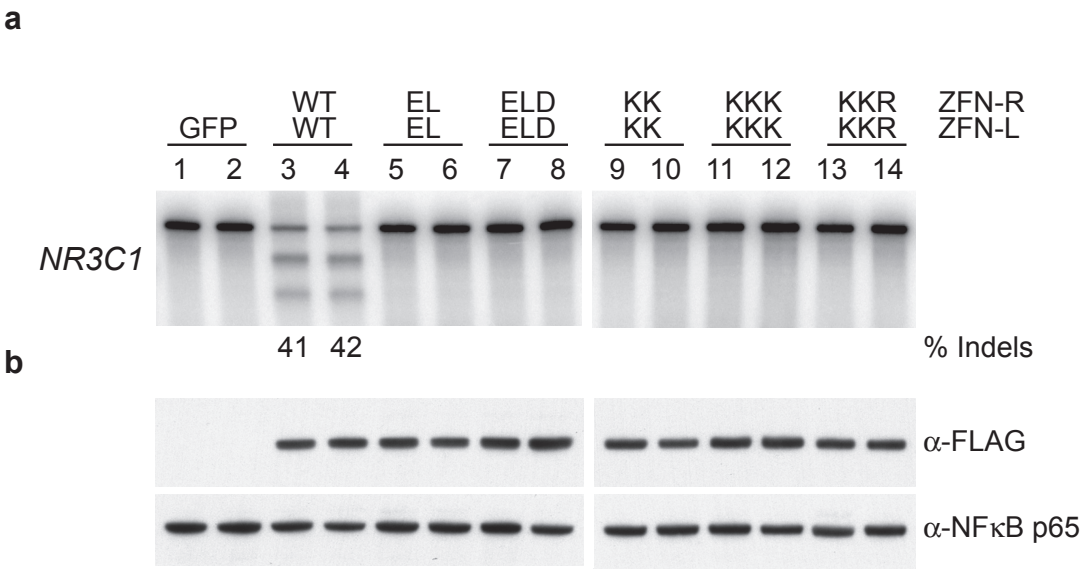


b *RSK4*



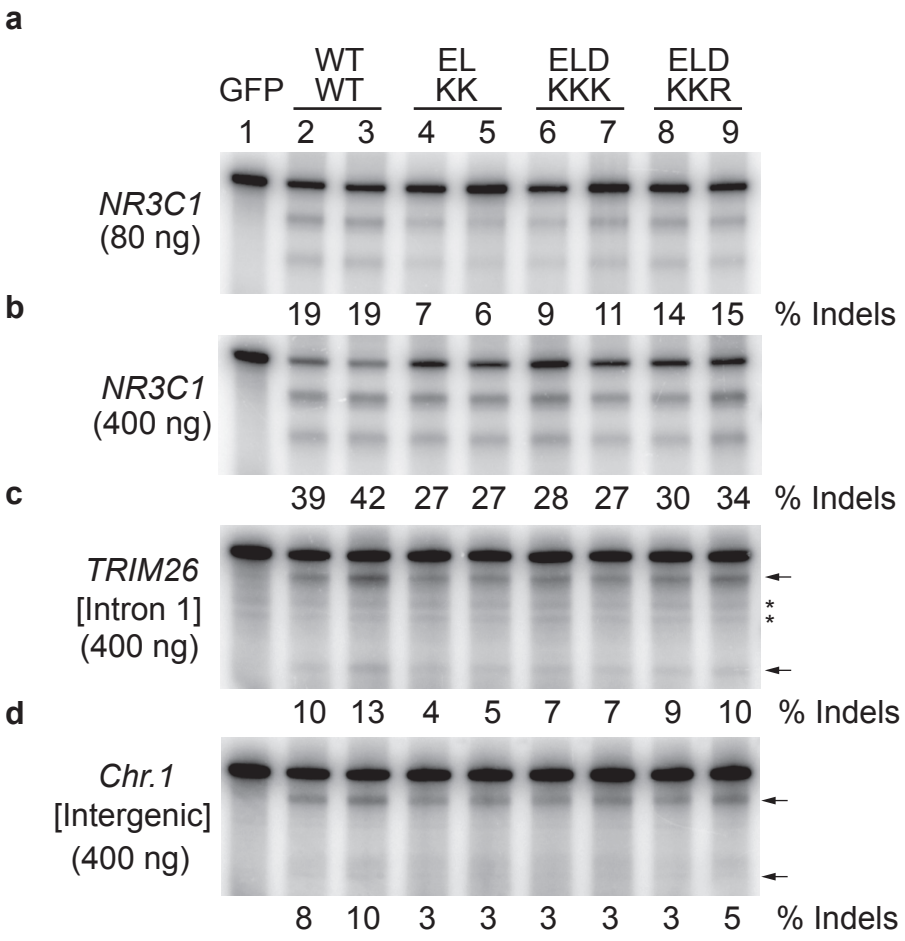
(**a**) Expression plasmids encoding the indicated ZFN variants targeting two different sites in the *RIPK1* gene were transfected (400 ng) into K562 and the cells were harvested 3 days post-transfection. Non-radioactive Cel-1 assay was used to determine the frequency of ZFN-induced insertions and deletions (indels). (**b**) Expression plasmids encoding ZFN pairs (A-B-C-D-E) targeting five different sites in the *RSK4* gene were transfected (400 ng) into K562 and the cells were incubated for 3 days at 37°C (left panel) or 30°C (right panel) for 3 days before harvest. Cel-1 assay was used to determine the frequency of ZFN-induced insertions and deletions (indels). * Indicates an underestimated value due to the saturation of the assay.

Supplementary Figure 4 | Preservation of the obligate heterodimer specificity.



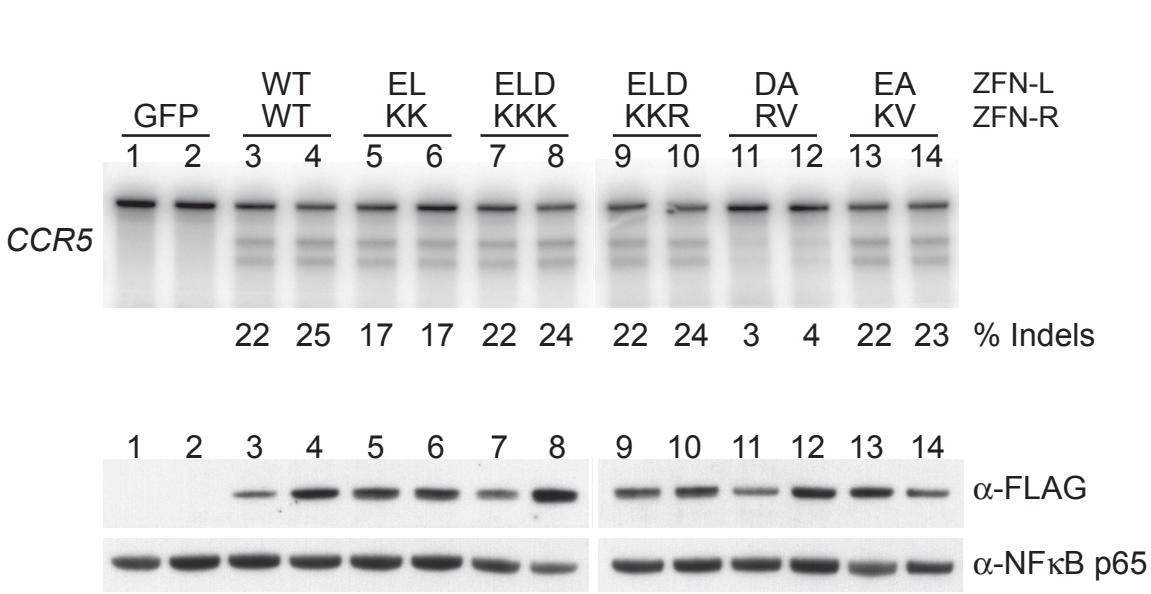
(**a**) Forced homodimerization of the indicated FokI domains using the *NR3C1* targeting ZFNs was assayed by transfection of the expression plasmids encoding the indicated ZFN variants (400 ng) in duplicate into K562 cells. The Cel-1 assay was used to determine the frequency of ZFN-induced indels at the *NR3C1* heterodimer site. (**b**) ZFN expression was monitored by western blot using the FLAG antibody and α-NFκB p65 was used as a loading control.

Supplementary Figure 5 | Characterization of ZFN variants activity at heterodimer off-target sites.



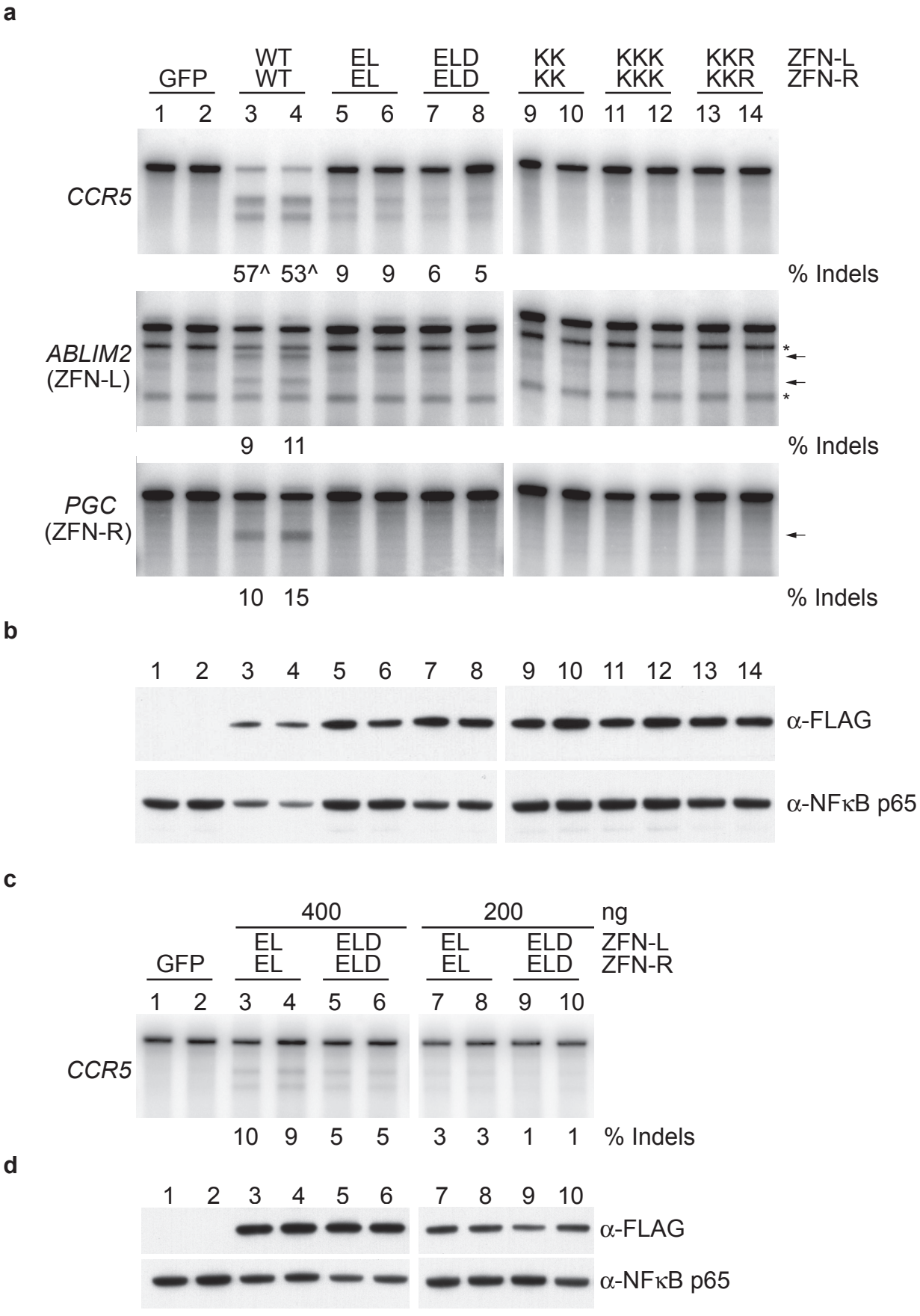
(**a**) and (**b**) are the same gels shown in Figure 3c. (**c**) and (**d**) Cleavage at *TRIM26* and *Chr.1* in the cells transfected with 400 ng of the indicated ZFN expression plasmids was assayed by the Cel-1 assay to determine the frequency of ZFN-induced indels at these heterodimer off-target sites for the *NR3C1*-targeting ZFNs. * Indicates non-specific cutting by the Cel-1 nuclease. Arrows indicate the cleavage products of the predicted size.

Supplementary Figure 6 | Activity of obligate heterodimeric FokI variants at the *CCR5* locus.



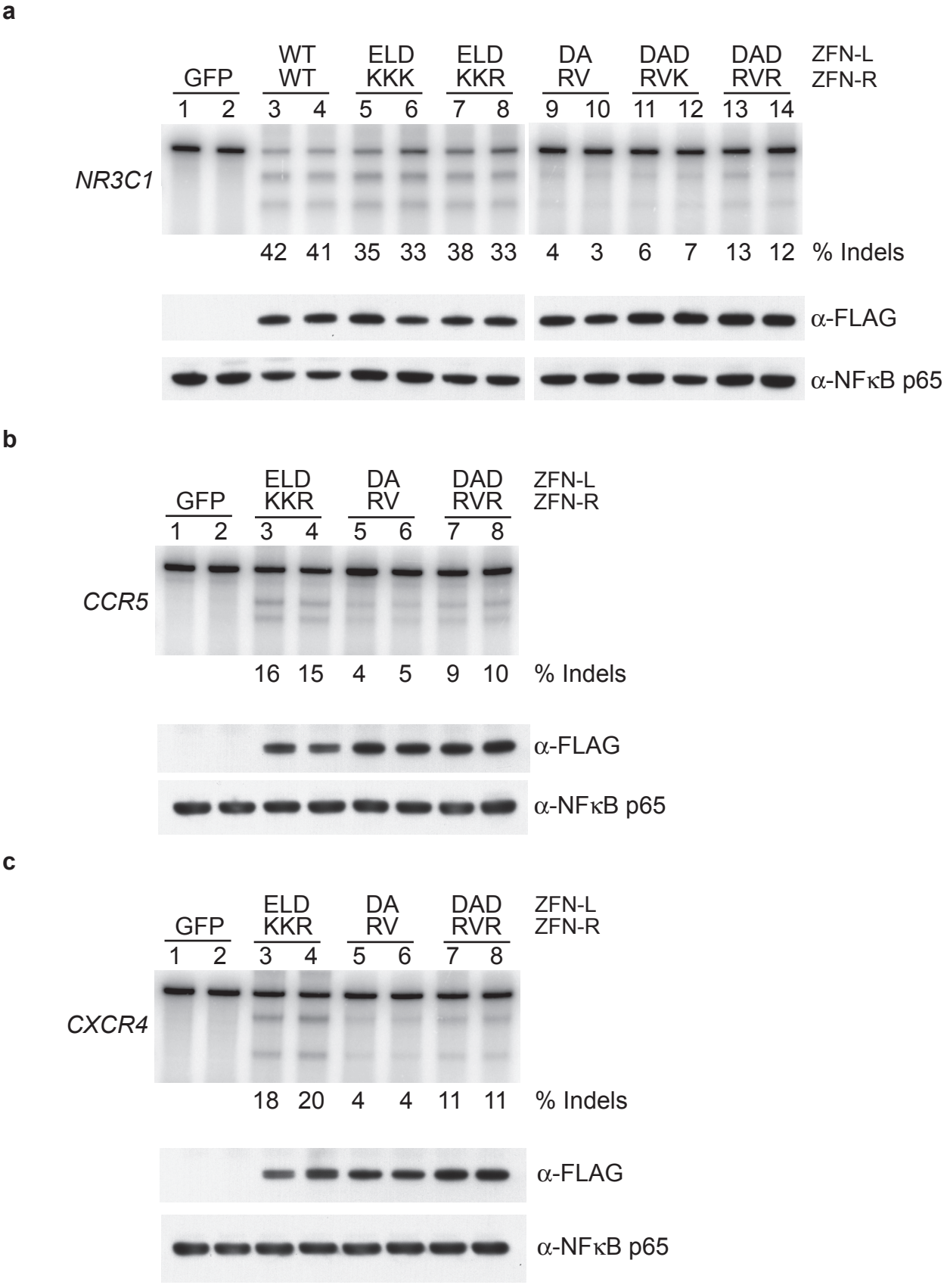
(**a**) Expression plasmids encoding the indicated ZFN variants targeting *CCR5* were transfected (80 ng) into K562 and the cells were harvested 3 days post-transfection. The Cel-1 assay was used to determine the frequency of ZFN-induced insertions and deletions (indels). (**b**) ZFN expression was monitored by western blot using the FLAG antibody and α-NFκB p65 was used as a loading control.

Supplementary Figure 7 | Anti-homodimerization is further improved by the introduced mutations.



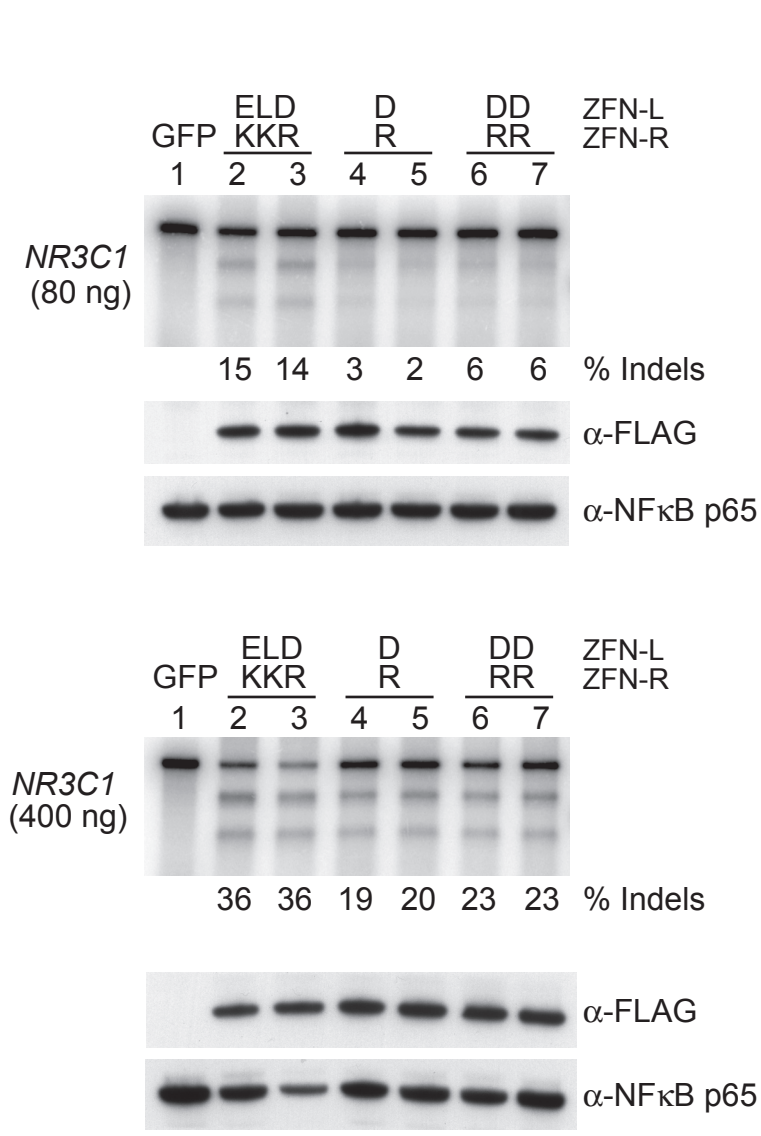
(**a**) Forced homodimerization of *CCR5* targeting ZFNs was assayed by transfection of expression plasmids encoding the indicated ZFN variants (400 ng) in duplicate into K562 cells and the Cel-1 assay was used to determine the frequency of ZFN-induced indels at the *CCR5* heterodimer target, a *CCR5*-L ZFN homodimer (*ABLIM2*), and a *CCR5*-R homodimer (*PGC*) off-target sites. * Indicates the presence of a SNP that results in non-specific cleavage by the Cel-1 nuclease. Arrows indicate the cleavage products of the predicted size. ^ Indicates an underestimated value due to the saturation of the assay **(b)** ZFN expression was monitored by western blot using the FLAG antibody and α-NFκB p65 was used as a loading control. **(c)** High (400 ng) and low (200 ng) amounts of expression plasmids encoding the *CCR5*-targeting ZFN variants were transfected in duplicate into K562s, the cells were harvested 3 days post-transfection, and the Cel-1 assay was used to determine the frequency of ZFN-induced indels at the *CCR5* heterodimer site. **(d)** ZFN expression was monitored as in **(b)**.

Supplementary Figure 8 | The N496D:H537R mutant pair enhances the activity of the DA:RV architecture.



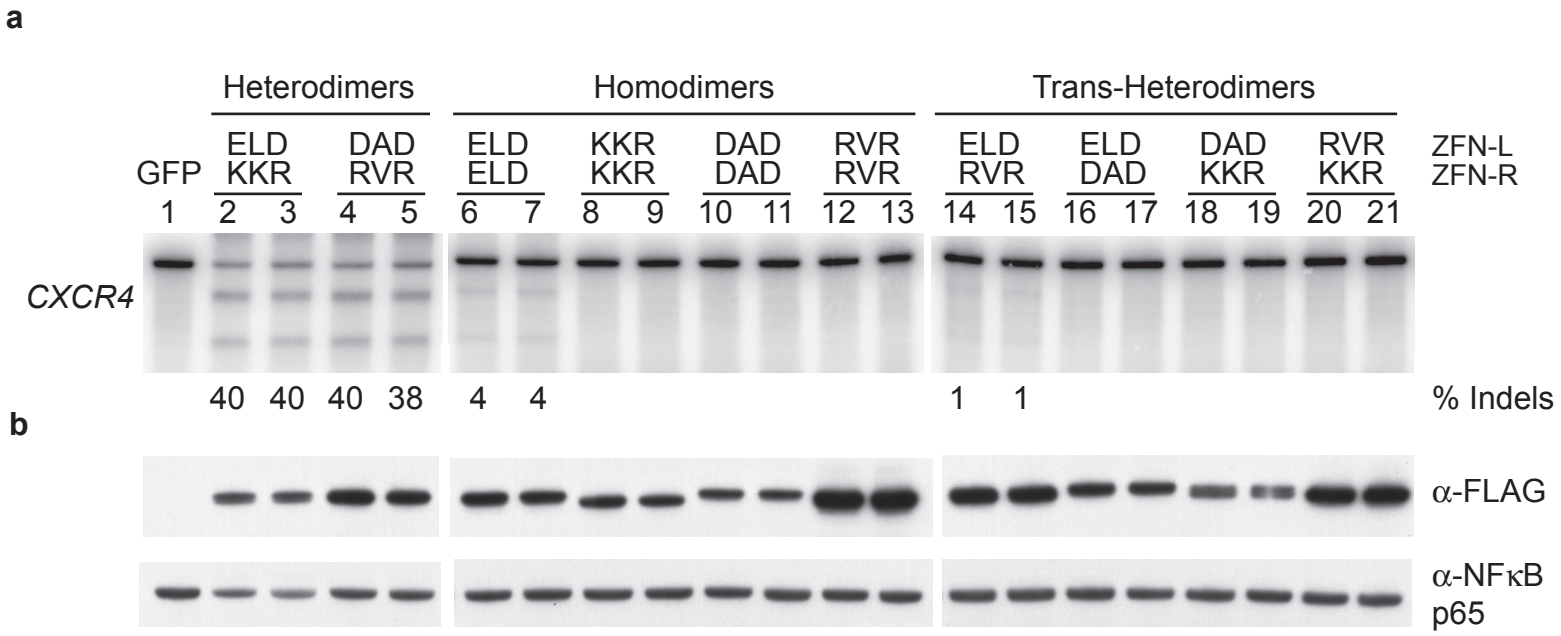
(**a**) Expression plasmids encoding the indicated *NR3C1* ZFN variants (400 ng) were transfected in duplicate into K562 and the cells were harvested 3 days post-transfection. The Cel-1 assay was used to determine the frequency of ZFN-induced insertions and deletions. ZFN expression level was monitored by western blot using the FLAG antibody and α-NFκB p65 was used as a loading control. (**b**) and (**c**) Expression plasmids encoding the indicated *CCR5* and *CXCR4* ZFN variants (80 ng) were transfected in duplicate into K562 and processed as in (**a**).

Supplementary Figure 9 | The N496D:H537R mutant pair enhances the activity of the D:R architecture.



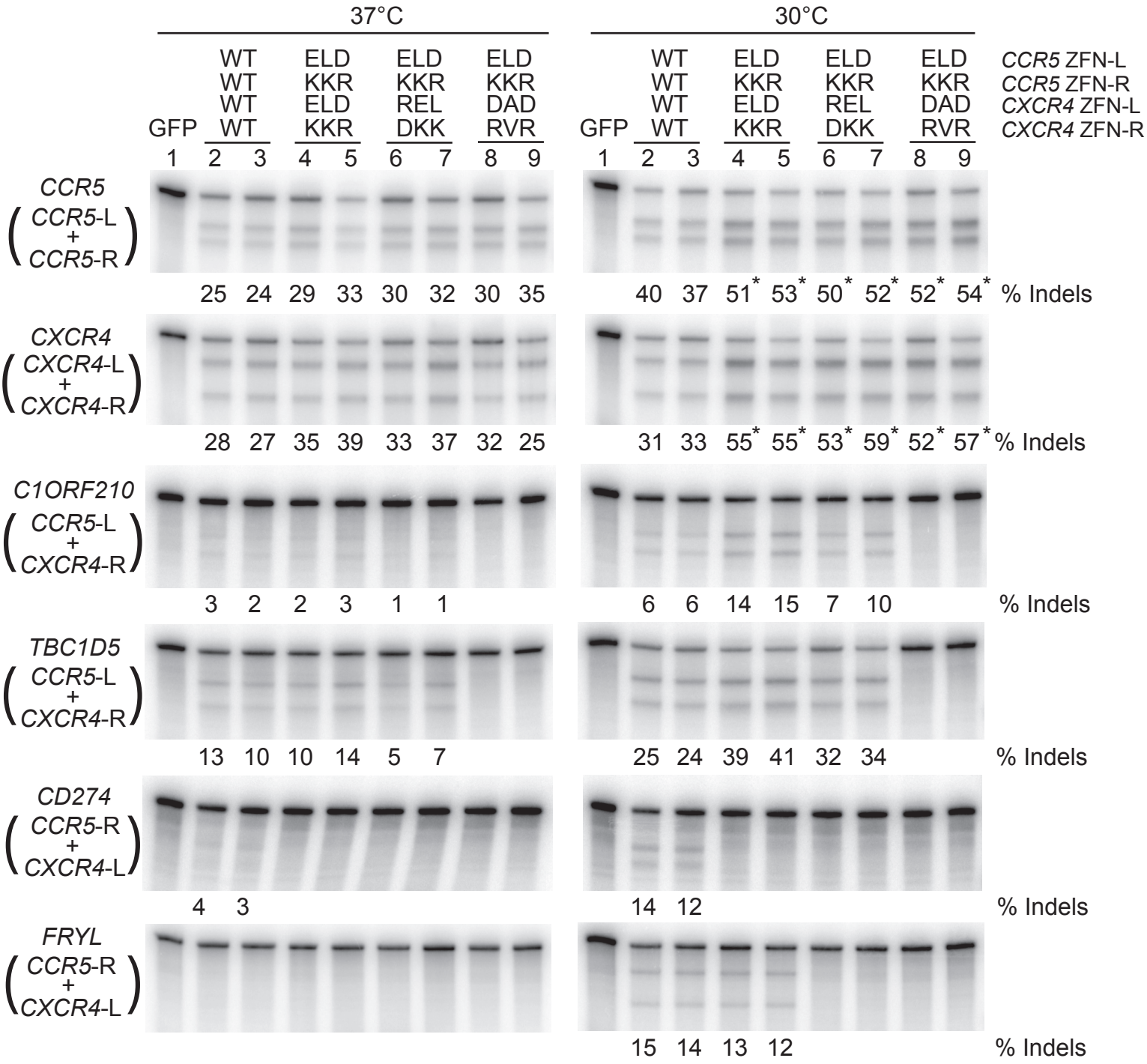
(**a**) and (**b**) Expression plasmids encoding the indicated *NR3C1* ZFN variants (80 and 400 ng) were transfected in duplicate into K562 and the cells were harvested 3 days post-transfection. The Cel-1 assay was used to determine the frequency of ZFN-induced insertions and deletions. ZFN expression level was monitored by western blot using the FLAG antibody and α-NFκB p65 was used as a loading control.

Supplementary Figure 10 | Orthogonal behavior of the ELD:KKR and DAD:RVR architectures.



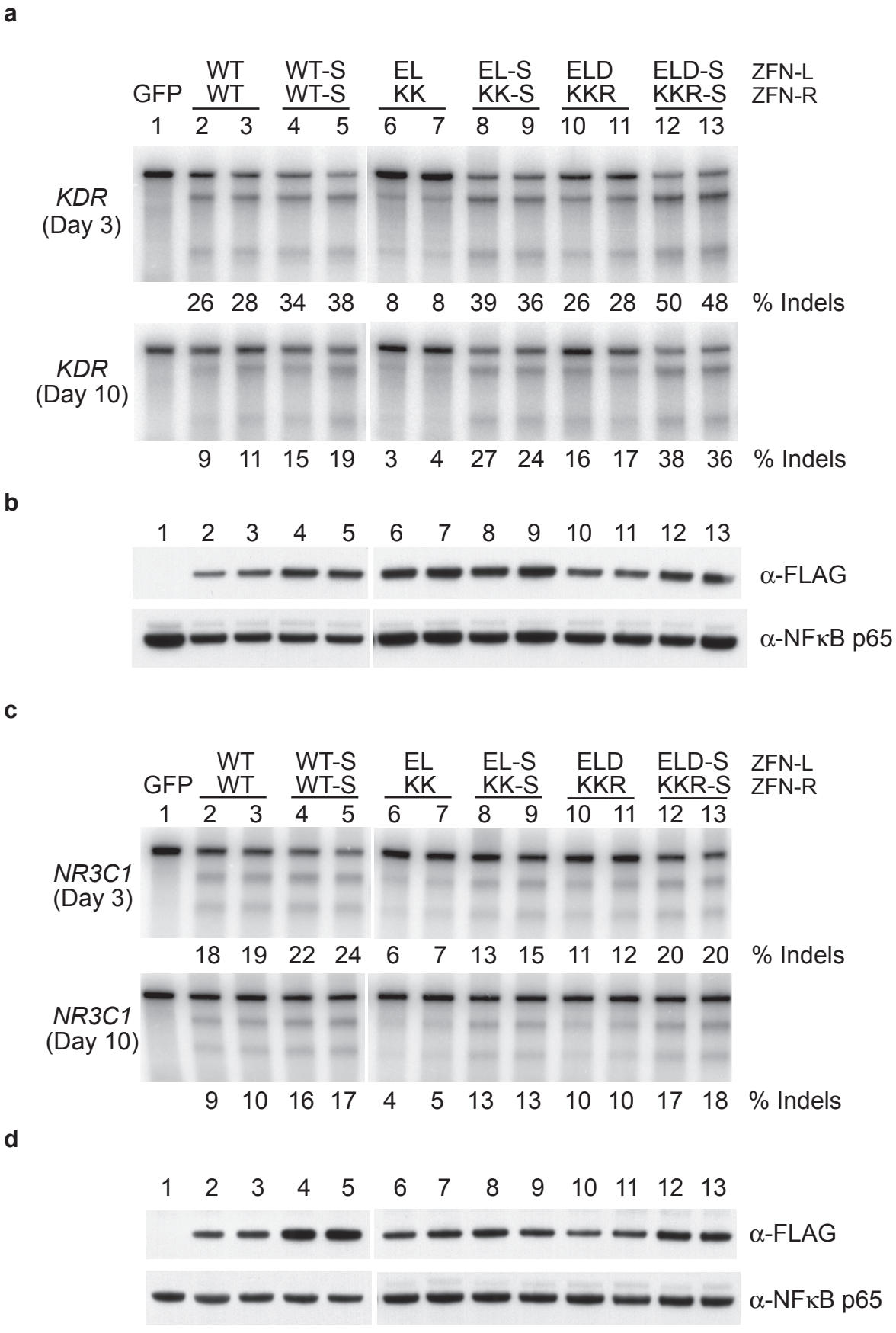
(**a**) Forced homodimerization and transheterodimerization of *CXCR4* targeting ZFNs was assayed by transfection of expression plasmids encoding the indicated FokI variants (400 ng) in duplicate into K562 cells and the Cel-1 assay was used to determine the frequency of ZFN-induced indels 3 days post transfection. (**b**) ZFN expression was monitored by western blot using the FLAG antibody and α-NFκB p65 was used as a loading control.

Supplementary Figure 11 | Elimination of transheterodimer off-target activity using combinations of the ELD:KKR and DAD:RVR architectures.



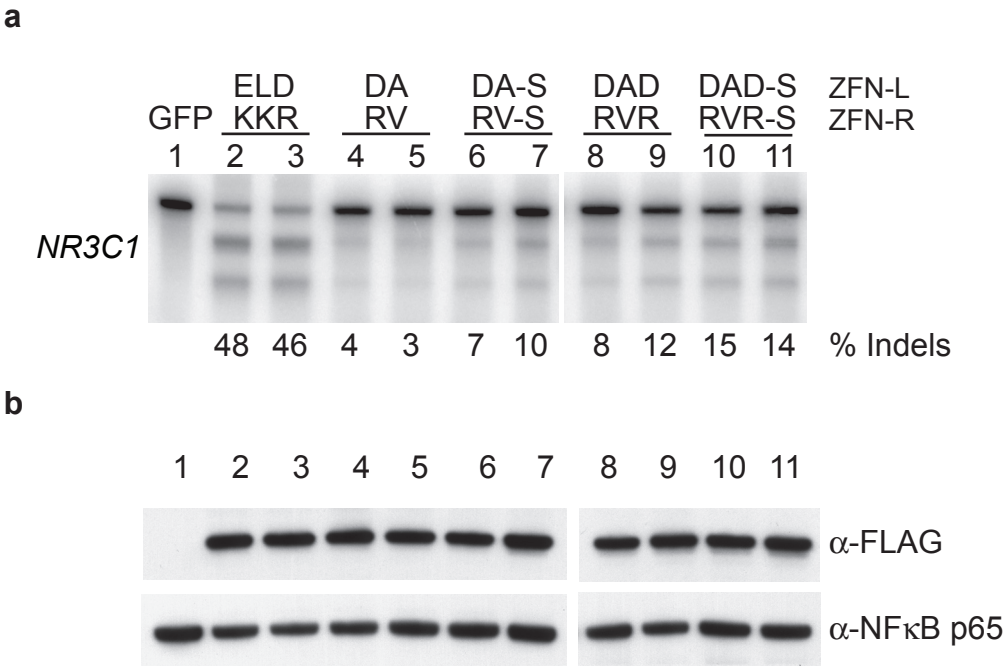
K562 cells were transfected in duplicate with various combinations of expression plasmids encoding the the *CCR5* and *CXCR4* targeting ZFNs (400 ng) and incubated at either 37°C or 30°C for 3 days. The Cel-1 assay was used to determine the frequency of ZFN-induced indels at both targets and four transheterodimer off-targets. The monomers that can potentially combine to form transheterodimers are indicated in parenthesis below the cleavage site. Note that in lanes 6 and 7 we attempted to block transheterodimerization by adding the N496D mutation on the KK domain (DKK) and the H537R mutation on the EL domain (REL) on the *CXCR4*-targeting ZFN. This created a situation where unwanted transheterodimer formation would be decreased by the presence either two N496D or H537R mutations in the heterodimer. As such, the following potential heterodimers ELD:DKK and KKR:REL could form with a decrease efficiency. This strategy resulted in a modest but detectable reduction of transheterodimerization, especially at weak off-target sites (e.g. *FRYL* at 30°C and *C1ORF210*). Note that the transfection of the two wild-type ZFN pairs resulted in decreased cell growth/survival and a resulting drop in on target activity. This was more pronounced under cold shock conditions. * Indicates an underestimated value due to the saturation of the assay.

Supplementary Figure 12 | Additive effect of the ELD:KKR and "Sharkey" FokI domains on ZFN activity.



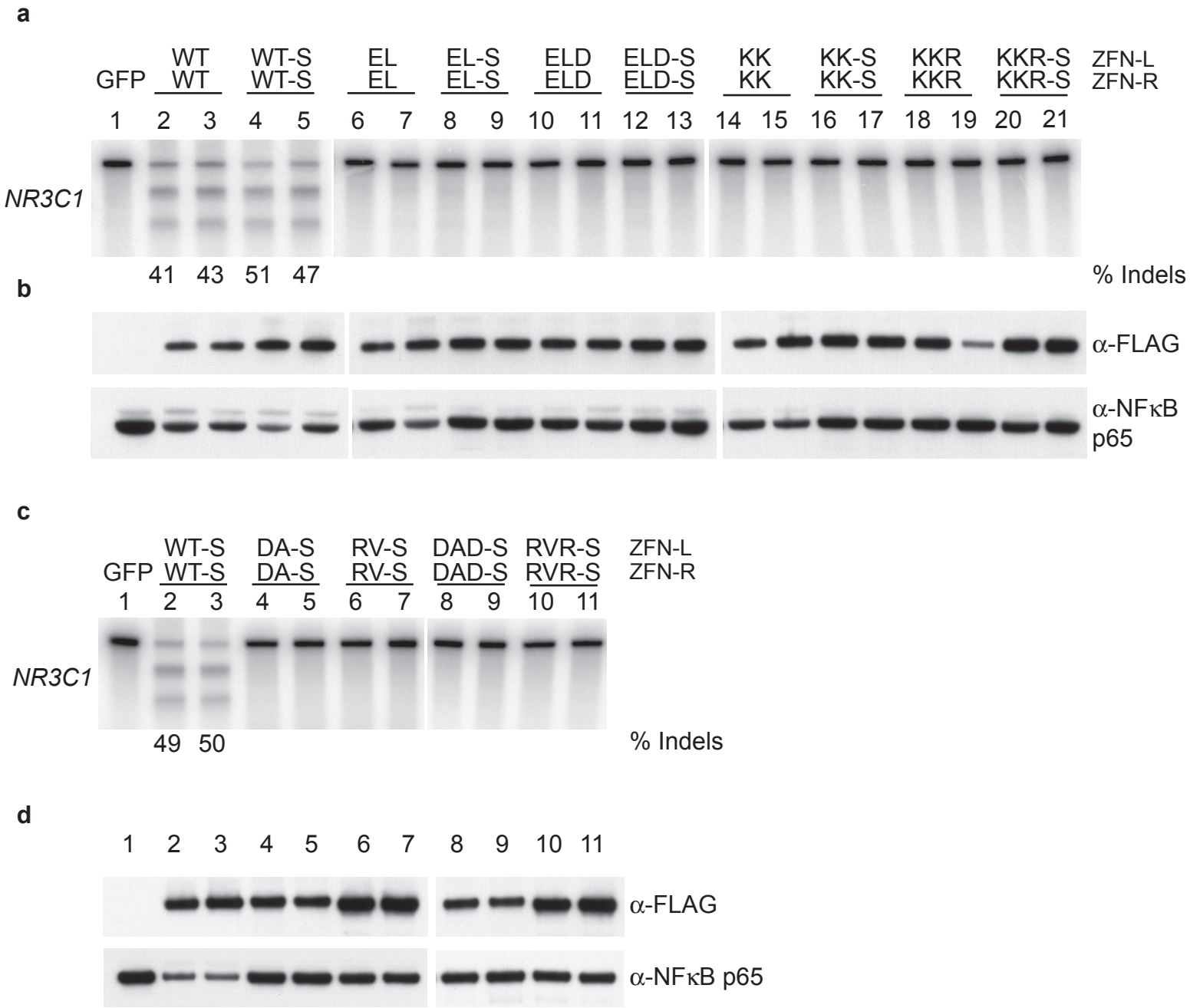
(**a**) Expression plasmids encoding the indicated *KDR* ZFN variants (400 ng) were transfected in duplicate into K562 and the cells were harvested 3 days post-transfection. The Cel-1 assay was used to determine the frequency of ZFN-induced insertions and deletions. An aliquot of the cells was also cultivated for an additional week to determine the stability of the modified cells in extended cultures. (**b**) ZFN expression level was monitored by western blot using the FLAG antibody and α-NFκB p65 was used as a loading control. (**c**) and (**d**) Same as (**a**) and (**b**) but using the expression plasmids encoding the *NR3C1* -targeting ZFNs at a 80 ng dose. –S indicates the "Sharkey" mutations (S418P, K441E).

Supplementary Figure 13 | Additive effect of the DAD:RVR and "Sharkey" FokI domains on ZFN activity.



(**a**) Expression plasmids encoding the indicated *NR3C1* ZFN variants (400 ng) were transfected in duplicate into K562 and the cells were harvested 3 days post-transfection. The Cel-1 assay was used to determine the frequency of ZFN-induced insertions and deletions. (**b**) ZFN expression level was monitored by western blot using the FLAG antibody and α-NFκB p65 was used as a loading control. –S indicates the "Sharkey" mutations (S418P, K441E).

Supplementary Figure 14 | Preservation of the obligate heterodimer specificity in the ELD:KKR and DAD:RVR architectures combined with the "Sharkey" FokI domains.



(**a**) and (**c**) Forced homodimerization of the indicated FokI domains using the *NR3C1* targeting ZFNs was assayed by transfection of expression plasmids encoding the indicated ZFN variants (400 ng) in duplicate into K562 cells. The Cel-1 assay was used to determine the frequency of ZFN-induced indels at the *NR3C1* heterodimer site 3 days post transfection. (**b**) and (**d**) ZFN expression was monitored by western blot using the FLAG antibody and α-NFκB p65 was used as a loading control. –S indicates the "Sharkey" mutations (S418P, K441E).

Supplementary Figure 15 | Unprocessed scans of full-length gels and blots for the primary figures.

Figure 1d

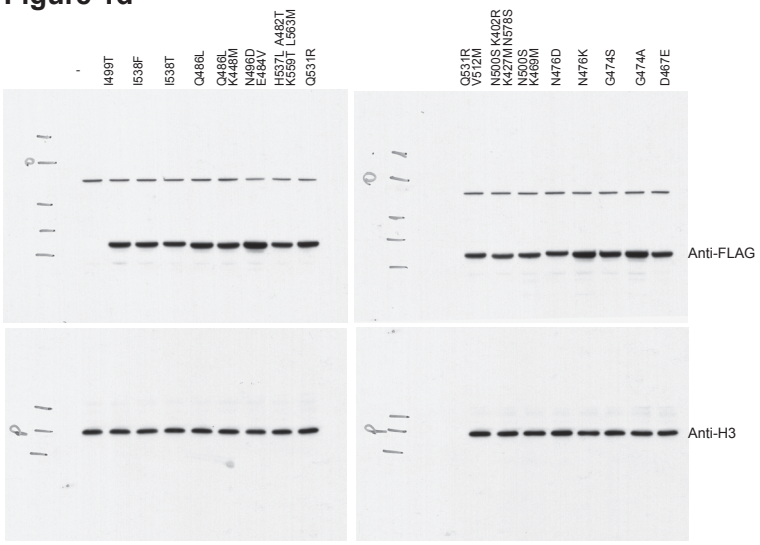


Figure 3a

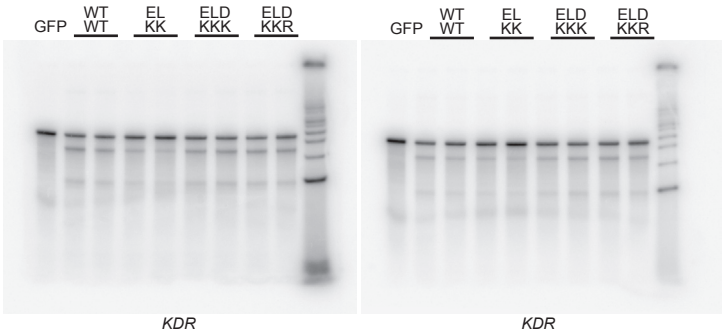
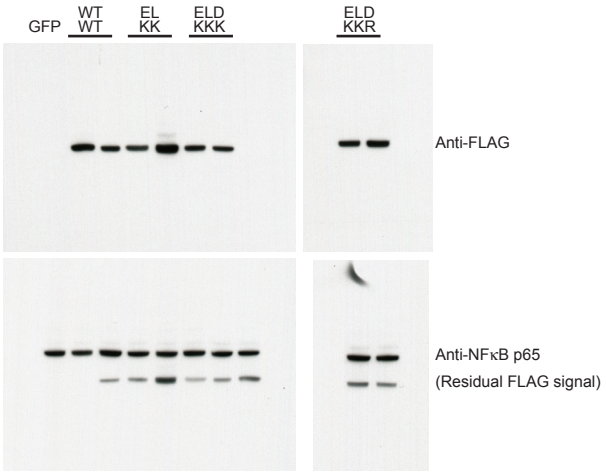


Figure 3b



Membrane was first probed with anti-FLAG, and then incubated with anti-NFkB p65.

Figure 4a

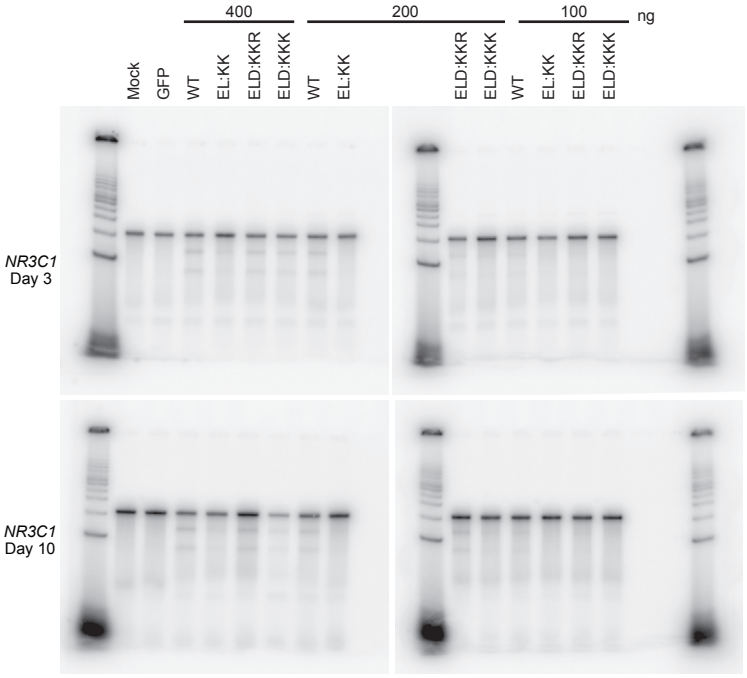
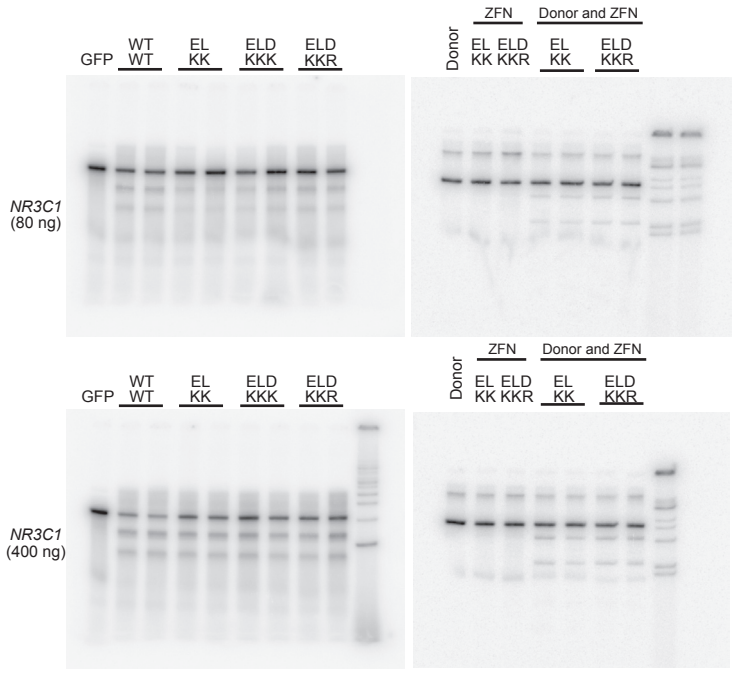


Figure 3c,d



Supplementary Figure 15 | Unprocessed scans of full-length gels and blots for the primary figures.

Figure 5b

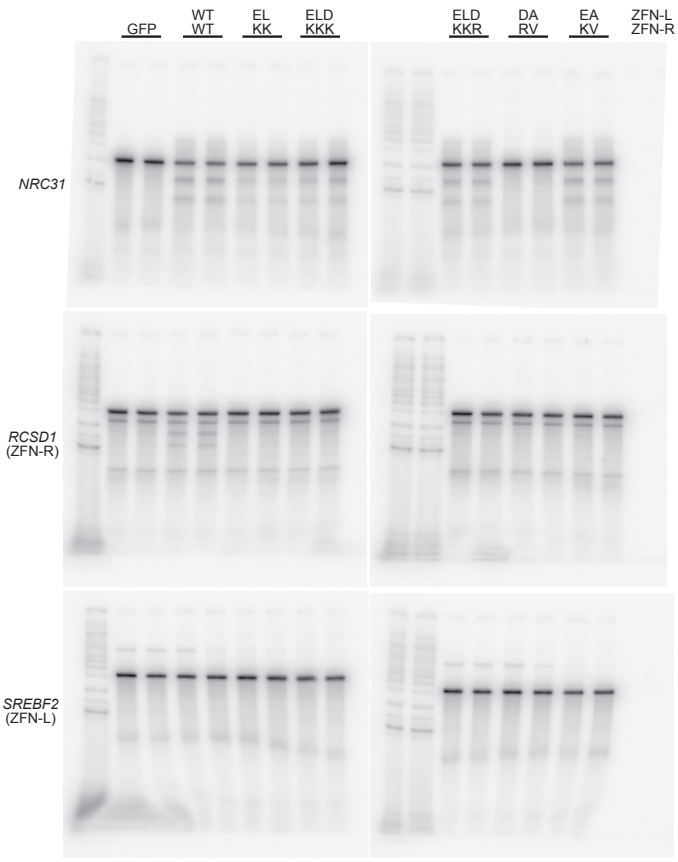


Figure 5d

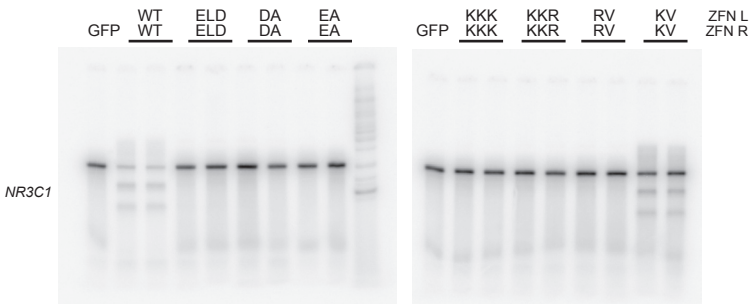
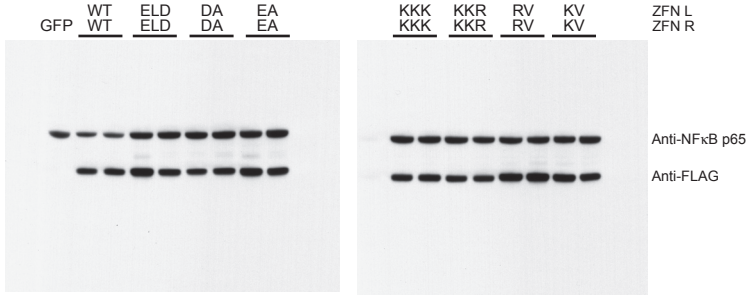
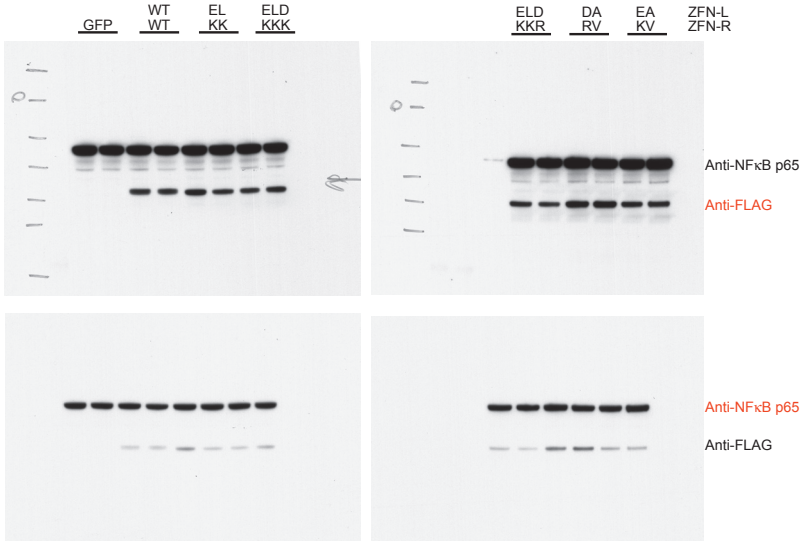


Figure 5e



Membrane was coincubated with anti-FLAG and anti-NFκB p65.

Figure 5c



Membrane was coincubated with anti-FLAG and anti-NFκB p65. Two different exposures of the same membrane are shown. Top gels were scanned for the FLAG and bottom gels scanned for NFκB p65

Supplementary Table 1 | Isolated cold-sensitive mutants.

Mutations				Isolates	Secondary structure
I499T[7]				1	loop P2
I538F[11]				1	α5
I538T[11]				3	α5
Q486L[5]				2	α4
Q486L[5]	K448M			4	α4, loop P1
N496D[6]	E484V[4]			3	loop P2, α4
H537L[10]	A482T[3]	K559T	L563M	1	α5, α4
Q531R[9]				2	α5
Q531R[9]	V512M			1	α5
N500S[8]	K402R	K427M	N578S	1	loop P2
N500S[8]	K469M*			1	loop P2
N476D[2]				1	unannotated
N476K[2]				1	unannotated
G474S[1]				4	unannotated
G474A[1]				5	unannotated
D467E*				1	catalytic triad

The numbers in the "Isolates" column refer to the number of independent isolates of each mutation. The position of the mutations (indicated by a superscript number and shown in Figure 2a,b) relative to the secondary structure is based on the FokI crystal structure[19].
* Denotes a residue in the catalytic triad.

The G474A/S and N476D/K mutations are predicted to occur in a region of the protein that is in close proximity to the DNA, but are also near residue Ile479 which makes a van der Waals contact with Arg447 of the opposite monomer[19].

Supplementary Table 2 | Primers used for the Cel-1 assays.

KDR	F	AGTATGGGGCCCTGTTGAAT
	R	TCCCACTGACCTTCTATTATGAAA
NR3C1	F	TCATAACACTGTTCTTCCCCTTCTTTAGCC
	R	TCAAAACACACACTACCTTCCACTGCTC
RCSD1	F	TCCTTGCCCTCAAGATTCAT
	R	CTGCTCCCACAGCTTAGGTC
SrebF2	F	AGCGACTGCAGAAGCAAGTT
	R	TCTAACCAGCTGGGTGACCT
53BP1	F	GGGGACAGATAGCTTTAAACACC
	R	TTGGTGAGTGATACCTTGTTTGA
RIPK1-A	F	TTCTAACGCTTCTGGCCTGT
	R	ATGCTAACGAGCTGCAAACA
RIPK1-B	F	TGTGGGAAGAGGACCATCTC
	R	GGTAGTTGGCTTCGTCTTGG
CCR5	F	AAGATGGATTATCAAGTGTCAAGTCC
	R	CAAAGTCCCACTGGGCG
ABLIM2	F	CGATGACTCTGAGGTCTACTCG
	R	CAAGTGAACACATGGTTTGCAG
PGC	F	AAGGCAGGAGACCCAGCATTTC
	R	CTACACAGGACTTTCCCTTGGAGC
RSK4	F	GGACAAAAGACAGAAAATGTGAAA
	R	TGCAAAAATTCATGAAATACACTG
TRIM26	F	GGAGTGGTACTGGGCGTGTC
	R	TTCAGGAGGTTTAGAGACCATCAAA
Chr. 1	F	GCAGCAACCTGCCAGCTCTA
	R	CCCATTTGCCAACCAAGAGA
CXCR4	F	CAGTCAACCTCTACAGCAGTGTCC
	R	GGAGTGTGACAGCTTGGAGATG
C1orf210	F	CCTCTGTCCCTGAGGTTCAA
	R	TGCTGGTAGGATTTGTGCTG
TBC1D5	F	TTCCCTTCTTGAAGGCTCAC
	R	TTGTCAGCATGGCATTCAGT
CD274	F	TGGAGAGGCACTAAGAGGGA
	R	CTCACAGCCACTCTTCCAGA
FRYL	F	CCAACCCAACTGCAGGTATATTA
	R	GGATTAGCTTTGAAAAGGGAGG

Supplementary Note 1

ZFN coding sequences

TP53BP1-L (ELD FokI domain)

Target sequence (5’ to 3’) GTTCAGGATTGG

```
atggactacaaagaccatgacggtgattataaagatcatgacatcgattacaaggatgac
 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D
gatgacaagatggcccccaagaagaagaggaaggtcggcatccacggggtacccgccgct
 D  D  K  M  A  P  K  K  K  R  K  V  G  I  H  G  V  P  A  A
atggctgagaggcccttccagtgtcgaatctgcatgcgtaacttcagtcgctccgaccac
 M  A  E  R  P  F  Q  C  R  I  C  M  R  N  F  S  R  S  D  H
ctgtccacccacatccgcacccacaccggcgagaagccttttgcctgtgacatttgtggg
 L  S  T  H  I  R  T  H  T  G  E  K  P  F  A  C  D  I  C  G
aggaaatttgccacctccgccaacctgtcccgccataccaagatacacacgggatctcag
 R  K  F  A  T  S  A  N  L  S  R  H  T  K  I  H  T  G  S  Q
aagcccttccagtgtcgaatctgcatgcgtaacttcagtcgctccgacaacctgtccgag
 K  P  F  Q  C  R  I  C  M  R  N  F  S  R  S  D  N  L  S  E
cacatccgcacccacaccggcgagaagccttttgcctgtgacatttgtgggaggaaattt
 H  I  R  T  H  T  G  E  K  P  F  A  C  D  I  C  G  R  K  F
gccacctccggctccctgacccgccataccaagatacacctgcggggatcccagctggtg
 A  T  S  G  S  L  T  R  H  T  K  I  H  L  R  G  S  Q  L  V
aagagcgagctggaggagaagaagtccgagctgcggcacaagctgaagtacgtgccccac
 K  S  E  L  E  E  K  K  S  E  L  R  H  K  L  K  Y  V  P  H
gagtacatcgagctgatcgagatcgccaggaacagcacccaggaccgcatcctggagatg
 E  Y  I  E  L  I  E  I  A  R  N  S  T  Q  D  R  I  L  E  M
aaggtgatggagttcttcatgaaggtgtacggctacaggggaaagcacctgggcggaagc
 K  V  M  E  F  F  M  K  V  Y  G  Y  R  G  K  H  L  G  G  S
agaaagcctgacggcgccatctatacagtgggcagccccatcgattacggcgtgatcgtg
 R  K  P  D  G  A  I  Y  T  V  G  S  P  I  D  Y  G  V  I  V
gacacaaaggcctacagcggcggctacaatctgcctatcggccaggccgacgagatggag
 D  T  K  A  Y  S  G  G  Y  N  L  P  I  G  Q  A  D  E  M  E
agatacgtggaggagaaccagacccgggataagcacctcaaccccaacgagtggtggaag
 R  Y  V  E  E  N  Q  T  R  D  K  H  L  N  P  N  E  W  W  K
gtgtaccctagcagcgtgaccgagttcaagttcctgttcgtgagcggccacttcaagggc
 V  Y  P  S  S  V  T  E  F  K  F  L  F  V  S  G  H  F  K  G
aactacaaggcccagctgaccaggctgaaccacatcaccaactgcaatggcgccgtgctg
 N  Y  K  A  Q  L  T  R  L  N  H  I  T  N  C  N  G  A  V  L
agcgtggaggagctgctgatcggcggcgagatgatcaaagccggcaccctgacactggag
 S  V  E  E  L  L  I  G  G  E  M  I  K  A  G  T  L  T  L  E
gaggtgcggcgcaagttcaacaacggcgagatcaacttcagatct
 E  V  R  R  K  F  N  N  G  E  I  N  F  R  S
```

TP53BP1-R (KKR FokI domain)

Target sequence (5’ to 3’) GCTGGAGAAGAAcGAGGAG

```
atggactacaaagaccatgacggtgattataaagatcatgacatcgattacaaggatgac
 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D
gatgacaagatggcccccaagaagaagaggaaggtcggcattcatggggtacccgccgct
 D  D  K  M  A  P  K  K  K  R  K  V  G  I  H  G  V  P  A  A
```

```
atggctgagaggcccttccagtgtcgaatctgcatgcgtaacttcagtcagtccggcgcc
 M  A  E  R  P  F  Q  C  R  I  C  M  R  N  F  S  Q  S  G  A
ctggcccgccacatccgcacccacaccggcgagaagccttttgcctgtgacatttgtggg
 L  A  R  H  I  R  T  H  T  G  E  K  P  F  A  C  D  I  C  G
aggaaatttgcccgctccgacaacctgacccgccataccaagatacacacgggcggaggc
 R  K  F  A  R  S  D  N  L  T  R  H  T  K  I  H  T  G  G  G
ggaagccaacggcccttccagtgtcgaatctgcatgcgtaacttcagtcagtccggcaac
 G  S  Q  R  P  F  Q  C  R  I  C  M  R  N  F  S  Q  S  G  N
ctggcccgccacatccgcacccacaccggcgagaagccttttgcctgtgacatttgtggg
 L  A  R  H  I  R  T  H  T  G  E  K  P  F  A  C  D  I  C  G
aggaaatttgcccagtccggcaacctggcccgccataccaagatacacacgggatctcag
 R  K  F  A  Q  S  G  N  L  A  R  H  T  K  I  H  T  G  S  Q
aagcccttccagtgtcgaatctgcatgcgtaacttcagtcagtccggccacctgcagcgc
 K  P  F  Q  C  R  I  C  M  R  N  F  S  Q  S  G  H  L  Q  R
cacatccgcacccacaccggcgagaagccttttgcctgtgacatttgtgggaggaaattt
 H  I  R  T  H  T  G  E  K  P  F  A  C  D  I  C  G  R  K  F
gcccagtcctccgacctgcgccgccataccaagatacacctgcggggatcccagctggtg
 A  Q  S  S  D  L  R  R  H  T  K  I  H  L  R  G  S  Q  L  V
aagagcgagctggaggagaagaagtccgagctgcggcacaagctgaagtacgtgccccac
 K  S  E  L  E  E  K  K  S  E  L  R  H  K  L  K  Y  V  P  H
gagtacatcgagctgatcgagatcgccaggaacagcacccaggaccgcatcctggagatg
 E  Y  I  E  L  I  E  I  A  R  N  S  T  Q  D  R  I  L  E  M
aaggtgatggagttcttcatgaaggtgtacggctacaggggaaagcacctgggcggaagc
 K  V  M  E  F  F  M  K  V  Y  G  Y  R  G  K  H  L  G  G  S
agaaagcctgacggcgccatctatacagtgggcagccccatcgattacggcgtgatcgtg
 R  K  P  D  G  A  I  Y  T  V  G  S  P  I  D  Y  G  V  I  V
gacacaaaggcctacagcggcggctacaatctgcctatcggccaggccgacgagatgcag
 D  T  K  A  Y  S  G  G  Y  N  L  P  I  G  Q  A  D  E  M  Q
agatacgtgaaggagaaccagacccggaataagcacatcaaccccaacgagtggtggaag
 R  Y  V  K  E  N  Q  T  R  N  K  H  I  N  P  N  E  W  W  K
gtgtaccctagcagcgtgaccgagttcaagttcctgttcgtgagcggccacttcaagggc
 V  Y  P  S  S  V  T  E  F  K  F  L  F  V  S  G  H  F  K  G
aactacaaggcccagctgaccaggctgaaccgcaaaaccaactgcaatggcgccgtgctg
 N  Y  K  A  Q  L  T  R  L  N  R  K  T  N  C  N  G  A  V  L
agcgtggaggagctgctgatcggcggcgagatgatcaaagccggcaccctgacactggag
 S  V  E  E  L  L  I  G  G  E  M  I  K  A  G  T  L  T  L  E
gaggtgcggcgcaagttcaacaacggcgagatcaacttc
 E  V  R  R  K  F  N  N  G  E  I  N  F
```

NR3C1-L (ELD FokI domain)

Target sequence (5’ to 3’) GTTGAGGAGCTG

```
atggactacaaagaccatgacggtgattataaagatcatgacatcgattacaaggatgac
 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D
gatgacaagatggcccccaagaagaagaggaaggtgggcatccacggggtacccgagagg
 D  D  K  M  A  P  K  K  K  R  K  V  G  I  H  G  V  P  E  R
cccttccagtgtcgaatctgcatgcgtaacttcagtgacagctggaacctggtccagcac
 P  F  Q  C  R  I  C  M  R  N  F  S  D  S  W  N  L  V  Q  H
atccgcacccacacaggcgagaagccttttgcttgcgacatttgtgggaggaagtttgcc
 I  R  T  H  T  G  E  K  P  F  A  C  D  I  C  G  R  K  F  A
cgctccgccaacctgacccgccataccaagatacacacgggatctcagaagcccttccag
 R  S  A  N  L  T  R  H  T  K  I  H  T  G  S  Q  K  P  F  Q
tgtcgaatctgcatgcgtaacttcagtacctccggcaacctgacccgccacatccgcacc
```

```
 C  R  I  C  M  R  N  F  S  T  S  G  N  L  T  R  H  I  R  T
cacacaggcgagaagccttttgcctgtgacatttgtgggaggaagtttgccacctccggc
 H  T  G  E  K  P  F  A  C  D  I  C  G  R  K  F  A  T  S  G
tccctgacccgccataccaagatacacctgcggggatcccagctggtgaagagcgagctg
 S  L  T  R  H  T  K  I  H  L  R  G  S  Q  L  V  K  S  E  L
gaggagaagaagtccgagctgcggcacaagctgaagtacgtgccccacgagtacatcgag
 E  E  K  K  S  E  L  R  H  K  L  K  Y  V  P  H  E  Y  I  E
ctgatcgagatcgccaggaacagcacccaggaccgcatcctggagatgaaggtgatggag
 L  I  E  I  A  R  N  S  T  Q  D  R  I  L  E  M  K  V  M  E
ttcttcatgaaggtgtacggctacaggggaaagcacctgggcggaagcagaaagcctgac
 F  F  M  K  V  Y  G  Y  R  G  K  H  L  G  G  S  R  K  P  D
ggcgccatctatacagtgggcagccccatcgattacggcgtgatcgtggacacaaaggcc
 G  A  I  Y  T  V  G  S  P  I  D  Y  G  V  I  V  D  T  K  A
tacagcggcggctacaatctgcctatcggccaggccgacgagatggagagatacgtggag
 Y  S  G  G  Y  N  L  P  I  G  Q  A  D  E  M  E  R  Y  V  E
gagaaccagacccgggataagcacctcaaccccaacgagtggtggaaggtgtaccctagc
 E  N  Q  T  R  D  K  H  L  N  P  N  E  W  W  K  V  Y  P  S
agcgtgaccgagttcaagttcctgttcgtgagcggccacttcaagggcaactacaaggcc
 S  V  T  E  F  K  F  L  F  V  S  G  H  F  K  G  N  Y  K  A
cagctgaccaggctgaaccacatcaccaactgcaatggcgccgtgctgagcgtggaggag
 Q  L  T  R  L  N  H  I  T  N  C  N  G  A  V  L  S  V  E  E
ctgctgatcggcggcgagatgatcaaagccggcaccctgacactggaggaggtgcggcgc
 L  L  I  G  G  E  M  I  K  A  G  T  L  T  L  E  E  V  R  R
aagttcaacaacggcgagatcaacttcagatct
 K  F  N  N  G  E  I  N  F  R  S
```

NR3C1-R (KKR FokI domain)

Target sequence (5’ to 3’) CAACAGGACCAC

```
atgagatctgactacaaagaccatgacggtgattataaagatcatgacatcgattacaag
 M  R  S  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K
gatgacgatgacaagatggcccccaagaagaagaggaaggtgggcattcatggggtaccc
 D  D  D  D  K  M  A  P  K  K  K  R  K  V  G  I  H  G  V  P
gccgccatggcggagaggccctacgcatgccctgtcgagtcctgcgatcgccgcttttct
 A  A  M  A  E  R  P  Y  A  C  P  V  E  S  C  D  R  R  F  S
acctcgagggcccttaccgcacatatccgcatccacaccggtgagaagcccttccagtgt
 T  S  R  A  L  T  A  H  I  R  I  H  T  G  E  K  P  F  Q  C
cgaatctgcatgcgtaacttcagtgacagggccaacctgagccgccacatccgcacccac
 R  I  C  M  R  N  F  S  D  R  A  N  L  S  R  H  I  R  T  H
acaggatctcagaagcccttccagtgtcgaatctgcatgcgtaacttcagtcgctccgac
 T  G  S  Q  K  P  F  Q  C  R  I  C  M  R  N  F  S  R  S  D
aacctgtccgagcacatccgcacccacacaggcgagaagccttttgcttgcgacatttgt
 N  L  S  E  H  I  R  T  H  T  G  E  K  P  F  A  C  D  I  C
gggaggaagtttgccgagcgcgccaaccggaactcgcataccaagatacacctgcgggga
 G  R  K  F  A  E  R  A  N  R  N  S  H  T  K  I  H  L  R  G
tcccagctggtgaagagcgagctggaggagaagaagtccgagctgcggcacaagctgaag
 S  Q  L  V  K  S  E  L  E  E  K  K  S  E  L  R  H  K  L  K
tacgtgccccacgagtacatcgagctgatcgagatcgccaggaacagcacccaggaccgc
 Y  V  P  H  E  Y  I  E  L  I  E  I  A  R  N  S  T  Q  D  R
atcctggagatgaaggtgatggagttcttcatgaaggtgtacggctacaggggaaagcac
 I  L  E  M  K  V  M  E  F  F  M  K  V  Y  G  Y  R  G  K  H
ctgggcggaagcagaaagcctgacggcgccatctatacagtgggcagccccatcgattac
 L  G  G  S  R  K  P  D  G  A  I  Y  T  V  G  S  P  I  D  Y
ggcgtgatcgtggacacaaaggcctacagcggcggctacaatctgcctatcggccaggcc
```

```
 G  V  I  V  D  T  K  A  Y  S  G  G  Y  N  L  P  I  G  Q  A
gacgagatgcagagatacgtgaaggagaaccagacccggaataagcacatcaaccccaac
 D  E  M  Q  R  Y  V  K  E  N  Q  T  R  N  K  H  I  N  P  N
gagtggtggaaggtgtaccctagcagcgtgaccgagttcaagttcctgttcgtgagcggc
 E  W  W  K  V  Y  P  S  S  V  T  E  F  K  F  L  F  V  S  G
cacttcaagggcaactacaaggcccagctgaccaggctgaaccgcaaaaccaactgcaat
 H  F  K  G  N  Y  K  A  Q  L  T  R  L  N  R  K  T  N  C  N
ggcgccgtgctgagcgtggaggagctgctgatcggcggcgagatgatcaaagccggcacc
 G  A  V  L  S  V  E  E  L  L  I  G  G  E  M  I  K  A  G  T
ctgacactggaggaggtgcggcgcaagttcaacaacggcgagatcaacttc
 L  T  L  E  E  V  R  R  K  F  N  N  G  E  I  N  F
```

CXCR4-L (ELD FokI domain)

Target sequence (5’ to 3’) GTAGAAGCGGTC

```
atggactacaaagaccatgacggtgattataaagatcatgacatcgattacaaggatgac
 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D
gatgacaagatggcccccaagaagaagaggaaggtcggcatccacggggtacccgccgct
 D  D  K  M  A  P  K  K  K  R  K  V  G  I  H  G  V  P  A  A
atggctgagaggcccttccagtgtcgaatctgcatgcgtaacttcagtgaccgctccgcc
 M  A  E  R  P  F  Q  C  R  I  C  M  R  N  F  S  D  R  S  A
ctgtcccgccacatccgcacccacacaggcgagaagccttttgcctgtgacatttgtggg
 L  S  R  H  I  R  T  H  T  G  E  K  P  F  A  C  D  I  C  G
aggaagtttgcccgctccgacgacctgacccgccataccaagatacacacgggatctcag
 R  K  F  A  R  S  D  D  L  T  R  H  T  K  I  H  T  G  S  Q
aagcccttccagtgtcgaatctgcatgcgtaacttcagtcagtccggcaacctggcccgc
 K  P  F  Q  C  R  I  C  M  R  N  F  S  Q  S  G  N  L  A  R
cacatccgcacccacacaggcgagaagccttttgcctgtgacatttgtgggaggaagttt
 H  I  R  T  H  T  G  E  K  P  F  A  C  D  I  C  G  R  K  F
gcccagtccggctccctgacccgccataccaagatacacctgcggggatcccagctggtg
 A  Q  S  G  S  L  T  R  H  T  K  I  H  L  R  G  S  Q  L  V
aagagcgagctggaggagaagaagtccgagctgcggcacaagctgaagtacgtgccccac
 K  S  E  L  E  E  K  K  S  E  L  R  H  K  L  K  Y  V  P  H
gagtacatcgagctgatcgagatcgccaggaacagcacccaggaccgcatcctggagatg
 E  Y  I  E  L  I  E  I  A  R  N  S  T  Q  D  R  I  L  E  M
aaggtgatggagttcttcatgaaggtgtacggctacaggggaaagcacctgggcggaagc
 K  V  M  E  F  F  M  K  V  Y  G  Y  R  G  K  H  L  G  G  S
agaaagcctgacggcgccatctatacagtgggcagccccatcgattacggcgtgatcgtg
 R  K  P  D  G  A  I  Y  T  V  G  S  P  I  D  Y  G  V  I  V
gacacaaaggcctacagcggcggctacaatctgcctatcggccaggccgacgagatggag
 D  T  K  A  Y  S  G  G  Y  N  L  P  I  G  Q  A  D  E  M  E
agatacgtggaggagaaccagacccgggataagcacctcaaccccaacgagtggtggaag
 R  Y  V  E  E  N  Q  T  R  D  K  H  L  N  P  N  E  W  W  K
gtgtaccctagcagcgtgaccgagttcaagttcctgttcgtgagcggccacttcaagggc
 V  Y  P  S  S  V  T  E  F  K  F  L  F  V  S  G  H  F  K  G
aactacaaggcccagctgaccaggctgaaccacatcaccaactgcaatggcgccgtgctg
 N  Y  K  A  Q  L  T  R  L  N  H  I  T  N  C  N  G  A  V  L
agcgtggaggagctgctgatcggcggcgagatgatcaaagccggcaccctgacactggag
 S  V  E  E  L  L  I  G  G  E  M  I  K  A  G  T  L  T  L  E
gaggtgcggcgcaagttcaacaacggcgagatcaacttcagatct
 E  V  R  R  K  F  N  N  G  E  I  N  F  R  S
```

CXCR4-R (KKR FokI domain)

Target sequence (5’ to 3’) GACTTGTGGGTG

```
atggactacaaagaccatgacggtgattataaagatcatgacatcgattacaaggatgac
 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D
gatgacaagatggccccaagaagaagaggaaggtcggcattcatggggtacccgccgct
 D  D  K  M  A  P  K  K  K  R  K  V  G  I  H  G  V  P  A  A
atggctgagaggcccttccagtgtcgaatctgcatgcgtaacttcagtcgctccgactcc
 M  A  E  R  P  F  Q  C  R  I  C  M  R  N  F  S  R  S  D  S
ctgctgcgccacatccgcacccacacaggcgagaagccttttgcctgtgacatttgtggg
 L  L  R  H  I  R  T  H  T  G  E  K  P  F  A  C  D  I  C  G
aggaagtttgcccgctccgaccacctgaccacccataccaagatacacacgggatctcag
 R  K  F  A  R  S  D  H  L  T  T  H  T  K  I  H  T  G  S  Q
aagcccttccagtgtcgaatctgcatgcgtaacttcagtcgctccgactccctgtccgcc
 K  P  F  Q  C  R  I  C  M  R  N  F  S  R  S  D  S  L  S  A
cacatccgcacccacacaggcgagaagccttttgcctgtgacatttgtgggaggaagttt
 H  I  R  T  H  T  G  E  K  P  F  A  C  D  I  C  G  R  K  F
gccgaccgctccaacctgacccgccataccaagatacacctgcggggatcccagctggtg
 A  D  R  S  N  L  T  R  H  T  K  I  H  L  R  G  S  Q  L  V
aagagcgagctggaggagaagaagtccgagctgcggcacaagctgaagtacgtgccccac
 K  S  E  L  E  E  K  K  S  E  L  R  H  K  L  K  Y  V  P  H
gagtacatcgagctgatcgagatcgccaggaacagcacccaggaccgcatcctggagatg
 E  Y  I  E  L  I  E  I  A  R  N  S  T  Q  D  R  I  L  E  M
aaggtgatggagttcttcatgaaggtgtacggctacaggggaaagcacctgggcggaagc
 K  V  M  E  F  F  M  K  V  Y  G  Y  R  G  K  H  L  G  G  S
agaaagcctgacggcgccatctatacagtgggcagccccatcgattacggcgtgatcgtg
 R  K  P  D  G  A  I  Y  T  V  G  S  P  I  D  Y  G  V  I  V
gacacaaaggcctacagcggcggctacaatctgcctatcggccaggccgacgagatgcag
 D  T  K  A  Y  S  G  G  Y  N  L  P  I  G  Q  A  D  E  M  Q
agatacgtgaaggagaaccagacccggaataagcacatcaaccccaacgagtggtggaag
 R  Y  V  K  E  N  Q  T  R  N  K  H  I  N  P  N  E  W  W  K
gtgtaccctagcagcgtgaccgagttcaagttcctgttcgtgagcggccacttcaagggc
 V  Y  P  S  S  V  T  E  F  K  F  L  F  V  S  G  H  F  K  G
aactacaaggcccagctgaccaggctgaaccgcaaaaccaactgcaatggcgccgtgctg
 N  Y  K  A  Q  L  T  R  L  N  R  K  T  N  C  N  G  A  V  L
agcgtggaggagctgctgatcggcggcgagatgatcaaagccggcaccctgacactggag
 S  V  E  E  L  L  I  G  G  E  M  I  K  A  G  T  L  T  L  E
gaggtgcggcgcaagttcaacaacggcgagatcaacttc
 E  V  R  R  K  F  N  N  G  E  I  N  F
```

The recognition helices are shown in red and the FokI mutations are labeled in blue.

Supplementary Note 2

RFLP donor vector sequence

Ggaagttaaagcccatgtttctaatacaatgaacattatgttatgcccaaacttaacaccatcatttcatatgatagcactttcttatag

tgttaccttatgctccctgaccaaactcccagacatcaacttgtacttttctattttattctagatctttttgtattgttgttttaaatactttcc

tgcccattagaggacctaggagccaccctcctctcccctcttaactgatatttagcctttcatgggctttgcatataatggaaatttca

aaatccaccctgagaaatgaaaaccaagtagaggaaaaataaactcttcaaaacacacactaccttccactgctcttttgaagaaa

actttacagcttccacaagttaagactccataatgacatcctgaagcttcatcagagcacaccaggcagagtttgggaggtggtcc

tgttgttgaggcatccagtccagacgggatccagccatactcactgctgttgaggagctggatggaggagagcttacatctggtc

tcatgctggggctaaagaaggggaagaacagtgttatgatttaactgtcaaaggaatatcaaaatacagttctcttagcttctcactt

catagtcagaatgctcacagtgaactctggcttcaagtgctagcaggcactaaaatatcctagctaaatatattcaaatcatgttatat

tcttctttaaacaaaattaagaatgaggtcatttcttttgaagtgtctccaaaatagaatggtgtggttctggttcacttcttcttctttttttt

ttttttttagatgcttaggatttatttttataatcacg

The tag sequence is shown in red and the BamHI site is underlined. The ZFN binding sites are labeled in blue.