Electronic Supplementary Material (ESI) for Molecular BioSystems

A single-chain TALEN architecture for genome engineering

Ning Sun and Huimin Zhao

Fig S1. Schematic of an eGFP-based single-strand annealing reporter system in yeast. The reporter construct integrated into yeast genome contains a divided *eGFP* gene harboring scTALEN target site and I-CreI target site. The DSB generated by the scTALEN can induce single-strand annealing of the two truncated *eGFP* DNA fragments and reconstitute a complete and functional *eGFP* gene. *ADH1p*, promoter of the *ADH1* gene; *ADH1t*, transcription terminator of the *ADH1* gene; *eGF*, 5'-fragment of the *eGFP* gene; and *GFP*, 3'-fragment of the *eGFP* gene.

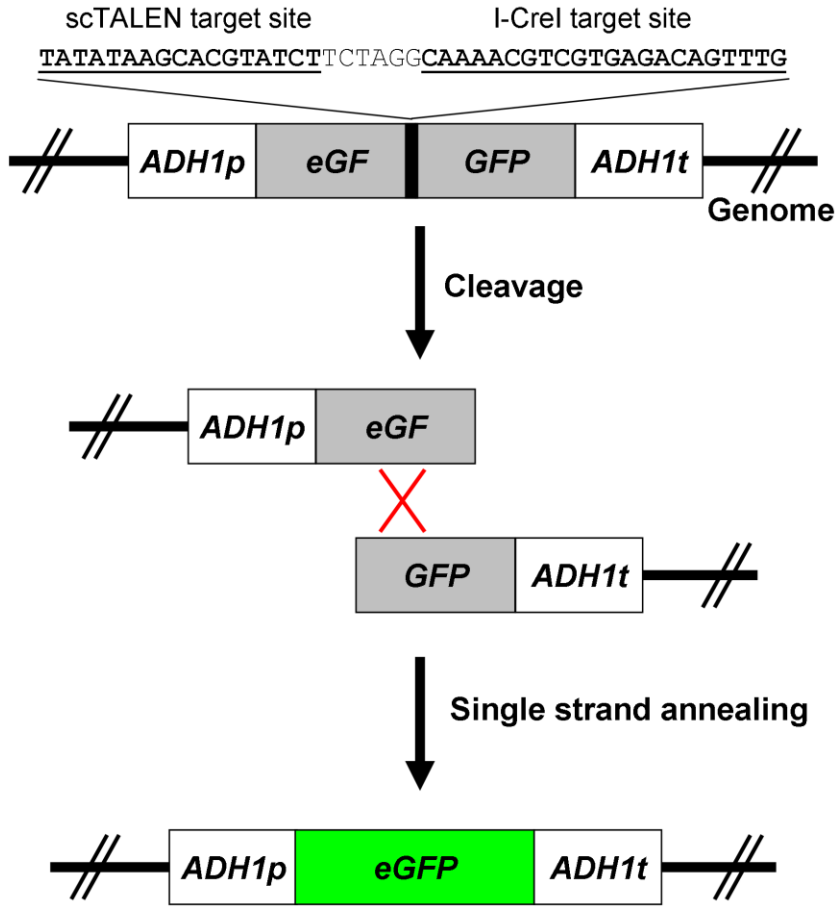

Fig. S2. Iterative high-throughput of scTALENs in the yeast screening system. The eGFP-positive cells in Gate P3 were collected by FACS and applied for the next round of screening. Autofluorescence was measured using a phycoerythrin (PE)-fluorescence channel.

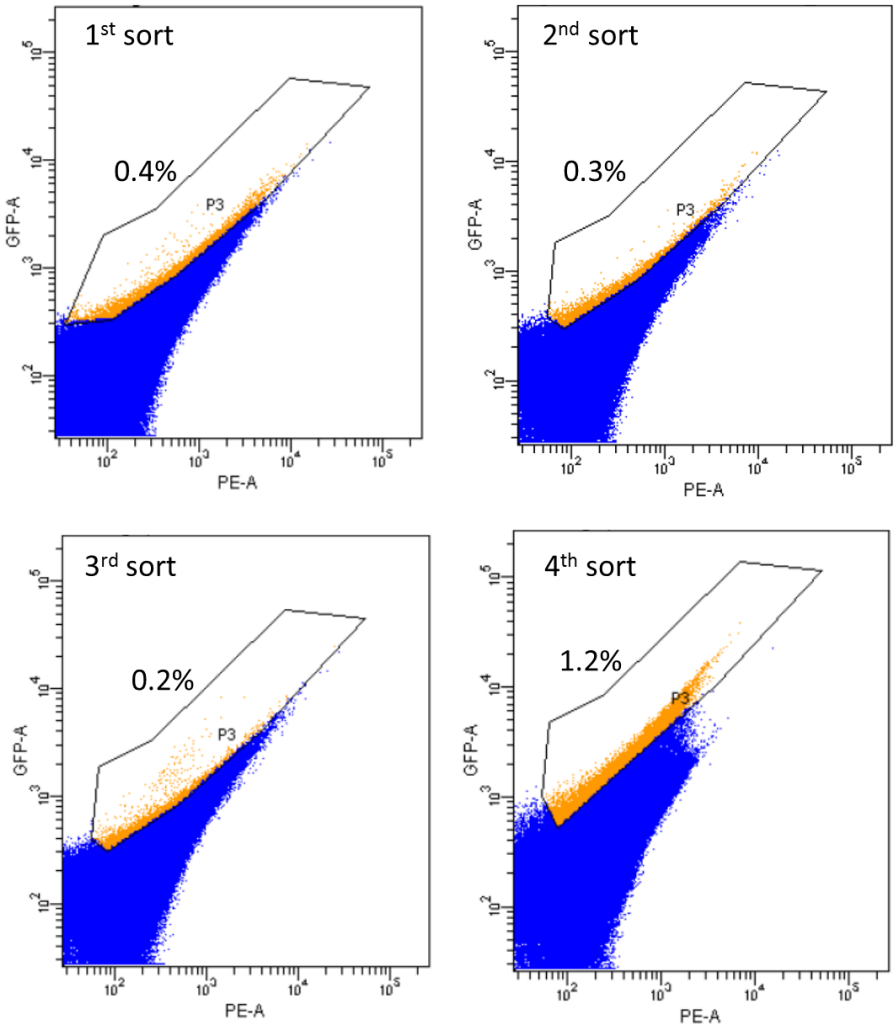

Fig. S3. The amino acid sequence of the scTALEN architecture. The FokI nuclease domains are shown in orange. The NTS is highlighted in magenta. The CTS is highlighted in purple. The polypeptide linker isolated from high-throughput screening is highlighted in red.

LDTSLLDSMPAVGTPHTAAAPAECDEVQSGLRAADDPPPTVRVAVTAARPPRAKPAPRRRAAQP
SDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAV
TYQDIIRALPEATHEDIVGVGKQWSGARALEALLTEAGELRGPPLQLDTGQLLKIAKRGGVTAV
EAVHAWRNALTGAPLN

----------Central repeat domain----------

SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPELIRRINRRIPERTSHRVAL
KQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGS
RKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNEWWKVYPS
SVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNN
GEINFGSGSGSGSITRTTNPRNVVPKIYMSAGSIPLTTHITNSIQPTLWTIGSINGVAPLAKSI
KLGIPVTGSAYTDQTTAMVRKKVSVFMGSGSGSGSSQLVKSELEEKKSELRHKLKYVPHEYIEL
IEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGY
NLPIGQADEMQRYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHIT
NCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

Fig. S4. Distribution of amino acid composition in the original linker library and the isolated linker from the scTALEN architecture. Cysteine was originally excluded from the original linker library to limit entropic costs resulting from intra-loop disulfide bond formation.[1]

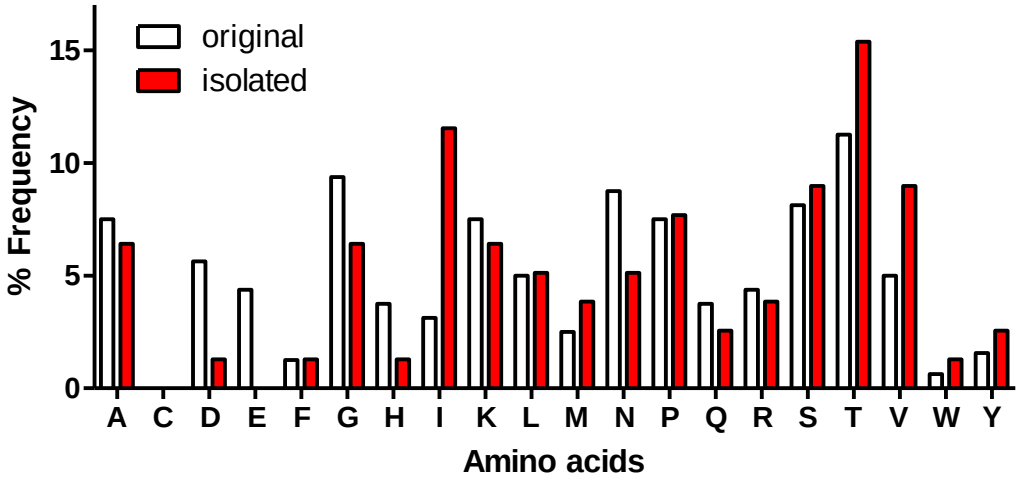

Fig. S5. TALEN target sequences within *ABL1* and *ERCC2* sites of human genome. The recognition sequences of the designer TALEN heterodimer and scTALEN monomers are highlighted in yellow. Repeat variable di-residues (RVDs) of the repeat units are shown above the DNA target sequences.

ABL1

```
5′-TACCTATTATTACTTTATGGGGCAGCAGCCTGGAAAAGTACTTGGGGACCAA-3′
   ||||||||||||||||||||||||||||||||||||||||||||||||||||
3′-ATGGATAATAATGAAATACCCCGTCGTCGGACCTTTTCATGAACCCCTGGTT-5′
```

ABL1-F

RVD	NI	HD	HD	NG	NI	NG	NG	NI	NG	NG	NI	HD	NG	NG	NG	NI	NG
	A	C	C	T	A	T	T	A	T	T	A	C	T	T	T	A	T

ABL1-R

RVD	NG	NN	NN	NG	HD	HD	HD	HD	NI	NI	NN	NG	NI	HD	NG	NG
	T	G	G	T	C	C	C	C	A	A	G	T	A	C	T	T

ERCC2

```
5′-TCCGGCCGGCGCCATGAAGTGAGAAGGGGGCTGGGGGTCGCGCTCGCTA-3′
   ||||||||||||||||||||||||||||||||||||||||||||||||||||
3′-AGGCCGGCCGCGGTACTTCACTCTTCCCCCGACCCCCAGCGCGAGCGAT-5′
```

ERCC2-F

RVD	HD	HD	NN	NN	HD	HD	NN	NN	HD	NN	HD	HD	NI	NG	NN	NI
	C	C	G	G	C	C	G	G	C	G	C	C	A	T	G	A

ERCC2-R

RVD	NI	NN	HD	NN	NI	NN	HD	NN	HD	NN	NI	HD	HD	HD
	A	G	C	G	A	G	C	G	C	G	A	C	C	C

Fig. S6. Histograms to quantify the gel images in SURVEYOR nuclease assay. The relative quantity of each substrate peak was indicated by a black arrow. The relative quantity of each product peak was indicated by a red arrow. *, a non-specific product from genomic PCR.

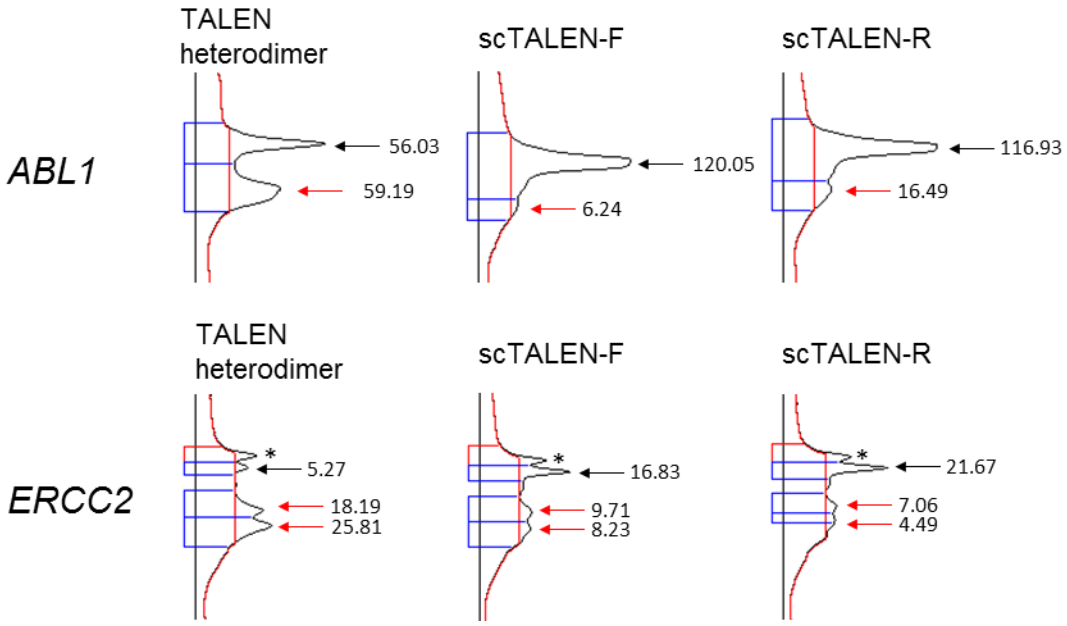

Fig. S7. Representative flow cytometry histograms for the analysis of cellular γ-H2AX expression. HEK293 cells were transfected with nuclease-expression vectors as indicated in targeting *ABL1* site (**A**) and *ERCC2* site (**B**). A gate was drawn (dashed line) to determine the percentage of γ-H2AX staining positive cells.

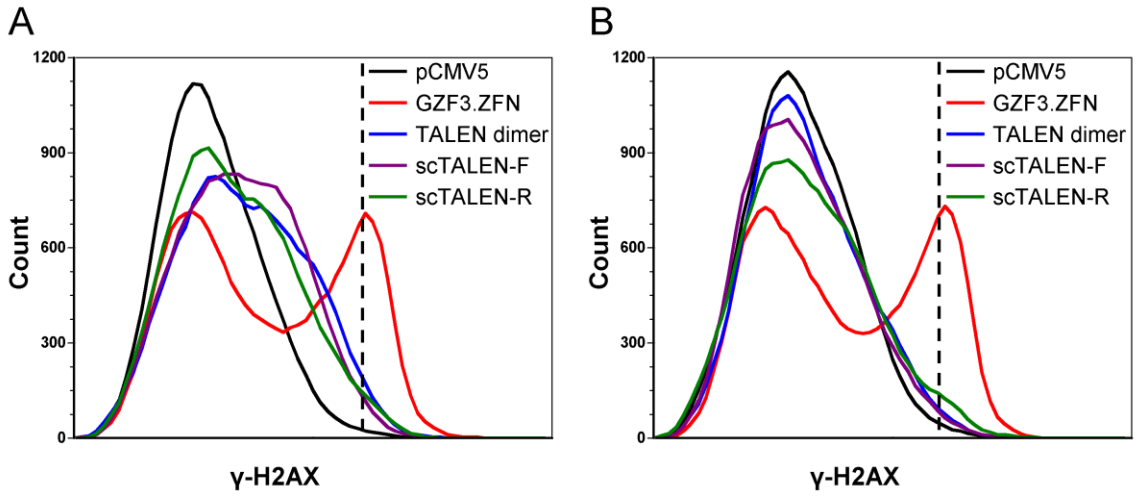

Fig. S8. The activity of a rational-designed scTALEN architecture, in which the two FokI nuclease domains are connected by a flexible $(Gly\text{-}Gly\text{-}Gly\text{-}Gly\text{-}Ser)_{10}$ polypeptide linker. (A) Schematic of the scTALEN-$(G_4S)_{10}$ architecture. The flexible $(Gly\text{-}Gly\text{-}Gly\text{-}Gly\text{-}Ser)_{10}$ polypeptide linker was introduced to facilitate the formation of an intra-molecular FokI dimer, which was successfully applied to make a single-chain zinc finger nuclease.[2] (B) *In vitro* cleavage analysis of the scTALEN-$(G_4S)_{10}$ architecture. Various concentrations of purified protein were incubated with linear DNA substrate, which contains a TALE recognition sequence. The standard TALEN served as a positive control.

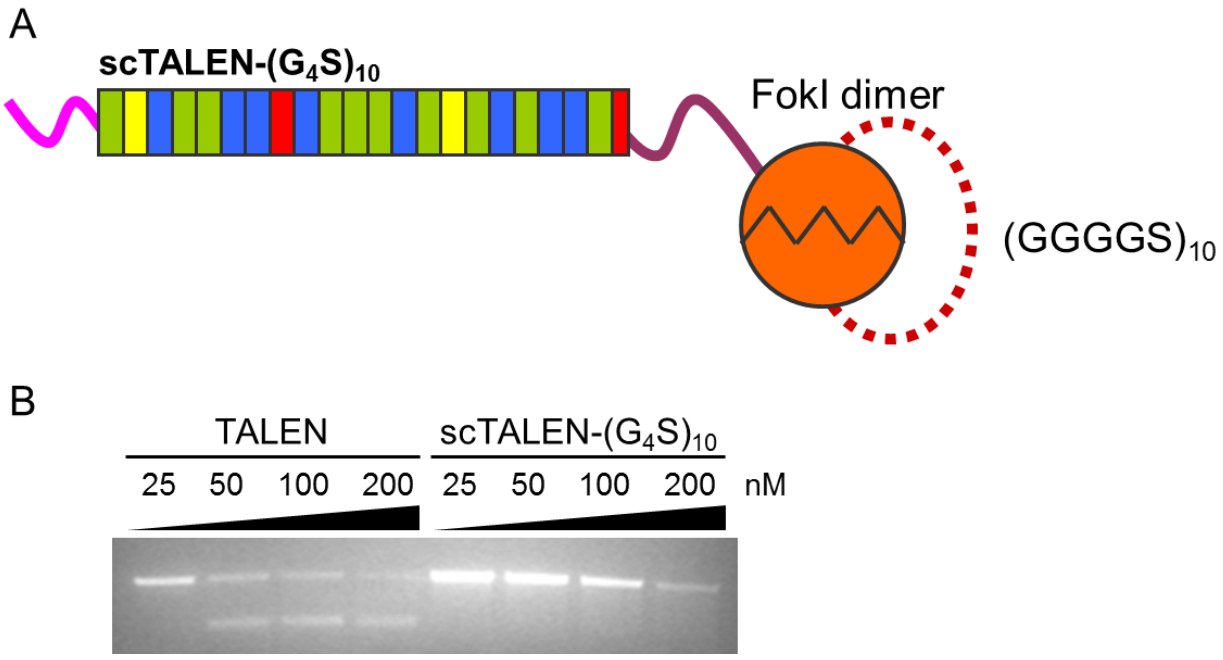

Fig S9. Plasmid map of the recipient plasmid for the construction of scTALEN library.

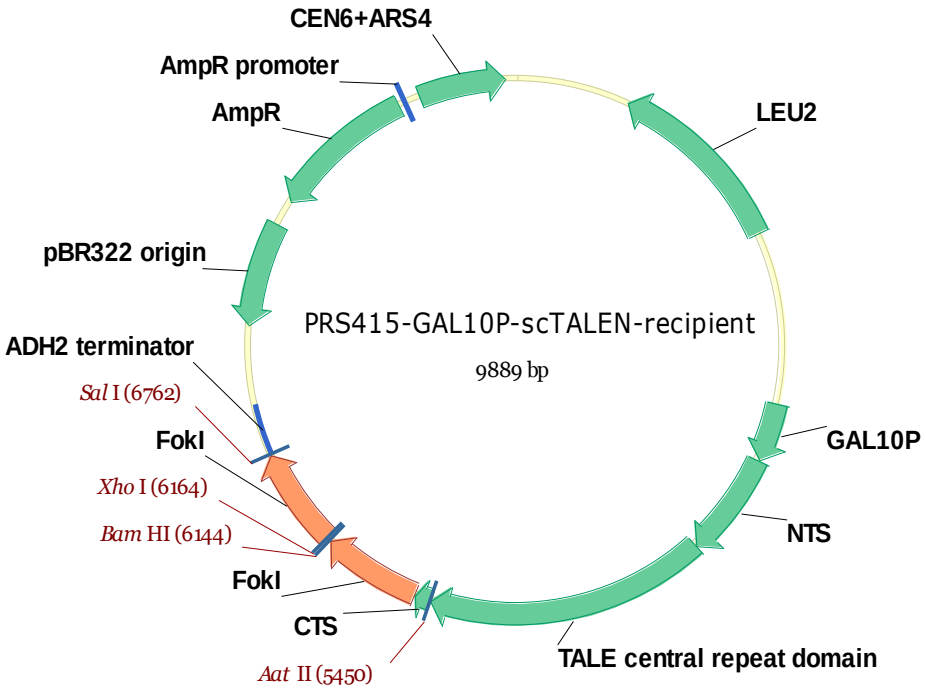

Table S1. Amino acid sequences of linker regions in scTALENs isolated from the third and fourth round of high-throughput screening. The linkers were labeled as S**X**C**Y** (clone number **Y** after **X** rounds of sorting). The activity of each scTALEN was tested in a *lacZ*-based yeast reporter system (Fig. 3A). Only the scTALEN with linker S3C6 (highlighted) showed significant activity in yeast (Fig. 3B&C) and was characterized in detail in human cells (Fig. 4).

Label	Linker amino acid sequence
S3C2	GSGSRSAPEAMYAPDAVKMVHEFGGSNYTGMTIKMSTSLTVTDEGSSHTTQPHAE NTRRDDQNIGSRFAGESHVNNTTKTTKLEGSGSGSG
S3C5	GSGSGSTPWKPIIARHDRRRPLSTGSGSGSG
S3C6	GSGSGSITRTTNPRNVVPKIYMSAGSIPLTTHITNSIQPTLWTIGSINGVAPLAKSIKLG IPVTGSAYTDQTTAMVRKKVSVFMGSGSGSG
S3C7	GSGSGSMNKMQPNWTFTWQANSLIGSGSGSG
S3C14	GSRFAGESHVNNTTKTTKLEGSGSGSGSS
S4C3	GSGSGSTHRKRHPPMTKEVIPPTAGSKVVKPNLPTNNARIFWNIGSTLTVWTSVTN MQQEFTTTGSGSGSG
S4C5	GSGSGSNYAAKPIPSAGQLETSHNGSGSGSG
S4C6	GSGSGSTTRRWPPGRPNQLRNLTTGSGSGSG
S4C7	GSGSGSATMTANFQVNSKPITSPDGSGSG
S4C13	GSGSGSAPEAMYAPDAVKMVHEFGGSNYTGMTIKMSTSLTVTDEESSHTTQPRAE NTRRDDQNIGSRFAGESHVNNTTKTTKLEGSGSGKRLEPLPDPSD
S4C15	GSGSGSTILKRYSLKNEQPRALHIGSGSG
S4C16	GSGSGSVLMTLIDRATASTKPKDAGSAHKGMPHARPRPWASSTVGSGTLTKIF

References

1. H. Li, S. Pellenz, U. Ulge, B. L. Stoddard and R. J. Monnat, Jr., *Nucleic Acids Res*, 2009, **37**, 1650-1662.
2. T. Mino, Y. Aoyama and T. Sera, *J Biotechnol*, 2009, **140**, 156-161.