DNA sense-and-respond protein modules for mammalian cells

DNA sense-and-respond protein modules for mammalian cells

Shimyn Slomovic[1,2,3] and James J. Collins[1-6*]

1. Institute for Medical Engineering & Science, Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, USA.

2. Department of Biological Engineering, MIT, Cambridge, Massachusetts, USA.

3. Synthetic Biology Center, MIT, Cambridge, Massachusetts, USA.

4. Harvard-MIT Program in Health Sciences and Technology

5. Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

6. Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts, USA.

* email: jimjc@mit.edu phone: 617-324-6607 fax: 617-253-7498

We have generated synthetic protein components that can detect specific DNA sequences and subsequently trigger a desired intracellular response. These modular sensors exploit the programmability of zinc finger DNA-recognition to drive the intein-mediated splicing of an artificial trans-activator that signals to a genetic circuit containing a given reporter or response gene. We utilize the sensors to mediate sequence recognition-induced apoptosis, as well as to detect and report a viral infection. This work establishes a synthetic biology framework for endowing mammalian cells with sentinel capabilities, providing a programmable means to cull infected cells, and for identifying positively transduced or transfected cells, isolating recipients of intentional genomic edits, and increasing the repertoire of inducible parts in synthetic biology.

Cys_2His_2 ZFs zinc finger proteins (ZFs) have long been recognized for their potential as artificial DNA binding proteins. Various methods, such as partially randomized libraries or single-finger modular assembly[1,2], may be used to design ZFs to bind any DNA sequence, enabling a range of applications. For example, when attached to the *Fok* I endonucleolytic domain, ZF pairs can facilitate genomic editing[3]. Through SEquence Enabled Reassembly (SEER), ZFs joined to split b-lactamase can function as *in vitro* diagnostic tools[4]. Also, fusion to the viral trans-activating domain, VP64, generated synthetic transcription factors (TFs) that drive engineered regulatory circuits in yeast or mammalian cells, furthering our understanding of transcriptional cooperativity in eukaryotes[5, 6]. Additionally, ZFs fused to chromatin regulators have been used to investigate epigenetics, revealing synergistic activation and spatial regulation[7].

Commonly, whether utilizing ZFs or other genome engineering tools, such as TALEs or CRISPR-Cas, the tethered, delivered activity acts locally on the DNA target itself, e.g., endonucleolytic cleavage or transcriptional modulation[8]. There are however instances in which it is desirable that the detection of a specific endogenous or synthetic DNA sequence trigger a tailored response. Responses could include an amplified report signal indicating provirus presence, a chromosomal transposition, or an intentional genomic alteration. Others might be immune system recruitment to the site of infection or tumor, or kill-switch activation that induces apoptosis to prevent pathogen spread. Accordingly, we sought to forward engineer a system that could utilize the programmable nature of ZFs for target detection and deliver the recognition signal to an easily modifiable response circuit.

To bridge the gap between DNA recognition and trans-response, we used inteins; peptides that splice their flanking regions, referred to as exteins[9,10]. In pioneering work on Conditional Protein Splicing (CPS), the contiguous *Saccharomyces cerevisiae Sce*VMA intein was artificially split, rendering the separate halves inactive such that they require a condition for co-localization and reassembly [11]. After fusing one member of the rapamycin-binding heterodimeric pair, FRB-FKBP, to each of the split *Sce*VMA intein-extein halves, rapamycin served as a condition for their co-localization[12]. The repertoire of trans-splice-inducing conditions has expanded and now includes temperature, light, and protein scaffold activation[13,14,15]. DNA-recognition can also induce splicing, as shown in a previous study wherein ZFs were fused to a split intein/luciferase construct and their binding signal measured drug-induced DNA demethylation[16].

Here, we sought to use DNA-sensing modules as dimerizing domains (**Fig. 1**). Programming two ZF-based sensors to bind adjacently within the targeted DNA sequence would allow for conditional intein co-localization and trans-splicing of a response factor that activates any gene installed in a response circuit. Below, we describe the engineering of a modular sequence-recognition system capable of producing a customizable response signal, and demonstrate its ability to mediate sequence recognition-induced apoptosis and virus detection in infected, human cells.

Results

Modular ZF sensor design and screening

At the onset of this work, we utilized a library of ten tridactyl ZFs designed using Oligomerized Pool ENgineering (OPEN) and tested for intracellular binding (**Supplementary Fig. 1** and **Supplementary Note 1**)[1,5]. To generate two hexadactyl ZF sensors of high affinity and specificity, we first applied an extended linker that permits finger multimerization without excessive DNA strain, and merged two pairs of the tridactyl ZFs[17, 18]. The hexadactyl ZFs were fused to VP64 and transformed into synthetic TFs. Reporters were prepared for each ZF-TF encoding GFP modulated by a single copy of the two original nine bp sites with 0-2 unrelated bps between (**Supplementary Fig. 2a**). We transiently co-transfected human embryonic kidney (HEK 293FT) or HeLa cells with each ZF-TF and its reporter, assessing intracellular binding by measuring GFP fluorescence via flow cytometry. The hexadactyl ZF-TFs were markedly more active than their tridactyl counterparts (**Supplementary Fig. 2b,c**). To assess whether the hexadactyl ZFs could overcome potential binding inhibition imposed by intein fusion in a relevant, intracellular context, we attached the I_C (C-terminal) or I_N (N-terminal) intein domains, C-terminally mounted VP64, and retested activation (**Fig. 2** and **Supplementary Fig. 3a,b**). Although binding of the hexadactyl ZF-intein proteins was encumbered by intein fusion, the chimers were yet able to achieve a much greater output than the tridactyl-intein fusions. We found that a 2× $(GGGS)_3$ linker between the ZF and intein domains yielded a higher level of GFP activation than the 1× or 3× forms, and used this linker thereafter (**Supplementary Fig. 3c-e**).

Response circuit optimization

ZF9 (**Supplementary Fig. 1b,c**) was chosen as the eventual split response factor, as it performed well in tested cell lines and its operator yielded low basal expression. GFP was selected as a reporter for response circuit characterization. To improve the dynamic range of GFP activation, ZF9 operators were multimerized (**Supplementary Fig. 4a**). The output signal strength increased with the number of sites, basal expression (in ZF9-TF's absence) decreased (resulting in a ~300x fold change using the 6x operator circuit), and there was no cross-activation of the response circuit by direct binding of the sensor components (**Supplementary Fig. 4b-d**).

Split extein design, construction, and validation

We hypothesized that an effective way to split the ZF9-TF extein would be to separate its DNA recognition and trans-activation properties, also allowing for the introduction (and retention) of a splice-junction (SJ) that could promote splicing without impairing ZF9-TF's activity[19,20]. In the SJ design, we adopted several extein residues that natively flank both sides of the *Sce*VMA intein or from two previously split proteins (Luciferase and TEV protease), reasoning that the intein might successfully ligate the ZFs and VP64 within this context (**Fig. 3a** and **Supplementary Fig. 5a**)[12,21].

First, we ensured that the retained SJs did not interfere with ZF9-TF's ability to activate GFP expression by expressing the ZF-TF variants in their post-splice forms (**Supplementary Fig. 5b**). Next, we devised a cis-splicing, test platform to gauge whether the intein could splice the split extein and produce a functional TF. The split-extein domains were fused to the corresponding divided inteins, which

were directly linked to one another via a flexible linker (**Fig. 3b**)[22]. Results indicated that all three variants underwent successful cis-splicing, resulting in GFP activation (**Fig. 3c** and **Supplementary Fig. 5c**). To confirm that cis-splicing yielded a product of anticipated size, protein was isolated and subjected to immuno blotting. FLAG- and HIS-tag probing of the N- and C-termini revealed a product that co-migrated with the ZF9-TF positive control for each of the variants but not for three splice-null mutants, which displayed greater accumulation of unspliced precursor (**Fig. 3d** and **Supplementary Fig. 5d-f**)[22]. Semi-quantitative analysis, by comparison of splice-product signal intensity to that of increasing amounts (volumes) of the ZF9-TF positive control, displayed >20% cis-splice efficiency (**Supplementary Fig. 5g**).

Intracellular DNA detection and trans-splicing

To form the complete sensors we fused the three SJ variants of the split-extein with their hexadactyl sensing components. Cells were transfected with a plasmid encoding the sensor pair, a plasmid carrying the target sequences, and one containing the response circuit (**Fig. 4a**). Flow cytometry analysis showed substantial fluorescence output for all three variants, indicating target sequence detection (**Fig. 4b** and **Supplementary Fig. 6a**). Fluorescence measured in the absence of target sequence, the N- or C-terminal sensor, or when a C-terminal deletion mutant (lacking the I_C domain) was tested, was comparably low, suggesting extremely low levels of spontaneous trans-splicing, confirming earlier reports and that both intact intein domains are required in order to generate a response (**Supplementary Fig. 6b**)[12]. Sensor pair V2 (containing the Luciferase SJ) displayed

the greatest activity and was chosen for further characterization. Targets containing zero, four, eight, or twelve bp gaps (nonrelated bps) between the two 18 bp binding sites produced response signals, with the contiguous 36 bp site displaying the highest output (**Supplementary Fig. 6c**). Increasing amounts (nanograms) of target plasmid resulted in a dose-response effect correlating between plasmid amount and signal strength (**Fig. 4c**).

Sensor binding strength is essential to overcome off-target dilution but could possibly limit the interchange of the post- and pre-splice sensors upon the DNA. We reasoned that by moderately lowering affinity but not specificity, interchange might be improved. Eight R→A mutations expected to lower ZF binding affinity were introduced into the ZF backbones and output signals of wildtype and mutant sensor pairs were compared[23]. Equal-molar target plasmids containing a gradient of target sequence instances (0-4, 8) produced a correlative effect for wildtype and mutant sensors but the output of the latter was higher, suggesting improved interchange (**Fig. 4d**). We also assessed the difference between a non-replicative and replicative target and found the signal to be substantially higher in the latter case (**Supplementary Note 2** and **Supplementary Fig. 7a,b).** To demonstrate the platform's extensibility, a second sensor pair was constructed (ZF3/ZF2 and ZF4/ZF5) and co-transfected with either its own target or that of the original sensors. Each pair detected its specific sequence while ignoring the other's, highlighting the platform's robustness and orthogonality (**Fig. 4e**).

DNA sequence recognition-induced apoptosis

Gene-directed enzyme prodrug therapy (GDEPT) involves the metabolism of a nontoxic prodrug into a cytotoxic form[24]. One example is the NTR+CB 1954 pair, wherein the toxicity of CB 1954 (5-(aziridin-1-yl)-2,4-dinitrobenzamide) is dependent upon its reduction by the bacterial nitroreductase, NTR, which transforms it into an agent of DNA interstrand crosslinking and apoptosis[25]. Diffusion into neighboring cells leads to increased potency through a "bystander" effect. This dual-component method was used to test the system's ability to link sequence recognition with triggered apoptosis. With GFP replaced by NTR in the response circuit and samples grown in prodrug-supplemented media, NTR expression and the ensuing CB 1954 reduction should, in principle, only occur in the presence of the target sequence[26].

HEK 293FT cells were plated in media containing a range of prodrug concentrations and transfected (at 50% rate, based on a control) with either the sensor system containing an "empty" response circuit or the full system, in the presence or absence of target sequences (**Fig. 5a** and **Supplementary Fig. 8a**). Apoptosis was measured using flow cytometry after 48 h and 72 h via Annexin V-FITC and propidium iodide (PI) staining. Cells with the empty response circuit were not notably affected by the prodrug, and cells transfected with the full system but no target, were only moderately affected. However, target detection led to markedly elevated levels of apoptotic cells displaying a clear trend in line with the drug gradient. Microscopy revealed that at 96 h, control populations were excessively confluent (**Fig. 5b** and **Supplementary Fig. 8b**). Conversely, target-positive samples at 16 or 32 (μM) displayed severe or complete cell death, providing evidence of

sequence-recognition and a bystander effect. To demonstrate a dual-output response, a GFP-IRES-NTR construct was also generated and simultaneously showed target reporting and apoptosis (**Supplementary Fig. 8c**).

Virus detection in infected mammalian cells

To test the system's capability of detecting viral infection, we adopted adenovirus type 5 (Ad5) as a model[27]. Virus was generated with constitutive expression of blue fluorescent protein (BFP) to track infected cells. Two sub-strains were prepared in which target sequences of sensor pair 1 or 2 (T1, T2) were embedded into the viral genome adjacent to the packaging signal (ψ) (**Fig. 6a**). Replication-permissive HEK 293FT cells were transfected with either of the sensor pairs, followed by infection with virus containing T1, T2, or neither target. Results revealed that both sensor pairs solely detected virus carrying their specific targets (**Fig. 6b,c**). Finally, the expression, reporting, and marker elements were condensed into a single construct. Only cells receiving both the detection system and virus displayed a substantial GFP response (**Supplementary Fig. 9**). Together, these studies demonstrate the feasibility of using this system for sequence-specific DNA detection in mammalian cells.

Discussion

Designable DNA-binding proteins have been developed to regulate gene expression in endogenous pathways and insulated exogenous gene networks. Using such proteins, we have developed a system that is sufficiently robust so as to facilitate sequence detection within mammalian cells but also highly modular in order to

allow its application to a range of scientific settings, including synthetic biology, diagnostics, and basic research. Below, we discuss the system's underlying features as well as possible expansions through which the system could be applied.

The recognition stage may be adjusted to detect any target sequence by installing the appropriate ZF component. Also, alternative DNA-binding technologies could be tested, including TALEs or the CRISPR/Cas system[28]. Another option would be the use of designer RNA-binding proteins for the detection of pathogens with RNA-based genomes[29]. Cells infected with a particular pathogen, retaining viral latency or undergoing carcinogenesis could be culled through the expression of pro-apoptotic proteins, cell death-activating surface molecules, or cytokines for immune system recruitment [30,31,32, 33].

This system has the potential to address many research challenges. For example, it could be adapted to identify or rescue only cells that have received a construct during transfection or transduction, using sensors programmed to recognize the vector backbone sequence. Sensors could be standardized, integrated into stable cell lines, and paired with specific delivery vectors, streamlining the delivery process and preserving packaging space.

Another possible use of this platform regards the CRISPR-Cas system, which is commonly utilized to introduce mutations or replace genomic sequences via homology directed repair[8]. An intrinsic part of such studies is the identification and isolation of cells containing the edit. PCR or DNA sequencing can measure frequencies but cannot be applied to live cells. Furthermore, to establish a clonal cell line containing the edit, multiple lines must undergo screening. Instead, cells could

be equipped with sensors designed to bind within the insertion or deletion site, easily identifying on a single-cell basis cells that have undergone the desired editing, showing gain or loss of the report signal. This approach could greatly benefit studies of genome editing, increasing throughput and efficiency.

Similarly, this concept could be used to produce clonal cell lines harboring HIV provirus, which are essential for studying viral latency. Such cell lines are screened through a tedious process that involves infection, serial dilution, and viral activation to identify clones that contain the provirus[34]. Using a sense-and-respond system programmed to recognize the known proviral sequence and produce a reporter signal, allowing for the rapid identification of positive cells and their subsequent characterization, could markedly shorten this process.

This system could also contribute to studies of chromosomal aberrations leading to cancer and other diseases. In some instances, fluorescence in situ hybridization may be applied and in others, genomic DNA may be subjected to PCR to indicate the presence of chromosomal inversions[35]. However, PCR-based methods cannot be used to study real-time biology nor disclose the aberration frequency within cell populations. Instead, by designing sensors whose proximity would be dependent on the occurrence of a specific chromosomal aberration, researchers could study the biology of living cells and also measure the impact of genetic predisposition or carcinogens.

The study of whole-genome 3D architecture relies on high-throughput chromosomal confirmation capture techniques (3C, 4C, 5C, Hi-C) to map long-range interactions[36]. Stages include DNA crosslinking, digestion, deep-sequencing,

microarrays, and the computational reconstruction of the 3D structure. With appropriate sensor design, a proximity-based DNA sense-and-respond approach could confirm the validity of proposed long-range interactions, in living cells, also providing insight about whole-genome real-time dynamics.

The sensors presented here may also expand the limited range of orthogonal inducible components for synthetic biology researchers, especially in cell-free platforms. Although engineered TFs can bind any sequence, providing limitless operators, the lack of inducible parts constrains the level of complexity and tunability synthetic circuits can achieve. Here, each sensor pair corresponds to its own binding site, and numerous pairs can be generated. In the form of short dsDNA duplexes, their binding sites could be used as inducers to control the reconstitution of orthogonal split ZF-TFs, yielding highly tunable and layered gene networks.

ACKNOWLEDGMENTS

We are grateful to S. Modi for helpful discussions and critical reviews of the manuscript. HEK 293FT cells were donated by the R. Weiss lab, MIT. This work was supported by funding from the Defense Advanced Research Projects Agency grant DARPA-BAA-11-23, Defense Threat Reduction Agency grant HDTRA1-14-1-0006, and AFOSR-BRI grant FA9550-14-1-0060.

AUTHOR CONTRIBUTIONS

S.S. and J.J.C. conceived the study, analyzed data and wrote the paper. S.S. designed and performed the experiments.

COMPETING FINANICIAL INTERESTS

The authors declare no competing financial interests.

References

1. Maeder, M.L. et al. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Molecular cell* **31**, 294-301 (2008).
2. Kim, S., Lee, M.J., Kim, H., Kang, M. & Kim, J.S. Preassembled zinc-finger arrays for rapid construction of ZFNs. *Nature methods* **8**, 7 (2011).
3. Tebas, P. et al. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *The New England journal of medicine* **370**, 901-910 (2014).
4. Ooi, A.T., Stains, C.I., Ghosh, I. & Segal, D.J. Sequence-enabled reassembly of beta-lactamase (SEER-LAC): A sensitive method for the detection of double-stranded DNA. *Biochemistry* **45**, 3620-3625 (2006).
5. Khalil, A.S. et al. A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647-658 (2012).
6. Lohmueller, J.J., Armel, T.Z. & Silver, P.A. A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. *Nucleic acids research* **40**, 5180-5187 (2012).
7. Keung, A.J., Bashor, C.J., Kiriakov, S., Collins, J.J. & Khalil, A.S. Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. *Cell* **158**, 110-120 (2014).
8. Gaj, T., Gersbach, C.A. & Barbas, C.F., 3rd ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology* **31**, 397-405 (2013).
9. Kane, P.M. et al. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* **250**, 651-657 (1990).
10. Hirata, R. et al. Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of Saccharomyces cerevisiae. *The Journal of biological chemistry* **265**, 6726-6733 (1990).
11. Mootz, H.D. & Muir, T.W. Protein splicing triggered by a small molecule. *Journal of the American Chemical Society* **124**, 9044-9045 (2002).
12. Schwartz, E.C., Saez, L., Young, M.W. & Muir, T.W. Post-translational enzyme activation in an animal via optimized conditional protein splicing. *Nature chemical biology* **3**, 50-54 (2007).
13. Zeidler, M.P. et al. Temperature-sensitive control of protein activity by conditionally splicing inteins. *Nature biotechnology* **22**, 871-876 (2004).

14. Tyszkiewicz, A.B. & Muir, T.W. Activation of protein splicing with light in yeast. *Nature methods* **5**, 303-305 (2008).
15. Selgrade, D.F., Lohmueller, J.J., Lienert, F. & Silver, P.A. Protein scaffold-activated protein trans-splicing in mammalian cells. *Journal of the American Chemical Society* **135**, 7713-7719 (2013).
16. Huang, X. et al. Sequence-specific biosensors report drug-induced changes in epigenetic silencing in living cells. *DNA and cell biology* **31 Suppl 1**, S2-10 (2012).
17. Kim, J.S. & Pabo, C.O. Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 2812-2817 (1998).
18. Liu, Q., Segal, D.J., Ghiara, J.B. & Barbas, C.F., 3rd Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 5525-5530 (1997).
19. Nogami, S., Satow, Y., Ohya, Y. & Anraku, Y. Probing novel elements for protein splicing in the yeast Vma1 protozyme: A study of replacement mutagenesis and intragenic suppression. *Genetics* **147**, 73-85 (1997).
20. Ozawa, T., Nishitani, K., Sako, Y. & Umezawa, Y. A high-throughput screening of genes that encode proteins transported into the endoplasmic reticulum in mammalian cells. *Nucleic acids research* **33** (2005).
21. Sonntag, T. & Mootz, H.D. An intein-cassette integration approach used for the generation of a split TEV protease activated by conditional protein splicing. *Mol Biosyst* **7**, 2031-2039 (2011).
22. Chong, S.R. & Xu, M.Q. Protein splicing of the Saccharomyces cerevisiae VMA intein without the endonuclease motifs. *Journal of Biological Chemistry* **272**, 15587-15590 (1997).
23. ElrodErickson, M., Rould, M.A., Nekludova, L. & Pabo, C.O. Zif268 protein-DNA complex refined at 1.6 angstrom: A model system for understanding zinc finger-DNA interactions. *Structure* **4**, 1171-1180 (1996).
24. Cobb, L.M. et al. 2,4-dinitro-5-ethyleneiminobenzamide (CB 1954): a potent and selective inhibitor of the growth of the Walker carcinoma 256. *Biochemical pharmacology* **18**, 1519-1527 (1969).
25. Knox, R.J., Friedlos, F. & Boland, M.P. The bioactivation of CB 1954 and its use as a prodrug in antibody-directed enzyme prodrug therapy (ADEPT). *Cancer metastasis reviews* **12**, 195-212 (1993).
26. Grohmann, M. et al. A mammalianized synthetic nitroreductase gene for high-level expression. *BMC cancer* **9**, 301 (2009).
27. Lion, T. Adenovirus Infections in Immunocompetent and Immunocompromised Patients. *Clin Microbiol Rev* **27**, 441-462 (2014).
28. Bogdanove, A.J. & Voytas, D.F. TAL Effectors: Customizable Proteins for DNA Targeting. *Science* **333**, 1843-1846 (2011).
29. Filipovska, A. & Rackham, O. Designer RNA-binding proteins New tools for manipulating the transcriptome. *Rna Biol* **8**, 978-983 (2011).

30. Maxwell, I.H., Maxwell, F. & Glode, L.M. Regulated Expression of a Diphtheria-Toxin a-Chain Gene Transfected into Human-Cells - Possible Strategy for Inducing Cancer Cell Suicide. *Cancer Res* **46**, 4660-4664 (1986).
31. Yu, J., Zhang, L., Hwang, P.M., Kinzler, K.W. & Vogelstein, B. PUMA induces the rapid apoptosis of colorectal cancer cells. *Molecular cell* **7**, 673-682 (2001).
32. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science* **333**, 1307-1311 (2011).
33. Koch, J., Steinle, A., Watzl, C. & Mandelboim, O. Activating natural cytotoxicity receptors of natural killer cells in cancer and infection. *Trends Immunol* **34**, 182-191 (2013).
34. Banerjee, C. et al. BET bromodomain inhibition as a novel strategy for reactivation of HIV-1. *J Leukoc Biol* **92**, 1147-1154 (2012).
35. Nambiar, M., Kari, V. & Raghavan, S.C. Chromosomal translocations in cancer. *Biochim Biophys Acta* **1786**, 139-152 (2008).
36. Dekker, J., Marti-Renom, M.A. & Mirny, L.A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet* **14**, 390-403 (2013).

Figure Legends

Figure 1. A DNA sense-and-respond system. A toolbox of parts (left) comprising the system and a schematic (right) of the detection of a specific DNA sequence in the mammalian cell nucleus are shown. Two modular ZF-based sensors are programmed to dock adjacently upon target DNA. Each sensor contains an artificially split *Sce*VMA intein domain that is fused to the split extein. The extein is a ZF-TF, split between the binding and trans-activating domains. Target detection facilitates intein co-localization followed by extein trans-splicing. The spliced ZF-TF activates the expression of an interchangeable response gene downstream of its operator in the response circuit.

Figure 2. ZF-intein sensor design. Tridactyl and hexadactyl (generated by merging tridactyl units) sensors were fused to corresponding intein domains (lacking

exteins) and intracellular binding activity was tested. Operators for hexadactyl proteins consisted of two contiguous nine bp sites (0gap) or two nine bp sites interrupted with one (1gap) or two (2gap) unrelated bps, with GFP as a reporter. Mean GFP fluorescence (blue bars) of biological duplicate samples is displayed as overlaid bars and compared to control samples in the absence of ZF-TFs (grey bars).

Figure 3. Split extein and splice-junction (SJ) design and cis-splice validation. (**a**) Schematic: Extein residues from split luciferase introduced into ZF9-TF between its ZF and VP64 domains form splice junctions (SJ). Successful extein ligation would result in the ZF9-TF-V2 version (see also **Supplementary Fig. 5**). (**b**) Schematic: Split extein cis-splicing was tested in a condition-free manner by tethering the divided intein domains to one another with a flexible linker, applied to three extein variants (CIS V1-3), and measured through GFP trans-activation by the splice product, shown in (**c**) and **Supplementary Fig. 5c**. Biological duplicate samples are displayed as overlaid bars. (**d**) A α-HIS-probed immuno blot of proteins extracted from HEK 293FT cells expressing the three variants and splice-null, double mutants (CIS-V1-3m), (see also **Supplementary Fig. 5d,e,f,g** for mutation schematic, uncropped blot version of (**d**), α-FLAG probing, and quantification). Lane 1 is a ZF9-TF-V2 size control distinguishing between unspliced precursor and splice product. Total protein was normalized with α-Tubulin.

Figure 4. Intracellular DNA sense-and-respond. (**a**) The three components of the trans-splice platform: (*i*) expression plasmid encoding the sensor pair, (*ii*) target

plasmid carrying 0-8 instances of the 36 bp target sequence, (*iii*) response circuit modulated by spliced ZF9-TF. (**b**) Transfection of three sensor pair variants (V1, 2, 3) with (blue) and without (grey) target plasmid (see also **Supplementary Fig. 6a,b**). (**c**) Increasing amounts (ng) of transfected target plasmid yield a dose-response output. N-terminal sensor is absent in column 1. (**d**) Signal comparison between wildtype (*wt*) and mutant sensors using a series of target plasmids containing 0-4, 8 target instances. (**e**) Orthogonality testing for two sensor pairs and their target plasmids (T1, T2). Mean GFP fluorescence of biological triplicate samples (for **b,c,e**) and duplicate samples (for **d**) are displayed as overlaid bars.

Figure 5. Sequence recognition-induced apoptosis. (**a**) Schematic: target sequence recognition leads to NTR expression, CB 1954 reduction, and apoptosis. Transfection (50% rate, see **Supplementary Fig. 8a**, right) of sensor system with target (blue), without target (grey), or in the absence of the NTR response circuit (orange). Media contained [CB 1954] (0-32 μM). Samples collected after 72 h, stained with Annexin V, and measured via flow cytometry. Numbers represent % of cells staining positive for Annexin V (apoptosis), (see also **Supplementary Fig. 8a** for 48 h measurement). Biological duplicate samples are displayed as overlaid bars of matching color. (**b**) Phase-contrast microscopy 96 h post-treatment for cell population viability (see also **Supplementary Fig. 8b** for full prodrug range). Scale bar is 50 μM.

Figure 6. Adenovirus detection and reporting. (**a**) Schematic: Instances of targets 1 or 2 (T1, T2) (or no target) inserted into the Ad5 genome downstream of the packaging signal (ψ). Constitutive BFP expression serves as a viral infection marker. The detection system (mCherry as a transfection marker), based on sensor pair 1 or 2, is present in the cell nucleus prior to infection. Cells were infected with virus containing T1, T2, or no target (Vir (–T)). Sequence-specific detection was measured through GFP trans-activation. (**b**) HEK 293FT cells were transfected and infected as described above. Mean GFP fluorescence in $mCherry^+/BFP^+$ cells of biological duplicate samples is displayed as overlaid bars. (**c**) Epifluorescence microscopy of cells subjected to transfection → infection (mCherry = transfection, BFP = infection, and GFP = detection). Scale bar is 50 μM.

ONLINE METHODS

OPEN ZF library and split *Sce*VMA intein.

The 10 tridactyl ZFs used in this study were developed using OPEN, as previously described[5] (See also **Supplementary Note 1** and **Supplementary Table 2**). In all screening, VP64 was C-terminally fused. The *Sce*VMA intein was simultaneously amplified and divided from gDNA purified from *Saccharomyces cerevisiae*, as previously described[11]. I_N and I_C consist of residues 1-184 and 390-454, respectively, in the full *Sce*VMA sequence. Glycine/serine insolating linkers were introduced by gBlock synthesis (Integrated DNA Technologies (IDT)).

General cloning procedures.

Standard cloning procedures were applied (see also **Supplementary Note 3**). All PCR amplification stages used Phusion polymerase (NEB), according to the manufacturer's protocol. All restriction enzymes used throughout this study were Type II, NEB enzymes. Alkaline phosphatase (CIP) was used to prevent linear plasmid re-ligation to reduce background (except during the ligation of annealed oligos, described below). Bacterial selection of NEB 10-beta electro-competent cells during cloning was performed on 1% LB-agar plates with 120 ug/mL ampicillin or 50 ug/mL kanamycin.

Bacterial strains, over-expression, and EMSA

For cloning, 10-beta electrocompetent *E. coli* (NEB) cells were used, and for protein over-expression, MG1655 Pro was used. ZF-intein fusion proteins were expressed using the pZE21 plasmid, allowing for kanamycin selection and induction with 100 ng/ml anhydrotetracycline (aTc). ZF5-I_C and I_N-ZF5 fusion proteins were N-terminally FLAG-tagged, extracted with B-Per II (Thermo Scientific), purified with α-FLAG M2 affinity gel, and eluted with FLAG peptide (Sigma Aldrich). Purified protein was subjected to EMSA using biotinylated ZF5 DNA duplex (IDT), a streptavidin-HRP conjugate (Thermo Scientific, 21130), chemiluminescent substrate (Bio-Rad), and read with a Gel Logic system (Carestream Mol. Imaging).

Plasmid construction, oligo annealing and ligation (oligo drop)

The reporter constructs and synthetic response circuit were based on pGL4.26 (Promega). Luciferase was replaced with EGFP. pVITRO1/MCS/Neo (InvivoGen) was used to express ZF-TFs and sensor chimers in human cells. pBW121 served as a co-transfection marker with mCherry driven by pCAG. It also functioned as a target-carrier plasmid (with and w/o SV40 Ori). All instances in which ZF operator sequences and sensor target sequences were ligated into reporter plasmid or target-carrying plasmid, respectively, were performed via "oligo drop": single-stranded DNA oligos were designed to contain post-cut overhangs once annealed with their complimentary strand, compatible with ligation into a linearized plasmid. Lyophilized oligos were resuspended in annealing buffer (10mM Tris-HCl, pH 7.5, 50mM NaCl), equal volumes of sense and anti-sense were mixed at 100°, cooled to 24° over 60 min, diluted and ligated into linearized plasmid. Tandem repeats of operator or target sequences were directionally cloned in cycles, using oligos designed to contain a full 5' cut site within which would replace the abolished (by design) 5' ligation site at cloning cycle.

Cell culture, cell lines, and transfection.

The three cell lines used in this study are HeLa (ATCC CCL2), HEK 293FT (donated by the Weiss lab, MIT), and 293AD (Cell Biolabs). HeLa and HEK 293FT were selected for their convenience as transformed cell lines (also HEK 293FT expresses the SV40 large T-antigen allowing transient replication of plasmids containing the

SV40 ori). 293AD has enhanced adhering qualities for Ad5 production. Cells were not tested for mycoplasma contamination or authenticated. Cells were cultured in 4.5 g/L glucose, L-glutamine DMEM (Corning) supplemented with 10% FBS, penicillin 100 I.U./mL, and streptomycin 100 µg/mL. The standard transfection protocol was as follows, briefly: 100-200K cells were plated in 12-well format and transfected 24h later with 1-1.5 ug combined DNA using polyethylenimine (PEI) as a transfection reagent. 48h post-transfection, cells were examined using a Nikon Eclipse T*i* epifluorescence microscope, trypsinized, and prepared for flow cytometry analysis.

Mammalian protein extraction and immunoblots.

48h post-transfection, total protein was extracted using RIPA (Thermo Scientific) from HEK 293FT cells expressing cis-splice test proteins or ZF9-TF, tagged with FLAG or 3xHIS at the N- or C- terminus, respectively. Protein was loaded on a Mini-PROTEAN TGX gel, transferred to a 0.45 µm nitrocellulose membrane using a Trans-Blot SD cell and normalized via α-Tubulin (abcam). Membranes were probed with either α-FLAG-M2 (Sigma Aldrich, F3165) or α-HIS (GenScript, A00174-40) primary antibodies and secondary HRP-conjugated antibodies (Bio-Rad, #1721011 and #1706515), and imaged with a Gel Logic system (Carestream Mol. Imaging).

Sequence recognition-induced apoptosis via CB 1954/NTR. CB 1954 (Sigma Aldrich) was resuspended in DMSO to 0.2M and a working stock [500µM] was prepared by dilution into DMEM before each experiment. Cells were transfected as

described above and supplemented with 0-32 [μM] prodrug. After 48h or 72h, cells were collected, including growth media, and subjected to Annexin V-FITC and PI staining according to manufacturer's instructions (BD Biosciences), followed by flow cytometry. Cells were also analyzed via light microscopy at 96h.

Adenovirus preparation, titration, and infection. Ad5 was prepared using the RAPAd CMV adenoviral expression system (Cell Biolabs). Briefly, TagBFP was cloned downstream of a CMV promoter as an infection marker. Target sequence (0-8 copies) for sensor pair 1 or 2 was ligated into the viral genome adjacent to the Ad5 ψ. Crude viral extract was prepared via co-transfection of the linearized viral plasmids into a 293AD line (Cell Biolabs), and after collection, was amplified in the same cells and titrated through serial dilution and plaque counting. For transfection/infection experiments, HEK 293FT cells were transfected with the sensor system as described above and after 24h, infected with virus at an MOI of 5. 48h later, cells were trypsonized, fixed and subjected to flow cytometry analysis.

Flow cytometry and data analysis. All transfections were measured with a BD FACSAria II flow cytometer (BD Biosciences). Transfection/infections were measured using a BD Fortessa High Throughput Sampler. Means of fluorescence distributions were calculated with FlowJo. In transfections, GFP arbitrary units were collected from cells gated to mCherry and in transfection/infection experiments, gated to mCherry and BFP fluorescence. All experiments were repeated multiple times and data are displayed either as an average of biological triplicates or

duplicates with standard deviation, or as fold change with error propagation calculations. Due to experimental variation, while data trends may be compared between experiments, absolute values should be compared within single experiments.

Supplementary Notes

Note 1. Tridactyl ZF screening.

At the onset of this work, we utilized ten tridactyl ZFs designed using OPEN (*O*ligomerized *P*ool *EN*gineering) and screened with bacterial 2-hybrid methods. In order to select two members to serve as DNA sensors and another to function as the split extein, the ZFs were fused to VP64 and transformed into synthetic TFs (**Supplementary Fig. 1a**). Reporters were prepared for each ZF-TF encoding GFP modulated by a single copy of the respective nine bp binding site. We transiently co-transfected human embryonic kidney (HEK 293FT) or HeLa cells with each ZF-TF and its reporter, assessing intracellular binding by measuring GFP fluorescence via flow cytometry (**Supplementary Fig. 1b,c**). While a number of ZF-TFs showed marked activation, several displayed fluorescence only moderately higher than their negative controls. Also, a purified tridactyl I_N-ZF fusion protein subjected to a gel retardation assay could not bind its operator to any detectable measure (**Supplementary Fig. 1d**). Therefore, hexadactyl ZF proteins were constructed; to both acquire higher affinity and overcome binding hindrance imposed by fusion to the intein domains, as described in the main text (**Fig. 2**).

Note 2. Non-replicative vs. replicative target.

A replicating target-carrying plasmid may more accurately represent an active viral infection compared to a non-replicative target plasmid. Therefore, to assess whether the system could distinguish between these two options, the Simian vacuolating virus 40

origin of replication (SV40 ori) was inserted into the target plasmid and tested in HEK 293FT cells (capable of replicating DNA containing this sequence). Epifluorescence microscopy revealed successful target recognition as GFP-positive cells were visible only in the presence of the target and even markedly more so, in that of its replicative form (**Supplementary Fig. 7**).

Note 3. Plasmids, oligos, and component sequences

Plasmid sequences have been deposited to addgene.com and accession numbers have been provided in Supplementary Table 1. Additional sequences of individual components, e.g. - ZFs and their operator sites, *Sce*VMA intein domains and the primers used to amplify them, linker sequences merging tridactyl ZFs to create hexadactyl sensors, Glycine/Serine-rich insulating linkers distancing the intein domains from the ZF domains, target sites of all gap sizes for both sensor pairs, localization and immuno blot tags, genes encoding fluorescent proteins, gBlocks used to create a second sensor pair, gBlocks to introduce R→A mutations into ZFs, splice junction linkers- can be found in the Supplementary Tables 2-4.

Note 4. Equipment and settings.

Images of blots shown in **Fig. 3** and **Supplementary Fig. 1,5**, were acquired using the Gel Logic 6000 Pro system (Carestream Mol. Imaging) with the manufacturer's image acquisition software (Carestream MI) for chemiluminescent imaging. Microscopy images in **Fig. 5**,**6** and **Supplementary Fig. 4,7,8** were acquired with the following equipment: Nikon Eclipse T*i* epifluorescence microscope, Nikon Intensilight C-HGFIE, Sutter

Instrument *Smart*shutter controller, Photometrics CoolSnap HQ^2, Nikon TI-PS 100W, Prior ProScan II. DAPI, TxRED, FGP(R)BP, ANALY filters. A Plan Fluor 10×/0.30 objective was used for all images, at a pixel dimension of 1392×1040 and resolution of 72 pixels/inch. Images were analyzed using the manufacturer's NIS-Elements software.

Supplementary Tables

Supplementary Table 1

Name	Description	Resistance	Marker	Addgene accession number
pVITRO1-SS-132	ZF1-TF + mCherry	kan	G418	68729
pVITRO1-SS-115	ZF2-TF + mCherry	kan	G418	68730
pVITRO1-SS-126	ZF3-TF + mCherry	kan	G418	68731
pVITRO1-SS-117	ZF4-TF + mCherry	kan	G418	68732
pVITRO1-SS-127	ZF5-TF + mCherry	kan	G418	68733
pVITRO1-SS-128	ZF6-TF + mCherry	kan	G418	68734
pVITRO1-SS-130	ZF7-TF + mCherry	kan	G418	68735
pVITRO1-SS-131	ZF8-TF + mCherry	kan	G418	68736
pVITRO1-SS-113	ZF9-TF + mCherry	kan	G418	68737
pVITRO1-SS-129	ZF10-TF + mCherry	kan	G418	68738
pGL4.26-SS-140	ZF1 Op - GFP	Amp	Hygro	68739
pGL4.26-SS-79	ZF2 Op - GFP	Amp	Hygro	68740
pGL4.26-SS-134	ZF3 Op - GFP	Amp	Hygro	68741
pGL4.26-SS-80	ZF4 Op - GFP	Amp	Hygro	68742
pGL4.26-SS-135	ZF5 Op - GFP	Amp	Hygro	68743
pGL4.26-SS-136	ZF6 Op - GFP	Amp	Hygro	68744
pGL4.26-SS-138	ZF7 Op - GFP	Amp	Hygro	68745
pGL4.26-SS-139	ZF8 Op - GFP	Amp	Hygro	68746
pGL4.26-SS-78	ZF9 Op - GFP	Amp	Hygro	68747
pGL4.26-SS-137	ZF10 Op - GFP	Amp	Hygro	68748
pVITRO1-SS-148	ZF8/ZF7-TF + mCherry	kan	G418	68749
pVITRO1-SS-149	ZF6/ZF10-TF + mCherry	kan	G418	68750
pGL4.26-SS-160	ZF10-ZF6 Op 0gap - GFP	Amp	Hygro	68751

pGL4.26-SS-162	ZF10-ZF6 Op 1gap - GFP	Amp	Hygro	68752
pGL4.26-SS-164	ZF10-ZF6 Op 2gap - GFP	Amp	Hygro	68753
pGL4.26-SS-161	ZF7-ZF8 Op 0gap - GFP	Amp	Hygro	68754
pGL4.26-SS-163	ZF7-ZF8 Op 1gap - GFP	Amp	Hygro	68755
pGL4.26-SS-165	ZF7-ZF8 Op 2gap - GFP	Amp	Hygro	68756
pGL4.26-SS-190	ZF9 Op x2 - GFP	Amp	Hygro	68757
pGL4.26-SS-191	ZF9 Op x4 - GFP	Amp	Hygro	68758
pGL4.26-SS-192	ZF9 Op x6 - GFP	Amp	Hygro	68759
pVITRO1-SS-246	ZF9-TF-V1 + mCherry	kan	G418	68760
pVITRO1-SS-247	ZF9-TF-V2 + mCherry	kan	G418	68761
pVITRO1-SS-248	ZF9-TF-V3 + mCherry	kan	G418	68762
pBW121-SS-273	ZF9-TF-V2 - HIS	Amp	(-)	68763
pBW121-SS-276	CIS1-V1 - HIS	Amp	(-)	68764
pBW121-SS-277	CIS-V1m - HIS	Amp	(-)	68765
pBW121-SS-278	CIS-V2 - HIS	Amp	(-)	68766
pBW121-SS-279	CIS-V2m - HIS	Amp	(-)	68767
pBW121-SS-280	CIS-V3 - HIS	Amp	(-)	68768
pBW121-SS-281	CIS-V3m - HIS	Amp	(-)	68769
pVITRO1-SS-267	Sensor pair 1 V1	kan	G418	68770
pVITRO1-SS-269	Sensor pair 1 V2	kan	G418	68771
pVITRO1-SS-271	Sensor pair 1 V3	kan	G418	68772
pVITRO1-SS-255	C-Terminal sensor 1 V1	kan	G418	68773
pVITRO1-SS-257	C-Terminal sensor 1 V2	kan	G418	68774
pVITRO1-SS-259	C-Terminal sensor 1 V3	kan	G418	68775
pBW121-SS-282	(-) Target no SV40 ori	Amp	(-)	68776
pBW121-SS-309	8x 0gap Target 1 no Sv40 Ori	Amp	(-)	68777
pBW121-SS-310	8x 4gap Target 1 no Sv40 Ori	Amp	(-)	68778
pBW121-SS-287	8x 8gap Target 1 no Sv40 Ori	Amp	(-)	68779
pBW121-SS-311	8x 12gap Target 1 no Sv40 Ori	Amp	(-)	68780
pBW121-SS-289	1x 0gap Target 1 no Sv40 Ori	Amp	(-)	68781
pBW121-SS-294	2x 0gap Target 1 no Sv40 Ori	Amp	(-)	68782
pBW121-SS-299	3x 0gap Target 1 no Sv40 Ori	Amp	(-)	68783
pBW121-SS-301	4x 0gap Target 1 no Sv40 Ori	Amp	(-)	68784
pVITRO1-SS-345	Sensor pair 1 V2 R-->A + mCherry	kan	G418	68785
pBW121-SS-315	8x 0gap Target 1 with Sv40 Ori	Amp	(-)	68786
pBW121-SS-321	8x 0gap Target 2 with Sv40 Ori	Amp	(-)	68787
pVITRO1-SS-288	N-Terminal sensor 1 V2	kan	G418	68788

pVITRO1-SS-323	Sensor pair 2 V2	kan	G418	68789
pGL4.26-SS-358	ZF9 Op x6 - NTR	Amp	hygro	68790
pGL4.26-SS-352	ZF9 Op x6 - GFP-IRES-NTR	Amp	hygro	68791
pacAd5 CMVK-NpA-SS-324	pShuttle TagBFP	blast	blast	68792
pacAd5 CMVK-NpA-SS-328	pShuttle TagBFP 8x 0gap Target 1	blast	blast	68793
pacAd5 CMVK-NpA-SS-332	pShuttle TagBFP 8x 0gap Target 2	blast	blast	68794
pVITRO1-SS-338	Sensor pair 1 V2, ZF9 Op x6 - GFP, + mCherry	kan	G418	68795

Supplementary Table 2

ZF number	previous name	Binding site 5'-3'	ZF sequence
ZF1	172-5	aGGAGGGGCTc	GTCTCTAGACCCGGGGAGCGCCCCTTCCAGTGTCGCATTTGCATGCGGAACTTTTCGATGAAAAATACTTTGACTAGACATACCCGTACTCATACCGGTGAAAAACCGTTTCAGTGTCGGATCTGTATGCGAAATTTCTCATCTACGTACGCACACCGGCGAGAAGCCATTCCAATGCCGAATATGCAATGCCGAATATGCATGCGCAACTTCAGTCAAAAACCACATTTGTCAAGACACCTAAAAACCCACCTGAGAGGATCC
ZF2	62-1	gGCCGAAGATa	GTCTCTAGACCCGGGGAGCGCCCCTTCCAGTGTCGCATTTGCATGCGGAACTTTTCGACTGGTCAAAGATTGAGAATTCATACCCGTACTCATACCGGTGAAAAACCGTTTCAGTGTCGGATCTGTATGCGAAATTTCTCCCAAAATCAAAATTTGGCTAGACATCTACGTACGCACACCGGCGAGAAGCCATTCCAATGCCGAATATGCATGCGCAACTTCAGTGATAAATCTGTTTTGGCTAGACACCTAAAAACCCACCTGAGAGGATCC
ZF3	43-8	aGAGTGAGGAc	GTCTCTAGACCCGGGGAGCGCCCCTTCCAGTGTCGCATTTGCATGCGGAACTTTTCGCGCCAGGACAGGCTTGACAGGCATACCCGTACTCATACCGGTGAAAAACCGTTTCAGTGTCGGATCTGTATGCGAAATTTCTCCCAGAAGGAGCACTTGGCGGGGCATCTACGTACGCACACCGGCGAGAAGCCATTCCAATGCCGAATATGCATGCGCAACTTCAGTCGCCGCGACAACCTGAACCGGCACCTAAAAACCCACCTGAGAGGATCC
ZF4	129-3	cGGGGACGTCa	GTCTCTAGACCCGGGGAGCGCCCCTTCCAGTGTCGCATTTGCATGCGGAACTTTTCGACTGCTGCTGTTTTGACTAGACATACCCGTACTCATACCGGTGAAAAACCGTTTCAGTGTCGGATCTGTATGCGAAATTTCTCCGATAGAGCTAATTTGACTAGACATCTACGTACGCACACCGGCGAGAAGCCATTCCAATGCCGAATATGCATGCGCAACTTCAGTAGAATTGATAAATTGGGTGATCACCTAAAAACCCACCTGAGAGGATCC
ZF5	150-4	gGTGTAGGGGt	GTCTCTAGACCCGGGGAGCGCCCCTTCCAGTGTCGCATTTGCATGCGGAACTTTTCGAAAGGTGAAAGATTGGTTAGACATACCCGTACTCATACCGGTGAAAAACCGTTTCAGTGTCGGATCTGTATGCGAAATTTCTCCAGAATGGATAATTTGTCTACTCATCTACGTACGCACACCGGCGAGAAGCCATTCCAATGCCGAATATGCATGCGCAACTTCAGTAGAAAAGATGCTTTGAATAGACACCTAAAAACCCACCTGAGAGGATCC
ZF6	151-1	tGCAGGAGGTg	GTCTCTAGACCCGGGGAGCGCCCCTTCCAGTGTCGCATTTGCATGCGGAACTTTTCGATTCCAAATCATTTGGCTAGACATACCCGTACTCATACCGGTGAAAAACCGTTTCAGTGTCGGATCTGTATGCGAAATTTCTCCCAATCTGCTCATTTGAAAAGACATCTACGTACGCACACCGGCGAGAAGCCATTCCAATGCCGAATATGCATGCGCAACTTCAGTCAAGATGTTTCTTTGGTTAGACACCTAAAAACCCACCTGAGAGGATCC
ZF7	63-4	aGCTGGAGGGt	GTCTCTAGACCCGGGGAGCGCCCCTTCCAGTGTCGCATTTGCATGCGGAACTTTTCGAAAAAAGATCATTTGCATAGACATACCCGTACTCATACCGGTGAAAAACCGTTTCAGTGTCGGATCTGTATGCGAAATTTCTCCCAAAGACCACATTTGACTAATCATCTACGTACGCACACCGGCGAGAAGCCATTCCAATGCCGAATATGCATGCGCAACTTCAGTGTTGGTGCTTCTTTGAAAAGACACCTAAAAACCCACCTGAGAGGATCC
ZF8	128-2	cGAAGTGGTCc	GTCTCTAGACCCGGGGAGCGCCCCTTCCAGTGTCGCATTTGCATGCGGAACTTTTCGACTATGGCTGTTTTGAGAAGACATACCCGTACTCATACCGGTGAAAAACCGTTTCAGTGTCGGATCTGTATGCGAAATTTCTCCAGAAGAGAAGTTTTGGAAAATCATCTACGTACGCACACCGGCGAGAAGCCATTCCAATGCCGAATATGCATGCGCAACTTCAGTCAAACTGTTAATTTGGATAGACACCTAAAAACCCACCTGAGAGGATCC
ZF9	158-2	tGTAGATGGAg	GTCTCTAGACCCGGGGAGCGCCCCTTCCAGTGTCG

			CATTTGCATGCGGAACTTTTCGGATAAAACTAAAT TGAGAGTTCATACCCGTACTCATACCGGTGAAAAA CCGTTTCAGTGTCGGATCTGTATGCGAAATTTCTC CGTTAGACATAATTTGACTAGACATCTACGTACGC ACACCGGCGAGAAGCCATTCCAATGCCGAATATG CATGCGCAACTTCAGTCAATCTACTTCTTTGCAAAG ACACCTAAAAACCCACCTGAGAGGATCC
ZF10	159-3	tGAAGAAGCTg	GTCTCTAGACCCGGGGAGCGCCCCTTCCAGTGTCG CATTTGCATGCGGAACTTTTCGTCTGCTCAAGCTTT GGCTAGACATACCCGTACTCATACCGGTGAAAAAC CGTTTCAGTGTCGGATCTGTATGCGAAATTTCTCC CAAGGTGGTAATTTGACTAGACATCTACGTACGCA CACCGGCGAGAAGCCATTCCAATGCCGAATATGC ATGCGCAACTTCAGTCAACATCCAAATTTGACTAG ACACCTAAAAACCCACCTGAGAGGATCC
N-ZF8/ZF7-C		0 gap- aGCTGGAGGGGAAGTGGTCc	GTCTCTAGACCCGGGGAGCGCCCCTTCCAGTGT CGCATTTGCATGCGGAACTTTTCGACTATGGCT GTTTTGAGAAGACATACCCGTACTCATACCGG TGAAAAACCGTTTCAGTGTCGGATCTGTATGC GAAATTTCTCCAGAAGAGAAGTTTTGGAAAAT CATCTACGTACGCACACCGGCGAGAAGCCATTC CAATGCCGAATATGCATGCGCAACTTCAGTCA AACTGTTAATTTGGATAGACACCTAAAAACCC ACCTGAGACAGAAGGACGGGGAGCGCCCCTTC CAGTGTCGCATTTGCATGCGGAACTTTTCGAA AAAAGATCATTTGCATAGACATACCCGTACTC ATACCGGTGAAAAACCGTTTCAGTGTCGGATC TGTATGCGAAATTTCTCCCAAAGACCACATTT GACTAATCATCTACGTACGCACACCGGCGAGA AGCCATTCCAATGCCGAATATGCATGCGCAAC TTCAGTGTTGGTGCTTCTTTGAAAAGACACCT AAAAACCCACCTGAGAGGATCC
		1 gap- aGCTGGAGGGtGAAGTGGTCc	
		2 gap- aGCTGGAGGGtcGAAGTGGTCc	
N-ZF6/ZF10-C		0 gap- tGAAGAAGCTGCAGGAGGTg	GTCTCTAGACCCGGGGAGCGCCCCTTCCAGTGT CGCATTTGCATGCGGAACTTTTCGATTCCAAA TCATTTGGCTAGACATACCCGTACTCATACCGG TGAAAAACCGTTTCAGTGTCGGATCTGTATGC GAAATTTCTCCCAATCTGCTCATTTGAAAAGA CATCTACGTACGCACACCGGCGAGAAGCCATTC CAATGCCGAATATGCATGCGCAACTTCAGTCA AGATGTTTCTTTGGTTAGACACCTAAAAACCC ACCTGAGACAGAAGGACGGGGAGCGCCCCTTC CAGTGTCGCATTTGCATGCGGAACTTTTCGTCT GCTCAAGCTTTGGCTAGACATACCCGTACTCAT ACCGGTGAAAAACCGTTTCAGTGTCGGATCTG TATGCGAAATTTCTCCCAAGGTGGTAATTTGA CTAGACATCTACGTACGCACACCGGCGAGAAG CCATTCCAATGCCGAATATGCATGCGCAACTTC AGTCAACATCCAAATTTGACTAGACACCTAAA AACCCACCTGAGAGGATCC
		1 gap- tGAAGAAGCTgGCAGGAGGTg	
		2 gap- tGAAGAAGCTgtGCAGGAGGTg	
N-ZF4/ZF5-C			GTGTCCAGACCTGGGGAGAGGCCATTCCAGTGTC GGATCTGTATGAGAAATTTTAGCAAAGGTGAAAG ATTGGTTAGACATACAAGAACCCATACAGGCGAG AAGCCCTTTCAGTGCAGGATCTGCATGCGAAACTT TTCTAGAATGGATAATTTGTCCACTCACCTGCGGA CTCATACCGGCGAAAAACCATTTCAGTGTCGAATT TGCATGCGCAACTTTAGCAGAAAAGATGCTTTGAA TAGACACCTTAAGACTCATCTGAGGCAGAAGGAC GGGGAGAGACCTTTCCAGTGCAGAATCTGCATGA GAAACTTTTCTACTGCTGCTGTTTTGACTAGACATA CCCGCACCCACACCGGCGAAAAACCCTTCCAGTGC CGCATCTGTATGCGAAATTTTAGCGATAGAGCTAA TTTGACTAGACACCTTAGAACACACACTGGAGAAA AACCATTTCAGTGTCGGATCTGCATGCGCAACTTC TCCAGAATTGATAAATTGGGTGATCACCTGAAAAC CCACCTGAGGGGATCT
N-ZF2/ZF3-C			ACACGTCCCGGGGAGCGCCCCTTCCAGTGCAGGA TCTGCATGAGGAATTTTAGCACTGGTCAAAGATTG AGAATTCACACACGGACTCACACAGGGGAGAAGC CCTTCCAGTGTAGAATTTGCATGCGGAACTTCAGC CAAAATCAAAATTTGGCTAGACACCTGCGGACCCA TACTGGCGAAAAGCCCTTTCAGTGCCGGATCTGCA TGCGCAACTTCTCCGATAAATCTGTTTTGGCTAGA CACCTGAAGACACATCTTCGGCAGAAAGATGGAG AGAGACCATTCCAGTGTCGGATTTGTATGAGGAA CTTTAGTCGCCAGGACAGGCTTGACAGGCATACCC GGACCCACACAGGTGAGAAGCCATTCCAGTGTCG CATCTGCATGCGGAATTTCTCCCAGAAGGAGCACT TGGCGGGGCACCTGCGGACTCATACAGGGGAGA AACCTTTTCAGTGTAGGATCTGCATGCGCAATTTTT CACGCCGCGACAACCTGAACCGGCACCTGAAAAC CCACCTGAGAGGATCC

6-Finger sensor pair 1		**Target site**	
ZF6/ZF10 + ZF8/ZF7		0 gap- tGAAGAAGCTGCAGGAGGTGCTGGAGGGGAAGTGGTCc	
		4 gap- tGAAGAAGCTGCAGGAGGTgtcaGCTGGAGGGGAAGTGGTCc	
		8 gap- tGAAGAAGCTGCAGGAGGTgtcagtcaGCTGGAGGGGAAGTGGTCc	
		12 gap- tGAAGAAGCTGCAGGAGGTgtcagtcagtcaGCTGGAGGGGAAGTGGTCc	
		Example of 8x 0 gap Target site	
		TGAAGAAGCTGCAGGAGGTGCTGGAGGGGAAGTGGTCCGGATCTTGAAGAAGCTGCAGGAGGTGCTGGAGGG GAAGTGGTCCGGATCTTGAAGAAGCTGCAGGAGGTGCTGGAGGGGAAGTGGTCCGGATCTTGAAGAAGCTGCA GGAGGTGCTGGAGGGGAAGTGGTCCGGATCTTGAAGAAGCTGCAGGAGGTGCTGGAGGGGAAGTGGTCCGGA TCTTGAAGAAGCTGCAGGAGGTGCTGGAGGGGAAGTGGTCCGGATCTTGAAGAAGCTGCAGGAGGTGCTGGAG GGGAAGTGGTCCGGATCTTGAAGAAGCTGCAGGAGGTGCTGGAGGGGAAGTGGTCC	
6-Finger sensor pair 2			
ZF2/ZF3 + ZF5/ZF4		0 gap- aGAGTGAGGAGCCGAAGATGGGGACGTCGTGTAGGGGt	

Supplementary Table 3

name of sequence	sequence 5'-3'	comment
FLAG/NLS	ATGGACTACAAAGACGATGACGACAAGCCTAAGAAGAAACGGAAG	N-terminal FLAG tag for immunoblot and/or purification and NLS (nulcear localization signal) for guiding protein to nucleus
6xHIS Tag	GGATCTGGCCACCATCACCATCACCAT	Short GSG linker followed by 6xHIS
extended linker merging 2x 3-finger ZFs	CGGCAGAAAGATGGAGAGAGACCA	Extended version of the canonical linker. Used to merge two 3-finger ZFs into one 6-finger ZF. The canonical linker found between fingers is :..(HT)**TGEKP**(FQC)... The extended linker used to merge two 3-finger proteins is: ..(HL)**RQKDGERP**(FQC)...
1x Glyince/Serine-rich linker	GGGGGCGGATCAGGTGGCGGATCAGGAGGCGGATCC	situated between intein domains and ZF sensor domains. The 2x and 3x versions of the linker may contain variations in nucleic acid sequence (via codon shuffling) to allow cloning
VP64	GGGCGCGCCGACGCGCTGGACGATTTCGATCTCGACATGCTGGGTTCTGATGCCCTCGATGACTTTGACCTGGATATGTTGGGAAGCGACGCATTGGATGACTTTGA TCTGGACATGCTCGGCTCCGATGCTCTGGACGATTTCGATCTCGATATGTTAATTAATTGT	
mCherry	ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAG GGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCA TGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGAC GGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCA GAAGAAGACCATGGGCTGGCAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCC ACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGAC TACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAG	used as transfection marker
eGFP	ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAG GGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGC TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTAC AAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTG GAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGC AGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAAC GAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG	used as response gene
TagBFP	ATGAGCGAGCTGATTAAGGAGAACATGCACATGAAGCTGTACATGGAGGGCACCGTGGACAACCATCACTTCAAGTGCACATCCGAGGGCGAAGGCAAGCCCTACG AGGGCACCCAGACCATGAGAATCAAGGTGGTCGAGGGCGGCCCTCTCCCCTTCGCCTTCGACATCCTGGCTACTAGCTTCCTCTACGGCAGCAAGACCTTCATCAACC ACACCCAGGGCATCCCCGACTTCTTCAAGCAGTCCTTCCCTGAGGGCTTCACATGGGAGAGAGTCACCACATACGAAGACGGGGGCGTGCTGACCGCTACCCAGGAC ACCAGCCTCCAGGACGGCTGCCTCATCTACAACGTCAAGATCAGAGGGGTGAACTTCACATCCAACGGCCCTGTGATGCAGAAGAAAACACTCGGCTGGGAGGCCTT CACCGAGACGCTGTACCCCGCTGACGGCGGCCTGGAAGGCAGAAACGACATGGCCCTGAAGCTCGTGGGCGGGAGCCATCTGATCGCAAACATCAAGACCACATAT AGATCCAAGAAACCCGCTAAGAACCTCAAGATGCCTGGCGTCTACTATGTGGACTACAGACTGGAAAGAATCAAGGAGGCCAACAACGAGACCTACGTCGAGCAGC ACGAGGTGGCAGTGGCCAGATACTGCGACCTCCCTAGCAAACTGGGGCACAAGCTCAAT	used as Adenovirus infection marker
SV40 ORI	GGATCCGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCG AGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTAACTTGTTTATTGCAGCTTATAATGGTTACAA ATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCC	
NTR (mammalian codon optimized)	ATGGACATCATCAGCGTGGCTCTGAAGAGGCACTCCACCAAGGCTTTCGACGCTTCCAAGAAACTGACCCCTGAACAGGCCGAGCAGATCAAGACCCTGCTCCAGTA CAGCCCTAGCTCCACCAACAGCCAGCCTTGGCACTTCATCGTGGCTAGCACCGAGGAAGGCAAAGCTAGGGTGGCTAAGAGCGCCGCTGGCAACTACGTGTTCAACG AGAGGAAGATGCTGGATGCTAGCCACGTGGTGGTGTTCTGCGCTAAGACCGCCATGGACGATGTGTGGCTGAAGCTGGTGGTGGATCAGGAAGATGCTGATGGCA GGTTCGCTACCCCTGAAGCTAAGGCCGCTAACGACAAGGGCAGGAAGTTCTTCGCCGACATGCACAGGAAGGATCTGCACGATGATGCTGAGTGGATGGCCAAGCA GGTGTACCTGAACGTGGGCAACTTCCTGCTCGGCGTGGCTGCCCTGGGCCTCGATGCTGTGCCCATCGAAGGCTTCGATGCTGCTATCCTGGATGCCGAGTTCGGCCT GAAGGAGAAAGGCTACACCAGCCTGGTGGTGGTGCCTGTGGGCCACCACAGCGTGGAGGACTTCAACGCTACCCTGCCTAAGAGCAGGCTGCCCCAGAACATCACC CTGACCGAGGTGTGA	**A mammalianized synthetic nitroreductase gene for high-level expression** Maik Grohmann, Nils Paulmann, Sebastian Fleischhauer, Jakob Vowinckel, Josef Priller and Diego J Walther*
1x ZF9	TGTAGATGGAGAAGCTTAGACACTAGAGGGTATATAATGGAAGCTCGACTTCCAGCTTGGCAATCCGGTACTGTTGGTAAAGCCACCATG	ZF9 Operator. Minimal

Operator site in reporter construct		promotor. GFP Start Codon. Reporter plasmids used for screening 3-finger ZFs and 6-finger ZFs were constructed in the same manner, with their specific binding site installed.
6x ZF9 Operator site in reporter construct	TGTAGATGGAGAAATGTAGATGGAGAAATGTAGATGGAGAAATGTAGATGGAGAAATGTAGATGGAGAAGCTTAGACACTAGAGGGTATAT AATGGAAGCTCGACTTCCAGCTTGGCAATCCGGTACTGTTGGTAAAGCCACCATG	ZF9 Operator. Minimal promotor. GFP Start Codon.
FMDV IRES	AGCAGGTTTCCCCAATGACACAAAACGTGCAACTTGAAACTCCGCCTGGTCTTTCCAGGTCTAGAGGGGTAACACTTTGTACTGCGTTTGGCTCCACGCTCGATCCACT GGCGAGTGTTAGTAACAGCACTGTTGCTTCGTAGCGGAGCATGACGGCCGTGGGAACTCCTCCTTGGTAACAAGGACCCACGGGGCCAAAAGCCACGCCCACACGG GCCCGTCATGTGTGCAACCCCAGCACGGCGACTTTACTGCGAAACCCACTTTAAAGTGACATTGAAACTGGTACCCACACACTGGTGACAGGCTAAGGATGCCCTTCA GGTACCCCGAGGTAACACGCGACACTCGGGATCTGAGAAGGGGACTGGGGCTTCTATAAAAGCGCTCGGTTTAAAAAGCTTCTATGCCTGAATAGGTGACCGGAGG TCGGCACCTTTCCTTTGCAATTACTGACCCTATGAATACA	used in the single plasmid construct to add mCherry transfection marker
Primers used to amplify SceVMA intein from yeast gDNA	GTTTTGCTTAACGTTCTTTCG	Forward primer for C-terminal split Intein
	ATTATGGACGACAACCTGG	Reverse primer for C-terminal split Intein
	TGCTTTGCCAAGGGTACC	Forward primer for N-terminal split Intein
	ataaagAATTGGAGCGTAAGTC	Reverse primer for N-terminal split Intein (lower case nts are part of original linker)
SceVMA N-Terminal portion	TGCTTTGCCAAGGGTACCAATGTTTTAATGGCGGATGGGTCTATTGAATGTATTGAAAACATTGAGGTTGGTAATAAGGTCATGGGTAAAGATGGCAGACCTCGTGA GGTAATTAAATTGCCCAGAGGAAGAGAAACTATGTACAGCGTCGTGCAGAAAAGTCAGCACAGAGCCCACAAAAGTGACTCAAGTCGTGAAGTGCCAGAATTACTC AAGTTTACGTGTAATGCGACCCATGAGTTGGTTGTTAGAACACCTCGTAGTGTCCGCCGTTTGTCTCGTACCATTAAGGGTGTCGAATATTTTGAAGTTATTACTTTTG AGATGGGCCAAAAGAAAGCCCCCGACGGTAGAATTGTTGAGCTTGTCAAGGAAGTTTCAAAGAGCTACCCAATATCTGAGGGGCCTGAGAGAGCCAACGAATTAGT AGAATCCTATAGAAAGGCTTCAAATAAAGCTTATTTTGAGTGGACTATTGAGGCCAGAGATCTTTCTCTGTTGGGTTCCCATGTTCGTAAAGCTACCTACCAGACTTAC GCTCCAATT	
SceVMA C-Terminal portion	GTTTTGCTTAACGTTCTTTCGAAGTGTGCCGGCTCTAAAAAATTCAGGCCTGCTCCCGCCGCTGCTTTTGCACGTGAGTGCCGCGGATTTTATTTCGAGTTACAAGAATT GAAGGAAGACGATTATTATGGGATTACTTTATCTGATGATTCTGATCATCAGTTTTTGCTTGCCAACCAGGTTGTCGTCCATAAT	
gBlocks used to create R-->A mutant sensors		
C-terminal sensor R-->A mutant	GATATCAAGGATCCATGGACTACAAAGACGATGACGACAAGCCTAAGAAGAAACGGAAGGTGTCCAGACCTGGGGAGgccCCATTCCAGTGTCGGATCTGT ATggcTAATTTTAGCACTATGGCTGTACTGAGGAGACATACAAGAACCCATACAGGCGAGAAGCCCTTTCAGTGCAGGATCTGCATGgcaAACTTTTCTCG CAGAGAAGTGCTTGAAAACCACCTGCGGACTCATACCGGCGAAAAACCATTTCAGTGTCGAATTTGCATGgccAACTTTAGCCAGACAGTGAACTTGGATA GACACCTTAAGACTCATCTGAGGCAGAAGGACGGGGAGgctCCTTTCCAGTGCAGAATCTGCATGgcaAACTTTTCTAAGAAAGATCATCTGCACCGCCATA CCCGCACCCACACCGGCGAAAAACCCTTCCAGTGCCGCATCTGTATGGCCAATTTTAGCCAGAGGCCACATCTTACAAATCACCTTAGAACACACACTGGA GAAAAACCATTTCAGTGTCGGATCTGCATGgctAACTTCTCCGTGGGCGCCAGCCTCAAACGCCACCTGAAAACCCACCTGAGG	EcoRV and Bsu36I were used for cloning. Lower-case orange codons are those that were modified in order to facilitate an R-->A point mutation.
N-terminal sensor R-->A mutant	CACGTCCCGGGGAGgccCCCTTCCAGTGCAGGATCTGCATGgctAATTTTAGCATTCCAAACCACCTTGCCCGGCACACACGGACTCACACAGGGGAGAAGC CCTTCCAGTGTAGAATTTGCATGgcaAACTTCAGCCAGTCAGCCCACTTGAAAAGACACCTGCGGACCCATACTGGCGAAAAGCCCTTTCAGTGCCGGATC TGCATGgccAACTTCTCCCAGGACGTTTCCCTTGTCAGGCACCTGAAGACACATCTTCGGCAGAAAGATGGAGAGgctCCATTCCAGTGTCGGATTTGTATG gcaAACTTTAGTTCTGCACAGGCTTTGGCTAGGCATACCCGGACCCACACAGGTGAGAAGCCATTCCAGTGTCGCATCTGCATGgccAATTTCTCCCAGGGA GGCAACCTCACTAGGCACCTGCGGACTCATACAGGGGAGAAACCTTTCAGTGTAGGATCTGCATGGCTAATTTTTCACAGCATCCAAATTTGACGCGTCA CCTGAAAACCCACCTGAGAGGATCCTAAAAGCTTGATGGGGGATCCTAGG	BmgBI and AvrIII were used for cloning. Lower-case orange codons are those that were modified in order to facilitate an R-->A point mutation.
gBlocks used to create second sensor pair		
C-Terminal sensor changing ZF8/ZF7 to ZF2/ZF3	CACGTCCCGGGGAGCGCCCCTTCCAGTGCAGGATCTGCATGAGGAATTTTAGCACTGGTCAAAGATTGAGAATTCACACACGGACTCACACAGGGGAGAA GCCCTTCCAGTGTAGAATTTGCATGCGGAACTTCAGCCAAAATCAAAAATTTGGCTAGACACCTGCGGACCCATACTGGCGAAAAGCCCTTTCAGTGCCGGA TCTGCATGCGCAACTTCTCCGATAAATCTGTTTTGGCTAGACACCTGAAGACACATCTTCGGCAGAAAGATGGAGAGAGACCATTCCAGTGTCGGATTTGT ATGAGGAACTTTAGTCGCCAGGACAGGCTTGACAGGCATACCCGGACCCACACAGGTGAGAAGCCATTCCAGTGTCGCATCTGCATGCGGAATTTCTCCCA GAAGGAGCACTTGGCGGGGCACCTGCGGACTCATACAGGGGAGAAACCTTTCAGTGTAGGATCTGCATGCGCAATTTTTCACGCCGCGACAACCTGAACC GGCACCTGAAAACCCACCTGAGAGGATCCTAAAAGCTTGATGGGGGATCCTAGG	EcoRV and Bsu36I were used for cloning.
N-Terminal sensor changing ZF6/ZF10 to ZF5/ZF4	GATATCAAGGATCCATGGACTACAAAGACGATGACGACAAGCCTAAGAAGAAACGGAAGGTGTCCAGACCTGGGGAGAGGCCATTCCAGTGTCGGATCTG TATGAGAAATTTTAGCAAAGGTGAAAGATTGGTTAGACATACAAGAACCCATACAGGCGAGAAGCCCTTTCAGTGCAGGATCTGCATGCGAAACTTTTCT AGAATGGATAATTTGTCCACTCACCTGCGGACTCATACCGGCGAAAAACCATTTCAGTGTCGAATTTGCATGCGCAACTTTAGCAGAAAAGATGCTTTGA ATAGACACCTTAAGACTCATCTGAGGCAGAAGGACGGGGAGAGACCTTTCCAGTGCAGAATCTGCATGAGAAACTTTTCTACTGCTGCTGTTTTGACTAG ACATACCCGCACCCACACCGGCGAAAAACCCTTCCAGTGCCGCATCTGTATGCGAAATTTTAGCGATAGAGCTAATTTGACTAGACACCTTAGAACACACA CTGGAGAAAAACCATTTCAGTGTCGGATCTGCATGCGCAACTTCTCCAGAATTGATAAATTGGGTGATCACCTGAAAACCCACCTGAGG	BmgBI and AvrIII were used for cloning.

Supplementary Table 4					
Splice junction variant	**Source**	**N-terminal junction**	**C-terminal junction**	**post-splice retained splice junction linker**	**sited previous works**
V1	native SceVMA	..gsGGIIYVG/CFAKGT..	VVVHN/CGERGNGSGg..	Zinc finger domain..gsGGIIYVGCGERGNGSGg..VP64	
V2	split luciferase	..gsGGVVLEKG/CFAKGT..	VVVHN/CTMTEKGSGg..	Zinc finger domain....gsGGVLEKGCTMTEKGSGg..VP64	"Post-translational enzyme activation in an animal via optimized conditional protein splicing" Edmund C Schwartz, Lino Saez, Michael W Young & Tom W Muir
V3	split TEV protease	..gsGGFQTVG/CFAKGT..	VVVHN/CGEKSMGSGg..	Zinc finger domain....gsGGVLEKGCTMTEKGSGg..VP64	"An intein-cassette integration approach used for the generation of a split TEV protease activated by conditional protein splicing" Tim Sonntag and Henning D. Mootz
*lowercase Amino Acids preexist in the original ZF9 or VP64 domains. In orange: residues that were addded in addition to the "borrowed" splice junction Amino Acids in order to ensure flexibility of the post-splice, retained linker.					

Supplementary Table legends

Table 1: Plasmids deposited at addgene are listed according to name, brief description, bacterial selection marker, mammalian selection marker, and accession code.

Table 2: ZFs are listed by nomenclature used in related works and that used here along with their nucleic acid sequences. ZF binding sites consist of nine core nucleotides (nt) flanked on each side by a single (lower-case) nt noted for optimal binding. Hexadactyl ZFs

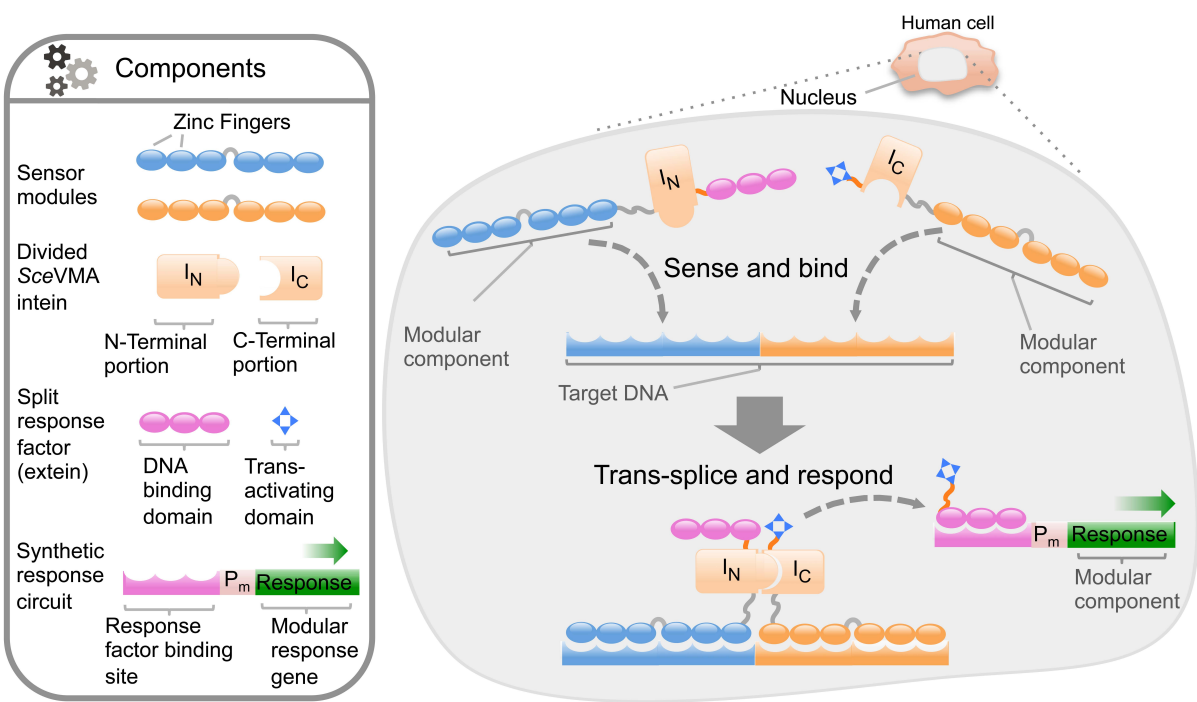
are listed as, for instance, ZF8/ZF7 (a merger of two tridactyl proteins, ZF8 and ZF7). The core contiguous 18 bp site is shown as well as sequences with one or two unrelated nts separating the two nine bp core sites. The target site for sensor pair 1 and for sensor pair 2 is shown. An example of a target site containing eight repeats of target site 1 is shown, as well. *Please be mindful of the ZF protein binding orientation on DNA target.

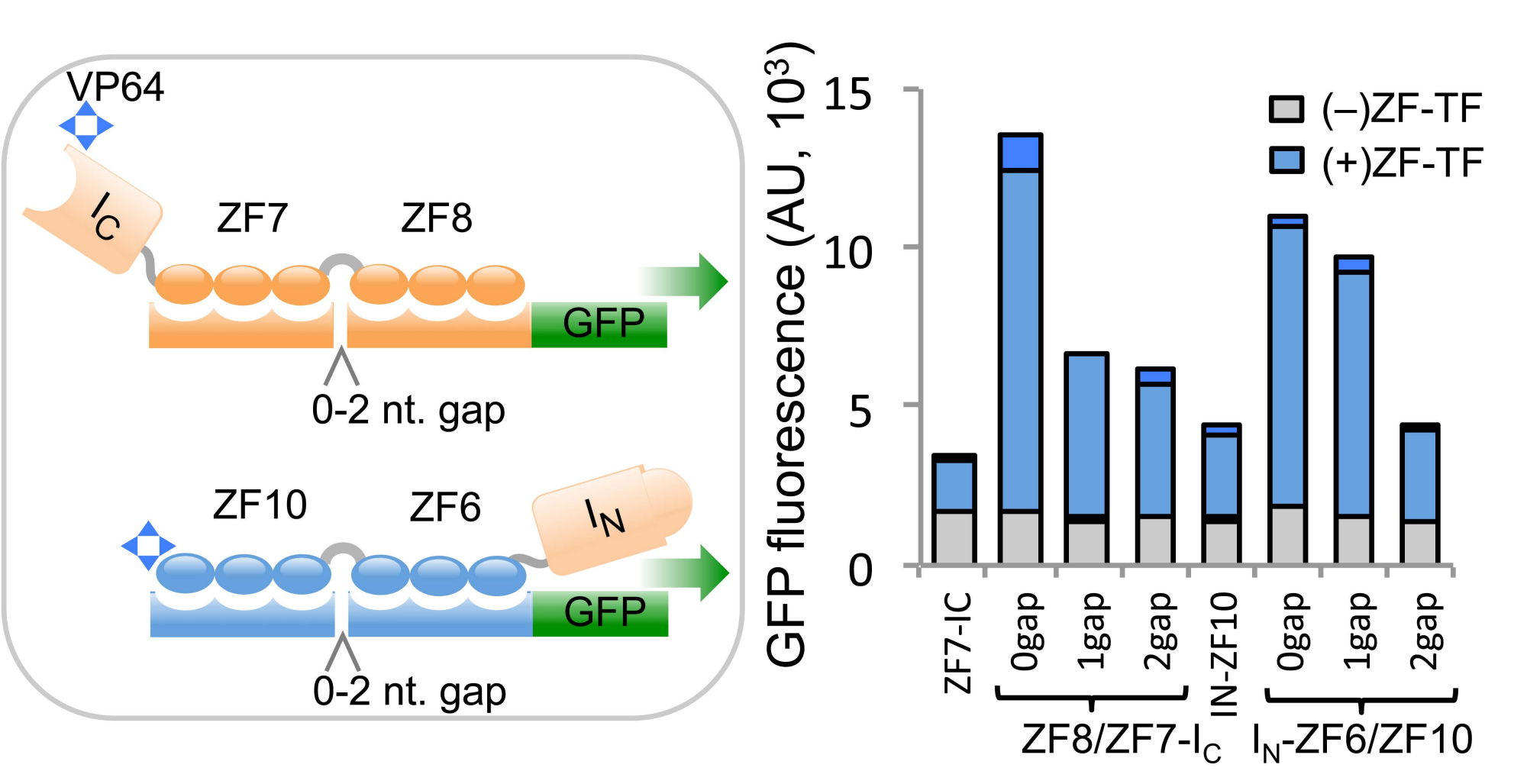
Table 3: The sequences of the central components used in this system are listed. These include those of the nuclear localization signal (NLS), trans-activating domain (VP64), the three fluorescent proteins used in this study, the SV40 ORI, NTR, linkers, and Tags used to purify protein and detect via immuno blot (FLAG and HIS). An example of the design of a reporter construct containing a single instance of the ZF9 operator is shown, as such was generated for each of the tridactyl and hexadactyl ZFs screened in this study, using their corresponding operator sites. Additionally, the relevant sequence of the response circuit containing six tandem repeats of the ZF9 binding site is provided; displaying the spaces between each repeat, the position and sequence of the minimal promotor, and the START codon of the response gene, eGFP.

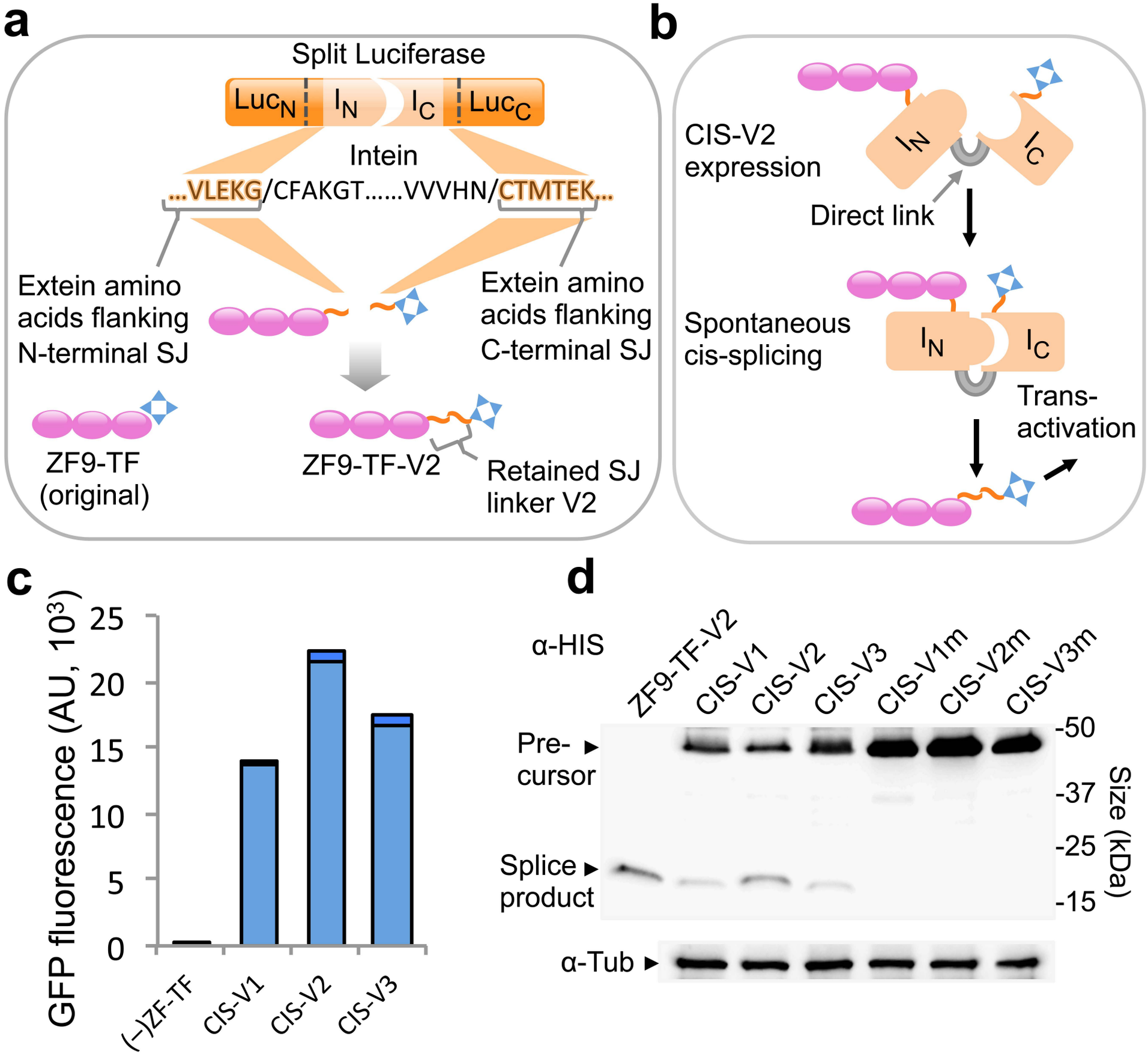
Also included are the primers used to amplify the intein portions from yeast gDNA, as well as their sequences. Additionally, in order to create a second hexadactyl sensor pair, capable of detecting a second target, the DNA recognition domains of the individual "fingers" were modified (while retaining conserved, structural regions not involved in DNA recognition); this was achieved by synthesizing gBlocks spanning the relevant area with the modified sequences, which are shown here. Furthermore, gBlock synthesis was

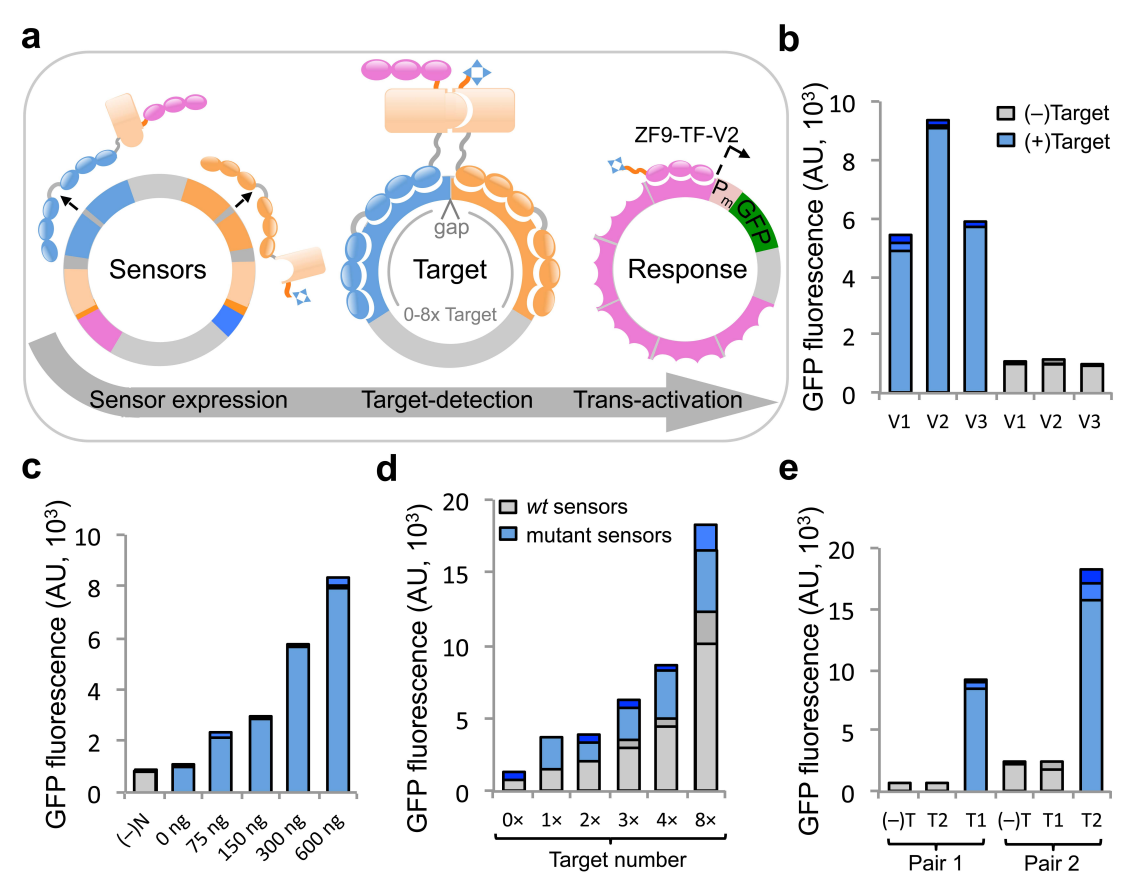
employed in order to easily alter the sequences of the *wt* ZF sensors to introduce R→A mutations. The gBlock sequences are shown here.

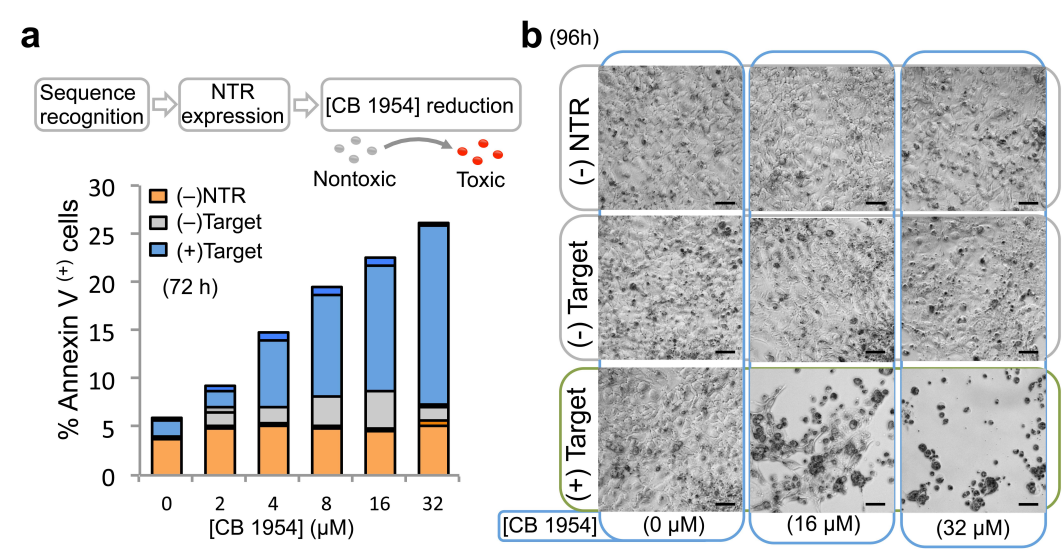
Table 4: Three splice junctions were tested for their ability to afford splicing of the split extein by the divided intein. These junctions (one between the E_N and I_N, and one between the E_C and I_C) were designed by introducing amino acid residues taken from three sources. The sources are listed, as is the specific amino acid content of each of the splice junctions. The post-splice junction linker sequence, as it exists within the extein after the intein has removed itself and ligated the split extein, is shown.

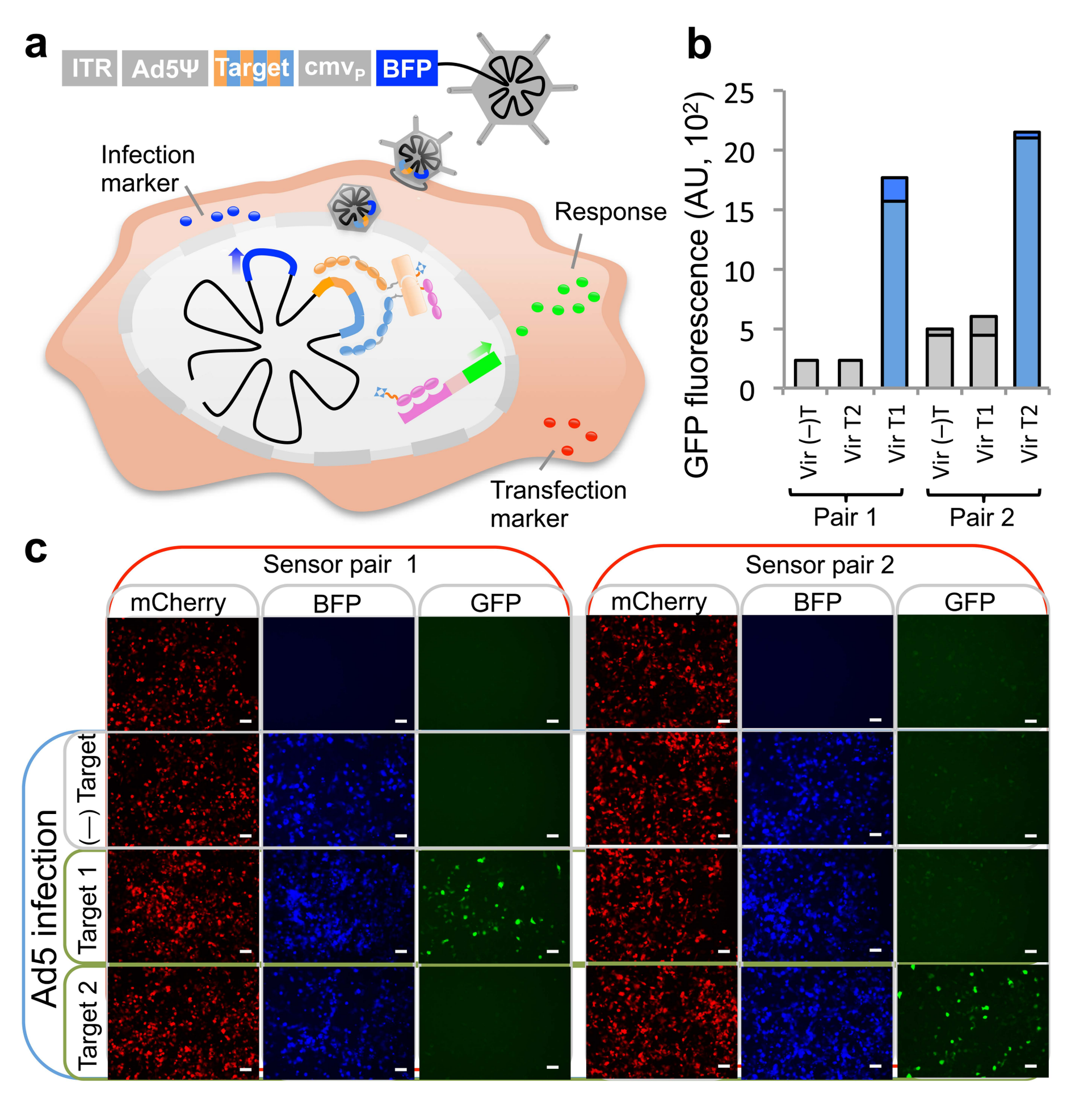

Components
Zinc Fingers
Sensor modules
Divided SceVMA intein
I_N
I_C
N-Terminal portion
C-Terminal portion
Split response factor (extein)
DNA binding domain
Trans-activating domain
Synthetic response circuit
P_m
Response
Response factor binding site
Modular response gene
Human cell
Nucleus
Sense and bind
Modular component
Target DNA
Modular component
Trans-splice and respond
P_m
Response
Modular component


VP64
I_C
ZF7
ZF8
GFP
0-2 nt. gap
ZF10
ZF6
I_N
GFP
0-2 nt. gap
GFP fluorescence (AU, 10^3)
15
10
5
0
(–)ZF-TF
(+)ZF-TF
ZF7-IC
0gap
1gap
2gap
IN-ZF10
0gap
1gap
2gap
ZF8/ZF7-I_C
I_N-ZF6/ZF10


a
Split Luciferase
Luc$_N$
I$_N$
I$_C$
Luc$_C$
Intein
...VLEKG/CFAKGT.....VVVHN/CTMTEK...
Extein amino acids flanking N-terminal SJ
Extein amino acids flanking C-terminal SJ
ZF9-TF (original)
ZF9-TF-V2
Retained SJ linker V2
b
CIS-V2 expression
Direct link
Spontaneous cis-splicing
Trans-activation
c
GFP fluorescence (AU, 10^3)
25
20
15
10
5
0
(−)ZF-TF
CIS-V1
CIS-V2
CIS-V3
d
α-HIS
ZF9-TF-V2
CIS-V1
CIS-V2
CIS-V3
CIS-V1m
CIS-V2m
CIS-V3m
Pre-cursor
Splice product
α-Tub
-50
-37
-25
-15
Size (kDa)


a
Sensors
gap
Target
0-8x Target
ZF9-TF-V2
Pm
GFP
Response
Sensor expression
Target-detection
Trans-activation
b
GFP fluorescence (AU, 10^3)
(–)Target
(+)Target
V1 V2 V3 V1 V2 V3
c
GFP fluorescence (AU, 10^3)
(–)N
0 ng
75 ng
150 ng
300 ng
600 ng
d
GFP fluorescence (AU, 10^3)
wt sensors
mutant sensors
0× 1× 2× 3× 4× 8×
Target number
e
GFP fluorescence (AU, 10^3)
(–)T T2 T1 (–)T T1 T2
Pair 1
Pair 2


a
Sequence recognition
NTR expression
[CB 1954] reduction
Nontoxic
Toxic
(–)NTR
(–)Target
(+)Target
(72 h)
% Annexin V (+) cells
30
25
20
15
10
5
0
0
2
4
8
16
32
[CB 1954] (μM)
b (96h)
(-) NTR
(-) Target
(+) Target
[CB 1954]
(0 μM)
(16 μM)
(32 μM)


a
ITR
Ad5Ψ
Target
cmv$_P$
BFP
Infection marker
Response
Transfection marker
b
GFP fluorescence (AU, 10^2)
25
20
15
10
5
0
Vir (–)T
Vir T2
Vir T1
Vir (–)T
Vir T1
Vir T2
Pair 1
Pair 2
c
Sensor pair 1
Sensor pair 2
mCherry
BFP
GFP
mCherry
BFP
GFP
Ad5 infection
(—) Target
Target 1
Target 2