

DNA sense-and-respond protein modules for mammalian cells

Shimyn Slomovic¹⁻³ & James J Collins¹⁻⁶

We 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 used 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, which provides a programmable means to cull infected cells. It may also be used to identify positively transduced or transfected cells, isolate recipients of intentional genomic edits and increase the repertoire of inducible parts in synthetic biology.

Cys₂His₂ 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 FokI endonucleolytic domain, ZF pairs can facilitate genomic editing³. Through sequence-enabled reassembly, ZFs joined to each domain of split β-lactamase can function as *in vitro* diagnostic tools⁴. Also, fusion to the viral trans-activating domain VP64 generates 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 epigenetic control, revealing synergistic activation and spatial regulation⁷.

Whether ZFs or other genome-engineering tools such as transcription activator–like effectors (TALEs) or clustered, regularly interspaced, short palindromic repeats (CRISPR)–Cas are used, 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 that indicates provirus presence, a chromosomal transposition or an intentional genomic alteration. Other desired responses might be immune system recruitment to the site of infection or tumor or kill-switch activation that induces apoptosis to prevent the spread of a pathogen. 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 the *trans* response, we used inteins, which are peptides that splice together their flanking regions, referred to as exteins^{9,10}. In pioneering work on conditional protein splicing, the contiguous *Saccharomyces cerevisiae* SceVMA intein was artificially split, rendering the separate halves inactive such that they required a condition for colocalization and reassembly¹¹. After one member of the rapamycin-binding heterodimeric pair, FRB–FKBP, was fused to each of the split SceVMA intein-extein halves, rapamycin binding served as a condition for their colocalization¹². The repertoire of *trans*-splice–inducing conditions has expanded and now includes temperature, light and protein-scaffold activation^{13–15}. 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 was used to measure drug-induced DNA demethylation¹⁶.

Here we sought to use DNA-sensing modules as dimerizing domains (Fig. 1). Programming two ZF-based sensors to bind adjacently in the targeted DNA sequence would allow for conditional intein colocalization 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.

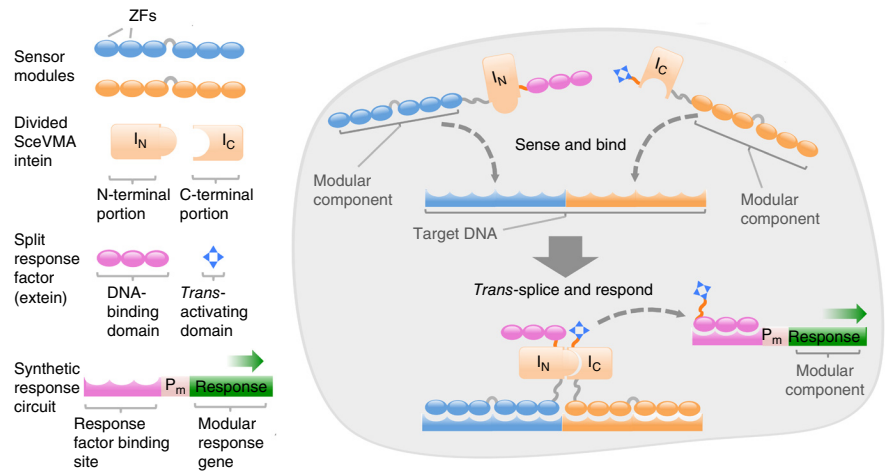
RESULTS

Modular ZF sensor design and screening

We used a library of ten tridactyl ZFs designed using oligomerized pool engineering (OPEN) and tested it for intracellular

¹Institute for Medical Engineering & Science, Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, USA. ²Department of Biological Engineering, MIT, Cambridge, Massachusetts, USA. ³Synthetic Biology Center, MIT, Cambridge, Massachusetts, USA. ⁴Harvard-MIT Program in Health Sciences and Technology, Cambridge, Massachusetts, USA. ⁵Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. ⁶Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts, USA. Correspondence should be addressed to J.J.C. (jimjc@mit.edu).

Figure 1 | A DNA sense-and-respond system. Schematic of the detection of a specific DNA sequence in a mammalian cell nucleus. Two modular ZF-based sensors are programmed to dock adjacently on target DNA. Each sensor contains an artificially split SceVMA intein fused to the split extein. The extein is a ZF-TF, split between the binding and *trans*-activating domains. Target detection facilitates intein colocalization 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.



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}. We fused the hexadactyl ZFs to VP64 and transformed the sequence encoding this construct into synthetic TFs. We prepared reporters for each ZF-TF encoding GFP modulated by a single copy of the two original 9–base pair (bp) sites with 0–2 unrelated base pairs between them (**Supplementary Fig. 2a**). We transiently cotransfected 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 C-terminal intein (I_C) or N-terminal intein (I_N) 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 chimeras nevertheless achieved a much greater output than the tridactyl-intein fusions. We found that a $2 \times (\text{GGGS})_3$ linker between the ZF and intein domains yielded higher GFP activation than the $1 \times$ or $3 \times$ forms, and we used this linker thereafter (**Supplementary Fig. 3c–e**).

Response-circuit optimization

We chose ZF9 (**Supplementary Fig. 1b,c**) as the split-response factor because it performed well in tested cell lines, and its operator yielded low basal expression. We selected GFP as a reporter for response-circuit characterization. To improve the dynamic range of GFP activation, we multimerized ZF9

operators (**Supplementary Fig. 4a**). The output signal strength increased with the number of sites, basal expression decreased (in the absence of ZF9-TF, resulting in a change of ~ 300 -fold using the $6 \times$ 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 activity of ZF9-TF^{19,20}. In the SJ design, we adopted several extein residues that natively flank both sides of the SceVMA intein or from two previously split proteins (luciferase and TEV protease), reasoning that the intein might successfully ligate the ZFs and VP64 in 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. We created sequences encoding split-extein domains fused to the corresponding divided inteins, directly linked to one another via a flexible linker (**Fig. 3b**)²². 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 the anticipated size, we isolated and immunoblotted protein. Flag-tag and HIS-tag probing of the N and C termini revealed a product that migrated with the ZF9-TF positive control for each of the variants but not for three splice-null mutants, which displayed greater accumulation of the unspliced precursor (**Fig. 3d** and **Supplementary Fig. 5d–f**)²².

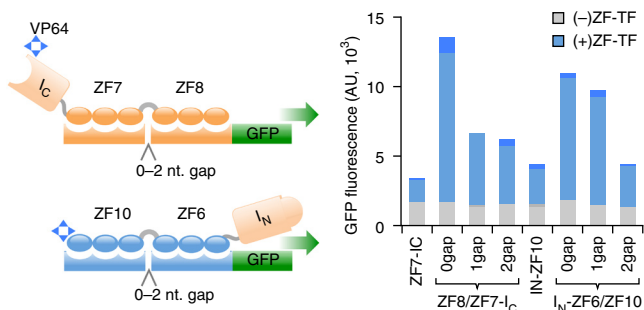


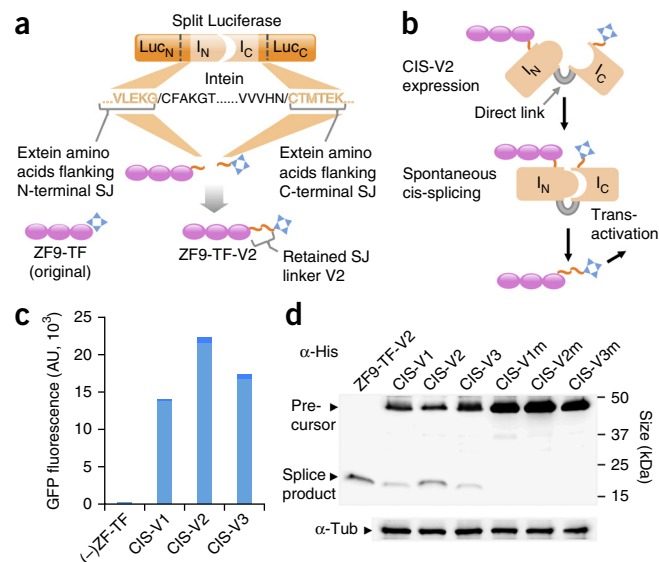
Figure 2 | ZF-intein sensor design. Left, schematic of hexadactyl (generated by merging tridactyl units) sensors, fused to corresponding intein domains (lacking ex-teins). Right, assay of intracellular binding activity. ZF7-IC and IN-ZF10 represent tridactyl proteins fused to the C- or N-terminal intein domains, respectively. Operators for hexadactyl proteins consisted of two contiguous 9-bp sites (0gap) or two 9-bp sites interrupted by one (1gap) or two (2gap) unrelated base pairs, with GFP as a reporter. GFP fluorescence of biological duplicate samples is displayed as overlaid bars, with (+)ZF-TF samples (tones of blue) compared to control samples (tones of gray), in which ZF-TFs were absent.

Figure 3 | Split-extein and SJ design, and *cis*-splice validation. (a) Extein residues from split luciferase introduced into ZF9-TF between its ZF and VP64 domains form SJs. Successful extein ligation would result in the ZF9-TF-V2 version. (b) Schematic of split extein *cis*-splicing testing in a condition-free manner by tethering the divided extein domains to one another with a flexible linker, applied to three extein variants (CIS-V1–3) and measured through GFP expression by the splice product. (c) Results of *trans*-activation assay (Supplementary Fig. 5c). Biological duplicate samples are displayed as overlaid bars. (d) Anti-His immunoblot of proteins extracted from HEK293FT cells expressing the three variants and splice-null double mutants (CIS-V1-3m) (Supplementary Fig. 5d–g shows a mutation schematic, an uncropped version of the blot in d, anti-Flag probing and quantification). Lane 1 is a ZF9-TF-V2 size control distinguishing between unspliced precursor and splice product. Total protein was normalized to anti-tubulin (α -Tub).

Semiquantitative 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 created constructs encoding fusions of the three SJ variants of the split extein with their hexadactyl sensing components. We transfected cells with a plasmid encoding the sensor pair, a plasmid carrying the target sequences, and a plasmid 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, in the absence of the N- or C-terminal sensor, or with a C-terminal deletion mutant (lacking the I_C domain) was comparably low, suggesting extremely low levels of spontaneous *trans*-splicing, and confirming earlier reports and that both intact intein domains are required to generate a response (Supplementary Fig. 6b)¹². Sensor pair V2 (containing the luciferase SJ) displayed the greatest activity, and we chose it for further characterization. Targets containing 0-, 4-, 8- or 12-bp gaps (nonrelated base pairs) 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-splice and pre-splice sensors on the DNA. We reasoned that with a moderate decrease in affinity but not specificity, interchange might be improved. We introduced eight R-to-A mutations expected to lower ZF-binding affinity into the ZF backbones and compared output signals of wild-type and mutant sensor pairs²³. Equimolar target plasmids containing a gradient of target sequence instances (0–4 and 8) produced a correlative effect for wild-type and mutant sensors, but the output of the latter was higher, suggesting improved interchange (Fig. 4d). We also assessed the difference between a nonreplicative and replicative target and found the signal to be substantially higher in the latter case (Supplementary

Fig. 7a,b and Supplementary Note 2). To demonstrate the platform's extensibility, we constructed a second sensor pair (ZF3/ZF2 and ZF4/ZF5) corresponding to

Figure 4 | Intracellular DNA sense-and-respond system.

(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, and (iii) response circuit modulated by spliced ZF9-TF. (b) Transfection of three sensor pair variants (V1, V2 and V3) with and without target plasmid (Supplementary Fig. 6a,b). (c) Output with indicated amounts of transfected target plasmid. N-terminal sensor is absent in (-)N. (d) Signal comparison between wild-type (WT) and mutant sensors using a series of target plasmids containing 0–4 or 8 target instances. (e) Orthogonality testing for two sensor pairs and their target plasmids (T1 and T2). GFP fluorescence of biological triplicate samples (b,c,e) and duplicate samples (d) are displayed as overlaid bars of blue tones or gray tones.

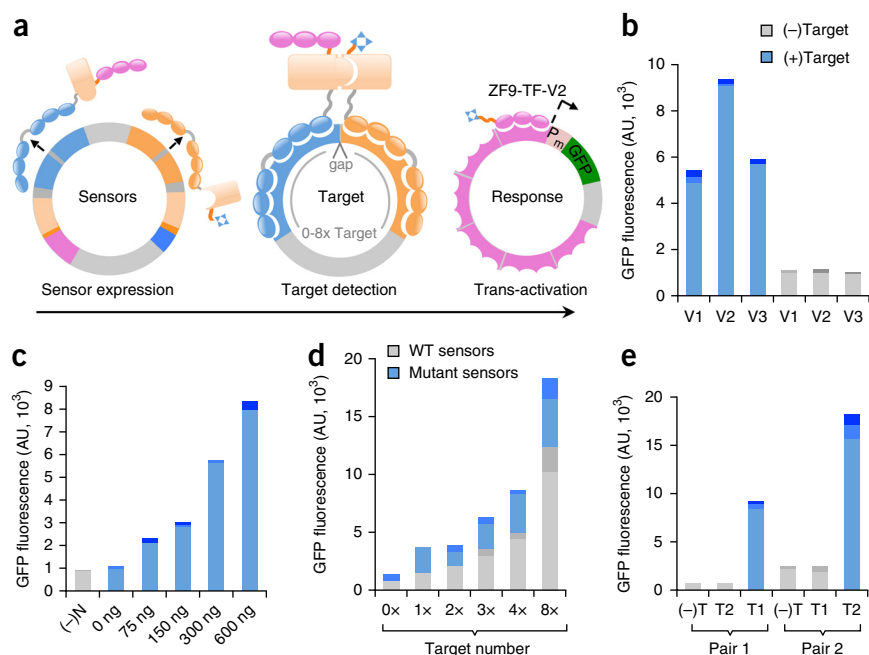
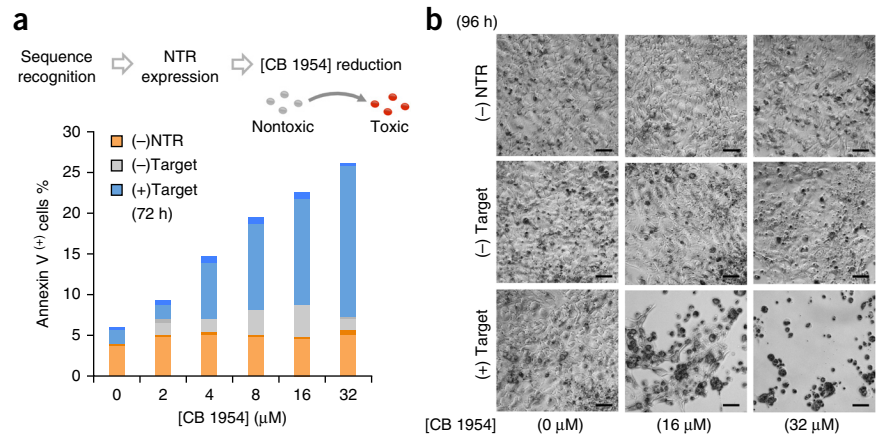


Figure 5 | Sequence recognition–induced apoptosis. **(a)** Schematic: target sequence recognition leads to NTR expression, CB 1954 reduction, and apoptosis. Transfection (50% rate; **Supplementary Fig. 8a**) of a sensor system with target, without target or in the absence of the NTR response circuit.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



a second target sequence, and tested for orthogonality between the two pairs. 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 involves the metabolism of a nontoxic prodrug into a cytotoxic form²⁴. 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 a bacterial nitroreductase (NTR), which transforms it into an agent of DNA interstrand cross-linking and apoptosis²⁵. Diffusion into

neighboring cells leads to increased potency through a 'bystander' effect. We used this dual-component method 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 medium, NTR expression and the ensuing CB 1954 reduction should, in principle, occur only in the presence of the target sequence²⁶.

We plated HEK293FT cells in medium containing a range of prodrug concentrations and transfected them (at a 50% rate, relative to 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**). We measured apoptosis 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

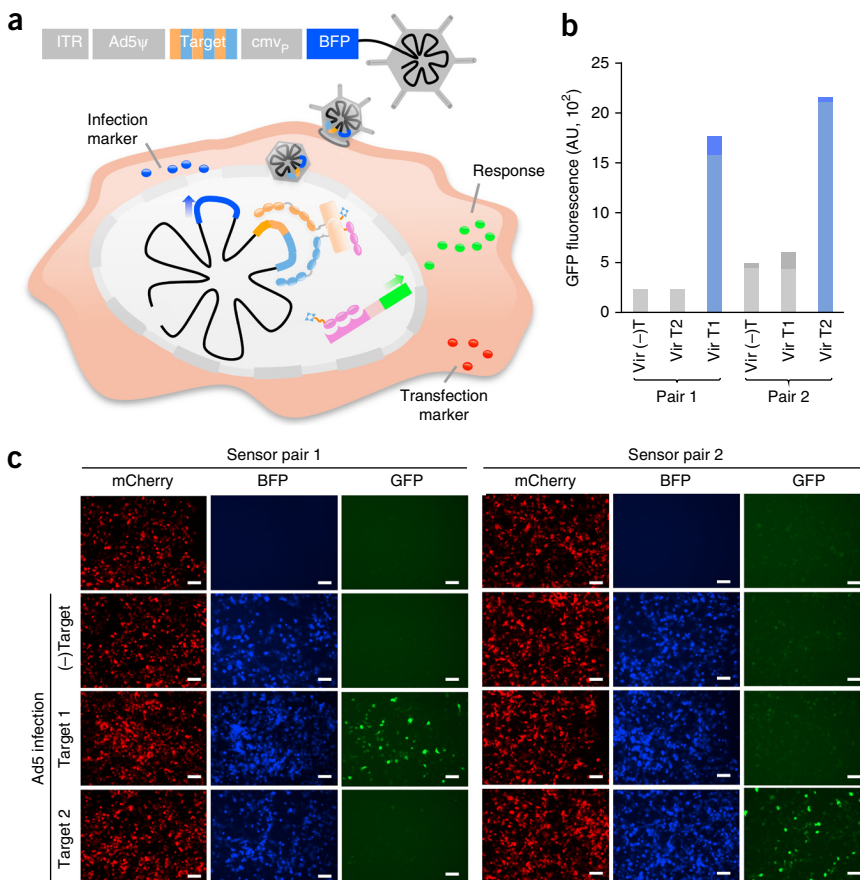


Figure 6 | Adenovirus detection and reporting. **(a)** Instances of targets 1 or 2 (T1 or T2), or no target, inserted into the Ad5 genome downstream of the packaging signal (ψ). Constitutive BFP expression served as a viral infection marker. The detection system (mCherry as a transfection marker), based on sensor pair 1 or 2, is present in the cell nucleus before infection. Cells were infected with virus containing T1, T2 or no target (Vir (-)T). Sequence-specific detection was measured through GFP *trans*-activation. **(b)** GFP fluorescence in mCherry+BFP+ cells of biological duplicate samples is displayed as overlaid bars. HEK293FT cells were transfected and infected as described above. **(c)** Epifluorescence microscopy of cells subjected to transfection followed by infection (mCherry indicates transfection, BFP indicates infection, and GFP indicates detection). Scale bars, 50 μM .

drug gradient. Microscopy showed that at 96 h, control populations were excessively confluent (Fig. 5b and Supplementary Fig. 8b). Conversely, target-positive samples at 16 μ M or 32 μ M CB 1954 displayed severe or complete cell death, providing evidence of sequence recognition and a bystander effect. To demonstrate a dual-output response, we generated a construct encoding GFP-IRES-NTR, which simultaneously showed target reporting and apoptosis (Supplementary Fig. 8c).

Virus detection in infected mammalian cells

To test the system's capability for detecting viral infection, we adopted adenovirus type 5 (Ad5) as a model²⁷. We generated virus with constitutive expression of blue fluorescent protein (BFP) to track infected cells. We prepared two substrains with target sequences of sensor pair 1 or 2 (T1 or T2) embedded into the viral genome adjacent to the packaging signal (ψ) (Fig. 6a). We transfected replication-permissive HEK293FT cells with either of the sensor pairs and then infected them with virus containing T1, T2 or neither target. The results showed that both sensor pairs detected only virus carrying their specific targets (Fig. 6b,c). Finally, we condensed the expression, reporting and marker elements 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 developed a system that is sufficiently robust to facilitate sequence detection in mammalian cells but is also sufficiently modular to allow for its application in a range of scientific settings, including synthetic biology, diagnostics and basic research.

The recognition stage can be adjusted to detect any target sequence by installation of the appropriate ZF component. Alternate DNA-binding technologies could be tested, including TALEs and the CRISPR-Cas system²⁸. Another option would be to use designer RNA-binding proteins to detect pathogens with RNA-based genomes²⁹. Cells infected with a particular pathogen, retaining viral latency or undergoing carcinogenesis could be culled through the expression of proapoptotic proteins, cell death-activating surface molecules or cytokines for immune-system recruitment^{30–33}.

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 used to introduce mutations or replace genomic sequences via homology-directed repair⁸. An intrinsic part of such studies is the identification and isolation of cells containing the edit. PCR or DNA sequencing can be used to measure frequencies but cannot be applied to live cells.

Furthermore, to establish a clonal cell line containing the edit, multiple lines must be screened. Instead, cells could be equipped with sensors designed to bind in the insertion or deletion site, allowing for easy identification of cells that have undergone the desired editing on a single-cell basis, 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 provirus³⁴. The use of 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 inversions³⁵. However, PCR-based methods cannot be used to study real-time biology or disclose the aberration frequency in 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-conformation capture techniques (3C, 4C, 5C or Hi-C) to map long-range interactions³⁶. Stages include DNA cross-linking, 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.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank S. Modi for helpful discussions and critical reviews of the manuscript. HEK293FT cells were donated by members of the R. Weiss laboratory (MIT, Cambridge, Massachusetts, USA). This work was supported by funding from the Defense Advanced Research Projects Agency (grant DARPA-BAA-11-23), the Defense Threat Reduction Agency (grant HDTRA1-14-1-0006) and the Air Force Office of Scientific Research (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 FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Maeder, M.L. *et al.* Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell* **31**, 294–301 (2008).
- Kim, S., Lee, M.J., Kim, H., Kang, M. & Kim, J.S. Preassembled zinc-finger arrays for rapid construction of ZFNs. *Nat. Methods* **8**, 7 (2011).
- Tebas, P. *et al.* Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N. Engl. J. Med.* **370**, 901–910 (2014).
- 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).
- Khalil, A.S. *et al.* A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647–658 (2012).
- Lohmueller, J.J., Armel, T.Z. & Silver, P.A. A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. *Nucleic Acids Res.* **40**, 5180–5187 (2012).
- 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).
- Gaj, T., Gersbach, C.A. & Barbas, C.F. 3rd ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**, 397–405 (2013).
- 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).
- 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*. *J. Biol. Chem.* **265**, 6726–6733 (1990).
- Mootz, H.D. & Muir, T.W. Protein splicing triggered by a small molecule. *J. Am. Chem. Soc.* **124**, 9044–9045 (2002).
- Schwartz, E.C., Saez, L., Young, M.W. & Muir, T.W. Post-translational enzyme activation in an animal via optimized conditional protein splicing. *Nat. Chem. Biol.* **3**, 50–54 (2007).
- Zeidler, M.P. *et al.* Temperature-sensitive control of protein activity by conditionally splicing inteins. *Nat. Biotechnol.* **22**, 871–876 (2004).
- Tyszkiewicz, A.B. & Muir, T.W. Activation of protein splicing with light in yeast. *Nat. Methods* **5**, 303–305 (2008).
- Selgrade, D.F., Lohmueller, J.J., Liener, F. & Silver, P.A. Protein scaffold-activated protein trans-splicing in mammalian cells. *J. Am. Chem. Soc.* **135**, 7713–7719 (2013).
- Huang, X. *et al.* Sequence-specific biosensors report drug-induced changes in epigenetic silencing in living cells. *DNA Cell Biol.* **31** (suppl. 1), S2–S10 (2012).
- Kim, J.S. & Pabo, C.O. Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc. Natl. Acad. Sci. USA* **95**, 2812–2817 (1998).
- Liu, Q., Segal, D.J., Ghiara, J.B. & Barbas, C.F. III. Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc. Natl. Acad. Sci. USA* **94**, 5525–5530 (1997).
- 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).
- 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 Res.* **33**, e34 (2005).
- 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).
- Chong, S. & Xu, M.Q. Protein splicing of the *Saccharomyces cerevisiae* VMA intein without the endonuclease motifs. *J. Biol. Chem.* **272**, 15587–15590 (1997).
- Elrod-Erickson, 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).
- 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. *Biochem. Pharmacol.* **18**, 1519–1527 (1969).
- 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 Rev.* **12**, 195–212 (1993).
- Bogdanove, A.J. & Voytas, D.F. TAL effectors: customizable proteins for high-level expression. *BMC Cancer* **9**, 301 (2009).
- Lion, T. Adenovirus Infections in Immunocompetent and Immunocompromised Patients. *Clin. Microbiol. Rev.* **27**, 441–462 (2014).
- Bogdanove, A.J. & Voytas, D.F. TAL effectors: customizable proteins for DNA targeting. *Science* **333**, 1843–1846 (2011).
- Filipovska, A. & Rackham, O. Designer RNA-binding proteins New tools for manipulating the transcriptome. *RNA Biol.* **8**, 978–983 (2011).
- 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).
- Yu, J., Zhang, L., Hwang, P.M., Kinzler, K.W. & Vogelstein, B. PUMA induces the rapid apoptosis of colorectal cancer cells. *Mol. Cell* **7**, 673–682 (2001).
- 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).
- 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).
- Banerjee, C. *et al.* BET bromodomain inhibition as a novel strategy for reactivation of HIV-1. *J. Leukoc. Biol.* **92**, 1147–1154 (2012).
- Nambiar, M., Kari, V. & Raghavan, S.C. Chromosomal translocations in cancer. *Biochim. Biophys. Acta* **1786**, 139–152 (2008).
- 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).

ONLINE METHODS

General cloning procedures. Information on the cloning and deposition of plasmids constructed during this study can be found in **Supplementary Note 3** and **Supplementary Table 1**. 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 religation 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 µg/ml ampicillin or 50 µg/ml kanamycin.

OPEN ZF library and split SceVMA intein. The ten tridactyl ZFs used in this study were developed using OPEN, as previously described⁵ (**Supplementary Note 1** and **Supplementary Table 2**). In all screening, VP64 was C-terminally fused. The SceVMA intein was simultaneously amplified and divided from gDNA purified from *S. cerevisiae*, as previously described¹¹. I_N and I_C consist of residues 1–184 and 390–454, respectively, in the full SceVMA sequence. Glycine/serine insulating linkers were introduced by gBlock synthesis (Integrated DNA Technologies (IDT)) (**Supplementary Table 3**). For split extein and splice junction design, see **Supplementary Table 4**.

Bacterial strains, overexpression and electrophoretic mobility shift assay. For cloning, 10-beta electrocompetent *E. coli* (NEB) cells were used, and for protein overexpression, 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 electrophoretic mobility shift assay 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). Gene encoding luciferase was replaced with that encoding GFP. pVITRO1/MCS/Neo (InvivoGen) was used to express ZF-TFs and sensor chimeras in human cells. pBW121 served as a cotransfection 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 to their complementary strand, compatible with ligation into a linearized plasmid. Lyophilized oligos were resuspended in annealing buffer (10 mM Tris-HCl, pH 7.5, and 50 mM NaCl), equal volumes of sense and antisense DNA were mixed at 100 °C, cooled to 24 °C 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 members of the Weiss laboratory, 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: 100,000–200,000 cells were plated in 12-well format and transfected 24 h later with 1–1.5 µg combined DNA using polyethylenimine (PEI) as a transfection reagent. 48 h after transfection, cells were examined using a Nikon Eclipse Ti epifluorescence microscope, trypsinized and prepared for flow cytometry analysis.

Mammalian protein extraction and immunoblots. 48 h after 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 0.2 M and a working stock (500 µM) was prepared by dilution into DMEM before each experiment. Cells were transfected as described above and supplemented with 0–32 µM prodrug. After 48 h or 72 h, 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 96 h.

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 cotransfection 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 24 h, infected with virus at a multiplicity of infection of 5. 48 h later, cells were trypsinized, fixed and analyzed by flow cytometry.

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 the indicated number of times and data are displayed either as an average of biological triplicates or duplicates with s.d., or as fold change with

error propagation calculations. Owing to experimental variation, although data trends may be compared between experiments, absolute values should be compared within single experiments.

Imaging. Equipment and settings for blot imaging and microscopy are listed in **Supplementary Note 4**.