

# Barcoding cells using cell-surface programmable DNA-binding domains

Prashant Mali<sup>1,3</sup>, John Aach<sup>1,3</sup>, Je-Hyuk Lee<sup>1,2</sup>, Daniel Levner<sup>1,2</sup>, Lisa Nip<sup>2</sup> & George M Church<sup>1,2</sup>

**We report an approach to barcode cells through cell-surface expression of programmable zinc-finger DNA-binding domains (surface zinc fingers, sZFs). We show that sZFs enable sequence-specific labeling of living cells by dsDNA, and we develop a sequential labeling approach to image more than three cell types in mixed populations using three fluorophores. We demonstrate the versatility of sZFs through applications in which they serve as surrogate reporters, function as selective cell capture reagents and facilitate targeted cellular delivery of viruses.**

The ability to construct and interrogate complex tissues and cellular libraries at single-cell resolution requires methods that enable highly multiplexed *in situ* probing of living cells. Although the use of fluorescent proteins has revolutionized the analysis of biological phenomena, their multiplexed use is limited to combinations that can be spectrally resolved. To expand the repertoire of probing tools, we explored the possibility of using DNA-binding domains such as zinc-finger proteins and transcription activator-like effectors (TALEs). Our motivation stemmed from the observation that as receptor-ligand pairings, the zinc finger–DNA or TALE–DNA interaction is unique in that both the receptor (zinc-finger or TALE protein) and the ligand (DNA) are highly programmable, and hence the space of engineerable orthogonal interactions is huge. Consequently, they can potentially be leveraged for engineering macromolecular interactions beyond genome targeting<sup>1–5</sup>. Here we exploit the programmability of the ZF–DNA interaction to barcode and image multiple cell types by anchoring zinc-finger proteins to the outside of the cell membrane and thus making them accessible to DNA-based probes provided in the extracellular medium.

To express zinc-finger DNA-binding domains on the cell surface, we fused an immunoglobulin  $\kappa$ -chain leader sequence at their N terminus and a platelet-derived growth factor (PDGF) transmembrane domain at their C terminus (Online Methods)<sup>6</sup>. To test the ability of sZF-expressing cells to bind DNA, we exposed them to fluorophore-tagged DNA molecules. sZF-expressing cells were observed to be strongly fluorescent, whereas control cells exhibited very low signals, implying that the former could bind DNA and, thus, that functional zinc-finger proteins were successfully

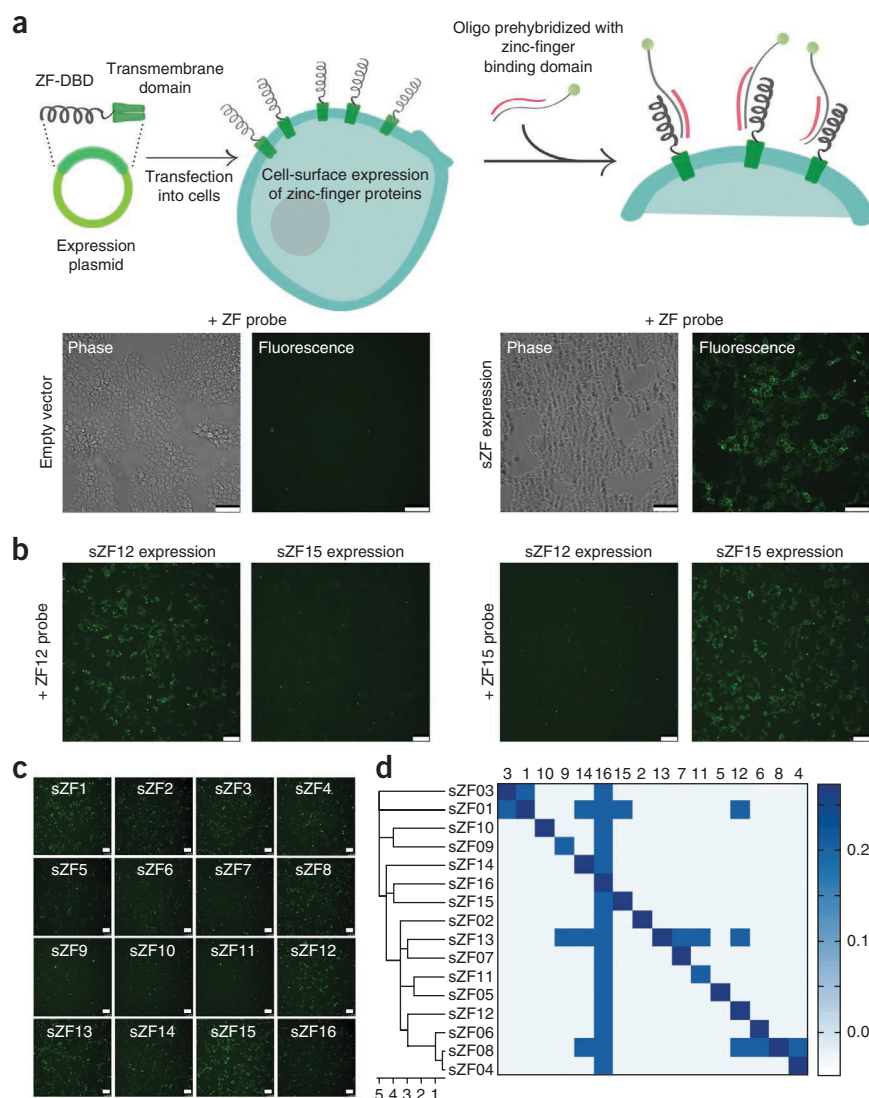
expressed on the cell surface (Fig. 1a). Two aspects of this sZF–DNA interaction are notable. First, sZFs could bind to both ssDNA<sup>7</sup> and dsDNA (Supplementary Fig. 1a); however, the former interaction was abrogated in the presence of competitor dsDNA (here, salmon sperm DNA). Second, sZFs bound nonspecifically to dsDNA; but again, binding to only their cognate target dsDNA was retained in the presence of competitor dsDNA (Supplementary Fig. 1b). We obtained similar results using FACS-based assays (Supplementary Fig. 2). Thus, in the presence of competitor dsDNA, sZF-expressing cells specifically bound their target dsDNA probe, and hence each zinc-finger protein barcoded the cell type expressing it (Fig. 1b).

We tested a total of 16 zinc-finger proteins<sup>8</sup> using this approach (protein sequences and target dsDNA sequences are provided in Supplementary Table 1). Several aspects of sZF–dsDNA interactions emerged from this analysis. First, different sZFs had different binding affinities for their target dsDNA (Fig. 1c). Specifically, whereas some bound their targets strongly (sZFs 1, 3, 8, 12, 13, 15 and 16), as assayed by both fluorescence intensity and duration of binding, some were moderately strong binders (sZFs 2, 4–7, 10 and 14), and others were only weak binders (sZFs 9 and 11). To enable these comparisons, we expressed all sZFs from identical expression vectors and under uniform transfection conditions. Next we evaluated the cross-reactivity profile for these 16 sZFs: we tested each sZF for its ability to bind its target dsDNA probe and also dsDNA probes corresponding to the other sZFs (Fig. 1d). Although most zinc fingers bound their target dsDNA specifically, some showed cross-reactivity (sZFs 1, 8 and 13). The strong sZF binders were particularly susceptible to this phenomenon<sup>9</sup>. Notably, we observed that almost all the zinc fingers bound the sZF16 target dsDNA, likely owing to its poly(G)-rich content. On the basis of these results, sZFs 2–7, 10, 12, 14 and 15 are moderate-to-strong binders that are orthogonal to each other and are thus good candidates for barcoding cells.

If sZFs are to serve as efficacious barcodes compatible with the analysis of structured tissues, they must enable differential labeling of cells in complex mixtures. To investigate this, we designed experiments to image and analyze mixtures of sZF-expressing cell populations. Cells expressing either sZF1, sZF2, sZF3 or sZF4 were mixed in pairs (sZF1 + sZF2 and sZF3 + sZF4) or in a pool of three (sZF1 + sZF2 + sZF3) and were probed using appropriate combinations of fluorophore-labeled target dsDNA molecules. We developed a suite of Matlab GUI applications to analyze the resulting images and compute quantitative measures of the specificity of binding of sZFs to their corresponding oligonucleotides at both the whole-cell and single-pixel levels (Online Methods; software is available on [http://arep.med.harvard.edu/sZF\\_cell\\_barcode/](http://arep.med.harvard.edu/sZF_cell_barcode/)). The processing flow for images is depicted

<sup>1</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. <sup>2</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, Massachusetts, USA. <sup>3</sup>These authors contributed equally to this work. Correspondence should be addressed to G.M.C. ([gchurch@genetics.med.harvard.edu](mailto:gchurch@genetics.med.harvard.edu)).

**Figure 1** | Labeling of sZF-expressing live cells. (a) Top, schematic of the approach. ZF-DBD, zinc-finger DNA-binding domain. Bottom, 293T cells transfected with an empty vector or with a sZF and then incubated with a fluorophore-labeled DNA probe. (b) 293T cells expressing the indicated sZF and labeled specifically with the indicated target dsDNA. (c) 293T cells individually expressing 1 of 16 sZFs labeled with the corresponding dsDNA probe. (d) Heat map showing the cross-reactivity profile of the 16 sZFs. Each sZF was tested for binding to its own target dsDNA sequence and to the sequences corresponding to the other sZFs (normalized strength of dsDNA binding assayed by relative fluorescence intensity in depicted). The sZFs were clustered according to target sequence similarity (**Supplementary Table 1**). Scale bars, 100  $\mu\text{m}$ .

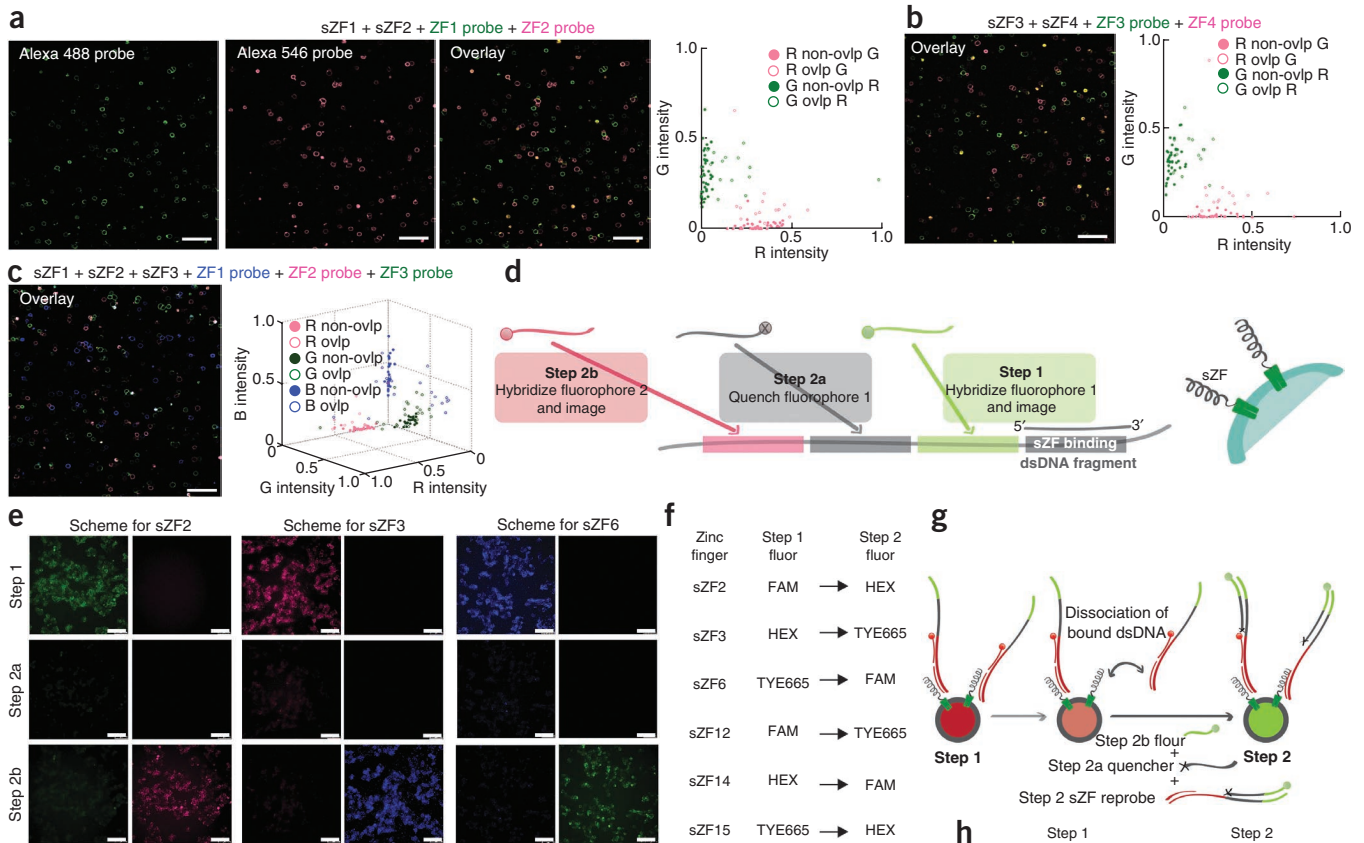


in **Supplementary Figure 3**. We confirmed with qualitative inspection and quantitative analysis of the images that the sZF-dsDNA interactions were sequence specific (**Fig. 2a–c**). Additional labeling examples and their corresponding analyses are presented in **Supplementary Table 2** and **Supplementary Figures 3–9**.

Exploring more zinc fingers or extending this approach to TALEs<sup>10,11</sup> will further expand and refine the list of orthogonal interaction pairs that can be used for cellular barcoding. But one is still limited by the small number of spectrally distinct fluorophores available for simultaneous cell imaging. To address this problem, we devised a sequential hybridization and imaging approach suitable for live adherent cells. It uses a modified two-part DNA probe that presents a double-stranded portion that binds the sZF and a single-stranded portion containing barcode sequences that can be read out by serial hybridizations (**Fig. 2d**). Specifically, in step 1, a fluorophore-tagged complementary oligonucleotide is hybridized to its target site, enabling a first fluorescence readout; in step 2, two adjacent complementary oligonucleotides are annealed, the first bearing a quencher that suppresses the step 1 fluorescence signal, and the second bearing another fluorophore that enables a second fluorescence readout. Extending this scheme to  $n$  steps enables barcoding of  $3^n$  cell types using just three fluorophores. We first demonstrated the scheme in a simplex setting (**Fig. 2e**). Here sZF-expressing cells were sequentially probed, with each sZF identity encoded by two colors: for instance, sZF2 by green in step 1 and red in step 2, sZF3 by red in step 1 and blue in step 2, and similarly for sZFs 6, 12, 14 and 15 (**Fig. 2f**). Next, we were also able to mix up to six distinctly labeled cells and identify their barcodes in the mixed population, using two hybridization cycles (**Fig. 2g,h**). In these multiplexed experiments, the zinc finger-binding probes were resupplied for each round of sequencing by hybridization. This reprobng compensated for the loss of fluorescence signal due to the dissociation of dsDNA probes from the sZFs in the interval between imaging

steps and also aided in the active displacement of the existing probes, thus mitigating effects of any incomplete quenching in the previous step (see **Supplementary Fig. 10** for representative dissociation kinetics of dsDNA probes). We also validated the genotype-to-labeling association in these experiments (**Supplementary Fig. 11**). Toehold-mediated strand exchange<sup>12</sup> to displace bound DNA probes can be used to refine this technique. Overall, our results suggest that a sequential tagging scheme can successfully identify constituent cells in complex mixtures of barcoded cell types using just three spectrally distinct fluorophores.

Finally we explored the versatility of sZFs for applications other than *in situ* labeling. First, we used sZFs as surrogate reporters of endogenous cellular activity. We constructed lentiviral vectors with small-molecule (tetracycline and cumate)-inducible promoters driving sZF expression, and we stably transduced 293T and HeLa cells with these constructs. Upon small-molecule induction, sZF expression was readily detected (**Fig. 3a** and **Supplementary Fig. 12**). Inducible expression of barcodes can also be used to minimize potential effects of sZF expression on cell physiology and toxicity (**Supplementary Fig. 13** provides a representative toxicity profile of sZFs in cells). Second, we examined sZFs as convenient cell-surface handles for DNA-mediated cell capture.



**Figure 2** | Multiplex labeling of sZF-expressing live cells. (a–c) Cells expressing either of four sZFs (sZF1, sZF2, sZF3 or sZF4) were mixed in pairs (a,b) or in a pool of three (c) as indicated and were probed using combinations of Alexa 488–, Alexa 546– and Alexa 647–labeled target dsDNA molecules. Scatter plots analyzing the cell populations in the micrographs are shown. G, green; R, magenta; B, blue; ovlp, overlap. (d) Schematic showing a sequential tagging technique for imaging of more than three barcoded cell types using three resolvable fluorophores (fluor). (e) Demonstration of the scheme in d in a simplex setting. The micrographs show sequentially probed cells expressing the indicated sZF; sZF identity is encoded by two colors. (f) Codes for additional sZFs. (g) Schematic for sequential imaging in a multiplex setting. (h) Demonstration of the sequential approach in a multiplex setting. Six sZFs were individually expressed in cells that were subsequently mixed and then sequentially imaged. Two sets of images using combinations of three fluorophores are shown. The sZF identity based on f is indicated in each case. Scale bars, 100  $\mu$ m.

We successfully enriched sZF-expressing cells from a mixture containing unlabeled K562 cells by performing a pull-down either using dsDNA probe–conjugated magnetic beads (Fig. 3b) or on dsDNA arrays (Supplementary Fig. 14). Third, we demonstrated sZF-mediated selective gene delivery by pseudotyping<sup>13</sup> lentiviruses with dsDNA probes (Supplementary Fig. 15). Modified lentiviruses successfully delivered genes to sZF-barcoded cells more efficiently than in the absence of the DNA pseudotyping (Fig. 3c). Thus, sZFs have uses beyond direct labeling of cells that include state probing, capture and targeting of cells.

sZF barcoding enables specific and quantifiable cellular labeling and is suited for applications requiring tracking of heterogeneous mixtures of cells. It can potentially synergize with existing methods for multiplexed cell probing such as elemental isotope–labeled antibody–based mass cytometry<sup>14</sup> and combinatorial fluorescent-protein expression<sup>15</sup>. We note that a threshold amount of sZF protein expression is needed to discernibly label a cell, and this will affect the sensitivity of the readout. Brighter probes such as quantum dots could be used to amplify the signal. Furthermore the

versatility of sZFs should make them a powerful tool enabling many synthetic biology applications. These could range from multiplexed tracking of endogenous gene activity for studying pathways and interacting gene networks<sup>16</sup>, to tissue engineering through control of physical cell arrangement<sup>17,18</sup> and to the engineering of designer macromolecular associations<sup>19,20</sup>.

## METHODS

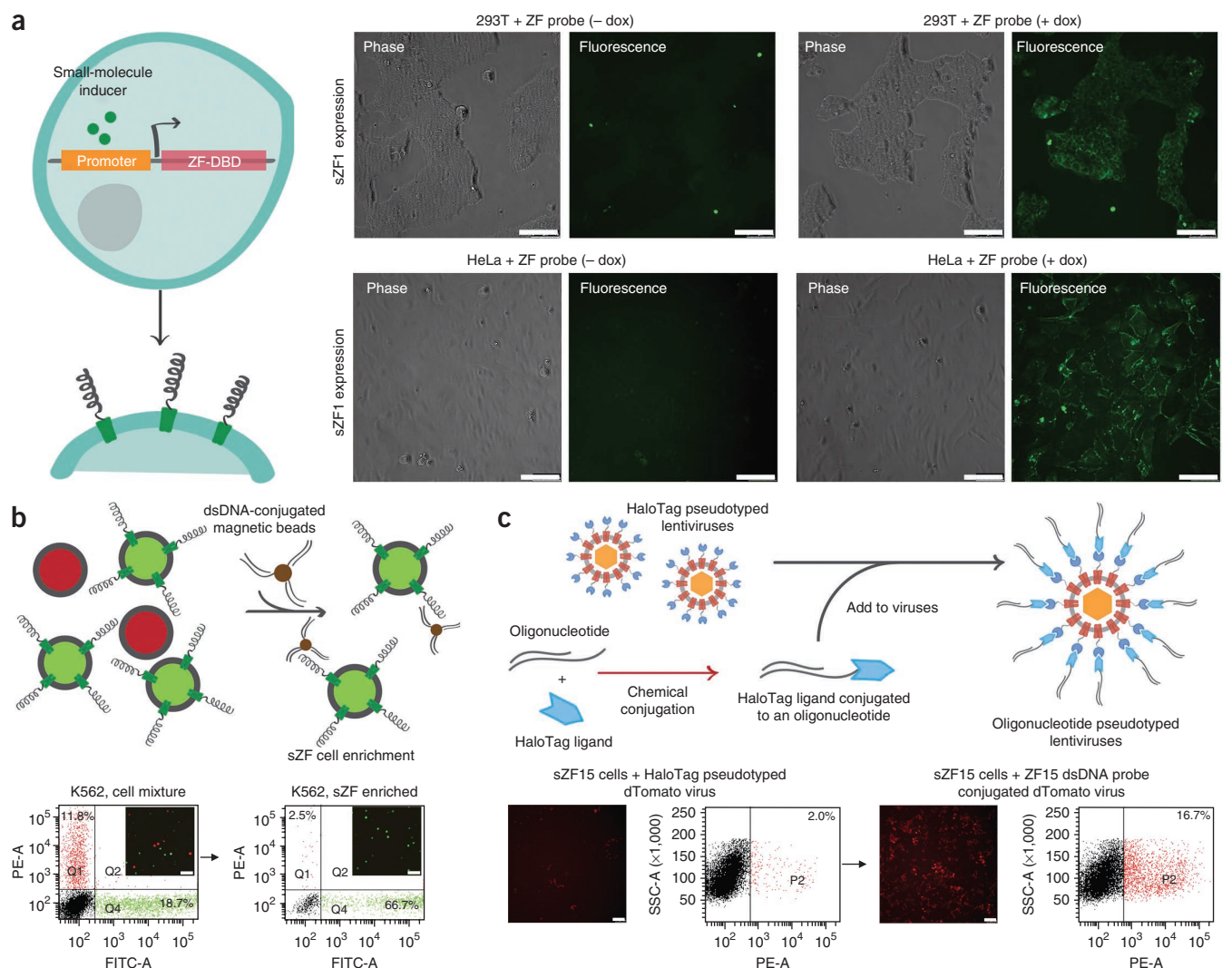
Methods and any associated references are available in the [online version of the paper](#).

**Accession codes.** Reagents developed in this study are available via Addgene (<http://www.addgene.org/browse/pi/765/articles/>).

*Note: Supplementary information is available in the online version of the paper.*

## ACKNOWLEDGMENTS

This work was supported by US National Human Genome Research Institute grant P50 HG005550. SINmu plasmid for lentivirus experiments was a gift from P. Wang (University of Southern California).



**Figure 3** | sZFs enable state probing, capture and targeting of live cells. **(a)** Schematic showing stable transduction of 293T and HeLa cells by a lentiviral vector with a tetracycline-inducible promoter to drive sZF expression. The micrographs show cells with and without small-molecule induction and after incubation with a labeled dsDNA probe. ZF-DBD, zinc-finger DNA-binding domain; dox, doxycycline (a tetracycline derivative). **(b)** Schematic showing selective enrichment of sZF expressing K562 cells (green) from a mixed population, including non-expressing cells (red), by pull-down with dsDNA-conjugated magnetic beads. The FACS plots show the relative enrichment of sZF-expressing cells after pull-down. **(c)** Schematic depicting the approach to create oligonucleotide-conjugated lentiviruses (Online Methods), and micrographs showing sZF15-expressing cells infected with control (left) or dsDNA-pseudotyped (right) lentivirus; FACS plots show quantification of this experiment. FITC and PE represent fluorochromes; SSC, side scatter. Scale bars, 100  $\mu$ m.

#### AUTHOR CONTRIBUTIONS

P.M. and G.M.C. conceived the study and designed the experiments. P.M. performed experiments. J.-H.L., D.L. and L.N. developed reagents. J.A. developed the image analysis suite and performed associated analyses. P.M., J.A. and G.M.C. wrote the manuscript with support from all authors.

#### 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>.

- McMahon, M.A., Rahdar, M. & Porteus, M. *Nat. Methods* **9**, 28–31 (2011).
- Kim, Y.G., Cha, J. & Chandrasegaran, S. *Proc. Natl. Acad. Sci. USA* **93**, 1156–1160 (1996).
- Rebar, E.J. & Pabo, C.O. *Science* **263**, 671–673 (1994).
- Volná, P. *et al. Nucleic Acids Res.* **35**, 2748–2758 (2007).
- Barbas, C.F. III. US patent 2007/0178499 A1 (2007).

- Chesnut, J.D. *et al. J. Immunol. Methods* **193**, 17–27 (1996).
- Brown, R.S. *Curr. Opin. Struct. Biol.* **15**, 94–98 (2005).
- Maeder, M.L. *et al. Mol. Cell* **31**, 294–301 (2008).
- Pattanyak, V., Ramirez, C.L., Joeng, J.K. & Liu, D.R. *Nat. Methods* **8**, 765–770 (2011).
- Boch, J. *et al. Science* **326**, 1509–1512 (2009).
- Moscou, M.J. & Bogdanove, A.J. *Science* **326**, 1501 (2009).
- Duose, D.Y. *et al. Nucleic Acids Res.* **40**, 3289–3298 (2012).
- Lei, Y., Joo, K.I., Zarzar, J., Wong, C. & Wang, P. *Virology* **7**, 35 (2010).
- Bendall, S.C. *et al. Science* **332**, 687–696 (2011).
- Livet, J. *et al. Nature* **450**, 56–62 (2007).
- Purnick, P.E. & Weiss, R. *Nat. Rev. Mol. Cell Biol.* **10**, 410–422 (2009).
- Boyce, M. & Bertozzi, C.R. *Nat. Methods* **8**, 638–642 (2011).
- Gartner, Z.J. & Bertozzi, C.R. *Proc. Natl. Acad. Sci. USA* **106**, 4606–4610 (2009).
- Delebecque, C.J., Lindner, A.B., Silver, P.A. & Aldaye, F.A. *Science* **333**, 470–474 (2011).
- Dueber, J.E. *et al. Nat. Biotechnol.* **27**, 753–759 (2009).

## ONLINE METHODS

**Plasmid construction.** The zinc-finger DNA-binding domains were synthesized as gBlocks from IDT. To express ZFs on the cell surface, we fused an immunoglobulin  $\kappa$ -chain leader sequence at their N terminus and the platelet-derived growth factor (PDGF) transmembrane domain at their C terminus (pDisplay system from Invitrogen). Additional endoplasmic reticulum import sequences based on the serotonin receptor 5HT3A and transmembrane domains from the neurokinin-1 receptor (NK1R) and  $\beta_2$  adrenergic receptor were also tried with similar success (relevant DNA fragments were cloned from NEB plasmids N9184S and N9216S). The lentiviral plasmids for inducible tetracycline expression and cumate expression were obtained respectively from Addgene (plasmids 20321, 20342) and System Biosciences (QM800A-1), and the sZF fusion constructs were directly cloned into these. Small-molecule inducers doxycycline and cumate were used at 1  $\mu\text{g}/\text{ml}$  and 30  $\mu\text{g}/\text{ml}$  concentrations, respectively. All reagents developed in this study are available via Addgene (<http://www.addgene.org/browse/pi/765/articles/>).

**Cell culture.** HEK 293T cells and HeLa cells were cultured in Dulbecco's modified Eagle's medium (DMEM), high glucose, supplemented with 10% FBS (FBS), penicillin and streptomycin (pen/strep), and non-essential amino acids (NEAA). K562 cells were cultured in RPMI-1640 medium supplemented with 10% FBS, pen/strep and NEAA. All cells were maintained at 37 °C and 5% CO<sub>2</sub> in a humidified incubator. Transfections of the sZF expressing plasmids were performed using Lipofectamine 2000 per the manufacturer's protocols (Invitrogen). For K562s, cells were resuspended in SF reagent and nucleofected according to manufacturer's instruction (Lonza). All reagents above were obtained from Gibco/Invitrogen.

**Cell labeling and DNA probes.** All cell labeling was performed in the following buffer: PBS (–CaCl<sub>2</sub>, –MgCl<sub>2</sub>) supplemented with 5% BSA (fraction V, fatty-acid free), 20  $\mu\text{M}$  ZnCl<sub>2</sub>, 1 mM MgCl<sub>2</sub> and 100  $\mu\text{g}/\text{ml}$  salmon sperm DNA. DNA probes (synthesized by IDT) had four phosphorothioate bonds on both the 5' and 3' ends to enhance protection against nucleases prevalent in extracellular medium. The fluorophores (conjugated to the probes) used for simplex and multiplex labeling experiments (Figs. 1–3, Supplementary Figs. 1, 9, 10, 12 and Supplementary Table 3) were Alexa Fluor 488, Alexa Fluor 546 and Alexa Fluor 647. The quencher-dye pairs used for the sequential labeling experiments (Fig. 2, Supplementary Fig. 11 and Supplementary Table 4) were Black Hole Quencher-1 with FAM, Iowa Black FQ with Hex and Iowa Black RQ with TYE 665. All imaging was conducted using a Leica AM TIRF MC microscope. All dsDNA probes were used at a 100–200 nM final concentration, and cells were labeled for 5–10 min with these, after which they were washed with buffer twice and subsequently imaged. Standard cell culture medium (containing 10% FBS) can also be used as a buffer to resuspend probes and successfully label cells; however, the dsDNA probes are rapidly degraded in this medium, and it is hence not ideal for long-term imaging applications. Cell-capture experiments were performed using streptavidin-coated Miltenyi beads conjugated to biotinylated dsDNA probes. The DNA arrays used in the study were synthesized using amine-conjugated oligonucleotides (synthesized by IDT) and were spotted onto epoxy-coated slides

using an Arrayit spotter. DNA arraying and slide passivation were performed per the manufacturer's instructions.

**Oligonucleotide-pseudotyped lentiviruses.** First we constructed a cell-surface HaloTag (sHaloTag) expressing the plasmid by fusing an immunoglobulin  $\kappa$ -chain leader sequence at its N terminus and a VSVG transmembrane domain at the C terminus. This plasmid was next used to produce the HaloTag protein–pseudotyped lentiviruses in 293T cells with the lentiviral and packaging plasmids transfected in the following ratio (per 150 mm cell culture dish): 15  $\mu\text{g}$  dTomato-expressing lentivirus vector, 15  $\mu\text{g}$  gag/pol plasmid, 7.5  $\mu\text{g}$  SINmu plasmid and 7.5  $\mu\text{g}$  of the above sHaloTag plasmid. Next, the HaloTag ligand succinimidyl ester (O4) building block was conjugated to an amine group bearing the DNA probe. Finally, the harvested lentiviruses were conjugated to the above sZF-specific dsDNA probes through the HaloTag protein–HaloTag ligand interaction (Promega), yielding the desired oligonucleotide-pseudotyped lentiviruses.

**Image processing and statistics.** With the goal of quantitating sZF behavior and specificity, JPG images acquired as single  $z$  slices from fluorescence confocal microscopy were processed using three in-house–developed user-interactive Matlab (MathWorks) applications whose use is depicted in **Supplementary Figure 3**. ImageNormalizer presents options for normalization of intensities and background subtraction in each channel to produce a standardized multichannel TIF image used by other applications. ImageMasker enables users to identify image regions containing cell debris and dead cells that are to be excluded from subsequent analysis. SegmentOverlapAnalysis gives users interactive control over parameters used to identify regions of the images occupied by cells in each channel (i.e., image segmentation) and then allows the user to submit the segmented image for statistical analysis. Statistical analysis is performed online by a fourth Matlab function invoked by SegmentOverlapAnalysis but is not itself interactive. In brief, for the images analyzed here, normalization entailed intensity clipping at the 99.6th–99.8th percentile intensity in each channel with no subsequent background subtraction, an average of  $37.0 \pm 12.7$  masks were created to mask out a total of  $2.94\% \pm 1.43\%$  of image area per image and  $47.1 \pm 9.8$  segments (**Supplementary Table 2**) were generated per channel (numbers are mean  $\pm$  s.d. in all cases). Details can be found in **Supplementary Notes 1–3**. To mitigate a potential bias in measuring sZF specificity by direct comparisons of intensity across channels caused by our use of intensity thresholds to segment images, we developed additional measures based on intensity correlations across channels and overlaps between segments of different channels. In brief, if sZFs are specific, correlations should be negative and large numbers of segments should be seen in each channel that do not overlap segments from other channels. Data substantiating these and other quantitative observations relevant to sZF behavior and specificity are summarized in **Supplementary Table 2** and **Supplementary Figures 4–8**. All actual processing and statistical data files and figures generated by our image processing applications for the images, and the Matlab image processing applications themselves along with instructions on their use, have been made freely available to the research community for noncommercial research on our website ([http://arep.med.harvard.edu/sZF\\_cell\\_barcode/](http://arep.med.harvard.edu/sZF_cell_barcode/)).