



Scalable biological signal recording in mammalian cells using Cas12a base editors

Hannah R. Kempton¹, Kasey S. Love¹, Lucie Y. Guo^{1,2} and Lei S. Qi^{1,3,4}✉

Biological signal recording enables the study of molecular inputs experienced throughout cellular history. However, current methods are limited in their ability to scale up beyond a single signal in mammalian contexts. Here, we develop an approach using a hyper-efficient dCas12a base editor for multi-signal parallel recording in human cells. We link signals of interest to expression of guide RNAs to catalyze specific nucleotide conversions as a permanent record, enabled by Cas12's guide-processing abilities. We show this approach is plug-and-play with diverse biologically relevant inputs and extend it for more sophisticated applications, including recording of time-delimited events and history of chimeric antigen receptor T cells' antigen exposure. We also demonstrate efficient recording of up to four signals in parallel on an endogenous safe-harbor locus. This work provides a versatile platform for scalable recording of signals of interest for a variety of biological applications.

The ability to convert transient signals encountered by cells into a permanent record in a scalable fashion has powerful implications for both understanding fundamental biological processes and developing the next generation of cell-based therapies. Molecular signal recording uses DNA-modifying enzymes to link biological cues of interest to targeted changes in DNA, which can be read out long after the time of exposure as a genetic archive of cellular history^{1–3}. Successful implementation of signal recording circuits in mammalian cells could be instrumental for studying events that occur in inaccessible locations or that are difficult to monitor in real time, including the progression of infection or immune cell differentiation *in vivo*. It could also be beneficial in a therapeutic context, through cells that provide a history of disease-relevant cues encountered within the body⁴. Many of these applications, however, are complex and multifactorial, and would require a highly scalable technology that is able to record multiple signals in parallel to fully understand disease context or signaling dynamics of interest.

Initial work demonstrated the capacity to permanently record biological signals in both prokaryotic and eukaryotic cells, applying autoregulatory transcriptional circuits^{5,6} or using recombinases^{7–10}. The development of CRISPR–Cas systems for programmable DNA manipulation has made it substantially easier to generate targeted DNA changes, facilitating their rapid application for recording cellular history. This has been done through Cas1/Cas2 sequence integration^{11–13}, or linking Cas9 activity to signals of interest^{14–18}. Although some of these approaches recorded multiple signals, they used components that are difficult to translate into mammalian contexts, such as Cas1/Cas2, or focused on artificial or bacterial-specific inputs^{2,15,16}. Therefore, despite these advances, there remains a need for robust recording of multiple biologically relevant signals in human cells.

To address these difficulties, we developed an approach using the type V Cas12a (Cpf1)¹⁹, which has been underused in the signal recording space. This Cas variant differs from the most widely used Cas9 protein in that it can process its own mature CRISPR RNA (crRNA) from a longer transcript^{20,21}. Because of this processing ability, Cas12a crRNAs can be controlled from the wide variety of inducible polymerase II (Polii) promoters^{22–24}. In this way, signals of interest

can be linked to the expression of a certain guide, which will make a programmed mutation in the cell's DNA. Here, we demonstrate the implementation of dCas12a base editors with inducible guides for highly scalable molecular recording in mammalian cells. Our design can be modified to include other effector functionality or logic gating, as well as easily extended to record multiple signals in parallel.

Results

Designing a fluorescent reporter for base editing. In establishing a Cas12a system for recording biological signals, we opted to use a nuclease-dead Cas12a (dCas12a) fused with a base editor^{25,26} rather than catalytically active Cas12a. Base editors enable targeted single-nucleotide resolution edits that can be multiplexed within a narrow window of DNA, and nickase-Cas9 base editors have been used successfully for signal recording^{15,16}. The dCas12a base editor should allow facile coupling of multiple inputs to crRNAs in parallel without inducing double-stranded breaks at multiple loci, which could induce DNA damage response or chromosomal translocations^{27,28}.

The initially developed dCas12a base editors had limited efficiency, especially compared to Cas9 versions, as there is no known nickase version of dCas12a that would enhance targeted nucleotide conversion^{29,30}. A recently evolved adenine base editor (ABE8e) demonstrated markedly increased A→G conversion compared to previous effectors when fused to dCas12a, which we hypothesized would enable robust signal recording³¹. We opted to use the nuclease-dead (D832A) ortholog from *Lachnospiraceae* bacterium (LbdCas12a), as for most published sites it worked equivalent to or better than AsdCas12a or denAsCas12a (refs. ^{31,32}).

We first set out to design a fluorescent reporter for dCas12a-mediated base editing, to induce green fluorescent protein (GFP) expression following an A→G (or, T→C depending on the targeted strand) conversion (Fig. 1a). We created four candidate constructs in which successful base editing would restore functional protein expression (Supplementary Fig. 1a,b) and tested their performance by transfecting the pairs of GFP reporters with corresponding guides and dCas12a-ABE8e in human embryonic kidney 293T (HEK293T) cells. Design 3, in which the base

¹Department of Bioengineering, Stanford University, Stanford, CA, USA. ²Department of Ophthalmology, Stanford University, Stanford, CA, USA.

³Sarafan ChEM-H, Stanford University, Stanford, CA, USA. ⁴Chan Zuckerberg BioHub, San Francisco, CA, USA. ✉e-mail: Stanley.qi@stanford.edu

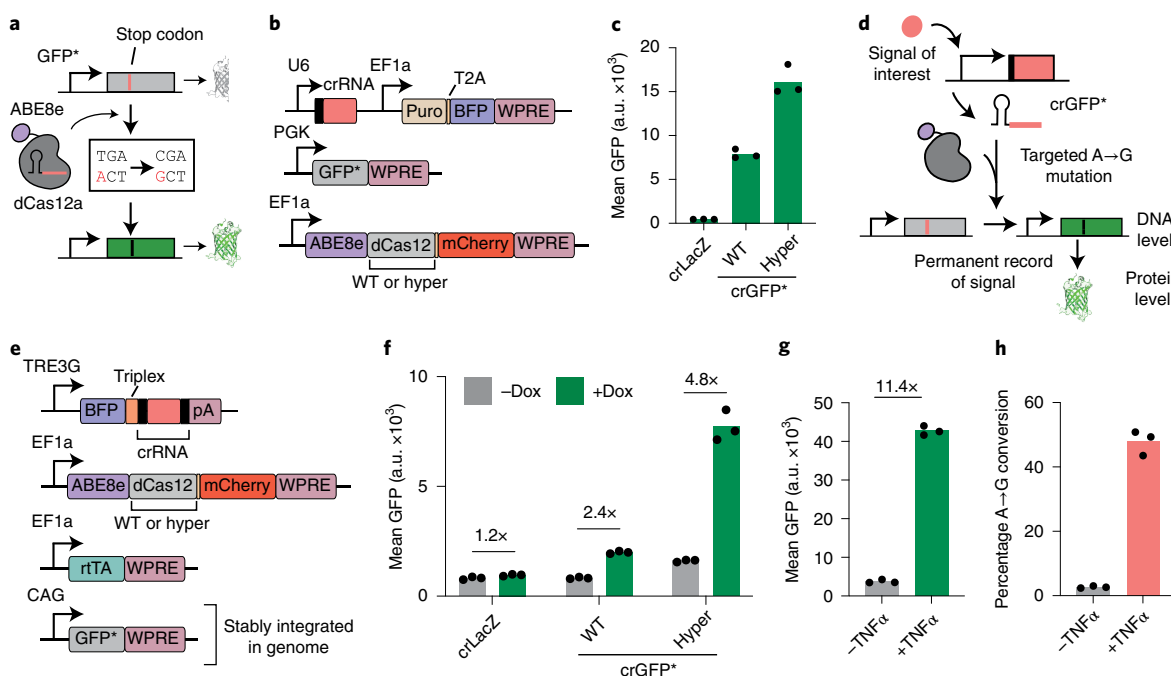


Fig. 1 | Base editing with hyperCas12a as a platform for signal recording. **a**, Schematic showing GFP*, the reporter used to generate a fluorescent readout for an adenine base editing event. **b**, Constructs transfected in HEK cells. **c**, Mean fluorescence as measured by flow cytometry 3 days posttransfection with GFP* reporter, U6-driven guide and wild-type (WT) or mutant (hyper) dCas12a-ABE. crLacZ, nontargeting control. a.u., arbitrary units. **d**, A simplified cartoon showing the use of an inducible guide with GFP* to record cellular history. **e**, Constructs used for dox-inducible signal recording. An additional plasmid with rtTA was included for proper function of the TRE3G promoter. **f**, Mean fluorescence as measured by flow cytometry for GFP* HEK cells transfected with the recording constructs in **e**, 3 days posttransfection and stimulation with dox. **g**, Mean fluorescence as measured by flow cytometry for GFP* HEK cells transfected with hyperdCas12a base editor and NFKB-driven crGFP*, 3 days posttransfection and stimulation with TNFα. **h**, Quantification of A→G mutations for the NFKB recorder at the targeted stop codon in the GFP* locus as measured by MiSeq. All data shown are for three independent replicates. Statistical analysis was performed using a two-sided t-test without adjustment for multiple comparisons, and P values are included in Supplementary Table 4.

editor corrected an introduced premature stop codon at residue 122, showed a 19-fold increase in GFP expression compared to nontargeting crLacZ control (Supplementary Fig. 1c, Supplementary Fig. 2a–c and Supplementary Tables 1–3). We termed this reporter and guide design GFP* and used it for experiments moving forward.

To optimize performance and obtain the largest dynamic range, we tested whether a recently developed LbCas12a mutant, termed hyperCas12a, could enhance base editing³³. This mutant was rationally engineered by introducing four additional positively charged amino acids (D156R/D235R/E292R/D350R) near the nucleic acid binding pocket, to strengthen the interaction with negatively charged crRNA and target DNA. With our GFP* reporter, we compared the original wild-type dCas12a to the hyperdCas12a mutant and saw a significant increase in both mean GFP fluorescence and percentage GFP* cells, indicating hyperdCas12a-ABE8e increased base editing activity (Fig. 1b,c, Extended Data Fig. 1a and Supplementary Table 4).

Inducible guides as a platform for signal recording. To have the induced A→G mutation be a permanent record within the cell we stably integrated the GFP* reporter locus using lentivirus, generating the GFP* HEK cell line. We confirmed that the stably integrated GFP* reporter cells still activated GFP expression when transfected with crGFP* and base editor (Supplementary Fig. 3a–c). Additionally, by extracting the genomic DNA and sequencing the GFP* locus using next-generation sequencing as previously described^{34,35}, we confirmed that the protein-level signal corresponded to a DNA-level A→G mutation (Supplementary Fig. 3d,e and Supplementary Tables 5 and 6).

To implement the system for biological recording, we transitioned from U6 to a polymerase II (Polii) promoter to drive expression of the crRNA. In this design, the signal of interest induces transient expression of the transcript containing the crRNA, which can be processed by dCas12a-ABE and used to generate a targeted mutation as a permanent record of the cell's history (Fig. 1d). We placed crGFP* under the control of the doxycycline (dox)-inducible TRE3G promoter; additionally, we added blue fluorescent protein (BFP) to directly measure the activity of the inducible promoter and a triplex sequence to ensure transcript stability^{24,36} (Fig. 1e). On transfecting GFP* HEK cells with base editor and inducible guide, there was induction of BFP after adding dox in all conditions, confirming similar expression of the transcript containing the crRNA (Supplementary Fig. 4a). However, the performance of the induced guide and base editor to permanently record the presence of dox into the GFP* DNA locus was significantly different between the wild-type and mutant hyperdCas12a (Fig. 1f, Extended Data Fig. 1b and Supplementary Fig. 4b). Using hyperdCas12a drastically improved the dynamic range for inducible Polii crRNAs, enabling robust signal recording not possible with wild-type dCas12a. This hyperCas12a base editor was used for all recording experiments moving forward.

Next, we showed that this approach could be expanded to record exposure to other, biologically relevant signals. We changed the promoter from TRE3G to an nuclear factor kappa B- (NFKB-) inducible promoter, to record the cell's history of inflammation through the NFKB pathway³⁷. On transfection with base editor and guide and induction with TNFα, the GFP* HEK cells had a 11.4-fold increase in GFP expression (Fig. 1g, Extended Data Fig. 1c

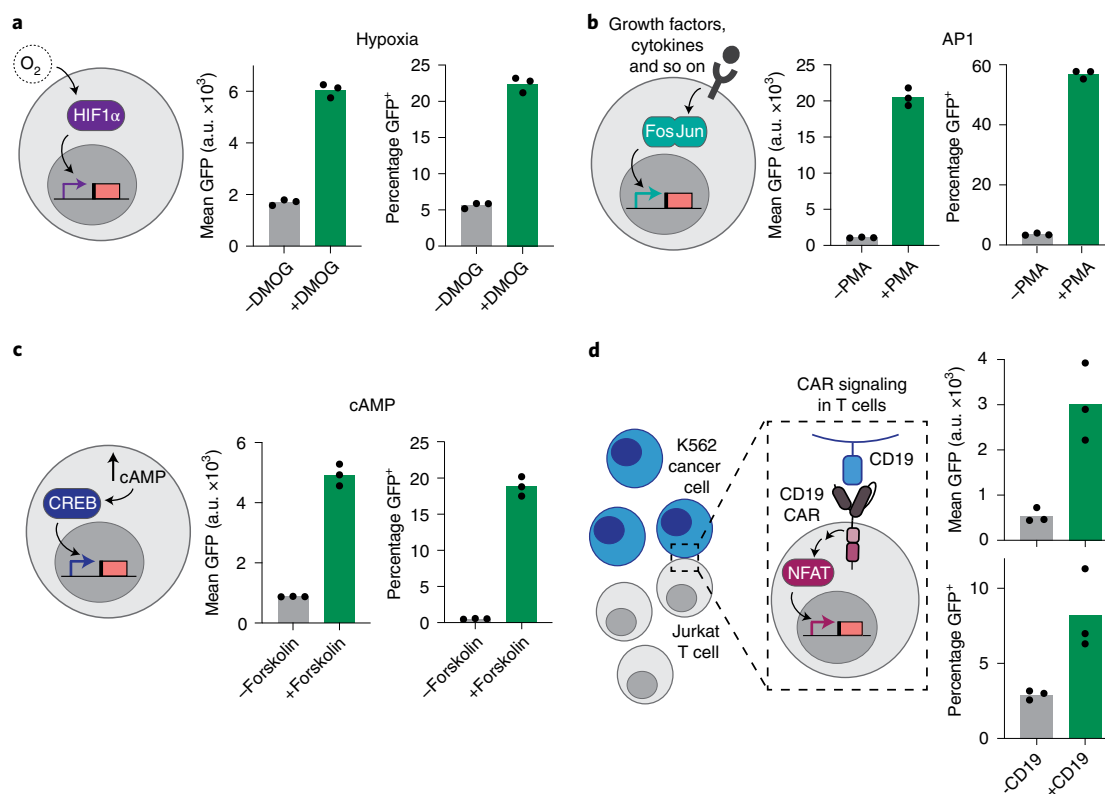


Fig. 2 | Modular recording of diverse biological signals. **a**, Cellular memory of hypoxia, stimulated by DMOG. **b**, Cellular memory of AP1 signaling, stimulated by PMA. **c**, Cellular memory of elevated cAMP, stimulated by forskolin. All experiments in **a–c** shown in GFP⁺ HEK cells stably transduced with respective inducible guide, 3 days posttransfection with base editor and induction with relevant signal. **d**, Recording circuit of CD19 antigen encounter and resulting T cell activation via CAR signaling in Jurkat cells. GFP fluorescence was measured after 2 days of coculture with $\pm$ CD19 K562 cells. All data shown are for three independent replicates. Statistical analysis was performed using a two-sided *t*-test without adjustment for multiple comparisons, and *P* values are included in Supplementary Table 4.

and Supplementary Fig. 4c,d). This record of the inflammatory event was further confirmed as a permanent change at the DNA level—on exposure to TNF α , 48% of loci had the targeted A $\rightarrow$ G mutation (Fig. 1h and Supplementary Fig. 4e,f). The small amount of leakiness of GFP ‘memory’ without input TNF α is due to basal expression of the NFkB promoter in unstimulated cells, which can be observed when comparing BFP expression among cells with or without guide plasmid and TNF α (Supplementary Fig. 4g).

For further characterization of dynamic range, we transferred the inducible guide construct into a lentiviral backbone so it could also be stably integrated (Extended Data Fig. 2a). Using a duplicate spacer-direct repeat cassette, such that two crRNAs could be processed from a single messenger RNA, further improved reporter response (Extended Data Fig. 2b)²³. This suggests the amount of available processed guide is a limiting factor in the system. Overall, comparing the same NFkB-inducible promoter for both GFP fluorescence and percentage A $\rightarrow$ G editing, the stably integrated lentiviral guide construct had lower leakiness in the OFF-state, but also a lower signal in the ON-state compared to the previous transfected version, and thus may be more suitable in certain applications (Fig. 1g,h and Extended Data Fig. 2c,d).

Multiplexed recording of two input signals. Using dCas12a with an inducible guide for signal recording enables flexible scale-up beyond a single input through multiple guides that can be independently controlled in parallel. For initial testing of dual recording, we designed an additional panel of GFP base editing constructs and identified GFP⁺ design 5 as another successful reporter

(Supplementary Fig. 5a,b). To generate a dual signal recorder, we combined the previous design 3 and the new design 5 into a novel fluorescent reporter, termed GFP^{**}. This dual recording reporter has two distinct premature stop codons, which can each be mutated out independently by their distinct crRNA (Extended Data Fig. 3a,b). Thus, detection of GFP fluorescence indicates that at some point in its cellular history, the cell population encountered both signals of interest.

To use GFP^{**} for stable signal recording, we transduced the reporter into HEK cells. We selected dox and TNF α as our two signals of interest, and placed crGFP^{*}3 and crGFP^{*}5 under the control of TRE3G and NFkB-inducible promoters, respectively (Extended Data Fig. 3c). We stably integrated these constructs into GFP^{**}HEK cells, transfected with base editor and exposed cells to combinations of dox and TNF α . We observed an increase in fluorescence when cells experienced both signals, with a 14-fold increase in GFP signal relative to unstimulated cells and confirmed that this induction of GFP signal corresponded to A $\rightarrow$ G mutations in both stop codons (Extended Data Fig. 3d–f).

Modular recording of diverse biological signals. To test the modularity of the system, we expanded to record a panel of additional signals of various types by swapping out the inducible promoter controlling the guide cassette (Extended Data Fig. 4a). We constructed individual recording circuits for hypoxia, cAMP signaling, nuclear factor of activated T cells (NFAT) signaling, AP1 signaling or oxidative stress—a panel of diverse pathways including secondary messengers, immune cell activation and local

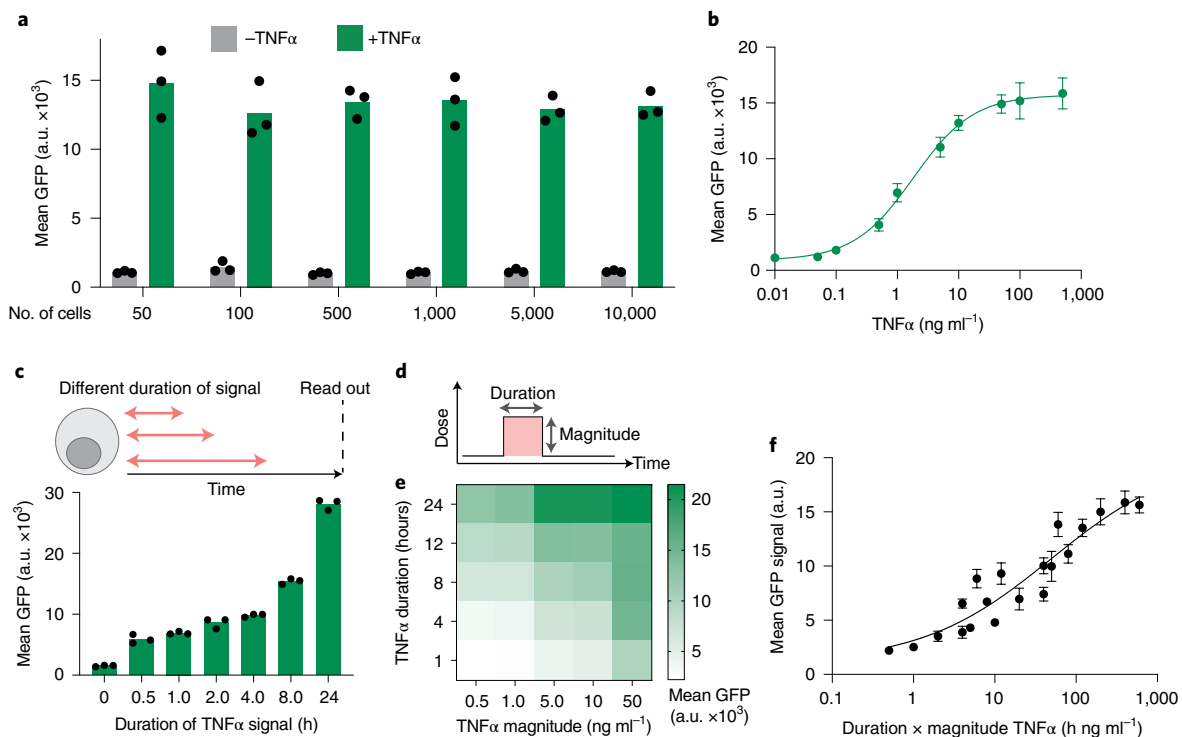


Fig. 3 | Recording of analog signal characteristics. **a**, GFP⁺ HEK cells stably transduced with lentiviral NFκB-guide and transfected with base editor were stimulated with 100 ng ml⁻¹ TNFα for 3 days, then analyzed via flow cytometry. Randomly downsampled subpopulations of cells of the indicated size were analyzed for progressively smaller cell numbers. **b**, Dose response for GFP⁺ HEK cells stably transduced with lentiviral NFκB-guide and transfected with base editor. Mean fluorescence is shown 3 days posttransfection and stimulation. **c**, Top, schematic showing cells exposed to different durations of signal. Bottom, mean GFP fluorescence 3 days posttransfection for NFκB-recording GFP⁺ HEK cells exposed to 10 ng ml⁻¹ of TNFα for varying lengths of time, h, hours. **d**, Schematic showing signal input graph, with a pulse of input defined by both signal magnitude and signal duration. The shaded pink area indicates the product of the two parameters, or AUC. **e**, Heatmap showing mean fluorescence for GFP⁺ HEK cells transduced with NFκB-guide, transfected with base editor, then stimulated with varying doses and times of TNFα. Flow cytometry was performed 3 days posttransfection. **f**, The mean GFP signal from **e** plotted against the product of the TNFα signal duration × magnitude. The solid curve indicates four parameter logistic regression. All data shown are for three independent replicates, and all error bars indicate standard deviation. Statistical analysis was performed using a two-sided *t*-test without adjustment for multiple comparisons, and *P* values are included in Supplementary Table 4.

environmental stress^{38–42}. For all five signals, we generated stable cell lines in GFP⁺ HEK cells and detected a significant increase in both GFP fluorescence and percentage of GFP⁺ cells on transfection with base editor and stimulation with relevant signal, demonstrating the ability to sense a versatile range of inputs (Fig. 2a–c and Extended Data Fig. 4b,c).

Recording of cell–cell interactions with tumor antigen. In addition to these soluble inputs, the ability of the circuit to record encounters with a given cell-surface protein would be of value for many applications. Therefore, we placed the guide cassette under control of a SynNotch receptor specific for CD19, a widely targeted cancer antigen⁴³. GFP⁺ HEK cells were transfected with CD19-tTA SynNotch and TRE3G-driven guide, and then cocultured with a second HEK cell line engineered with or without CD19. In the presence of CD19 antigen, there was an increase in GFP signal, indicating permanent memory of cell–cell interactions (Extended Data Fig. 5a,b).

To demonstrate this recording in a more clinically relevant setting we used a chimeric antigen receptor (CAR), which provides T cells with the ability to activate in response to a specific antigen⁴⁴. We introduced the base editor, CD19 CAR, GFP⁺ reporter and NFAT-inducible guide into Jurkat T cells and cocultured them with CD19⁺K562 cell lines (Fig. 2d and Extended Data Fig. 5c). On sensing CD19, the Jurkat cells induced downstream CAR signaling and activated the NFAT pathway, and the resulting guide expression

edited the GFP⁺ reporter (Fig. 2d and Extended Data Fig. 5d). This provided a proof of principle implementation of a circuit recording that an engineered CAR-T cell encountered and was activated by the antigen of interest.

Characterizing dCas12a analog signal recording. At a single GFP⁺ reporter, there is a binary response with only two states, the original A or mutated G. However, each cell can have more than one copy of the reporter in its genome, and we are reading out information from multiple cells in the population. Thus, we wanted to see whether this system could read out a more nuanced history than simply the digital exposure to a signal. First, before exploring the system parameters we characterized the robustness of the signal for different cell population sizes. For the NFκB-recording circuit, we implemented random downsampling to progressively decrease the number of cells used for analysis following flow cytometry and observed a consistent signal down to as few as 50 cells (Fig. 3a). For even smaller populations, the variability between replicates was high due to the small sample sizes, which would obfuscate analog signal features from biological differences (Extended Data Fig. 6a,b).

Next, we stimulated NFκB-recording cells with a titration of TNFα concentrations, ranging from 0.01 to 500 ng ml⁻¹. The cells exhibited clear dose-response behavior for both BFP (indicating how strongly the cell turned on the NFκB promoter) and GFP (indicating the permanent record of the input) (Fig. 3b and Extended

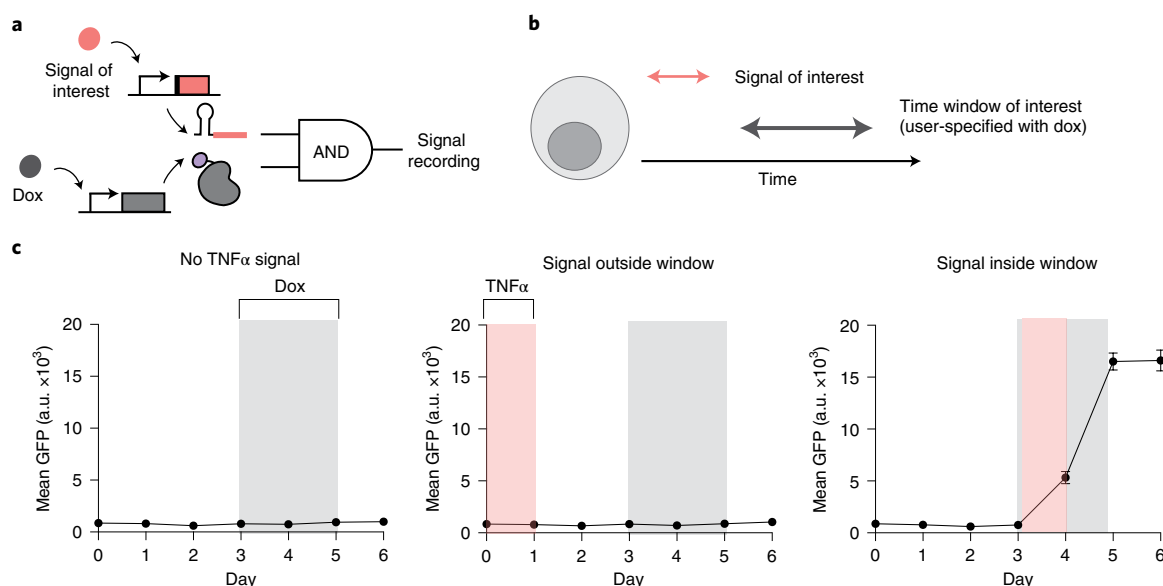


Fig. 4 | Temporal resolution of cellular history within recording window. **a**, Schematic showing AND-gate behavior of recording circuit, requiring simultaneous expression of guide and Cas protein effector to execute base editing. **b**, Schematic showing the concept of a recording window in which externally provided dox specifies the time period in which the user is interested in observing whether or not the biological signal of interest occurred. **c**, Time-course data of mean GFP as measured by daily flow cytometry for GFP⁺ HEK cells stably transduced with NFκB-crGFP⁺ and dox-inducible base editor. Dox was added from day 3–5 to define the recording window of interest. Left, no TNFα signal was added. Middle, TNFα was added from day 0 to 1. Right, TNFα was added from day 3 to 4. Data shown for three independent replicates, error bars indicate standard deviation. Statistical analysis was performed using a two-sided *t*-test without adjustment for multiple comparisons, and *P* values are included in Supplementary Table 4.

Data Fig. 6c,d). This suggests the dCas12a circuit can provide analog information about the magnitude of input encountered—in this case, the degree of inflammation experienced. Additionally, for a given concentration we wanted to determine whether the system could distinguish between different durations of exposure. At time 0, all cell populations were stimulated with 10 ng ml⁻¹ of TNFα. After the indicated times, the signal was removed and replaced with normal media for the remainder of the experiment (Extended Data Fig. 6e). The resulting GFP signal increased with increasing time of exposure (Fig. 3c)—for the NFκB-recording cells, this could enable distinction between acute and chronic inflammatory signals of a given magnitude.

To explore the interplay between the duration and magnitude of input signal, we performed an additional TNFα recording experiment simultaneously varying both parameters (Fig. 3d,e). For each stimulation condition, we then calculated the area under the curve (AUC) of the input signal graph, multiplying the signal duration by magnitude to obtain the overall TNFα exposure. We plotted these values against the resulting GFP fluorescence, excluding the 24-hour time point at which the system became quickly saturated (Fig. 3f). By applying four parameter logistic regression, we generated a sigmoidal curve that could be used to decode the input duration × magnitude from an observed GFP readout. However, with a single guide it would not be possible to distinguish two input signal pulses with different individual parameters but the same AUC.

Temporal sensitivity within a given recording window. An advantage of using dCas12a's processing ability for inducible crRNAs is that the guide and Cas protein can both be controlled independently. The CRISPR–Cas effectors are naturally a two-component system (AND gate), where Cas protein and guide must both be expressed at the same time to act on target DNA (Fig. 4a). We took advantage of this to introduce temporal sensitivity to the signal recording system. Rather than expressing dCas12a base editor constitutively, we placed it under the dox-inducible promoter. In this way, we can record not

just whether cells experienced a given biological signal of interest, but specifically whether the signal occurred during a given time period of interest in which the dCas12a effector is also expressed (Fig. 4b). For our purposes, this 'recording window' of interest is specified by the user through dosing of the small molecule dox.

We generated a new GFP⁺ HEK cell line with stably introduced TRE3G-dCas12a-ABE and the NFκB-inducible guide, then confirmed that we needed to turn on both dCas12a and crRNA using dox and TNFα, respectively, to see GFP (Supplementary Fig. 6a,b). To demonstrate the temporal dynamics of the system, we performed a time-course experiment in which we set our recording window to days 3–5 by adding dox to the media only during those days. With our circuit, we were able to determine whether our cell population experienced an inflammatory event specifically during the time period we were interested in, enabling a more sophisticated record of cellular history (Fig. 4c).

Dual-function effector and recorder cells. Because dCas12a can process its crRNA from a longer coding transcript, we can also include open reading frames on the inducible construct to be expressed with the guide. This allows for addition of effector molecules beyond the current BFP to combine the recording module with sense-and-respond circuits. In this way, an engineered cell could act for dual purposes—sensing of a disease-relevant signal would simultaneously turn on a transient effector to act therapeutically for the duration of the signal, as well as a crRNA to make a permanent record of the cell's history of that event (Fig. 5a). Sense-and-respond therapeutic cells have been of increasing interest in many fields, including for autoimmune diseases characterized by flare-ups of inflammatory signals^{45,46}. Thus, we constructed an inflammation sense-and-respond cell with memory functionality, modifying our NFκB-recording device to include antiTNF, an established biologic for autoimmune therapeutics⁴⁷. We incorporated an antiTNF nanobody upstream of the crRNA⁴⁸, adding an IgK signal peptide for secretion⁴⁹ and a His tag for detection.

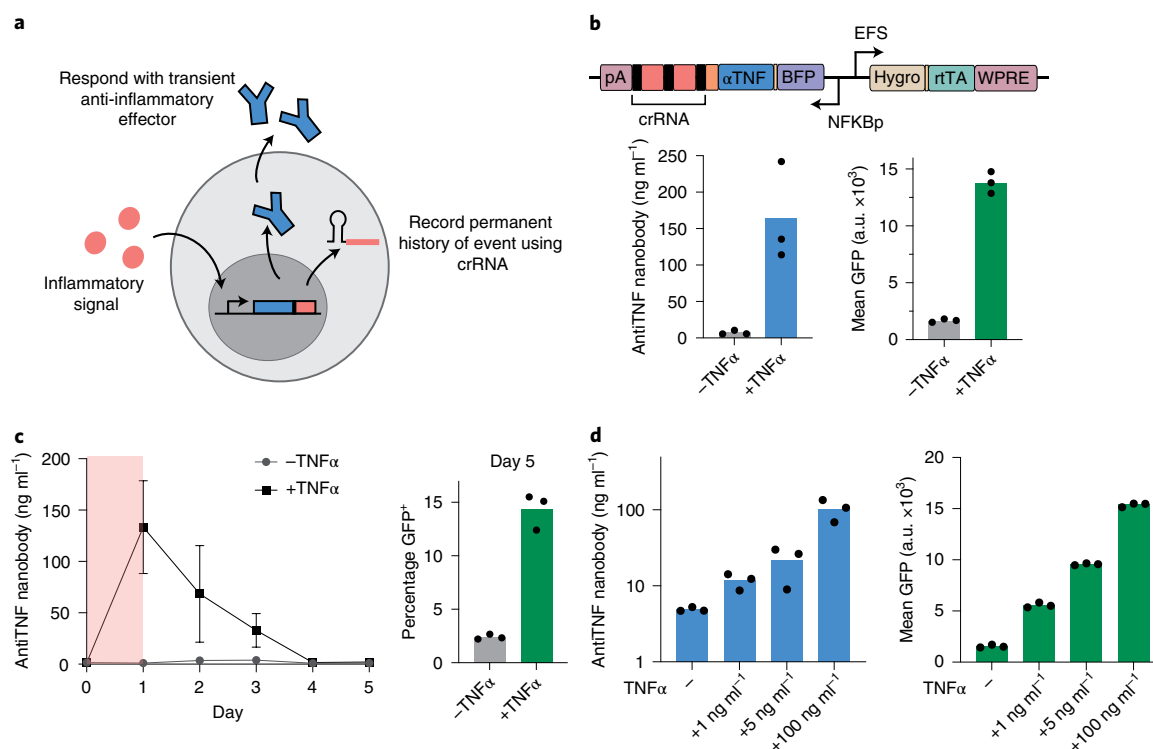


Fig. 5 | Dual-function effector and recorder cells. a, Schematic of engineered sense-and-respond cell with memory functionality. An inflammatory event induces expression of a single transcript, containing both a protein-coding effector to combat inflammatory state, as well as the crRNA that will make a permanent DNA change to record the inflammatory event. **b**, GFP⁺ HEK cells were stably transduced with inducible effector/guide (top), then transfected with base editor and stimulated with TNFα. Left, concentration of antiTNF nanobody as measured by anti-His ELISA. Right, mean GFP as measured by flow cytometry. **c**, Left, time course showing mean antiTNF nanobody expression as measured by ELISA. Error bars indicate standard deviation. Right, percentage of GFP⁺ cells as measured by flow cytometry on day 5. **d**, Dose-response behavior of memory/effector cells, one day after a 24-hour stimulation with TNFα. Left, concentration of antiTNF nanobody. Right, mean GFP. Data for all experiments for three independent replicates, with ELISA data additionally performed in technical duplicate. Statistical analysis was performed using a two-sided *t*-test without adjustment for multiple comparisons, and *P* values are included in Supplementary Table 4.

On stimulating cells with TNFα, we detected an average of 164 ng ml⁻¹ of antiTNF nanobody using enzyme-linked immunosorbent assay (ELISA), as well as dCas12a-mediated record of TNFα signal through the base edited GFP⁺ (Fig. 5b). By stimulating cells for 24 hours, then removing the TNFα, we tracked the temporal behavior of the circuit. Once the cell was no longer experiencing an inflammatory event, the effector nanobody concentration quickly decreased. At day 5, the experimental endpoint, the nanobody concentration decreased to baseline but the cells still maintained a permanent record of that inflammatory event, as indicated by GFP (Fig. 5c). We further demonstrated the sensitivity of the system to different input magnitudes. After exposing the cells to a range of increasing TNFα inputs, we observed that a higher concentration of input signal resulted in a higher nanobody concentration, to combat the more intense inflammatory event, as well as a higher GFP signal as a record of the signal (Fig. 5d).

Simultaneous recording of multiple signals in parallel. Although the GFP⁺ reporter discussed previously has use for dual recording, it is difficult to know information about individual signals from the fluorescent output and each additional signal requires considerable reporter redesign. Thus, for more scalable multi-input recording we transitioned our system to fully sequencing-based readout, with signals being recorded in the endogenous AAVS1 safe-harbor locus⁵⁰. In this way, increasing number of signals can be recorded in parallel by linking each additional signal to a new crRNA targeting

the AAVS1 locus that will execute a distinct A→G mutation. After sequencing the locus, the state of each targeted A will provide the cellular history of each signal of interest (Fig. 6a). To define a shorter AAVS1 scratchpad to read within a single next-generation sequencing run, we looked for subregions with the highest density of TTTV protospacer adjacent motifs (PAMs). We tested a set of U6-driven AAVS1 guides and selected the final recording scratchpad based the highest density of best performing guides (Supplementary Fig. 7a,b).

We next verified that this AAVS1 locus could be used for inducible guide mediated signal recording. We put the best guide, crAAVS1_4, under the control of the NFKB promoter and transfected the guide construct with base editor in HEK293T cells. The targeted protospacer had two As in the editing window, A11 and A14 relative to PAM. On exposing cells to TNFα, there was a significant increase in A→G% for both target As, with A11 going from an average 14% A→G in unstimulated cells to 68% in stimulated cells and A14 going from 4 to 44%, indicating the cells were successfully recording their inflammatory history (Extended Data Fig. 7a,b). By electroporating the NFKB-crAAVS1_4 construct into Jurkat T cells expressing the base editor, we showed this approach can also successfully record inflammatory signal in a more clinically relevant cell type (Extended Data Fig. 7c,d).

We then set out to independently record the cells' history of two biologically relevant pathways related to infection and inflammation, NFKB and interferon regulatory factor (IRF)^{51,52}. To construct

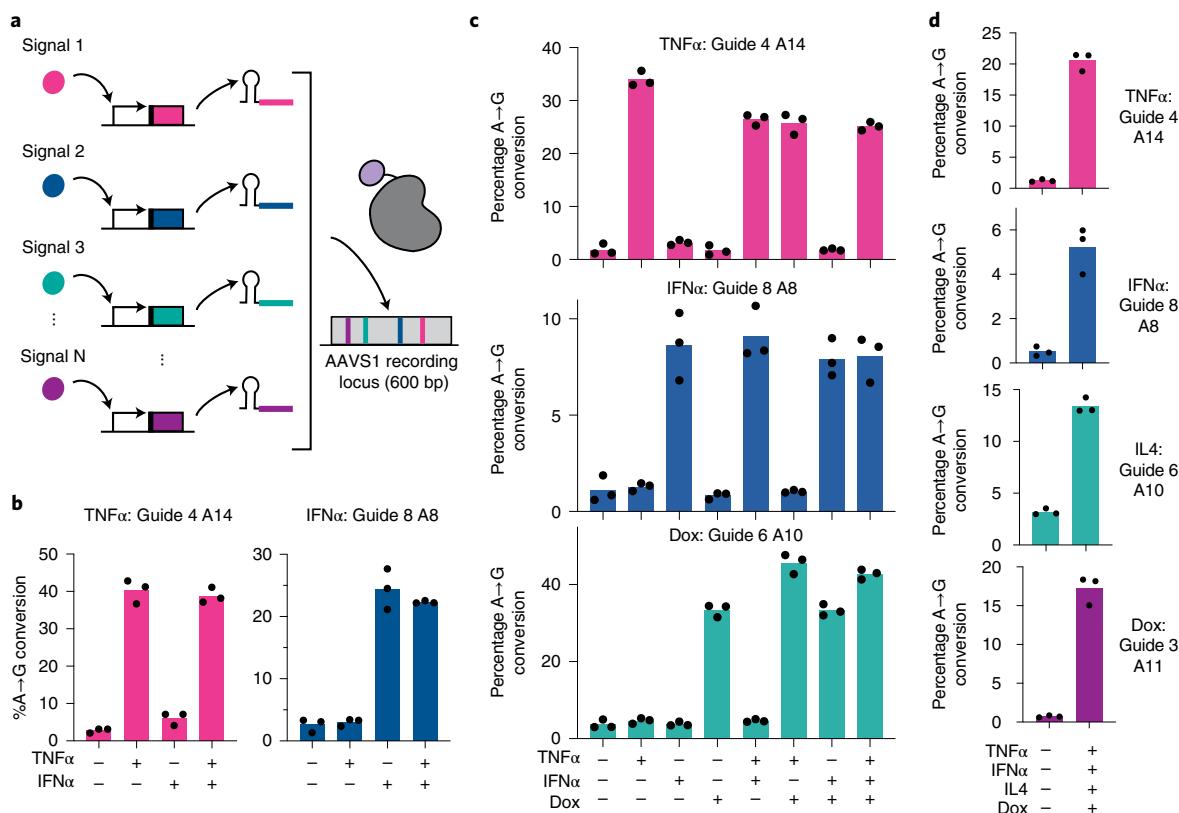


Fig. 6 | Simultaneous recording of multiple signals in parallel. **a**, Schematic showing highly scalable, parallel recording strategy into the endogenous AAVS1 safe-harbor locus. **b**, Dual-input recording at the AAVS1 locus in HEK cells transfected with base editor, NFKB-driven crAAVS1_4 and IRF-driven crAAVS1_8. Cells were stimulated with TNF α and/or IFN α for 3 days, followed by quantification of base editing at the specified A (number indicates base position relative to PAM) using MiSeq. **c**, Triple-input recording in HEK cells transfected with base editor, rTA and NFKB, IRF and TRE3G inducible guides. Cells were stimulated with combinations of TNF α , IFN α and/or dox for 3 days. **d**, Parallel recording of four input signals in HEK cells, using NFKB, IRF, TRE3G and STAT6 inducible guides. All data shown are for three independent replicates. Statistical analysis was performed using a two-sided *t*-test without adjustment for multiple comparisons, and *P* values are included in Supplementary Table 4.

the IRF-inducible construct, we generated an inducible promoter with interferon stimulated response elements, increasing the number of interferon stimulated response elements sites to get higher induction in HEK cells (Supplementary Fig. 8a). We transfected HEK cells with NFKB-inducible crAAVS1_4 and IRF-inducible crAAVS1_8, along with base editor, and stimulated with combinations of the respective orthogonal inputs TNF α and IFN α (Supplementary Fig. 8b). The frequency of A→G mutations at the respective sites within the amplicon indicated whether the cell activated neither, one or both pathways (Fig. 6b). Comparing the unstimulated to dual stimulated cells, Guide 4's targeted A14G increased from 2.8 to 39%, and guide 8's targeted A8G increased from 2.6 to 22%. Thus, these engineered cells enabled the robust detection of the history of two inflammatory signals in parallel through next-generation sequencing of the single AAVS1 amplicon.

To further demonstrate the scalability of the system for recording multiple signals, we placed a third AAVS1 guide, crAAVS1_6, under control of the dox-inducible promoter. We transfected HEK cells with base editor and the three inducible guide plasmids, and stimulated them with TNF α , IFN α and dox (Fig. 6c). By reading out the targeted AAVS1 amplicon, the respective A→G mutations informed the combination of input signals experienced by the cell population. By swapping the TRE3G promoter for the STAT6 promoter, inducible by IL4, we demonstrated that this triple recording can also work with a panel of three biologically relevant cytokines (Supplementary Fig. 8c–e). Finally, by incorporating an additional fourth guide we reported parallel recording of four inputs in HEK cells (Fig. 6d).

Multi-guide recording for deconvoluting input parameters. In addition to using parallel guides to record information about different signals, we hypothesized that implementing multiple inducible guides with different strengths or sensitivities could help provide a more detailed history of a single input. Previously, we observed that with a single readout it was difficult to distinguish between a short pulse of high dose TNF α and a longer pulse of a low dose with the same AUC (Extended Data Fig. 8a). To address this, we placed two AAVS1 guides under the control of two NFKB-inducible promoters, the second of which was engineered to be much less sensitive by decreasing the number of NFKB binding sites from five to two (Extended Data Fig. 8b,c). We transfected both guide constructs and base editor into HEK cells, and exposed them to either 100 ng ml⁻¹ TNF α for 0.25 h, or 0.5 ng ml⁻¹ TNF α for 50 h. On sequencing the AAVS1 amplicon, the A→G% of the first guide under standard promoter was similar between the two conditions. However, the second guide with less sensitivity had significantly reduced A→G% at the lower concentration condition (Extended Data Fig. 8d). This shows an initial proof of concept for how information from multiple guides could be integrated for more precise reconstruction of cellular history, where the first guide with a decoder provides the product of the duration × magnitude, and additional guides help to further detail the individual parameters.

Discussion

Here, we present a framework for scalable and modular cellular recording using a dCas12a base editor. Our development and

application of the hyperCas12a base editor uniquely enables this approach, providing a dynamic range for sensitive recording unachievable with the wild-type Cas12a. Although there has been some work using inducible control of guides with Cas9, these approaches use artificial inputs^{14,15}, or require incorporation of additional components such as Csy4 or ribozymes⁵³, limiting their scalability and wider use. By tethering inputs to the expression of a given crRNA, which can be directly processed by Cas12a itself, we record a variety of biologically relevant signals in human cells. We demonstrate that this recording operates in an analog fashion, able to record signal strength and duration. While this recorded history requires information from a population of cells and cannot be read out at a single-cell level, we demonstrate that fewer than 100 cells are required for robust signal, a small enough sample size to be achievable for many applications.

Throughout this work, we demonstrate recording both to an artificial reporter locus and an endogenous safe-harbor locus. The GFP reporter provides easy single-input readout and could enable fluorescence-based isolation of responding cells for further analysis, but is limited in scalability. Additionally, different cells have different numbers of GFP⁺ integrations, which may also express differently based on integration site, leading to potential variability between GFP metrics (fluorescence versus %GFP⁺). Reading out via sequencing is more costly but may alleviate some of these confounding factors, as well as provide easy scalability for recording of multiple signals on a single amplicon. For parallel recording, the performance of each input depends on the behavior of the inducible promoter, including the basal leakiness in resting state and strength of expression in stimulated conditions. As the number of inputs increases, there may also be some reduction in performance for a given promoter due to both transfection efficiency and limitations of cellular resources for transcription of multiple promoters at once⁵⁴.

The ability of this system to define the time boundaries of the signal of interest, as well as to scale up to multiple biologically relevant signals, makes it a powerful tool for studying complex processes. In particular, the temporal resolution may be useful for studying a number of time-critical processes, including organismal development and immune cell differentiation, where the question can now evolve from ‘did this cell experience signal of interest?’ to ‘did it experience this signal during the time subperiod of interest?’ The capacity for multi-signal recording could also be useful for functionally studying cellular processes behind intricate multifactorial pathways that would be difficult to dissect with a single-input recording device, including progression of infection in vivo.

In addition to recording signals to understand biology processes, this approach could also be used to benchmark performance and improve design of engineered cells. We demonstrated the ability of our recording system to interface with CAR-T cells to record successful activation in response to antigen. However, there are still challenges that preclude the successful application of CAR-T cells across tumor types, including cell trafficking to the tumor and difficulties with exhaustion. The incorporation of parallel dCas12a memory devices into engineered immune cells to record multiple transient signals they encounter while experimentally inaccessible in the body, to be read out at a later point by sequencing, could help improve the design-build-test cycle for these engineered cell therapies.

This dCas12a-based platform could be further expanded by integrating a variety of recently developed synthetic biology and genetic engineering tools. The guide-containing transcript could also be controlled directly at the RNA level, rather than transcriptionally, expanding the potential repertoire of cellular inputs⁵⁵. Additionally, the protein components could be controlled at the protein level using controllable degrons for more precise temporal resolution (for example, for a more sharply defined dCas12a-recording window)⁵⁶.

This type of logic gating for more precise signal recording could be further advanced using technology previously implemented in dCas12a-based activation gene circuits, including antiCRISPR proteins and split dCas12a (refs. ^{23,57}). Overall, we envision this as a modular, plug-and-play system to enable the next generation of sophisticated biological signal recording.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-022-01034-2>.

Received: 17 September 2021; Accepted: 6 April 2022;

Published online: 30 May 2022

References

- Burrill, D. R. & Silver, P. A. Making cellular memories. *Cell* **140**, 13–18 (2010).
- Sheth, R. U. & Wang, H. H. DNA-based memory devices for recording cellular events. *Nat. Rev. Genet.* **19**, 718–732 (2018).
- Farzadfard, F. & Lu, T. K. Emerging applications for DNA writers and molecular recorders. *Science* **361**, 870–875 (2018).
- Kotula, J. W. et al. Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc. Natl Acad. Sci. USA* **111**, 4838–4843 (2014).
- Burrill, D. R., Inniss, M. C., Boyle, P. M. & Silver, P. A. Synthetic memory circuits for tracking human cell fate. *Genes Dev.* **26**, 1486–1497 (2012).
- Ajo-Franklin, C. M. et al. Rational design of memory in eukaryotic cells. *Genes Dev.* **21**, 2271–2276 (2007).
- Bonnet, J., Subsoontorn, P. & Endy, D. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl Acad. Sci. USA* **109**, 8884–8889 (2012).
- Farzadfard, F. & Lu, T. K. Synthetic biology. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science* **346**, 1256272 (2014).
- Yang, L. et al. Permanent genetic memory with >1-byte capacity. *Nat. Methods* **11**, 1261–1266 (2014).
- Siuti, P., Yazbek, J. & Lu, T. Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31**, 448–452 (2013).
- Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175 (2016).
- Sheth, R. U., Yim, S. S., Wu, F. L. & Wang, H. H. Multiplex recording of cellular events over time on CRISPR biological tape. *Science* **358**, 1457–1461 (2017).
- Schmidt, F., Cherepkova, M. Y. & Platt, R. J. Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* **562**, 380–385 (2018).
- Perli, S. D., Cui, C. H. & Lu, T. K. Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* **353**, aag0511 (2016).
- Tang, W. & Liu, D. R. Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* **360**, eaap8992 (2018).
- Farzadfard, F. et al. Single-nucleotide-resolution computing and memory in living cells. *Mol. Cell* **75**, 769–780.e4 (2019).
- Loveless, T. B. et al. Lineage tracing and analog recording in mammalian cells by single-site DNA writing. *Nat. Chem. Biol.* <https://doi.org/10.1038/s41589-021-00769-8> (2021).
- Park, J. et al. II Recording of elapsed time and temporal information about biological events using Cas9. *Cell* **184**, 1047–1063 (2021).
- Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
- Zetsche, B. et al. Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2016).
- Fonfara, I., Richter, H., Bratović, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517–521 (2016).
- Zhong, G., Wang, H., Li, Y., Tran, M. H. & Farzan, M. Cpf1 proteins excise CRISPR RNAs from mRNA transcripts in mammalian cells. *Nat. Chem. Biol.* **13**, 839–841 (2017).
- Kempton, H., Goudy, L., Love, K. & Qi, L. Multiple input sensing and signal integration using a split Cas12a system. *Mol. Cell* **78**, 184–191.e3 (2020).
- Campa, C. C., Weisbach, N. R., Santinha, A. J., Incarnato, D. & Platt, R. J. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. *Nat. Methods* <https://doi.org/10.1038/s41592-019-0508-6> (2019).

25. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
26. Gaudelli, N. M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
27. Wang, X. et al. Cas12a base editors induce efficient and specific editing with low DNA damage response. *Cell Rep.* **31**, 107723 (2020).
28. Stadtmayer, E. A. et al. CRISPR-engineered T cells in patients with refractory cancer. *Science* **367**, eaba7365 (2020).
29. Li, X. et al. Base editing with a Cpf1–cytidine deaminase fusion. *Nat. Biotechnol.* **36**, 324–327 (2018).
30. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
31. Richter, M. F. et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-020-0453-z> (2020).
32. Kleinstiver, B. P. et al. Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282 (2019).
33. Guo, L. et al. Multiplexed genome regulation in vivo with hyper-efficient Cas12a. *Nat. Cell Biol.* **24**, 590–600 (2022).
34. Huang, T. P., Newby, G. A. & Liu, D. R. Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nat. Protoc.* <https://doi.org/10.1038/s41596-020-00450-9> (2021).
35. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
36. Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. & Lu, T. K. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol. Cell* **54**, 698–710 (2014).
37. Taniguchi, K. & Karin, M. NF- κ B, inflammation, immunity and cancer: coming of age. *Nat. Rev. Immunol.* **18**, 309–324 (2018).
38. Ede, C., Chen, X., Lin, M.-Y. & Chen, Y. Y. Quantitative analyses of core promoters enable precise engineering of regulated gene expression in mammalian cells. *ACS Synth. Biol.* **5**, 395–404 (2016).
39. Carlezon, W., Duman, R. & Nestler, E. The many faces of CREB. *Trends Neurosci.* **28**, 436–445 (2005).
40. Crabtree, G. R. & Olson, E. N. NFAT signaling: choreographing the social lives of cells. *Cell* **109**, S67–S79 (2002).
41. Shaulian, E. & Karin, M. AP-1 in cell proliferation and survival. *Oncogene* **20**, 2390–2400 (2001).
42. Nguyen, T., Sherratt, P. J. & Pickett, C. B. Regulatory mechanisms controlling gene expression mediated by the antioxidant response element. *Annu. Rev. Pharmacol. Toxicol.* **43**, 233–260 (2003).
43. Morsut, L. et al. Engineering customized cell sensing and response behaviors using synthetic notch receptors. *Cell* <https://doi.org/10.1016/j.cell.2016.01.012> (2016).
44. Lim, W. A. & June, C. H. The principles of engineering immune cells to treat cancer. *Cell* **168**, 724–740 (2017).
45. Schukur, L., Geering, B., Charpin-El Hamri, G. & Fussenegger, M. Implantable synthetic cytokine converter cells with AND-gate logic treat experimental psoriasis. *Sci. Transl. Med.* **7**, 318ra201 (2015).
46. Smole, A., Lainšček, D., Bezeljak, U., Horvat, S. & Jerala, R. A synthetic mammalian therapeutic gene circuit for sensing and suppressing inflammation. *Mol. Ther.* **25**, 102–119 (2017).
47. Monaco, C., Nanchahal, J., Taylor, P. & Feldmann, M. Anti-TNF therapy: past, present and future. *Int. Immunol.* **27**, 55–62 (2015).
48. Beirnaert, E. et al. Bivalent llama single-domain antibody fragments against tumor necrosis factor have picomolar potencies due to intramolecular interactions. *Front. Immunol.* **8**, 867 (2017).
49. Roybal, K. T. et al. Engineering T cells with customized therapeutic response programs using synthetic notch receptors. *Cell* **167**, 419–432.e16 (2016).
50. Papapetrou, E. P. & Schambach, A. Gene insertion into genomic safe harbors for human gene therapy. *Mol. Ther.* **24**, 678–684 (2016).
51. Lawrence, T. The nuclear factor NF- κ B pathway in inflammation. *Cold Spring Harb. Perspect. Biol.* **1**, a001651 (2009).
52. Tamura, T., Yanai, H., Savitsky, D. & Taniguchi, T. The IRF family transcription factors in immunity and oncogenesis. *Annu. Rev. Immunol.* **26**, 535–584 (2008).
53. Frieda, K. L. et al. Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).
54. Frei, T. et al. Characterization and mitigation of gene expression burden in mammalian cells. *Nat. Commun.* **11**, 4641 (2020).
55. Chang, A. L., Wolf, J. J. & Smolke, C. D. Synthetic RNA switches as a tool for temporal and spatial control over gene expression. *Curr. Opin. Biotechnol.* **23**, 679–688 (2012).
56. Banaszynski, L. A., Chen, L., Maynard-Smith, L. A., Ooi, A. G. L. & Wandless, T. J. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995 (2006).
57. Watters, K. E., Fellmann, C., Bai, H. B., Ren, S. M. & Doudna, J. A. Systematic discovery of natural CRISPR–Cas12a inhibitors. *Science* **362**, 236–239 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

Methods

Plasmid cloning. Standard molecular cloning techniques were used to assemble constructs in this paper. The crRNA plasmids were cloned using ligation cloning. Oligos were annealed and phosphorylated with PNK (NEB) and inserted into backbone using T4 DNA Ligase (NEB), then transformed into Stellar Competent cells (Takara Bio). The scaffold used (Ns denote the spacer sequence) is as follows: 5'-AATTTCTACTAAGTGTAGATNNNNNNNNNNNNNNNNNNNNNNNN-3'. Spacer sequences for all crRNAs can be found in Supplementary Table 1. A similar annealing and ligation strategy was used to synthesize and clone repetitive response elements upstream of minimal promoters to construct inducible promoters.

Other plasmids were cloned using InFusion (Takara Bio). Nuclease-dead (D832A) dCas12a from *Lachnospiraceae bacterium*, its crRNA backbone and intein sequences were modified from plasmids from previous work³³. The hyperdCas12a mutations were obtained from previous work³³. The ABE8e and antiTNF nanobody sequence were cloned from ordered gBlocks (Integrated DNA Technologies)^{43,48}. The triplex sequence was a gift from R. Platt²⁴. The SynNotch sequence was a gift from W. Lim.⁴³ The split hygromycin construct was designed as previously described⁵⁸. DNA sequences and full reference of plasmids used can be found in Supplementary Tables 2 and 3.

GFP reporter design. The reporters and guides were designed based on sfGFP so that the targeted A was in the rough 'editing window' 7–15 bp from TTTV PAM, and there were as few other As in that window as possible. When permitted, silent mutations were made in the GFP sequence to introduce a PAM or remove other As in editing window, without changing final amino acid sequence. For the first four designs, the final 'corrected' GFP amino acid sequence was identical to the original starting sequence. However, due to limited availability of arginine, tryptophan and glutamine residues in correct proximity to a TTTV PAM within the GFP sequence, the second panel of reporters included an amino acid conversion from the original sequence. We looked for candidate lysine or asparagine residues to mutate to stop codon, which could then be base edited to arginine or glutamine, respectively, hypothesizing that a similar size and charge of amino acid would minimize the chance of disrupting GFP protein folding or stability.

Cell culture. LentiX HEK293T cells (Clontech) were cultured in DMEM + GlutaMAX (Thermo Fisher) supplemented with 10% Tet System Approved fetal bovine serum (FBS) (Clontech) and 100 U ml⁻¹ of penicillin and streptomycin (Gibco). HEK cell cultures were passaged every 2–3 d as cells approached confluence. Jurkat cells and K562 cells (ATCC) were cultured in RPMI 1640 (Thermo Fisher) supplemented with 10% Tet-FBS and 100 U ml⁻¹ Pen/Strep. Jurkat cell cultures were passaged every 2–3 days as cell culture densities approached 1 × 10⁶ cells per ml. Cells were maintained at 37 °C and 5% CO₂ and passaged using standard cell culture techniques. Cells were not tested for mycoplasma contamination.

Signal induction. To turn on the TRE3G promoter, cells were stimulated with 100 ng ml⁻¹ doxycycline (Gold Biotech). To activate NFκB signaling, cells were stimulated with TNFα (R&D Systems), at 100 ng ml⁻¹ unless indicated otherwise. To activate hypoxia, cells were stimulated with 2 mM DMOG (Cayman). To increase cAMP, cells were stimulated with 100 μM Forskolin (Selleckchem). To stimulate NFAT signaling, cells were stimulated with 50 ng ml⁻¹ PMA (STEMCELL). To induce oxidative stress, cells were exposed to 10 μM tBHQ (Sigma). To stimulate AP1 signaling, cells were stimulated with 50 ng ml⁻¹ PMA (STEMCELL). To stimulate the IRF pathway, cells were stimulated with 100 ng ml⁻¹ IFNα (STEMCELL). To stimulate the STAT6 pathway, cells were stimulated with 100 ng ml⁻¹ IL4 (Biologend). For most HEK transfection/stimulation experiments, inducer was added at same time as transfection, except indicated otherwise. For oxidative stress, inducer was added 24 h posttransfection to mitigate toxicity. For experiments with variable duration of signal exposure, the media change to add inducer was performed 24 h posttransfection to ensure all samples were exposed to transfection complexes for the same time for consistent transfection efficiency. For experiments with parallel recording including IL4, IL4 was added at same time as transfection, and the other two cytokines were added 24 h posttransfection. For Jurkat AAVS1-targeting experiments, TNFα was added 24 h postelectroporation.

To turn on the SynNotch system, 24 h posttransfection HEK cells were dissociated with 0.05% Trypsin-EDTA (Life Technologies) and reseeded with sender HEK cells expressing the CD19 antigen at a 1:5 ratio. To turn on CAR signaling, 24 h postelectroporation Jurkat cells were cocultured with K562 cells expressing the CD19 antigen at a 1:3 ratio. Cocultures were performed for 48 h before analysis.

Transfections. For transient transfection of HEK293Ts, cells were seeded the day before transfection at 1 × 10⁵ cells per ml. Transient transfections were performed using 3 μl of TransIT-LT1 transfection reagent (Mirus) per μg of plasmid. Briefly, for transfections into 24-well plates the calculated volumes of plasmid DNA were added to 50 μl of Opti-MEM reduced serum media (Thermo Fisher), followed by the addition of an appropriate volume of TransIT-LT1 and subsequent mixing through vigorous pipetting. The transfection mixture was then incubated for 15 min at room temperature before transfection into cells using drop-wise pipetting. For most experiments, 250 ng of each construct were transfected into

24-well plates. For dual guide experiments, 125 ng of each guide plasmid was transfected for a total of 250 ng guide. For triple or quadruple guide experiments, 125 ng of each guide plasmid and 125 ng of rtTA and/or STAT6 plasmid (as needed) were transfected. For genomic DNA extraction, 4 μg ml⁻¹ of puromycin (Gold Biotech) was added to the media 1 d posttransfection to select for cells transfected with base editor.

For transient transfection of Jurkats, cells were electroporated with guide plasmid using the Neon Transfection System 10 μl Kit (Invitrogen). Electroporations were performed using 2 × 10⁵ Jurkat cells per well of a 24-well plate. For AAVS1 editing Jurkat experiments, 2.25 μg of guide plasmid, prepped using endotoxin-free maxiprep kit (Qiagen), was electroporated. For CAR Jurkat coculture experiments, 2.25 μg of guide plasmid and 0.75 μg of reporter plasmid were electroporated. The electroporation parameters used were as follows: pulse voltage 1,325 V, pulse width 10 ms, pulse number 3. Cells were cultured in antibiotic-free media postelectroporation for improved viability. Dox was added at 100 ng ml⁻¹ postelectroporation to induce expression of stable base editor in Jurkat cell lines.

Stable cell generation. Stable cell lines were generated using lentiviral transduction. For lentiviral production, HEK293T cells were transfected with 1.51 μg of pHR vector with construct of interest, 1.32 μg of dR8.91 and 165 ng of pMD2 with 7.5 μl of Mirus TransIT-LT1 (for a six-well plate format, scaled up or down as appropriate). Two days posttransfection, lentivirus was harvested and filtered through a 0.45-μm polyvinylidene fluoride filter (Millipore). Following filtration, 1 volume of lentivirus was mixed with 4 volumes of Lentivirus Precipitation Solution (Alstem) and refrigerated overnight. The next day, lentivirus was pelleted at 1,500g for 30 min at 4 °C. HEK cells were seeded the day before transduction at 1 × 10⁵ cells per well of a 12-well plate and transduced with precipitated lentivirus resuspended in 1/100 of the original volume of PBS. For Jurkat cell transduction, cells were seeded the day of transduction at 1 × 10⁶ cells per ml in 100 μl in a 96-well plate, and transduced with precipitated lentivirus resuspended in 100 μl of RPMI. When appropriate, transduced cells were selected for with 2 μg ml⁻¹ Puromycin (Gold Biotech) or 150 μg ml⁻¹ Hygromycin (Sigma).

Flow cytometry. HEK cells were dissociated using 0.05% Trypsin-EDTA (Life Technologies), resuspended in PBS + 10% FBS, and analyzed for fluorescence using a CytoFLEX S flow cytometer (Beckman Coulter). 10,000 cells from the transfected population of interest were collected for each sample. For flow cytometry experiments with constitutive base editor and constitutive U6-guide, cells were gated for mCherry⁺/BFP⁺ (presence of base editor and guide). For flow cytometry experiments with the constitutive base editor and inducible guide, cells were gated for mCherry⁺ (presence of base editor). For flow cytometry experiments with the inducible base editor and inducible guide, cells were gated for single cells. For flow cytometry to generate the dose/time heatmap, cells were analyzed in DMEM in 96-well plate format and 10,000 total cells were collected. Data were analyzed using FlowJo v.10.8.1 (BD Biosciences). For analysis with decreasing number of cells in Fig. 3a, the FlowJo Downsampling Plugin was used.

ELISA. Supernatants from cell cultures were harvested at the indicated time point, and samples were stored at -80 °C until analysis. Secreted antiTNF nanobody was measured using the His Tag ELISA Detection Kit (Genscript), with all samples run in technical duplicate on the ELISA plate. Absorbance was measured at 450 nm using a Synergy H1 plate reader (BioTek) and the BioTek Gen5 v.3.02 software, and protein concentrations were calculated by standard curves fitted to a power law.

High throughput sequencing. Genomic DNA was isolated using the DNeasy Blood&Tissue Kit (Qiagen), and the targeted loci of interest were amplified with Q5 Hot Start High-Fidelity Mastermix (NEB), using a two-round PCR strategy to add Illumina adaptors for each sample as previously described³⁴. For the first round PCR, 1–4 μl of purified genomic DNA were added as template based on concentration from the nanodrop, to get 50–100 ng of input DNA template. This PCR was performed for 30 cycles at *T_m* = 68, and successful amplification was confirmed on a 1% agarose gel. Next, 2 μl of unpurified PCR1 were added as template for the second round PCR, along with unique combinations of forward and reverse primers to barcode individual samples. This second PCR was performed for 14 cycles at *T_m* = 60, and successful amplification was confirmed on a 1% agarose gel. Following confirmation of amplification of all individual samples, samples were pooled by experiment and purified through gel extraction using the Macherey-Nagel NucleoSpin Gel and PCR Clean-up Kit. Sample concentration was then quantified via Qubit Fluorometric quantification (Thermo Fisher). Up to 96 samples were pooled into a single library and sequenced using a paired-end reads for a 2 × 300-cycle run with MiSeq Reagent Kit v.3 (600-cycle) (Illumina). The primers used for generating amplicons are shown in Supplementary Table 5, and the amplicon sequences are in Supplementary Table 6. Fastq files were analyzed using CRISPResso2, filtering out reads with a phred33 quality score less than 20.

Data analysis and statistics. For bar graphs, data are displayed as individual points. For other graphs, data are displayed as mean value with error bars showing standard deviation. Sample size is indicated in figure legends. No randomization

or blinding was performed. Sample sizes used are consistent with those used by similar cell recording studies. Statistical analysis was performed with Prism 9 (Graphpad), using Welch's two-sided *t*-test without adjustment for multiple comparison. Statistical values are included in Supplementary Table 4.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data supporting main figures and extended data figures have been included in source data files. Plasmids are available from Addgene with the following accession codes: 183098, 183203, 183204, 183205, 183206, 183207, 183623, 183624, 183625, 183626, 183627, 183628 and 183629. Raw deep sequencing data are available at National Center for Biotechnology Information Bioproject [PRJNA818698](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA818698). No custom code was used in this study. Source data are provided with this paper.

References

58. Jillette, N., Du, M., Zhu, J. J., Cardoz, P. & Cheng, A. W. Split selectable markers. *Nat. Commun.* **10**, 4968 (2019).

Acknowledgements

We thank members of the Qi laboratory, including M. Chavez, V. Tieu, P. Finn, T. Abbott, A. Chemparathy, X. Xu, J. Bian and M. Nakamura, as well as E. Gonzalez Diaz and F. Yang for advice and helpful discussions. The authors also thank the Stanford Shared Protein and Nucleic Acid Facility for technical support. H.R.K. acknowledges support from the National Science Foundation Graduate Research Fellowship Program, Stanford Bio-X Fellowship Program and Siebel Scholars Foundation. L.S.Q. acknowledges support from the Li Ka Shing Foundation, the Stanford Maternal and Child Health

Research Institute through the Uytengsu-Hamilton 22q11 Neuropsychiatry Research Award Program, National Science Foundation CAREER award (award no. 2046650) and California Institute for Regenerative Medicine (CIRM, DISC2-12669). L.S.Q. is a Chan Zuckerberg Biohub investigator. The work is supported by National Science Foundation CAREER award (award no. 2046650) and the Li Ka Shing Foundation.

Author contributions

H.R.K. and L.S.Q. conceived the idea. H.R.K. designed and performed experiments for all figures. K.S.L. designed constructs and performed experiments for initial testing of Cas12 base editor constructs. L.Y.G. designed constructs and contributed for testing of hyperCas12 mutants. H.R.K. analyzed the data. H.R.K. and L.S.Q. wrote the manuscript. All authors read and commented on the manuscript.

Competing interests

The authors have filed a provisional patent via Stanford University partly related to the work (US patent no. 63/148,652; international application patent no. PCT/US2022/016223). L.S.Q. is a founder of Epicrispr Biotechnologies.

Additional information

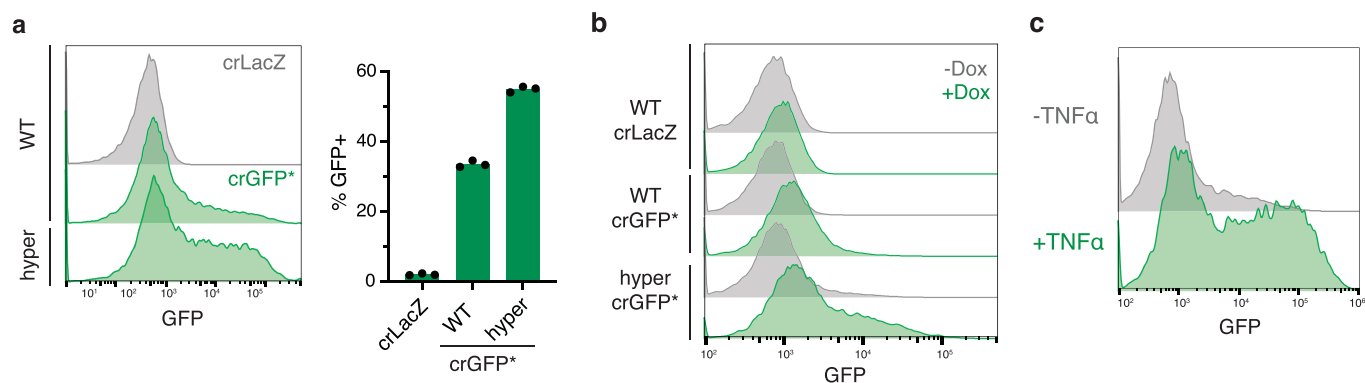
Extended data are available for this paper at <https://doi.org/10.1038/s41589-022-01034-2>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41589-022-01034-2>.

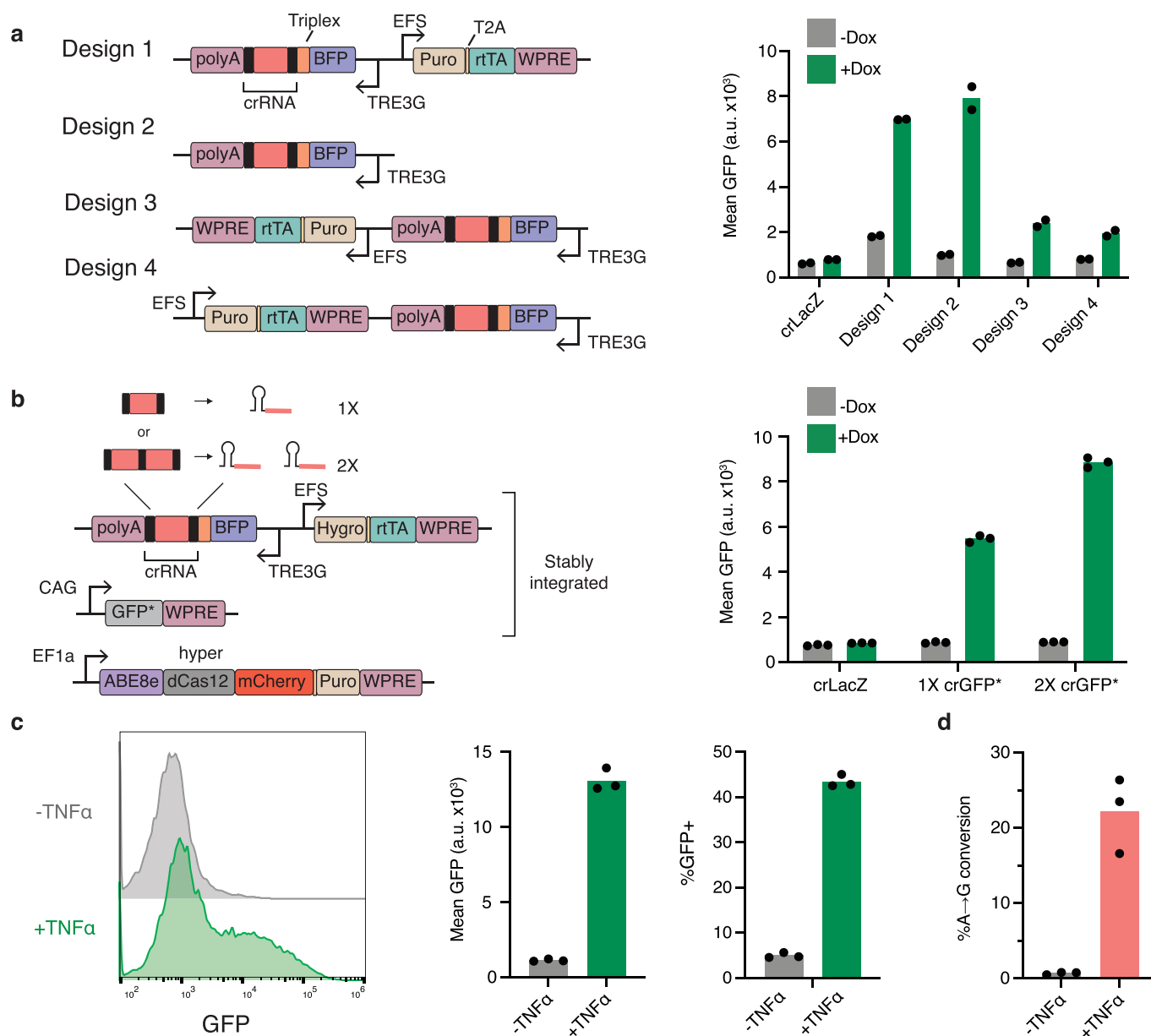
Correspondence and requests for materials should be addressed to Lei S. Qi.

Peer review information *Nature Chemical Biology* thanks the anonymous reviewers for their contribution to the peer review of this work.

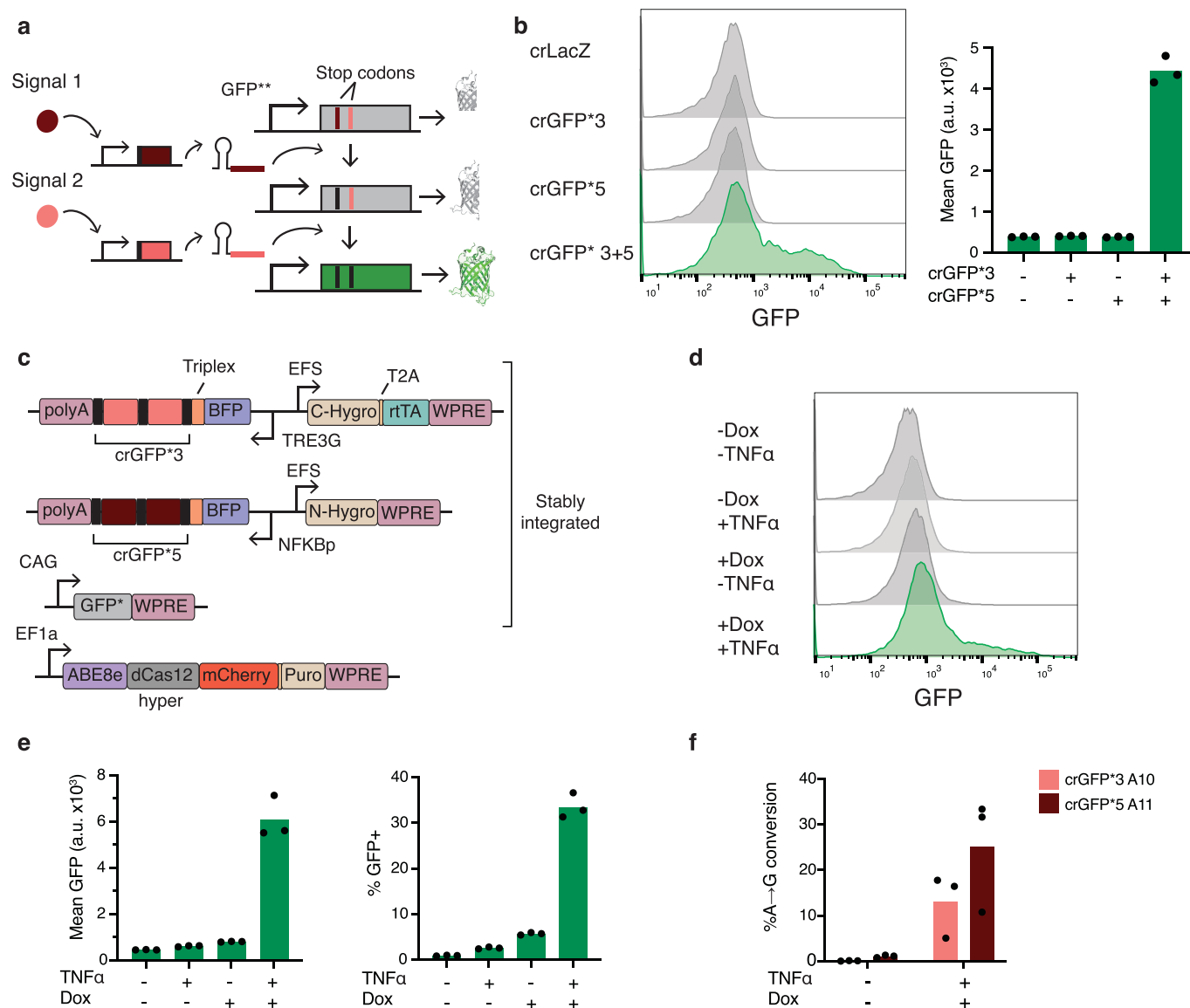
Reprints and permissions information is available at www.nature.com/reprints.



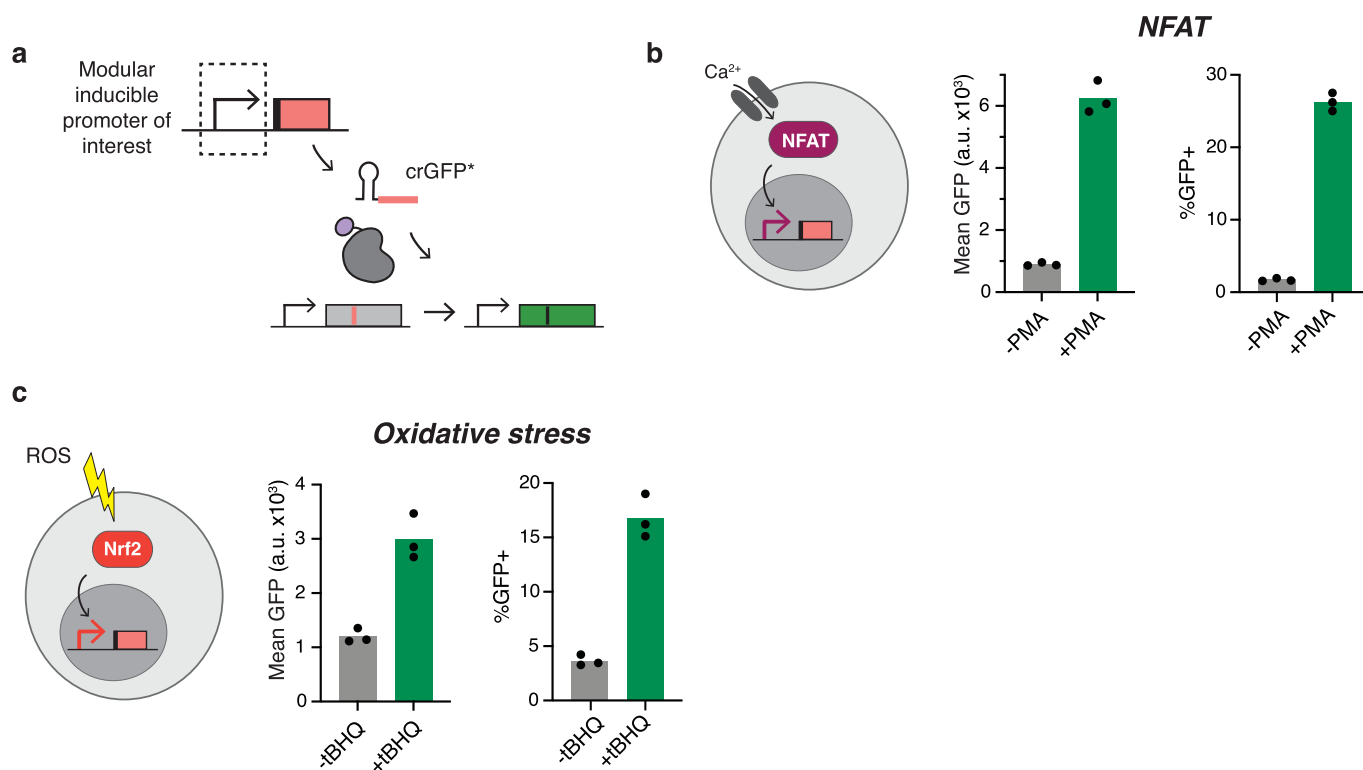
Extended Data Fig. 1 | Base editing with hyperCas12a for signal recording. **a**, Left, representative flow cytometry histograms showing GFP fluorescence in HEK cells 3 days post-transfection with GFP* reporter, U6-driven guide and wild type (WT) or mutant (hyper) dCas12a-ABE. crLacZ, non-targeting control. Right, quantification of percentage GFP+ cells for 3 independent replicates. To go with Fig. 1c. **b**, Representative flow cytometry histograms for GFP* HEK cells transfected with the dox-recording constructs, 3 days post-transfection and stimulation with dox. To go with Fig. 1f. **c**, Representative flow cytometry histograms for GFP* HEK cells transfected with hyperdCas12a base editor and NFκB-driven crGFP*, 3 days post-transfection and stimulation with TNFα. To go with Fig. 1g.



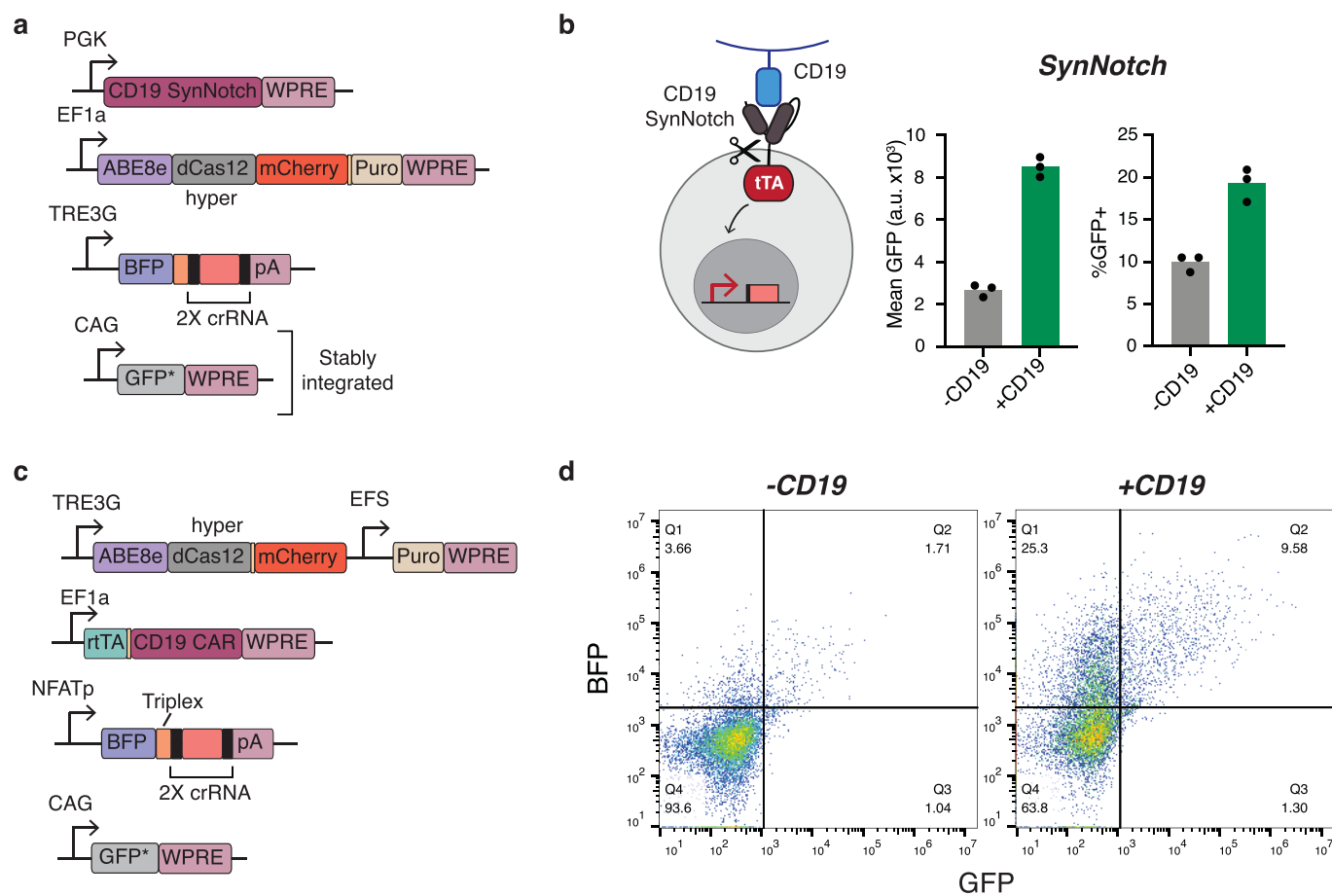
Extended Data Fig. 2 | Optimizing inducible crRNAs for lentiviral transduction. **a**, Left, designs tested for transferring the inducible guide cassette into a lentiviral backbone. The inducible construct was placed in reverse orientation, to prevent the polyA from disrupting lentiviral packaging. Right, quantification of reporter GFP fluorescence in HEK cells transiently transfected with inducible guide, mutant base editor, and reporter, as well as additional rtTA for design 2. Data shown for 2 independent replicates, 3 days post-transfection and stimulation with dox. Design 1 was selected moving forward, as despite slightly worse performance than design 2 it had benefit of including constitutive open reading frame for selection. **b**, Left, schematic showing incorporation of additional repeat-spacer cassette to increase amount of guide available per mRNA transcript. Right, GFP fluorescence in GFP* HEK cells stably transduced with dox-inducible guide, 3 days post-transfection with base editor and stimulation with dox. Data shown for 3 independent replicates. **c**, Left, representative histogram of GFP* HEK cells stably transduced with NFKB-crRNA, 3 days post-transfection with base editor and stimulation with TNF α . Right, quantification of mean fluorescence and percent GFP+ for 3 independent replicates. **d**, Quantification of A→G mutations in stably transduced NFKB-crRNA cells at the targeted stop codon in the GFP* locus as measured by MiSeq. Data are shown for 3 independent replicates.



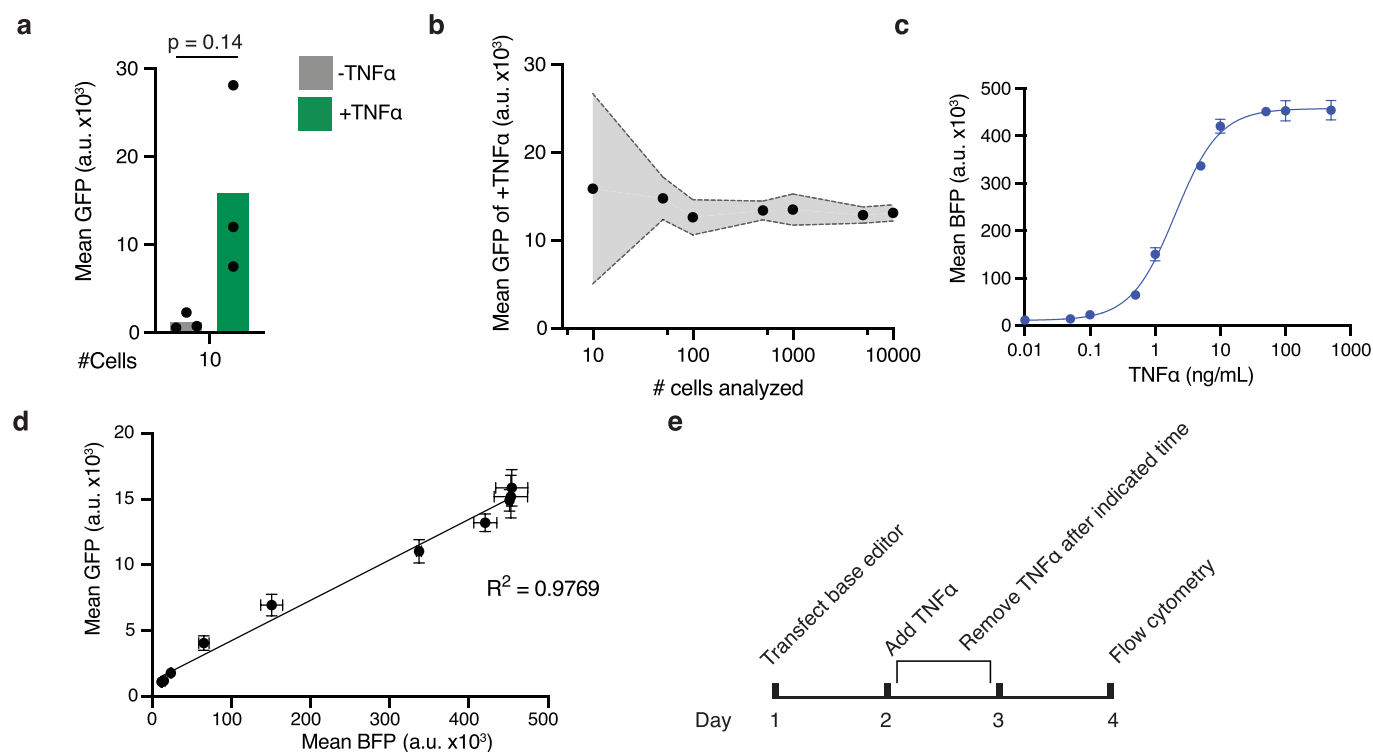
Extended Data Fig. 3 | Dual signal recording using the GFP reporter. a**, Schematic showing the GFP** dual recording reporter, which expresses GFP upon the activity of two distinct GFP* guides removing two stop codons. **b**, Left, representative histograms of HEK cells 3 days post-transfection with GFP**, base editor, and combinations of U6-crGFP*. Right, quantification of mean GFP for 3 independent replicates. **c**, Modified lentiviral constructs used to transduce two inducible guides into cells and select with a single antibiotic, through use of split hygromycin. **d**, Representative histograms of GFP** HEK cells stably transduced with NFKB and TRE3G inducible guides shown in **c** and transfected with base editor, 3 days posttransfection and stimulation with TNF α /dox. **e**, Quantification of mean fluorescence (left) and percentage GFP+ (right) of GFP** HEK cells stably transduced with NFKB and TRE3G inducible guides and transfected with base editor, 3 days post transfection and stimulation with TNF α /dox. All data shown for 3 independent replicates. **f**, Quantification of A→G mutations at each targeted stop codon in the GFP** locus as measured by MiSeq. Data are shown for 3 independent replicates.



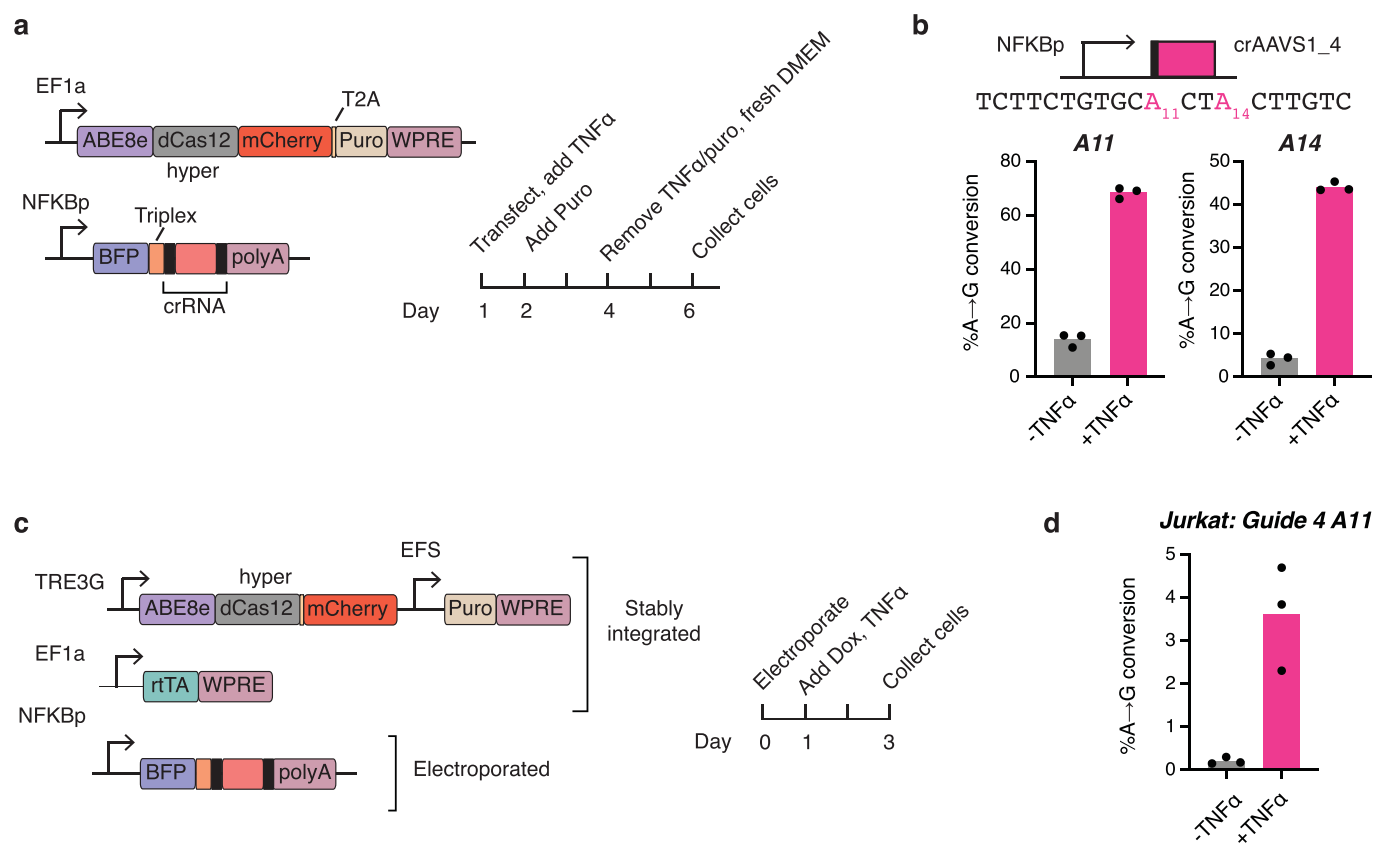
Extended Data Fig. 4 | Modular recording of diverse biological signals. a, Schematic showing modular recording by incorporating different inducible promoters into the guide construct. **b**, Cellular memory of NFAT signaling, stimulated by PMA. **c**, Cellular memory of oxidative stress, stimulated by tBHQ. ROS = reactive oxygen species. All experiments in **b-c** shown in GFP* HEK cells stably transduced with respective inducible guide, for 3 independent replicates 3 days post transfection with base editor.



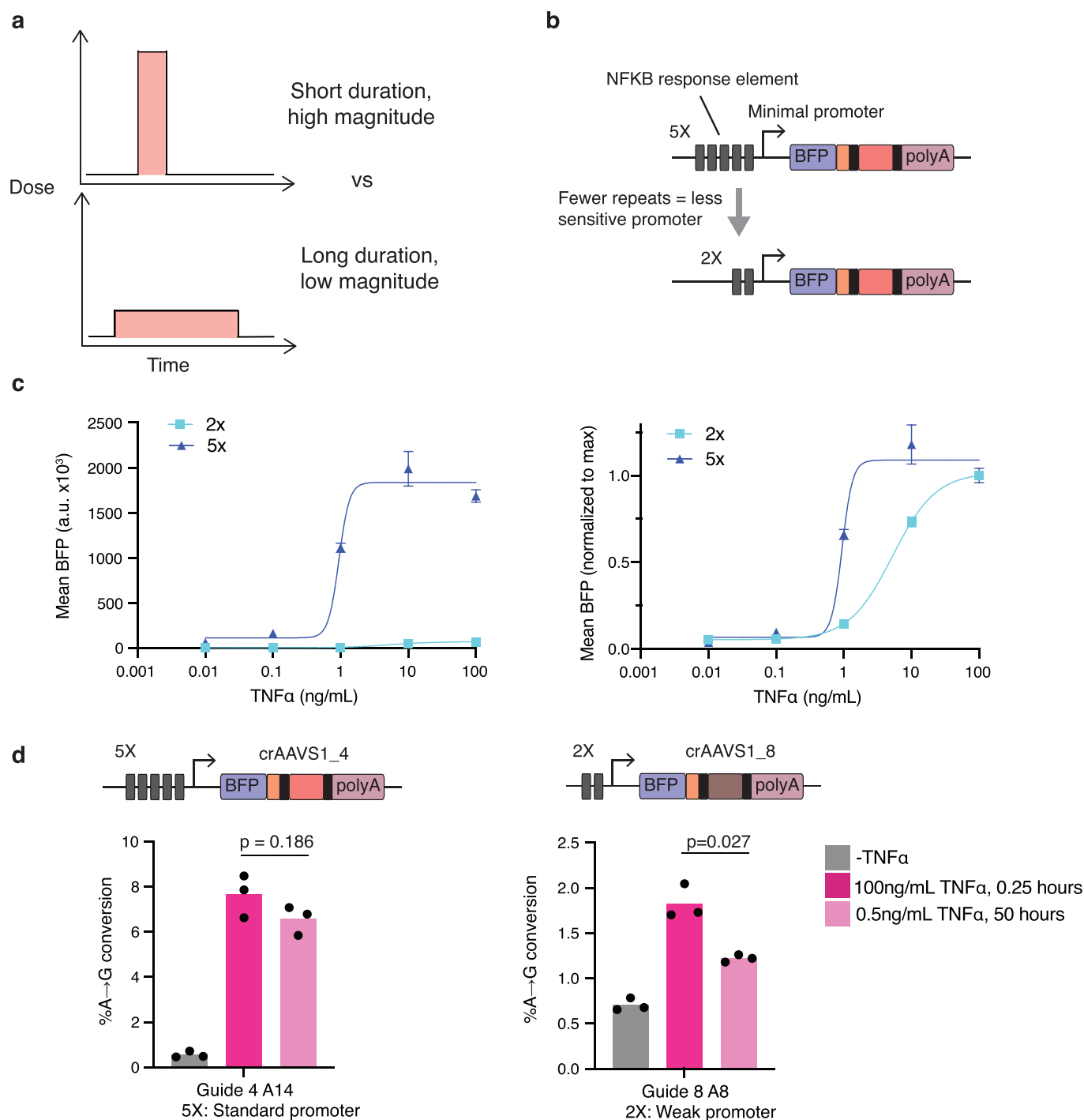
Extended Data Fig. 5 | Recording memory of antigen encounter. a, Constructs used for SynNotch-induced guide expression. **b**, Cellular memory of antigen encounter via SynNotch, in GFP* HEK cells transfected with guide, SynNotch, and base editor, then cocultured with -/+CD19 HEK cells. Data shown for 3 independent replicates. **c**, Constructs used for recording CD19-CAR activation in Jurkat cells. The top two constructs were stably integrated, and the bottom two were electroporated. **d**, Representative flow cytometry histograms for signal recording of CD19 antigen encounter in Jurkat cells. Histograms show BFP (induced by NFAT promoter) vs GFP (from successful base editing event of GFP* reporter) in engineered Jurkat CAR-T cells, after 48 hours of co-culture with K562 cells -/+CD19. To go with Fig. 2d.



Extended Data Fig. 6 | Characterization of recording parameters. **a**, GFP⁺ HEK cells stably transduced with lentiviral NFkBp-guide and transfected with base editor were stimulated with 100 ng/mL TNF α for 3 days, then analyzed via flow cytometry. Randomly selected sub-populations of cells of the indicated size were analyzed using downsampling. Statistical analysis was performed using a two-sided t test without adjustment for multiple comparisons. Data shown for 3 independent replicates. To go with Fig. 3a. **b**, The mean GFP fluorescence for each downsampled cell population for the NFkB-recording circuit, +TNF α condition is plotted against the number of cells analyzed, to go with Fig. 3a. Dotted lines with gray shading indicate standard deviation of 3 independent replicates. **c**, Dose response for GFP⁺ HEK cells stably transduced with lentiviral NFkBp-guide and transfected with base editor. Mean BFP fluorescence is shown 3 days post-transfection and stimulation with TNF α . Error bars indicate standard deviation of 3 independent replicates. To go with Fig. 3b. **d**, Mean BFP (expressed under NFkB promoter) plotted against mean GFP ("memory" reporter) for TNF α dose-response experiment. Data were fitted using a linear regression. Error bars indicate standard deviation of 3 independent replicates. **e**, Experimental timeline used for testing different durations of TNF α stimulation.



Extended Data Fig. 7 | Recording history of NFKB signaling at the AAVS1 locus. **a**, Left, constructs used for signal recording using inducible AAVS1 targeting guides in HEK cells. Right, experimental timeline used for genomic DNA (gDNA) collection. **b**, Single-input recording at the AAVS1 locus in HEK cells transfected with base editor and NFKB-driven crAAVS1_4, and stimulated with TNF α for 3 days. The %A $\rightarrow$ G conversion as measured by MiSeq is shown for A11 (left) and A14 (right), with number indicating position of A relative to PAM. Data shown for 3 independent replicates. **c**, Left, constructs used for recording at the AAVS1 locus in Jurkat cells. Right, experimental timeline. **d**, Recording of NFKB-mediated inflammation in Jurkat cells at the AAVS1 locus, as measured by MiSeq. Data shown for 3 independent replicates.



Extended Data Fig. 8 | Utilizing two parallel guides for more detailed record of input signal characteristics. **a**, Schematic showing two input signals with distinctive patterns but the same area under the curve. **b**, Promoter engineering strategy used to reduce the sensitivity of NFKB-responsive promoter, to generate two inducible guide constructs that respond differently to input TNF α signal. **c**, Mean BFP fluorescence of HEK cells 3 days post-transfection with NFKB promoter variants and stimulation with TNF α . Left, raw mean BFP values are plotted. Right, mean BFP values normalized to the maximum of each respective promoter are plotted. Error bars indicate standard deviation of 3 independent replicates. **d**, Quantification of %A $\rightarrow$ G conversion as measured by MiSeq for HEK cells transfected with base editor, 5X-NFKB-crAAVS1_4, and 2X-NFKB-crAAVS1_8. One day after transfection, cells were stimulated with either 100 ng/mL TNF α for 0.25 hours, or 0.5 ng/mL TNF α for 50 hours. All data shown for 3 independent replicates. Statistical analysis was performed using a two-sided t test without adjustment for multiple comparisons.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| ☐ | ☒ | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☒ | ☐ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Plate reader data for ELISA experiments was collected using BioTek Gen5 v3.02 software.

Data analysis CRISPResso2 was used for analysis of next generation sequencing data. FlowJo v10.8.1 was used for analysis of flow cytometry data. Prism v9 was used for statistical analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data supporting main figures and extended data figures has been included in Source data files. Plasmids are available from Addgene with the following accession codes: 183098, 183203, 183204, 183205, 183206, 183207, 183623, 183624, 183625, 183626, 183627, 183628, 183629. Raw deep sequencing data are available at NCBI Bioproject: PRJNA818698.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No calculations were performed to predetermine sample size. Experiments were performed in triplicate unless indicated otherwise
Data exclusions	No data were excluded.
Replication	All experiments were repeated for three independent biological replicates unless otherwise noted. All attempts at replication were successful.
Randomization	Randomization was not relevant to this study. All experiments were performed with in vitro cell culture, in which cells were seeded identically in 24-well plates (or other size, if indicated) from the same starting population of homogeneous cells the day prior to experimentation. At time of data collection, cells from an individual well were thoroughly mixed prior to analysis, to ensure a well-mixed population with random subset of cells being analyzed via flow cytometry or other indicated experimental methods.
Blinding	Blinding was not relevant due to this study. The data across given samples were generated in an automated fashion using equipment like flow cytometry or high throughput sequencing. This data generation did not involve any assignment of values from the person running the experiment that could have introduced subjective bias and thus would have required blinding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	LentiX HEK293T cells were purchased from Clontech. Jurkat cells were purchased from ATCC. K562 cells were purchased from ATCC.
Authentication	Cell lines used in this study were not authenticated.
Mycoplasma contamination	Cell lines used in this study were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in this study.

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Samples were prepared as indicated in Methods. Briefly, cells were trypsinized, pelleted, and resuspended in PBS+10% FBS prior to flow cytometry. No staining was performed.

Instrument

Beckman Coulter CytoFLEX S

Software

Data was analyzed using FlowJo 10

Cell population abundance

At least 10,000 cells of the gated population of interest were collected for all experiments, unless indicated otherwise.

Gating strategy

Samples were gated as indicated in Methods. Briefly, first live cells were gated for using FSC/SSC, then single cells were gated for using FSC-A vs H. Then, when required based on what constructs were transfected (see Methods), cells were further gated for mCherry+ cells or mCherry+/BFP+ cells. Gates were constructed compared to a non-transfected control sample

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.