

High-performance chemical and light-inducible recombinases in mammalian cells and mice

Benjamin H. Weinberg¹, Jang Hwan Cho¹, Yash Agarwal¹, N. T. Hang Pham¹, Leidy D. Caraballo¹, Maciej Walkosz¹, Charina Ortega¹, Micaela Trexler¹, Nathan Tague¹, Billy Law¹, William K. J. Benman¹, Justin Letendre¹, Jacob Beal², and Wilson W. Wong^{1*}

¹Department of Biomedical Engineering and Biological Design Center, Boston University, Boston, MA 02215, USA.

²Raytheon BBN Technologies, Cambridge, MA 02138, USA

*Correspondence: wilwong@bu.edu.

ABSTRACT

Site-specific DNA recombinases are some of the most powerful genome engineering tools in biology. Chemical and light-inducible recombinases, in particular, enable spatiotemporal control of gene expression. However, the availability of inducible recombinases is scarce due to the challenge of engineering high performance systems with low basal activity and sufficient dynamic range. This limitation constrains the sophistication of genetic circuits and animal models that can be created. To expand the number of available inducible recombinases, here we present a library of >20 orthogonal split recombinases that can be inducibly dimerized and activated by various small molecules, light, and temperature in mammalian cells and mice. Furthermore, we have engineered inducible split Cre systems with better performance than existing inducible Cre systems. Using our orthogonal inducible recombinases, we created a “genetic switchboard” that can independently regulate the expression of 3 different cytokines in the same cell. To demonstrate novel capability with our split recombinases, we created a tripartite inducible Flp and a 4-Input AND gate. We have performed extensive quantitative characterization of the inducible recombinases for benchmarking their performances, including computation of distinguishability of outputs in terms of signal-to-noise ratio (SNR). To facilitate sharing of this set of reagents, we have deposited our library to Addgene. This library thus significantly expands capabilities for precise and multiplexed mammalian gene expression control.

INTRODUCTION

Site-specific recombinases (SSRs) are enzymes that perform site-specific DNA recombination through their ability to carry out DNA cutting and ligation operations¹. SSRs can effectively regulate gene expression based on the placement of their corresponding set of recombination sites around a gene of interest or transcription termination sequence. Furthermore, chemical and light-inducible systems have been engineered to control the expression or activity of SSRs, enabling spatiotemporal control of gene expression^{2,3}. Due to their unique chemistry and performance, SSRs have been used extensively in genetic circuit designs⁴⁻⁸ and animal model development⁹⁻¹¹ to interrogate biological processes and design principles. While inducible recombinases have proven to be invaluable in many studies¹²⁻¹⁴, especially in mammalian cells, the number of orthogonal recombinases available is limited. Most inducible systems are optimized solely for Cre recombinase and few, if any, orthogonal systems are available for other recombinases. This limitation constrains the sophistication of the genetics that can be developed and ultimately the biological and engineering problems that can be addressed. Furthermore, some small molecule inducers (e.g., antibiotics or hormone analogs) may not be compatible with some experimental objectives. As such, there is an urgent need to develop a larger collection of inducible recombinases that employ an array of orthogonal inducers.

A major challenge in developing novel inducible recombinases in mammalian cells is to ensure both low basal activity and high induced activity, which requires orthogonal SSRs and induction systems to work effectively together. We and others have shown that many recombinases are active in mammalian cells^{6,15}, thus serving as a potential source for engineering orthogonal inducible recombinases. Furthermore, a variety of inducible systems has been used to control the activity or expression of Cre. While the availability of diverse classes of SSRs and inducible systems provide a rich pool of source materials, engineering optimal pairings of SSRs and

inducible systems is challenging and requires systematic testing and characterization. One proven and attractive approach to generating inducible recombinases has been to utilize chemical-inducible dimerization domains.

Chemical-inducible dimerization (CID) domains, whereby a protein of interest is split into two fragments and each one fused to a complementary chemical-inducible heterodimerization domain, are an attractive system to control the activity of recombinases. The most prominent CID system is the FKBP/FRB system, which forms a heterodimer upon the addition of the rapamycin analog, AP21967, termed “rapalog.” A study utilized the FKBP/FRB system to generate a rapamycin-inducible Cre recombinase¹⁶. Since this work, orthogonal CID systems have been identified, including one that uses phytohormone abscisic acid¹⁷ (ABA) and another that uses phytohormone gibberellin (GIB)^{18,19}. CID systems are advantageous because split proteins tagged with CIDs can be rendered non-functional until the molecule is administered, having the potential to yield low basal activities. Thus, high levels of the split protein can be expressed in advance in the desired cellular compartment, enabling fast responses. This is in contrast to other post-translational systems that are either actively degrading the protein until a drug is administered to stabilize the protein²⁰ or sequestering the protein to the cytoplasm until drug exposure permits translocation of the protein to the nucleus^{21,22}. Furthermore, the expression of CID-regulated recombinases can be controlled by conditional promoters (e.g., tissue specific promoters) independent of the CID system. This feature enhances spatiotemporal control of gene expression²³, which is not achievable if recombinase expression is controlled by a drug-inducible promoter (e.g., Tet-ON/doxycycline²⁴).

In this work, we address the shortage of inducible recombinases by developing a suite of inducible recombinases through screening for functional split locations to use with the rapalog, abscisic acid, and gibberellin CID systems. In total, we are able to achieve functional chemical

induction in six recombinases. Moreover, we demonstrate the extensibility of this approach by engineering two light and temperature inducible recombinases using light inducible dimerization domains (LID) for conditional recombination. Finally, we demonstrate the functionality of these engineered systems by implementing the first 4-input drug inducible logic gates and tripartite inducible Flp recombinase in mammalian cells. To ensure accessibility of the inducible recombinases we have developed, we have deposited our library to DNA-repository Addgene. We have also performed detailed characterizations of select recombinases using community-derived standards and metrics for benchmarking purposes and quantitative comparison. Together, our library of characterized inducible recombinases will enable the creation of novel genetic circuits and animal models for both fundamental studies and biotechnological applications.

RESULTS

Selecting split locations:

CID systems were utilized for small-molecule control of recombinases. Four recombinases from the tyrosine recombinase family were chosen, including two from phage (Cre and VCre), one from yeast (Flp), and three from the large serine integrase family (PhiC31, TP901, and Bxb1). These recombinases were chosen due to their high activity in mammalian cells and orthogonality with one another^{6,15}. Primary and secondary sequences were first identified in order to identify potential split sites. A crystal structure was available for Cre (PDB: 1XO0) was utilized to determine secondary structure of analogous VCre through HHPred alignment (**Supplementary Figure 1**). Secondary structure annotations for serine integrases were determined from a literature source²⁵ that utilized TP901 integrase N-terminus domain structure and A118 integrase C-terminus domain structure. In general, split sites were chosen to avoid secondary structures and catalytic residues, as we hypothesized introducing CID domains into these locations could permanently disrupt these structures and their roles in recombination (DNA-binding, cleavage, exchange and ligation); however, many splits sites within these structures in Cre recombinase were chosen to counter-test this hypothesis. After a split location (S) was chosen, a “starting fragment” was cloned from the N-terminus of the protein (i.e., beginning from the starting methionine in DNA) to S and a “terminating fragment” from S+1 to the C-terminus of the protein (i.e., to the last base-pair in DNA). Fragments were inserted into vectors containing sequences to code for flexible glycine and serine-rich linkers, nuclear localization signals, and CID domains (**Supplementary Figure 2**). For the rapalog (RAP) system, FRB and FKBP CID domains were used; for the abscisic acid (ABA) system, PYL and ABI; and for the gibberellin (GIB) system, GID1 and GAI. All secondary structure annotations,

plasmid numbers, Addgene deposit codes, and fold changes are provided in **Supplementary File 1**.

Screening for small-molecule inducible dimerization:

Screens for identifying functional split locations in tyrosine recombinases were performed using the GIB CID system (**Figure 1**). In total, 42 locations were tested for Cre recombinase, 20 for Flp, and 20 for VCre. The RAP CID system used for 13 splits of PhiC31 and 11 for TP901, while the GIB CID system was used for screening of two Bxb1 split locations.

A variety of qualitative responses were found: “functionally dead,” yielding no activity with or without addition of the drug; “constitutively active,” yielding partial or full recombination with no significant dependency on the drug; and “responders,” that saw a mild to strong dependency on the drug. A majority of the splits fall into the “functionally dead” category and these are thought to yield no activity due to improper formation of complexes capable of effective recombination. In these cases, CID domains may have been inserted into places critical for recombination processes or addition of CID domains caused too much three-dimensional contortion to perform recombination. Some locations yielded “constitutively active” responses, where CID domains apparently played little or no role in inhibiting and activating recombination. These configurations may splice off a region of the recombinase that may not be critical for a full recombination response. Moreover, resultant split halves of the recombinase could have enough self-association or recruitment through DNA-binding to a recombination site to elicit a drug-independent response as well. Finally, some were successful “responders,” which triggered drug-dependent responses. Within this group, a select few “good responders” yielded high fold-activations with high GFP in the “on” state and low GFP in the “off” state. However,

some less desirable “poor/moderate responders” also produced low levels of GFP expression in the “on” state or high levels of GFP in the “off state.”

Four split Flps with high responses to gibberellin (S = 27, 168, 374, 396) were studied in more detail through transfection of all combinations of each split component. All three components (N-terminal split half, C-terminal split half and drug) were found to be necessary for recombination (**Supplementary Figure 3**). Testing these split configurations without fusion to dimerization domains yielded no appreciable recombination except for one split location (S = 374), which was found to have high recombination activity in contrast to its low recombination response when fused with GIB-responsive GAI and GID1 chemical inducible dimerization domains.

Despite each recombinase having analogous structures, with Cre identifying most closely to VCre, successful split locations did not seem show significant correspondence between recombinases. Interestingly, for Flp, PhiC31, and Bxb1, some locations that yielded strong drug-dependent responses were surprisingly close (<35 amino acids) to either the N-terminus or C-terminus. This result highlights the importance of the native protein structures in these terminal areas for recombination. Of the seven Cre splits (S = 190, 251, 184, 144, 270, 229, and 256) that yielded fold changes over ten, two (S = 144 and 270) were within secondary structures. Overall, 2 out of 19 (10.52%) of the Cre splits within secondary structures yielded fold changes over ten, and 5 out of 23 (21.74%) Cre splits outside secondary structures yielded fold changes over ten. Four of the best performing Cre GIB candidates were compared against a split location determined from a previous work¹⁶ where amino acids 19-59 were used for the “starting fragment” and 60-343 for the “ending fragment.” Despite placing GID1 and GAI domains at varying N-terminal or C-terminal ends for the 19-59/60-343 split, none were close to reaching the low “off” and high “on” responses of the newly identified splits (**Supplementary Figure 4**).

Domain swapping and benchmarking performance:

To gain better insight into the functionality of some of the inducible recombinases, quantitative flow cytometry analyses were performed to characterize performance. Fluorescence channel measurements were converted to units of molecules of equivalent fluorescein (MEFL)^{26,27}, a community-derived standard used to normalize measurements across instruments and fluorescent molecules (see **Methods** section for details). Analyses focused around splits 270, 27 and 233 of Cre, Flp and PhiC31, respectively, with all three CID systems. In addition, the comparison also included a commonly used system in which Cre is fused to a mutated estrogen receptor (ER^{T2})²; this complex remains in the cytoplasm until addition of 4-hydroxytamoxifen (4OHT), which induces translocation to the nucleus and results in site-specific recombination. A signal-to-noise (SNR) metric was used to evaluate how well the plus and minus drug conditions can be distinguished from each other; better separation between signals involves both a higher absolute difference in mean fluorescence intensities but also a lower spread of fluorescence values (noise) within those cell population distributions (**Figure 2a**). Among the three recombinase systems, PhiC31 integrase yielded the highest SNR values; this could be attributed to the “quieter” reporter that is used, which necessitates the complete inversion of the GFP sequence; this is in contrast to using a polyA deletion strategy in the case of tyrosine recombinases, which yields a low but detectable amount of read-through for highly-transfected cells (**Figure 2b, Supplementary Figures 5 - 9**). In the case of Cre and Flp RAP systems, however, SNR reached a maximum and then declined due to “leaky” recombination and noisier GFP expression in the “off” state. Over a time-course of 100 hours, leaky “off” behavior was especially apparent in Cre RAP, Cre 4OHT, and Flp RAP systems, which yielded high SNR values early on, but began to decline after 21 hours. The GIB system consistently resulted in strong performance over all three recombinases and PhiC31 integrase performed well with all three CID systems. With respect to plasmid dosage, as correlated by transfection marker

expression level, Fold Activation (mean “on” divided by mean “off” expression) and SNR values increased as “on” state GFP fluorescence grew due to increased copies of reporter while “off” state GFP fluorescence remained low. However, RAP and 4OHT systems had noticeable values of transfection marker expression where Fold Activation and SNR peaked and then declined. This behavior is attributed to the significant rise in leaky “off” state GFP expression with higher plasmid copy. Moreover, the 4OHT-inducible Cre performance had notable time dependence as well; even at mid-levels of transfection marker (1×10^7 - 4×10^7 MEFL), this system yielded significant decreasing performance over time. These observations highlight the importance of characterized operation ranges with respect to expression level of biological device components and dynamic behavior over time.

An advantage of identifying viable split locations is that they can be applied to other into dimerization systems. Four splits for Cre, four for Flp, three for VCre and four for PhiC31 having moderate to good responses from the initial screens were selected to swap for the other two CID domains, RAP and ABA (**Figures 2c through 2n, Supplementary Figures 10-14**). In general, swapping dimerization domains between GIB/RAP/ABA domains or fusion to different protein termini for the case of Flp recombinase (**Supplementary Figure 15**) yielded drug-dependent responses for the selected split sites; however, many tyrosine recombinase split locations using the RAP system yielded surprisingly high levels of recombination without addition of the drug, converting good responders with the GIB system into near constitutively active systems with RAP, particular for higher values of transfection marker expression.

Light and temperature inducible dimerization:

A variety of light-inducible dimerization (LID) systems are available for applications that necessitate minimal perturbation and precise stimulation. Here, we chose to use a blue LID

system, termed “Magnet”, which uses heterodimerization domains (pMag and nMag) engineered from the blue light photoreceptor VVD from the filamentous fungus *Neurospora crassa*. A programmable blue-light emitting device was used to induce four splits of Cre or Flp that use the pair of Magnet LIDs (**Figure 3**). In addition, a construct that expresses an established³ Cre Magnet (termed PA-Cre) using the 19-59/60-343 split was tested in parallel. Sixty-second blue light pulses were administered every hour over a 24-hour period, 18-hours post-transfection. Blue-light inducible dimerization was achieved for all constructs for Cre and Flp, albeit with different performances amongst splits. Serendipitously, we found that exposing the cells to colder temperatures could also induce recombination, in some cases more than blue light. Temperature induction was performed by exposing the cells to either room temperature (22°C) or refrigerated (4°C) conditions for four hourly 15-minute pulses 18 hours after transfection. The temperature sensitivity of this system may not be completely surprising as VVD has been found to play a role in temperature regulation^{28,29} and others have engineered another closely related blue LID system to be more sensitive to colder temperature³⁰.

Protein and DNA-based logic:

With this repertoire of inducible recombinases capable of responding to different drugs and having various functional split locations, we hypothesized that we could generate multi-input logic gates. The first approach taken was to use multiple CIDs within one recombinase protein. Flp recombinase was split into three segments using two functional split locations found previously ($S_1 = 27$, $S_2 = 396$); the GIB and ABA CID domains were utilized for small-molecule mediated assembly of this tripartite double-split complex (**Figure 4a**). A strong response was found only when both gibberellin and abscisic acid were administered, demonstrating a 2-input AND gate functionality with 125-fold increase in expression upon application of both drugs (**Figure 4b**); however, this did not work for another pair ($S_1 = 168$, $S_2 = 396$) of vetted split

locations (**Supplementary Figure 16**). A second approach taken to achieve multi-input logic operations was to use multiple inducible recombinase systems and integrate those signals with a multi-input reporter DNA plasmid (**Figure 4c**). Three inducible split recombinase systems and a 4OHT-inducible Cre construct were selected. These signals were then integrated using a 4-input AND gate reporter where GFP is expressed only when two termination sequences are deleted and both promoter and GFP sequences are inverted. All combinations of drugs were applied to the cells and the highest GFP expression was found when all four drugs were applied together (>250-fold compared to no drugs) (**Figure 4d**).

Applications of split-recombinases:

The inducible expression of genes using recombinases has wide applications. We demonstrate the multiplexed control of three genes that code for therapeutically relevant cytokines using a “genetic switchboard” consisting of three inducible recombinases and three reporter genes. In this system, we utilize high-performing ABA-inducible Cre, GIB-inducible Flp, and RAP-inducible PhiC31 systems we developed along with Cre-controlled IFN-gamma, Flp-controlled IL-10, and PhiC31-controlled IL-2. (**Figure 5a**). IFN-gamma and IL-2 are important inflammatory cytokines, especially against cancer. IL-10 is a pleiotropic cytokine with both anti-cancer and anti-inflammatory potential. Supernatant cytokine measurement reveals strong activation of indicated cytokines with drugs added separately or in combination (**Figure 5b**). In order to verify if our inducible recombinase systems could operate in the context of a living animal, we developed three xenograft luciferase mouse models (**Figure 5c**). The models utilized ABA, GIB, or RAP-induced activity of Cre, Flp, and PhiC31, respectively, to activate the expression of firefly luciferase. Inducible recombinase systems and reporters were transfected into HEK293FT cells (**Figure 5d**). The following day, transfected cells were transplanted subcutaneously into NSG mice. After another 24 hours, drugs or vehicle-only were injected intraperitoneally. Finally, 24 hours later, mice were given D-luciferin and imaged for luminescence. Imaging reveals

elevation of luminescence for all three inducible recombinase systems (**Figure 5e**) with ABA-Cre and RAP- PhiC31 systems yielding higher fold expression changes than the GIB-Flp system.

DISCUSSION

In this work, we have greatly increased the number of inducible recombinase systems available in mammalian cells. We demonstrated the first successful splitting of VCre, PhiC31, TP901 and Bxb1 recombinases and reconstitution of functionality with chemical-inducible dimerization (CID) systems; this vastly enhances the potential usage of these recombinases, as few or zero inducible systems were available for each recombinase previously. Additional split locations in Cre recombinase were found to yield higher performance than a previously recognized location; using the new split locations, we demonstrate the first gibberellin and abscisic acid inducible Cre systems. Using viable split locations of Cre and Flp recombinases, we incorporate the blue-light inducible Magnet heterodimerization system. Moreover, we also found that we could use a reduction of temperature rather than light illumination to induce recombination with this system, providing yet another way of conditionally activating recombinase responses.

Not only did we find single functional split locations, we found multiple viable candidates for most recombinases. Interestingly, each location yielded a different degree of performance depending on which CID system was utilized. For instance, Cre Magnet (S = 270) yielded surprisingly low sensitivity to blue light despite this split location having strong performance compared with other tested Cre splits using the CID domains. This finding highlights the importance of finding multiple functional split locations for each protein, as they may behave differently for different introduced dimerization domains. Moreover, the usage of multiple sites within one protein permits creation of multi-input logic gates. We demonstrate this idea with a

doubly-split Flp construct that requires both the addition of abscisic acid and gibberellin to function. A tripartite complex such as this will be especially useful for intersectional targeting of cells. Each of the three fragments could be individually expressed by different promoters, thereby further tightening the desired spatial and temporal targeting of recombination. Moreover, we demonstrate a multi-output genetic switchboard that triggers control of three therapeutic proteins upon drug stimulation. Finally, we demonstrate that our inducible systems perform in live animal models, opening the doors to a wide-range of applications.

In sum, we have created the largest-to-date library of well-curated inducible recombinases in mammalian cells for the genetic engineering community. Due to the high prevalence of site-specific recombinase technology, we anticipate many applications and discoveries will come about using our inducible recombinase systems. As such, to ensure ease of access to our reagents, and encourage further development of these tools, we have deposited our reagents in open repository Addgene. Newly discovered recombinases are being increasingly isolated³¹, better high-throughput DNA-assembly techniques could provide faster and automated split site identification capabilities, and ligand-responsive dimerization domains are waiting to be found or engineered.

METHODS

Protein structural analysis and identification:

Secondary structure assignments for Cre were obtained from Protein Database (PDB) crystal structure 1X00_A. HHPred³² was utilized to predict secondary structures for VCre using Cre crystal structure data. ASAView³³ was used to obtain absolute solvent area (solvent accessibility) from Cre. Secondary structure assignments for serine integrases PhiC31, TP901 and Bxb1 were adapted from a literature source²⁵. 3D crystal structure images were created using the Visual Molecular Dynamics software package.

DNA assembly:

A series of sub-cloning vectors containing CID or LID domains were assembled using standard molecular cloning techniques or Gibson isothermal assembly. For each CID or LID, two sub-cloning vectors were made such that a recombinase could be tagged with the domain on the C-terminus or N-terminus and a glycine-serine rich linker L1 (SGGSGSGSSGGSGT). For insertion into a C-terminal CID or LID sub-cloning vector, recombinase “starting fragments” were first PCR-amplified out of a full recombinase sequence template to contain a 5’ sequence to code for an MluI restriction site and a Kozak consensus sequence, and a 3’ sequence to code for a BspEI restriction site. These PCR fragments were then isolated using agarose gel extraction and purification (Epoch Life Science). Finally, the isolated fragments were digested with MluI and BspEI restriction enzymes (New England Biolabs), purified by PCR cleanup (Epoch Life Science) and ligated into an MluI/BspEI-digested CID and LID sub-cloning vectors that were isolated by PCR cleanup (Epoch Life Science). Reaction products were transformed into chemically competent Top10 E. coli cells and selected on LB-agar plates with carbenicillin at

37°C static conditions. Colonies were picked the following day, inoculated into carbenicillin-containing LB medium, and grown in a 37°C shaking incubator (Infors). Once the cells were near or at stationary phase, plasmid DNA was isolated using mini plasmid preparation (Epoch Life Science). Analytical digests with MluI/BspEI were performed and run on gel electrophoresis to assay if a correct product was made. Plasmid clones with correct gel bands were sent for sequencing (Quintara Biosciences). For insertion into a N-terminal CID or LID sub-cloning vector, recombinase “terminating fragments” were first PCR-amplified out of a full recombinase sequence template to contain a 5’ sequence to code for an KpnI-HF restriction site and a 3’ sequence to code for an SV40 nuclear localization signal, stop codon, and an EcoRI-HF restriction site. A similar process was followed as before with the C-terminal CID and LID sub-cloning vectors; however, KpnI-HF and EcoRI-HF restriction digests (New England Biolabs) were carried out rather than with MluI and BspEI. To swap between CID and LID domains, sequenced fragments were gel-isolated from sequenced vectors using MluI/BspEI or KpnI-HF/EcoRI-HF and inserted into corresponding CID or LID sub-cloning vectors. Occasionally, the BspEI site was blocked by *dam* methylation and PCR fragments were utilized instead. The extra central fragment for the trimeric Flp system was created with two different linkers L1 and L2 (GPAGGGSGGGSGGPAG) separating the Flp “central fragment” from the CID domains.

Maintenance and transient DNA transfection of HEK293FT cells:

HEK293FT cells were kept in a static 37°C humidified incubator with 5% CO₂, except for specified periods during temperature induction. Cells were maintained in DMEM medium (Corning) containing 5% fetal bovine serum (ThermoFisher), 50 UI/mL penicillin and 50 µg/mL streptomycin mixture (Corning), 2 mM glutamine (Corning) and 1mM sodium pyruvate (Lonza). A polyethylenimine (PEI) protocol was followed to transfect HEK293FT cells. A master stock of PEI was made by dissolving linear PEI (Polysciences Cat. No. 23966-2) in deionized water at a

concentration of 0.323g/L. Concentrated hydrochloric acid was used to drop the pH to 2-3 while stirring to ensure complete dissolution; sodium hydroxide was added over time to bring the pH to 7.5. The final solution was filter-sterilized (0.22 μ m), aliquoted, and stored at -80°C. For transfections, HEK293FT cells were plated (250 μ L) in tissue culture-treated 48-well plates (Fisher Scientific) one- or two-days prior such they would be 50-70% confluent on the day of transfection. For most induction experiments in this work, 2000 ng (at 50ng/ μ L) of DNA was prepared with specific combinations of plasmids and brought up to a volume of 100 μ L using 0.15 M sodium chloride (NaCl, Fisher Scientific). Next, a PEI/NaCl mixture of 16 μ L PEI stock and 84 μ L NaCl, was mixed with the DNA-NaCl solution and incubated at room temperature for at least ten minutes. Finally, 25 μ L of this solution was carefully added to each well of cells that were to be assayed; three wells for non-treated cells and three wells for drug or light-treated. For time-course experiments, this was similarly followed, rather with eight times the amount of each component prior to dispensing of 25 μ L into each well. DNA mass values for transfected plasmids are described in **Supplementary File 1**.

Small-molecule chemical induction:

1000X stocks of abscisic acid (100mM, Gold Biotechnology Cat. No. A-050-500) and gibberellin ester (10mM, Toronto Research Chemicals Cat. No. G377500) were made in 100% ethanol and stored at -20°C. A 1000X stock of rapalog (A/C heterodimerizer AP21967) was obtained commercially (0.5mM, Takara Cat. No. 635056) and stored at -20°C. A 1000X stock of 4-hydroxytamoxifen (1mM, Sigma Aldrich) was made in methanol and stored at -20°C or -80°C. These molecules were added to cell cultures such that the final concentration was 1X at the time of induction. For mouse experiments, stocks of abscisic acid (40mg/mL), gibberellin ester (4mg/mL), and rapalog (0.6mg/mL, Takara Cat. No. 635055) were dissolved into vehicle solutions containing 16.7% propanediol, 22.5% PEG-400 and 1.25% Tween-80 through

vigorous vortexing and water bath sonication. Vehicle components were purchased from Sigma Aldrich (Cat. No. 82280, 91893, and P4780).

Light illumination and temperature induction.

A modified large blue light transilluminator (IO Rodeo) was used in order to illuminate samples with 470 nm light. In order to control the timing and duration of light exposure, the included 18V wall adapter was cut and exposed leads were used to power both an ARDUINO® UNO microcontroller and the light array. The negative terminal of the array was connected to the collector of two NPN N2222 transistors in parallel. The signal pins of each transistor were then connected to pin 9 of the microcontroller. The UNO was programmed to send a signal through pin 9 to the transistors, activating the light box for a set duration at specific time intervals (e.g., turn on for one minute once every hour). PWM Pin 9 was used so that one could turn on the array at varying intensities. At maximum intensity, the light array achieved an illumination of 1 mW/m² as measured by a photodiode power sensor (S120C, Thorlabs) placed at the sample distance.

In order to adjust the timing and intensity of light exposure, an HC-06 ARDUINO® Bluetooth module was installed, allowing the device to communicate with an Android™ smartphone application made with MIT App Inventor. The application allows the user to set the length of exposure, time between exposures, number of exposures, the delay time before the exposure cycles began, and the relative intensity of the LED array.

Plates containing the samples that were to be illuminated were placed onto the light array directly after transfection in the 37°C humidified incubator and a sheet of aluminum foil was used to cover the entire apparatus. The plates not treated for light illumination were wrapped in

aluminum foil directly after transfection. These “dark” plates were either kept constantly in the 37°C incubator until flow cytometry or removed starting at 18 hours post-transfection and placed outside the incubator at room temperature (22°C) or in a refrigerator at 4°C; these temperature pulses were performed four times, once every hour for fifteen minute durations.

Enzyme-linked immunoabsorbance assay:

Cytokine concentrations for **Figure 5b** were measured using enzyme-linked immunoabsorbance assays (ELISA) from Becton Dickinson (Cat. No. 555142, 555190, 555157 and 550534).

Absorbance measurements of supernatant cytokine concentration were made using a Molecular Devices SpectraMax M5 plate reader at 450nm with wavelength correction at 570nm.

Flow cytometry:

For all cytometry experiments, except for given times during time-course experiments, plates were removed from the incubator two days post-transfection. Cells were trypsinized using 0.05% trypsin/0.53 mM EDTA (Corning) and resuspended with cell culture medium to a volume of 200 μ L in 96 U-bottom plates. All cell populations were analyzed using a ThermoFisher Attune Nxt cytometer equipped with four lasers. GFP was detected using a 488 nm blue laser and a 510/10 bandpass emission filter; a 530/30 bandpass filter was used for experiments that converted to MEFL units. mTagBFP was detected using a 405 nm violet laser and 440/50 bandpass emission filter. iRFP720 was detected using a 638 nm red laser and 720/30 bandpass emission filter. Experiments in **Figure 2** were analyzed using TASBE Flow Analytics, an open source flow cytometry software tool^{26,34}. In short, fluorescence channel measurements were converted to units of MEFL using 8-peak rainbow beads (Spherotech RCP-30-5A) and using linear color translation and compensation between fluorophores. Gaussian mixture models were

used to isolate viable cell populations and remove debris. Results were analyzed above a cutoff of 10^6 MEFL on the constitutive transfection marker. For all other experiments, results were analyzed using FlowJo 9 (TreeStar), using a polygonal FSC-A/SSC-A gate to remove debris from the mammalian cell populations. An mTagBFP fluorescent protein was used as a transfection marker and positive BFP cells were first gated, then mean GFP fluorescence determined afterwards.

Xenograft Mouse Models:

Female NSG mice, 4-6 weeks of age, were purchased from Jackson Laboratories (#005557) and maintained in the BUMC Animal Science Center (ASC). All protocols were approved by the Institutional Animal Care and Use Committee at BUMC. In order to carry out the xenograft models, NSG mice were initially injected with 20×10^6 of transiently transfected HEK293FT cells subcutaneously. After 24hr, mice were randomly grouped, and different drugs were injected intraperitoneally (200mg/kg abscisic acid, 20mg/kg gibberellin ester, 0.06mg/kg rapalog). After 24hr of drug injection, luciferase signal was measured by IVIS Spectrum (Xenogen) and was quantified as total flux (photons per sec) in the region of interest to verify split-recombinase activity. Images were acquired within 30 minutes following intraperitoneal injection of 150mg/kg of D-luciferin (PerkinElmer #122799).

Statistical methods:

All transient gene expression experiments involved transfection of DNA into $n = 3$ separate cell cultures. Fluorescence intensities for each cell culture population was averaged and the s.e.m. or s.d. was taken as noted.

Data availability:

Data are available from the corresponding authors upon reasonable request. Plasmids and construction resources are available at <https://www.addgene.org/browse/article/28193023>

ACKNOWLEDGEMENTS

B.H.W. acknowledges funding from the NSF Graduate Research Fellowship Program (DGE-1247312), a NIH/NIGMS fellowship (T32-GM008764) and the NSF Expeditions in Computing Award No. 1522074, which is part of the “Living Computing Project” (<https://www.programmingbiology.org/>). W.W.W. acknowledges funding from the NIH Director's New Innovator Award (1DP2CA186574), NSF Expeditions in Computing (1522074), NSF CAREER (162457), NSF BBSRC (1614642), NSF EAGER (1645169), and Boston University College of Engineering Dean's Catalyst Award. J.B. acknowledges funding from the NSF Expeditions in Computing (1522074) grant. We acknowledge S. Koplik and T. Vantrimpont for contributions to preparations of molecular reagents. We thank D. Densmore and D. Chakravarti for suggestions on the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization, B.H.W. and W.W.W.; Methodology, B.H.W. and B.L.; Software, W.K.J.B. and J.B.; Validation, B.H.W., C.J.H., Y.A. and N.T.; Formal Analysis, B.H.W., C.J.H., Y.A. J.T. and J.B.; Investigation, B.H.W., C.J.H., Y.A., N.T., M.W., M.T., C.O., J.T. and B.L.; Resources, B.H.W., Y.A., N.T.H.P., L.D.C., N.T., M.W., M.T., C.O., B.L., J.T. and W.K.J.B.; Writing – Original Draft, B.H.W.; Writing – Review & Editing, B.H.W., C.J.H., N.T.H.P., J.B. and W.W.W.; Visualization, B.H.W.; Supervision, B.H.W., J.B., and W.W.W.; Funding Acquisition, B.H.W., J.B., and W.W.W.

COMPETING INTERESTS

W.W.W. and B.H.W. have a patent application pending USPTO 15/838,598 whose value may be affected by the publication of this paper. W.W.W has consulted for and own shares in Senti Biosciences.

REFERENCES

- 1 Nagy, A. Cre recombinase: the universal reagent for genome tailoring. *Genesis* **26**, 99-109, doi:10.1002/(SICI)1526-968X(200002)26:2<99::AID-GENE1>3.0.CO;2-B [pii] (2000).
- 2 Feil, R., Wagner, J., Metzger, D. & Chambon, P. Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem Biophys Res Commun* **237**, 752-757, doi:10.1006/bbrc.1997.7124 (1997).
- 3 Kawano, F., Okazaki, R., Yazawa, M. & Sato, M. A photoactivatable Cre-loxP recombination system for optogenetic genome engineering. *Nat Chem Biol* **12**, 1059-1064, doi:10.1038/nchembio.2205 (2016).
- 4 Ham, T. S., Lee, S. K., Keasling, J. D. & Arkin, A. P. Design and construction of a double inversion recombination switch for heritable sequential genetic memory. *PLoS One* **3**, e2815, doi:10.1371/journal.pone.0002815 (2008).
- 5 Roquet, N., Soleimany, A. P., Ferris, A. C., Aaronson, S. & Lu, T. K. Synthetic recombinase-based state machines in living cells. *Science* **353**, aad8559, doi:10.1126/science.aad8559 (2016).
- 6 Weinberg, B. H. *et al.* Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat Biotechnol* **35**, 453-462, doi:10.1038/nbt.3805 (2017).
- 7 Siuti, P., Yazbek, J. & Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nat Biotechnol* **31**, 448-452, doi:10.1038/nbt.2510 nbt.2510 [pii] (2013).
- 8 Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P. & Endy, D. Amplifying genetic logic gates. *Science* **340**, 599-603, doi:10.1126/science.1232758 (2013).
- 9 Hingorani, S. R. *et al.* Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **4**, 437-450 (2003).
- 10 Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev* **15**, 3243-3248, doi:10.1101/gad.943001 (2001).
- 11 Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci* **13**, 133-140, doi:10.1038/nn.2467 (2010).
- 12 Schonhuber, N. *et al.* A next-generation dual-recombinase system for time- and host-specific targeting of pancreatic cancer. *Nat Med* **20**, 1340-1347, doi:10.1038/nm.3646 (2014).
- 13 Ventura, A. *et al.* Restoration of p53 function leads to tumour regression in vivo. *Nature* **445**, 661-665, doi:10.1038/nature05541 (2007).
- 14 Guenther, C. J., Miyamichi, K., Yang, H. H., Heller, H. C. & Luo, L. Permanent genetic access to transiently active neurons via TRAP: targeted recombination in active populations. *Neuron* **78**, 773-784, doi:10.1016/j.neuron.2013.03.025 (2013).
- 15 Xu, Z. *et al.* Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnol* **13**, 87, doi:10.1186/1472-6750-13-87 1472-6750-13-87 [pii] (2013).

- 16 Jullien, N., Sampieri, F., Enjalbert, A. & Herman, J. P. Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. *Nucleic Acids Res* **31**, e131 (2003).
- 17 Liang, F. S., Ho, W. Q. & Crabtree, G. R. Engineering the ABA plant stress pathway for regulation of induced proximity. *Sci Signal* **4**, rs2, doi:10.1126/scisignal.2001449 4/164/rs2 [pii] (2011).
- 18 Razavi, S., Su, S. & Inoue, T. Cellular signaling circuits interfaced with synthetic, post-translational, negating Boolean logic devices. *ACS Synth Biol* **3**, 676-685, doi:10.1021/sb500222z (2014).
- 19 Miyamoto, T. *et al.* Rapid and orthogonal logic gating with a gibberellin-induced dimerization system. *Nat Chem Biol* **8**, 465-470, doi:10.1038/nchembio.922 nchembio.922 [pii] (2012).
- 20 Sando, R., 3rd *et al.* Inducible control of gene expression with destabilized Cre. *Nat Methods* **10**, 1085-1088, doi:10.1038/nmeth.2640 nmeth.2640 [pii] (2013).
- 21 Kellendonk, C. *et al.* Regulation of Cre recombinase activity by the synthetic steroid RU 486. *Nucleic Acids Res* **24**, 1404-1411 (1996).
- 22 Metzger, D., Clifford, J., Chiba, H. & Chambon, P. Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase. *Proc Natl Acad Sci U S A* **92**, 6991-6995 (1995).
- 23 Hirrlinger, J. *et al.* Split-cre complementation indicates coincident activity of different genes in vivo. *PLoS One* **4**, e4286, doi:10.1371/journal.pone.0004286 (2009).
- 24 Backman, C. M. *et al.* Generalized tetracycline induced Cre recombinase expression through the ROSA26 locus of recombinant mice. *J Neurosci Methods* **176**, 16-23, doi:10.1016/j.jneumeth.2008.08.024 (2009).
- 25 Van Duynes, G. D. & Rutherford, K. Large serine recombinase domain structure and attachment site binding. *Crit Rev Biochem Mol Biol* **48**, 476-491, doi:10.3109/10409238.2013.831807 (2013).
- 26 Beal, J. Bridging the gap: a roadmap to breaking the biological design barrier. *Front Bioeng Biotechnol* **2**, 87, doi:10.3389/fbioe.2014.00087 (2014).
- 27 Wang, L. L., DeRose, P. & Gaigalas, A. K. Assignment of the Number of Equivalent Reference Fluorophores to Dyed Microspheres. *J Res Natl Inst Stan* **121**, 264-281, doi:10.6028/jres.121.012 (2016).
- 28 Hunt, S. M., Elvin, M., Crosthwaite, S. K. & Heintzen, C. The PAS/LOV protein VIVID controls temperature compensation of circadian clock phase and development in *Neurospora crassa*. *Genes Dev* **21**, 1964-1974, doi:10.1101/gad.437107 (2007).
- 29 Castrillo, M. *et al.* Transcriptional basis of enhanced photoinduction of carotenoid biosynthesis at low temperature in the fungus *Neurospora crassa*. *Res Microbiol*, doi:10.1016/j.resmic.2017.11.003 (2017).
- 30 Richter, F. *et al.* Engineering of temperature- and light-switchable Cas9 variants. *Nucleic Acids Res* **44**, 10003-10014, doi:10.1093/nar/gkw930 (2016).
- 31 Yang, L. *et al.* Permanent genetic memory with >1-byte capacity. *Nat Methods*, doi:10.1038/nmeth.3147 (2014).
- 32 Soding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**, W244-248, doi:10.1093/nar/gki408 (2005).
- 33 Ahmad, S., Gromiha, M., Fawareh, H. & Sarai, A. ASAView: database and tool for solvent accessibility representation in proteins. *BMC Bioinformatics* **5**, 51, doi:10.1186/1471-2105-5-51 (2004).

- 34 Beal, J., Weiss, R., Yaman, F., Davidsohn, N. & Adler, A. A Method for Fast, High-Precision Characterization of Synthetic Biology Devices. (Massachusetts Institute of Technology, Cambridge, MA, 2012).
- 35 Guo, F., Gopaul, D. N. & van Duyne, G. D. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389**, 40-46, doi:10.1038/37925 (1997).

FIGURE LEGENDS

Figure 1. Screens of split locations yield small-molecule dependent recombination responses.

Cre, Flp, VCre, PhiC31, and TP901 recombinases (**a** through **f**) were split into two fragments at amino acid S and fused to gibberellin-inducible (GAI/GID1) or rapalog (FKBP/FRB) heterodimerization domains. Splits were made at various amino acid locations; pink, blue and gray shaded regions indicate alpha helices, beta sheets, or undetermined structures, respectively. Split recombinases were transfected along with reporters that yield green fluorescence protein (GFP) expression upon site-specific recombination (**g** through **m**). Constitutive recombinases (black square) or a blank vector (gray squares) were transfected to indicate highest or lowest expected GFP expression. Split recombinases were also transfected, and drug was either added (green squares) or not added (white squares) to the cell culture medium after two hours. M.F.I. indicates mean fluorescence intensity measured with arbitrary units (a.u.). Errors bars represent arithmetic standard error of the mean of three transfected cell cultures.

Figure 2. Swapping chemical-inducible dimerization domains yield a large repertoire of small-molecule inducible recombinases.

(**a**) A Signal-to-Noise (SNR) metric can be used to capture distinguishability of “on” and “off” states, accounting for both the absolute difference in mean signal expression and spread (“noise”) of the distributions. SNR dynamics in units of decibels (dB) were captured over 100 hours (**b**) post-transfection for particular amino acid splits of inducible Cre, Flp and PhiC31 systems incorporating the gibberellin (GIB), rapalog (RAP) and abscisic acid (ABA) dimerization domains GAI/GID1, FKBP/FRB and PYL/ABI, respectively. A 4-hydroxytamoxifen (4OHT)-inducible pCAG-ER^{T2}-Cre-ER^{T2} construct is also included in (**b**). Various splits of Cre (**c** and **d**)

Flp (**e** and **f**), VCre (**g** and **h**), PhiC31 (**i** and **j**) TP901 (**k** and **l**) and Bxb1 (**m** and **n**) were incorporated with the GIB-, RAP- and ABA-associated dimerization domains. Measurements of GFP mean fluorescence intensity (M.F.I.) in units of molecules of equivalent fluorescein (MEFL) and SNR were made for plus drug (colored blue, purple, and pink squares corresponding to gibberellin, rapalog, and abscisic acid added two hours post-transfection) and minus drug (white squares) conditions. Constitutive recombinases (black squares) and blank vectors (gray squares) were also transfected as a comparison. Red diamonds indicate SNR score. Error bars of M.F.I. indicate the geometric standard deviation of means for three transfected cell cultures.

Figure 3. Light and temperature-inducible recombinases.

(**a**) A schematic showing a split Cre fused to nMag and pMag blue light-inducible dimerization domains. (**b**) Pulses indicating the changing light (left) and temperature (center and right) conditions. Plots indicate GFP mean fluorescence intensity (M.F.I.) of various nMag/pMag tagged split Cre (**c**) and Flp (**d**) recombinases under blue light (blue squares), room temperature (pink squares) and refrigeration (purple squares). Black squares indicate samples treated with constant temperature and dark conditions. PA-Cre indicates expression of a previously developed blue-light inducible Cre recombinase³. Inducible recombinases were compared with samples transfected with constitutive recombinases or a blank vector. Error bars of M.F.I. indicate the arithmetic standard error of the mean between three transfected cell cultures.

Figure 4. Protein and DNA-based recombinase logic gates.

(**a**) A two-input protein-based AND gate is created by splitting Flp recombinase at two locations ($S_1 = 27$, $S_2 = 396$) and fusing to the gibberellin (GIB) and abscisic acid (ABA)-associated chemical inducible dimerization domains. (**b**) Plotted results indicate GFP mean fluorescence intensity (M.F.I.) of four conditions of GIB and ABA. (**c**) Four chemical inducible recombinases can be used together to generate a DNA-based 4-input logic gate; these include a gibberellin-

inducible split Bxb1 (S = 468), 4-hydroxytamoxifen (4OHT)-inducible ER^{T2}-Cre-ER^{T2}, abscisic acid (ABA)-inducible split Flp (S = 396), and rapalog (RAP)-inducible split PhiC31 (S = 571). Drugs were added two hours post-transfection. **(d)** Plotted results indicate GFP M.F.I. of sixteen conditions of GIB, 4OHT, ABA and RAP. Error bars of M.F.I. indicate the arithmetic standard error of the mean between three transfected cell cultures.

Figure 5. Applications of split recombinases.

(a) A genetic switchboard using three inducible split recombinases to individually control expression of three therapeutically relevant human cytokines, interferon gamma (IFN-gamma), interleukin 2 (IL2) and interleukin 10 (IL10). Drugs were added one day post-transfection. **(b)** Measurement of supernatant cytokine concentration using enzyme-linked immunosorbent assays (n = 3). **(c)** A xenograft mouse model of split recombinases used in **(a)** that were transfected in human embryonic kidney (HEK) cells alongside a corresponding luciferase reporter that expresses upon site-specific recombination. Transfected cells are transplanted subcutaneously (s.c.) and drugs or vehicle-only control are injected intraperitoneally (i.p.). **(d)** Time-course depicting time of transfection, injections and imaging. **(e)** Images and averaged values of emitted luminescence from three split recombinase systems. Error bars of mean cytokine and luminescence values indicate the arithmetic standard error of the mean.

FIGURE 1

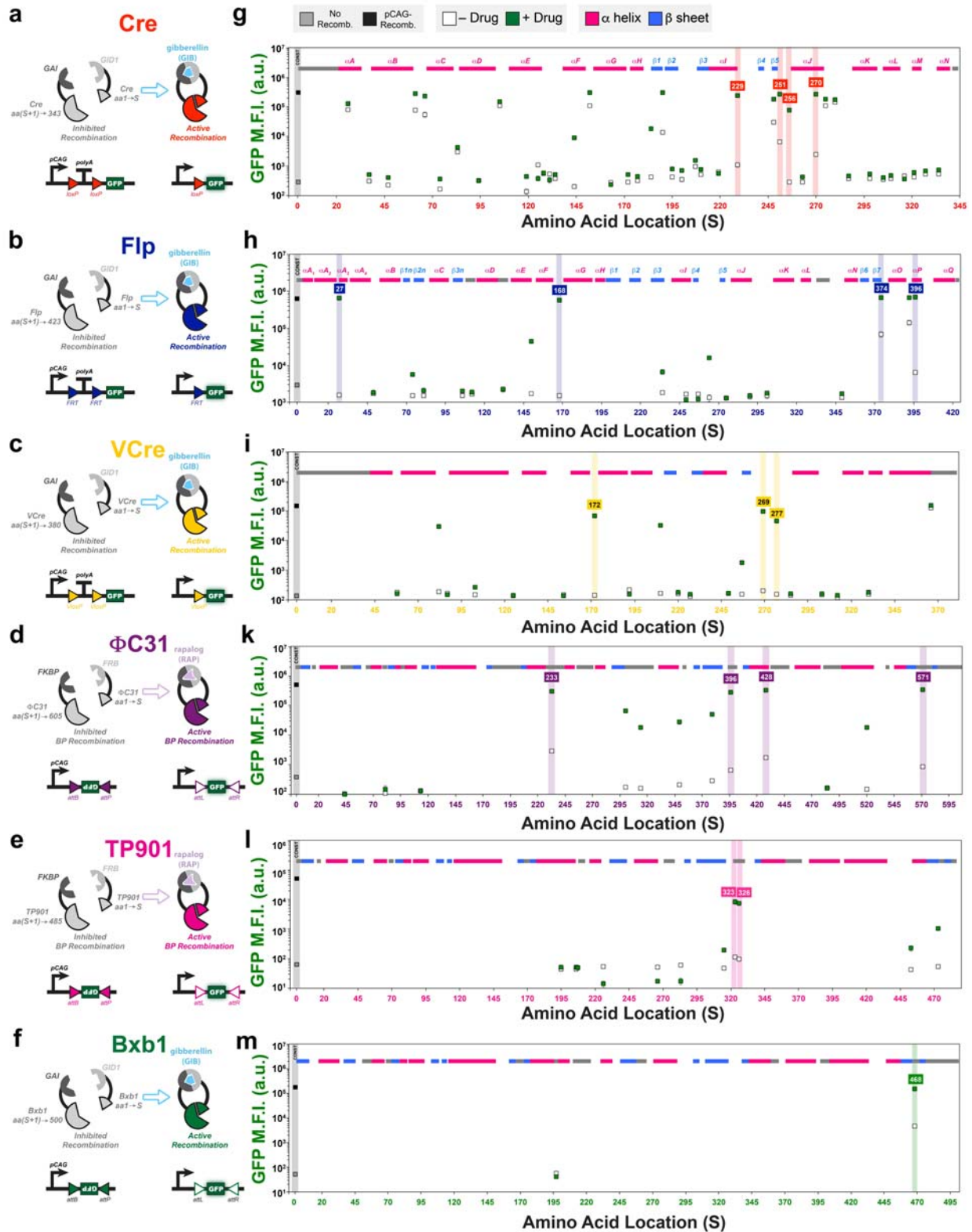


FIGURE 2

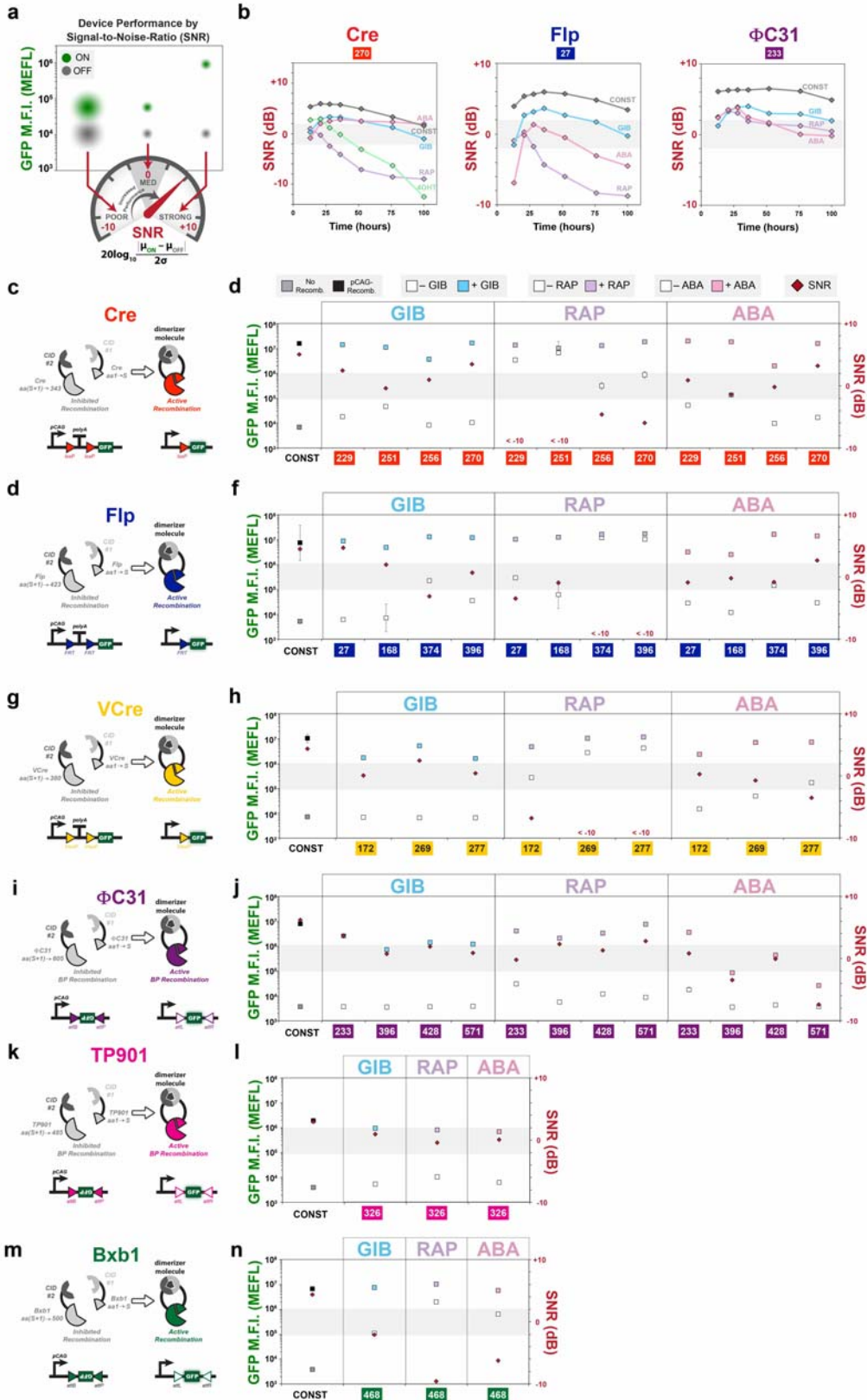


FIGURE 3

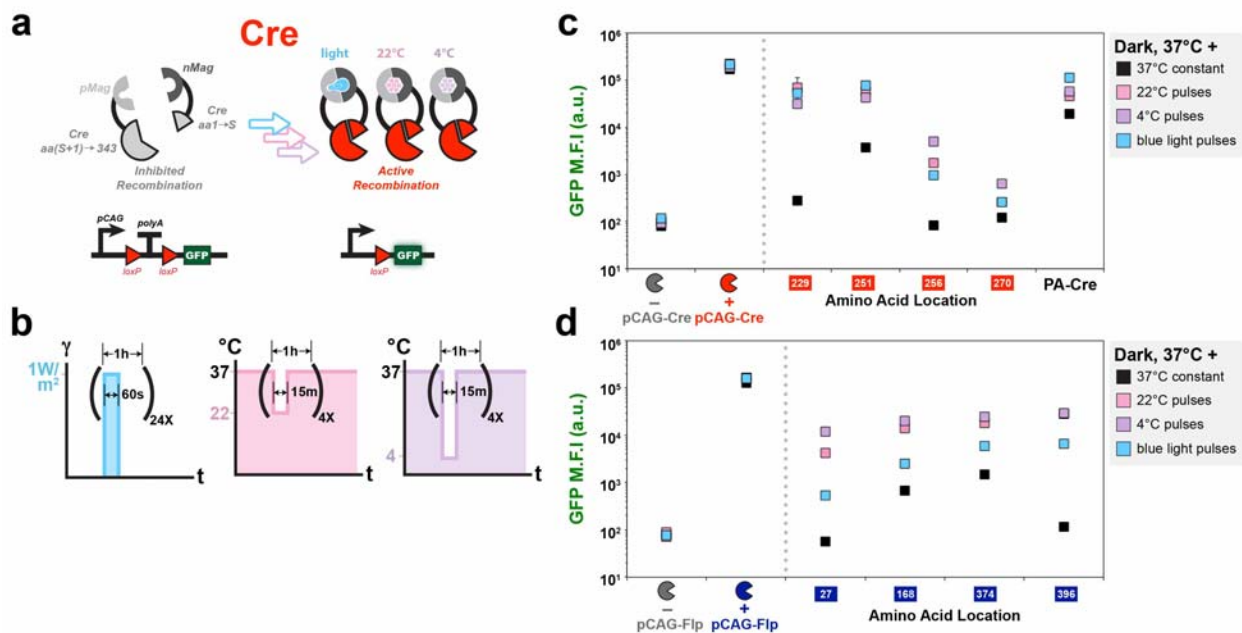


FIGURE 4

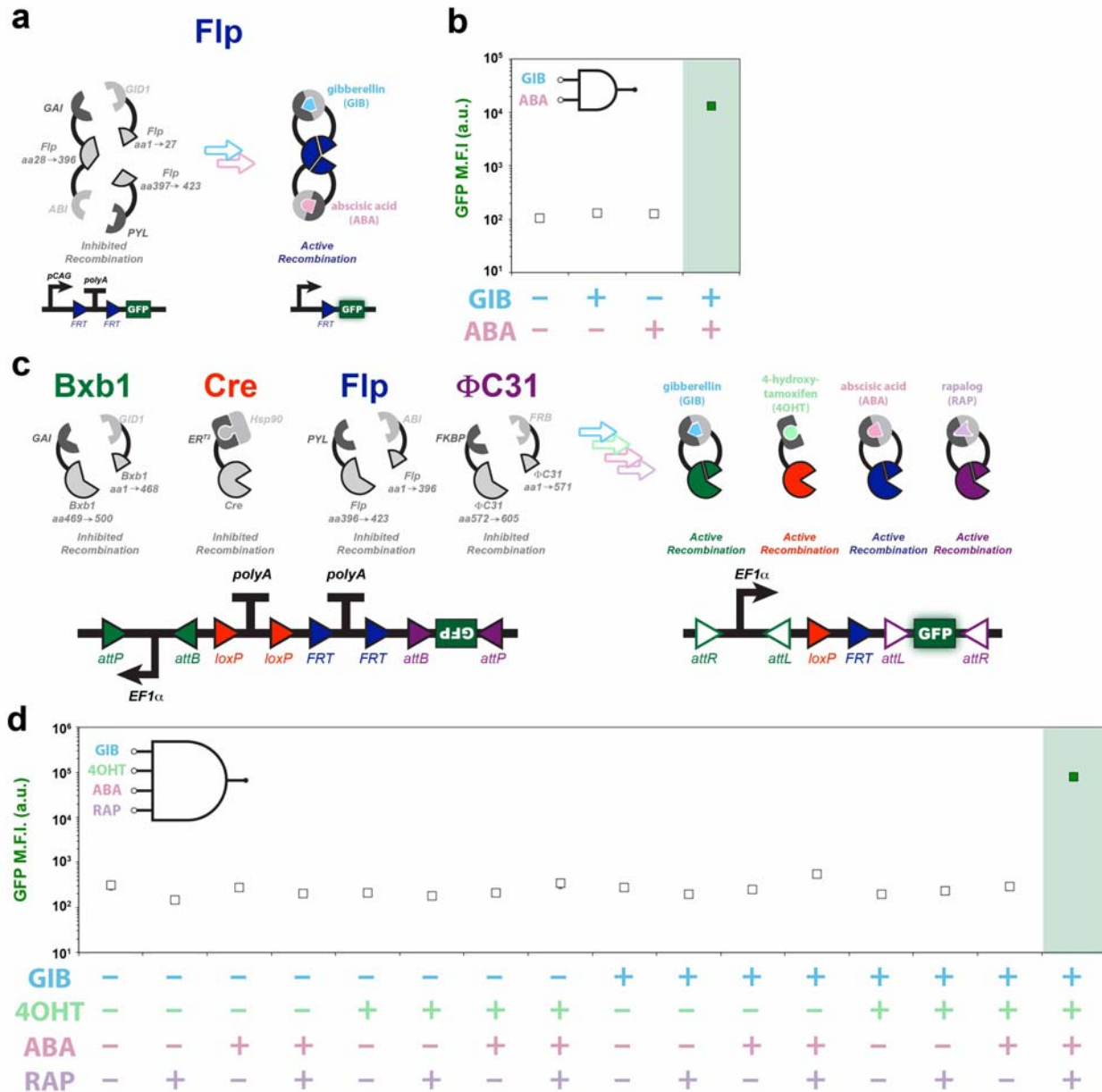


FIGURE 5

