THE UNIVERSITY OF CHICAGO

DEVELOPMENT OF CHEMICAL AND PROTEIN TOOLS TO INTERROGATE AND HARNESS BIOMOLECULAR INTERACTIONS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

KAITLIN MARY KENTALA

CHICAGO, ILLINOIS

AUGUST 2021

This work is dedicated to my family, whose unconditional love and unwavering support have continuously enabled me to dream big and never give up.

“Striving toward a Ph.D. is like running a marathon…when you have never run one before…you have never trained for one…and you don’t even realize you are running one until you are already several paces in. There are uphill climbs, but there are also downhill breathers; there are serious leg cramps…particularly around mile 24…but there are also refueling stations. And in the end, in spite of, or perhaps because of it all, you cross the finish line having accomplished something incredible.”

-Me, gasping along somewhere around mile 24

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF NOTES

ACKNOWLEDGEMENTS

I would like to thank my advisor, Prof. Bryan Dickinson, for his academic and financial support throughout my graduate career. As one of the first students in the group, I was continuously inspired by Bryan's enthusiasm for science, and it was a privilege to watch the lab grow and change over time. Bryan's support throughout the ups and downs of the PhD proved invaluable. My labmates (special shout-out to Simone Rauch, Jeff Dewey, and Saara-Anne Azizi) also deserve my never-ending gratitude for making work enjoyable and for the wonderful discussions—both scientific and life-centric—that shaped me as a scientist and individual. Many thanks go to my collaborators: Jinyue Pu, Krysten Jones, and Tian Qiu, in particular, for providing wonderful mentorship and partnership and without whom the many projects I was involved with would not have been possible.

I would also like to thank Profs. Joe Piccirilli and Yamuna Krishnan for serving on my dissertation committee and for their support during my time at The University of Chicago. Joe was instrumental in my being awarded the NIH CBI Training Grant as well as first introducing me to the world of RNA biology in his class, and Yamuna was always counted on for wonderful discussions about science, and at department functions and beyond. Her many insights shared during my candidacy exam helped mold my approach to my research going forward.

Finally, I would like to thank my family and friends, without whom I would not be where I am, or who I am, today. They were my bedrock throughout the PhD journey, urging me to push through the difficulties and celebrate the successes, all while reminding me not to lose sight of what is truly most important.

ABSTRACT

When we consider the roles biopolymers play in our cells, the central dogma of biology generally comes to mind: DNA stores information to be read out by cellular machinery transcribing it into RNA which encodes proteins responsible for catalysis. While these functions are critical, they do not account for the numerous secondary and tertiary phenomena responsible for the bulk of cellular function: all three biopolymers can be decorated with various modifications, and, in addition to or in concert with secondary structural changes, interact transiently with each other through protein-protein, RNA-protein, and DNA-protein interactions. These dynamic processes dictate whether DNA is accessible for transcription, whether resultant RNA will be processed normally or degraded/upregulated, and which competition-mediated signaling pathways might be activated downstream, to name a few. Due to this complexity, it has been historically challenging to measure or harness such interactions. Therefore, in this thesis we will cover the development of several tools designed to tackle these challenges. The first exploits the RNA output created by two small molecule-sensing T7 polymerase-based biosensors in order to control Cas9 activity in cells. The second harnesses the interchangeability-potential of ester-caged imaging and bioactive molecules using a split esterase biosensor to study protein-protein interactions. The third further employs the esterase-based technology toward the creation of an RNA proximity labeling method via a unique ester-masked enol ester acylating reagent. Finally, we will briefly discuss efforts toward an esterase-based “synthetic” bioluminescent sensitive *in vivo* imaging system. Taken together, this work generated technologies that will help better understand biomolecular interactions and harness them to control cellular processes.

LIST OF PUBLICATIONS BASED ON WORK IN THIS THESIS

1. Pu, J.; Kentala, K.; Dickinson, B. C. Multidimensional control of Cas9 by evolved RNA polymerase-based biosensors. *ACS Chem Biol* **2018**, *13* (2), 431-437. (Adapted, Chapter 2)

2. Jones, K. A.; Kentala, K.; An, W; Lippert, A.; Lewis, J.; Dickinson, B. C. Development of a split esterase for protein-protein interaction-dependent small-molecule activation. *ACS Cent Sci* **2019**, *5*(11), 1768-1776. (Adapted, Chapter 3)

3. Kentala, K*.; Qiu, T*.; Dickinson, B.C. Development of an RNA proximity labeling method that deploys a unique ester-masked enol ester acylating reagent. *Manuscript in preparation*. (Adapted, Chapter 4) **Denotes equal contribution.*

4. Kentala, K.; Azizi, S.-A.; An, W.; Beck, M. W.; Lippert, A.; Dickinson, B. C. Toward a synthetic bioluminescent reporter system for sensitive in vivo imaging. *Unpublished manuscript*. (Adapted, Appendix A)

CHAPTER 1

INTRODUCTION

1.1 Overview

At the core of our understanding of biology lies the central dogma, which defines the flow of cellular information: DNA, to RNA, to protein. Guided by this principle, modern genomics, transcriptomics, and proteomics approaches to identify and quantify the presence of these biopolymers have delivered us valuable and comprehensive information on presence, abundance, and activity. It has become increasingly clear, however, that nucleic acids do more than simply encode proteins, and proteins do more than just catalyze reactions: transient interactions involving proteins and nucleic acids, as well as their asymmetric subcellular organizations, are critical actors in cell signaling pathways, disease progression, and the general modulation of biological systems on a molecular level (**Figure 1.1**).[1-6] Indeed, they underpin *most* cellular processes, and yet have been historically challenging to interrogate in the absence of tools that can capture them. As such, methods that can report on, or even control, such interactions have exploded in scope and demand.

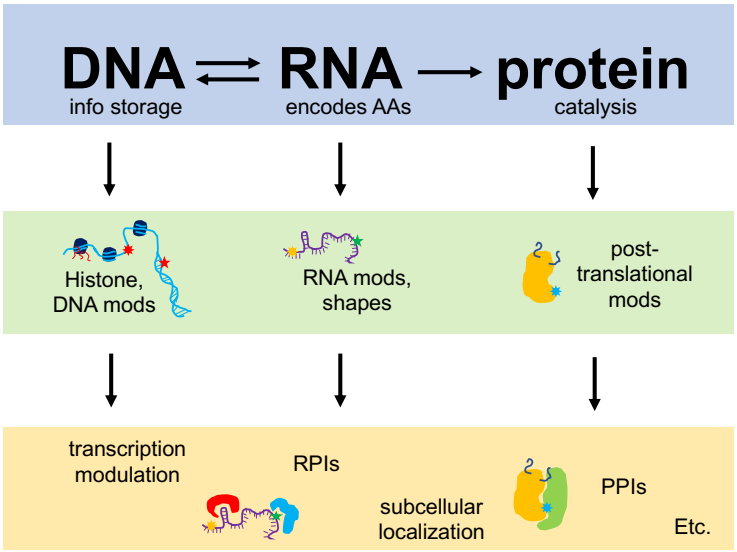


Figure 1.1 Looking beyond the central dogma. Secondary & tertiary phenomena drive outcomes.

1.2 Protein-Fragment Complementation Technologies

One approach to studying biomolecular interactions, and in particular protein-protein interactions (PPIs), is protein fragment complementation.[7,8] These systems report on a given interaction via split reconstitution, in which each half of a given "reporter" is fused to a respective protein of interest, and upon the interaction of those proteins the reporter becomes whole and active and can measure the event (**Figure 1.2**). Examples include split fluorescent protein-based reporters like split GFP,[9,10] as well as split enzymatic reporters such as split horseradish peroxidase[11] and split luciferases.[12,13] Enzymatic reporters, generally optimized for imaging applications, offer the advantage of multiple turnover of reporting molecules and so have found more broad use owing to improved signal-to-noise. Split reporters whose readouts are not imaging based, such as split T7 RNA polymerase biosensors,[14-16] split proteases,[17,18] and split Cas systems,[19] have provided expanded functionalities in which each interaction event can induce downstream cellular effects such as driving gene expression, controlling gene editing, and other cell engineering applications.

Figure 1.2 General protein-fragment complementation schematic. PPIs drive the reconstitution of whole, active reporter systems.

1.3 Proximity Labeling Technologies

Another tactic for studying biomolecular interactions centers on a family of technologies designed to interrogate interactions and localizations without the requirement of knowing exactly which biomolecules you are interested in assessing.

Proximity labeling is conceptually based upon the idea that if you localize a "catalyst," most often an enzyme, to a protein or region of interest in a cell and introduce some type of reagent that only becomes reactive upon activation by said catalyst, you will label biomolecules surrounding the catalyst as the reactive moiety radiates outward from the point of activation (**Figure 1.3**).[20,21] Designing reagents that contain functional handles, such as azides or biotins, allow for downstream analysis of the labeled biomolecules, giving us a picture of which biomolecules were where at a given time in a cell. Most proximity labeling technologies have been developed and widely deployed to study previously unknown protein localizations and interactions. Most notably, BioID/TurboID/miniTurbo,[22-24] and APEX2-based systems[25-27] have harnessed evolved biotin ligase BirA and an engineered ascorbate peroxidase, respectively, to label and capture proteins proximal to cellular spaces and proteins of interest to which the enzymes have been genetically encoded. The former directly biotinylates proteins while the latter generates a phenoxyl radical labeling probe. μMAP approaches this challenge differently by deploying a catalyst that is not enzyme-based, but rather conjugates an iridium blue-light activated photocatalyst

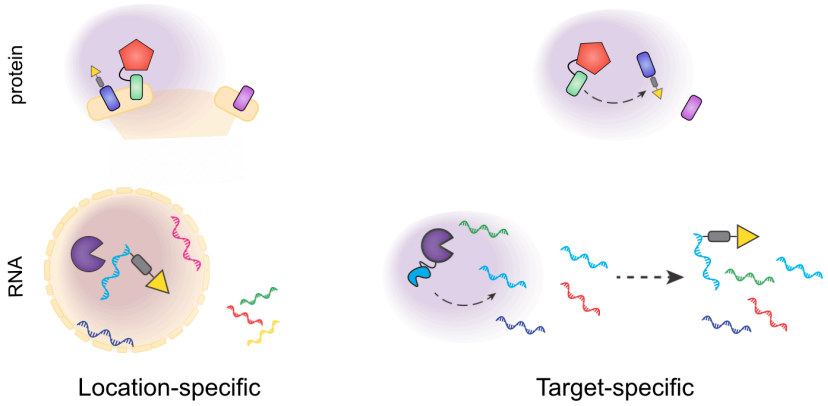


Figure 1.3 General proximity labeling schematic. A catalyst is loaded onto a protein or location of interest. Labeling probe radiates outward from the activation point, tagging proximal biopolymers before being otherwise quenched and unable to label those further away.

to antibodies directed toward proteins of interest which subsequently activate diazirines to form triplet carbenes which rapidly label proximal proteins.[28] These systems have collectively been used to study critical cellular actors such as oncogenic interactomes,[29] organellar- and sub-organellar protein-protein contacts,[30] and ligand-dependent GPCR signaling,[31,32] to name a few. Additionally, the genetic encodability and rapid reactivity of APEX2, in particular, has allowed dCas9-directed studies of proteins at specific genetic loci.[33,34]

1.4 Current Challenges, Opportunities, and the Scope of this Thesis

As discussed above, most protein-fragment complementation technologies are designed for improved imaging applications. This is very useful for quick, sensitive readouts of whether or not a given interaction is occurring, and split NanoLuciferase,[35] in particular, is widely regarded as the gold standard for this application owing to its remarkably good signal-to-noise ratio and rapid scalability for high throughput screening. However, this system is quite limited beyond the scope of imaging because it has been highly evolved for a complex natural substrate that provides little room for multiplexability or interchangeability, though work has been done toward the former.[36] As such, our group has thematically sought to expand on the wide variety of PFC technologies available in order to tackle PPI interrogation by developing reporters with increased functionalities and the potential for easy interchangeability. Prior to work outlined in this thesis, and as mentioned briefly above, we developed split T7 RNA polymerase (RNAP) biosensors specific to different promoters and capable of transcription only upon reconstitution. These could be multidimensionally deployed to measure the interactions between entire

families of proteins simultaneously, with each given transcribed RNA reporting on a different interaction.[14,15] This system is unique because the readout is not the traditional fluorescence or luminescence we tend to associate with PFCs: it is RNA itself, the quantity of which can be related to the degree of interaction of a given PPI. Because of this, we realized that we had an opportunity to not only assess PPIs, but to use the resultant RNA readout itself to affect cellular change. As such, my first project in the lab, and what I will cover in **Chapter 2**, focused on the application of our split T7 RNAP system toward Cas9 control, in which we ultimately show that Cas9 can be controlled exogenously through small molecule induction and endogenous PPI event measurements. We hope this system can be easily adapted for a range of physiological events, and successful modulation of Cas9 at the gRNA level could in-principle open many new doors for Cas9 activation responses using multiple different input signals.

For my second project, and what I will discuss in **Chapter 3**, we sought to expand on traditional imaging-based PFCs. In thinking about the fantastic imaging utility of split NanoLuc, as discussed above, we considered what enzyme might be ideal to replicate this process while also affording expanded function. We ultimately developed a split esterase-based biosensor, which we deployed in concert with selectively-masked fluorescent and chemiluminescent imaging reagents as well as a bioactive molecule for both imaging- and biological change-based readouts. This tool as an all-in-one type of PFC technology, which we envision will find widespread use, particularly in future directions as an encodable pro-drug unmasker.

Moving away from technologies that require the experimentalist to know exactly which PPI events he or she wishes to measure, the world of proximity labeling has proven an exciting one to move into when trying to understand and illuminate unknown interactions. Building off of the abundance of protein proximity labeling technologies currently available, work has recently begun in the area of RNA proximity labeling. APEX2 labeling has been developed into APEX-seq[37] by taking advantage of the guanine base's propensity to participate in ROS-based reactions. While this method has been pioneering, its broad utility has been limited by the requirement of toxic reaction components that preclude *in vivo* utility and induce cellular stress, making perturbative studies difficult to deconvolute. To this end, two new methods, Cap-seq[38] and Halo-seq,[39] circumvent the toxic H_2O_2 requirement by instead relying on the activity of a blue- and green light-activated enzyme (miniSOG) and ligand (Halo), respectively, to achieve ROS-based labeling of the G base. A common theme, this ROS-based labeling of the G base ultimately arises as a byproduct of intended protein proximity labeling targets, and so has been applied to RNA labeling out of convenience. However, this chemistry is neither optimized for RNA labeling, nor does it take advantage of the unique reactivity of RNA. As such, we considered how we might marry the versatility of our selective esterase/ester-masking system with a reagent designed to label RNA. My third project in the group, and what I will highlight in **Chapter 4**, centers on ongoing, nearly complete, efforts toward the development of an esterase-based RNA proximity labeling strategy that deploys a unique ester-masked enol ester acylating reagent inspired by RNA SHAPE chemistry. We ultimately deploy this technology to assess both organellar- and sub-organellar- transcript

localizations as well as to assess RNA binding protein (RBP) engagement to hopefully be applied more broadly to illuminate RNA-protein interactions (RPIs).

To briefly **summarize**, then, in this thesis, I will focus on the development of three technologies to interrogate and harness biomolecular interactions, as well as briefly discuss in an appendix efforts toward the application of one of the technologies for animal imaging applications:

Chapter 2 details the development of a dual small molecule-activated split T7 RNA polymerase-based biosensor system that is deployed to control Cas9 in an on/off fashion.

Chapter 3 discusses the creation of a new protein-fragment complementation technology, a split esterase, and its deployment both as an imaging-based split reporter in addition to as a mediator of cell death based on a given interaction event.

Chapter 4 presents the making of an esterase-based RNA proximity labeling method that deploys a unique ester-masked enol-thioester acylating reagent and its deployment to organelles and RNA binding proteins of interest.

Chapter 5 summarizes this thesis and provides a look forward toward the future of the technologies described herein.

Appendix A examines efforts made toward a synthetic bioluminescent *in vivo* imaging system predicated on evolved bioorthogonal esterase activity.

CHAPTER 2

MULTIDIMENSIONAL CONTROL OF CAS9 BY EVOLVED RNA POLYMERASE-BASED BIOSENSORS

This chapter is reproduced and adapted from the publication: Pu, J.; K. Kentala, K.; Dickinson, B. C. “Multidimensional control of Cas9 by evolved RNA polymerase-based biosensors.” *ACS Chem Biol.* **2018**, *13* (2), 431-437.

2.1 Abstract

Systems to control Cas9 with spatial and temporal precision offer opportunities to decrease side effects, protect sensitive tissues, and create gene therapies that are only activated at defined times and places. Here, we present the design of new Cas9 controllers based on RNA polymerase (RNAP)-based biosensors that produce gRNAs, thereby regulating target knockout. After development and validation of a new abscisic acid-inducible biosensor to control Cas9, we lowered the background of the system using continuous evolution. To showcase the versatility of the approach, we designed biosensors that measure medically relevant protein−protein interactions to drive knockout. Finally, to test whether orthogonal RNAP biosensors could integrate multiple input signals to drive multiple gRNA-based outputs with a single Cas9 protein, we designed an “on- switch/off switch” controller. The addition of one input activates the “on switch” and induces knockout, while the addition of a second input activates the “off switch” and produces a gRNA that directs the Cas9 protein to degrade the “on switch” gRNA vector, thereby deactivating it. This combined activation and deactivation system displayed very low background and inducible target knockout using different combinations

of small-molecule treatment. Our results establish engineered RNAP biosensors as deployable Cas9 control elements and open up new opportunities for driving genetic editing technologies by diverse input signals.

2.2 Introduction

The RNA-guided endonuclease Cas9 from type II clustered regularly interspaced short palindromic repeats (CRISPR)−CRISPR-associated (Cas) systems have revolutionized the practicality and therapeutic potential of genome engineering.[40-42] Cas9 is programmed through complex formation with a single guide RNA (gRNA) that directs the protein to a specific DNA locus by base pair complementarity.[43] Deployment of Cas9-based systems in the laboratory or the clinic could be substantially improved by general methods to activate the gene editing function of Cas9 only in target cells and tissues with spatial and temporal control, as well as methods to shut the system down if side effects begin to emerge due to off-target activities.[44]

The conditional regulation of Cas9 through external activation signals has been shown to lessen off-target cleavage events compared to constitutively active Cas9.[45] To-date, approaches to regulate Cas9/gRNA genome editing activity have generally focused on engineering the Cas9 proteins. For example, Cas9 has been split and each half fused to dimerization domains[46-49] and allosteric regulatory domains have been inserted into the Cas9 protein.[45,50,51] Cas9 has been fused to small-molecule-regulated protein degrons,[52] and unnatural amino acids have been inserted into Cas9 for control purposes.[53] One issue with these approaches is that they each need to be carefully optimized for a specific Cas9 variant. As new Cas9 systems continue to be discovered and optimized,[54,55] each

of these control systems in turn needs to be reinvented. More importantly, controlling Cas9 at the protein level does not permit leveraging the potential for multidimensionality through different activation signals driving different editing responses, which is trivial for Cas9 through the use of different gRNAs. Therefore, systems to control Cas9/gRNA genome editing activity at the gRNA level could offer advantages in terms of generality with any Cas9 variant and the opportunity for multidimensional responses. Thus far, gRNAs have been engineered with strand displacement-based riboswitches[56] and ribozymes that respond to ligands.[57]

In this work, we develop new small-molecule control systems for Cas9 using evolved RNA polymerase (RNAP)-based biosensors that drive target gRNAs when activated (**Figure 2.1A**). We show that the biosensors can detect interactions between medically relevant proteins, such as the B cell lymphoma 2 (Bcl-2) and truncated Bcl-2 homology 3 (BH3)-interacting domain death agonist (tBID), and drive gene knockout based on those interactions. We go on to deploy two small-molecule-activated biosensors in a logic gate, with one small-molecule input activating targeted genome engineering and the other small-molecule input deactivating the system. The “turn on,” “turn off” system maintained reasonable levels of targeted genome engineering but displayed very low background editing in the absence of “turn on” activation. Together, these results establish that regulating Cas9 at the gRNA level using engineered RNAPs is a viable strategy for engineering multidimensional, responsive genome engineering control systems. The ease of engineering the new ABA biosensor further solidifies the “plug-and-

play" nature of our evolved split RNAP scaffold and presages future utility of these approaches for sensing and responding to diverse inputs.

2.3 Results

2.3.1 Generation of an Abscisic Acid-Inducible RNAP Biosensor for Cas9 Control

Recently, we reported the generation of a versatile biosensor platform based on evolved, proximity-dependent split RNAPs.[58,59] In order to generate a new Cas9 control system, we aimed to leverage the evolved split RNAP system to create a ligand-inducible transcriptional controller for gRNA production in mammalian systems. We chose to generate an RNAP biosensor for abscisic acid (ABA), due to its lack of targets in the human proteome, using previously reported ABA-inducible dimerization systems based on the ABI and PYL proteins.[60] We fused ABI and PYL onto the N-terminal split RNAP ($RNAP_N$) and C-terminal split RNAP ($RNAP_C$) halves of our evolved split RNAP system and measured ABA-induced transcriptional activation of the biosensor in E. coli (**Figure 2.1B**, **Table 2.1**, and **Figure 2.5**). Induction with ABA resulted in a robust, dose-dependent increase in RNA synthesis, with a 468-fold dynamic range after just 3 h of induction (**Figure 2.1C**).

Given that the new ABA biosensor performed well in E. coli, we next tested whether we could use the ABA biosensor to control Cas9/gRNA genome editing in mammalian cells. We cloned a GFP-targeting gRNA controller gene circuit using the engineered ABA

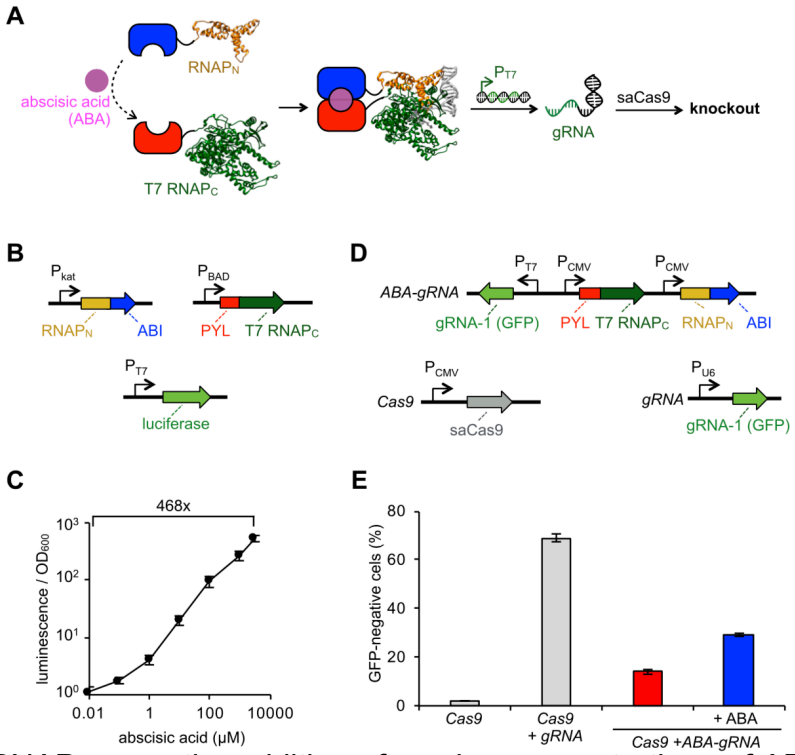


Figure 2.1 Design and application of an abscisic acid (ABA) inducible biosensor to control Cas9. (A) Schematic of Cas9 split RNAP biosensor control concept. Evolved proximity-dependent split RNAP halves ($RNAP_N$ and $RNAP_C$) are fused to sensor domains, such that when the domains interact, the RNAP assembles and produces a gRNA, which then assembles with Cas9 and knocks out a target gene. (B) Vector system to test ABA-induced split RNAP detection system in E. coli. (C) Transcriptional output of split RNAPs upon the addition of varying concentrations of ABA assayed in E. coli using the vectors shown in B. Cells were induced for 3 h with arabinose and then analyzed for luminescence. Error bars are ± SEM, n = 5. (D) Vector system to deploy ABA biosensors in mammalian cells. (E) Knockout efficiency of ABA biosensor with N-29-1 $RNAP_N$ compared to a constitutive gRNA. HEK293-GFP cells transfected with varying combinations of the vectors shown in D, grown for 6 days, and then analyzed for GFP knockout by flow cytometry. The addition of 10 μM ABA for 5 days induces knockout. Error bars are ±SEM, n = 3.

split RNAP system (**Figure 2.1D**), which we deployed along with Staphylococcus aureus Cas9 (saCas9).[61] Assaying the system in GFP-expressing HEK293 (HEK293- GFP) cells showed that the background level of genome editing in the absence of inducer was 14%, compared to 1.7% experimental background for cells transfected with an RFP vector alone as a transfection control. However, induction with 10 μM ABA for 5 days resulted in a knockout efficiency of 29%, compared with 69% knockout from a control vector with constitutive gRNA production under these experimental conditions (**Figure 2.1E**). These results demonstrate the viability of using RNAP-based biosensors to control Cas9 at the

gRNA level, but the high background knockout activity in the absence of inducer precludes utility of the approach. Although the evolved split RNAP biosensor platform appeared to have a very low background in E. coli and our previous mammalian experiments, the irreversible nature of genome engineering accentuates any level of leaky activation, revealing a deficit and an opportunity for improvement of the split RNAP biosensors. Therefore, we next turned our attention to lowering the background activity of the split RNAP-based biosensor platform.

2.3.2 Evolution and Deployment of RNAP Biosensors with Lower Background

In our original work to create the proximity-dependent split RNAPs, we used Phage-Assisted Continuous Evolution (PACE)[62] to evolve proximity dependent assembly of the split RNAP using leucine zipper peptides as a model dimerization system.[59] In brief, the phage carried an evolving $RNAP_N$ fused to one leucine zipper, and E. coli cells housed two $RNAP_C$'s fused to either a binding peptide or a nonbinding peptide. Each $RNAP_C$ drove either positive or negative selection pressure. At the conclusion of the reported continuous evolutions, we found that we could not enhance the negative selection any further without extinction of the phage. Upon further analysis of the evolved gene products subsequent to the published studies, we discovered that the $RNAP_N$ leucine zipper peptide had evolved a low but measurable affinity for the off-target control peptide. Therefore, we reasoned that we could improve the negative selection by simply using an off-target $RNAP_C$ without any fusion protein. To test this concept, we cloned a new negative accessory plasmid (Apneg) with stronger ribosome binding sites (RBSs) controlling the $RNAP_C$ and dominant negative gIII (gIIIneg), without any fusion protein

(**Figure 2.5**, **Table 2.1**). We initiated PACE with libraries of phages from the end of our previous evolutions[59] and found that the phage could now replicate with the stronger Apnegs. Full evolutionary details are outlined in **Figure 2.6**. Briefly, after 5 days of PACE with sequentially stronger selection pressure using combinations of posAP and negAPs (**Figure 2.2A**), the populations converged on several new $RNAP_N$ mutations (**Figure 2.7**). Several of the new mutations exist at the split protein interface, but some are in unstructured regions of the protein (**Figure 2.8**), illustrating why unbiased evolution is a powerful design strategy for this type of protein engineering challenge.

Transcriptional reporter assays revealed that most of the newly evolved variants had a lower background than the starting evolved variant (N-29-1), with some variants having up to a 2100-fold dynamic range based on the fused protein−protein interaction (**Figure 2.2B**), indicating the evolution dramatically improved the split RNAP platform. We selected a subset of variants (**Figure 2.2C**) to assay in a mammalian cell GFP transcriptional reporter assay (**Figure 2.2D**), which generally reproduced the trends observed in E. coli, further demonstrating that the evolved variants had a lower background and robust ABA responsiveness in mammalian cells (**Figure 2.2E**, **Figure 2.9**, and **Figure 2.10**). We also assayed the optimal inducer concentration and found that induction with 1−100 μM ABA resulted in similar levels of activation, similar to the 10 μM concentration used thus far (**Figure 2.11**, **Table 2.2**). Finally, we cloned the new variants into the gRNA controller gene circuit (**Figure 2.1D**) and found that they overall maintained reasonable genome engineering activation upon induction with ABA (10−34%), but in the absence of ABA the background level of knockout was decreased (2.3% - 8.8%), in some

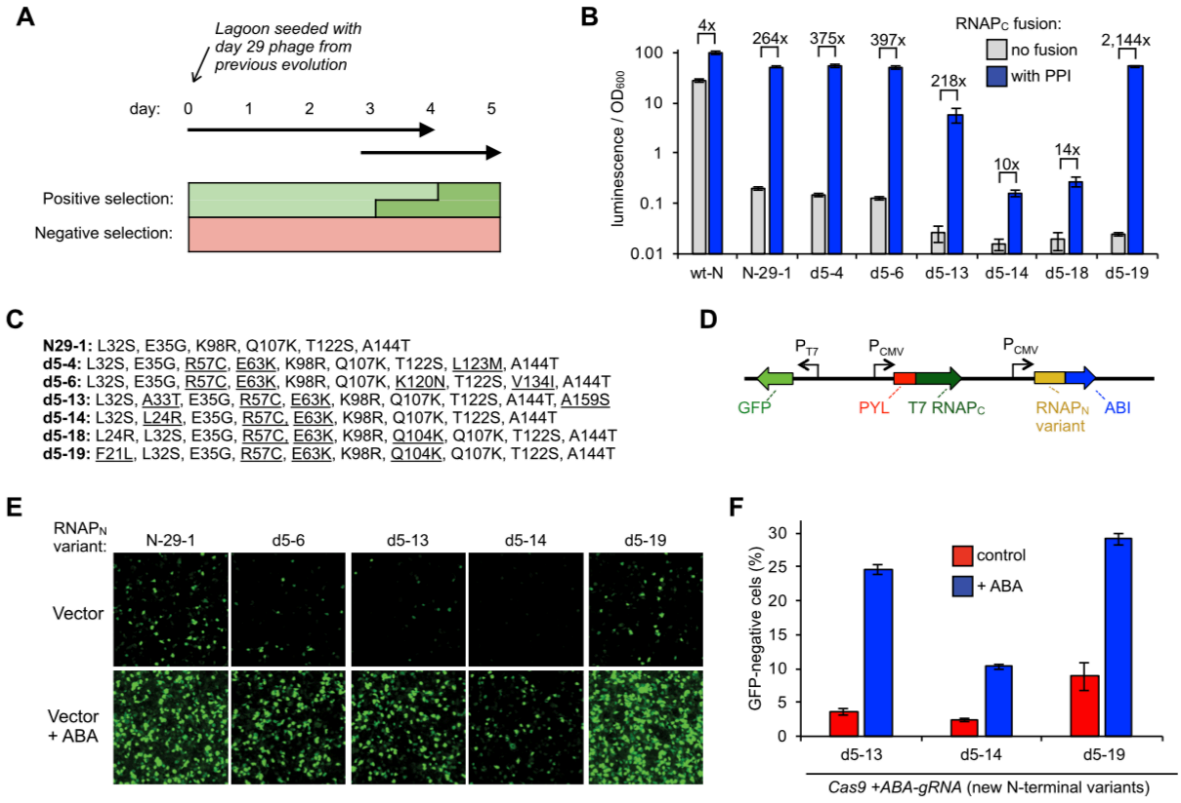


Figure 2.2 Evolution of improved split RNAP biosensors and deployment as Cas9 controllers. (A) Overview of PACE design. (B) Transcriptional reporter assays of a series of $RNAP_N$ variants that emerged from the evolution. *E. coli* cells were transformed with (1) a vector expressing a zipper peptide-fused $RNAP_N$ variant, (2) a vector expressing $RNAP_C$ either fused to a zipper peptide or without a fusion, and (3) a T7 promoter-driven luciferase reporter vector. The cells were induced for 3 h with arabinose. Error bars are ± SEM, *n=5*. (C) Mutations of the $RNAP_N$ variants. (D) Vectors designed to test new $RNAP_N$ variants in (E). (E) HEK293T cells were transfected with thee vectors shown in (D) containing different $RNAP_N$ variants, stimulated with either nothing or 10 μM ABA 7 h post-transfection, grown for 23 h, and then analyzed by fluorescence microscopy (quantification and full imaging shown in Figures 2.9 and 2.10). (F) Knockout efficiency of new variants. HEK293T-GFP cells were transfected with the vector system shown in Figure 2.1D. Then 10 μM ABA or DMSO control was added 1 day after transfection and grown for 5 days more and then analyzed for GFP knockout by flow cytometry. Error bars are ± SEM. *n=3*.

cases near the background of the assay (1.7%; **Figure 2.2F**). Collectively, these results demonstrate that the newly evolved RNAP biosensors perform better as ABA-inducible Cas9 controllers.

2.3.3 *Using BCL-2 Protein-Protein Interactions (PPIs) to Drive Cas9*

One advantage of the protein-based split-RNAP biosensor scaffold is that one could, in principle, detect medically relevant PPIs and drive responses based on those interactions. The eventual concept would then be to use measurements about endogenous biochemical events to control Cas9, thereby selecting cell types based on molecular markers. To explore this possibility, we designed RNAP-based biosensors that detect interactions between the apoptotic regulatory protein, Bcl-2, and its ligand, tBID,[63] using methods we recently developed.[64] We designed vectors that express the $RNAP_C$ fused to Bcl-2 and $RNAP_N$ fused to either the binding domain of BID (tBID) or a control peptide with key binding components omitted (dBID; **Figure 2.3A**). Delivery of only the Bcl-2-$RNAP_C$ vector along with Cas9 resulted in very low knockout of 1.9%, while adding in the control, $RNAP_N$-dBID vector resulted in a background knockout without the PPI of 7.2% (**Figure 2.3B**). However, when the interacting $RNAP_N$-tBID vector is deployed, the knockout efficiency rises to 21.6%. To

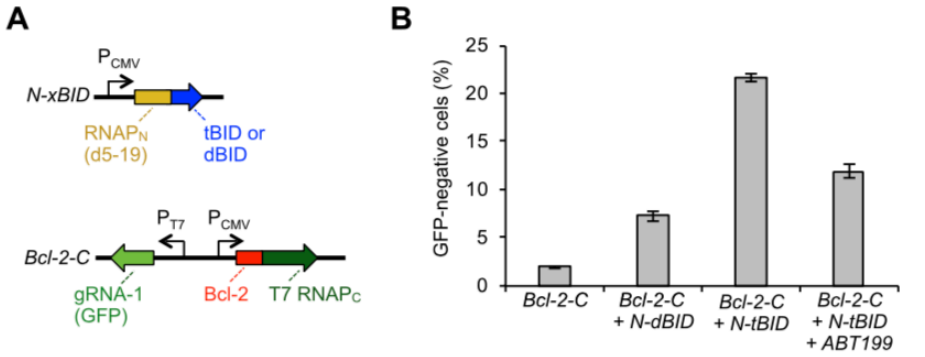


Figure 2.3 Biosensors detect Bcl-2 PPIs and drive Cas9 knockout. (A) Vectors designed to express Bcl-2 fused to T7 $RNAP_C$ and the evolved $RNAP_N$ fused to either an interacting BID (tBID) or a noninteracting BID (dBID), along with T7 promoter-drive gRNA to GFP. (B) HEK293-GFP cells were transfected with different combinations of the vector system shown in A, grown for 6 days, and then analyzed for GFP knockout by flow cytometry. The addition of 500 nM ABT199 6 h after transfection and maintained for the following 2 days lowered the PPI-dependent knockout. Error bars ±SEM, n = 3.

confirm that this effect is due to the Bcl-2/tBID PPI, we treated the cells with a Bcl-2 PPI inhibitor, ABT199,[65] which lowered the level of knockout to 11.9%. Additionally, we observed some dose-dependent alterations in gRNA output levels due to ABT199 inhibition as assayed by RT-qPCR (**Table 2.3**). Collectively, these results highlight the versatility of the RNAP biosensor platform and illustrate the possibility of using biosensors of disease-relevant, regulated PPIs as endogenous molecular markers to control Cas9.

2.3.4 Deploying Dual Biosensors for Small-Molecule Logic Gate Cas9 Control

Although the newly evolved split RNAP variants displayed lower levels of knockout in the absence of induction, there was still measurable background activity for the most active variants. To remedy this, we envisioned that a second, “off switch” gRNA controller gene circuit, which produces a gRNA that targets and degrades the “on switch” gRNA controller upon activation by a different small-molecule inducer, could lower the background and “protect” cells by turning off the system prior to induction. We had previously generated a rapamycin-inducible split RNAP biosensor and an orthogonal RNAPC that drives transcription from a unique DNA promoter sequence.[59] Using these tools, we generated a rapamycin-inducible gRNA controller gene circuit that instead produces a gRNA that targets saCas9 to the ABA-inducible gRNA controller gene circuit (**Figure 2.4A**). In this way, the “off switch” can be activated to shut down the entire system prior to activation of the “on-switch”, thereby lowering the background knockout in target cells. Both controllers utilize the same saCas9 protein, so all of the control emerges from differential gRNA generation.

To test the dual controller system, we delivered the “on switch” and “off switch” vectors, along with an saCas9 expression vector, into mammalian cells, and treated the cells with ABA and/or rapamycin at various time points (**Figure 2.4B**). Under the experimental conditions, a constitutive gRNA vector induced knockout at 34% relative to 2.3% background without a gRNA.

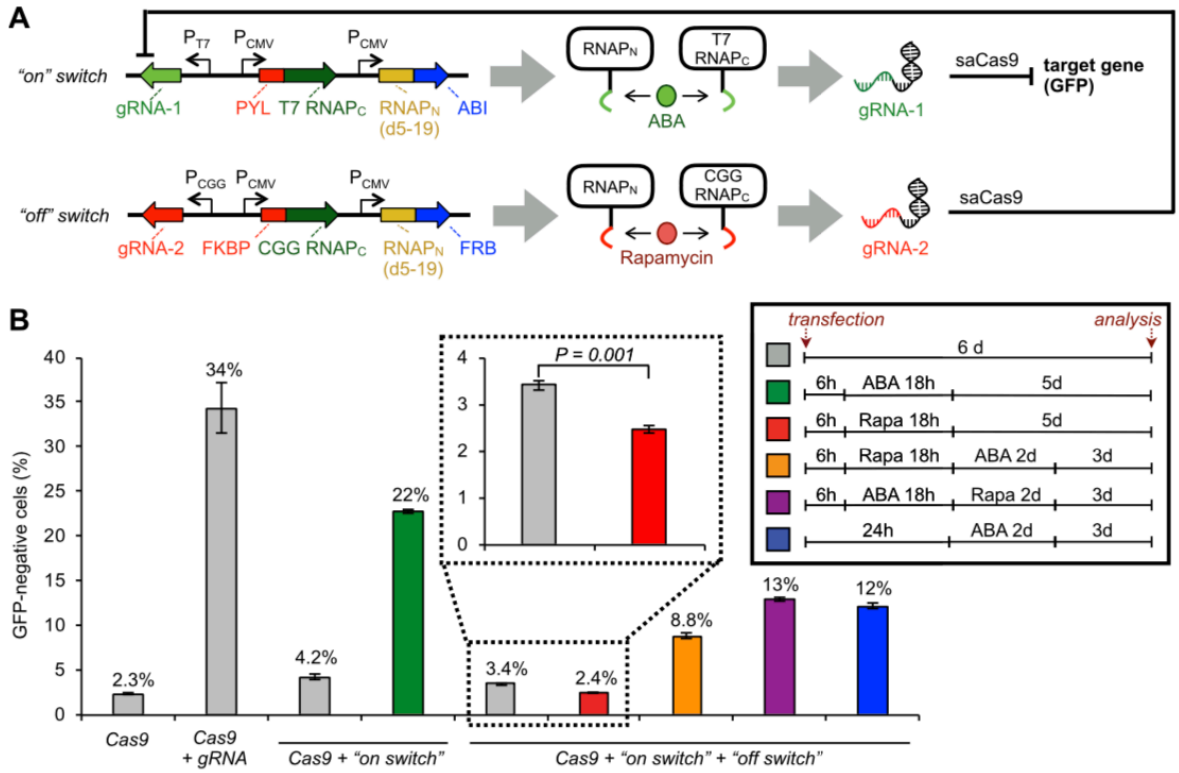


Figure 2.4 Dual control of Cas9 using two small-molecule biosensors. (A) Schematics of the “on switch” vector that produces an ABA-inducible RNAP biosensor system and triggers gRNA production for a target gene (GFP) and the “off switch” vector that produces a rapamycin-inducible RNAP biosensor system and triggers gRNA production that targets the “on switch” vector. (B) HEK293-GFP cells were transfected with a Cas9 expression vector and the vectors shown in A, grown for 6 days, then analyzed for GFP knockout by flow cytometry. Treatment with 10 μM ABA and/or 10 nM rapamycin (“Rapa”) for the given times shown in the inset induces either the “on switch” or “off switch” vectors and corresponding target knockout or system deactivation. Cells were washed between each treatment change to remove the molecule from the previous condition. Error bars are ±SEM, n = 3.

The “on-switch” vector alone, without activation, raised the background knockout to 4.2% but was readily activated by 18 h of ABA treatment to 22% knockout. However, with both the “on switch” and “off switch” vectors, rapamycin treatment lowered the background to 2.4%, which is equivalent to the background of the experimental cells treated with RFP transfection control vectors, thereby demonstrating the ultra- low background in the dually controlled system. ABA treatment in the absence of rapamycin still induced reasonable knockout (12%), though to a lower level than the “on-switch” alone, possibly due to leaky activation of the “off switch.” Pretreatment with rapamycin and then ABA still induced knockout (8.8%), but to a lower extent, showcasing the ability of the “off switch” to protect cells. The incomplete deactivation of the rapamycin-induced “off switch” is possibly due to heterogeneity of DNA plasmid uptake in the transient transfections, leading to incomplete degradation. In any case, as a proof-of-concept, these observations demonstrate the ability of multicontrolled Cas9 systems to protect cells from subsequent Cas9 knockout, while also lowering the background in a small-molecule- dependent manner.

2.4 Discussion

Our current system displays a very low background but also suffers from a diminished level of activation, which has been observed in other reported Cas9 control systems. Because our system uses the full-length, wild-type Cas9 protein, all of our performance loss is due to gRNA production inefficiencies from the RNAP biosensors. The heterogeneity in response could be due to the transient transfections used in this study. In future work, using viral delivery systems or direct protein delivery could offer

better control over the Cas9 response. Additionally, the localization of the DNA substrate for the split RNAP biosensors may be heterogeneous and suboptimal. Interestingly, we found that the split RNAPs actually produce more gRNA than a constitutive PU6-driven gRNA, which is commonly used for constitutive Cas9 knockout, indicating that further optimization of our system is certainly possible (**Figure 2.12**). For example, using small, chemically modified, gRNA-generating DNA substrates could increase potency and result in a more homogeneous response.

Recently, a “self-limiting Cas9 circuit” was developed,[66] in which a self-targeting gRNA regulates the levels of Cas9 protein, thereby decreasing off-target activities. Our “on switch/ off switch” approach described here is conceptually similar but has the added benefit of temporal regulation afforded by small- molecule control. Finally, we note that we chose the rapamycin biosensor for validation in cell culture, but since rapamycin has targets in human cells, it would likely not be the best biosensor system for therapeutic applications.

In summary, this work demonstrates that engineered RNAP- based biosensors can control Cas9 in a multidimensional manner using exogenously added small molecules or endogenous PPI measurements. Moreover, these studies validate that multidimensional control of Cas9 at the gRNA level opens up new opportunities for engineering Cas9 activation responses using multiple input signals. An advantage of the RNAP-based scaffold is that biosensors for a range of physiological activities can be easily created by leveraging the wealth of precedent on how to engineer fluorescent protein-based biosensors based on split GFP and FRET probes.[67,68] Finally, the newly

evolved $RNAP_N$ variants generated here will likely find utility in other synthetic biology applications due to their exceptionally low background activity and very large dynamic range.

2.5 Materials and Methods

Cloning. All plasmids were constructed by Gibson assembly from PCR products generated using Q5 DNA Polymerase (NEB) or Phusion Polymerase. pSPgRNA was a gift from Professor Charles Gersbach (Addgene plasmid #47108),[69] which was used to construct the constitutive gRNA vector. The gene for saCas9 was a gift from Professor Feng Zhang and was cloned into a custom CMV-driven mammalian expression vector. The genes for ABI and PYL were gifts from Professor Fu-Sen Liang and were cloned into our custom vectors. Phages used in this research were constructed and evolved in previous work.[59] Briefly, the $RNAP_N$-ZA fusion had been cloned into an SP phage backbone and underwent 29 days of evolution before the evolution in this current work. All plasmids were sequenced at the University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping Facility. All used vectors are described in **Table 2.1**, and maps for each plasmid are shown in **Figure 2.5**. The constructed gRNA sequences for GFP knockout (gRNA-1) and "off switch" (gRNA-2) are listed in **Table 2.4**. Full vector sequences and annotated vector maps are available upon request.

Phage-Assisted Continuous Evolution (PACE). PACE was carried out using previously described methods[20] with slight modifications. The key difference was that the

negative AP contained no fusion protein on the $RNAP_C$. During the evolution, phage samples were collected every 24 h, then boiled for 10 min to lyse the phage and release the genomes. PCR was then used to amplify the DNA library containing the $RNAP_N$ variants, which was then subcloned into vector p3-7. Single colonies were picked from the transformation and subjected to analysis by Sanger sequencing. The results of the sequence analysis during the course of the evolution are shown in **Figure 2.7**.

Luciferase-based in Vivo Transcription Assays of Split RNAPs. Experiments were conducted as previously described.[59] Briefly, S1030 cells were transformed with three plasmids: (i) a constitutive $RNAP_N$-expression plasmid, (ii) an arabinose induced $RNAP_C$-expression plasmid, and (iii) a T7 promoter driven luciferase expression plasmid. Single colonies were then grown in a 96-deep-well plate overnight at 37 °C, and 50 µL of the culture was transferred to a new 96-deep-well plate containing 0.5 mL of LB with antibiotics and 10 mM arabinose. After growth with shaking at 37 °C for 3 h, 150 µL of each culture was transferred to a 96-well black wall, clear bottom plate (Nunc), and luminescence and OD600 were measured on a Synergy Neo2 Hybrid Multi-Mode Reader (BioTek). The data were analyzed by dividing the luminescence values by the background- corrected OD600 value. All values were then normalized to the wild- type split RNAP fused to ZA and ZB, which was assigned an arbitrary value of 100. For the ABA induced system, the experiment was performed identically, except upon outgrowth, different concentrations of ABA were added together with the arabinose.

Mammalian Fluorescence Imaging and Quantification. HEK293T cells (ATCC) were maintained in DMEM (glucose, GlutaMAX, phenol red, sodium pyruvate, obtained from Gibco) supplemented with 10% fetal bovine serum (FBS, Gibco/Life Technologies, Qualified US origin) and 1% penicillin/streptomycin (P/S, Gibco/Life Technologies). For the data shown in **Figure 2.2E** and **Figure 2.9**, HEK293T cells were plated on an eight-well coverglass slide (Cellvis) and transfected the next day with 500 ng of split RNAP GFP expression vector (**Table 2.1** and **Figure 2.5**) using 1.5 µL of Lipofectamine 3000 (ThermoFisher Scientific) following the standard protocol. Then, 10 µM ABA was added 7 h after transfection. The cells were imaged 23 h later on an Olympus BX53 microscope using a GFP filter set and a 10× objective. For the data shown in **Figure 2.11**, HEK293T cells were plated on an eight-well coverglass slide and transfected the next day with 500 ng of pJin 278 (Table S1 and Figure S1) using 1.5 µL of Lipofectamine 3000 (ThermoFisher Scientific) following the standard protocol. Then, ABA was added during transfection. The cells were imaged 46 h later on a Leica Dmi8 microscope using a GFP filter set and a 20× objective. For quantification, identical settings were used for a given condition to adjust brightness and contrast in ImageJ (Wayne Rasband, NIH). The fluorescent spots in each GFP image were counted as fluorescent cell numbers in ImageJ for quantification analysis, and macro batch scripts were used for each analysis in the "Batch Process" of ImageJ. The data for **Figure 2.10** were obtained using the script "setThreshold(1000, 5000); setOption("BlackBackground", false); run("Convert to Mask"); run("Watershed"); run("Analyze Particles...", "size = 100-Infinity pixel include summarize

in_situ” ”). For **Figure 2.11**, the threshold was set to (6000, 65 535) instead as images were taken as 16-bit rather than 8-bit. The “Count” results were then used for quantification analysis.

GFP Knockout. HEK293-GFP cells (GenTarget), which genomically encode GFP, were used to test knockout efficiencies and were maintained in the same DMEM media as described above for HEK293T cells. Seven ×104 cells were plated on a 48-well plate (NEST) for each experiment. One day after plating, the cells were transfected with 1.8 μL of Lipofectamine 3000 (Life Technologies) and corresponding vector DNAs. In the assays shown in **Figure 2.1E** and **Figure 2.2F**, 20 ng of RFP expression plasmid, 300 ng of saCas9 expression plasmid,22 and 300 ng of GFP-knockout gRNA plasmid were used. In the assays shown in **Figure 2.3B**, 20 ng of RFP expression plasmid, 200 ng of saCas9 expression plasmid,22 300 ng of $RNAP_N$-xBID, and 300 ng of Bcl-2-$RNAP_C$ vector were used. In the assays shown in **Figure 2.4B**, 20 ng of RFP expression plasmid, 250 ng of saCas9 expression plasmid, 150 ng of GFP-knockout gRNA(gRNA-1) plasmid, and 250 ng of GFP knockout switch-off gRNA(gRNA-2) plasmid were used. For single small-molecule treatment, 10 μM ABA was added 24 h after transfection (**Figure 2.1E** and **Figure 2.2F**). For the double small-molecule treatment experiments (**Figure 2.4B**), ABA or rapamycin were added sequentially, and the cells were washed between each treatment. For all assays, 6 days after transfection, the cells were trypsinized and resuspended in DMEM supplemented with 10% FBS and analyzed on an LSR Fortessa 4−15 flow cytometer (BD digital instrument, 488 nm laser with 530/30 nm filter for GFP,

561 nm laser with 610/20 nm filter for RFP). GFP negative or positive cells were analyzed on RFP- gated cells as a transfection control.

Quantitative Reverse Transcription PCR (RT-qPCR) Analysis of gRNA Production. HEK293T cells were transfected with 1 μg of DNA of corresponding vectors for each sample in 12-well plates (Denville). Lipofectamine 3000 (ThermoFisher Scientific) was used for the transfection following standard protocol. For the ABA-induced gRNA production test in **Table 2.2**, 1 μg of pJin 264 was used for each sample, and 0, 1, 10, or 100 μM ABA was added when transfecting cells. For the concentration-dependent test of ABT199 in **Table 2.3**, 500 ng of each plasmid (pJin 310 and p12−34) was used for each sample, and 0, 50, 250, 500 or 1000 nM ABT199 was added when transfecting cells. The cells were harvested 40 h after transfection for these assays. For analysis of the amount of gRNA produced by split RNAP in **Figure 2.12**, 1 μg of plasmid was used, and DMSO or corresponding small molecules were added 15 h after transfection. The cells were harvested 26 h later. After cell lysis, RNA was purified using an Rneasy Kit (Qiagen), then reverse-transcribed using a PrimeScript TM RT reagent Kit (TaKaRa). The transcribed cDNA libraries were analyzed by qPCR on a LightCycler 96 Instrument (Roche) using FastStart Essential DNA Green Master (Roche). The primer sequences for qPCR are listed in **Table 2.4**.

2.6 Supplemental Information

Table 2.1 List of plasmids used in this work.

Vector name	Antibiotic resistance	Origin	Purpose	Map
p2-22	carb	sc101	P_{T7} luciferase reporter plasmid	a
p3-7	spec	P15A	T7 $RNAP_N$ (wt)-linker-ZA expression plasmid	b
p2-55	chlr	CloDF13	T7 $RNAP_C$ expression plasmid	c
p2-39	chlr	CloDF13	ZB-linker-T7 $RNAP_C$ expression plasmid	d
p5-74	spec	P15A	T7 $RNAP_N$ (N-29-1)-linker-ZA (L13I, L20I) expression plasmid	b
p8-15	spec	P15A	T7 $RNAP_N$ (N-29-1)-linker-ABI expression plasmid	e
p8-17	chlr	CloDF13	PYL-linker-T7 $RNAP_C$ expression plasmid	f
pJin 274	kan	pBR322	P_{T7}-GFP mRNA expression, PYL-linker-T7 $RNAP_C$, T7 $RNAP_N$ (d5-14)-linker-ABI	g
pJin 278	kan	pBR322	P_{T7}-GFP mRNA expression, PYL-linker-T7 $RNAP_C$, T7 $RNAP_N$ (d5-19)-linker-ABI	g
pJin 286	kan	pBR322	P_{T7}-GFP mRNA expression, PYL-linker-T7 $RNAP_C$, T7 $RNAP_N$ (N-29-1)-linker-ABI	g
pJin 287	kan	pBR322	P_{T7}-GFP mRNA expression, PYL-linker-T7 $RNAP_C$, T7 $RNAP_N$ (d5-6)-linker-ABI	g
pJin 288	kan	pBR322	P_{T7}-GFP mRNA expression, PYL-linker-T7 RNAPC, T7 $RNAP_N$ (d5-13)-linker-ABI	g
pJin 239	kan	pBR322	pT7-gRNA GFP expression, PYL-linker-T7 $RNAP_C$, T7 $RNAP_N$ (N-29-1)-linker-ABI	h
pJin 264	kan	pBR322	P_{T7}-gRNA GFP expression, PYL-linker-T7 $RNAP_C$, T7 $RNAP_N$ (d5-19)-linker-ABI	h
pJin 272	kan	pBR322	P_{T7}-gRNA GFP expression, PYL-linker-T7 $RNAP_C$, T7 $RNAP_N$ (d5-13)-linker-ABI	h
pJin 263	kan	pBR322	P_{T7}-gRNA GFP expression, PYL-linker-T7 $RNAP_C$, T7 $RNAP_N$ (d5-14)-linker-ABI	h
pJin 290	kan	pBR322	P_{CGG}-gRNA GFP-off expression, FKBP-linker-T7 $RNAP_C$, T7 $RNAP_N$ (d5-19)-linker-FRB	i
pJin 310	kan	pBR322	P_{T7}-gRNA GFP expression, Bcl-2-linker-T7 $RNAP_C$	j
p12-34	kan	pBR322	T7 $RNAP_N$ (d5-19)-linker-tBID mammalian expression plasmid	k
p12-33	kan	pBR322	T7 $RNAP_N$ (d5-19)-linker-dBID mammalian expression plasmid	k
pJin 178	carb	sc101	Positive AP (RBS1 – SD8, RBS2 – sd6)	l
pJin 179	carb	sc101	Positive AP (RBS1 – SD8, RBS2 – sd5)	l
pJin 179	kan	pBR322	Negative AP (RBS3 – SD4, RBS4 – sd6)	m
p9-18	carb	CloE1	*Staphylococcus aureus* Cas9 expression vector	n
p5-54	carb	CloE1	P_{U6} driven gRNA GFP expression vector	o
P3-62	kan	pUC	P_{CMV}-driven RFP expression vector	p

Table 2.2 ABA-induced gRNA production from pJin 264. HEK293T cells were transfected with pJin 278 (**Figure 2.1D**, variant d5-19). During transfection, the cells were treated with 0, 1, 10, or 100 µM ABA. After 40 h of growth, the cells were lysed, total RNA isolated, and the levels of gRNA production analyzed by RT-qPCR. Ct values shown. GAPDH RNA levels analyzed as a control for RNA isolation.

[ABA]	0	1 µM	10 µM	100 µM
GAPDH	20.29	20.22	20.31	20.17
gRNA-1	21.12	16.6	16.72	16.61

Table 2.3 Concentration-dependence of ABT199. HEK293T cells transfected with pJin 310 and p12-34 (**Figure 2.3A**). During transfection, the cells were treated with 0, 50, 250, 500 or 1,000 nM ABT199. After 40 h of growth, the cells were lysed, total RNA isolated, and the levels of gRNA production was analyzed by RT-qPCR. Ct values shown. GAPDH RNA levels analyzed as a control for RNA isolation.

[ABT-199]	0	50 nM	250 nM	500 nM	1,000 nM
GAPDH	19.95	19.67	20.06	19.88	19.94
gRNA-1	20.12	22.89	22.7	21.99	23.11

Table 2.4 saCas9 gRNA and RT-qPCR primer sequences.

Name	DNA or oligo sequences
gRNA-1 (GFP)	GCCCTCGAACTTCACCTCGGCGTTTTAGTACTCTGTAATGAAAATTA CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAAC TTGTTGGCGAGATTTTTTT
gRNA-2 (off switch)	AGTTCGAGGGCTCTCCCTATAGTTTTAGTACTCTGTAATGAAAATTA CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAAC TTGTTGGCGAGATTTTTTT
gRNA-1-for	CCTCGAACTTCACCTCGGCGT
gRNA-2-for	AGTTCGAGGGCTCTCCCTATAGT
gRNA-rev	CTCGCCAACAAGTTGACGAGATAAACAC
GAPDH-for	TGCACCACCAACTGCTTAGC
GAPDH-rev	GGCATGGACTGTGGTCATGAG

Figure 2.5 Vector maps for all constructs used in this work. Correspond to Table 2.1.

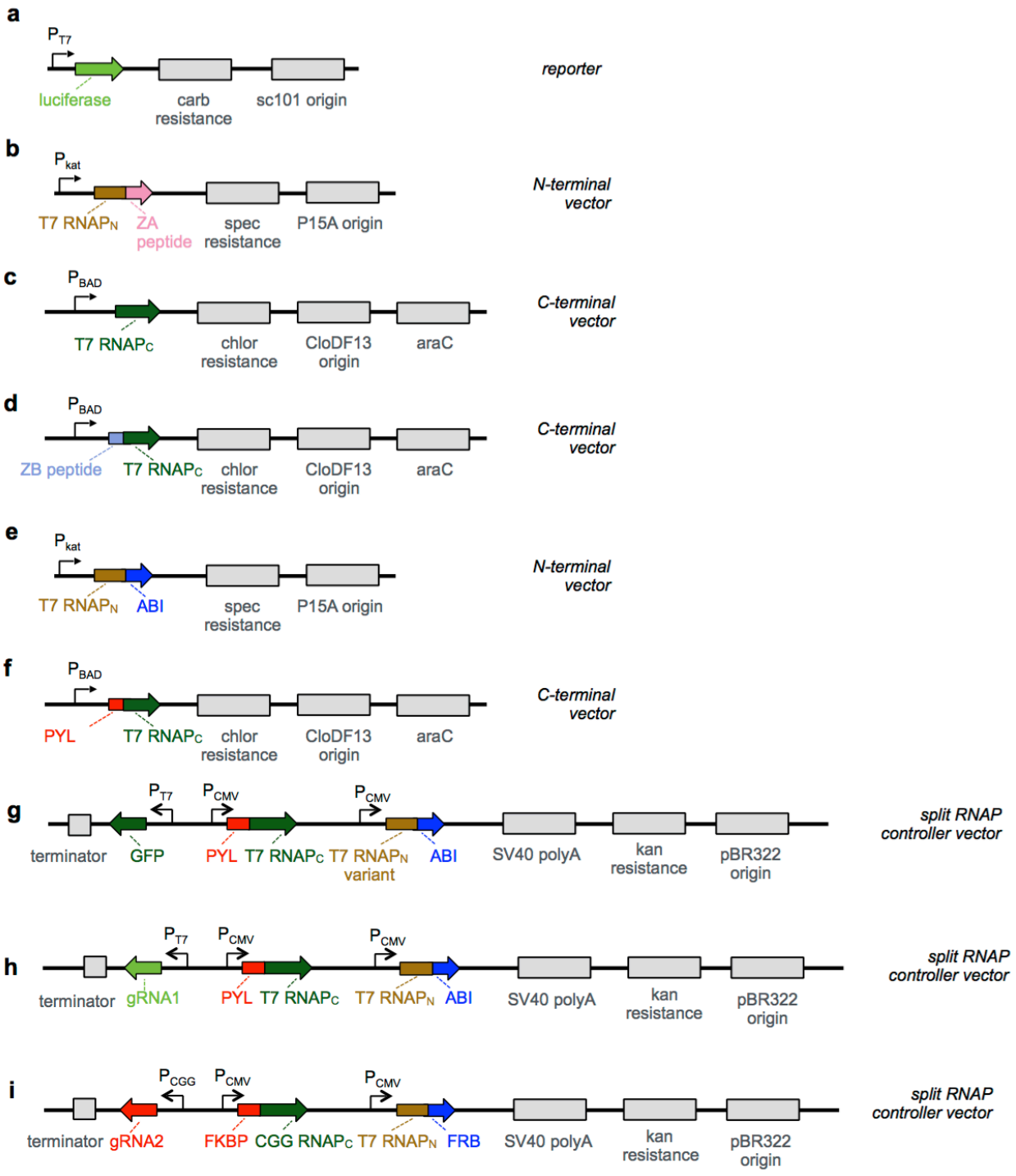


Figure 2.5 – continued from previous page

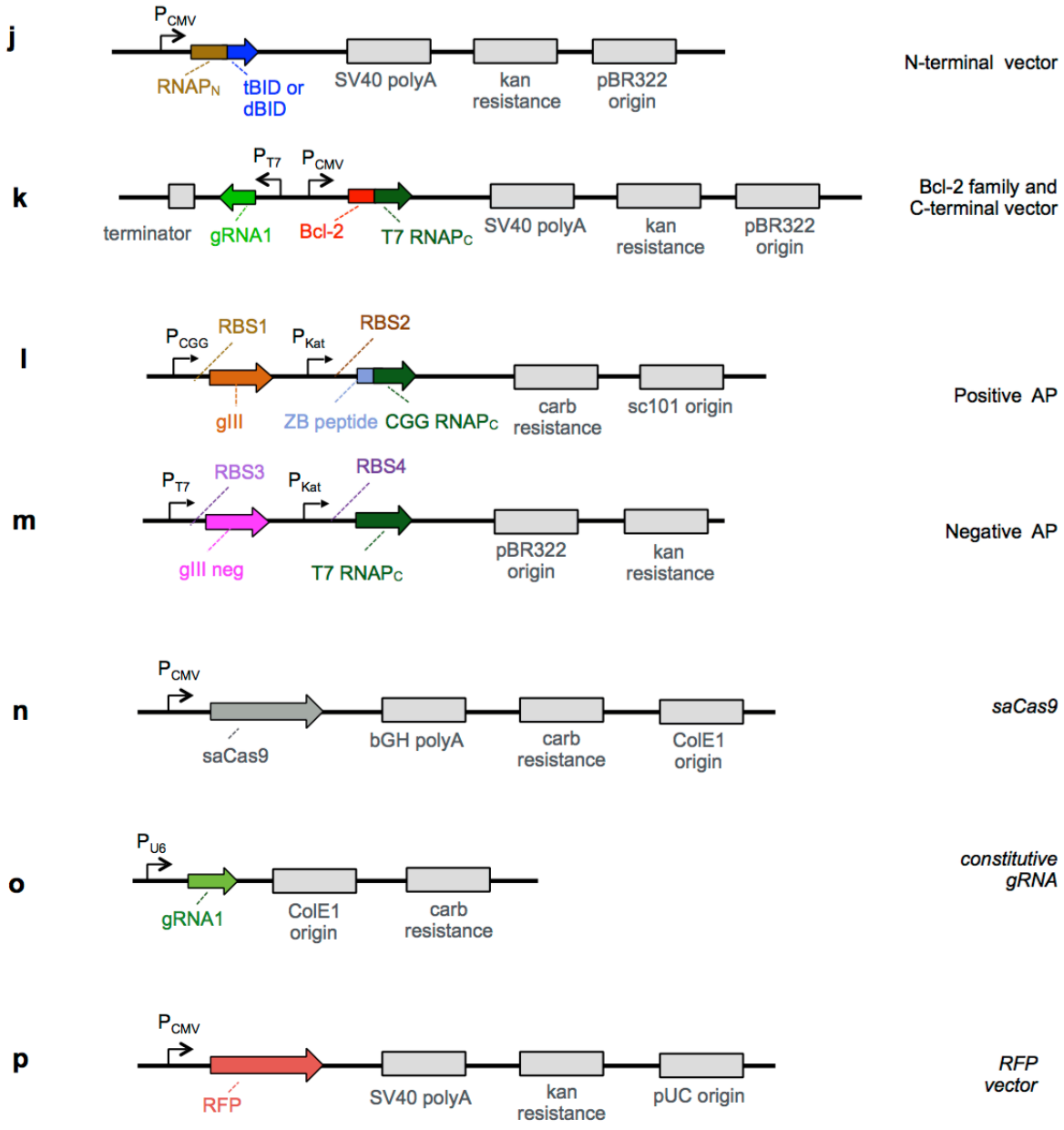


Figure 2.6 Evolutionary protocol for PACE experiment. Vector names and details are provided for each day of PACE. Vector maps for the posAP and negAP vectors are shown in **Figure 2.5**, l and m. Two sets of posAP/negAP vectors listed on the same day indicates that a mixed selection pressure was utilized, in which two types of host cells, each containing one set of the posAP/negAP plasmids were added to a lagoon simultaneously. The relative RBS strengths were obtained from previous studies.[64]

evolutionary date	positive AP					negative AP				
	Vector name	Map	origin	RBS1	RBS2	Vector name	Map	origin	RBS3	RBS4
day1	pJin178	I	sc101	SD8	sd6	pJin153	m	pBR322	SD4	sd6
day2	pJin178	I	sc101	SD8	sd6	pJin153	m	pBR322	SD4	sd6
day3	pJin178	I	sc101	SD8	sd6	pJin153	m	pBR322	SD4	sd6
day4	pJin178	I	sc101	SD8	sd6	pJin153	m	pBR322	SD4	sd6
	pJin179	I	sc101	SD8	sd5	pJin153	m	pBR322	SD4	sd6
day5	pJin179	I	sc101	SD8	sd5	pJin153	m	pBR322	SD4	sd6

Figure 2.7 Mutational analysis of evolving split RNAP. Single phage sequenced during the course of the PACE experiment and coding mutations are shown for a set of variants assayed at each time point. The final variant selected for further assay (d5-19) is highlighted yellow.

day3																			
d3-1			L32S		E35G		R57C		E63K				K98R	Q107K		T122S			A144T
d3-2			L32S		E35G				E63K				K98R	Q107K		T122S			A144T
d3-4			L32S		E35G		R57Y		E63K				K98R	Q107K		T122S			A144T
d3-5			L32S		E35G		R57C		E63K				K98R	Q107K		T122S			A144T
d3-7			L32S		E35G		R57C		E63K				K98R	Q107K		T122S			A144T
d3-8			L32S		E35G				E63K				K98R	Q107K		T122S			A144T
d3-9			L32S		E35G				E63K				K98R	Q107K		T122S			A144T
d3-10			L32S		E35G		R57C		E63K				K98R	Q107K		T122S	T127A		A144T
d3-11			L32S		E35G				E63K				K98R	Q107K		T122S		A136S	A144T
d3-12			L32S		E35G				E63K				K98R	Q107K		T122S		A136S	A144T
d3-14			L32S		E35G				E63K				K98R	Q107K		T122S			A144T
d3-15			L32S		E35G		R57C		E63K				K98R	Q107K		T122S	T127A		A144T
d3-16			L32S		E35G				E63K				K98R	Q107K		T122S			A144T
day5																			
d5-1			L32S		E35G		R57C		E63K	N67T			K98R	Q107K		T122S			A144T
d5-2			L32S		E35G		R57C		E63K	N67T	T76S	R96S	K98R	Q107K		T122S			A144T
d5-3			L32S		E35G		R57C	A61	E63K	N67T	L77I		K98R	Q107K		T122S			A144T
d5-4			L32S		E35G		R57C		E63K				K98R	Q107K		T122S	L123M		A144T
d5-5		D26Y	L32S		E35G	M46I	R57C		E63K				K98R	Q107K		T122S			A144T
d5-6			L32S		E35G		R57C		E63K				K98R	Q107K	K120N	T122S		V134I	A144T
d5-8			L32S		E35G		R57C		E63K	N67			K98R	Q107K		T122S			A144T
d5-9			L32S		E35G		R57C		E63K				K98R	Q107K		T122S			A144T
d5-10			L32S		E35G		R57C		E63K				K98R	Q107K		T122S			A144T
d5-12			L32S		E35G		R57C		E63K				K98R	Q107K		T122S			A144T
d5-13			L32S	A33T	E35G		R57C		E63K				K98R	Q107K		T122S			A144T A159S
d5-14	L24R		L32S		E35G		R57C		E63K				K98R	Q107K		T122S			A144T
d5-15		D26Y	L32S		E35G		R57C		E63K				K98R	Q107K		T122S			A144T
d5-16			L32S		E35G		R57C		E63K				K98R	Q107K		T122S			A144T
d5-18	L24R		L32S		E35G		R57C		E63K				K98R	Q104K	Q107K	T122S			A144T
d5-19	F21L		L32S		E35G		R57C		E63K				K98R	Q104K	Q107K	T122S			A144T

Figure 2.8 Mutations from d5-19 variant mapped onto the T7 RNAP initiation complex crystal structure. Mutations and split site mapped onto T7 RNAP crystal structure. $RNAP_N$ shown in orange and $RNAP_C$ shown in green. Mutations from previous evolution (N-29-1) shown in blue and new mutations from d5-19 shown in red. Mutations in regions of the protein that do not show up in the structure omitted. (PDB 1QLN).[70,71]

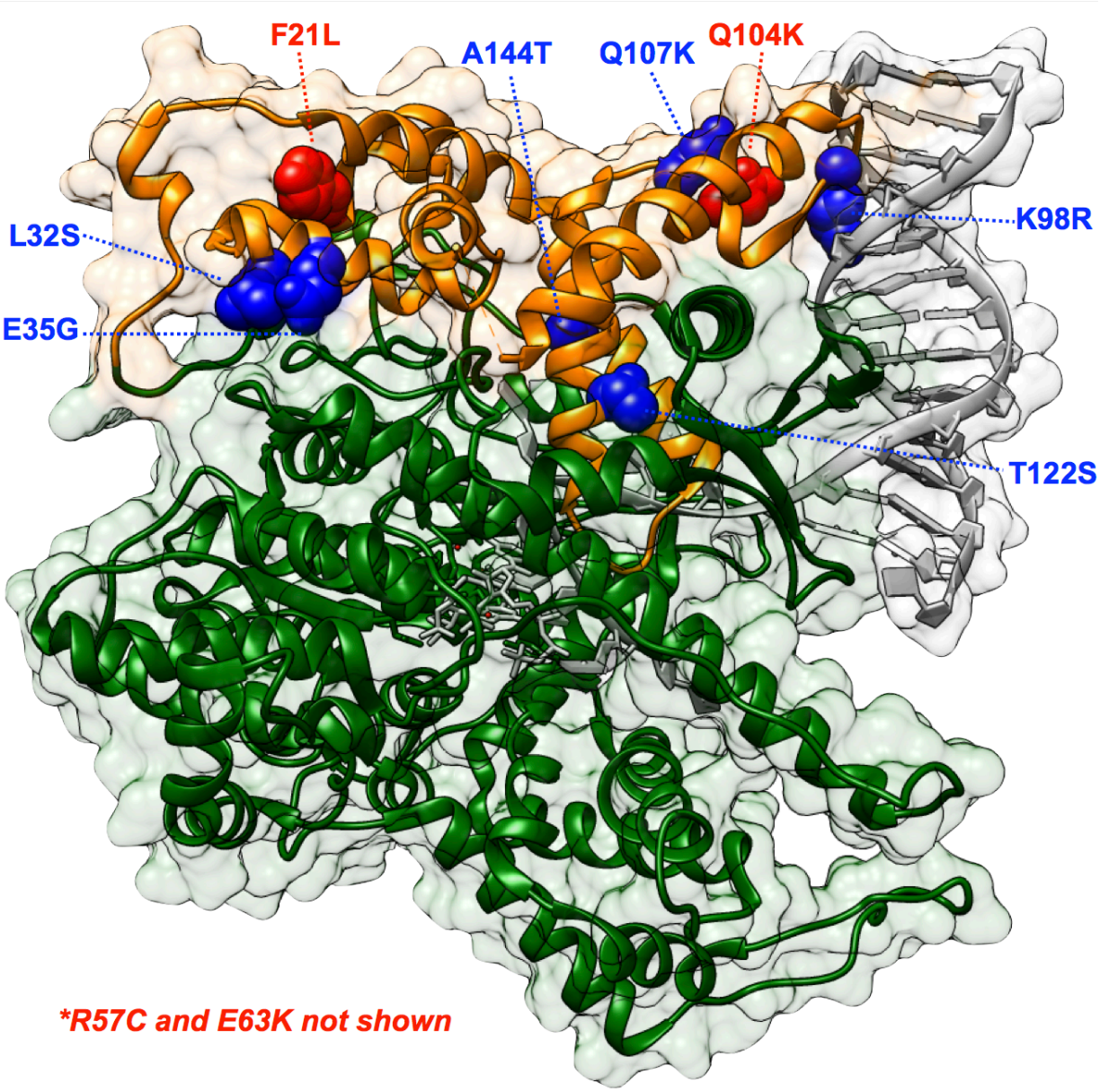


Figure 2.9 Complete imaging series of Figure 2.2E. HEK293T cells transfected with the plasmids shown in **Figure 2.2D**. 7 h after transfection, the cells were treated with either nothing or 10 μM ABA. After an additional 23 h of growth, the cells were loaded with 1 μM Hoechst 33342 and analyzed by fluorescence microscopy. 100 μm scale bar shown.

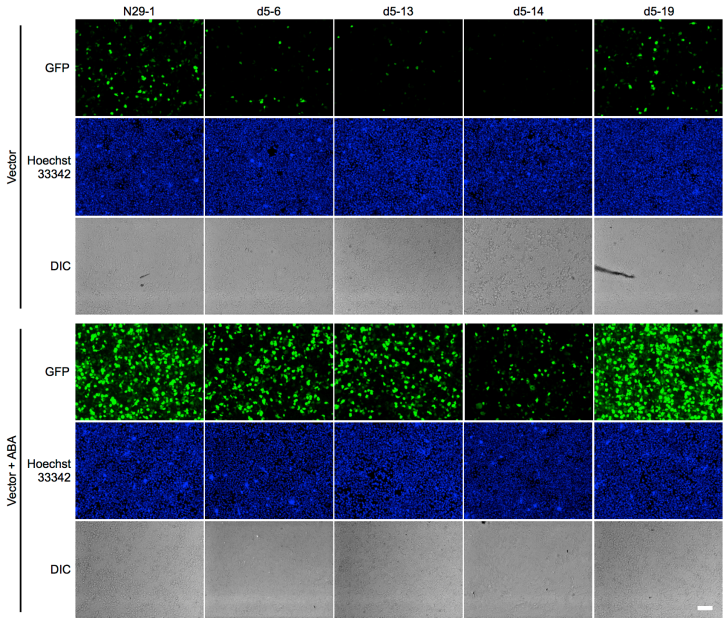


Figure 2.10 Quantification of experiment shown in Figures 2.2E and 2.9. Error bars are ± SEM, n = 5.

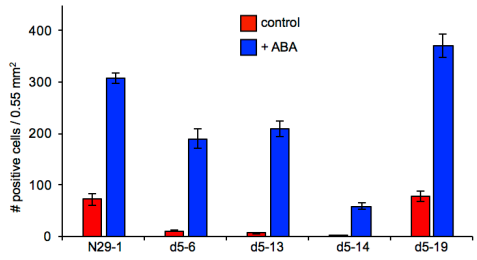


Figure 2.11 Dose response to ABA. HEK293T cells transfected with pJin 278. During transfection, the cells were treated with 0, 1, 10, or 100 μM ABA. After 46 h of growth, the cells were loaded with 1 μM Hoechst 33342 and analyzed by fluorescence microscopy. 100 μm scale bar shown. For quantification of imaging, error bars are ± SEM, n = 4.

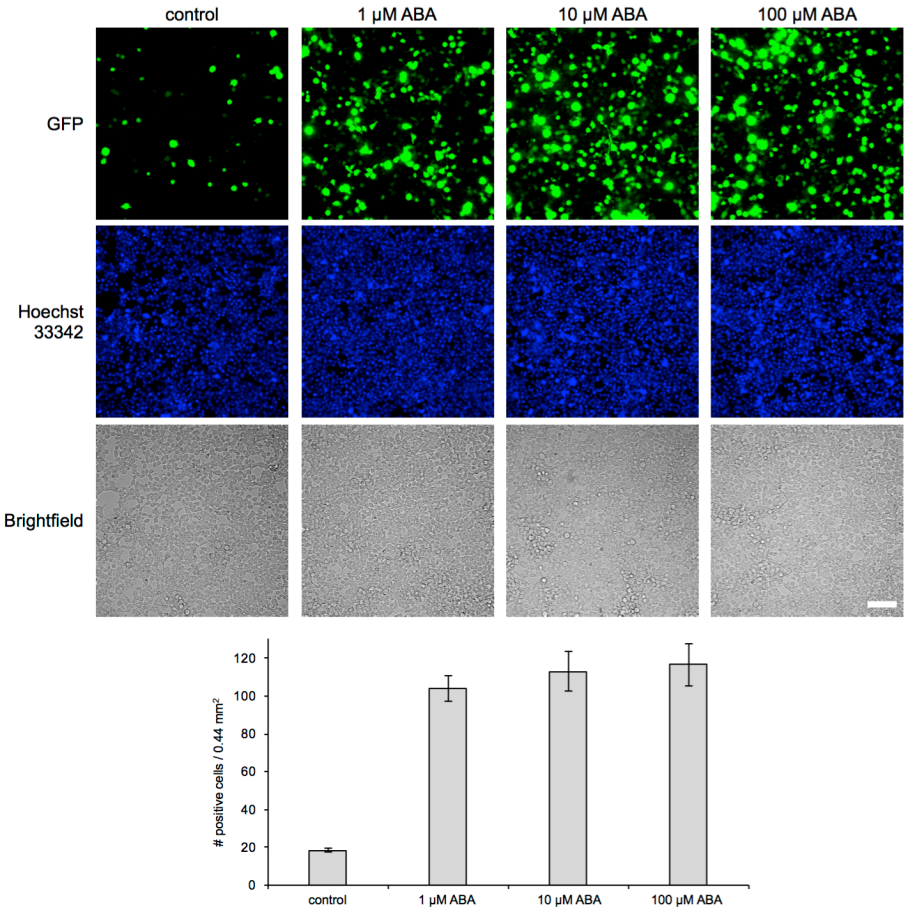
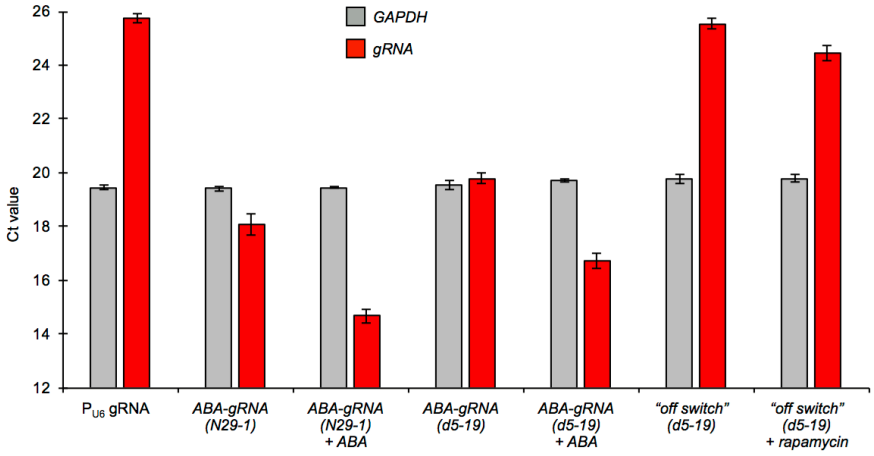


Figure 2.12 Analysis of gRNA made by split RNAP vectors compared to constitutive gRNA vectors. HEK293T cells transfected with either a PU6-driven gRNA vector (p5-54), the N-29-1 ABA-inducible vector shown in **Figure 2.1D** (pJin 239), the "on switch" d5- 19 ABA-inducible vector shown in **Figure 2.4A** (pJin 264), or the rapamycin-inducible "off switch" vector shown in **Figure 2.4A** (pJin 290). 15 h after transfection, the cells were treated with DMSO control, 10 μM ABA, or 10 nM rapamycin. 26 h after treatment, total RNA was collected from the cells and the amount of gRNA was analyzed by RT-qPCR. GAPDH was also analyzed as a control for RNA isolation. As seen in the plot, both the N-29-1 and d5-19 produce more gRNA than the constitutive PU6-driven vector, even though they display lower background and lower levels of target knockout. Additionally, the rapamycin-inducible vector shows a very low background of gRNA production, but also modest rapamycin-induced gRNA production, providing an explanation as to why the "off switch" vector did not completely block the Cas9 response. This is possibly due to the diminished activity of the CGG-RNAPC variant. Error bars are ± SEM, n = 3.



CHAPTER 3

DEVELOPMENT OF A SPLIT ESTERASE FOR PROTEIN-PROTEIN INTERACTION-DEPENDENT SMALL-MOLECULE ACTIVATION

This chapter is reproduced and adapted from the publication: Jones, K. A.; Kentala, K.; An, W; Lippert, A.; Lewis, J.; Dickinson, B. C. “Development of a split esterase for protein-protein interaction-dependent small-molecule activation.” *ACS Cent Sci* **2019**, *5*(11), 1768-1776.

3.1 Abstract

Split reporters based on fluorescent proteins and luciferases have emerged as valuable tools for measuring interactions in biological systems. Relatedly, biosensors that transduce measured input signals into outputs that influence the host system are key components of engineered gene circuits for synthetic biology applications. While small-molecule-based imaging agents are widely used in biological studies, and small-molecule-based drugs and chemical probes can target a range of biological processes, a general method for generating a target small molecule in a biological system based on a measured input signal is lacking. Here, we develop a proximity-dependent split esterase that selectively unmasks ester-protected small molecules in an interaction-dependent manner. Exploiting the versatility of an ester-protected small- molecule output, we demonstrate fluorescent, chemiluminescent, and pharmacological probe generation, each created by masking key alcohol functional groups on a target small molecule. We show that the split esterase system can be used in combination with ester-masked fluorescent or luminescent probes to measure protein−protein interactions and protein−protein interaction inhibitor engagement. We demonstrate that the esterase-

based reporter system is compatible with other commonly used split reporter imaging systems for the simultaneous detection of multiple protein−protein interactions. Finally, we develop a system for selective small-molecule-dependent cell killing by unmasking a cytotoxic molecule using an inducible split esterase. Presaging utility in future synthetic biology-based therapeutic applications, we also show that the system can be used for intercellular cell killing via a bystander effect, where one activated cell unmasks a cytotoxic molecule and kills cells physically adjacent to the activated cells. Collectively, this work illustrates that the split esterase system is a valuable new addition to the split protein toolbox, with particularly exciting potential in synthetic biology applications.

3.2 Introduction

Protein−protein interactions (PPIs) are critical regulators of diverse cellular processes[72-75] and are increasingly recognized as viable therapeutic targets for the treatment of multiple disease states.[76-78] Synthetic biology-based biosensor systems that drive cell fate changes based on measured PPIs are increasingly critical components of engineered gene circuits. For example, split proteases can be used to engineer cell receptors[79,81] and drive gene expression, and split Cas9[81] can be used to control gene editing, both based on fused sensor domains triggered by a PPI. Our group has developed proximity-dependent split T7 RNAPs as a versatile strategy for encoding PPIs in RNA signals for applications in biosensing, cell engineering, and directed evolution.[82-85] In a similar approach, a bioluminescence resonance energy transfer-based system that enables transcriptional activation with improved PPI specificity was recently engineered.[86] While methods to generate genetic output responses based on PPI inputs are continuing

to improve, a simple and general method to create a small-molecule output based on measured PPIs is lacking, despite the versatility of small molecules as biological indicators and mediators.

Aside from exploiting PPIs for engineering purposes, methods for detecting the interactions between proteins and the disruption of those interactions due to therapeutic target engagement in live cells are critical for understanding PPIs.[87] One of the most widely used approaches to measure PPIs in live cells is protein fragment complementation assays (PFCs), which involves fusing interacting proteins to complementary fragments of a split protein reporter.[88-90] Interactions between the fused proteins drive assembly of the split reporter, which in turn generates an output signal. In the context of analysis, a variety of reporters have been developed, including luciferases,[91] fluorescent proteins,[92] and horseradish peroxidase,[93] which produce photons, fluorescence, or reactive molecule outputs, respectively.

We envisioned a new split reporter system that would be capable of unmasking small molecules in a PPI-dependent manner. Small molecules provide substantial flexibility as an output signal for both analysis and generating bioactive molecules for synthetic biology purposes. Selective enzyme/ substrate pairs that are orthogonal to endogenous cellular machinery have found utility for neuronal imaging,[94] cell- specific pharmacology,[95-97] and imaging of cellular inter- actions.[98,99] For example, porcine liver esterase (PLE) was discovered to be able to process 1-methylcyclopropyl (CM)-masked phenol substrates when expressed in human cells, whereas endogenous esterases are not capable of processing this bulky protecting group.[94] We reasoned that selective

enzyme/substrate pairs may also be amenable to the creation of biosensor platforms by identifying proximity-dependent split sites of the enzymes that can ultimately unmask small molecules based on PPI-driven enzyme assembly. While it is conceptually possible to use the pre-existing split β-lactamase[100] with β-lactam prodrugs[101] or split β-galactosidase[102] with galactoside prodrugs,[103,104] this has not yet been demonstrated. We envisioned using a selective esterase−ester pair for this strategy. Exogenous esterases have previously been expressed in mammalian cells for prodrug metabolism,[105] analysis of calcium signaling,[106] cell-type specific pharmacology,[94,107] and neuronal imaging.[94] In addition, ester masking strategies for a variety of functional groups, including CM-masked phenol substrates, are well-established.[94,108-111] On the basis of this work, we aimed to engineer a selective split esterase−ester substrate pair that would be capable of unmasking small molecules in a PPI-dependent manner (**Figure 3.1A**).

In this work, we develop a split BS2 esterase system using a fluorescence-based screen to identify interaction-dependent esterase fragments in Escherichia coli. We then demonstrate that the split esterase is capable of detecting interactions between multiple PPIs in mammalian cells, including small molecule-induced dimerization domains, leucine zipper peptides, and medically relevant proteins of the B cell lymphoma (Bcl-2) family of apoptotic regulatory proteins and Bcl-2 homology 3 (BH3)-only interacting domains. We validate that the split esterase system is capable of measuring time-dependent engagement of a small-molecule PPI inhibitor in live cells. Moreover, we go on to multiplex our split esterase reporter with existing PCA technologies to simultaneously monitor two PPIs in both the intracellular and extracellular environment. In addition to demonstrating outputs from fluorescent and chemiluminescent probes, we go on to show that the split

esterase can generate a bioactive molecule output to control cellular cytotoxicity, highlighting the versatility of small-molecule signals. Taken together, these results establish the split esterase as a versatile new addition to the PCA toolbox.

3.3 Results

3.3.1 Development of a Proximity-Dependent Split Esterase

To develop a split esterase reporter, we aimed to leverage the fluorescein α-cyclopropyl ester (fluorescein-CM2) fluorogenic molecule to rapidly screen esterase cut sites for PPI- dependent esterase activity in E. coli. We selected BS2 esterase from Bacillus subtilis as our exogenous esterase, which, like PLE, acts on sterically hindered esters. BS2, however, can be efficiently expressed in E. coli and has been successfully used in plate-based screens with high enzymatic activity.[112] First, to confirm BS2-

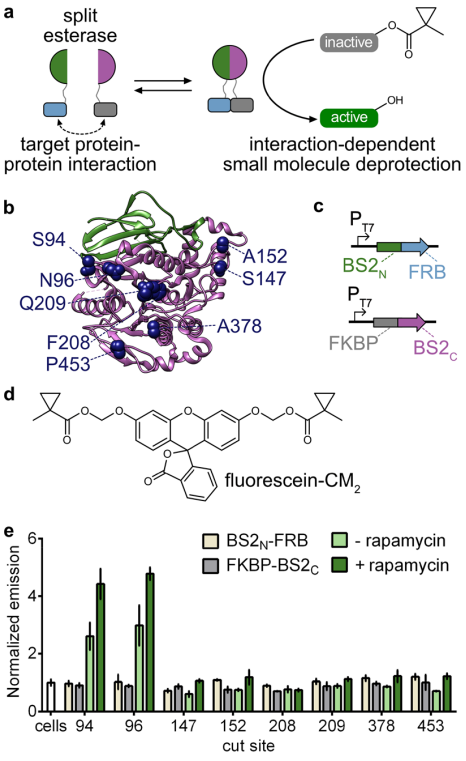


Figure 3.1 Development of a split esterase sensor. (a) Schematic of a PPI-driven split esterase assembly to unmask methylcyclopropyl (CM)-masked molecules. (b) Mapping the cut sites onto a homologous B. subtilis esterase structure (PDB 1QE3). The $BS2_N$ fragment (green) and $BS2_C$ (magenta) from the lead split site, S94, are shown. Split sites occur after the designated amino acids. (c) Vector system to identify PPI-dependent esterase fragments in E. coli. (d) Chemical structure of masked fluorophore fluorescein-CM2. (e) Fluorescence output of split esterase fragments. E. coli expressing $BS2_N$-fused FRB (tan), FKBP-fused $BS2_C$ (gray), or both in the absence (light green) or presence (green) of rapamycin were incubated with fluorescein-CM_2 for 4 h and then analyzed for fluorescence. Error bars are the standard deviation for n = 3 replicates.

mediated unmasking of fluorescein-CM2, we incubated fluorescein-CM2 with E. coli expressing BS2, PLE, or a negative control protease. Both BS2 and PLE showed an enhanced fluorescent signal due to fluorescein-CM2 unmasking (**Figure 3.7A**), but BS2 showed significantly greater activity, likely due to improved expression in E. coli. More importantly, when expressed in mammalian cells, both BS2 and PLE show a robust fluorescent signal (**Figure 3.7B-D**). On the basis of these results, we moved forward with BS2 as our target for split esterase development.

To develop a split esterase, we screened eight potential split sites on BS2, each located on surface-exposed loop regions (**Figure 3.1B**). We fused the split esterase fragments via flexible linkers to the rapamycin dimerization domains, FRB and FKBP,[113] co-expressed the fragments in E. coli, and measured the activity on fluorescein-CM2 in the absence and presence of rapamycin (**Figure 3.1C, D**). Splitting of BS2 at two of the eight sites, positions 94 and 96, produced robust enzymatic activity, which was enhanced in the presence of rapamycin (**Figure 3.1E**). We selected cut site 94 (resulting in a 10.5 kDa N-terminal fragment and a 43.6 kDa C-terminal fragment) as our lead and further validated esterase assembly with an abscisic-acid- inducible dimerization system by fusing the fragments to the ABI and PYL proteins.[114] Similarly, esterase assembly was enhanced with the addition of the small-molecule dimerization trigger in E. coli (**Figure 3.8**). Given the performance of the system, we imaged lysate from E. coli co-transformed with N- terminal split BS2 ($BS2_N$)-fused FRB and FKBP-fused C- terminal split BS2 ($BS2_C$) as a proof-of-principle experiment toward a cell-free synthetic enzymatic detection system.[115] Indeed, the split esterase assembly was robust enough to detect fluorescein-CM2 cleavage by eye and to discern the presence of rapamycin (**Figure 3.9**).

3.3.2 Monitoring Small-Molecule-Activated PPIs in Mammalian Cells with Split BS2 Esterase

Given that the split esterase performed well in E. coli, we next sought to determine if the system could function in mammalian cells. We first fused the split esterase fragments to tightly binding leucine zipper domains ZA and ZB[116] ($BS2_N$-ZA and ZB-$BS2_C$) to optimize the deployment of the reporter in mammalian cells, including vector concentrations and experimental timing (**Figure 3.10**). Once we obtained optimized conditions, we then tested the rapamycin-dependent dimerization system in mammalian cells. A fluorescent signal, and therefore BS2 activity, was only observed in the presence of rapamycin in both imaging and plate-reader-based cell assays (**Figure 3.2** and **Figure 3.11**). While esterase activity was observed in the absence of rapamycin in E. coli, esterase assembly was fully PPI-dependent in mammalian cells. We hypothesized that this discrepancy was due to a substantial overexpression of esterase fragments in E. coli, as compared with the relatively lower concentrations achieved in mammalian cells, which enables purely PPI-dependent assembly. Similar to the rapamycin system, we also observed robust PPI-dependent BS2 activity in mammalian cells with the ABA-inducible dimerization system (**Figure 3.12**).

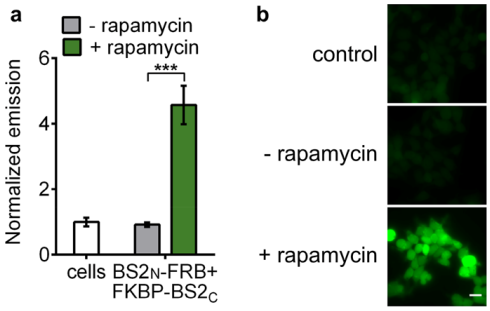


Figure 3.2 Split BS2 can detect small-molecule-activated PPIs. (a) HEK293T cells co-transfected with $BS2_N$-fused FRB and FKBP-fused $BS2_C$ or HEK293T control cells (white) were incubated with rapamycin (green) or a DMSO control (gray). After 24 h, fluorescein-CM_2 was applied, and the cells were analyzed for fluorescence. (b) HEK293T cells were treated identically to conditions in part a and analyzed by fluorescence microscopy. Error bars are the standard deviation for n = 4 replicates. Unpaired t test; ***P < 0.0001. Scale bars shown are 20 μm.

3.3.3 Monitoring BCL-2 Family PPIs and Inhibitor Engagement

Having established small-molecule-dependent activation of split BS2 in mammalian cells, we next examined whether the system could detect therapeutically relevant PPIs. We selected the Bcl-2 family of apoptotic regulatory proteins to detect both the interaction network with BH3 domains and their pharmacological engagement.[117] We generated fusions of $BS2_C$ to the antiapoptotic proteins Bcl-2 and Mcl-1 and cloned fusions of $BS2_N$ to the BH3 binding domains of tBID, which binds Bcl-2 and Mcl-1, NOXA, a selective Mcl-1 ligand, or deadBID, a control peptide without the necessary BH3 domain (**Figure 3.3A**). Co-transfection of BS2N-fused deadBID with either of the antiapoptotic protein-fused $BS2_C$ showed no significant signal over non-transfected cells incubated with fluorescein-CM2 as measured by either imaging or plate reader assays (**Figure 3.3B, C** and **Figure 3.13**). As expected, Mcl-1-fused $BS2_C$ showed significant esterase assembly when combined with either tBID- or NOXA-fused $BS2_N$, while Bcl-2-fused $BS2_C$ only showed robust esterase activity in combination with tBID- fused $BS2_N$. Collectively, these results indicate that split BS2 is capable of measuring biologically relevant PPIs in mammalian cells in a manner that recapitulates the well-studied selectivity profiles of these PPI targets.

We next aimed to measure pharmacological engagement of PPI inhibitors, a common use of PCA systems, which would also allow us to test the time-dependency of split BS2 disassembly. We selected the FDA approved Bcl-2 inhibitor, ABT-199 (Venetoclax),[118] and monitored tBID-fused BS2N and Bcl-2-fused BS2C disassembly over time (**Figure 3.3D** and **Figure 3.14A, B**). Treatment resulted in significantly

decreased esterase activity within 1 h with no appreciable esterase activity detected after 6 h. We compared this to Nanobit, a structurally optimized split NanoLuc (Nluc) luciferase reporter, which has been used to monitor numerous PPIs and their modulation by small molecules (**Figure 3.14C**).[119] While both systems reported on the ABT-199-mediated blockade of tBID-Bcl-2 assembly with approximately 10-fold reduction in signal on comparable time scales, a high level of signal remained with Nanobit. Nonetheless, these results confirm that split BS2 performs similarly to an established PPI inhibitor screening system, such as the state-of-the-art Nluc system.

Since both Nluc and split BS2 reporters measure pharmacological engagement in mammalian cells, and the two PCA systems should be orthogonal to one another, we reasoned that we could multiplex the reporters to simultaneously detect two PPIs and Bcl-2 inhibition (**Figure 3.3E**). We co-transfected one cell population with $BS2_N$-fused tBID and Bcl-2-fused $BS2_C$ and another population with Nluc11S-fused tBID and Mcl-1-fused Nluc114. The two cell populations were then mixed and treated with ABT-199. As expected, only the Bcl-2 and tBID interaction was blocked, as esterase activity (as measured by fluorescence) was significantly decreased in the presence of the inhibitor while Nanobit activity (as measured by luminescence) remained constant (**Figure 3.3F**). We also swapped the fusion partners on Nluc and split BS2, with Nanobit reporting on the Bcl-2/tBID interaction and split BS2 reporting on the Mcl-1/tBID interaction, and observed a decrease in luminescence, but not fluorescence, with ABT-199 treatment.

3.3.4 *Generation of a Chemiluminescent Signal Output for Detecting Extracellular PPIs*

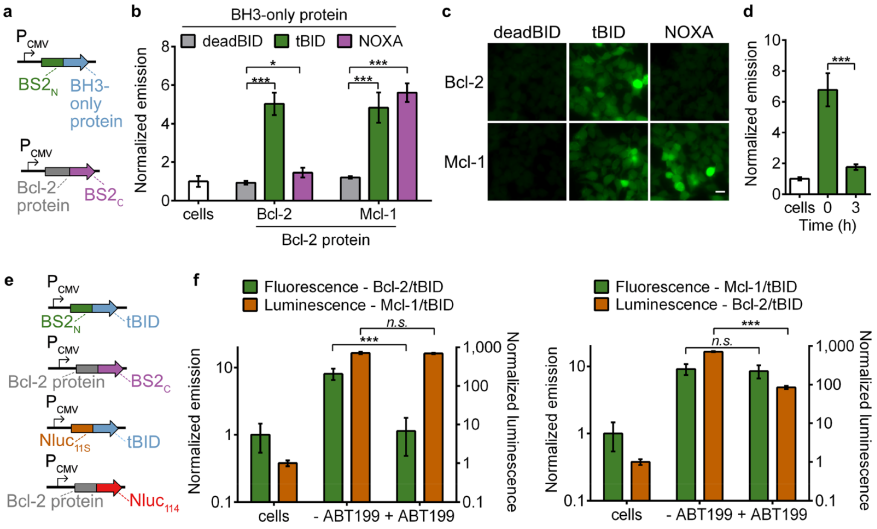


Figure 3.3 Split BS2 can detect Bcl-2 family PPIs and inhibitor engagement. (a) Vector system to test Bcl-2 split esterase PPI detection. (b) HEK293T cells cotransfected with the plasmids shown in part a were incubated with fluorescein-CM_2 and analyzed for fluorescence by a plate reader. The normalized emission for interactions between deadBID/Bcl-2 protein (gray), tBID/Bcl-2 protein (green), and NOXA/Bcl-2 protein is shown. HEK293T control cells (white) were similarly analyzed. (c) HEK293T cells cotransfected and incubated with fluorescein-CM2 as in part b were analyzed by fluorescence microscopy. (d) HEK293T cells cotransfected with $BS2_N$-fused tBID and Bcl-2-fused $BS2_C$ were treated with DMSO (time = 0) or ABT199 for 0−3 h (green) followed by incubation with fluorescein-CM2 and analyzed for fluorescence. (e) Vector system to simultaneously detect two PPIs and Bcl-2 inhibition. (f) HEK293T cells were cotransfected with split BS2 or Split Nluc plasmids shown in part e. The two cell populations were mixed and treated with ABT- 199 or a DMSO control. After 24 h, the cells were incubated with fluorescein-CM_2 and analyzed (green). Immediately after analysis, furimazine was added, and the cells were reimaged (orange). The Bcl-2/tBID interaction was selectively blocked and detected with the esterase reporter (left) or Nluc reporter (right). HEK293T control cells were similarly analyzed. Error bars are the standard deviation for n = 4 (b), n = 6 replicates (d), and n = 8 replicates (f). Unpaired t test; *P < 0.01, ***P < 0.0001. Scale bars shown are 20 μm.

On the basis of the ability of the split esterase to sensitively and selectively monitor intracellular PPIs in mammalian cells, we next aimed to detect an extracellular PPI. Such an extracellular system could be useful for investigating both ligand- and receptor-mediated dimerization of transmembrane cell-surface receptors, such as receptor tyrosine kinases or G protein-coupled receptors. Additionally, measuring extracellular interactions also presented us with an opportunity to

develop another signal output of split BS2. Specifically, we aimed to develop a masked chemiluminescent molecule that directly generates photons in an activity-dependent manner by split BS2. To this end, we first synthesized Chemilum-CM (**Figure 3.4A**, **Figure 3.15** and **Note 3.1** and **Note 3.2**), a methylcyclopropyl ester-masked pro-chemiluminescent substrate based on a previously reported scaffold.[120,121] We reasoned that esterase activity on Chemilum-CM would release the chemiluminescent form of the probe, which spontaneously reacts to generate a photon and emit steady-state luminescence. To confirm that Chemilum-CM can act cooperatively with BS2 to generate luminescence, we incubated Chemilum-CM with HEK293T cells transfected with a glycophosphatidyl inositol (GPI)-anchored[122] full-length BS2 esterase and observed a robust luminescence signal (**Figure 3.16**). We next examined whether the split esterase could provide a sensitive and fast readout on extracellular PPIs. HEK293T cells co-transfected with GPI-anchored $BS2_N$-fused FRB and FKBP-fused $BS2_C$ were cultured in the presence and absence of rapamycin over time and subsequently incubated with Chemilum-CM (**Figure 3.4B, C**). Luminescence increased with rapamycin concentration and was observed 30 min post- rapamycin-addition and increased over time, suggesting that split esterase assembly is occurring at the cell membrane (**Figure 3.4C, D**). Luminescence also corresponded to the amount of split esterase used for transfection (**Figure 3.16B**). The BS2/Chemilum-CM system therefore functioned as a "synthetic luciferase", allowing us to perform side-by-side direct comparisons between split BS2 and Nluc.

We next tested whether we could monitor intracellular and extracellular PPIs simultaneously, using the selective luminescent signals generated from Nluc/luciferin and

BS2/Chemilum-CM2. We used $BS2_N$-ZA and ZB-$BS2_C$ with GPI anchors to monitor extracellular interactions (**Figure 3.4E**) and used Zbneg,[116] a triple mutant of ZB with weakened affinity for ZA, as a negative BS2C fusion control. To monitor intracellular PPIs, we used Nluc11S-fused FRB and FKBP-fused Nluc114. All four plasmids were simultaneously transfected into mammalian cells, followed by treatment with rapamycin or DMSO carrier as a control. After 24 h, we added Chemilum-CM2, measured luminescence, and observed robust esterase activity only in cells expressing the ZA/ZB interaction fusions (**Figure 3.4F**). Cells expressing the Zbneg fusion did not exhibit significant chemiluminescence over un-transfected control cells, indicating that

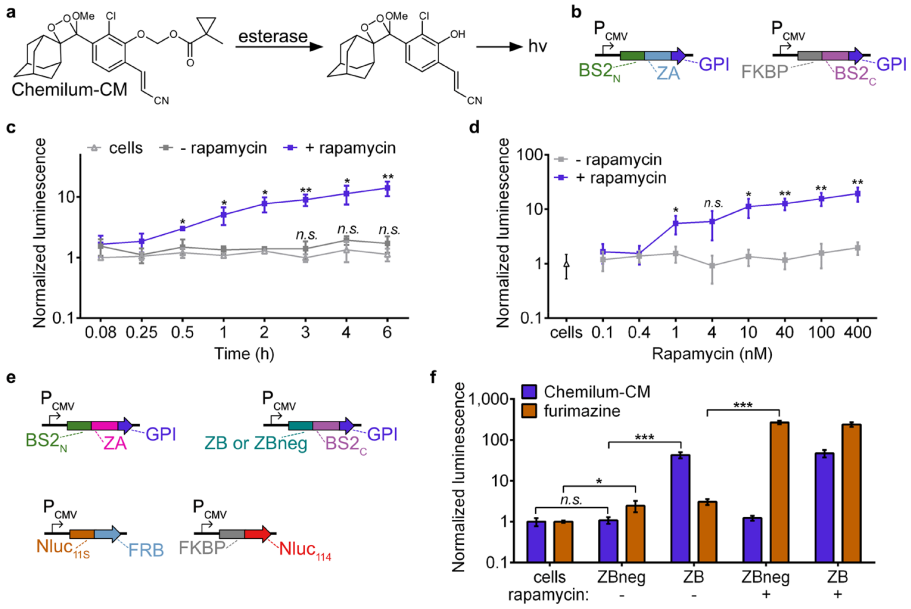


Figure 3.4 Multiplexed PPI analysis with split BS2. (a) Chemilum-CM is unmasked by esterase activity and generates a photon. (b) System to monitor extracellular PPIs. (c) HEK293T cells were cotransfected with plasmids shown in part b. Rapamycin (blue) or a DMSO control (gray) were added to cells for 0−6 h. Media was replaced with Chemilum-CM and analyzed for luminescence. (d) HEK293T cells cotransfected as in part c or HEK293T control cells (triangle) were incubated with rapamycin (blue) or a DMSO control (gray). After 6 h, the cells were analyzed with Chemilum-CM as in part c. (e) System to simultaneously monitor extracellular and intracellular PPIs. (f) HEK293T cells were transfected with all four plasmids shown in part d. Rapamycin or a DMSO control was added to the cells for 24 h. Media was replaced with Chemilum-CM (10 μM) and analyzed for luminescence (blue). The cells were then rinsed, incubated with furimazine, and analyzed for bioluminescence (orange). Error bars are the standard deviation for n = 4 replicates. Unpaired t test; *P < 0.01, **P < 0.001, ***P < 0.0001.

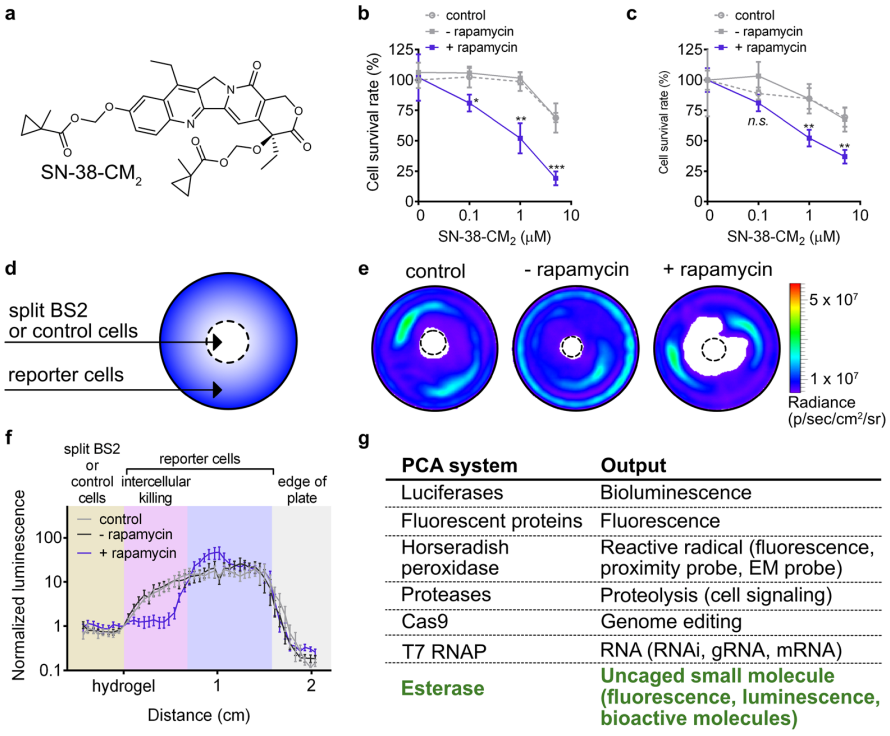
esterase assembly was interaction-dependent, even on the cell surface. We then rinsed the cells, administered the Nanobit substrate, furimazine, and again measured luminescence. As expected, cells that were treated with rapamycin showed significantly enhanced luminescence activity over DMSO- treated cells. While luminescence was also observed in DMSO- treated cells, the signal was comparable to cells expressing just Nluc11S-fused FRB or mismatched PPI controls and, moreover, is a previously reported artifact attributed to the background of the N-terminal fragment of Nluc (**Figure 3.17**).[119] Collectively, these results confirm not only that split BS2 functions as a proximity-dependent split reporter at the cell surface but also that the BS2/Chemilum-CM system is fully compatible with commonly used split luciferase reporters.

3.3.5 Proximity-Dependent Uncaging of Bioactive Molecules

Finally, we tested the utility of the split esterase to unmask a pharmacological agent in a proximity-dependent manner. The prodrug irinotecan (CPT-11) is converted to the cytotoxic active metabolite SN-38, a 1000-fold more potent topoisomerase-1 inhibitor, by carboxylesterases.[105,123,124] Irinotecan is currently in clinical trials for neuroblastoma,[125] colon cancer,[126] and other solid malignancies.[126-132] New strategies to more effectively activate irinotecan by modified rabbit and human carboxylesterases are in development.[105,123] We reasoned that our new split esterase system could potentially be used to activate irinotecan analogues in an orthogonal and programmable manner. We therefore synthesized a new methylcyclopropyl-masked version of irinotecan to generate

SN-38- CM2 (**Figure 3.5A**, **Note 3.3**). Incubation of SN-38-CM2 with BS2 in vitro showed >95% conversion to SN-38 within 5 min (**Figure 3.18**).

To test the PPI-dependent activity of the BS2/SN-38-CM2 system, we incubated SN-38-CM2 with MDA-MB-231 cells stably expressing luciferase and co-transfected with GPI-anchored $BS2_N$-fused FRB and FKBP-fused $BS2_C$ in the presence or absence of rapamycin. After incubation with the masked pharmacological agent over time, cells were rinsed and cultured for 40 h prior to measuring cell viability via luciferase assay. Increased cytotoxicity was observed in cells treated with rapamycin over DMSO-treated cells and was



PCA system	Output
Luciferases	Bioluminescence
Fluorescent proteins	Fluorescence
Horseradish peroxidase	Reactive radical (fluorescence, proximity probe, EM probe)
Proteases	Proteolysis (cell signaling)
Cas9	Genome editing
T7 RNAP	RNA (RNAi, gRNA, mRNA)
Esterase	**Uncaged small molecule (fluorescence, luminescence, bioactive molecules)**

Figure 3.5 Small-molecule-induced intra- and intercellular cell death. (a) SN-38-CM_2. (b) MDA-MB-231 luciferase cells were cotransfected with GPI-anchored $BS2_N$-FRB and FKBP-$BS2_C$. Rapamycin or DMSO control was added to cells for 12 h prior to addition of SN-38-CM_2. After 40 h, they were imaged. (c) MDA-MB-231 cells were cotransfected with GPI-anchored $BS2_N$-FRB and FKBP-$BS2_C$ and coplated with MDA-MB-453 luciferase cells. Rapamycin or DMSO control and SN-38-CM_2 were added and imaged as in part b. (d) Coculture cytotoxicity assay. Split BS2 or control cells were plated in the center of a dish. MDA-MB-453 luciferase cells surround them. Cleavage of SN-38-CM_2 results in induced killing of neighboring cells. (e) Images of cocultures after treatment as in part d. (f) Quantification of d. (g) Commonly used PCA systems and their outputs. Error bars: standard deviation for n = 4 replicates (b, c) and standard error of the mean for n = 8 replicates (e). Unpaired t test; *P < 0.01, **P < 0.001, ***P < 0.0001.

comparable to activity observed with BS2 (**Figure 3.5B**, **Figure 3.19**). In addition, DMSO-treated cells and un-transfected control cells showed no significant cytotoxicity with up to 1 μM SN-38- CM2. The reduced cell survival observed at higher concentrations of SN-38-CM2 for control conditions is likely due to slow background hydrolysis of the CM motif as shorter incubation times do not show any significant changes to cell proliferation (**Figure 3.19**). If background hydrolysis unmasks only 2−5% of SN-38-CM2 at the highest concentration, significant cell toxicity would be expected on the basis of reported IC50 values with comparable treatment times.[131] While careful tuning of the SN-38-CM2 incubation is necessary to ensure maximum PPI-dependent esterase-mediated cell death, our experiments suggest that further optimization of the BS2/ SN-38-CM2 system could effectively result in complete reduction of cell survival rate.

Since the split esterase/SN-38-CM2 system induced robust cytotoxicity in cells expressing the esterase system, we aimed to detect intercellularly mediated cell death. Such an approach, if successful, could lead to “sentinel” cells that respond to activation signals and release cytotoxic molecules that kill adjacent cells only when activated. To test this idea, we first co-plated MDA-MB-231 cells co-transfected with GPI-anchored $BS2_N$-fused FRB and FKBP-fused $BS2_C$ alongside MDA-MB-453 luciferase cells. We incubated the cells in the presence or absence of rapamycin and SN-38-CM2 and then imaged with D-luciferin. Indeed, we observed rapamycin-dependent cytotoxicity of the second cell population (**Figure 3.5C**).

To further visualize intercellular-mediated cell death based on the relative location of a second cell population using a bystander effect, we used MDA-MB-231 cells co-

transfected with GPI-anchored $BS2_N$-fused FRB and FKBP-fused $BS2_C$ suspended in a hydrogel, plated in the middle of a 3.5 cm^2 tissue culture dish in the presence or absence of rapamycin with SN-38-CM2 (**Figure 3.5D**). After the matrix solidified, MDA-MB-453 luciferase expressing cells were plated around the hydrogel. The cells were then rinsed and assayed for cytotoxicity via light emission after 40 h. We then quantified luminescence from the hydrogel to the edge of the plate where the most cell killing was observed (**Figure 3.5E, F**, **Figure 3.20**). The reporter cells nearest the rapamycin-treated split BS2 cells showed significant cytotoxicity indicated by reduced light emission. Light emission was comparable between control cells and split BS2 cells in the absence of rapamycin. These results confirm that cells engineered with the split esterase reporter system can generate a cytotoxic molecule and kill adjacent cells in a small-molecule-dependent manner.

3.4 Discussion

In this work, we developed a strategy for proximity-dependent uncaging of small molecules using an engineered esterase biosensor system. This approach comprises split versions of BS2 esterase that assemble to form a functional esterase when brought into contact by fused PPIs, which subsequently activates ester-masked small molecules. By utilizing small-molecule products as the output of BS2 PPI detection, we show that we can generate fluorescent, chemiluminescent, and pharmacological output signals, thereby adding a versatile new addition to the PFC toolbox (**Figure 3.5G**). The split esterase platform can enable a range of new applications in mammalian cells, including both PPI detection and synthetic biology. For in vitro applications with split BS2, current

efforts have shown that the individual fragments express poorly in E. coli, precluding adequate protein purification (**Figure 3.21**). Optimization of structural stability after enzyme fragmentation has previously been reported,[119] and we are currently investigating evolution strategies[133,134] to improve stability and solubility of split BS2.

The small-molecule output made possible by BS2 can be used to generate fluorescent or luminescent signals and is compatible with other common PCA detection tools such as luciferase for analysis purposes. By modulating the properties of the ester-masked imaging agents, it should be possible to rapidly alter the dynamic range and sensitivity of BS2 imaging. Indeed, cell permeability, pharmacokinetics, and emission maximum have all been investigated for D-luciferin, enabling improved tissue distribution, sensitivity, and dynamic range in bioluminescence imaging.[135,136]

While Chemilum-CM enabled sensitive detection of extracellular PPIs, it can currently only detect full-length BS2 intracellularly, suggesting that improved cellular uptake is necessary to detect intracellular split BS2 (**Figure 3.22**). Chemiluminescent probes with improved cellular uptake and increased quantum yields have recently been reported,[137,138] and these strategies should readily extend to our chemiluminescent scaffold. Fluorophores have also been extensively designed for enhanced brightness and photostability for cellular imaging,[139-141] and we are currently investigating ester-masking strategies for new imaging scaffolds. For in vivo imaging, masked red-shifted fluorophores should be easily accessible, as well as masked versions of other imaging modalities such as magnetic resonance-based imaging reagents.

While the CM-masked substrates were not efficiently hydrolyzed by endogenous esterases in these studies or in a panel of cell lines previously tested (including HeLa, CHO, and neuronal cells),[94] cell types that express more hydrolytic enzymes could be a concern for this strategy. When fluorescein-CM2 was incubated with different cell lines, including metastatic cancer cells, immune cells, and liver cells, varying endogenous activity on the probe was observed (**Figure 3.23** and **Figure 3.24**). We are therefore currently investigating new esters with increased steric bulk for improved biorthogonality to native esterases across a wider range of cellular environments and contexts.

In principle, any bioactive molecule that relies on an alcohol position for function can be masked and therefore unmasked by BS2 in an activity-dependent manner, which we demonstrated in this work through controlled cellular cytotoxicity of SN-38. Aside from activating killing in the cells expressing the biosensor, we also show that split BS2 in combination with an ester-masked pharmacological agent induced intercellular cell killing based on the proximity of the second cell population to the esterase expressing cells. While robust cytotoxicity is also observed in cells expressing split BS2, this could be mitigated against by expressing known resistance proteins to SN-38[142,143] and would enable a constant esterase source for adjacent cell killing. We are currently pursuing integrating split BS2 with cell engineering approaches, such as chimeric antigen receptor T cell therapy, to integrate activity-dependent small-molecule activation at cancer sites. We also anticipate additional new synthetic biology opportunities for cell engineering and cell signaling- based prodrug release with PPI-dependent split esterase small- molecule activation.

3.5 Materials and Methods

General synthetic procedures. All reactions were performed in dried glassware under an atmosphere of dry N2. Silica gel P60 (SiliCycle) was used for column chromatography, and Analytical Chromatography TLC Silica gel 60 F254 (Merck Millipore, Darmstadt, Germany) or SiliCycle 60 F254 silica gel (precoated sheets, 0.25 mm thick) was used for analytical thin layer chromatography. Plates were visualized by fluorescence quenching under UV light or by staining with iodine. Other reagents were purchased from Sigma-Aldrich (St. Louis, MO), Alfa Aesar (Ward Hill, MA), EMD Millipore (Billerica, MA), Oakwood Chemical (West Columbia, SC), TCI (Tokyo, Japan) and used without further purification. 1H NMR and 13C NMR spectra for characterization of new compounds and monitoring reactions were collected in CDCl3 (Cambridge Isotope Laboratories, Cambridge, MA or Sigma- Aldrich) on a JEOL 500 MHz spectrometer in the Department of Chemistry at Southern Methodist University or using a 500 MHz Bruker Avance II+ spectrometer with 5 mm QNP probe at the Department of Chemistry NMR Facility at the University of Chicago. All chemical shifts are reported in the standard notation of parts per million using the peak of residual proton signals of the deuterated solvent as an internal reference. Coupling constant units are in Hertz (Hz) Splitting patterns are indicated as follows: br, broad; s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet; dd, doublet of doublets; dt, doublet of triplets. High resolution mass spectroscopy was performed on a Shimadzu IT-TOF (ESI source) at the University of Texas, Arlington or Agilent 6224 TOF High Resolution Accurate Mass Spectrometer (HRA-MS, combination of APCI and ESI) at the Department of Chemistry Mass Spectrometry Facility at the University of Chicago. Low resolution mass spectral analyses and liquid chromatography

analysis of SN-38-CM2 were carried out on an Advion Expression L mass spectrometer (Ithaca, NY) coupled with an Agilent 1220 Infinity LC System (Santa Clara, CA). Synthesis of fluorescein-CM2 was carried out using literature procedures.[94]

(3-(((1*r*,3*r*,5*R*,7*S*)-adamantan-2-ylidene)(methoxy)methyl)-2-chloro-6-(-2-cyanovinyl phenoxy)methyl1-methylcyclopropane-1-carboxylate(1). Chloromethyl 1-methyl cyclopropane carboxyate1 (116.3 mg, 0.7832 mmol, 1.5 equiv) was added to the dry, N2 filled flask, then dissolved with 2.1 mL anhydrous acetone. NaI (128.6 mg, 0.8580 mmol, 1.7 equiv) was added to the solvent and the mixture was stirred for 24 h at ambient temperature. The reaction was concentrated under reduced pressure. Purification by silica column chromatography (CH2Cl2) yielded a pale yellow oil. 3-(4-(((1*r*,3*r*,5*R*,7*S*)-adamantan-2-ylidene)(methoxy)methyl)-3-chloro-2-hydroxyphenyl)acrylonitrile2,3 (184.4 mg, 0.5182 mmol, 1.0 equiv) was dissolved with 2.0 mL anhydrous DMF in an dry, N2 filled flask and anhydrous *N*,*N*-diisopropylethylamine (DIPEA, 0.19 mL, 1.1 mmol, 2.1 equiv) was added. Iodomethyl 1- methylcyclopropane-1-carboxylate was dissolved with 3.0 mL anhydrous DMF and added to the flask. The reaction mixture was stirred for 21 h at ambient temperature. The reaction was concentrated under reduced pressure. Purification by silica column chromatography (1:20 EtOAc/Hexane) yielded **1** as a white solid (107.4 mg, 44%). 1H NMR (500 MHz, CDCl3) d 7.63 (d, 1H, *J* = 16.6 Hz), 7.36 (d, 1H, *J* = 8.1 Hz), 7.11 (d, 1H, *J* = 8.1 Hz), 5.95 (d, 1H, *J* = 16.6 Hz), 5.70 (m, 2H), 3.29 (s, 1H), 3.25 (s, 1H), 2.04 (s, 1H), 1.94–1.63 (m, 12H), 1.28 (s, 3H), 1.23 (m, 2H), 0.77 (m, 2H); 13C NMR (500 MHz, CDCl3) d 175.07, 152.02, 144.76, 139.33, 139.04, 133.30,

129.21, 128.62, 128.54, 123.91, 117.83, 98.59, 89.32, 57.44, 39.19, 39.03, 38.61, 38.56, 36.96, 32.91, 31.59, 29.73, 28.29, 28.11, 19.10, 18.54, 17.60; HRMS calcd for C27H30ClNO4 (M+Na+) 490.1756, found 490.1755.

(2-chloro-6-(2-cyanovinyl)-3-((1*r*,3*r*,5*r*,7*r*)-4'-methoxyspiro[adamantane-2,3'[1,2] 56engal56ne]-4'-yl)phenoxy)methyl-1-methylcyclopropane-1-carboxylate (**Chemilum-CM**). Enol ether **1** (107.4 mg, 0.2295 mmol, 1.0 equiv) was dissolved with 5 mL THF at 0 °C in a two- neck flask and Rose 56engal (12.4 mg, 0.0122 mmol, 0.050 equiv) was added to the solvent. O2 was bubbled through the solvent when illuminated with a 120W light bulb (Home Depot, Dallas, TX). The reaction was monitored by TLC. After 3 h 15 min, the mixture was concentrated under reduced pressure. Purification by silica column chromatography (1:20 EtOAc/Hexane) yielded chemi-CM2 as a white solid (76.9 mg, 67%). 1H NMR (500 MHz, CDCl3) d 7.97 (d, 1H, *J* = 8.6 Hz), 7.66 (d, 1H, *J* = 16.7 Hz), 7.52 (d, 1H, *J* = 8.6 Hz), 6.03 (d, 1H, *J* = 17.2 Hz), 5.67 (dd, 2H), 3.21 (s, 3H), 3.01 (s, 1H), 1.96 (s, 1H), 1.85–1.58 (m, 12H), 1.28 (s, 3H), 1.23 (m, 2H), 0.79 (m, 2H); 13C NMR (500 MHz, CDCl3) d 175.11, 152.70, 144.42, 136.53, 130.51, 129.80, 127.19, 124.13, 117.63, 111.63, 99.99, 96.48, 89.32, 49.87, 36.57, 33.99, 33.66, 32.67, 32.24, 31.62, 31.58, 29.82, 26.17, 25.83, 19.18, 18.59, 17.74; HRMS calcd for C27H30ClNO6 (M+Na+) 522.1654, found 522.1654.

[(19S)-10,19-diethyl-7-[(1-methylcyclopropane carbonyloxy)methoxy]-14,18-dioxo-17-oxa-3,13-diazapentacyclo[11.8.0.02,11.04,9.015,20]henicosa1(21),2,4(9),

5,7,10,15(20)-heptaen-19-yl]oxy}methyl 1-methylcyclopropane-1-carboxylate (SN-38-CM2). In a 250 mL round bottom flask equipped with a magnetic stir bar, freshly prepared Chloromethyl 1- methylcyclopropanecarboxyate1 (1.5141 g, 10.2 mmol, 20 equiv) was dissolved in 50 mL dry DMF. SN-38 (0.2002 g, 0.51 mmol, 1 equiv) was then added followed by the sequential addition of sodium iodide (5.3476 g, 35.7 mmol, 70 equiv) and potassium carbonate (0.3522 g, 2.5 mmol, 4.9 equiv). The resulting solution allowed to stir at room temperature. After 5 h, when side products began to appear as measured by LC-MS, the solution was diluted with EtOAc, filtered through a pad of celite, and concentrated under vacuum on a rotary evaporator before drying under high vacuum until the reaction mixture was completely solid. The resulting solid was then suspended in EtOAc and filtered through another pad of celite. The resulting mixture was concentrated before purifying via silica column chromatography (0% to 100% EtOAc in Hexanes) and purifying via silica column chromatography again (0% to 5% MeOH in DCM) yielding 12.3 mg (3.9 % yield) as a light yellow solid. Purity was determined by LC-MS. Rf (5% MeOH in DCM) = 0.47. HRA-MS(+) Calculated for C34H36N2O9 [M+]: 616.2421; found 616.2431. 1H-NMR (500 MHz; CDCl3) δ: 8.23 – 8.17 (m, 1H), 7.64 – 7.60 (m, 1H), 7.55 – 7.52 (m, 1H), 5.94 (s, 1H), 5.77 (d, *J* = 15 Hz, 1H), 5.30 (m, 4H), 5.12 (s, 1H), 3.71 (s, 1H), 3.35 (s, 1H), 3.16 (m, 1H), 1.90 (m, 2H), 1.63 (s, 1H), 1.52 (m, 2H), 1.40 (m, 3H), 1.34 (s, 2H), 1.30 (m, 1H), 1.25 (s, 1H), 1.04 (m, 3H), 0.94 (m, 1H) 0.76 (m, 1H). 13C-NMR (126 MHz; CDCl3): 174.7, 174.1, 157.8, 156.1, 151.9, 150.4, 150.2, 147.5, 147.1, 145.4, 132.6, 132.2, 127.6, 127.4, 125.7, 122.6, 118.7, 114.7, 105.8, 98.2, 97.9, 85.4, 73.0, 66.5, 49.6, 31.9, 31.8, 23.4, 19.6, 19.3, 18.0, 14.2, 14.0, 8.0.

Cloning. All plasmids were constructed by Gibson assembly from PCR products generated using Phusion Polymerase or Q5 DNA Polymerase (NEB). The plasmids were sequenced by the University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping Facility. Maps for each plasmid are shown in **Figure 3.6**. All plasmids are described in **Table 3.1**, which includes links to fully annotated sequence maps. Full vector sequences are also available upon request. The original GPI-anchor plasmid was a gift from Evan W. Miller (UC Berkeley).

Esterase screening in *E. coli*. BL21 cells were transformed by heat shock with either BS2 esterase, PLE esterase, or TEV protease. Single colonies were grown to saturation overnight at 37 °C and then each well of a 96- well deep well plate containing 0.185 mL of LB with antibiotics, 10 mM arabinose, and 50 μM fluorescein-CM2 was inoculated with 15 μL of the overnight culture. After growth with shaking for 4 h at 37 °C, the cells were pelleted and the LB was removed. The cells were then resuspended in 0.2 mL PBS and transferred to a 96-well black wall, clear bottom plate (Nunc), and fluorescence intensities (λex 485/20 nm, λem 528/20 nm) and OD600 were measured on a Synergy Neo2 Hybrid Multi-Mode Reader (BioTek). The data were analyzed by dividing the emission values by the background-corrected OD600 value. All values were then normalized to the emission of cells expressing TEV protease, which was assigned an arbitrary value of 1 to allow for values from each fluorescence plot to be compared to each other. All experiments were performed in three technical replicates.

Cut site screening in *E. coli*. BL21 cells were transformed by heat shock with either BS2N-linker-FRB, FKBP-linker-BS2C, or co-transformed with both. Single colonies were grown to saturation overnight at 37 °C and then each well of a 96-well deep well plate containing 0.185 mL of LB with antibiotics, 0.1 mM IPTG, and 0.5 mM fluorescein-CM2 was inoculated with 15 µL of the overnight culture. As indicated, 20 µM rapamycin or 500 µM abscisic acid was added per well. After growth with shaking for 4 h at 37 °C, the cells were pelleted and the LB was removed. The cells were then resuspended in 0.2 mL PBS and transferred to a 96-well black wall, clear bottom plate (Nunc), and fluorescence intensities (λex 485/20 nm, λem 528/20 nm) and OD600 were measured on a Synergy Neo2 Hybrid Multi-Mode Reader (BioTek). The data were analyzed by dividing the emission values by the background-corrected OD600 value. All values were then normalized to the emission of non- transformed BL21 cells incubated with fluorescein-CM2 as indicated above, which was assigned an arbitrary value of 1 to allow for values from each fluorescence plot to be compared to each other. All experiments were performed in at least two technical replicates.

Split esterase assay in *E. coli* cell lysate. BL21 cells were co-transformed by heat shock with $BS2_N$-linker-FRB and FKBP-linker-$BS2_C$. Single colonies were grown to saturation overnight at 37 °C and then each well of a 24-well deep well plate containing 3 mL of LB with antibiotics, was inoculated with 30 µL of the overnight culture. When cells reached an OD600 of 0.7-0.8, 0.1 mM IPTG and 20 µM rapamycin was added and then cultured at 30 °C for 16 h. Cells were collected by centrifugation and the supernatant was

removed. Cell pellets were either stored at -20 °C for 5 days prior to lysis or freshly lysed in 0.3 mL lysis buffer (50 mM sodium phosphate buffer, 0.5% (vol/vol) Tween, pH 7.4). Lysed bacterial suspensions were incubated with 0.5 mM fluorescein-CM2 and 1 μM rapamycin for 30 minutes at room temperature with rocking prior to visualization. A portion (8 μL) of this mixture was then spotted on filter paper (Whatman Grade) and imaged on a ChemiDoc Imaging System (Bio-Rad) at the University of Chicago BioPhysics Core Facility. All experiments were performed in three technical replicates.

Mammalian cell culture and plasmid transfection. HEK293T (ATCC), MDA-MB-231 (courtesy of Harikrishna Nakshatri, Indiana University), MDA- MB-453 (courtesy of the Greene laboratory, Uchicago), HT-29, RAW264.7 (courtesy of the Esser- Kahn laboratory, Uchicago) cells were cultured in DMEM (L-glutamine, high glucose, sodium pyruvate, phenol red; obtained from Corning) supplemented with 10% (vol/vol) fetal bovine serum (FBS, Gemini Benchmark), and 1% (vol/vol) penicillin/streptomycin (Gibco/Life Technologies). Jurkat (ATCC), PC-3, and THP-1 (courtesy of the Esser-Kahn laboratory, Uchicago) were cultured in RPMI1640 (L-glutamine, phenol red; obtained from Corning) supplemented with 10% FBS and 1% penicillin/streptomycin. HEPG2 cells (courtesy of the Krishnan laboratory, Uchicago) were cultured in DMEM Glutamax (high glucose, sodium pyruvate, phenol red; obtained from ThermoFisher Scientific) supplemented with 10% FBS and 1% penicillin/streptomycin. Cells were maintained in a 5% CO2, water-saturated incubator at 37 °C. Transient transfections were performed using cationic lipid formulations (Lipofectamine 2000 or Lipofectamine 3000; Invitrogen)

following the manufacturer's protocol. HEK293T cells are listed in the database of commonly misidentified cell lines maintained by ICLAC (http://iclac.org/databases/cross-contaminations/). We obtained fresh cells from ATCC or early passage aliquots from the Cellular Screening Center, University of Chicago, which were frozen down at an early passage (passage 5) in individual aliquots. The cells were then used for less than 25 passages for all experiments. Multiple biological replicates were performed with cells from different passages and freshly thawed aliquots. There was no testing for mycoplasma infection or further authentication because early passage cells were used for all experiments.

Mammalian cell fluorescence imaging. HEK293T cells were plated in a 48-well plate and transfected the next day with 500 ng PLE or BS2 (to test full length esterase activity) or 2 µg $BS2_N$-POI and 500 ng POI-$BS2_C$ (to test split esterase activity) using 1.5 µL of Lipofectamine 2000 following the manufacturer's protocol. For bulk and single cell assays, cells were plated 24 h after transfection in a 96-well black wall, glass bottom plate (Cellvis) pre-treated with 0.1 mg/mL Poly-D-lysine for 2 h. When indicated, 400 nM rapamycin, 10 µM abscisic acid, or a DMSO carrier control was added. After 24 h of culturing, the media was replaced with 10 µM fluorescein-CM2 in Live Cell Imaging Solution (ThermoFisher Scientific) for bulk cell assays and fluorescence intensities (λex 490/20 nm, λem 545/20 nm) were measured on a Synergy Neo2 Hybrid Multi-Mode Reader every minute for 1 h. For single cell assays, the media was replaced with 10 µM fluorescein-CM2 and 1 µM Hoechst 33342 in Live Cell Imaging Solution. The cells were incubated for 1 h at room

temperature and then imaged on an inverted epifluorescence microscope (Leica Dmi8) equipped with a camera (Hamamatsu Orca- Flash 4.0) with either 20x air objective or 63x oil objective (N/A 1.4) and light source (Sutter Lamdba XL, 300 W Xenon) for fluorescein-CM2 (ET 490/20x, Quad-S, ET 525/36 m), Hoechst 33342 (ET 402/15x, Quad-S, ET 455/50 m), and brightfield using Leica LAS X software. For image analysis of fluorescent emission, low resolution (20x) microscopy images were analyzed in order to obtain more cells in the field and provide a better representation of the cell population. Each image for a given condition was processed to adjust for maximum brightness and areas containing cells were selected and measured in ImageJ.[144] All experiments were performed in three technical replicates.

Mammalian luciferase assay. HEK293T cells were plated in a 48-well plate and transfected the next day with 500 ng Nluc11S- POI and 500 ng POI-Nluc114 using 1.5 µL of Lipofectamine 2000 following the manufacturer's protocol. Cells were plated 24 h after transfection in a 96-well white wall plate. When indicated, 400 nM rapamycin or a DMSO carrier control was added was added. After 24 h, luciferase activity was assayed using the Nano-Glo Live Cell Assay System (Promega) on a Synergy Neo2 Hybrid Multi-Mode Reader. All experiments were performed in three technical replicates.

Inhibitor assay. HEK293T cells were plated in a 48-well plate and transfected the next day with 2 µg $BS2_N$-tBID and 500 ng Bcl-2-$BS2_C$ or 500 ng Nluc11S-tBID and 500 ng Bcl-2-Nluc114 using 1.5 µL of Lipofectamine 2000 following the manufacturer's protocol.

Cells were plated 24 h after transfection in a 96-well black wall, glass bottom plate pre-treated with 0.1 mg/mL Poly-D-lysine for 2 h for fluorescence imaging or 96-well white wall plate for luminescence imaging. For dual fluorescence and luminescence imaging, cells were plated in a 96-well black wall, glass bottom plate pre-treated with 0.1 mg/mL Poly-D-lysine for 2 h. Then, 1 μM ABT-199 or a DMSO carrier control was added during cell culture for indicated lengths of time prior to imaging. Fluorescence and luminescence imaging were performed as described above. All experiments were performed in three technical replicates.

Extracellular chemiluminescent assay. HEK293T cells were plated in a 48-well plate and transfected the next day with 500 ng $BS2_N$-FRB-GPI anchor and 500 ng FKBP-$BS2_C$-GPI anchor using 1.5 μL of Lipofectamine 2000 following the manufacturer's protocol. Cells were plated 24 h after transfection in a 96-well white wall plate for luminescence imaging. For DNA titrations, the same procedure was followed with 0.2–1000 ng total DNA (1 $BS2_N$-FRB-GPI anchor:1 FKBP-$BS2_C$-GPI anchor) used for transfections. Cells were treated with 400 nM rapamycin or a DMSO control for 24 h. For the rapamycin time course, cells were treated with 400 nM rapamycin or a DMSO control for 0.08 – 6 h. For rapamycin dose response, cells were treated with 0.1–400 nM rapamycin or a DMSO control for 6 h. For all experiments, after indicated rapamycin treatment, cells were incubated with 10 μM Chemilum-CM2 in Live Cell Imaging Solution and luminescence was measured on a Synergy Neo2 Hybrid Multi-Mode Reader every minute for 1 h. All experiments were performed in three technical replicates.

Extracellular/intracellular luminescence assay. HEK293T cells were plated in a 48-well plate and transfected the next day with 500 ng $BS2_N$-POI- GPI anchor and 500 ng POI-$BS2_C$-GPI anchor or 500 ng Nluc11S-FRB and 500 ng FKBP-Nluc114 using 1.5 µL of Lipofectamine 2000 following the manufacturer's protocol. Cells were plated 24 h after transfection in a 96-well white wall plate for luminescence imaging. When indicated, 400 nM rapamycin was added. After 24 h of culturing, the media was replaced with 10 µM Chemilum-CM2 in Live Cell Imaging Solution and luminescence was measured on a Synergy Neo2 Hybrid Multi- Mode Reader every minute for 1 h. Cells were then gently rinsed in 0.2 mL PBS and subsequently assayed using the Nano-Glo Live Cell Assay System. All experiments were performed in three technical replicates.

Recombinant protein expression and purification. The pET-BS2 plasmid was transformed into chemically competent E. coli BL21 cells. Cultures (1 L) were grown at 37 °C in LB broth (with 40 µg/mL Kan) to mid log-phase (OD600 = 0.6 – 0.8), induced with 0.2 mM IPTG, and incubated at 30 °C for 18 – 20 h. The cells were harvested by centrifugation at 4 °C and resuspended in 50 mM sodium phosphate buffer (pH 7.4), 300 mM NaCl, and 1 mM PMSF. Lysozyme (1 mg) was added and the cells were sonicated and centrifuged at 10000 x *g* for 1 h at 4 °C. BS2 was purified from clarified supernatants using nickel affinity chromatography and dialyzed into 50 mM sodium phosphate buffer (pH 7.4). Prior to storage at -20 °C, 15% glycerol was added to the sample. Final protein concentrations were determined using standard BCA assay or UV spectroscopy. SDS-PAGE was also performed and gels were stained with Coomassie R-250.

BS2 *in vitro* reactions with SN-38-CM2. Bioconversions contained purified BS2 (2 μM) and SN-38-CM2 (1 mM) in reaction buffer (50 mM sodium phosphate buffer, pH 7.4) and were run at room temperature. Aliquots of the *in vitro* reaction were removed over time (5–30 min) and quenched 1:1 with methanol. Liquid chromatography analysis was carried out on an Agilent 1220 Infinity LC System.

BS2/SN-38-CM2 cytotoxicity assay. MDA-MB-231 or MDA-MB-231 luciferase cells were plated in a 6-well plate and transfected the next day with 2500 ng $BS2_N$-FRB and 2500 ng FKBP-$BS2_C$ or 2500 ng BS2 using 5 μL of P3000 Reagent and 5 μL Lipofectamine 3000 following the manufacturer's protocol. MDA-MB-231 cells were plated 24 h after transfection in a 96-well white wall plate with MDA-MB-453 luciferase cells (1:1) for luminescence imaging. MDA-MB-231 luciferase cells were plated 24 h after transfection in a 96-well white wall plate. When indicated, 400 nM rapamycin or DMSO control was added. After 12 h of culturing, SN-38-CM2 (0.1 – 5 μM) was added. A stock solution of D-luciferin (500 μM final) was added after 40 h and luminescence was measured on a Synergy Neo2 Hybrid Multi- Mode Reader. All experiments were performed in three technical replicates.

BS2/SN-38-CM2 coculture cytotoxicity assay. MDA-MB-231 cells were plated in a 6-well plate and transfected the next day with 2500 ng $BS2_N$-FRB and 2500 ng FKBP-$BS2_C$ or 2500 ng BS2 using 5 μL of P3000 Reagent and 5 μL Lipofectamine 3000 following the manufacturer's protocol. When indicated, 400 nM rapamycin or DMSO control was added

24 h after transfection. After 12 h of culturing, cells (1 x 105) were suspended in Hydrogel (Sigma, TrueGel3D Hydrogel Kits, SLO-DEXTRAN, PEG crosslinker) and cell culture media according to the manufacture's protocol with SN-38-CM2 (1 μM) and plated in a 3.5-cm tissue culture plate (Corning). The matrix was solidified at 37 °C for 1 h and MDA-MB- 453 luciferase cells were then added (50% confluent) around the hydrogel-cell mix. After 12 h, the cells were rinsed with PBS. A stock solution of D-luciferin (250 μM final) was added after 40 h and luminescence was measured on a Xenogen IVIS 200 at the University of Chicago Optical Imaging Core Facility.

Endogenous esterase screening. Metastatic cells and liver cells: MDA-MB-231, HT-29, PC-3, and HEPG2 cells were plated in a 96-well plate and transfected the next day with 100 ng BS2 using 0.1 μL P3000 reagent and 0.15 μL Lipofectamine 3000 following the manufacture's protocol. For bulk and single cell assays, cells were plated 24 h after transfection in a 96-well black wall, glass bottom plate pre-treated with 0.1. mg/mL Poly-D-lysine for 2 h. After 24 h of culturing, the media was replaced with Live Cell Imaging Solution with or without 10 μM fluorescein-CM2 and imaged and quantified as in Mammalian cell fluorescence imaging method. Immune cells: Jurkat, RAW264.7, and THP-1 cells were plated in a 96-well black wall, glass bottom plate pre-treated with 0.1 mg/mL Poly-D-lysine for 2 h for bulk and single cell assays. For cell activation, ultrapure LPS-EB was incubated at 100 ng/mL for 18 h (RAW264.7), 100 ng/mL for 1 h (Jurkat), and 1 μg/mL for 3 h (THP-1). After culturing in the presence and absence of LPS, the

media was replaced with Live Cell Imaging Solution with or without 10 μM fluorescein-CM2 and imaged and quantified as in Mammalian cell fluorescence imaging method.

Statistical analyses. Data were analyzed using unpaired two-tailed t-tests (GraphPad Prism 6, GraphPad Software, La Jolla, CA, USA).

Data availability. All data generated or analyzed during this study are included in the published article (and its supplementary information) or are available from the corresponding author on reasonable request.

Safety statement. No unexpected or unusually high safety hazards were encountered.

3.6 Supplemental Information

Table 3.1 List of plasmids used in this work.

Vector Name	Res.	Origin	Purpose	Map*	Benchling Link
KJ89	Chlor	CloDF13	BS2 expression plasmid	a	https://benchling.com/s/seq-iefwLbJz8iB9HIDlaUsY
p9-50	Chlor	CloDF13	PLE expression plasmid	a	https://benchling.com/s/seq-j18JbOR0iSPcjqxbecFF
p1-18	Chlor	CloDF13	TEV expression plasmid	a	https://benchling.com/s/seq-GOEUaX6hsarI55zkuFV3

Table 3.1 – continued from previous page

p9-70	Kan	ColE1	PLE mammalian expression plasmid	b	https://benchling.com/s/seq-ASiFsYMKIQ9AciVPXyRk
KJ128	Kan	ColE1	BS2 mammalian expression plasmid	b	https://benchling.com/s/seq-51TZnY3q5ZfFQi2XXUhS
KJ150	Kan	ColE1	$BS2_N$-linker-FRB expression plasmid	c	https://benchling.com/s/seq-pE2lgWMhm7M4Gf38dBzJ
KJ246	Kan	ColE1	$BS2_N$-linker-ABI expression plasmid	c	https://benchling.com/s/seq-KZyurXJXoeoPALkFmGru
KJ158	Spec	CloDF13	FKBP-linker-$BS2_C$ expression plasmid	d	https://benchling.com/s/seq-l20ugDxo7pzS2F1KZPps
KJ247	Spec	CloDF13	PYL-linker-$BS2_C$ expression plasmid	d	https://benchling.com/s/seq-vWQZoM900bv37RxQkUx9
KJ174	Kan	ColE1	$BS2_N$-linker-ZA mammalian expression plasmid	e	https://benchling.com/s/seq-5PPiZDoazSDpHu9Bar1f
KJ178	Kan	ColE1	$BS2_N$-linker-FRB mammalian expression plasmid	e	https://benchling.com/s/seq-WupOwWWn4LOeeWLtYEIm
KJ242	Kan	ColE1	$BS2_N$-linker-ABI mammalian expression plasmid	e	https://benchling.com/s/seq-M4jnRrFGcwar5tktRqbG
KJ207	Kan	ColE1	$BS2_N$-linker-deadBID mammalian expression plasmid	e	https://benchling.com/s/seq-1ywROLLeq1bu9D5DLUmM
KJ206	Kan	ColE1	$BS2_N$-linker-tBID mammalian expression plasmid	e	https://benchling.com/s/seq-wdtMHBWIkHn99fPFkojE
KJ208	Kan	ColE1	$BS2_N$-linker-NOXA mammalian expression plasmid	e	https://benchling.com/s/seq-atRLE3HszeiAygxcU85K
KJ176	Kan	ColE1	ZB-linker-$BS2_C$ mammalian expression plasmid	f	https://benchling.com/s/seq-hEK3wVLVoViAdCftONWM
KJ180	Kan	ColE1	FKBP-linker-$BS2_C$ mammalian expression plasmid	f	https://benchling.com/s/seq-itvF3x0l32B9uJ2VNQ5l
KJ243	Kan	ColE1	PYL-linker-$BS2_C$ mammalian expression plasmid	f	https://benchling.com/s/seq-ehAHfJXs6i7dGOQZuWEF

Table 3.1 – continued from previous page

KJ215	Kan	ColE1	(Bcl-2)-linker-$BS2_C$ mammalian expression plasmid	f	https://benchling.com/s/seq-lr5irLU8aaeW7U634C3R
KJ214	Kan	ColE1	(Mcl-1)-linker-$BS2_C$ mammalian expression plasmid	f	https://benchling.com/s/seq-8NIc418kHjasGUEPFIvV
KJ229	Kan	ColE1	$Nluc_{11S}$-linker-FRB mammalian expression plasmid	g	https://benchling.com/s/seq-W8cks59UP7xELm7YFOoC
KJ231	Kan	ColE1	$Nluc_{11S}$-linker-tBID mammalian expression plasmid	g	https://benchling.com/s/seq-m9YI3wtAWPrvH4jDFCO8
KJ230	Kan	ColE1	FKBP-linker-$Nluc_{114}$ mammalian expression plasmid	h	https://benchling.com/s/seq-XYXW8jXkpNhvd6r9Fve0
KJ232	Kan	ColE1	(Bcl-2)-linker-$Nluc_{114}$ mammalian expression plasmid	h	https://benchling.com/s/seq-6PsH8QWPYnbz0qYTkzoN
KJ240	Kan	ColE1	(Mcl-1)-linker-$Nluc_{114}$ mammalian expression plasmid	h	https://benchling.com/s/seq-kXMq76VD3Q7PYt3LIsui
KJ141	Carb	pUC	IgK-BS2-DAF mammalian expression plasmid	i	https://benchling.com/s/seq-P1W4xtabZVduQI07YH0C
KJ182	Carb	pUC	IgK-$BS2_N$-linker-ZA-DAF mammalian expression plasmid	j	https://benchling.com/s/seq-TECKAy2zWpm7cEUcg88P
KJ186	Carb	pUC	IgK-$BS2_N$-linker-FRB-DAF mammalian expression plasmid	j	https://benchling.com/s/seq-m0KkifQj6osisQI7K3Zo
KJ184	Carb	pUC	IgK-ZB-linker-$BS2_C$-DAF mammalian expression plasmid	k	https://benchling.com/s/seq-7IMEIMrznJqcxsiriJZW
KJ248	Carb	pUC	IgK-ZBneg-linker-$BS2_C$-DAF mammalian expression plasmid	k	https://benchling.com/s/seq-L5y9mDosE3A6xzApvMcr
KJ188	Carb	pUC	IgK-FKBP-linker-$BS2_C$-DAF mammalian expression plasmid	k	https://benchling.com/s/seq-PZyMomM636of3SVi9F6E
KJ122	Kan	pBR322	BS2 expression plasmid	l	https://benchling.com/s/seq-Cd1UfTOPGd4vJN4TMYjc

* Vector maps for each construct type shown in Figure 3.6

Figure 3.6 Vector maps for all constructs used in this work. (a-l). Vector maps corresponding to vectors listed in **Table 3.1**.

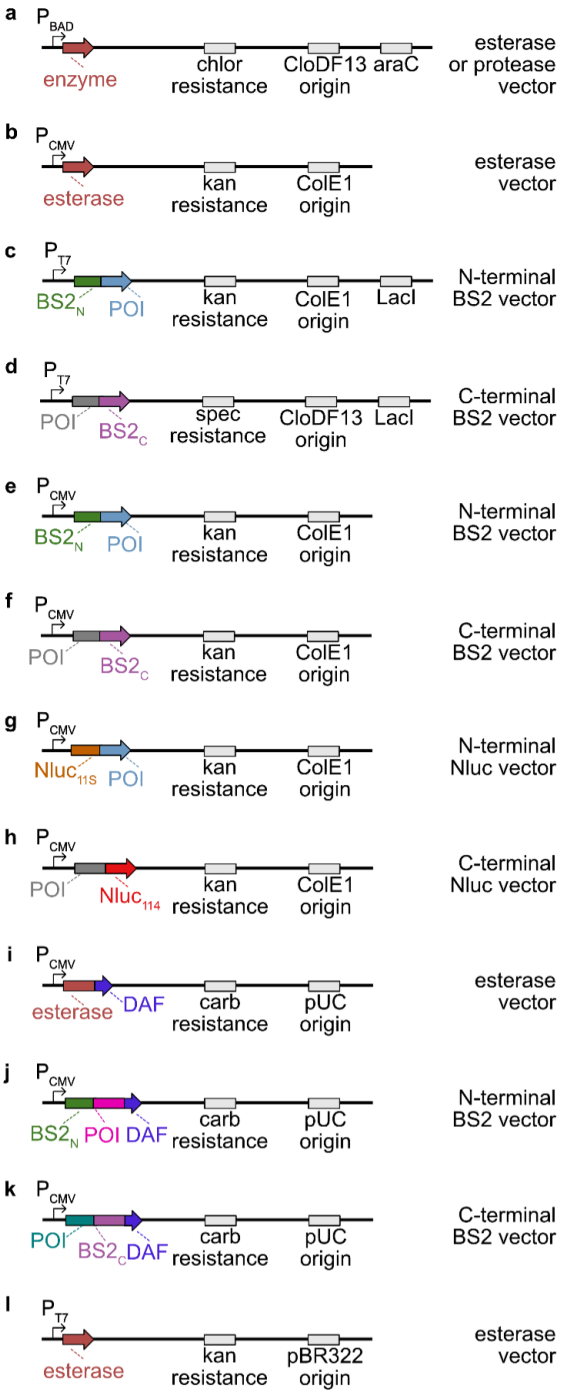


Figure 3.7 BS2 can efficiently unmask fluorescein-CM2. (a) *E. coli* transformed with a control TEV protease, PLE, or BS2 were incubated with IPTG (0.1 mM) in the absence (gray) or presence (green) of fluorescein-CM2 (50 μM) for 4 h. The cells were pelleted and resuspended in PBS prior to imaging. (b) HEK293T cells transfected with PLE or BS2 or non-transfected control cells (white) were incubated with fluorescein-CM2 (10 μM) for 1 h (green) and analyzed by plate reader. (c-d) HEK293T cells were transfected and treated with fluorescein-CM2 as in (b), loaded with Hoechst 33342 (1 μM) and analyzed by (c) high resolution or (d) low resolution fluorescence microscopy for quantification. HEK293T control cells (white) were similarly imaged. 20 μm scale bars shown. Error bars are the standard deviation for $n = 4$ replicates. Unpaired t-test; ***$P < 0.0001$.

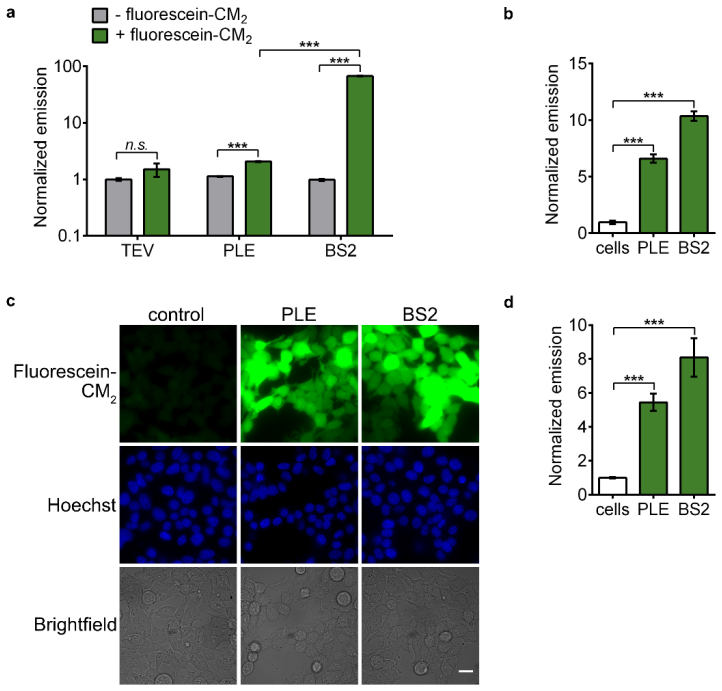


Figure 3.8 Validation of lead cut site with abscisic acid-inducible dimerization system. *E. coli* transformed with $BS2_N$-fused ABI, PYL-fused $BS2_C$ or both were cultured with IPTG (0.1 mM) in the absence (gray) or presence (green) of ABA with fluorescein-CM2 (500 μM) for 4 h. The cells were then centrifuged and the pellets were resuspended in PBS prior to imaging. Error bars are the standard deviation for n = 4 replicates.

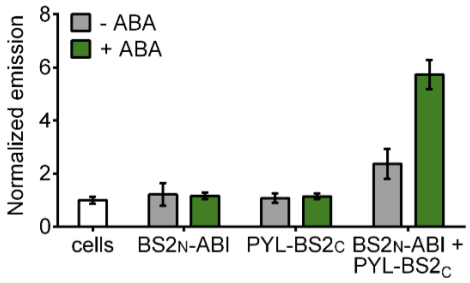


Figure 3.9 Imaging *E. coli* cell lysate from rapamycin-induced split esterase system. (a-b) *E. coli* cells cotransformed with $BS2_N$-fused FRB and FKBP-fused $BS2_C$ or control BL21 cells were induced for 16 h with IPTG (0.1 mM) in the absence and presence of rapamycin (20 μM). The cultures were centrifuged and the pellets were (a) lysed immediately after collection or (b) stored at -20 °C for 5 d prior to lysis. Fluorescein-CM2 (500 μM) was added to the lysate and incubated for 30 minutes at room temperature prior to imaging with a cell phone camera. (c-d) The lysate reaction mixtures from (a-b) were spotted on filter paper and analyzed on a gel imaging system. (c) Quantification of the images in (c). (f) Quantification of the images in (d).

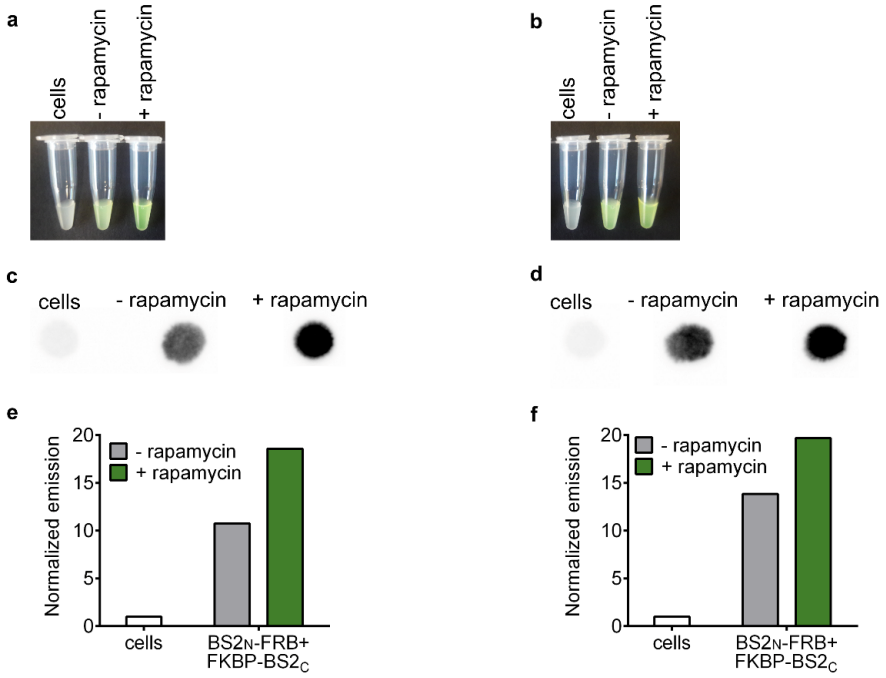


Figure 3.10 Optimization of split esterase in mammalian cells. (a) HEK293T cells were transfected with $BS2_N$-fused ZA (tan), ZB-$BS2_C$ (gray), or co-transfected with both (green) at varying ratios. The transfected cells and control HEK293T cells (white) were incubated with fluorescein-CM2 (10 μM) and analyzed for fluorescence. (b) HEK293T cells co-transfected with $BS2_N$-ZA and ZB-$BS2_C$ (4:1 ratio) or control HEK293T cells were incubated with fluorescein-CM2 (1-100 μM) and analyzed for fluorescence over time. The fold inductions in emission for co-transfected cells versus control cells are plotted. (c) Raw emission values from (b). Error bars are the standard deviation for *n* = 4 replicates.

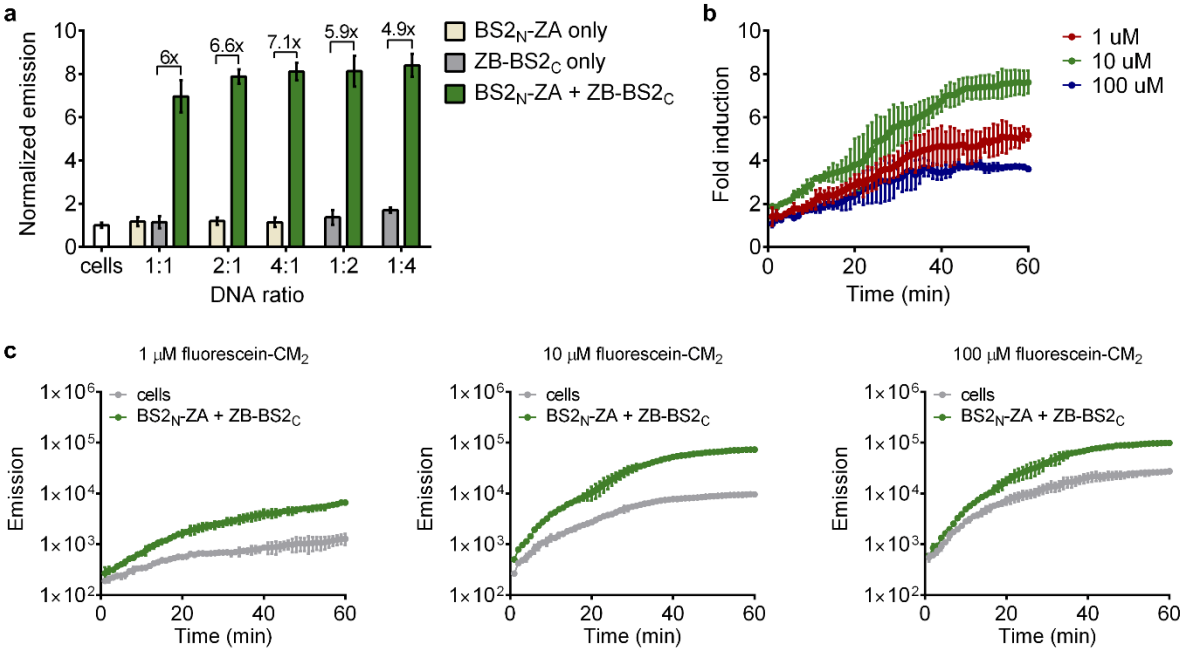


Figure 3.11 PPI-dependent esterase assembly with rapamycin-inducible dimerization system. (a) Complete imaging series from **Figure 3.2B**. HEK293T cells cotransfected with $BS2_N$- fused FRB and FKBP-fused $BS2_C$ were incubated with a DMSO carrier control or rapamycin (400 nM) for 24 h. The cells and control HEK293T cells were loaded with Hoechst 33342 (1 μM), incubated with fluorescein-CM2 (10 μM), and analyzed by high resolution fluorescence microscopy. (b) HEK293T cells cotransfected and treated identically to conditions in (a) were analyzed by low resolution fluorescence microscopy and quantified. 20 μm scale bar shown. Error bars are the standard deviation for n = 4 replicates. Unpaired t-test; ***P < 0.0001.

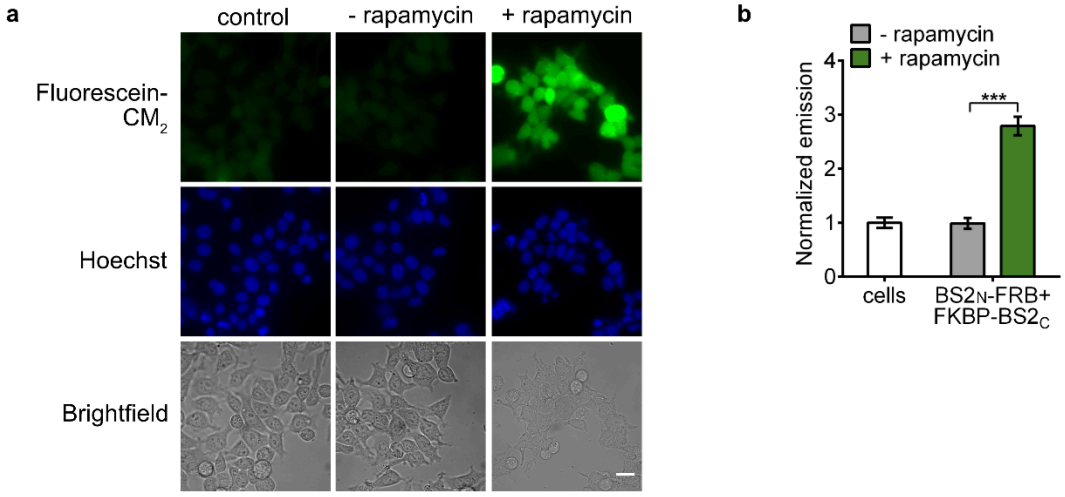


Figure 3.12 PPI-dependent esterase assembly with ABA-inducible dimerization system. (a) HEK293T cells cotransfected with $BS2_N$-fused ABI and PYL-fused $BS2_C$ were incubated with a DMSO carrier control (gray) or ABA (10 μM, green) for 24 h. Fluorescein-CM2 (10 μM) was added to the cells and control HEK293T cells (white) for 1 h and analyzed by plate reader for fluorescence. (b) HEK293T cells cotransfected and treated identically to conditions in (a) were loaded with Hoechst 33342 (1 μM) and analyzed by high resolution fluorescence microscopy. (c) HEK293T cells cotransfected and treated identically to conditions in (b) were analyzed by low resolution fluorescence microscopy and quantified. 20 μm scale bar shown. Error bars are the standard deviation for n = 4 replicates. Unpaired t-test; ***P < 0.0001.

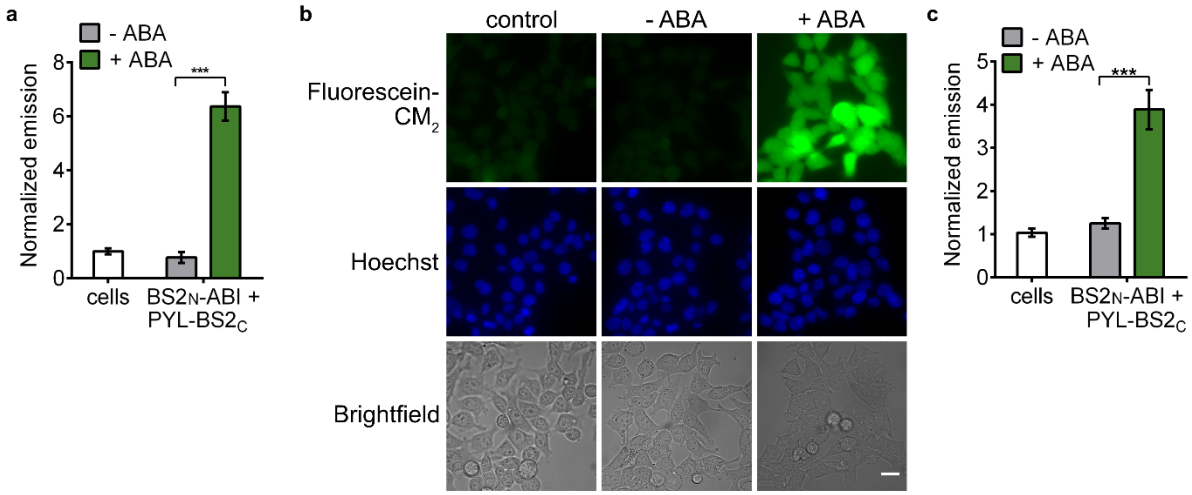


Figure 3.13 Measuring therapeutically relevant PPIs with split esterase in mammalian cells. (a) Complete imaging series from **Figure 3.3C**. HEK293T cells cotransfected with the plasmids shown in Figure 3a with $BS2_N$-fused BH3-only protein and Bcl-2-fused $BS2_C$ (top) or Mcl-1-fused $BS2_C$ (bottom). The cells and control HEK293T cells were loaded with Hoechst 33342 (1 μM), incubated with fluorescein-CM2 (10 μM), and analyzed by high resolution fluorescence microscopy. (b) HEK293T cells cotransfected and treated identically to conditions in (a) were analyzed by low resolution fluorescence microscopy and quantified. 20 μm scale bar shown. Error bars are the standard deviation for $n = 4$ replicates. Unpaired t-test; $^{**}P < 0.001$, $^{***}P < 0.0001$.

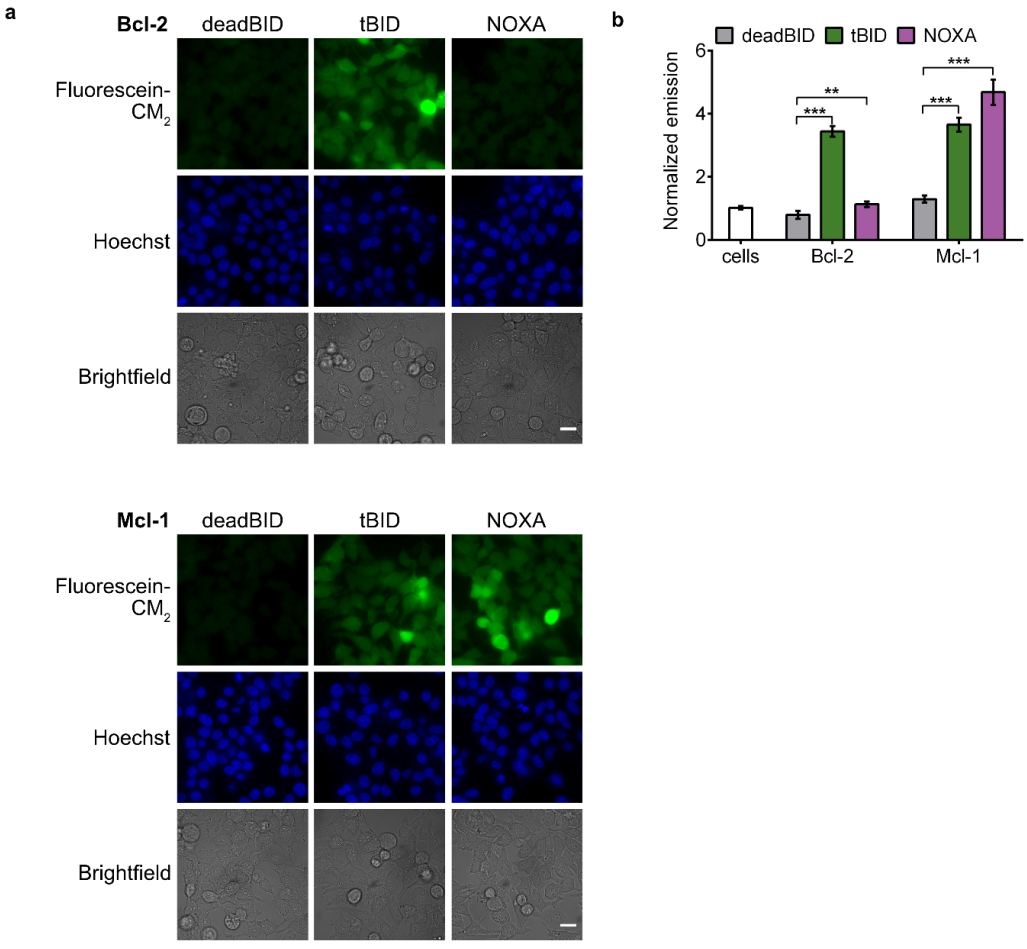


Figure 3.14 Monitoring pharmacological engagement of PPI inhibitors. (a) Complete inhibitor series analyzed by plate reader from **Figure 3.3D**. (b) HEK293T cells contransfected with $BS2_N$-fused tBID and Bcl-2-fused $BS2_C$ were incubated with a DMSO carrier (0 h) or ABT-199 (1 μM) over time. The cells and control HEK293T cells were loaded with Hoechst 33342 (1 μM), incubated with fluorescein-CM_2 (10 μM), and analyzed by high resolution fluorescence microscopy. (c) HEK293T cells cotransfected with Nluc11S-fused tBID and Bcl-2-fused Nluc114 were incubated with a DMSO carrier (0 h) or ABT-199 (1 μM) over time. The cells (orange) or control HEK293T cells (white) were incubated with furimazine and analzyed for bioluminescence. 20 μm scale bar shown. Error bars are the standard deviation for n = 4 replicates. Unpaired t-test; ***P < 0.0001.

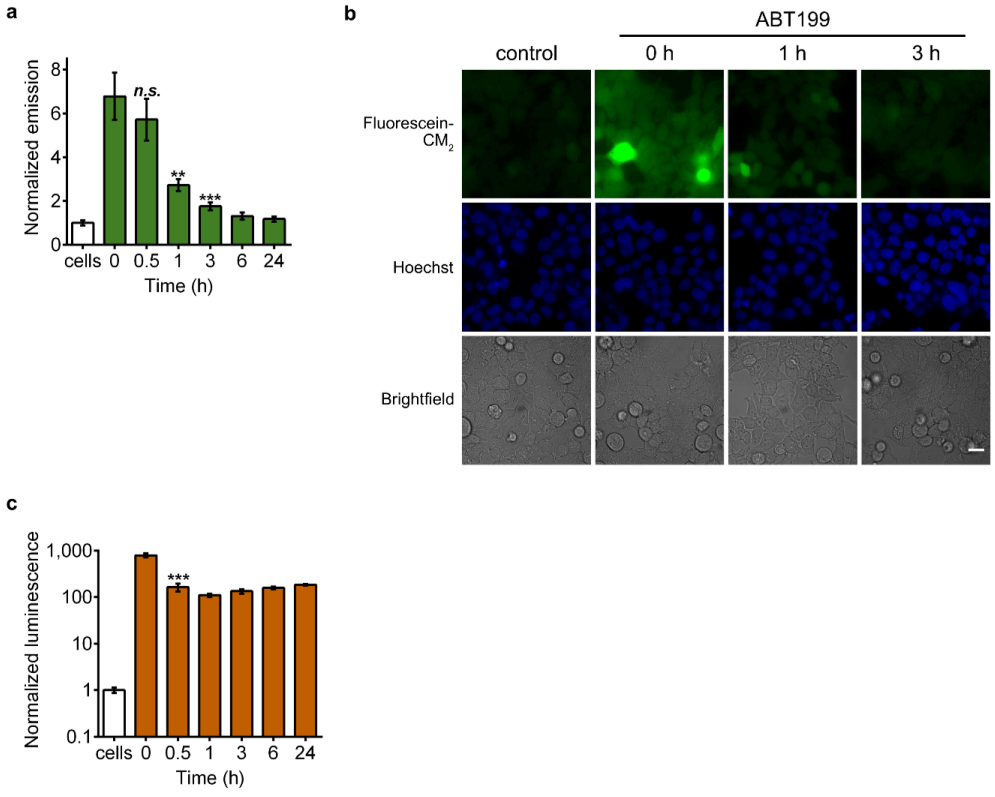


Figure 3.15 Synthetic scheme of Chemilum-CM.

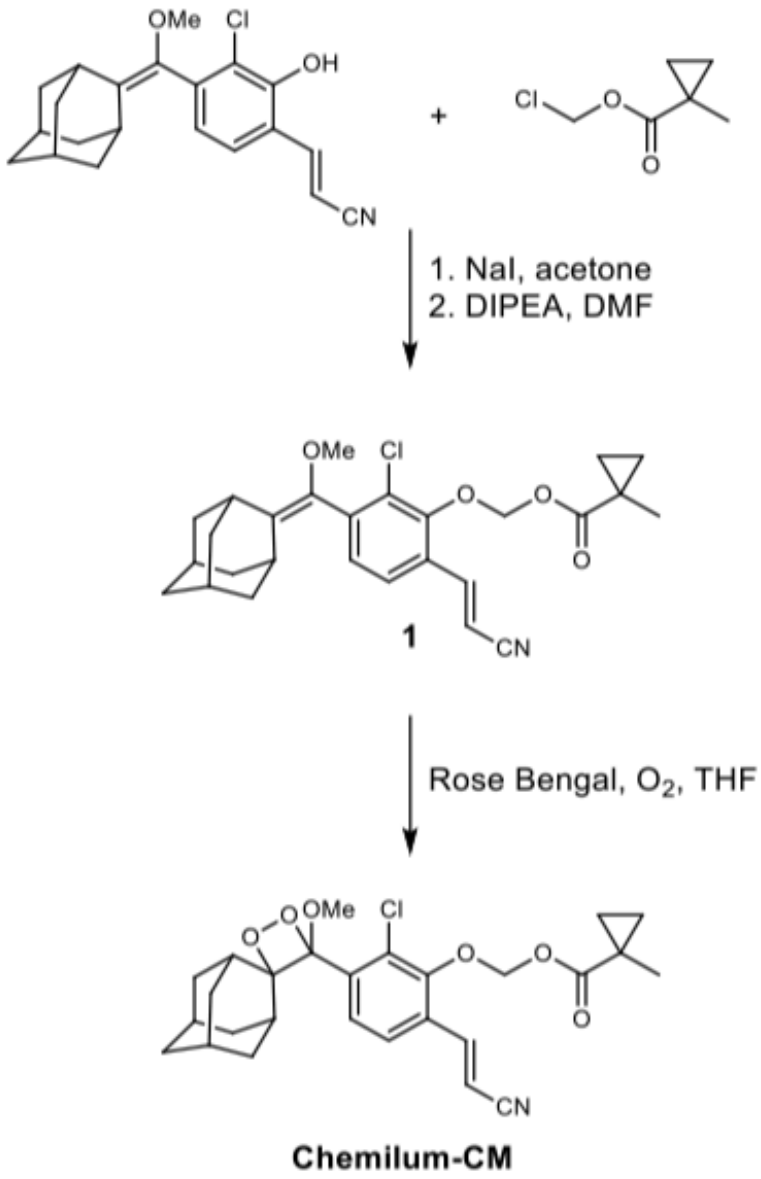


Figure 3.16 Cell surface localized esterase. (a) HEK293T cells transfected with GPI-anchored full length-BS2 esterase (blue) or HEK293T control cells (white) were incubated with Chemilum- CM2 (10 μM) and analyzed for luminescence. (b) HEK293T cells co-transfected with BS2N-fused FRB and FKBP-fused BS2C with GPI anchors were treated with a DMSO carrier control (gray) or rapamycin (blue) for 24 h. The cells or HEK293T control cells (white) were incubated with Chemilum-CM (10 μM) and analyzed for luminescence. Error bars are the standard deviation for n = 4 replicates. Unpaired t-test; *P < 0.01, **P < 0.001, ***P < 0.0001.

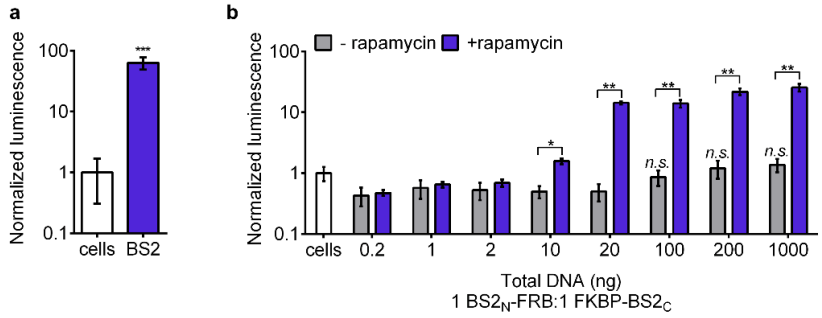


Figure 3.17 Imaging of Nanobit controls. HEK293T cells were transfected with Nluc11S-fused FRB or cotransfected with Nluc11S-fused tBID and FKBP-fused Nluc114 or Nluc11S-fused FRB and Bcl-2-fused Nluc114. The cells (gray) or HEK293T control cells (white) were incubated with furimazine and analyzed for bioluminescence. Error bars are the standard deviation for n = 4 replicates.

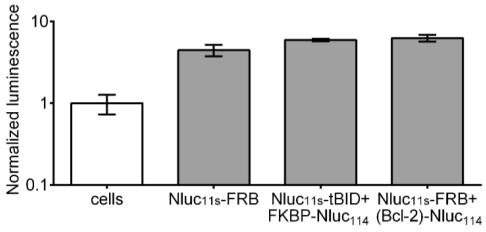


Figure 3.18 Bioconversion of SN-38-CM2 to SN-38 using BS2. Reaction of SN-38-CM2 and BS2 monitored by LC. LC traces showing elution times for bioconversion of SN-38-CM2 to SN- 38 using BS2, SN-38-CM2, and the commercially available SN-38.

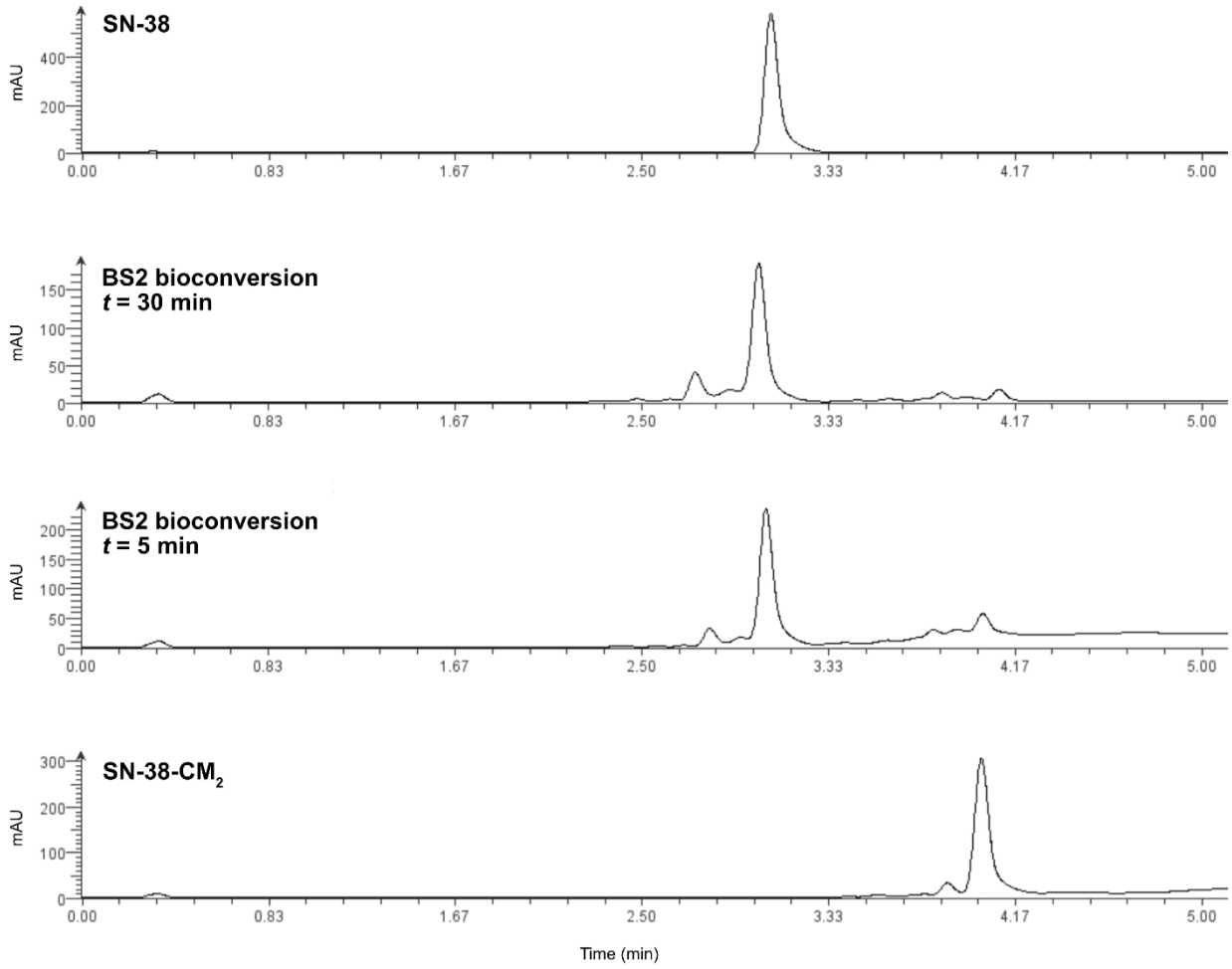


Figure 3.19 Monitoring esterase mediated cytotoxicity in mammalian cells over time. MDA-MB-231 luciferase cells were transfected with GPI anchored $BS2_N$-fused FRB and FKBP-fused $BS2_C$, cytosolic full length BS2 (purple), or GPI anchored full length BS2 (pink). $BS2_N$-fused FRB and FKBP-fused $BS2_C$ cells were treated with a DMSO carrier control (gray) or rapamycin (blue) for 12 h. The transfected cells or control cells (gray circle, dashed) were then incubated with SN-38-CM2 for (a) 1 h, (b) 3 h, or (c) 6 h. The cells were then rinsed and cultured for 40 h prior to addition of D-luciferin and imaging. Error bars are the standard deviation for n = 4 replicates. Unpaired t-test; *P < 0.01, **P < 0.001, ***P < 0.0001.

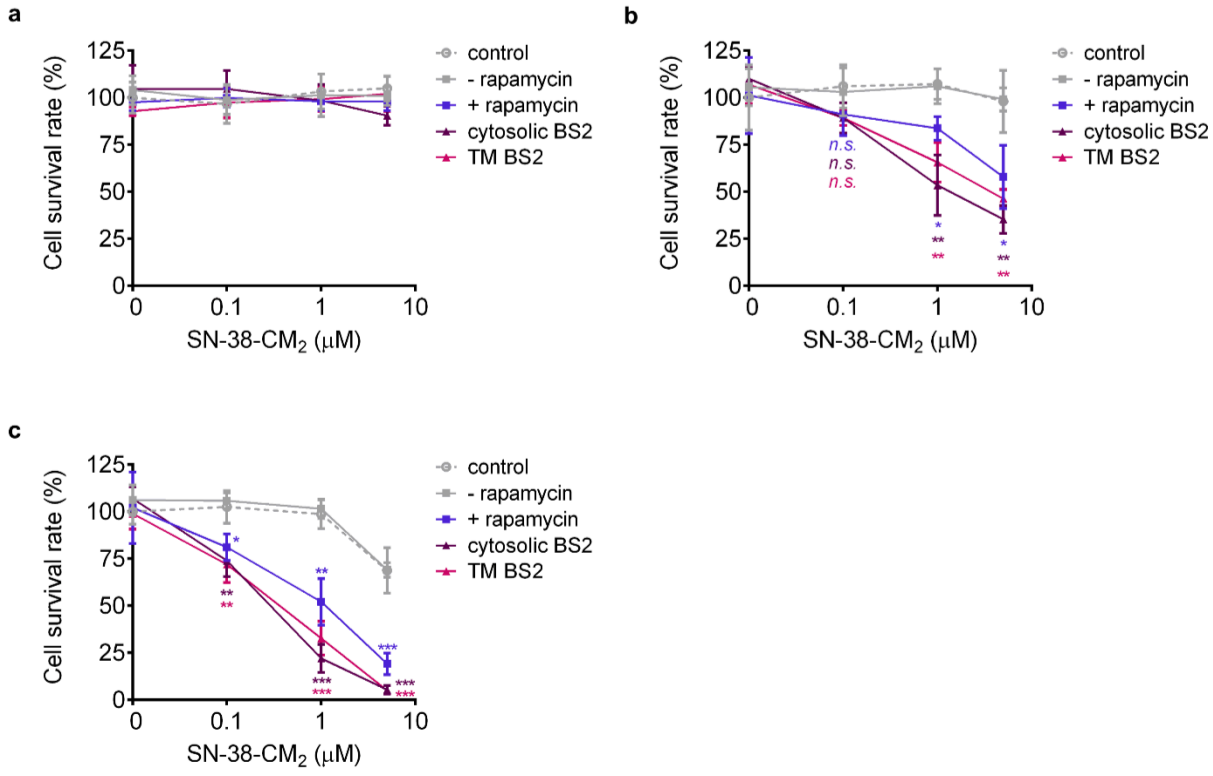


Figure 3.20 Coculture assays. (a) Complete coculture cytoxicity assays from **Figure 3.5E** and with GPI anchored BS2 expressing MDA-MB-231 cells. (b) Quantification of light emission from bioluminescence images of cocultures. Photon counts along a 2 cm line in the direction of most observed cell killing from the hydrogel embedded cells (hydrogel) to the edge of each dish are plotted. Error bars are the standard error of the mean for *n* = 8 replicates.

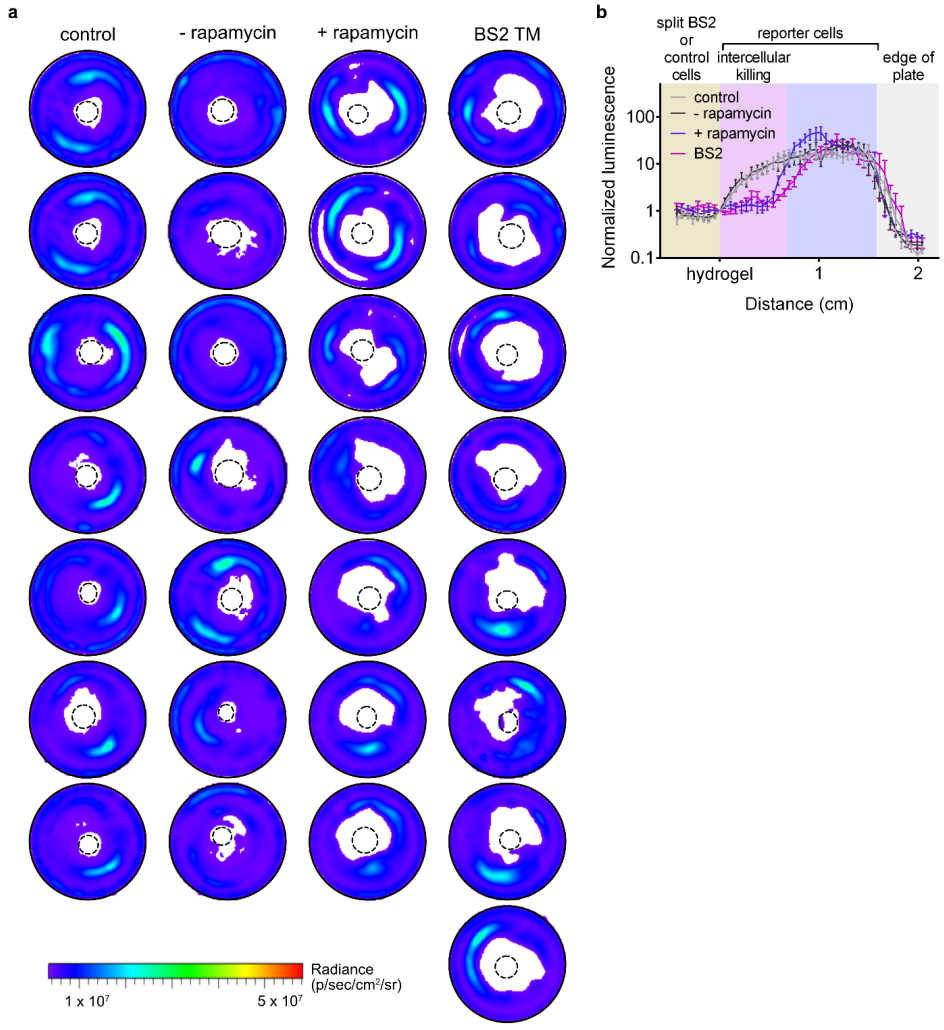


Figure 3.21 Preliminary conditions for split BS2 expression. Initial conditions tested for expression of split BS2 in *E. coli*.

	$BS2_N$	$BS2_c$
Fusion proteins	FRB, ABI +/- MBP	FKBP, PYL +/- MBP
Induction conditions	50 – 500 µM IPTG	50 – 500 µM IPTG
Expression temperature	16, 30, 37 °C	16, 30, 37 °C
Expression length	8 – 18 h	8 – 18 h
Media conditions	LB, 2xYT, Autoinduction media	LB, 2xYT, Autoinduction media

Figure 3.22 Split- vs. full length-BS2 activity with Chemilum-CM. HEK293T cells were transfected with cytosolic or GPI-anchored full length-BS2 esterase (pink) or co-transfected with cytosolic or GPI-anchored $BS2_N$-fused FRB and FKBP-fused $BS2_C$. Split esterase expressing cells were treated with a DMSO carrier control (gray) or rapamycin (blue) for 24 h. The cells or HEK293T control cells (white) were incubated with Chemilum-CM2 (10 µM) and analyzed for luminescence.

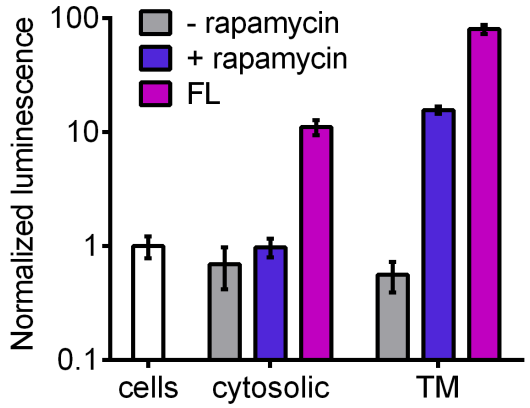


Figure 3.23 Endogenous esterase activity on CM group assessment (1). MDA-MB-231, HT-29, PC-3, and Hep G2 cells were transfected with BS2. The cells and non-transfected cells were loaded with Hoechst 33342 (1 μM), incubated in the presence and absence of fluorescein-CM2 (10 μM), and analyzed by high resolution fluorescence microscopy (left) or plate reader (right) for emission. 20 μm scale bars shown. Error bars are the standard deviation for $n = 4$ replicates. Unpaired t-test; *$P < 0.01$, **$P < 0.001$ ***$P < 0.0001$.

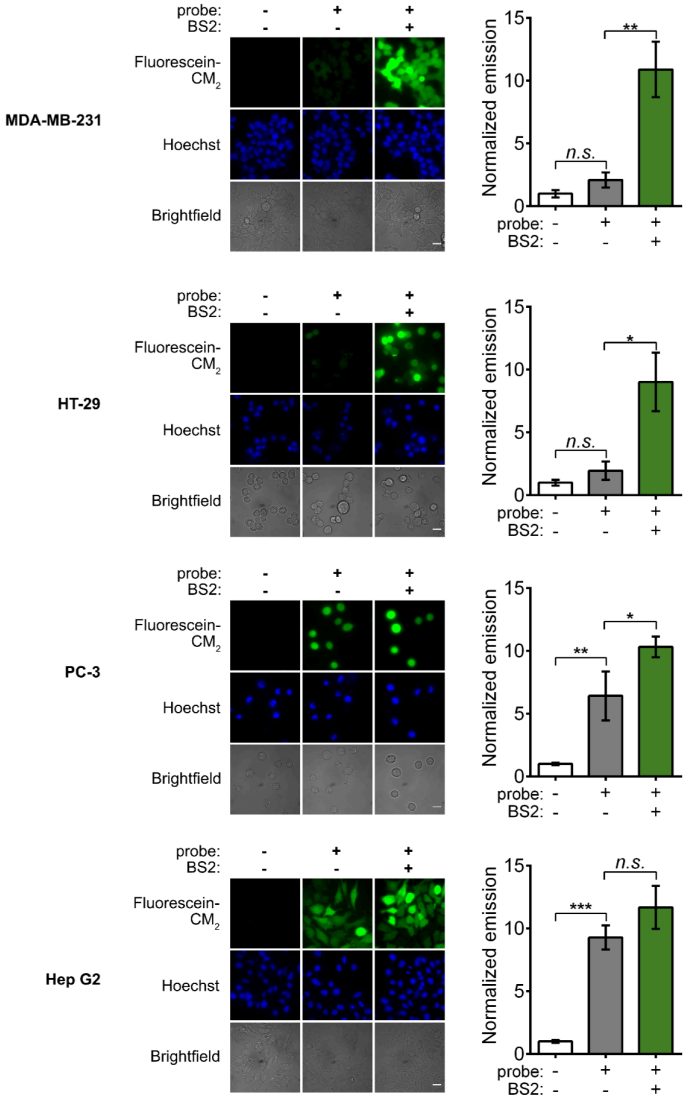
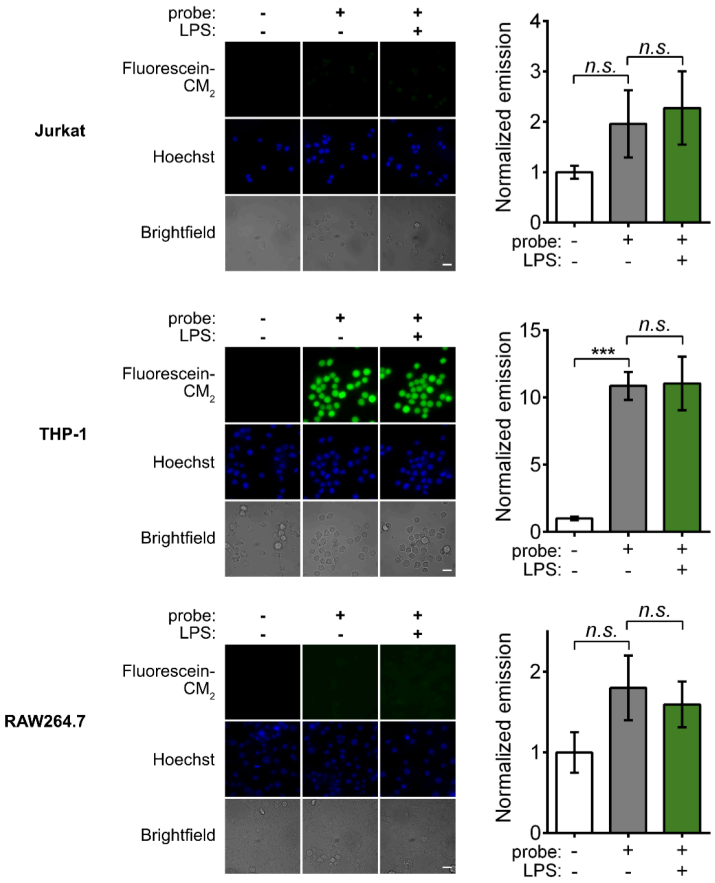
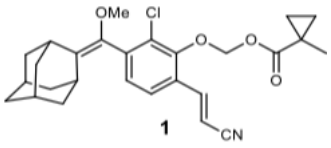
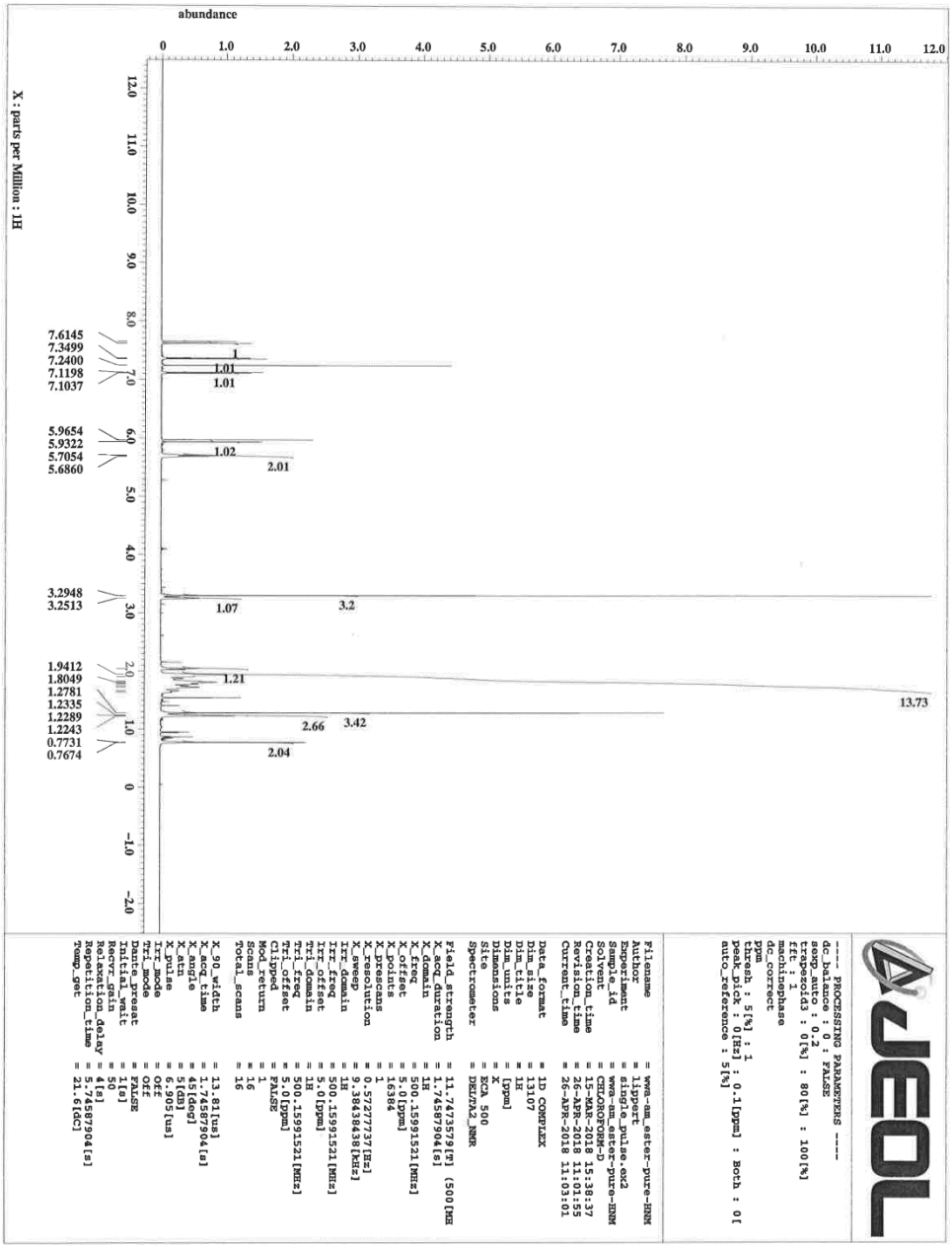


Figure 3.24 Endogenous esterase activity on CM group assessment (2). Jurkat, THP-1, and RAW264.7 cells were cultured in the presence or absence of LPS for 3-18 h. Media was replaced with Hoechst 33342 (1 μM) in the presence or absence of fluorescein-CM2 and analyzed by high resolution fluorescence microscopy (left) or plate reader (right) for emission. 20 μm scale bars shown. Error bars are the standard deviation for n = 4 replicates. Unpaired t-test; *P < 0.01, **P < 0.001 ***P < 0.0001.

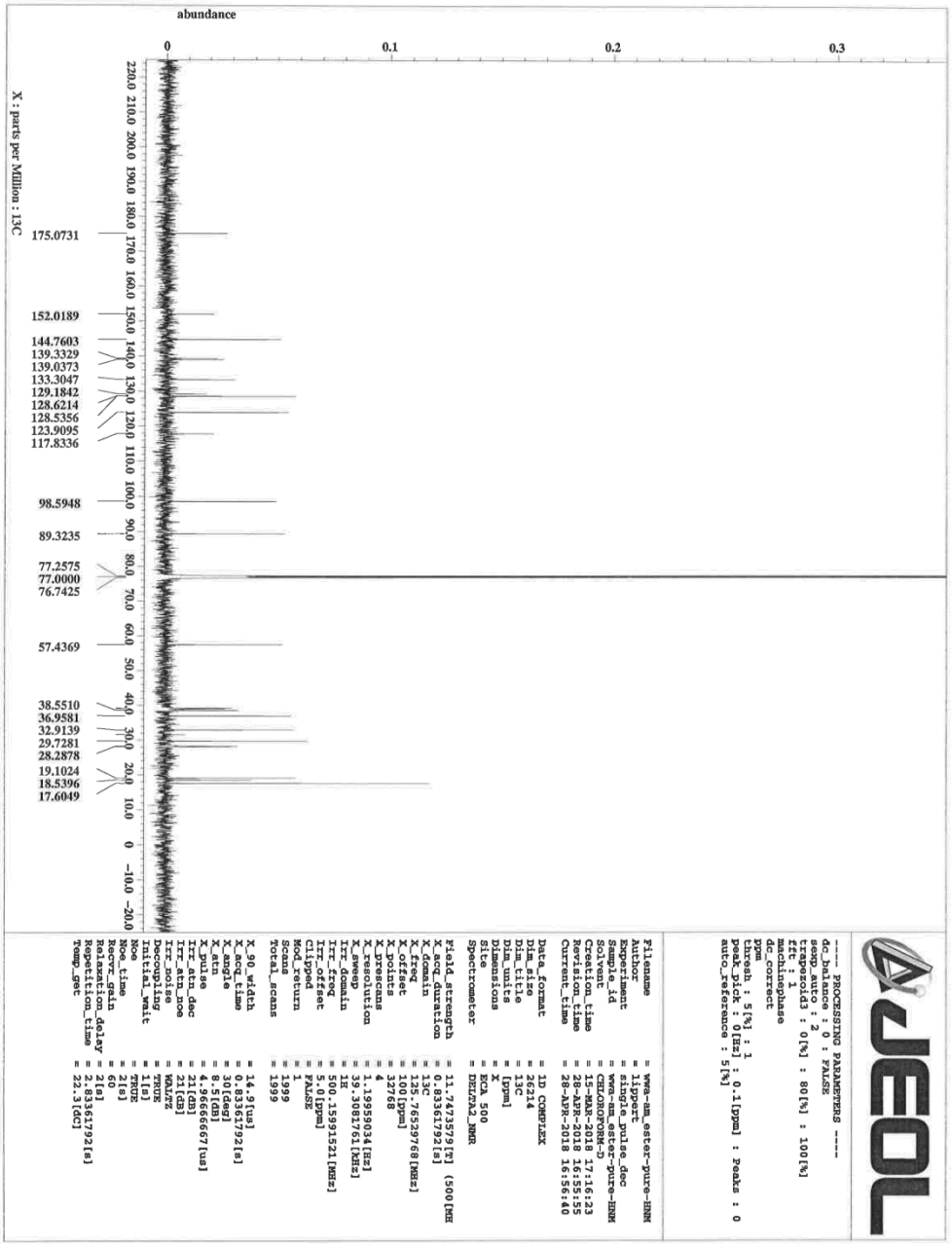


Note 3.1 Chemical characterization data for 1.



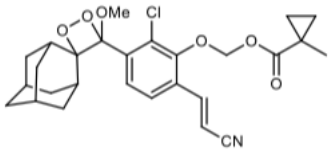


^{1}H NMR of 1.

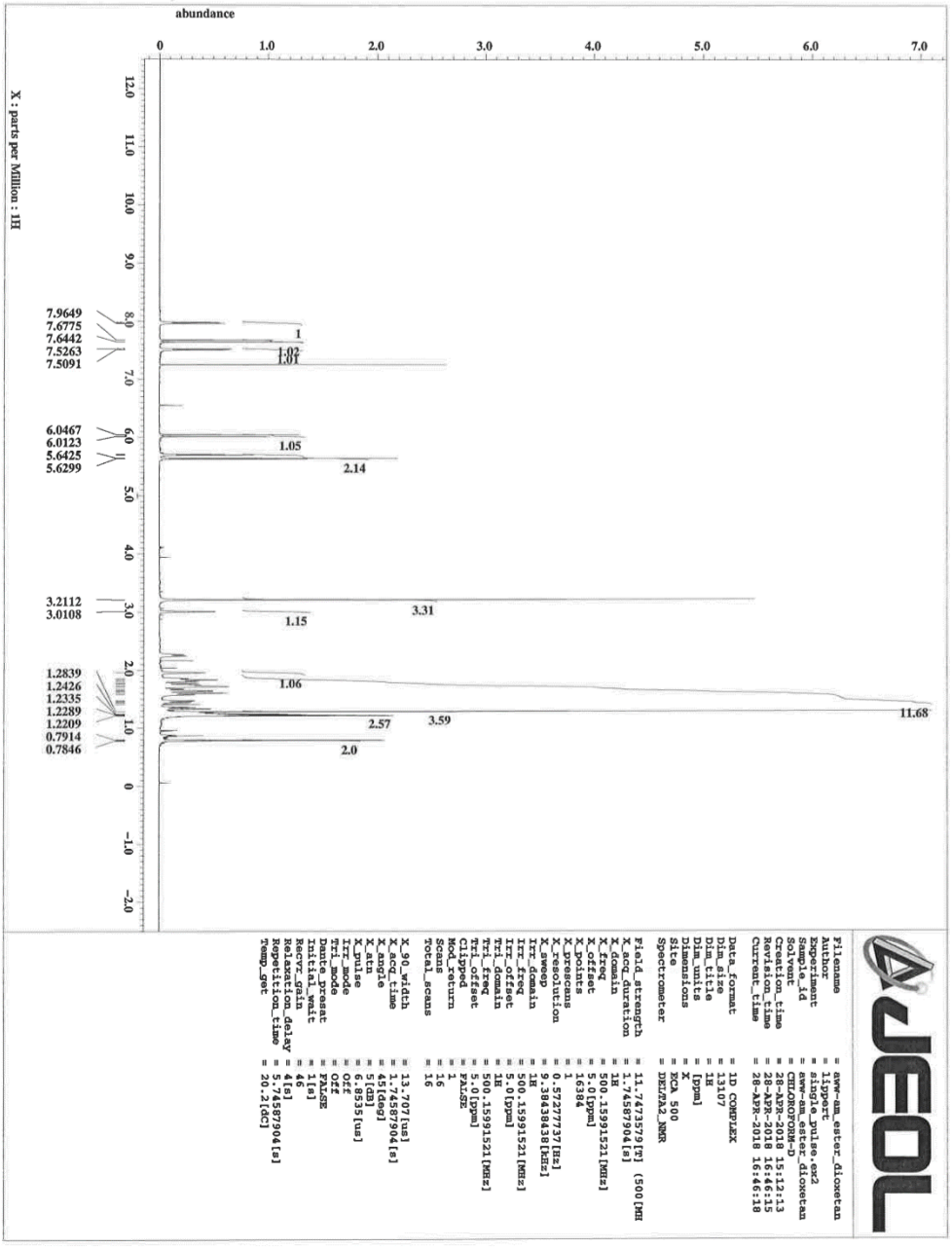

abundance
X : parts per Million : 13C
175.0731
152.0189
144.7603
139.3329
139.0373
133.3047
129.1842
128.6214
128.5356
123.9095
117.8336
98.5948
89.3235
77.2575
77.0000
76.7425
57.4369
38.5510
36.9581
32.9139
29.7281
28.2878
19.1024
18.5396
17.6049
JEOL
Filename = wwa-nm_ester-pure-DNM
Author = lippert
Experiment = single_pulse_dec
Sample_id = wwa-nm_ester-pure-DNM
Solvent = CHLOROFORM-D
Creation_time = 15-MAR-2018 17:16:23
Revision_time = 28-APR-2018 16:55:55
Current_time = 28-APR-2018 16:56:40
Data_format = 1D COMPLEX
Dim_size = 26214
Dim_title = 13C
Dim_units = [ppm]
Dimensions = X
Site = ECA 500
Spectrometer = DELTA2_NMR
Field_strength = 11.7473579[T] (500[MH
X_acq_duration = 0.833617792[s]
X_domain = 13C
X_freq = 125.7652976[MHz]
X_offset = 100.0[ppm]
X_points = 32768
X_prescans = 4
X_resolution = 1.19559034[Hz]
X_sweep = 39.3081761[kHz]
Irr_domain = 1H
Irr_freq = 500.1599152[MHz]
Irr_offset = 5.0[ppm]
Clipped = FALSE
Mod_return = 1999
Scans = 1
Total_scans = 1999
X_90_width = 14.9[us]
X_acq_time = 0.833617792[s]
X_angle = 30[deg]
X_atn = 8.5[dB]
X_pulse = 4.966666667[us]
Irr_atn_dec = 21[dB]
Irr_atn_noe = 21[dB]
Irr_noise = WALTZ
Decoupling = WALTZ
Initial_wait = 1[s]
Noe = TRUE
Noe_time = 2[s]
Recvr_gain = 60
Relaxation_delay = 2[s]
Repetition_time = 2.833617792[s]
Temp_get = 22.3[dC]
----- PROCESSING PARAMETERS -----
dc_balance : 2 : FALSE
expo_auto : 2
trapezoid3 : 0[%] : 80[%] : 100[%]
fft : 1
magnitude
auto_correct
ppm
thresh : 5[%] : 1
peak_pick : 0[Hz] : 0.1[ppm] : Peaks : 0
auto_reference : 5[%]

^{13}C NMR of 1.

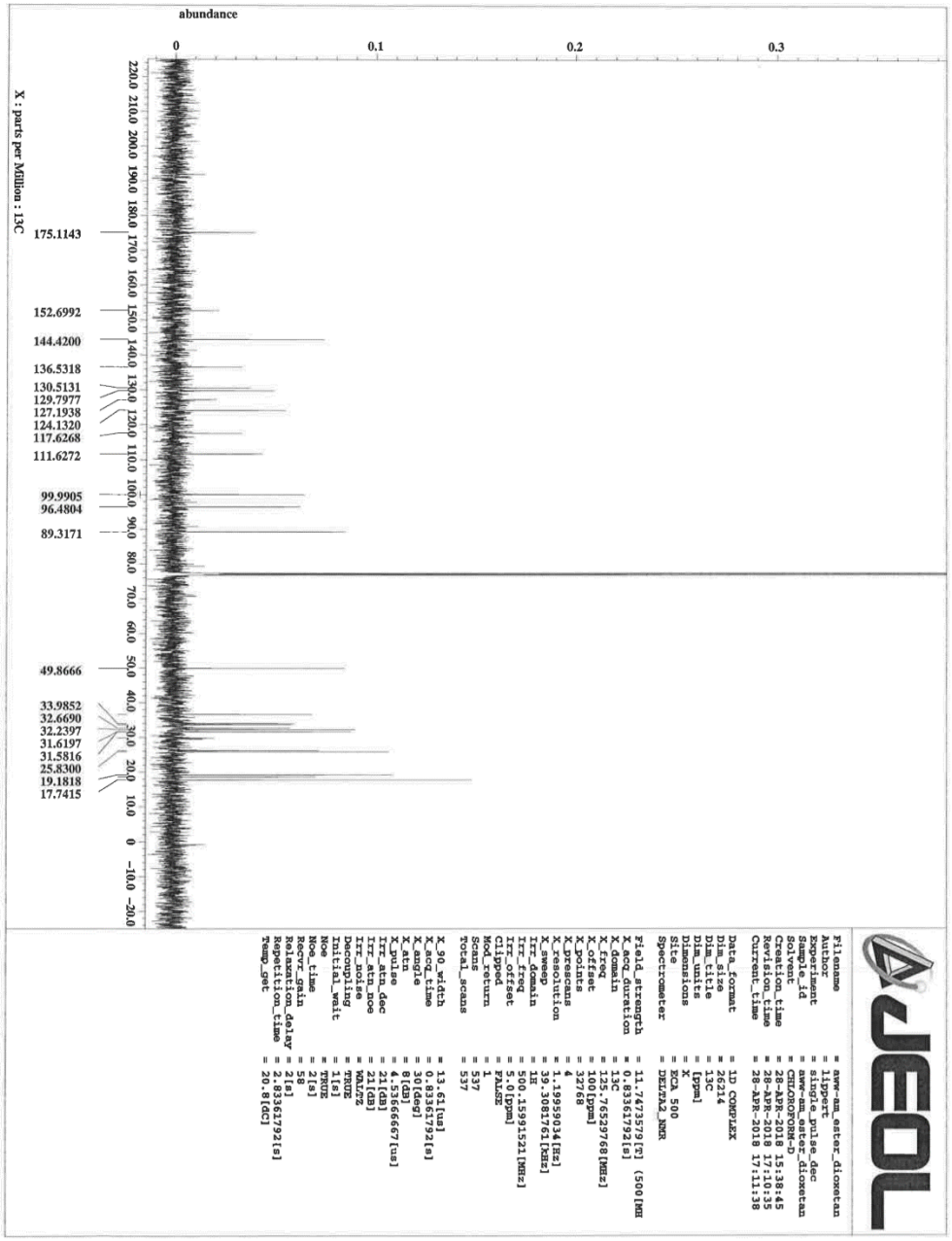
Note 3.2 Chemical characterization data for Chemilum-CM.



Chemilum-CM

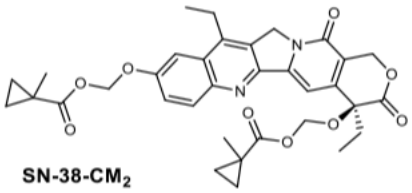


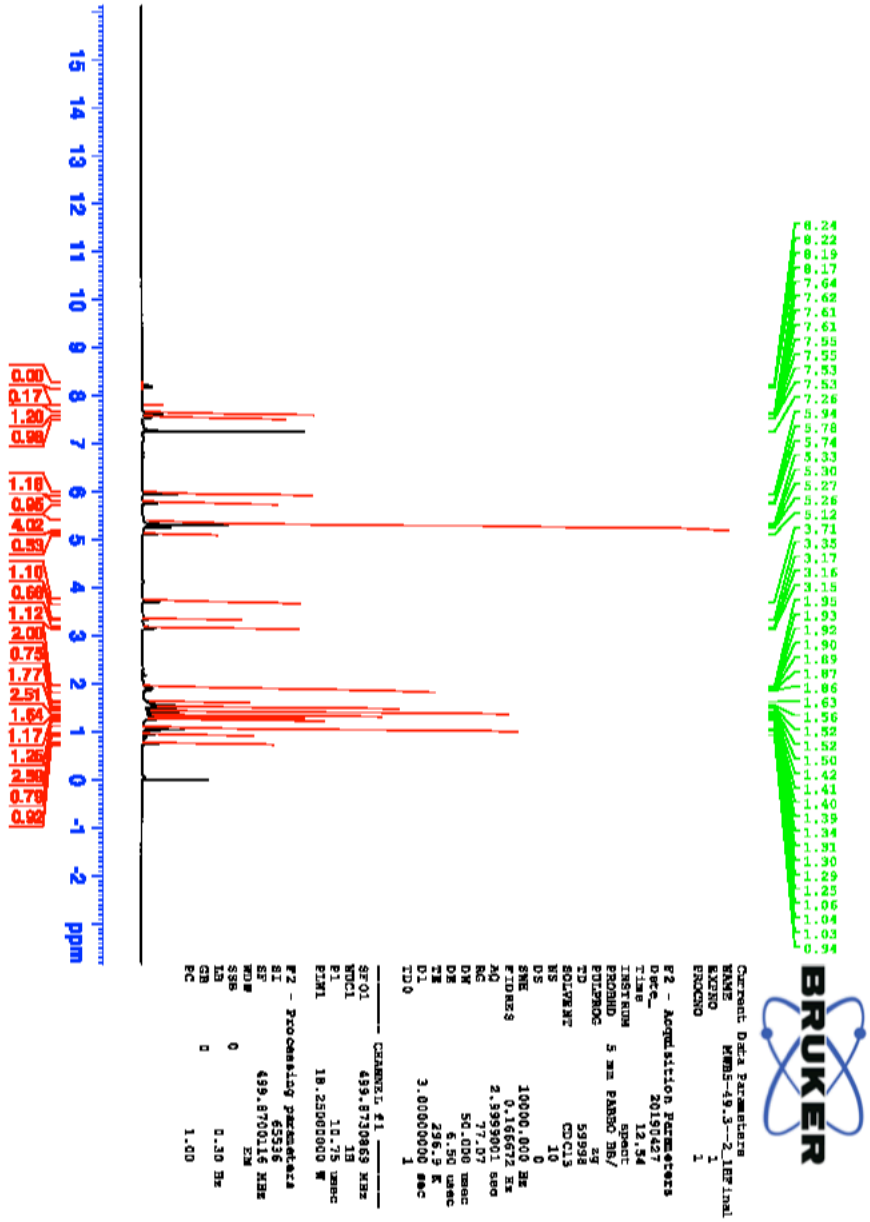
^{1}H NMR of Chemilum-CM.



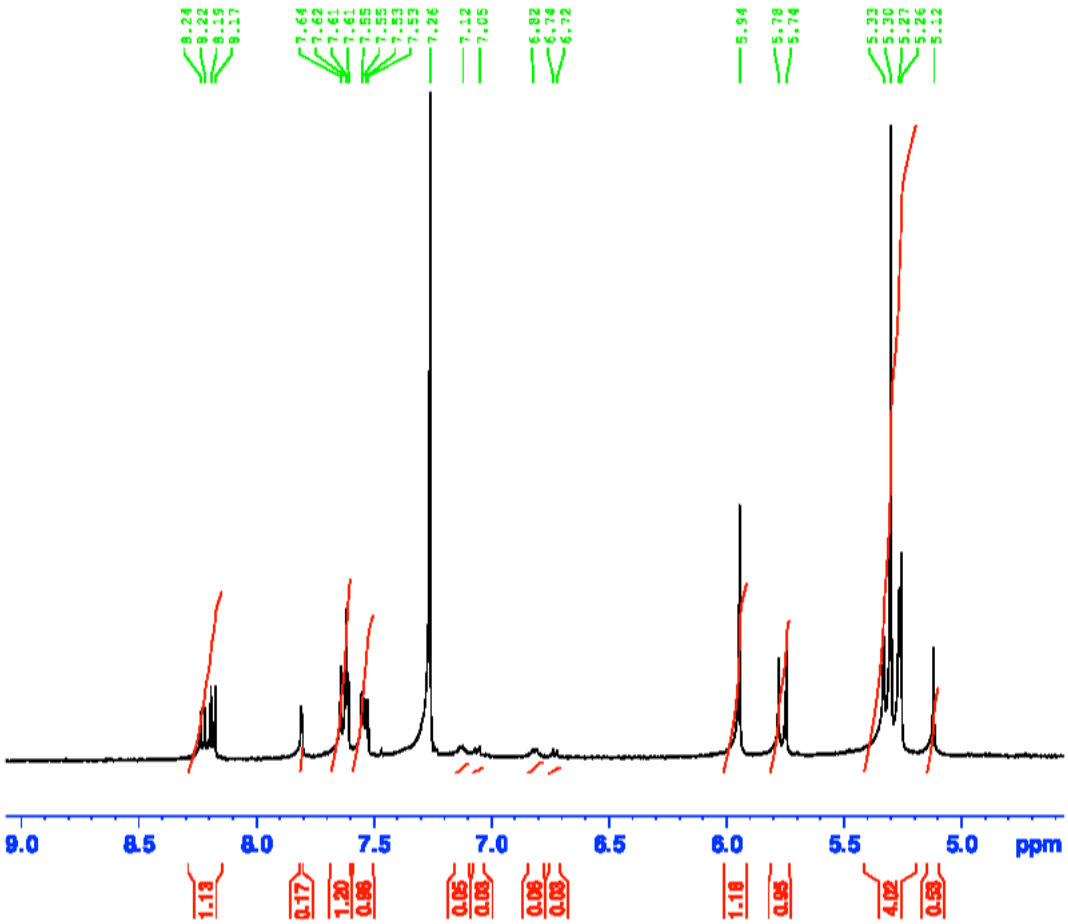
^{13}C NMR of Chemilum-CM.

Note 3.3 Chemical characterization data for SN-38-CM$_2$.



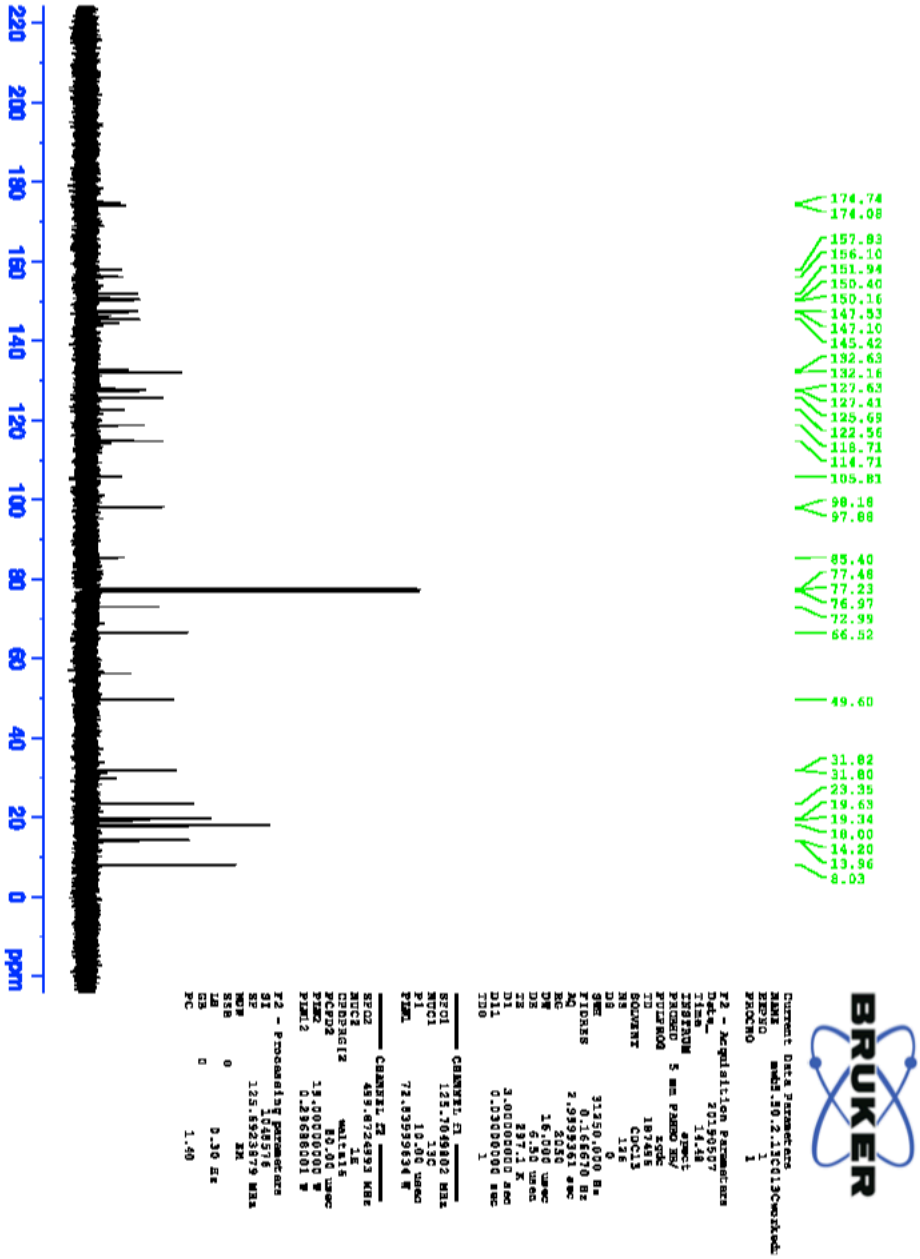


^{1}H NMR of SN-38-CM2.

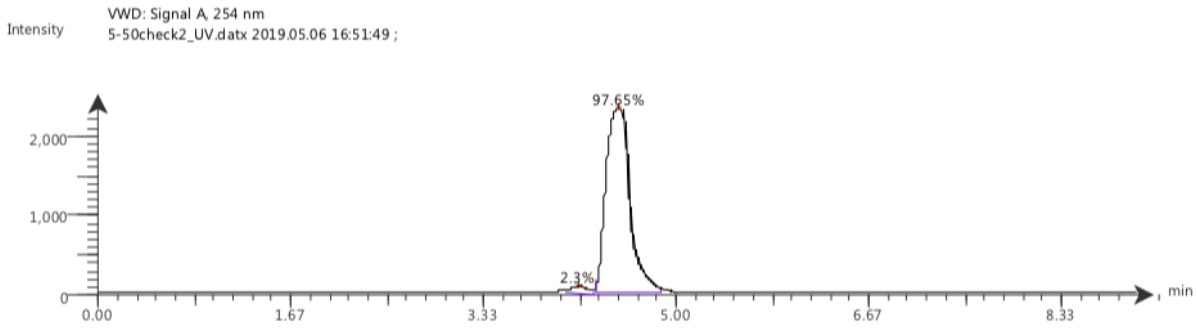


^{1}H NMR of Aromatic Region of SN-38-CM2.

Note 3.3 – continued from previous page



^{13}C NMR of SN-38-CM2.



LC UV (254 nm) Trace of SN-38-CM2.

CHAPTER 4

DEVELOPMENT OF AN RNA PROXIMITY LABELING METHOD THAT DEPLOYS A UNIQUE ESTER-MASKED ENOL ESTER ACYLATING AGENT

This chapter is reproduced and adapted from the manuscript in progress: Kentala, K.;* Qiu, T.;* Dickinson, B. C. Development of an RNA proximity labeling method that deploys a unique ester-masked enol ester acylating agent. *Manuscript in progress.*

4.1 Abstract

Transient interactions involving proteins and nucleic acids underpin most cellular processes, and yet have been historically challenging to interrogate in the absence of tools that can capture them. As such, methods for determining the subcellular organization and localization of these biopolymers have exploded in scope and demand. Seeking to add to the growing RNA proximity labeling toolbox and doing so in a way that harnesses the unique reactivity of RNA, our method takes advantage of well-established RNA 2'-OH reactivity toward acylating agents. In this work, we present BS2-seq, a genetically encodable RNA proximity labeling strategy that capitalizes on the unique design and reactivity of a novel ester-masked enol thioester acylating reagent 0106. We validate 0106's RNA labeling capacity *in vitro* and demonstrate its utility in labeling RNA in a proximity dependent manner within subcellular organelles. We further challenge the system by labeling transcripts in sub-organellar space. Fusion of BS2 to RNA binding proteins (RBPs) recapitulates known RNA-protein interactions (RPIs) and opens the door for the discovery of new ones. Collectively, this work presents a promising new RNA proximity labeling approach that is complementary to those currently available, and it

should be broadly applicable in the study of the spatial transcriptome and RNA-protein interactomes.

4.2 Introduction

Our understanding of the role RNA plays in our cells is continuously evolving. RNA transfers important molecular information not only through the bases it encodes, but through its own complex asymmetric subcellular organization as well as the secondary structure it takes. Dynamic RNA-protein interactions (RPIs) are responsible for the spatial translation of proteins as well as whether the RNA itself will be stored, processed, degraded, or spliced, all of which can inform critical cellular function and organismal development.[145-147] As such, there is considerable interest in the development of tools that can interrogate and map the spatial transcriptome.[148] A few RNA proximity labeling methods have been recently developed to tackle this challenge. APEX-seq deploys an organelle-localized or genetically-fused ascorbate peroxidase mutant (APEX2) to regions and proteins of interest, respectively, and labels the guanine base of proximal RNAs with a H_2O_2-mediated biotin-containing phenoxyl radical which can subsequently be captured and sequenced. This method has found broad utility in assessing RNA distribution across membrane-bound organelles, membrane-less organelles such as stress granules, and most recently in interrogating RNAs involved in RPIs as well as membrane-associated transcripts.[149-152]

Despite the numerous advances made by APEX-seq, there remain some disadvantages inherent to the method; the requirement of toxic H_2O_2 in achieving the reaction precludes its utility *in vivo*, and induces cellular stress, making perturbative

studies difficult to deconvolute. To this end, the light-inducible CAP-seq and Halo-seq methods bypass the need for toxic peroxide by utilizing photosensitizers.[153,154] Other work has been done to improve APEX-seq by optimizing the reactivity of the phenoxyl radical-generating reagent itself, through increased functionalization of the aromatic ring.[155]

Because the above methods are based on radical chemistry, they label only the guanine bases present in RNA. This imparts an inherent bias and may lead to decreased labeling efficiency of some transcripts. Additionally, while these methods have paved the way, they have ultimately arisen as protein proximity labeling-optimized systems that have been applied to RNA proximity labeling out of convenience due to the guanine base's ability to react similarly, and do not take advantage of the unique chemistries in which RNA can participate. Interested, therefore, in tackling the challenge of proximity labeling a different way, we considered what other types of molecules have already been used to interrogate RNA through reactivity, and we landed on RNA SHAPE chemistry. Selective 2'-hydroxyl acylation reagents have been well established to acylate any 2'-OH on a flexible RNA nucleoside while being unable to access those encumbered by engagement in more rigid structural- or base pairing-interactions.[156,157] Since the 2'-OH group is a universal component of RNA, designing a proximity labeling reagent that labels it could in principle decrease labeling bias as well as offer richer sequencing information based not only on location or RPI engagement, but also on secondary structure.

Recently, and described in **Chapter 3** of this thesis, our group developed a genetically-encodable protein-fragment complementation technology based on a split esterase.[158] We showed that BS2 esterase was capable of bioorthogonally unmasking

methylcyclopropyl ester-containing molecules in multiple cell lines, and this work was chiefly informed by the notion that many different molecules---from imaging reagents to bioactives---can be masked with esters such that activity requires an unmasking event. Seeking to continue to leverage this principle, we wondered whether we could mask and unmask an RNA acylation reagent with our selective ester-esterase pair, and whether the system could function in a proximity-dependent manner.

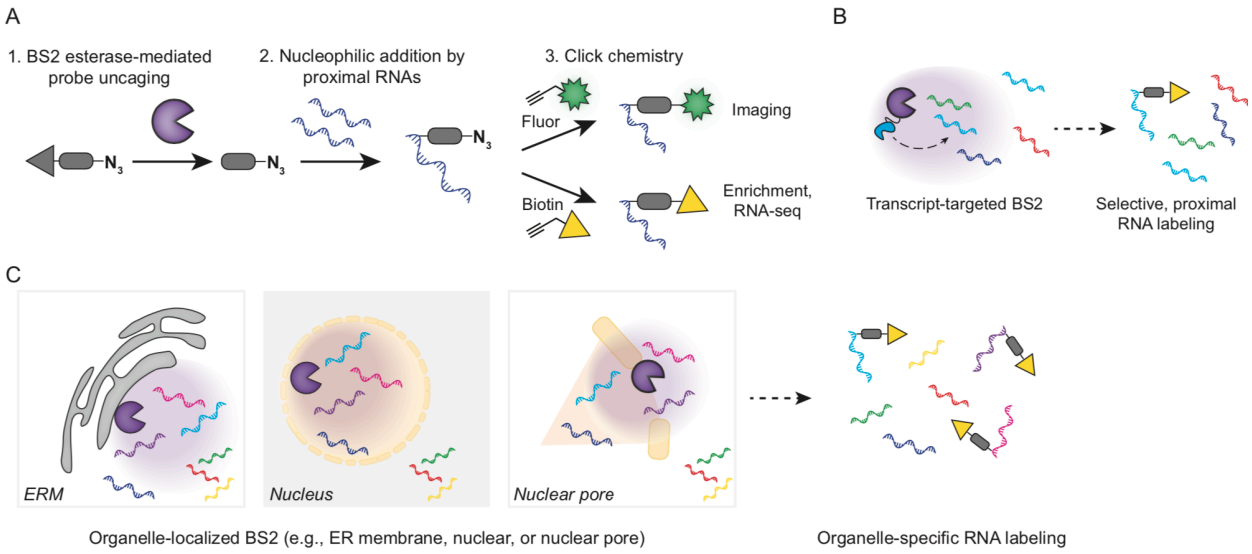


Figure 4.1 BS2-seq method overview. (a) BS2 esterase unmasks an acylating probe which is nucleophilically attacked by proximal RNAs. (b) BS2 is fused to RNA binding proteins and selectively labels bound RNAs. (c) BS2 is localized to various organelles and selectively labels organelle-associated transcripts.

In this work, we develop BS2-seq (**Figure 4.1**), an RNA proximity labeling method mediated by an azide-functionalized, ester-masked enol thioester acylating reagent, 0106, and a paired BS2 esterase. We validate 0106's RNA labeling ability *in vitro* and leverage 0106's functional click chemistry handle in multiple ways, demonstrating its versatility: fluorophores are clicked on in cells to confirm BS2 organelle locations and demonstrate labeling localization, or labeled RNAs are instead functionalized with biotin

for subsequent enrichment and analysis. RT-qPCR experiments on enriched transcripts confirm that BS2/0106-mediated RNA labeling occurs in a proximity dependent fashion in cells only when all components are present. We go on to perform high throughput RNA sequencing experiments in which we localize BS2 to various compartments and recapitulate expected transcript localizations. We fuse BS2 to RBPs and confirm known RPIs, and fascinatingly determine that when assessing proximity labeling in the subnuclear space via fusions to a sub-nuclear organellar protein, fusion to a nuclear RBP, and fusion to a nuclear histone acetyltransferase, each enrichment dataset is almost entirely unique, clearly demonstrating the power of this proximity labeling approach and opening the door to discover new biology. Altogether, these results establish BS2-seq as a viable new tool that should complement the current suite of proximity labeling strategies available.

4.3 Results

4.3.1 Design and synthesis of 0106 masked acylating reagent

In considering how to design a masked RNA acylating reagent, it is helpful to first break them down into their functional parts (**Figure 4.2**): NAI for example, contains an imidazole leaving group that causes it to be highly susceptible to nucleophilic displacement upon attack at the electrophilic carbonyl center. With the development of icSHAPE, Chang et al. add an azide functional handle to NAI, which enables enrichment-based *in vivo* SHAPE.[159,160] With this basic design in mind, and given that our masking group is an ester, we considered how we might marry the two. While we cannot mask a carbonyl with an ester, we can mask an alcohol with one, so we decided to exploit keto-

enol tautomerism. Masking the highly disfavored enol form of an RNA acylating reagent should in principle lead to rapid conversion to the keto form upon an unmasking event.

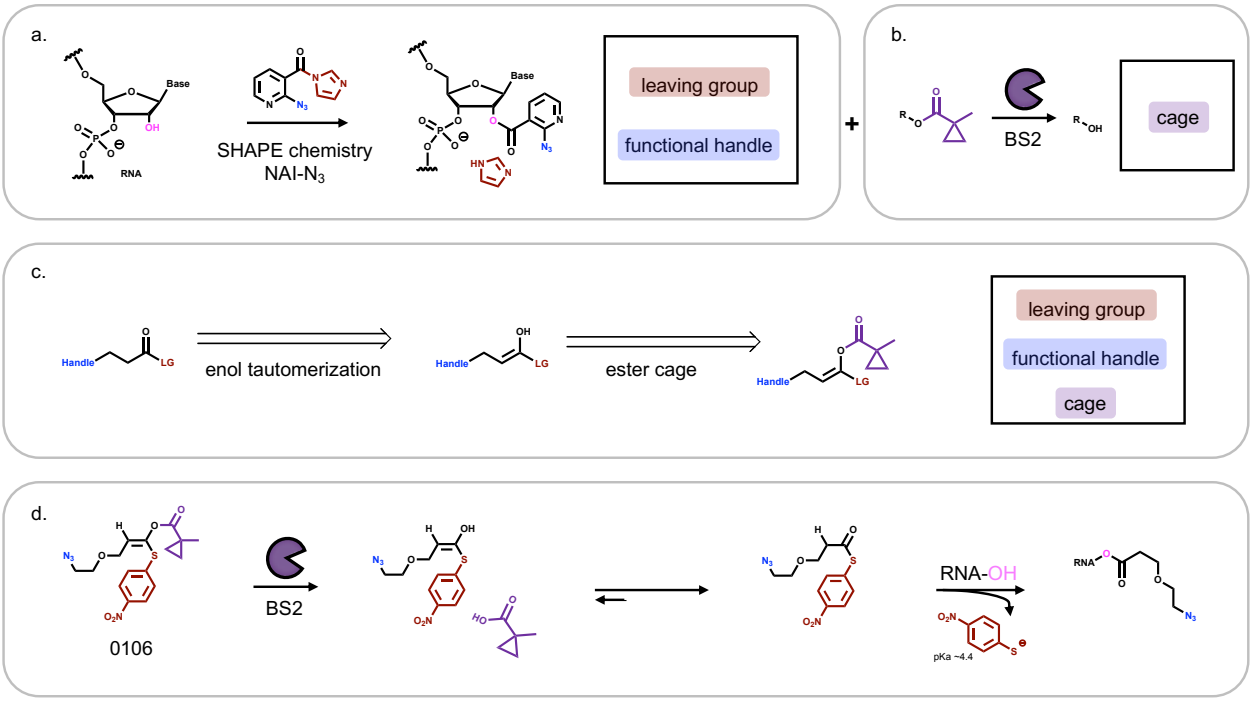


Figure 4.2 RNA acylation probe 0106: inspiration and design. (a) Functionalizable RNA 2-OH acylation reagents contain two components: a good leaving group (brown) and a handle (blue). (b) BS2 esterase chemistry and its potential as a caging group (purple). (c) Masking the enol tautomer of a ketone-containing molecule with the BS2 esterase-selective cage. (d) 0106 lead probe labeling of RNA.

Through the diligent work of my collaborator, Tian Qiu, who screened multiple reaction conditions and synthesized several probe iterations, we ultimately arrived at 0106 as our current lead probe design (**Figure 4.2D**). Unmasking with BS2 liberates a highly electrophilic thioester, which can then in principle undergo nucleophilic attack by surrounding biomolecules.

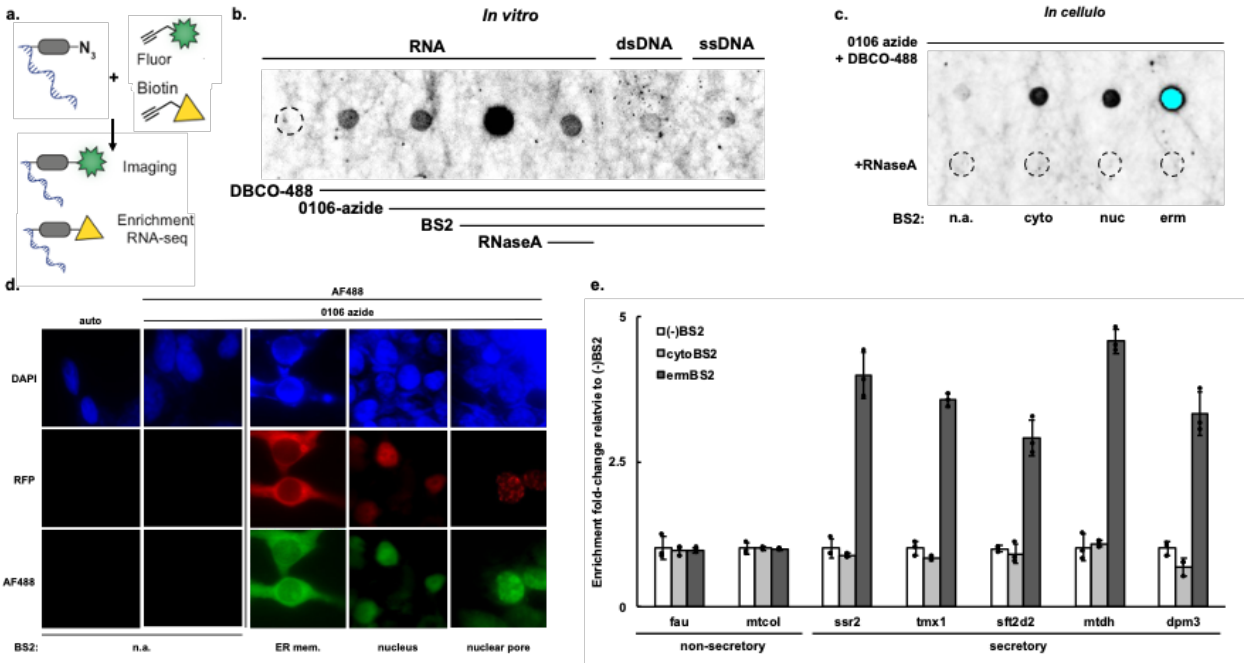
4.3.2 BS2 unmasks 0106 and labels RNA, but not DNA

Based on previous work, we have found that BS2 accepts and unmasks methylcyclopropyl ester-containing molecules robustly regardless of differences in overall

structural scaffold. We confirmed this activity, finding BS2 was capable of uncaging an acylating reagent scaffold completely in 3 minutes (**Figure 4.7**). We next performed a dot blot experiment using purified BS2 and leveraging our azide handle with an imaging reagent to determine whether the unmasked thioester electrophile would undergo a reaction with *in vitro* transcribed RNA. We found that only in the presence of 0106 and BS2 did we see robust clicked DBCO-488 signal, which was promptly abolished with the addition of RNAse A (**Figure 4.3A**). dsDNA and ssDNA corresponding to the RNA was not labeled, nor did we observe any DNAse 1 susceptibility, which we predicted based on the fact that DNA does not contain that 2-OH group (**Figure 4.3B**, **Figure 4.8**).

Encouraged by these results but recognizing that *in vitro* labeling experiments bypass any of the nucleophilic noise that would be present in an in-cell environment such as protein side chains or metabolites, we challenged our labeling system further by performing dot blots with RNA purified from cell lysates. Recombinant BS2s localized to three compartments of interest of various degrees of membrane containment (cytosol, ER membrane, and nucleus, accomplished via appended localization sequences) were transfected in HEK293T cells and 50 μM 0106 was introduced for 15 minutes. We then lysed the cells, purified out the RNA, and subjected it to a click reaction with DBCO-488. Again, excitingly, only the in the presence of both 0106 and BS2 did we observe robust labeling, which was susceptible to RNAse A degradation (**Figure 4.3C**, **Figure 4.9**). It was especially interesting to note that regardless of where BS2 was localized, strong labeling signal was observed, suggesting a tightly enclosed environment is not required for successful labeling. Unsurprisingly, while no background labeling was observed in an

in vitro setting, we did observe some background labeling in the in-cell context, which is presumably mediated by hydrolysis or weak activity by endogenous esterases.



4.3 BS2/0106 mediated labeling *in vitro* and in cells. (a) A fluorophore or biotin is clicked on to labeled RNA. (b) *In vitro* transcribed RNA, dsDNA, and ssDNA (500 ng) is subjected to BS2/0106 (50 μM) or DMSO reaction and subsequent DBCO-488 click reaction followed by RNase A or DMSO treatment and blotted. (c) RNA extracted from BS2-containing HEK293T cells is treated as in (b) and blotted. (d) Live HEK293T cells containing various RFP-tagged BS2s are subjected to 0106 labeling (50 μM) for 15 minutes and subsequently washed, fixed, and permeabilized. AF488 dye is copper clicked on and labeling events are imaged. (e) RT-qPCR analysis of secretory transcript enrichment by cytosolic and ER membrane-bound BS2s. n=3 bio. reps.

4.3.3 BS2 labels biomolecules in close proximity to its location in cells

Once we determined that BS2-mediated 0106 RNA labeling could occur in cells, we wanted to see if the labeling signal was occurring predominantly near where BS2 itself was localized. To do this, we leveraged our azide handle once again, this time through copper-mediated click chemistry of AF488 in cells. We localized BS2 to four compartments: cytosol, ER membrane, nucleus, and nuclear pore (through a SENP2 pore

protein fusion), with each containing an RFP tag. The live cells were labeled with 50 μM 0106 for 15 minutes, followed by fixation and permeabilization. The permeabilized cells were subjected to AF488 clicking and subsequently imaged (**Figure 4.3D**, **Figure 4.10**). Remarkably, there was very good co-localization between the BS2 signal and the labeling signal, even in the sub-organellar case, suggesting that proximity labeling was a viable option for deploying this technology.

We assessed the types of biomolecules being proximity labeled in cells by doing the same experiment, but instead, post-permeabilization, added Proteinase K, DNAse 1, and RNAse A on the cells and looked for degradation-based changes in the cells (**Figure 4.11**).[161] We found that the labeling signal was susceptible both to Proteinase K- and RNAse A-degradation, but not susceptible to DNAse 1 degradation. This suggests one of two things: either labeling can occur both on RNA and on proteins, or digestion of proteins engaged in interactions with labeled RNAs might cause labeled RNAs to diffuse away from the focal imaging plane, leading to decreased signal. Follow up protein labeling experiments demonstrated that while 0106 does label protein, it does not do so efficiently (**Figure 4.12**). As proteins are substantially higher in concentration in cells than RNA, less efficient protein labeling could still produce higher imaging signal.

Of additional interest to us was whether labeling was localized to the cell in which 0106 uncaging occurred only, as we would predict, or whether it could permeate nearby cells. To test this, we established a co-culture experiment consisting of mixed populations of HEK293T cells, RFP-positive cells, BS2-positive cells, and RFP-positive/BS2-positive

co-cultured cells and performed labeling (**Figure 4.13**). Gratifyingly, labeling events were constrained to the cells in which BS2 was present only.

4.3.4 Labeling of BS2-proximal RNAs assessed via RT-qPCR.

Having demonstrated both proximity labeling as well as the option to leverage 0106's azide handle for imaging applications, we next sought to utilize the handle for enrichment applications. Clicking on biotin would allow us to separate out labeled RNA populations from those that were unlabeled, enabling location-specific RNA studies. As proof-of-concept, we chose first to focus on ER membrane (cytosol-facing) ermBS2 labeling compared to general cytosolic cytoBS2 labeling. Transcripts associated with the ER membrane have been well-studied by previous methods,[149,162-165] and also allow us to test our method very stringently: since ermBS2 is cytosol-facing, comparing transcripts enriched by it vs. by general cytoBS2 should be maximally challenging. When comparing enrichment of non-secretory transcripts by cyto- and ermBS2, we observed no substantial differences, but when comparing enrichment of secretory transcripts, ermBS2 enriched them well relative to cytoBS2 in a dose-dependent manner (**Figure 4.3E**, **Figure 4.14**). We additionally wished to test our system against the APEX-seq system in this context with an added level of stringency: we looked at secretory transcript enrichment by cyto-, erm-, and nucBS2 compared to equally-localized APEX2s, wishing to assess the degrees of "off target" labeling by the nuc-enzymes, and found that, while the APEX-seq system enriches transcripts more greatly than the BS2 system, it also sees nucAPEX labeling of ER-proximal transcripts, whereas nucBS2 de-enriches ER-proximal transcripts (**Figure**

4.15). Encouraged by these results, we moved on to assessing enriched transcripts more broadly.

4.3.5 RNA-seq reveals expected organellar enrichment patterns

After assessing specific transcripts, we looked more wholistically at transcripts enriched by a given BS2 system. We chose to look first at the ER membrane and the nucleus, as these test cases have been well characterized by current methods.[149,162-165] In determining whether a given transcript was enriched specifically in a given organelle, we compared transcript abundances from erm- or nucBS2 with all requisite reaction components vs. general cytosolic BS2 with all requisite reaction components (**Figure 4.4B, C**). In this way we could assess true differences across compartments, rather than

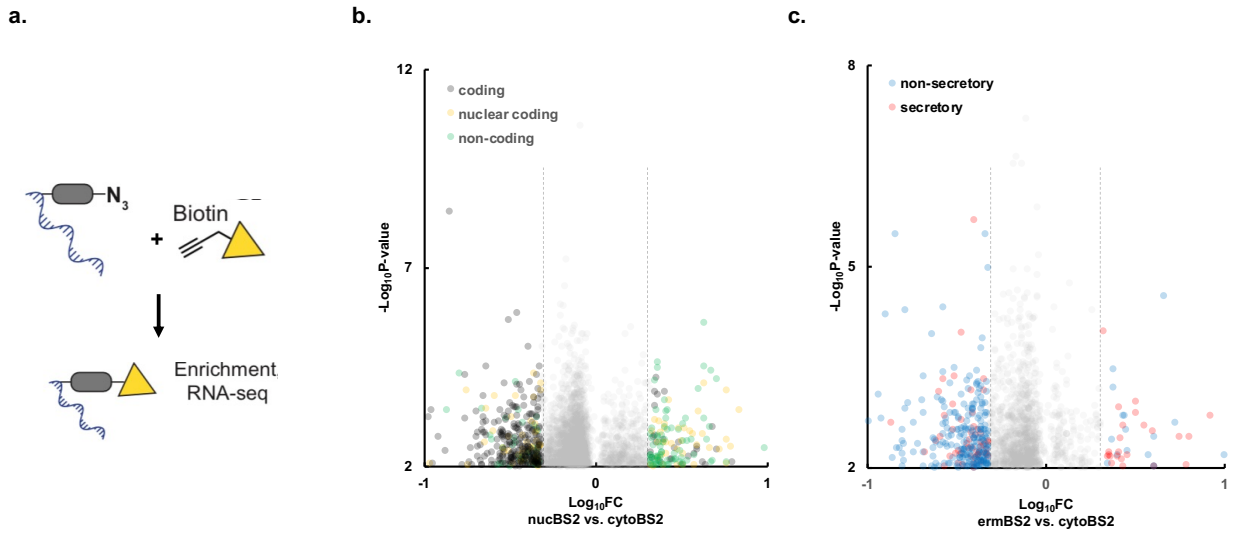
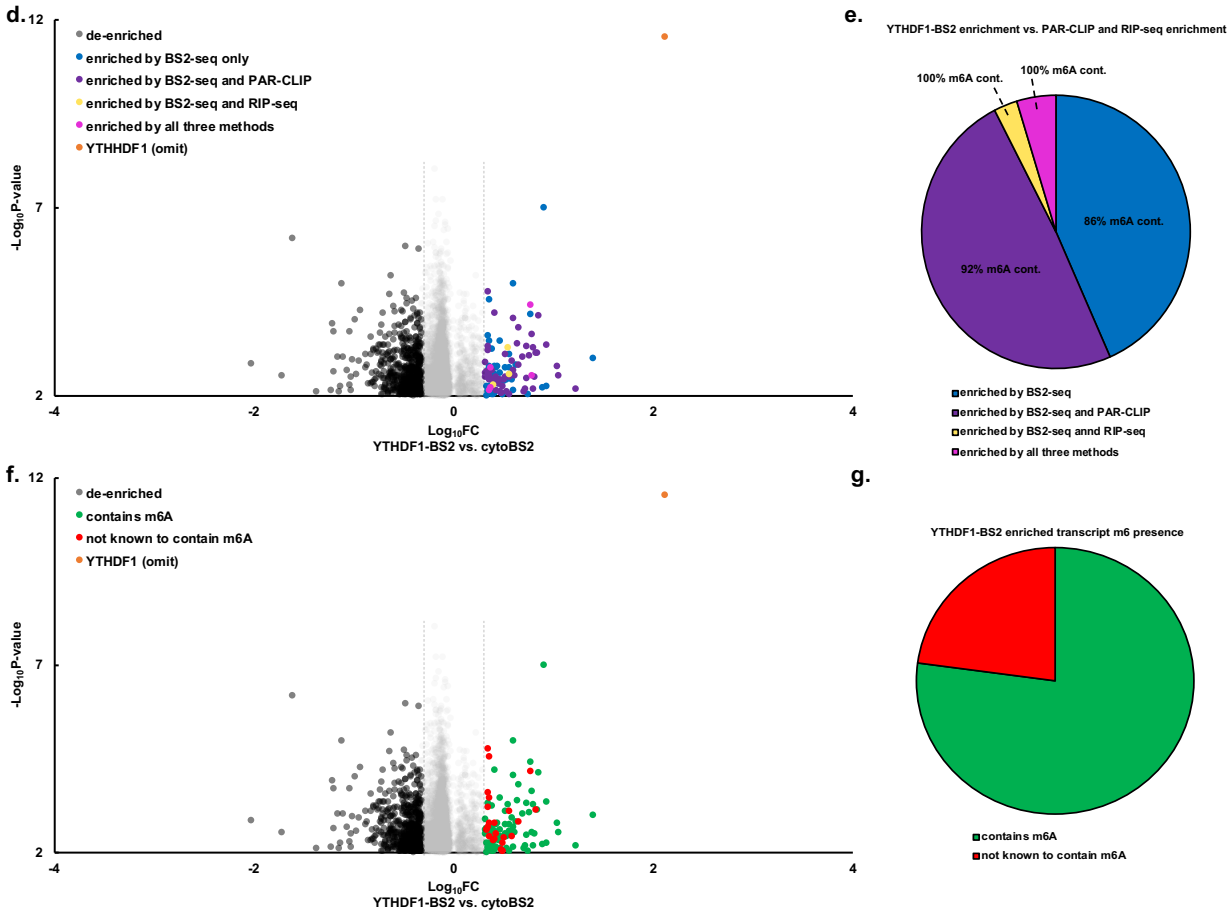


Figure 4.4 RNA-seq of organellar- and RBP-fused BS2 enriched transcripts. (a) Biotin is clicked onto labeled RNA for enrichment and RNA-seq analysis. (b) nucBS2 transcript enrichment. (c) ermBS2 transcript enrichment. (d) Transcripts enriched by YTHDF1-BS2 color-coded for enrichment with other methods (PAR-CLIP, RIP-seq). (e) Comparison of transcripts enriched by different combinations of the three methods. (f) YTHDF1-BS2 enriched transcripts color-coded for known presence of m6A. (g) Percentages of enriched transcripts that are known/not known to contain m6A. A transcript is considered enriched or de-enriched if it has a fold-change (FC) of ≥2 and is considered statistically significant if it has a p-value of ≤0.01.

Figure 4.4 – continued from previous page



simply comparing transcripts enriched in the presence vs. absence of some critical reaction component. We considered any transcript with a P-value of ≤ 0.01 as significant, while any transcript with a fold-change of ≥ 2 was considered enriched or de-enriched. Gratifyingly, in the ermBS2 vs. cytoBS2 case we observed general enrichment of secretory transcripts and general de-enrichment of non-secretory transcripts. In the case of nucBS2 vs. cytoBS2, we observed the general enrichment of non-coding transcripts and the general de-enrichment of coding transcripts. Interestingly, in tune with other current methods, we also observed enrichment of nuclear-associated protein-coding transcripts as well.[149] Given that this phenomenon has been observed with multiple

methods, it likely bears future investigation: is translation occurring in the nucleus? Or, more likely, is translational machinery co-localizing at nuclear pores to translate nuclear-associated proteins locally?

4.3.6 BS2-YTHDF1 fusion recapitulates known binding partners

Encouraged by our organelle-based proximity labeling results, we turned to an even more stringent test of our technology by fusing BS2 to an RBP whose interactions with RNA binding partners has been well-characterized: YTHDF1. YTHDF1 is an m6A reader protein that binds to m6A sites and promotes translation.[166] Its binding partners have been well studied through crosslinking- and antibody-based IP methods, PAR-CLIP and RIP-seq respectively.[166] We therefore appended BS2 to YTHDF1 and assessed transcript enrichment vs. general cytosolic BS2. Of note is how stringent this challenge is: both BS2s we compare are located in the cytosol. The only difference between the two conditions is that one is an RBP fusion. Excitingly, we saw high levels of overlap between the three datasets, with 43% of the transcripts enriched by BS2-seq only, vs. 57% enriched by multiple methods (50% enriched by both BS2-seq and PAR-CLIP, 2% enriched by both BS2-seq and RIP-seq, and 5% enriched by all three methods) (**Figure 4.4D, E**). Additionally, when assessing enriched transcripts based on m6A presence, we found that 77% of enriched transcripts were known to contain at least one m6A site (**Figure 4.4F, G**).[167]

4.3.7 Sub-organellar BS2 fusions reveal interesting and unique enrichment patterns

Having demonstrated that BS2-seq can discern RBP binders in the open environment of the cytosol, we focused our attention on assessing four different types of

BS2 fusions and their resultant enrichment profiles in a more confined space, the nucleus, to see whether we would still see stark differences or whether we would lose proximity labeling. We localized BS2 to the nuclear pore and nuclear lamina, fused it to a nuclear RBP, hnRNPC, and fused it to a histone acetyltransferase, ep300. Using general nucBS2 this time as our reference enrichment dataset, we compared transcripts enriched by each BS2 fusion. When comparing the types of transcripts enriched by nuclear pore BS2 (SENP2-BS2) and nuclear lamina BS2 (LMNA-BS2), 47% of transcripts enriched at the pore are coding, vs. 70% enriched at the lamina, while non-coding transcripts are enriched predominantly at the pore and de-enriched predominantly at the lamina (**Figure 4.5A, B**). When assessing the BS2-seq data vs. APEX-seq data at both sub-organelles and against nuclear MERFISH, there is about 30% overlap between the multiple methods at the pore and 35% overlap at the lamina (**Figure 4.16**). hnRNPC-BS2-enriched transcripts compared to PAR-CLIP-enriched transcripts reveal a 44% overlap, with 74% of the enriched transcripts known to contain m6A (**Figure 4.5C, D**). Inspired by examples of non-coding RNA interactions with chromosomes causing various interference events (xist X inactivation, for example), we chose to finally fuse BS2 to a histone/DNA-associated protein to see what types of RNAs, if any, we might find. In looking at ep300-BS2's enrichment profile, which does not itself bind RNA, we found the transcripts were composed chiefly of non-coding transcripts like lincRNAs, antisense RNAs, and intronic species (**Figure 4.5E**).

Perhaps the most validating observation we made amongst these sub-nuclear enrichment profiles was that, remarkably, each respective profile was almost entirely unique (85-87%) (**Figure 4.5F**). Given that all three are in very close proximity owing to

the confinement of the nucleus, this is a strong demonstration of the viability of the method in distinguishing proximal transcripts.

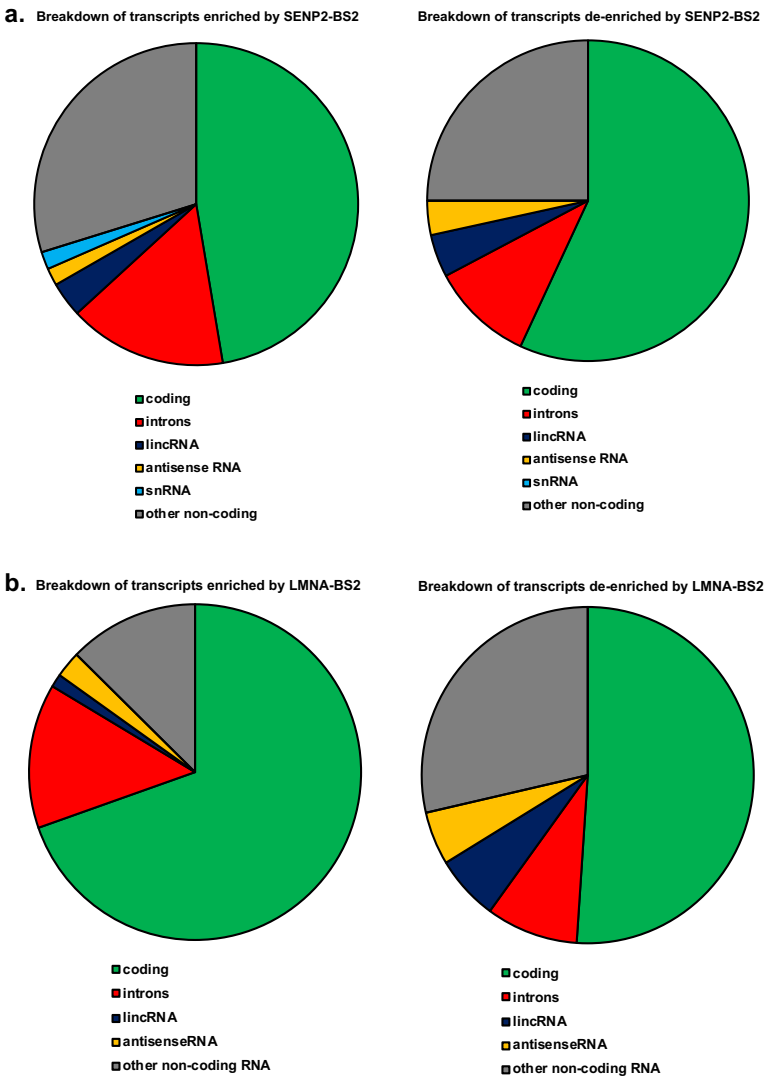
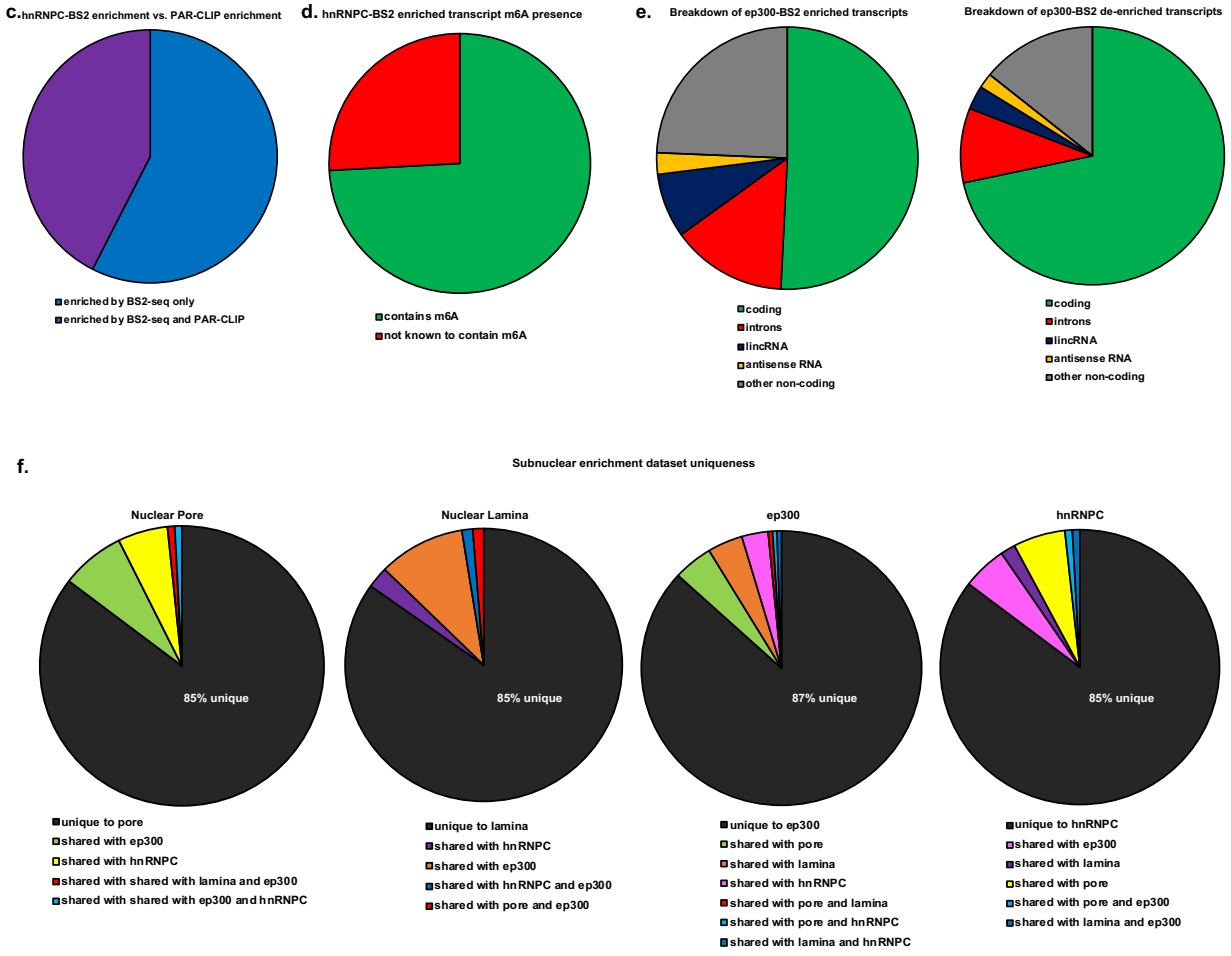


Figure 4.5 RNA-seq of sub-nuclear BS2s. (a) Breakdown of transcripts enriched/de-enriched by SENP2-BS2. (b) Breakdown of transcripts enriched/de-enriched by LMNA-BS2. (c) hnRNPC-BS2 enrichment vs. PAR-CLIP enrichment. (d) m6A presence in hnRNPC-BS2 enrichment profile. (e) Breakdown of transcripts enriched/de-enriched by ep300-BS2. (f) Assessment of the uniqueness of each sub-nuclear dataset.

4.4 Discussion

Given that much of the sub-nuclear data we have generated is something of a "fishing expedition," we are actively now exploring the significance of some of the transcripts we have found in these datasets. We have started, for example, in looking at transcripts that are shared between the datasets, since there are so few (**Figure 4.17**). Many of the shared transcripts have nuclear annotations, and have various functions such as encoding splicing factors, polymerase- and helicase-associated factors, as well as involvement in RNAi.[168]

Figure 4.5 – continued from previous page



And, interestingly, at the pore, many of the transcripts enriched there encode nuclear pore and pore-associated proteins themselves.

Due to its implications in multiple key cellular functions, the spatial organization and protein interaction-engagement of RNA is crucial to understand. Proximity labeling methods have just begun to scratch this surface, and current available methods, while they have paved the way, are based in being the “side product” of predominantly protein-targeted proximity labeling reactions and do not take advantage of the chemistry RNA is uniquely poised to undergo. Here, we report BS2-seq, a method that does just that by harnessing the reactivity of the 2’-OH group of the RNA.

We have shown that BS2/0106-mediated labeling of RNA occurs both *in vitro* and in an in-cell context, likely through a 2'-OH acylation mechanism as we see no DNA labeling. We have leveraged 0106's click handle through imaging experiments as well enrichment/sequencing experiments and demonstrated the method's potential in discovering new RPIs and other interesting nucleic acid-protein interactions. Our method is minimally perturbative when compared to current methods, and no 0106 toxicity has been observed, opening the door for the study of the influence of stress conditions on transcript localizations or RPI engagement, as well as *in vivo* studies.

One area our method and others struggle with is the requirement of significant RNA input (25-50 μg) owing to enrichment inefficiencies. Our method can be done with less input than others (25 vs. 50 μg) even with the requisite click reaction to install the biotin handle. Removing this clicking step would likely only improve our enrichment yields, allowing even less input, which is critical when using some cell lines. The synthesis of a directly biotinylated labeling probe is underway. Additionally, because changing the leaving group component of our labeling probe could in principle tune reactivity and labeling radius, work is underway developing second generation suites. We are also interested in pushing the temporal resolution of the system; current experiments have been done at the 15-minute timepoint, but we have found that even as low as 1 minute incubation with probe yields modest RNA enrichment (**Figure 4.18**). More reactive second-generation probes may in principle make these faster incubations viable. Another area of current focus lies in the deployment of the method for in-cell genetically encodable RNA SHAPE experiments. Since 0106 is in principle a SHAPE reagent, every proximity labeling event is also telling us something about secondary structure at the site(s) of

labeling. Therefore, by preparing BS2-seq RNA-seq libraries in the presence of an RT that inserts a mutational signature or stopping signature at each labeling site,[169,170] we could learn valuable information not only about the presence and abundance of a given transcript in a given area (or engaged in a given interaction) in a cell, but also about its secondary structural components.

As we continue to wrap up this story, we view this work as an important contribution to the RNA proximity labeling toolbox and hope that it is of broad use for those interested in understanding these types of complex interactions.

4.5 Materials and Methods

Cloning. All plasmids were constructed by Gibson Assembly with PCR products generated using Q5 or Phusion DNA polymerases (NEB). The plasmids were sequenced by the University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping facility. Plasmid maps can be found in **Figure 4.6**, and descriptions and links to fully annotated sequence maps for each plasmid can be found in **Table 4.1**. Full vector sequences are also available upon request. APEX2 vectors used in this work were obtained through Addgene (see **Table 4.1**).

***In vitro* transcription.** A gene block containing an upstream T7 promoter (95 bp) was obtained from IDT and served as the reaction template. The template was PCR amplified to yield about 50 μg DNA. The following components were mixed together for the IVT in the following final concentrations: 10x transcription buffer (NEB, M0251S), 100 μL (final

conc. 1x); 1 M $MgCl_2$, 25 μL (final conc. 25 mM); 1 M DTT, 10 μL (final conc. 10 mM); SUPERase-in 30U/μL (Thermo, AM2694), 1 μL (final conc. 30U/mL); 25 mM NTP mix (100 mM NTPs, 50 μL ea.), 200 μL (final conc. 4 mM/NTP); DNA template, 50 μg (final conc. 50 μg/mL); T7 RNA polymerase (NEB, M0251S), 6.25 μL (final conc. 40 μg/mL); RNAse/DNAse-free H_2O, to bring total reaction volume to 1 mL. All reaction components were mixed thoroughly by pipetting up and down several times, and incubated overnight at 37°C. 5 μL of Dnase was then added and incubated for an additional 30 min. at 37°C. The sample was then concentrated down to 150 μL and cleaned up using the RCC-5 RNA Clean & Concentrator Kit (Zymo, R1013). Loading dye was added and the sample was boiled for 5 min. at 90°C. The sample was then loaded onto a 8M urea gel (pre-run for 30 min. at 120V) for gel purification by splitting the sample evenly across 10 wells. After the run, the proper RNA band was excised from the gel using UV shadowing, and the RNA was extracted out of the urea gel using the ZR small-RNA PAGE Recovery Kit (Zymo, R1070).

Recombinant protein expression and purification. The pET-BS2 plasmid (**Table 4.1**) was transformed into chemically-competent *E. coli* BL21 cells. 1 L cultures were grown at 37°C in LB broth (40 μg/mL Kan) to mid log-phase (OD_{600} of 0.65). The cultures were then induced with 0.2 mM IPTG and incubated at 30°C overnight. Cells were harvested by 4°C centrifugation (4000xg for 15 min), the supernatants discarded, and the pellets were resuspended in lysis buffer (50 mM Tris (pH 7.5), 1 M NaCl, 10 mM TCEP, 20% glycerol) and protease inhibitor tablet (Thermo, A32963). The mixtures were then

sonicated (total on time: 10 min; 10s on/20s off; amplitude 30%) and spun down at 12,000 rpm for 40 min. BS2 was purified from the supernatants by nickel affinity chromatography and dialyzed into 50 mM sodium phosphate buffer (pH 7.5). Prior to storage at -80°C, 15% glycerol was added. Final BS2 concentrations were measured using a standard BCA assay.

***In vitro* dot blot.** IVT RNA obtained as described above, as well as corresponding dsDNA and ssDNA (IDT), was subjected to reaction with BS2 (purified, as described above) and 0106 as follows: 500 ng nucleic acid in 1x PBS was reacted with 50 μM 0106 or DMSO control in the presence or absence of 200 nM BS2 for 15 min. at 37°C in a thermomixer at 950 rpm. 100 μM DBCO-488 (Click Chemistry Tools, 1278-1) or DMSO control was added to the samples and they were incubated at 37°C in a thermomixer at 950 rpm for 30 min. Samples were RCC-5 or DCC-5 cleaned up. Samples were subjected to RNAse A (Thermo, EN0531), DNAse I (Thermo, EN0525), or DMSO control at 37°C in a thermomixer at 950 rpm for 30 min. Samples were RCC-5 or DCC-5 cleaned up, blotted, and imaged on a BioRad ChemiDoc imager.

Mammalian cell culture and plasmid transfection. HEK293T (ATCC) cells were cultured in DMEM (L-glutamine, high glucose, sodium pyruvate, phenol red; Corning) supplemented with 10% (vol/vol) fetal bovine serum (FBS, Gemini Benchmark) and 1% (vol/vol) penicillin/streptomycin (Gibco/Life Technologies). Cells were maintained in a

water-saturated, 5% CO_2-containing, 37°C incubator. Cells were used for experiments never exceeding passage number 25. Fresh HEK293T cells were obtained from ATCC and frozen down at an early passage number (5) in individual aliquots. There was no testing for mycoplasma infection as a result. Transient transfections were performed using Lipofectamine 3000 (Invitrogen; Thermo, L3000015) following the manufacturer's protocol.

***In vivo* dot blot.** HEK293T cells were transfected with cytoBS2 (KJ128), nucBS2 (47-17), ermBS2 (43-52), or a no BS2-containing dummy plasmid in biological duplicate. 24 h post-transfection, 50 μM 0106 was added to all wells and incubated for 15 min at 37°C. The cells were then washed 2x with 1x PBS, lysed, and the RNA was extracted and purified from the lysates using the Rneasy Plus Mini Kit (Qiagen, 74136). 100 μM DBCO-488 was added to the samples and they were incubated at 37°C in a thermomixer at 950 rpm for 30 min. Samples were RCC-5 cleaned up, and subjected to RNAse A or DMSO control at 37°C in a thermomixer at 950 rpm for 30 min. Samples were RCC-5 cleaned up, blotted, and imaged on a BioRad ChemiDoc imager.

Fluorescence microscopy. *Cell treatment and CuAAC reaction.* HEK293T cells were plated, on glass coverslips pre-treated with 0.1 mg/mL Poly-D-Lysine (for 2 h at RT) and contained within a 24-well dish, and transfected with cytoBS2-RFP (52-60), ermBS2-RFP (52-12), nucBS2-RFP (52-70), SENP2-BS2-RFP (nuclear poreBS2, 52-62), or no BS2-

containing dummy plasmid in biological duplicate. 24 h post-transfection, 50 μM 0106 or DMSO control was added for 15 min. at 37°C. The cells were then washed 2x with 1x PBS, fixed with 4% paraformaldehyde solution for 30 min. at RT, washed 2x with 1x PBS, permeabilized with ice cold MeOH for 5 min. at RT, and washed 2x with 1x PBS. The samples were then bathed in a CuAAC reaction solution composed in 1x PBS with the following final concentrations, adapted from a previously published method:[171] 2 mM BTTAA (Click Chemistry Tools, 1236-100), 1 mM $CuSO_4$, 10 μM 488-PEG4-alkyne (Sigma, 761621), and 10 mM sodium ascorbate (prepared fresh). The BTTAA, $CuSO_4$, and 488-PEG4-alkyne were pre-mixed in 1x PBS for 5 min., and the sodium ascorbate was added immediately prior to cell bathing. The cells were incubated in this bath for 1 h at RT, followed by 3x washes with 1x PBS for 5 min. each, the second of which contained DAPI. The last 1x PBS wash was removed, and the cell-containing coverslips were then lifted from the 24-well plate, inverted, and transferred to fixative pre-dotted on a glass slide. The slides were dried for 5 h away from light prior to imaging.

Fluorescence Imaging. The cells on slides prepared as above were imaged on an inverted epifluorescence microscope (Leica Dmi8) equipped with a camera (Hamamatsu Orca-Flash 4.0) with a 63x oil objective and light source (Sutter Lambda XL, 300W Xenon). The filters ET490/x, Quad-S, ET 525/x for 488-PEG4-alkyne, Etx/x, Quad-S, Etx/x for RFP, ET402/x, Quad-S ET455/x, and brightfield were used accordingly, and navigated utilizing the Leica LAS X software. Image analysis was performed in ImageJ.[172]

RT-qPCR experiment of ERM targets with BS2 panel. *Azide labeling and copper-free click reaction to install biotin on labeled RNAs.* HEK293T cells plated in a 6-well dish were transfected with cytoBS2 (KJ128), ermBS2 (43-52), or a no BS2-containing dummy plasmid in biological triplicate. 24 h post-transfection, 50 μM 0106 was added to all wells and incubated for 15 min at 37°C. The cells were then washed 2x with 1x PBS, lysed, and the RNA was extracted and purified from the lysates using the Rneasy Plus Mini Kit (Qiagen, 74136). The azide-labeled samples were eluted in 84 μL RNAase/DNAse-free H_2O, to which 10 μL 10x PBS, 1 μL SUPERase-in, and 5 μL 25 mM DBCO-PEG4-Biotin (Sigma, 760749) were added for a total reaction volume of 100 μL/sample. The samples were incubated at 37°C, 950 rpm on a thermomixer for 2 h, followed by an RCC-25 (Zymo, R1017) cleanup. The biotin-labeled samples were eluted with 125 μL RNAse/DNAse-free H_2O and, if not proceeding immediately, stored at -20°C overnight or -80°C long-term until ready to proceed with bead enrichment.

Enrichment of biotin-labeled RNAs. Enrichment of labeled RNAs was performed using the previously-reported APEX-seq enrichment method, with some modification.[173] Briefly, 10 μL Pierce Streptavidin Magnetic Beads (Thermo, 88816) were utilized per 25 μg RNA sample. Beads were washed three times on/off the magnet with 500 μL binding & wash buffer (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.1% Tween-20) per sample, followed by two 500 μL washes with solution A (100 mM NaOH, 50 mM NaCl) per sample, one 500 μL wash with solution B (100 mM NaCl), and a final resuspension of the beads in 125 μL solution B containing 1 μL SUPERase-in per sample. The resuspended beads

were mixed thoroughly with the eluted, biotin-labeled RNA from above and incubated at 4°C for 2 h on a rotator. After this incubation, samples were washed three times with binding & wash buffer as described above, and enriched RNAs were digested off the beads by proteinase K as follows: washed beads were resuspended in 54 μL RNAse/DNAse-free H_2O and 35 μL 3x digestion buffer (330 μL 10x PBS, pH 7.5, 330 μL 20% N-laurylsarcosine sodium solution (Sigma, L7414), 66 μL 0.5 M EDTA, 16.5 μL 1 M DTT, and 357.5 μL RNAse/DNAse-free H_2O) was added to each sample. 1μL SUPERase-in and 10 μL of 20 mg/mL proteinase K (Sigma, P2308) was also added, bringing the total volume for the digestion to 100 μL. The samples were then incubated at 950 rpm on a thermomixer at 42°C for 1 h, followed by 55°C for 1 h. The beads were then placed on the magnet, and the supernatant was extracted and cleaned up using the RCC-5 kit for downstream applications (samples eluted in 6.5 μL water). Concentrations of RNA/sample were recorded. If not proceeding immediately, the samples were stored at -20°C overnight or -80°C long-term until ready to proceed.

Reverse-transcription. The PrimeScript RT Reagent Kit was used to generate cDNA (Takara, RR037B). The RT reaction was assembled as follows: to each 6.5 μL enriched RNA sample, 2 μL 5x PrimeScript buffer, 0.5 μL random 6-mer primers, 0.5 μL oligo-dT primers, and 0.5 μL PrimeScript RT enzyme were added for a total of 10 μL/reaction. The samples were thoroughly mixed and placed on a thermocycler with the following protocol: 15 min. at 37°C followed by a 5 sec. 85°C RT inactivation step and a cooling to 4°C. If

not proceeding immediately, cDNA samples were stored at 4°C overnight or -20°C long-term until ready to proceed with the qPCR.

qPCR. The qPCR reaction was set up as follows: A 1:4 dilution of the cDNA from above was performed. Primer mixes were pre-prepared as 1:20 dilutions of 100 μM stocks of forward and reverse primers (IDT, see **Table 4.2** for primer sequences for each target). 3 μL RNAse/DNAse-free H_2O, 10 μL PowerUp SYBR Green (Thermo, A25742), 2 μL diluted primer mix, and 5 μL diluted cDNA were mixed together thoroughly for a total of 20 μL/reaction. Reactions were loaded onto a 384-well clear plate and the qPCR was performed on a QuantStudio6 Pro System (Thermo) with the standard cycling mode: UDG activation: 50°C, 2 min.; Dual-Lock DNA polymerase: 95°C, 2 min.; Denature: 95°C, 15 sec.; Anneal/Extend: 60°C, 1 min. (40 cycles). Enrichments were calculated by $2^{(Ct\_controlBS2 - Ct\_targetBS2)}$, where the control condition was all labeling components present -BS2 vs. target conditions containing a variably localized BS2; enrichments of the BS2s of expected vs. unexpected transcripts were then assessed.

NovaSeq library preparation. HEK293T cells plated in a 6-well dish were transfected with an array of BS2s or a no BS2-dummy plasmid in biological quadruplicate. The samples were treated as described in the above RT-qPCR experiment protocol, except the samples were not RT-ed or qPCR-ed. Instead, the total enriched RNA samples eluted in 6.5 μL H_2O were diluted to a total volume of 50 μL and transformed into NovaSeq

libraries using the KAPA mRNA HyperPrep Kit (Roche, KK8581, KK8441), containing an mRNA capture step prior to library generation. Typically we yield 50-150 ng RNA from our enrichments, which is on the low end of KAPA's recommended input, but we get high quality libraries using this kit. The manufacturer's protocol is followed exactly, with the following exception: typically, 18-22 PCR cycles are required to achieve sufficient amplification, contrary to the 13-16 cycles recommended by the manufacturer, due to the relatively low total RNA input. The number of amplification cycles required for each library was pre-determined using qPCR prior to amplification of the entire libraries. See **Table 4.2** for the sequences of the primers used (they correspond to the universal adapter portions of the adapter indexes used for each library).

Analysis of BS2-seq enrichment libraries. Kallisto[174] was used to quantify transcript-level abundances of the BS2-seq libraries. A fasta file corresponding to GRCh38 and hg38 was downloaded from the Ensembl website and a kallisto index was generated using the index command with default arguments. Fold-change calculations were performed for ermBS2 tpms+1 vs. cytoBS2 tpms+1 to assess transcripts enriched by BS2 at the ER membrane vs. those enriched by general cytosolic BS2. Transcripts with a FC ≥ 2 ($Log_{10}FC \geq 0.30$) and a P-value ≤ 0.01 ($-Log_{10}P\text{-value} \geq 2$) were considered significantly enriched. Transcripts were highlighted as secretory or non-secretory using previously published datasets[162-165] and GO ontology.[168] Similar analysis was performed for nucBS2 tpms+1 vs. cytoBS2 tpms+1 to assess transcripts enriched by BS2 in the nucleus vs. those enriched by general cytosolic BS2. Transcripts were highlighted as

coding or non-coding according to the GRCh38 transcriptome annotations, and transcripts were specified as nuclear coding using GO ontology. Similar analysis was performed to assess subnuclear BS2 targets, this time with BS2-SENP2 (nuclear pore BS2) or LMNA-BS2 tpms+1 vs. nucBS2 tpms+1 to assess transcripts enriched by BS2 in subnuclear space vs. those enriched by general nucBS2. Transcript types were identified according to the GRCh38 transcriptome annotations, and comparisons of hits were made against nucAPEX-seq and APEX-SENP2/LMNA-seq hits[149] and nuclear MERFISH hits.[175] Similar analysis was performed when assessing DNA/RNA binding protein-fused BS2 targets, with the BS2-YTHDF1 tpms+1 vs. cytoBS2 tpms+1 comparison made to assess transcripts enriched by BS2-YTHDF1 vs. general cytosolic BS2. Hits were compared to published PAR-CLIP and RIP-seq datasets[166] as well as an m6A transcript database.[167] The BS2-hnRNPC tpms+1 and BS2-ep300 tpms+1 were leveraged against nucBS2 tpms+1 to assess transcripts enriched by the RNA- and DNA-binding protein fusions, respectively, vs. general nucBS2. Transcript types were identified according to GRCh38 transcriptome annotations, and comparisons of hits were made against nucAPEX-seq hits,[149] nuclear MERFISH hits,[175] and in the case of BS2-hnRNPC, PAR-CLIP hits.[176]

Synthetic scheme of 0106. Full synthetic details for the synthesis of 0106 are currently unavailable.

4.6 Supplementary Information

Table 4.1 List of all plasmids used in this work.

Vector Name	Res.	Origin	Purpose	Map*	Benchling Link
KJ128	kan	ColE1	cytosolic BS2 mammalian expression plasmid	a	https://benchling.com/s/seq-51TznY3q5ZfFQi2XXUhS
47-17	kan	ColE1	nuclear BS2 mammalian expression plasmid	b	https://benchling.com/s/seq-vw7D8k8JHWGEsfwFqrJY
43-52	kan	ColE1	ER membrane BS2 mammalian expression plasmid	c	https://benchling.com/s/seq-sTsOqY2h7Yiqn16cE0rz
51-52	kan	ColE1	nuclear pore BS2-SENP2 mammalian expression plasmid	d	https://benchling.com/s/seq-z9nTTT9b7Y2ZmbIMHECc
52-60	kan	ColE1	cytosolic BS2-RFP mammalian expression plasmid	e	https://benchling.com/s/seq-Zli7fzosctWtJy42AowQ
52-70	kan	ColE1	nuclear BS2-RFP mammalian expression plasmid	f	https://benchling.com/s/seq-tq4P8pd2tUhAyVRC9oBr
52-12	kan	ColE1	ER membrane BS2-RFP mammalian expression plasmid	g	https://benchling.com/s/seq-64ME5tUGvKHuMRPirC38
52-22	kan	ColE1	BS2-hnRNPC mammalian expression plasmid	h	https://benchling.com/s/seq-L4bUlECk4HmlQGFwS826
52-21	kan	ColE1	BS2-ep300 mammalian expression plasmid	i	https://benchling.com/s/seq-5PGahUJXNylalP9ywgPk
49-45	kan	ColE1	BS2-YTHDF1 mammalian expression plasmid	j	https://benchling.com/s/seq-kmTHOf2DJrGb9d68Ctnf
KJ122	kan	pBR322	BS2 expression plasmid with His tag	k	https://benchling.com/s/seq-Cd1UfTOPGd4vJN4TMYjc
52-62	kan	ColE1	BS2-SENP2-RFP mammalian expression plasmid	l	https://benchling.com/s/seq-Zc8iBWAlcOwvAdFCrNIE
49-42	amp	ColE1	nuclear pore APEX2-SENP2 mammalian expression plasmid	n.a.	Addgene plasmid no. 129276
52-63	amp	ColE1	cytosolic APEX2 mammalian expression plasmid	n.a.	Addgene plasmid no. 49386
52-64	amp	ColE1	nuclear APEX2 mammalian expression plasmid	n.a.	Addgene plasmid no. 124617
52-65	amp	ColE1	ER membrane APEX2 mammalian expression plasmid	n.a.	Addgene plasmid no. 79055

*Vector maps for each construct type shown in Figure 4.6.

Table 4.2 List of all RT-qPCR primers used in this work.

Target (dir.)	Sequence
fau (f)	TCCTAAGGTGGCCAAACAGG
fau (r)	GTGGGCACAACGTTGACAAA
mtcol (f)	CGATGCATACACCACATGAA
mtcol (r)	AGCGAAGGCTTCTCAAATCA
ssr2 (f)	GTTTGGGATGCCAACGATGAG
ssr2 (r)	CTCCACGGCGTATCTGTTCA
tmx1 (f)	ACGGACGAGAACTGGAGAGA
tmx1 (r)	ATTTTGACAAGCAGGGCACC
illumina univ.-1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA
illumina index end-2	CAAGCAGAAGACGGCATACGAGAT

Figure 4.6 Vector maps for all plasmids used in this work.

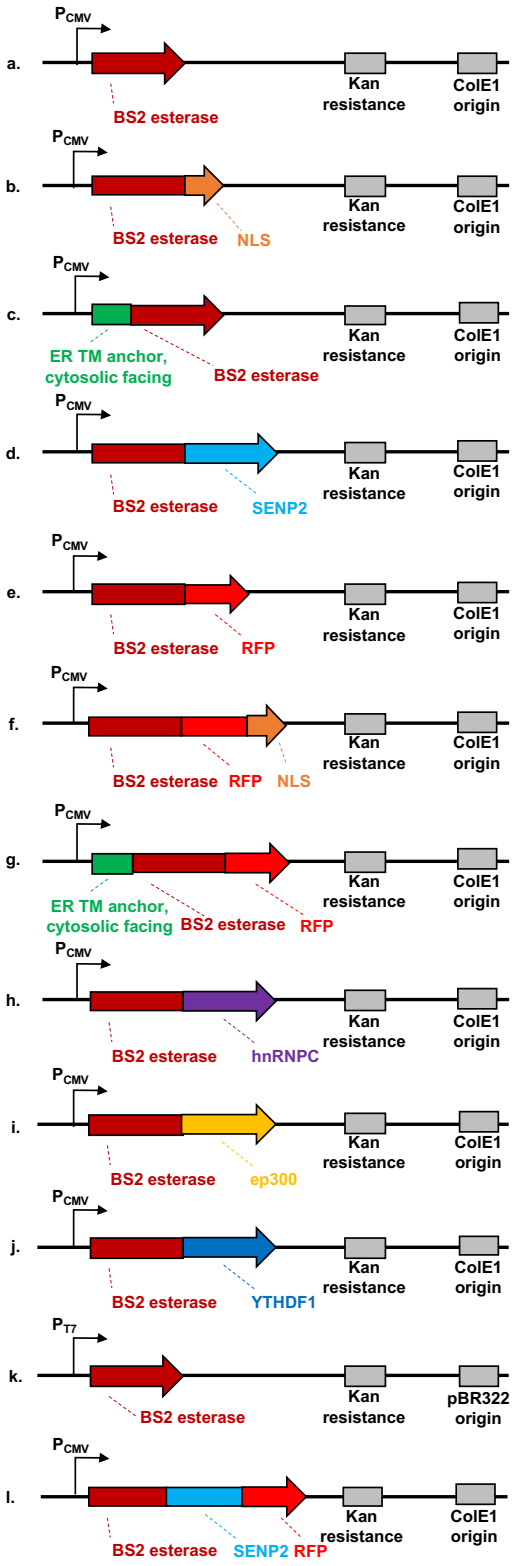
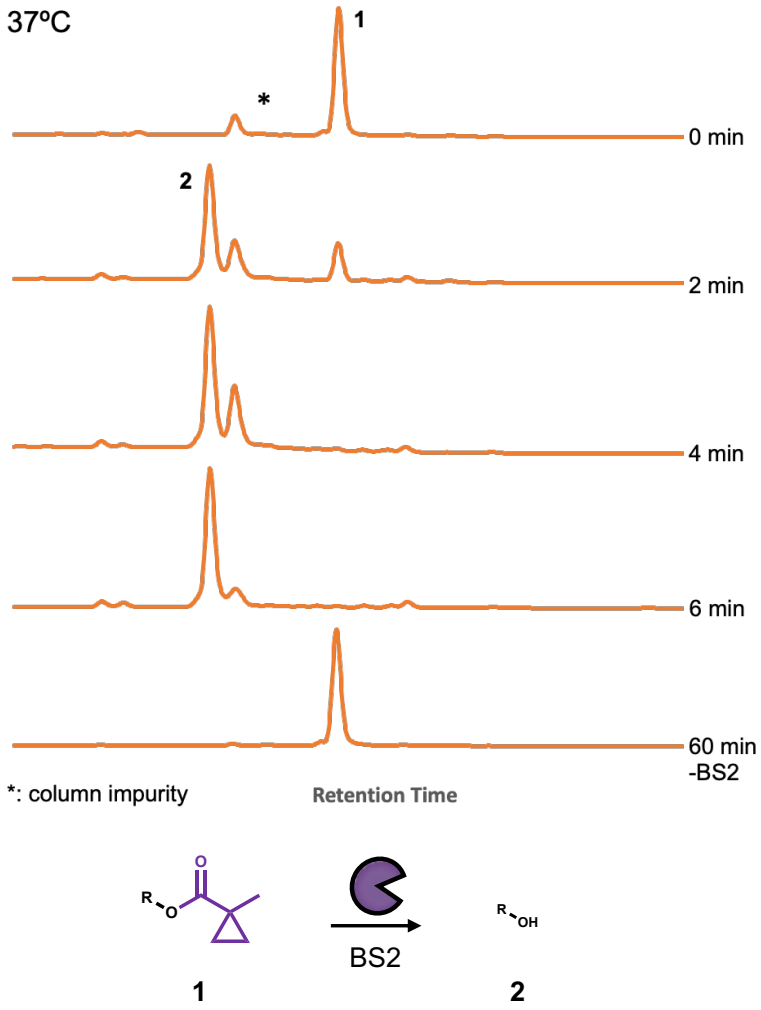


Figure 4.7 Preliminary BS2-mediated methylcyclopropyl unmasking. 2 μM purified BS2 was added to 0.75 mM masked acylating agent (1) in 15% DMSO in PBS at pH 7.4 and quenched with 1 volume MeOH at the indicated timepoint. Between 2 and 4 minutes at 37°C (1) is fully converted to unmasked (2).



4.8 Full dsDNA and ssDNA *in vitro* dot blot. As in **Figure 4.3B** and **C**, ds- or ssDNA was treated with 50 μM 0106 azide and/or DBCO-488, blotted, and read out in an imager. Any signal can be attributed to the stickiness of the fluorophore itself to the membrane and no labeling or sensitivity to DNAse I is observed.

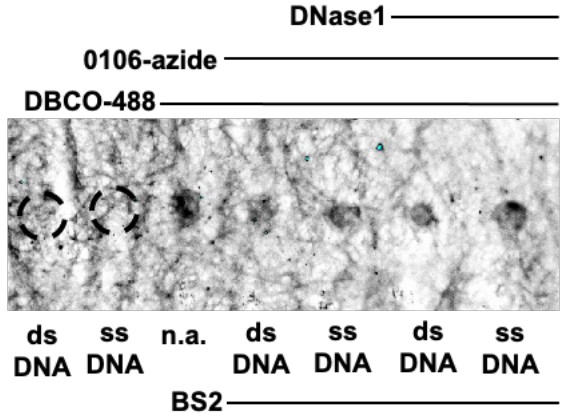


Figure 4.9 *In vivo* dot blot biological replicate. Biological replicate to supplement **Figure 4.3C**. HEK293T cells transfected with cyto-, nuc-, or ermBS2 or dummy were labeled for 15 minutes with 50 μM 0106 azide and the RNA was purified out and subsequently subjected to DBCO-488 click and treatment with RNAse A.

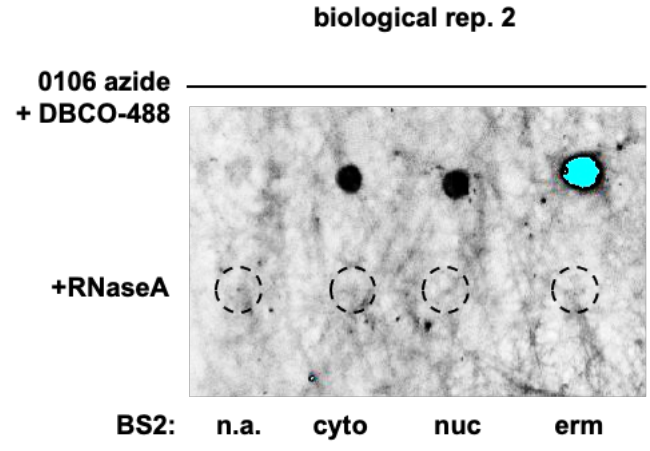


Figure 4.10 Full imaging panel, supplementary to Figure. 4.3D. Contains images supplementary to main text figure. Live, RFP-tagged BS2-transfected HEK293T cells were treated with 50 μM 0106 labeling for 15 minutes, washed, fixed, permeabilized, and subjected to AF488 CuAAC (see methods for more details).

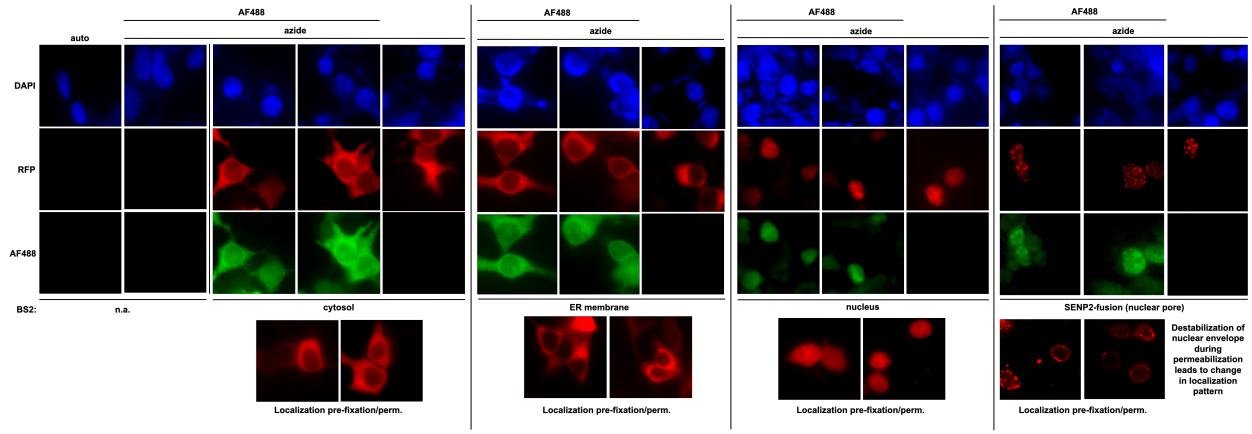


Figure 4.11 On-cell digestion of labeled biopolymers. HEK293T cells transfected with nucBS2 were treated as above. Post labeling, cells were focused on and treated with 5U Dnase I, RNAse A, or Proteinase K. Digestion was allowed to proceed and images were captured at sequential timepoints. Signal sensitivity to Proteinase K and RNAse A was observed.

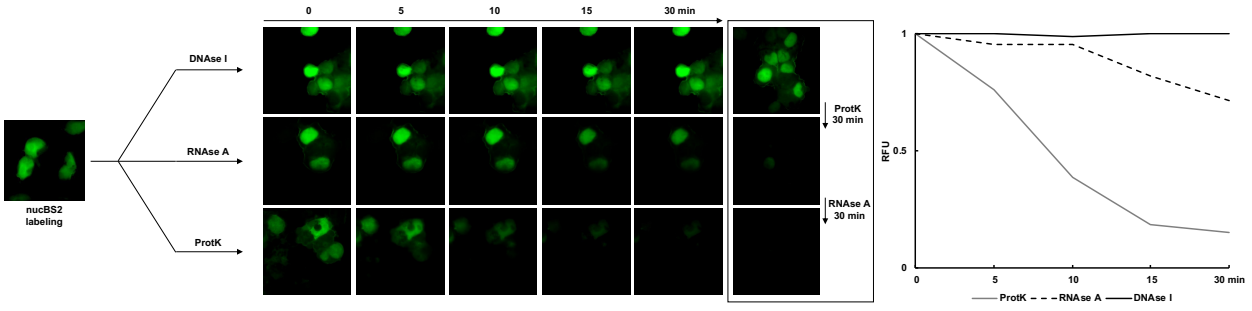


Figure 4.12 BS2/0106 protein labeling efficiency. Total protein lysates from 6-well dishes of HEK293T cells run on an SDS-PAGE that have been transfected with variably localized BS2s, as indicated by number below. 50 μM 0106 was added to the cells for 15 min prior to lysis. DBCO-488 clicked on as described previously and labeled protein shown below. No significant labeling is observed relative to no BS2-containing cells.

O106 labeling (50uM, 15min)
1: NT
2: ERM
3: Nuc-FLAG
4: Cyto
5: LMNA
6: SENP
7: Nuc-RFP
8: Nuc-FLAG 30min

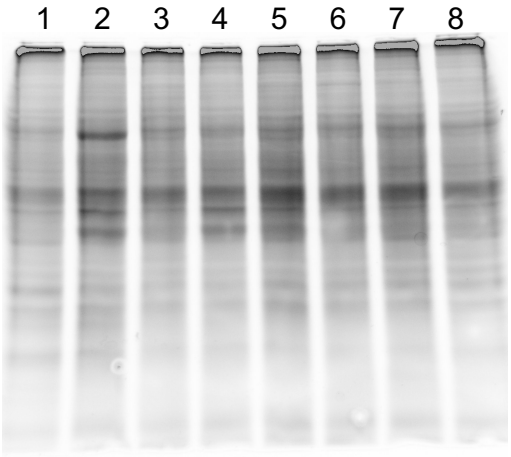


Figure 4.13 BS2/0106 labeling is limited to single cells. HEK293T cells transfected with nucBS2 were co-cultured with RFP-positive HEK293T cells and labeling was performed as above. No substantial labeling overlap between the two populations was observed. 7 biological replicates shown.

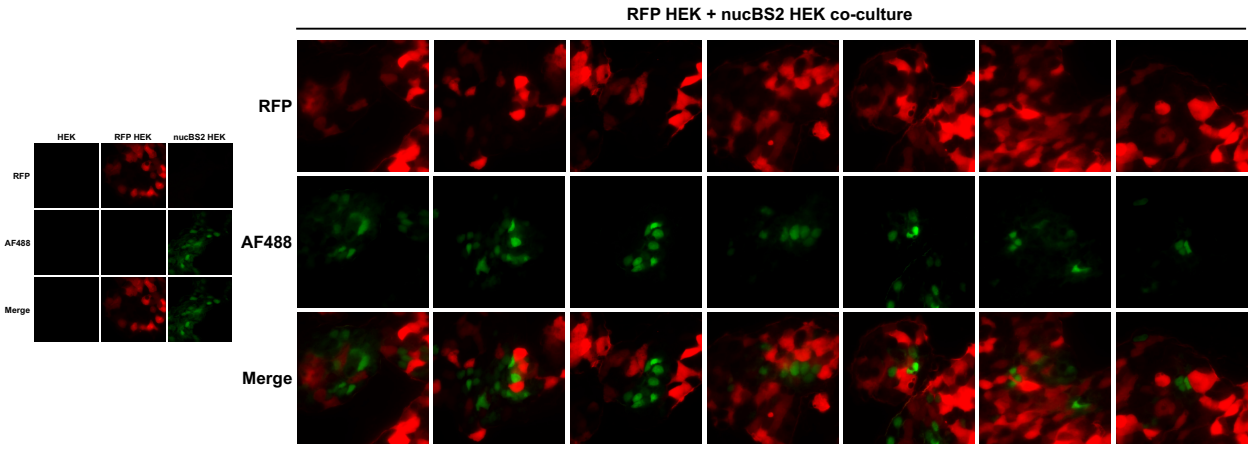


Figure 4.14 BS2/0106 dose response measured by RT-qPCR. See methods section for full experimental details. HEK293T cells transfected with variably localized BS2s subjected to labeling with 0, 10, 50, and 100 μM 0106 azide and subsequently enriched via DBCO-biotin click and streptavidin bead pulldown. Secretory transcripts (ssr2 and tmx1) are enriched to varying degrees while non-secretory transcripts (fau and mtco1) are not enriched differently across the sample types.

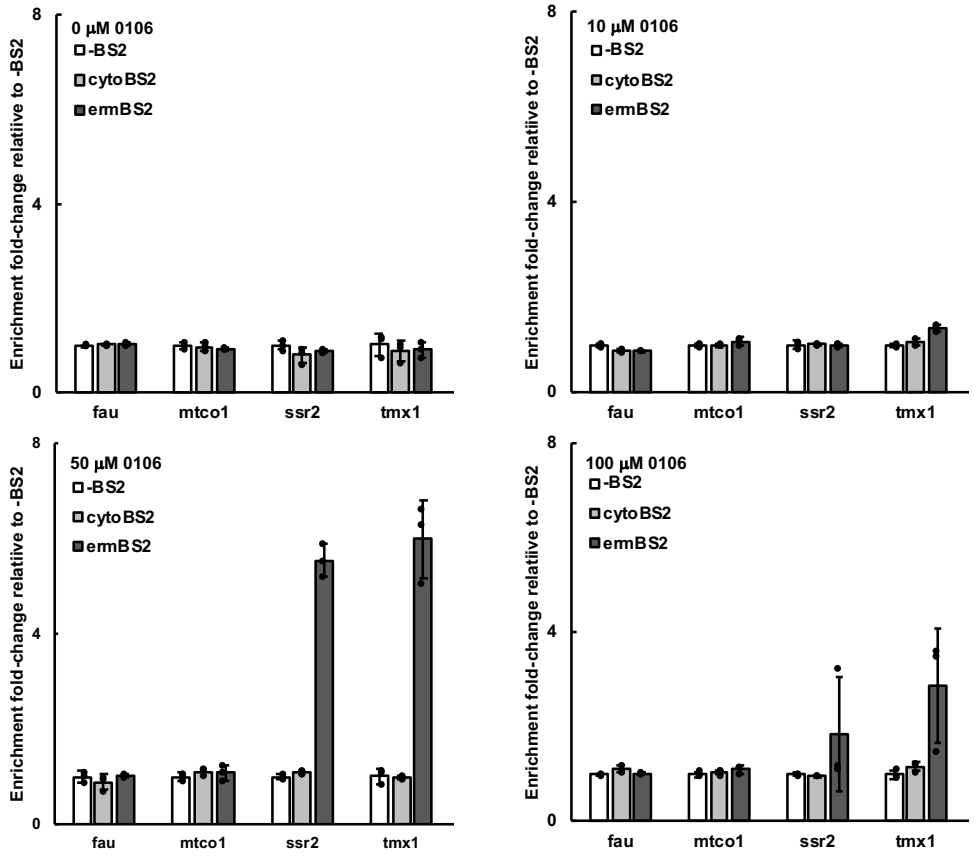


Figure 4.15 BS2 vs. APEX comparison assessed via RT-qPCR. Samples were prepared as above or according to the published APEX-seq protocol.[173] Degree of secretory transcript (ssr2 and tmx1) enrichment by variably localized BS2s and APEXs assessed.

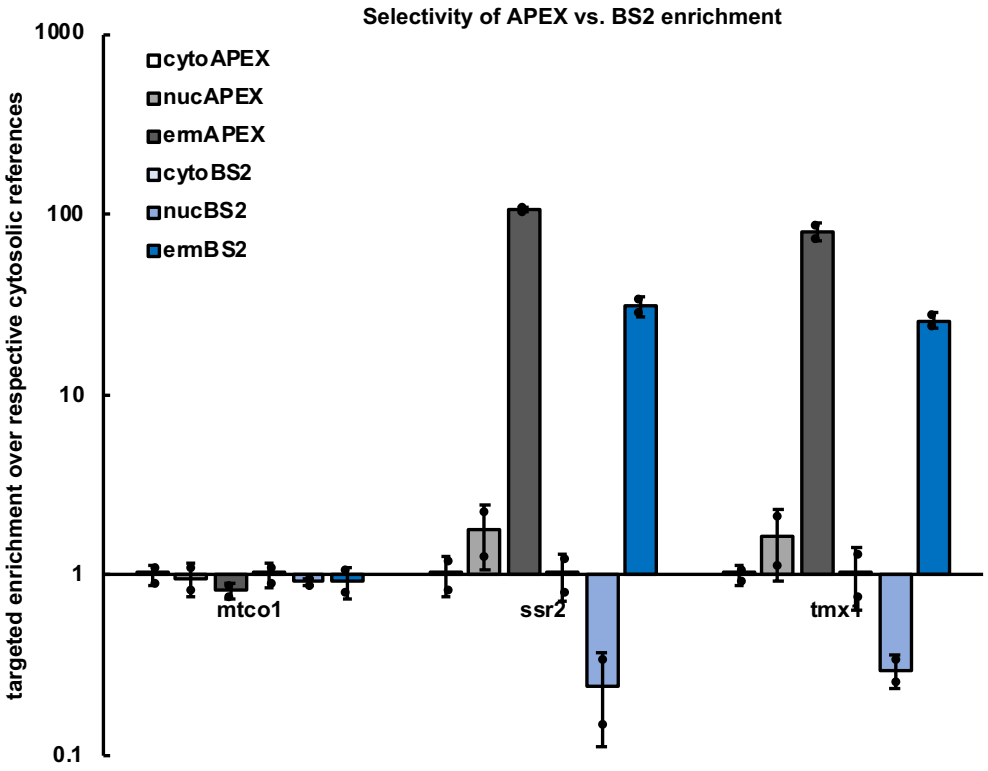


Figure 4.16 Pore, lamina enrichment by BS2-seq, APEX-seq, and MERFISH. Analysis of BS2-seq enrichment at the nuclear pore and nuclear lamina vs. the comparable APEX-seqs[149] and nuclear MERFISH performed.[175] 30% (pore) and 35% (lamina) overlap of any kind observed, with the bulk being overlap between APEX-seq and BS2-seq.

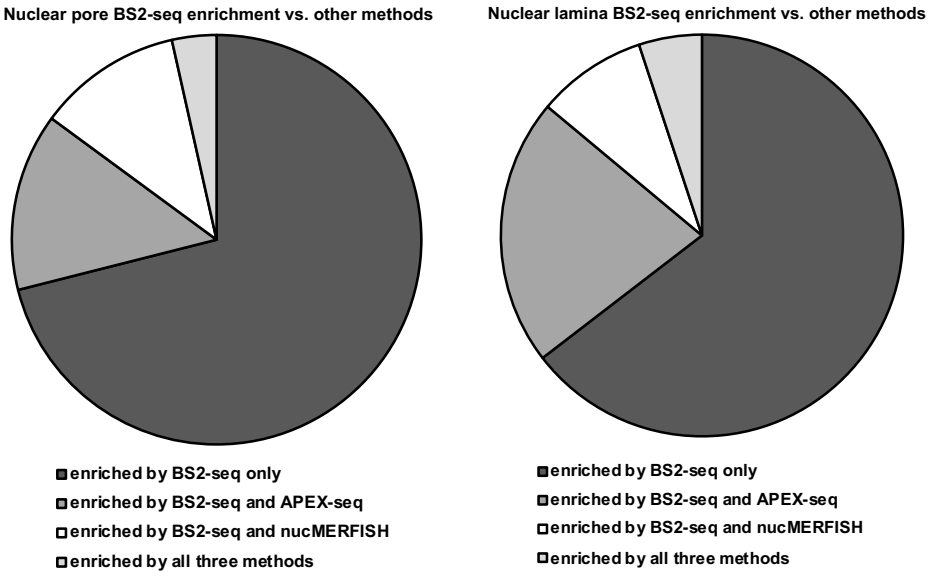
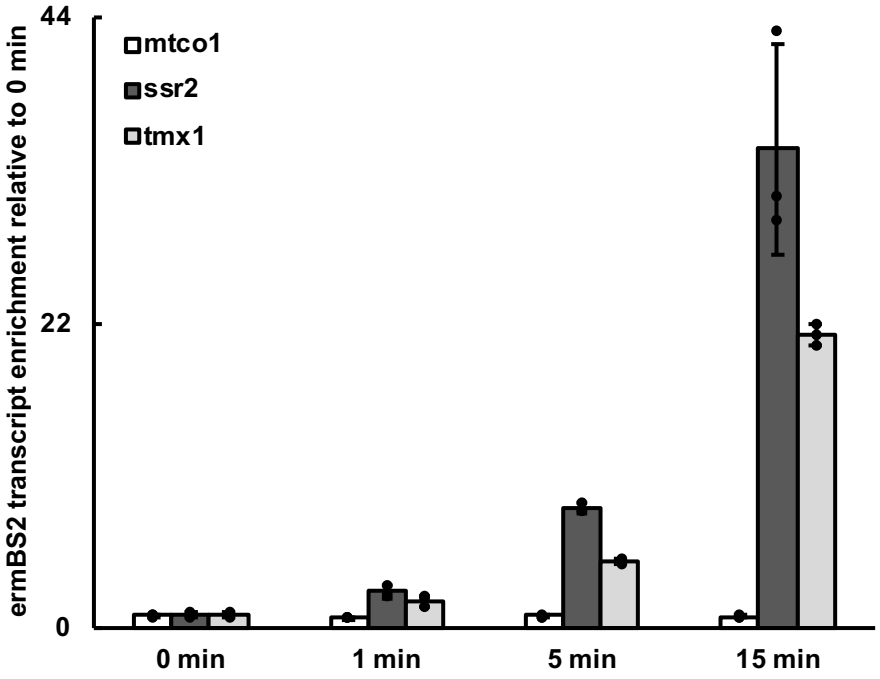


Figure 4.17 GO ontology[168] analysis of shared sub-nuclear transcripts.

sharing	gene name	GO ontology analysis
senp2/lmna/ep300 shared	carhsp1 (coding)	bnds mRNA, p-body assoc., assoc. with U6 snRNA
senp2/hnrnpc shared	wdr54 (intron)	involved in ERK signaling
	cops3 (intron)	nuclear assoc., kinase activity, assoc. with mRNA export from nucleus
	lrsam1 (coding)	ring finger protein, E3 ligase acitivty
	polr2j3 (coding)	rna polymerase subunit, nuclear assoc.
	enah (processed transcript)	actin assoc.
	tbc1d20. (NMD)	nuclear assoc., GTPase activator
	def8 (coding)	nuclear assoc., no other info
senp2/ep300 shared	ebpl (coding)	membrane protein between nucleus and ER
	fxn (coding)	mito/cyto
	eps15l1 (NMD)	EGFR pathway
	c18orf32 (coding)	unknown
	pms2p3 (trasncribed, unproce	nucleic acid binding, mismatch repeair
	nup188 (intron)	nuclear pore protein, assoc. with pore and envelope, gateway protein
	dhrs2 (coding)	nuclear assoc., nuclear pore complex assoc.
	zfand2b (coding)	zinc-finger, nuclear-ER membrane assoc.
	ints10 (coding)	nuclear assoc., RNA polymerase assoc., snRNA binder
senp2/ep300/hnrnpc share	tef (coding)	nuclear assoc., transcription factor
lmna/hnrnpc share	ttc17 (coding)	membrane assoc., actin assoc.
	rrbp1 (coding)	ribsosome binding protein, secretory function
lmna/ep300 share	mcoln1 (intron)	calcium cation channel transmembrane protein, nuclear assoc.
	calcoco2 (coding)	calcium binding, membrane, nuclear assoc.
	cers5 (coding)	ceramide synthase, nucleus-ER membrane assoc.
	iba57 (processed transcript)	mitochondiral, iron-sulfur complex former
	grn (coding)	secreted granulin, ER-assoc.
	lmo4 (processed transcript)	nuclear assoc., transcription factor
	nkd2 (coding)	nuclear assoc., wnt signaling negative regulator, calcium binding
	rab11b (antisense)	ras super family member
lmna/ep300/hnrnpc share	pigo (coding)	GPI anchor biosynthesis, nucleus-ER-sec assoc.
ep300/hnrnpc share	hist2h4a (coding)	histone, nuclear assoc.
	klc2 (coding)	kinesin light chain, assoc. everywhere in cell
	recql5 (coding)	DNA helicase, nuclear assoc.
	srsf3 (NMD)	splicing factor, nuclear assoc.
	znf524 (coding)	zinc finger, nuclear assoc.
	ago2 (NMD)	nucleus, p-body assoc., RNAi

Figure 4.18 BS2/0106 temporal response measured by RT-qPCR. Samples treated as previously described, with ermBS2 degree of enrichment of secretory transcripts (ssr2 and tmx1) assessed with 50 μM 0106 labeling at 0, 1, 5, and 15 minutes.



CHAPTER 5

SUMMARY AND OUTLOOK

5.1 Recap

In conclusion, we have covered the development of three tools to interrogate and harness biomolecular interactions. The first took advantage of the power of the RNA output of split T7 RNAP to control Cas9 through the deployment of small molecule-sensitive gRNAs in an on/off switch action. The second leveraged the interchangeability inherent to esterases and employed one as a new split reporter capable of unmasking a range of small molecules for imaging and cellular change applications. The third represents efforts toward the development of an RNA proximity labeling method that acylates the 2'-OH position of RNA nucleosides upon selective unmasking of an ester-masked enol-thioester agent. The latter two technologies are first-generation tools, and as such, have substantial room for improvement, expansion, and new applications, which we will discuss below.

5.2 Future directions: Split BS2

While we have successfully utilized split BS2 in its capacities both as an imaging-based PPI reporter and as an unmasker of bioactive molecules, we plan to further explore the latter with significantly more interest. Using a cell death-inducer (SN-38) at a high, non-pharmacologically relevant concentration (1 μM) was a great start for proof-of-concept, but ultimately we would like to integrate this technology with more interesting bioactive molecule masking and selective unmasking. Site-specific pro-drug unmasking is an obvious direction to pursue, and we are particularly excited by the potential to

uncage pro-drugs upon the interaction events of cell-cell contacts, as split BS2 functions well intercellularly where many other split systems fail. Additionally, thinking about marrying split BS2 with engineered therapies like CAR-T could in-principle afford the opportunity for BS2 activity-dependent small molecule activation at cancer associated sites. We are also actively exploring engineering multiplexable suites of split BS2s that are orthogonally active on different ester masks which could exponentiate the above applications and possibilities. This area is of strong interest not only with the split BS2 technology but with the BS2-seq technology and beyond, as we will discuss below.

5.3 Future directions: BS2-seq

As a first demonstration of using our BS2-based technology for proximity labeling applications, we chose to focus predominantly on proof-of-concept and basic demonstrations that the method is viable. However, there are many areas for improvement and expansion. 0106 was not the first probe we tested (about a half dozen others failed owing predominantly to reactivity and/or stability issues – some were too reactive and hydrolyzed immediately, while others were completely unreactive and had half-lives on the order of tens of days), but it is likely not the absolute best-in-class. Balancing reactivity in this case is fascinating and frustrating: RNA is not the most abundant material in cells, nor is it the most reactive. Designing a reagent that is reactive enough with RNA without being so reactive that it is immediately gobbled up by reactive metabolites etc. has proven very tricky, and 0106 has thus far proven to strike the best balance. Efforts are currently underway to continue to tune the leaving group (and thus the reactivity) of our masked acylating probe. In principle, a goal could be to develop a

suite of variably reactive probes to achieve different labeling radii. As was likely made clear in Chapter 4, we are also currently sitting on a substantial amount of yet un-mined data, particularly in the case of the sub-nuclear enrichment profiles. Further exploration will no doubt reveal interesting biology that will hopefully be followed up on with collaborators. Time will also tell what true cutoffs for "interesting" data are; at this stage, this field is in its infancy, often driven by fairly arbitrary statistical and enrichment cutoffs. Are transcripts that are enriched 2-fold actually real? 200-fold? Further validation with more traditional methods like IPs is critical in beginning to establish baselines for what is real and what is not, and as we saw in Chapter 4, the overlaps between BS2-seq, APEX-seq, and these more traditional methods are not close to 100%.

As I mentioned at the end of the previous chapter, a direction we are particularly excited by is the application of this technology as a genetically encodable *in vivo* SHAPE system. Our major inspiration for the current method was to design a

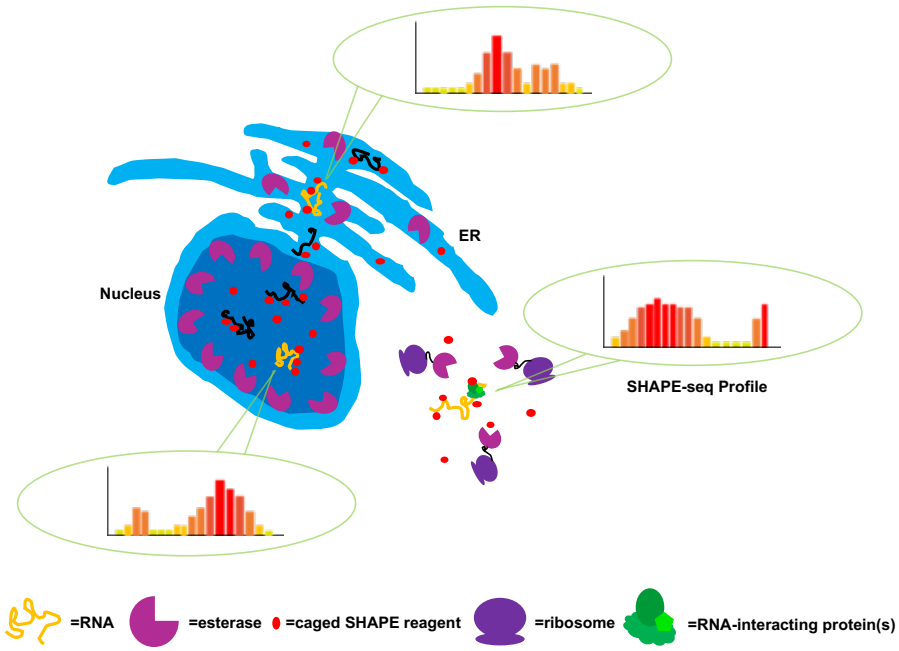


Figure 5.1 BS2-SHAPE-seq premise. BS2 is localized to various regions/proteins of interest, labeling is performed, and analyzed via SHAPE-seq. Following the changing SHAPE profile of a given transcript as it moves though cellular space would be possible in principle.

reagent that behaved in principle like a SHAPE reagent by labeling the 2'-OH of RNA rather than labeling the G base as is true for current methods. Labeling RNA but not DNA provides strong evidence for this mechanism, and given, then, that every proximity labeling event is installing a 2'-O-acylation, preparing RNA-seq libraries with an RT that reads through stop sites and installs mutations, or by processing the samples differently and looking at RT stops,[169,170] would give us not only information on where a given transcript is localized or which protein it is interacting with, but also information on the shape of that RNA in that given context (**Figure 5.1**).

5.4 Future directions: Addressing BS2-based technology pitfalls

While our BS2 esterase/methylcyclopropyl ester selective unmasking-based genetically encodable technologies have proven neat innovations, their widespread use will be substantially hampered by a key pitfall: the methylcyclopropyl mask has turned out to not be bio-orthogonal in many cell lines. This problem was first discovered over the course of work I did on developing the technology as a highly sensitive *in vivo* imaging modality, which I discuss in **Appendix A**. We found that, when we deployed our esterase/ester pair in mice, their livers rapidly uncaged methylcyclopropyl ester-caged Chemilum-CM and all signal we observed in the mice was due to these uncaging events. The original work on this bio-orthogonal ester by Lavis et al. as discussed in Chapter 3 focused predominantly on HEK cells and on a variety of neuronal cells and brain tissues, all of which demonstrated complete bio-orthogonality. Concerned by the background activity we observed in the mice, we were prompted to investigate background activity in other cell lines (primarily liver and immune lines) and found other examples of failed bio-

orthogonality (see Chapter 3). Based on all of these findings, it is clear that for these technologies to be broadly useful to researchers, who often need to do work in more therapeutically relevant cell lines than HEK293T cells, or to work *in vivo*, the esterase must be evolved to accept bulkier esters that cannot be cleaved endogenously across many more cell lines. In **Appendix A**, a directed evolution screening platform was developed based on NNK library generation around the active site for increased function, and should in-principle be immediately applicable for achieving these goals. Moreover, our group recently had some success evolving an esterase with novel substrate recognition, presaging further success in this direction.[203]

APPENDIX A

TOWARD A SYNTHETIC BIOLUMINESCENT REPORTER SYSTEM FOR SENSTIVIE IN VIVO IMAGING

This chapter is reproduced and adapted from the unpublished manuscript: Kentala, K.; Azizi, S.-A.; An, W.; Beck, M. W.; Lippert, A.; Dickinson, B. C. Toward a synthetic bioluminescent reporter system for sensitive in vivo imaging. *Unpublished manuscript*.

A.1 Abstract

Bioluminescent systems isolated from natural sources have proven powerful tools to track cells within organisms, but are challenging to engineer and optimize. Here, we develop an evolved esterase paired with an ester-caged chemiluminescent probe as a new genetically-encoded luminescent reporter system. We demonstrate the reporter is orthogonal to common luciferase systems and is capable of sensitively monitoring small populations of cells. When deployed *in vivo* to monitor tumors, however, we discover a loss in biorthogonality as a result of endogenous liver carboxylesterase activity, precluding its utility in this context. Still, this work lays the foundation for a new approach to bioluminescent imaging, and informs the evolution of esterases against bulkier caging groups for future deployment in mice.

A.2 Introduction

Genetically-encoded bioluminescent reporter systems allow for tracking cells *in vivo* due to the extremely low background of the approach.[177] For example, cancer cells engineered to express a luminescent reporter allow for easy monitoring of tumor growth in murine model organisms using readily available imaging equipment.[178] Naturally-

occurring bioluminescent systems produce light through the enzymatic action of a protein "luciferase" oxidizing a small molecule substrate "luciferin". Currently, firefly luciferase-based systems are the only bioluminescent reporters that have found widespread use *in vivo*.

While the natural firefly luciferase has proven powerful for *in vivo* imaging, there is a strong desire to engineer reporters with enhanced properties, such as a brighter output, faster kinetics, improved tissue distribution, and orthogonality for multidimensional experiments. Recent advances in luciferase directed evolution coupled with the synthesis of new luciferins has begun to address these challenges.[179-181] For example, new approaches for rapidly screening luciferin/luciferase pairs for orthogonality yielded a set of three orthogonal reporters[182] and a red-shifted, blood-brain barrier permeable luciferin analog has begun to make deep brain imaging possible.[183] However, the engineered luciferase enzymes still require luciferins that are structurally similar to the native substrate, forbidding the deployment of the diverse array of potentially synthetically-accessible chromogenic molecules. Additionally, the most commonly used firefly luciferase reporter requires both molecular oxygen and ATP, leading to complications in deploying these tools in low-oxygen or extracellular environments. Renilla luciferase, while it bypasses these requirements, suffers from suboptimal coelenterazine biodistribution in animals, significantly hindering its applicability in vivo.[184-186]

The design and synthesis of small molecule chemiluminescent reporters has rapidly progressed to the point where *in vivo* imaging of chemistry has been achieved for multiple target activities.[187,189,190] However, these reporters have not yet been adapted

into a biorthogonal genetically-encoded platform capable of monitoring gene expression, tracking cells, or measuring tumor growth, analogous to luciferase reporters. Porcine liver esterase (PLE) is capable of processing a methylcyclopropyl-ester substrate, which is not processed by any endogenous enzyme in human cells.[188] We envisioned that this orthogonal ester/esterase pair could be adopted into a chemiluminescent reporter system, in which an ester-masked pro-chemiluminescent substrate is only activated in cells expressing the orthogonal esterase. If possible, such a system could take full advantage of red-shifted and brighter chemiluminescent substrates as they continue to be developed, as ester caging small molecule probes would be a broadly applicable approach.[189,190,192] While luciferases have evolved high levels of selectivity for a given luciferin, making engineering difficult, the esterase would likely be capable of unmasking any synthetically-accessible ester-caged chemiluminescent substrate. Additionally, esterases have proven to be malleable targets through directed evolution,[191] providing a path forward for esterase substrate reprogramming to generate orthogonal ester/esterase pairs, ultimately allowing for multidimensional bioluminescent reporting. Bioluminescent systems based on β-galactosidase and alkaline phosphatase activity have been validated in cell culture[193] and whole animal enzymatic imaging;[193,194] however, β-galactosidase is naturally present in cells, precluding bioorthogonality, limiting dynamic range, and necessitating a cell-targeting peptide to achieve tumor specificity. Moreover, both β-galactosidase and alkaline phosphatase are challenging protein engineering targets for further optimization and development as bioluminescent platforms.

A.3 Results

A.3.1 Selection of mutational sites and NNK library generation

To test the feasibility of our proposed approach, we first confirmed the activity of expressed PLE in mammalian cells using a previously reported fluorescent fluorescein-CM_2 substrate (**Figure A.1A**).[188] As was reported, expression of PLE in HEK293T cells resulted in enhanced fluorescent signal from fluorescein-CM_2 compared to control cells. However, compared to the level of fluorescence achieved

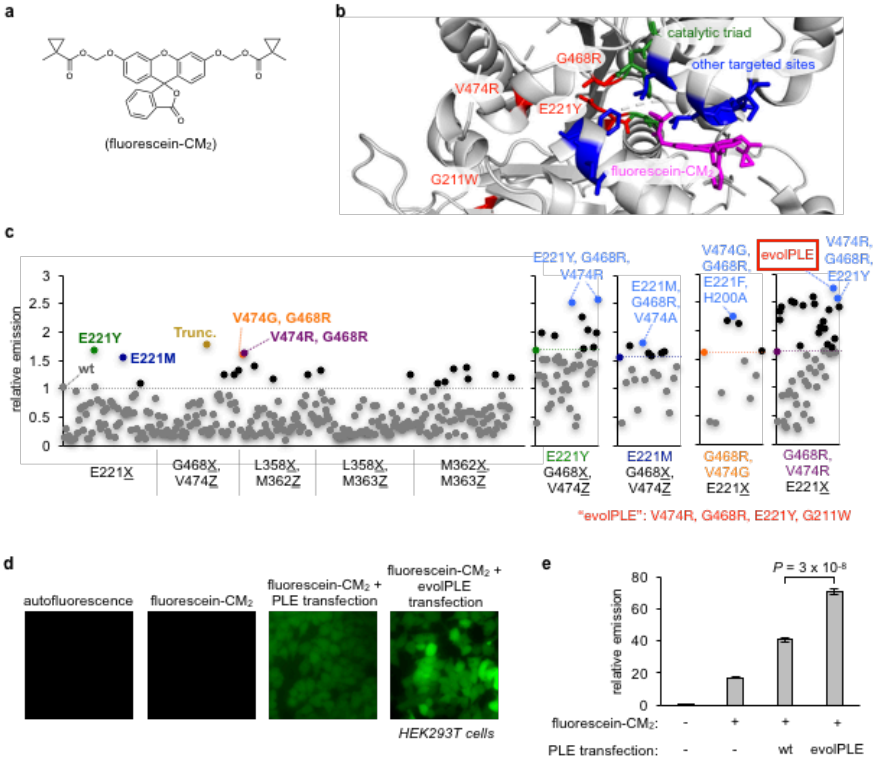


Figure A.1 Directed evolution of PLE to process methylcyclopropyl-ester substrates more effectively. (a) Structure of methylcyclopropyl-ester masked fluorescein (fluorescein-CM_2). (b) Fluorescein-CM_2 (magenta) docked to the active site of PLE-homologous hCE1 (PDB: 3K9B), with the catalytic triad shown (green). Residues targeted in the screen are shown in red and blue, with the final mutational sites from the most active variant shown in red. (c) Mammalian cell screening assay of libraries of PLE variants. Cells were loaded with fluorescein-CM_2 for 20 min and then analyzed by fluorescence microscopy. In round 1, five sets of sites were targeted, resulting in four genotypes with enhanced activity. In round two, each of those four genotypes were screened for additional mutations, targeting sites predicted to make contacts with the primary mutations, culminating in evolPLE. (d) Comparison of PLE and evolPLE. (e) Quantification of experiment shown in d. Error bars are ± s.e.m. (n = 6 data points, two biological replicates). Statistical analyses performed with a two-tailed Student's t-test with unequal variance.

from endogenous esterases on acetyl ester fluorescein substrates, we found that the overall signal was quite low, indicating that methylcyclopropyl-ester is not an optimal substrate for PLE. Because our ultimate goal is to deploy PLE *in vivo*, we wanted to first ensure we were maximizing the signal output of the system. As a bioluminescent reporter, each substrate turnover of the esterase results in the production one photon, meaning the faster the reaction the more light will be produced. Therefore, we sought to first optimize the activity of PLE on the methylcyclopropyl-ester substrate through directed evolution.

To optimize PLE, we sought to target putative residues that make substrate contacts in the active site by random mutagenesis, and then screen the resultant libraries of enzymes in live mammalian cells. There is no substrate-bound crystal structure of PLE,[194] but there is a high-resolution substrate-bound structure of a related enzyme, human liver carboxylesterase 1 (hCE1) (PDB: 3K9B), which is 80% identical. Therefore, we employed flexible ligand/rigid receptor docking to computationally model the fluorescein-CM_2 methylcyclopropyl-ester substrate in the enzyme active site (**Figure A.1B**). Using this model, we targeted six pairs of sites in the PLE mammalian expression vector, each by NNK mutagenesis, generating separate libraries each with 400 theoretical variants. We then used high-throughput imaging by fluorescence microscopy with fluorescein-CM_2 to measure the activity of 58 variants from each library (290 variants total, **Figure A.1C**). This first round of evolution yielded four genotypes that enhanced activity, including E221Y, E221M, V474G/G468R, and V474R/G468R. For the second round of directed evolution, we generated separate libraries of mutants for each genotype targeted at secondary sites that make contacts with the first-round

mutant sites. We screened between 10 and 50 mutants from each separate library, identifying several variants with enhanced activity over the parent genotypes. The most active variant from the screen, which we named “evolPLE”, contained four mutations: V474R, G468R, E221Y, and G211W. The G211W mutation, which is not predicted to lie in the active site, arose due to a random mutation in cloning, but did enhance activity slightly relative to the variant without the additional mutation (**Figure A.1C**). To confirm the enhanced activity of evolPLE, we performed quantitative imaging experiments with fluorescein-CM_2 comparing wild-type PLE with evolPLE (**Figure A.1D, Figure A.3**), which confirmed evolPLE can process the fluorescein-CM_2 methylcyclopropyl-ester substrate substantially better in live cells (**Figure A.1E**).

A.3.2 evolPLE sensitively detects cells in a dose-dependent manner

Next, we sought to explore whether evolPLE could be deployed as a bioluminescent system. We first synthesized Chemilum-CM (**Figure A.2A**), a methylcyclopropyl-ester-masked pro-chemiluminescent substrate based on a previously reported scaffold.[183,192] We reasoned that evolPLE action on Chemilum-CM would release the chemiluminescent form of the probe, thereby generating a photon and resulting in bioluminescence. To assess the function of the system in live mammalian cells, we expressed evolPLE on the surface of MBA-MB-231 cells, taking advantage of the lack of cofactor requirements of the system, and measured luminescence by plate reader. Addition of Chemilum-CM to control cells resulted in no significant luminescent signal (**Figure A.2B**, grey line), further confirming the bioorthogonality of the methylcyclopropyl-ester. However, cells expressing evolPLE and treated with Chemilum-

CM displayed substantial luminescent signal (**Figure A.2B**, green line), indicating evolPLE can process Chemilum-CM and generate light in live cells.

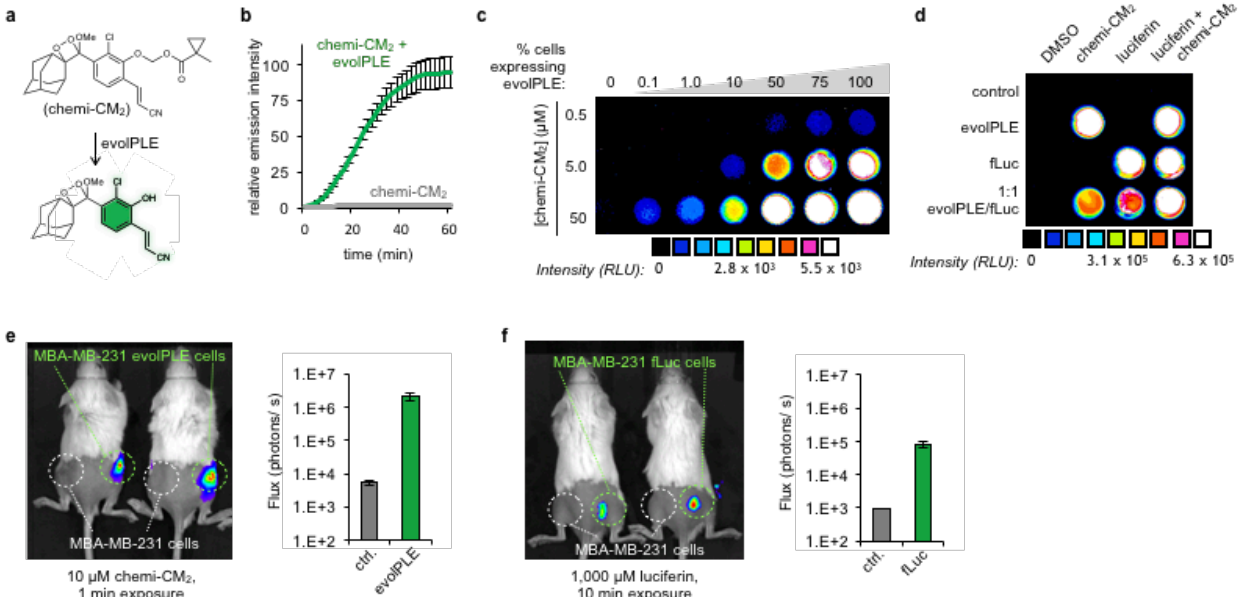


Figure A.2 Imaging evolPLE/Chemilum-CM bioluminescence in cell culture and animals. (a) Structure of Chemilum-CM. Deprotection results in photon release. (b) MBA-MB-231 cells with/without evolPLE are loaded with 50 μM Chemilum-CM and luminescence monitored by plate reader. Error bars ± s.e.m (n = 4 biological replicates). (c) Mixtures of MBA-MB-231 cells and MBA-MB-231 evolPLE added to each well of a 96-well assay plate, with a constant 1 x 10^5 cells in each well, varying the percentage of evolPLE cells from 0% to 100%. 0.5, 5.0, or 50 μM Chemilum-CM was then added to each well and the plate imaged for 30 min. (d) 1 x 10^5 MBA-MB-231 cells, MBA-MB-231 evolPLE cells, MBA-MB-231 fLuc cells, or a 1:1 mixture of evolPLE and fLuc cells added to each well of a 96 well plate. The cells were then treated with either DMSO carrier control, Chemilum-CM, D-luciferin, or a mixture of chemi-CM_2 and D-luciferin, and luminescence monitored for 30 min. (e) Control or evolPLE expressing MBA-MB-231 cells implanted in opposing flanks of mice. One day post implantation, 10 μM chemi-CM_2 was administered IP and the mice imaged for 1 min. Error bars ± s.e.m (n = 3 mice). (f) Control or Fluc expressing MBA-MB-231 cells implanted in opposing flanks of mice. One day post implantation, 1000 μM luciferin was administered IP and the mice imaged for 10 min. Error bars ± s.e.m (n = 3 mice).

With the evolPLE/Chemilum-CM concept validated, we next assessed the sensitivity and dose-responsiveness of the system in live cells by imaging. First, we added a constant number of MBA-MB-231 cells to each well of a 96-well plate, but varied the

percentage of MBA-MB-231 cells expressing evolPLE to assess sensitivity. We then added 0.5, 5.0, or 50 μM Chemilum-CM to each well and imaged the luminescence from the plate on a luminescence gel imaging system. We observed luminescence from the cells that was dependent on both the number of cells expressing evolPLE and the amount of Chemilum-CM added to the well (**Figure A.2C, Figure A.4**). At 0.5 μM Chemilum-CM, we could observe signal with 50% of the cells expressing evolPLE, which corresponds to 50,000 cells. At 50 μM Chemilum-CM, we could observe detectable signal even with 0.1% of the cells expressing evolPLE, which corresponds to detecting 100 cells. We next compared the evolPLE/Chemilum-CM to the commonly used firefly luciferase (Fluc) reporter system. We added control, evolPLE, Fluc, or a 50:50 mixture of evolPLE and Fluc expressing MBA-MB-231 cells to separate wells of a 96-well plate, and then treated each well with DMSO, Chemilum-CM, firefly luciferin, or a mixture of Chemilum-CM and firefly luciferin. As expected, we found that the two systems were completely orthogonal to one another (**Figure A.2D**). Moreover, 250 μM firefly luciferin was required to generate equal luminescent signal as 50 μM of Chemilum-CM, further illustrating the sensitivity of the new reporter. To confirm the relative signals from the two systems, we performed dose curves for Fluc/luciferin and evolPLE/Chemilum-CM (**Figure A.5**), which again confirmed that the evolPLE generates detectable signal even at 0.5 μM Chemilum-CM, while Fluc requires at least 50 μM to begin producing detectable signal under these assay conditions. These experiments confirm that the evolPLE/Chemilum-CM system works as a new orthogonal bioluminescent reporter in cell culture.

A.3.3 evolPLE is deployed in vivo and runs into a substantial road-block

Finally, we sought to test whether the evolPLE/Chemilum-CM system could be deployed *in vivo*, using a murine tumor model system as an exemplar. In the context of a whole animal, issues such as biodistribution of Chemilum-CM, stability and bioorthogonality of Chemilum-CM, and *in vivo* evolPLE activity are all critically important. We injected MBA-MB-231 cells subcutaneously into the right flank of mice and an equivalent number of MBA-MB-231 cells expressing evolPLE into the left flank. After 24 hours, we then injected 10 μM Chemilum-CM by intraperitoneal injection (IP), which was optimized through dose-response experiments (**Figure A.6**) and imaged the mice on a whole animal luminescent imager (IVIS). Initially, we observed what we thought was robust signal from the site injected with evolPLE-positive cells, and no signal from the control injection site or anywhere else in the mice. However, upon switching the orientations of the evolPLE-positive- and control-tumors for due diligence, we discovered that luminescent signal was radiating from near the control injection and not near the evolPLE injection. Further examination of the signal revealed that the predominant signal we were seeing in the mice, and thus the predominant uncaging events, were occurring in the liver and not in the tumor(s). These results were disappointing but did importantly alert us to the reality that the methylcyclopropyl-ester cage was not universally bioorthogonal, and that mouse liver carboxylesterases actually could process the group with ease, as signal was maxed out after only 60s. Since the cell lines ourselves and others had predominantly assessed for endogenous activity had been epithelial and neural, we wondered whether human liver cell lines and other types of cell lines might also see endogenous activity on the methylcyclopropyl-ester cage. Indeed, many cell

lines, particularly the liver ones, do express esterases endogenously capable of unmasking our probes (see Chapter 3).

A.4 Conclusions and future directions

In summary, we present the design, optimization, and validation of a completely new bioluminescent reporter system capable of sensitively and dose-dependently monitoring cells. Despite its present bioorthogonality shortcomings, the evolPLE/Chemilum-CM system offers many advantages compared to current naturally-isolated bioluminescent systems. For one, it is free of the structural constraints inherent to luciferase-based systems, thereby allowing full synthetic control over the caged substrate probe. Accessing brighter and red-shifted versions of Chemilum-CM should be possible by simply exchanging the small molecule scaffold and are currently being developed. The cell-based screening methods described here should provide a path forward toward evolving orthogonal evolPLE variants by using fluorescein-CM_2 derivatives with varied ester cages. This is of particular import and is currently being investigated due to the discovery that the bioorthogonality of the methylcyclopropyl-ester is not as universal as we initially thought, and PLE variants capable of uncaging much bulkier substrates should allow for the successful deployment of this system for *in vivo* imaging applications, as was originally intended. Since the evolPLE/Chemilum-CM system is fully compatible with other bioluminescent systems, it should provide a nice way to perform multidimensional cell-tracking experiments. Overall, we believe that once esterase variant(s) capable of uncaging universally-bioorthogonal substrate(s) are

finished being developed, they will provide a suite of very powerful imaging tools that can be deployed to multidimensionally track cells in animals.

A.5 Materials and Methods

Cloning and vectors. Plasmids were constructed via Gibson Assembly[196] from PCR products made using Phusion DNA Polymerase (homemade) or Q5 DNA Polymerase (NEB). The pcDNA3-IgK-PLE-HA-DAF-IRES-mCherry vector was a gift from Professor Evan W. Miller (University of California, Berkeley), which was used as a comparison and to construct the evolPLE expression vector. The gene for PLE was codon-optimized and synthesized (IDT), and cloned into a custom CMV-driven mammalian expression vector. Saturation mutagenesis studies were conducted using NNK-containing primers from IDT and Gibson Assembly to clone variants into the CMV-driven mammalian expression vector. pmirGLO vector for firefly luciferase expression was purchased from Promega (#FJ376737). All vectors utilized in this study are outlined in **Table A.1** and are available upon request.

Mammalian Fluorescence Imaging and Quantification. All *in cellulo* imaging was done on an inverted epifluorescence microscope (Leica DMi8) equipped with a camera (Hamamatsu Orca-Flash 4.0) with either 20x air objective or 63x oil objective (N/A 1.4) and light source (Sutter Lambda XL, 300 W Xenon) for fluorescein-CM_2 (ET 490/20x, Quad-S, ET 525/36 m), Hoechst 33342 (ET 402/15x, Quad-S, ET 455/50 m), RFP, and brightfield using Leica LAS X software. Either HEK293T cells (at 20x, for NNK mutant screening) or MDA-MB-231 cells (for 63x imaging of wt PLE/evolPLE comparison) were

transfected with the appropriate vectors using either Lipofectamine 3000 (ThermoFisher) or Lipofectamine LTX and PLUS reagent (ThermoFisher), respectively and following the "96-well" protocol for each. 24h after transection, the cells were imaged. Analyses were performed in ImageJ (Wayne Rasband, NIH).

Flexible ligand-rigid receptor molecular docking. Flexible ligand-rigid receptor molecular docking was performed with fluorescein-CM_2 using Autodock Vina 1.1.2[197] (exhaustiveness=1024) on a homebuilt server with an Intel Xeon E3-1220 V3 3.1GHz Quad-Core Processor and 16 GB of ECC RAM running the latest stable release of Debian 8 Linux. Receptor and ligand structures were prepared using previously reported methods.[198] Briefly, ligand structures were generated in Avogadro 1.1.1[199] and MGLTools 1.5.7rc1[200] with energy minimization using openbabel 2.3.2[201] with the MMFF94s forcefield. The receptor structure PDB: 3K9B was prepared in PyMOL 1.7 (Schrödinger, LLC), existing hydrogens, where present, were removed and replaced using MolProbity 4.4[202] to ensure the structure properly modeled protonation at pH 7.4. Grid parameters of size_x = 79.00, size_y = 71.00, size_z = 71.00, center_x = 3.18, center_y = 95.58, center_z = 48.91 were used to encompass the entire active site. Analysis of the lowest energy confirmation and generation of the figure was performed using PyMOL.

Plate Reader Imaging and Quantification. Plate reader data was collected on a BioTek Synergy Neo2. MDA-MB-231 cells were plated on a glass-bottom, black-walled 96-well plate (InVitro Scientific, #1 Cover Glass) and transfected with either evolPLE or an empty

vector using Lipofectamine LTX and PLUS reagent (ThermoFisher), according to the “96-well” protocol. After 24 h, the media was replaced with 1x PBS containing 50 μM chemi-CM_2, and the plate was immediately subjected to analysis by the plate reader. The following conditions and protocol were followed: Temperature setpoint: 37°C, Start Kinetic Runtime: 2:30:00 (HH:MM:SS), read interval 0:01:00 (151 reads total), Linear Shake: 0:10 (MM:SS) at frequency 567cpm, Read: Luminescence Endpoint (Integration time: 0:00.30 [MM:SS:ss], read height 6 mm, Filter Set 1, Emission: Hole). Four biological replicates were run for each condition and the data was averaged and normalized.

Cell culture luminescence imaging. MDA-MB-231 cells (ATCC) were grown in 3.5 cm dishes and transfected with either evolPLE, firefly luciferase, or an empty vector using Lipofectamine LTX and PLUS reagent (ThermoFisher), according to the “6-well” protocol. After 24 h, the cells were resuspended and spun down at 800 x g for 5 min. The supernatant was discarded, and the cell pellet was transferred to a clear-bottom, black-wall 96-well plate (1 3.5 cm well: 1 well in the 96-well plate). Solutions of the chemi-CM_2 in 1x PBS with no greater than 1% DMSO were added in 25 μL volumes to bathe the cells. The plate was then transferred to an imager and exposed for 30 min. The data was worked up in ImageJ by converting the image lut from “grays” to “16-colors” and each well was measured for RLU using the ImageJ “measure” tool. For **Figure A.2C**, the number of MDA-MB-231 cells was kept constant in each well, with the percentage of PLE-containing cells of the total varying. The appropriate volumes of each cell type were thoroughly mixed and spun down collectively before the pellet was transferred.

Stable cell generation. A G418 (geneticin) kill curve for MDA-MB-231 cells was generated and found to be ~750 ug/mL. The cells were then plated transfected in a 6-well plate using Lipofectamine LTX and PLUS reagent (ThermoFisher), according to the "6-well" protocol with the evolPLE vector containing a geneticin positive selection marker. 72h after transfection, the selection antibiotic was added to the cells and the cells were monitored for survival for 10 days, changing out the media and selection antibiotic every 3 days. Surviving cells were allowed to expand, then passaged sequentially to 10 cm and 15 cm dishes and allowed to expand. Media and selection antibiotic were changed every 3 days. Cells were then resuspended and diluted to ~1 cell:100 µL, and several 96-well plates were plated with 100 µL of the diluted cells per well, such that each well contained roughly one cell. Those wells found to contain one cell were allowed to expand. Monoclonal colonies were frozen down for later use after validation of stable protein expression.

Murine experiments. 6-week-old C57BL/6J mice were obtained from Jackson Laboratories (#000664). Mice were anesthetized and shaved dorsally so that both flanks were well clear of fur. 2.5 x 10^6 MDA-MB-231 cells either stably expressing evolPLE or Fluc, or control cells, suspended in a 1.5:1 mix of 1x PBS and Matrigel (Corning) were injected subcutaneously in opposite flanks. After 24 h, the mice were anesthetized and given an IP injection of chemi-CM_2 or luciferin in 1x PBS with no greater than 2% DMSO. The mice were then allowed to recover for 15 min before beginning imaging. Imaging was performed using an IVIS SPECTRUM imager. Exposure times were 60s or 10 min, with

an F-Stop of 2 and Binning: Medium. Images were analyzed using ImageJ (Wayne Rasband, NIH).

General synthetic procedures. All reactions were performed in dried glassware under an atmosphere of dry N_2. Silica gel P60 (SiliCycle) was used for column chromatography and Analytical Chromatography TLC Silica gel 60 F_{254} (Merck Millipore, Darmstadt, Germany) was used for analytical thin layer chromatography. Plates were visualized by fluorescence quenching under UV light or by staining with iodine. Other reagents were purchased from Sigma-Aldrich (St. Louis, MO), Alfa Aesar (Ward Hill, MA), EMD Millipore (Billerica, MA), Oakwood Chemical (West Columbia, SC), TCI (Tokyo, Japan) and used without further purification. ^{1}H NMR and ^{13}C NMR spectra for characterization of new compounds and monitoring reactions were collected in $CDCl_3$ (Cambridge Isotope Laboratories, Cambridge, MA) on a JEOL 500 MHz spectrometer in the Department of Chemistry at Southern Methodist University. All chemical shifts are reported in the standard notation of parts per million using the peak of residual proton signals of the deuterated solvent as an internal reference. Coupling constant units are in Hertz (Hz) Splitting patterns are indicated as follows: br, broad; s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet; dd, doublet of doublets; dt, doublet of triplets. High resolution mass spectroscopy was performed on a Shimadzu IT-TOF (ESI source) and low resolution mass spectroscopy was performed on a Shimadzu LCMS-8050 Triple Quadrupole LCMS (ESI source) or a Shimadzu Matrix Assisted Laser Desorption/Ionization MS (MALDI) at

the Shimadzu Center for Advanced Analytical Chemistry at the University of Texas, Arlington. Synthesis of fluorescein-CM_2 was carried out using literature procedure.[188]

(3-(((1*r*,3*r*,5*R*,7*S*)-adamantan-2-ylidene)(methoxy)methyl)-2-chloro-6-((*E*)-2-cyanovinyl)phenoxy)methyl 1-methylcyclopropane-1-carboxylate (1). Chloromethyl 1-methylcyclopropanecarboxyate (116.3 mg, 0.7832 mmol, 1.5 equiv) was added to the dry, N_2 filled flask, then dissolved with 2.1 mL anhydrous acetone. NaI (128.6 mg, 0.8580 mmol, 1.7 equiv) was added to the solvent and the mixture was stirred for 24 h at ambient temperature. The reaction was concentrated under reduced pressure. Purification by silica column chromatography (CH_2Cl_2) yielded a pale yellow oil. (*E*)-3-(4-(((1*r*,3*r*,5*R*,7*S*)-adamantan-2-ylidene)(methoxy)methyl)-3-chloro-2-hydroxyphenyl)acrylonitrile (184.4 mg, 0.5182 mmol, 1.0 equiv) was dissolved with 2.0 mL anhydrous DMF in an dry, N_2 filled flask and anhydrous *N*,*N*-diisopropylethylamine (DIPEA, 0.19 mL, 1.1 mmol, 2.1 equiv) was added. Iodomethyl 1-methylcyclopropane-1-carboxylate was dissolved with 3.0 mL anhydrous DMF and added to the flask. The reaction mixture was stirred for 21 h at ambient temperature. The reaction was concentrated under reduced pressure. Purification by silica column chromatography (1:20 EtOAc/Hexane) yielded **1** as a white solid (107.4 mg, 44%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.63 (d, 1H, *J* = 16.6 Hz), 7.36 (d, 1H, *J* = 8.1 Hz), 7.11 (d, 1H, *J* = 8.1 Hz), 5.95 (d, 1H, *J* = 16.6 Hz), 5.70 (m, 2H), 3.29 (s, 1H), 3.25 (s, 1H), 2.04 (s, 1H), 1.94–1.63 (m, 12H), 1.28 (s, 3H), 1.23 (m, 2H), 0.77 (m, 2H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 175.07, 152.02, 144.76, 139.33, 139.04, 133.30, 129.21, 128.62, 128.54, 123.91, 117.83, 98.59, 89.32, 57.44, 39.19, 39.03, 38.61, 38.56,

36.96, 32.91, 31.59, 29.73, 28.29, 28.11, 19.10, 18.54, 17.60; HRMS calcd for $C_{27}H_{30}ClNO_4$ ($M+Na^+$) 490.1756, found 490.1755.

(2-chloro-6-((*E*)-2-cyanovinyl)-3-((1*r*,3*r*,5*r*,7*r*)-4'-methoxyspiro[adamantane-2,3'-[1,2]dioxetan]-4'-yl)phenoxy)methyl 1-methylcyclopropane-1-carboxylate (**chemi-CM_2**). Enol ether **1** (107.4 mg, 0.2295 mmol, 1.0 equiv) was dissolved with 5 mL THF at 0 °C in a two-neck flask and Rose bengal (12.4 mg, 0.0122 mmol, 0.050 equiv) was added to the solvent. O_2 was bubbled through the solvent when illuminated with a 120W light bulb (Home Depot, Dallas, TX). The reaction was monitored by TLC. After 3 h 15 min, the mixture was concentrated under reduced pressure. Purification by silica column chromatography (1:20 EtOAc/Hexane) yielded chemi-CM_2 as a white solid (76.9 mg, 67%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.97 (d, 1H, *J* = 8.6 Hz), 7.66 (d, 1H, *J* = 16.7 Hz), 7.52 (d, 1H, *J* = 8.6 Hz), 6.03 (d, 1H, *J* = 17.2 Hz), 5.67 (dd, 2H), 3.21 (s, 3H), 3.01 (s, 1H), 1.96 (s, 1H), 1.85–1.58 (m, 12H), 1.28 (s, 3H), 1.23 (m, 2H), 0.79 (m, 2H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 175.11, 152.70, 144.42, 136.53, 130.51, 129.80, 127.19, 124.13, 117.63, 111.63, 99.99, 96.48, 89.32, 49.87, 36.57, 33.99, 33.66, 32.67, 32.24, 31.62, 31.58, 29.82, 26.17, 25.83, 19.18, 18.59, 17.74; HRMS calcd for $C_{27}H_{30}ClNO_6$ ($M+Na^+$) 522.1654, found 522.1654.

A.6 Supplemental Information

Table A.1 List of Plasmids used in this work.

Plasmid:	Description:	Benchling Link:
cmv WT PLE	cmv driven intracellular expression of wt PLE	https://benchling.com/s/seq-zu3h2l9Cvg9aMSnYYrb5
cmv evolPLE	cmv driven intracellular expression of evolPLE	https://benchling.com/s/seq-PtGRuYqyLyLznLiGbr4C
pcDNA3-Igk-evolPLE-HA-DAF-IRES-mCherry	membrane-localized evolPLE	https://benchling.com/s/seq-CmGBKBtcCVnQyZ9ZH3kb
pmirGLO	pgk driven intracellular expression of Fluc	https://benchling.com/s/seq-GV2I5Lu7QHCQIAG38ygH

Table A.2 Gibson Assembly NNK Primers used to generate PLE mutant libraries.

NNK Pos.(s)	Primer (forward)
E221	*TCCTGGCAGTGTCACGATTTTTGGG**NNK**TCCGCAGGCGGAGAGTCCGTA*
G468, V474	*ACCTAAGTCAGTTATCGGCGATCAC**NNK**GATGAGATCTTCTCT**NNK**TTCG GATT*
L358, M362	*TCAATAAGCAAGAATTCGGGTGG**NNK**TTGCCCACA**NNK**ATGGGCTT*
L358, M363	*TCAATAAGCAAGAATTCGGGTGG**NNK**TTGCCCACAATG**NNK**GGCTT*
M362, M363	*AGAATTCGGGTGGCTGTTGCCCACA**NNKNNK**GGCTTTCCACTGTCTGA*

Figure A.3 Imaging analysis of WT PLE and evolPLE *in cellulo*. MDA-MB-231 cells were transfected with an empty vector. Cells were nuclear stained with Hoechst33342 and loaded with DMSO carrier control. After 20min, images were gathered using an inverted epifluorescence microscope.

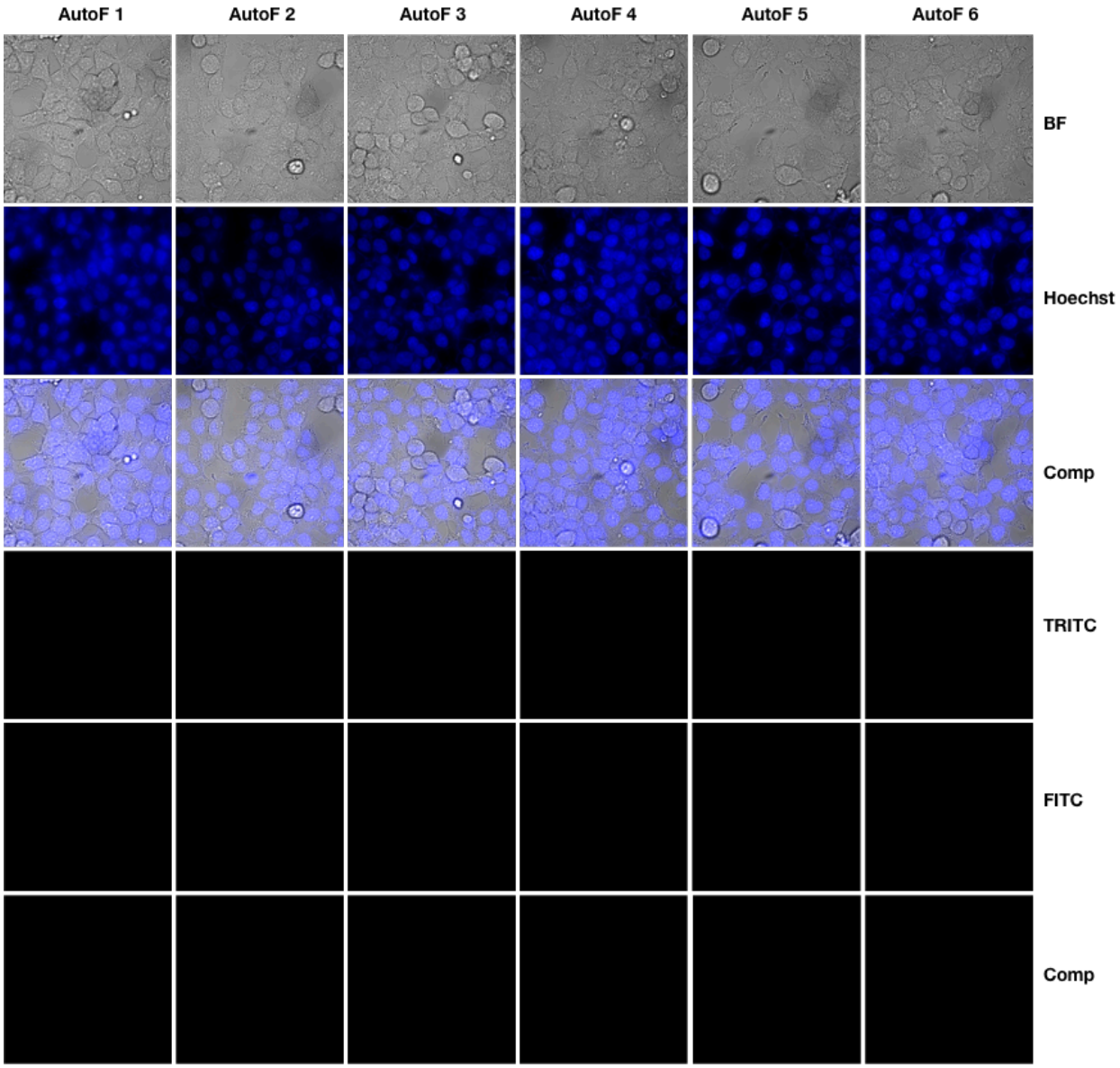


Figure A.3 – continued from previous page

Imaging analysis of WT PLE and evolPLE *in cellulo*. MDA-MB-231 cells were transfected with an empty vector. Cells were nuclear stained with Hoechst33342 and loaded with fluorescein-CM_2 (10uM). After 20min, images were gathered using an inverted epifluorescence microscope.

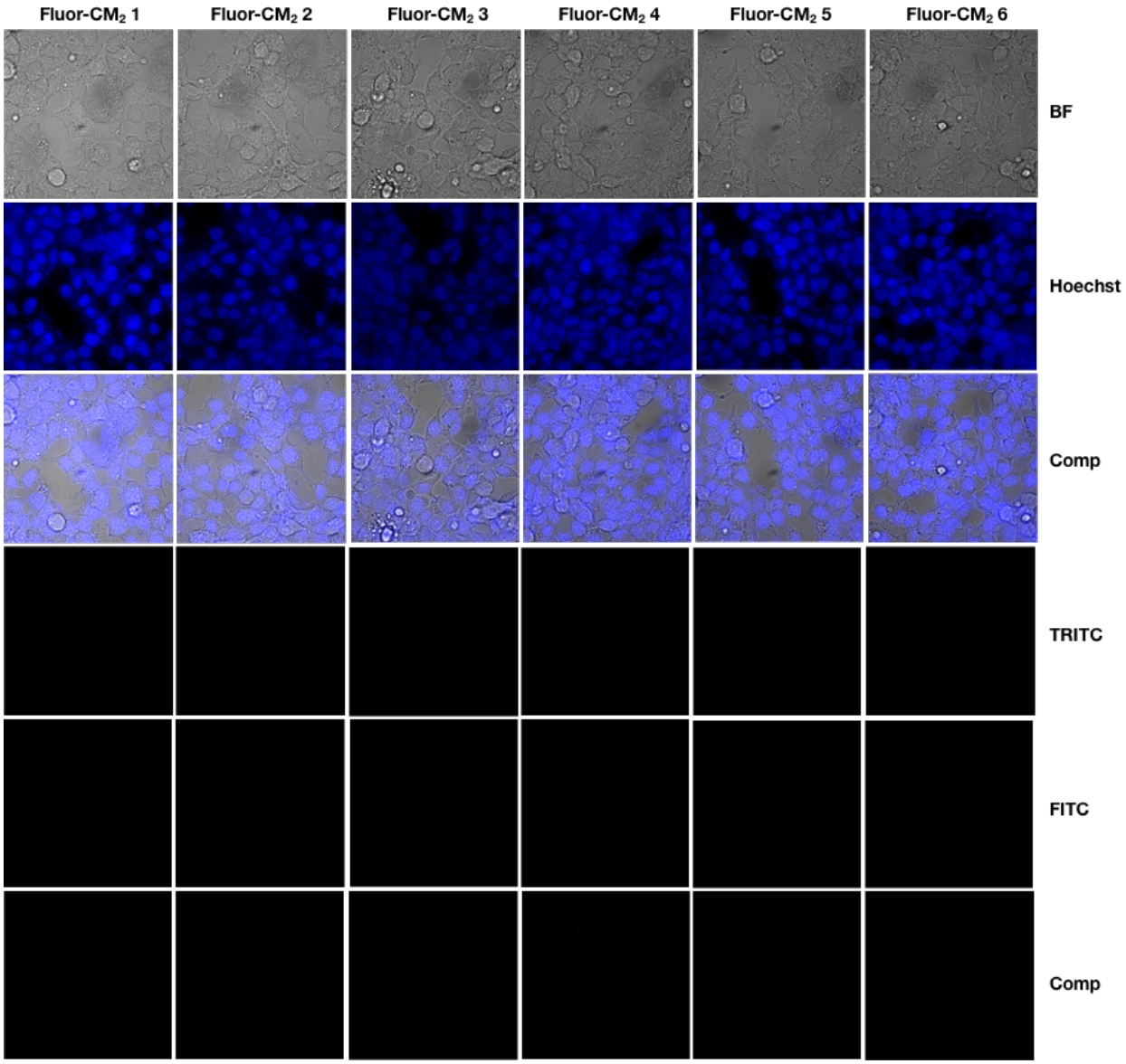


Figure A.3 – continued from previous page

Imaging analysis of WT PLE and evolPLE *in cellulo*. MDA-MB-231 cells were transfected with WT PLE. RFP transfection control was used. Cells were nuclear stained with Hoechst33342 and loaded with fluorescein-CM_2 (10uM). After 20min, images were gathered using an inverted epifluorescence microscope.

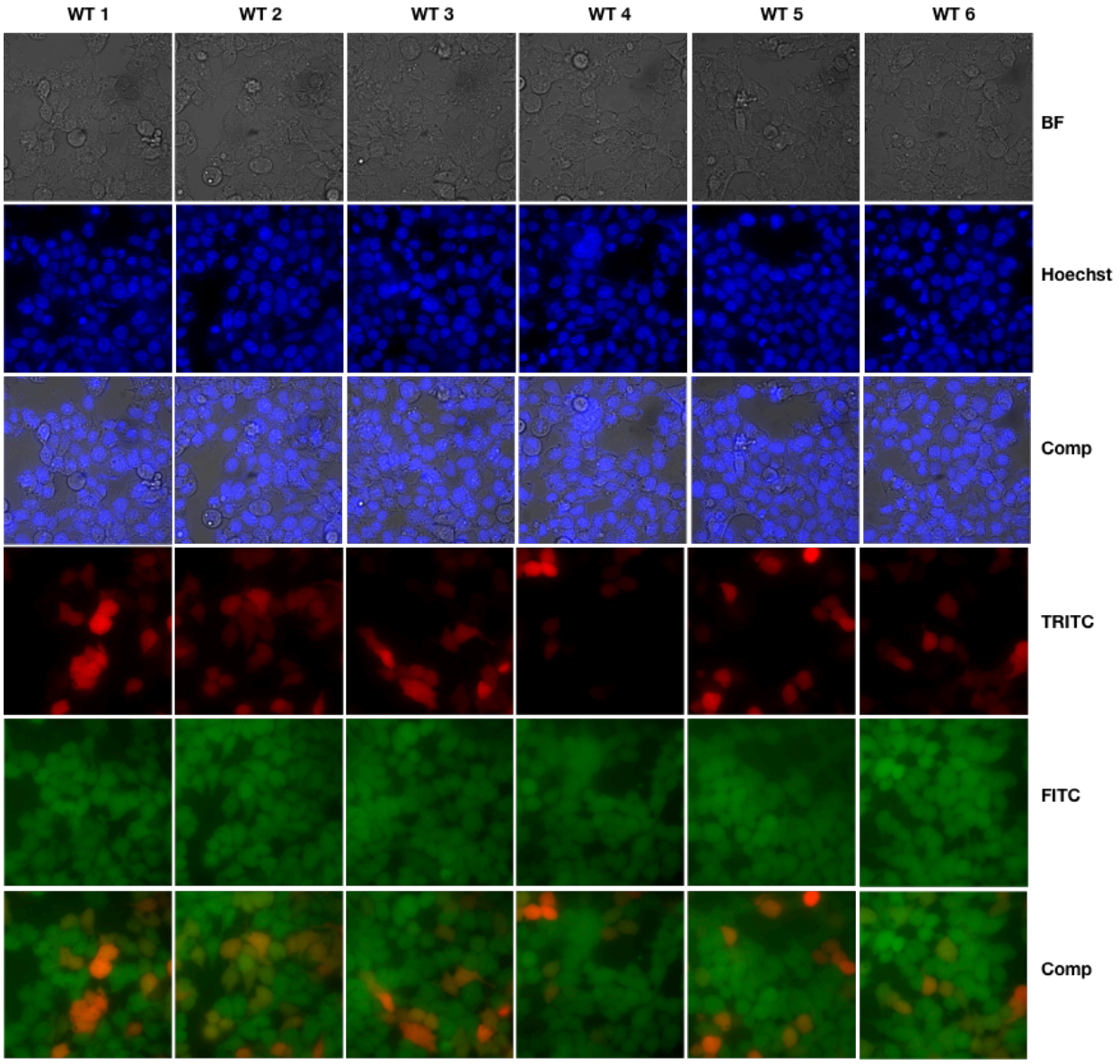


Figure A.3 – continued from previous page

Imaging analysis of WT PLE and evolPLE *in cellulo*. MDA-MB-231 cells were transfected with evolPLE. RFP transfection control was used. Cells were nuclear stained with Hoechst33342 and loaded with fluorescein-CM_2 (10uM). After 20min, images were gathered using an inverted epifluorescence microscope.

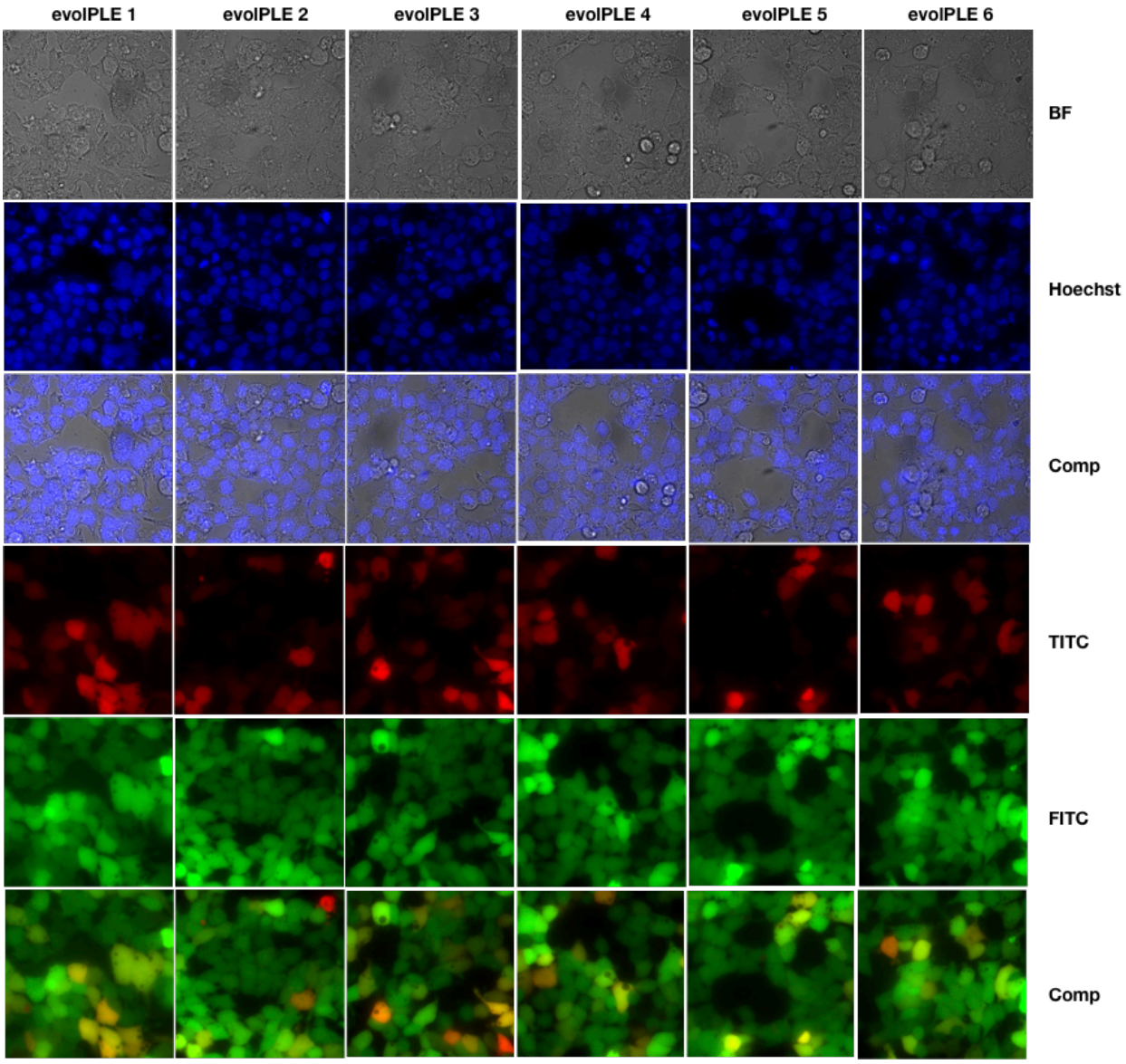


Figure A.4 evolPLE sensitivity. Mixtures of MBA-MB-231 cells and MBA-MB-231 cells expressing evolPLE were added to each well of a 96-well assay plate, with a constant number of 1 x 10^5 cells added to each well, varying the percentage of evolPLE expressing cells from 0% to 100%. 0.5, 5.0, or 50 μM Chemilum-CM was then added to each well and the plate imaged for 30 min. Well RLUs were quantified in ImageJ.

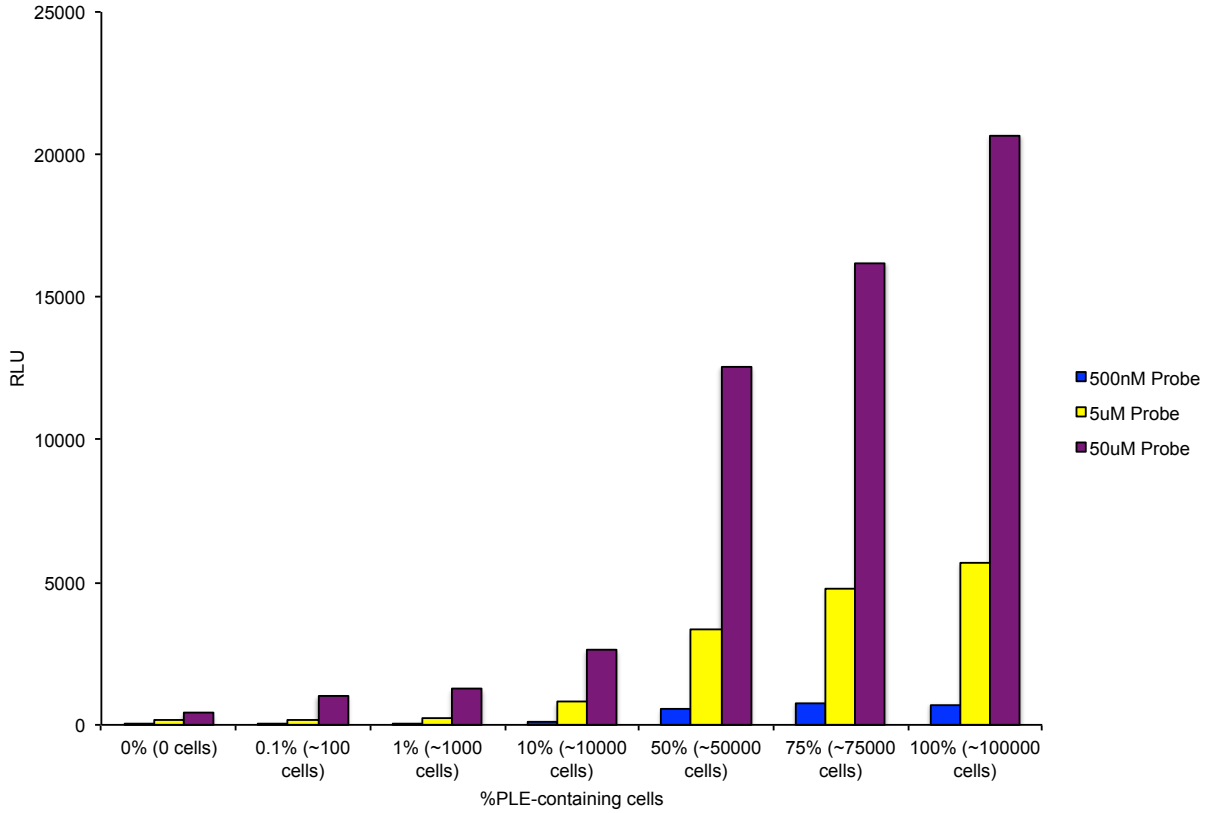


Figure A.5 Dose Response of evolPLE. ~1 x 10^5 MBA-MD-231 cells expressing either evolPLE or Fluc were added to each well of a 96-well plate. Increasing concentrations of either Chemilum-CM or D-luciferin were added, and the plate was imaged for 30min. Well RLUs were quantified in ImageJ.

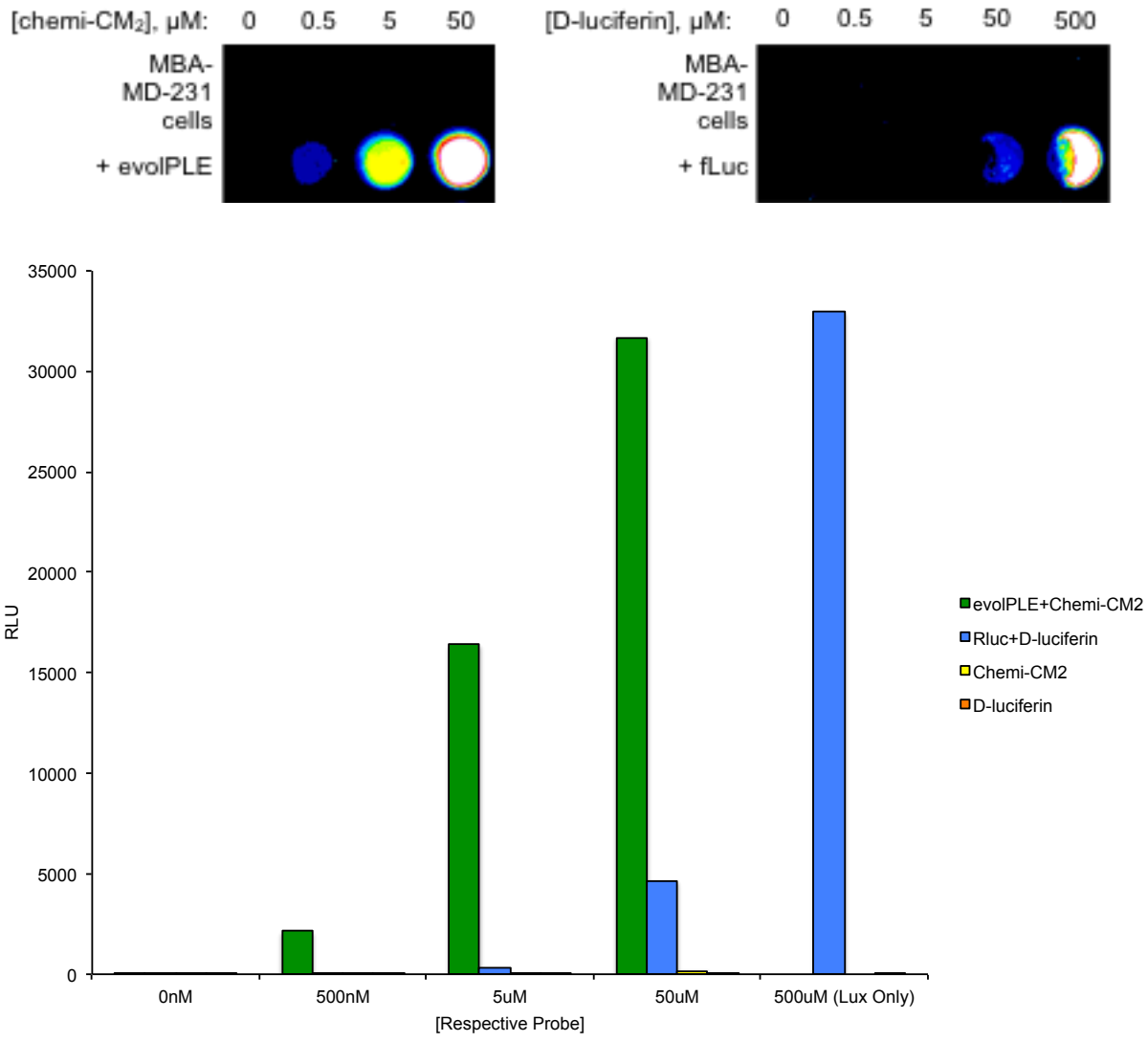
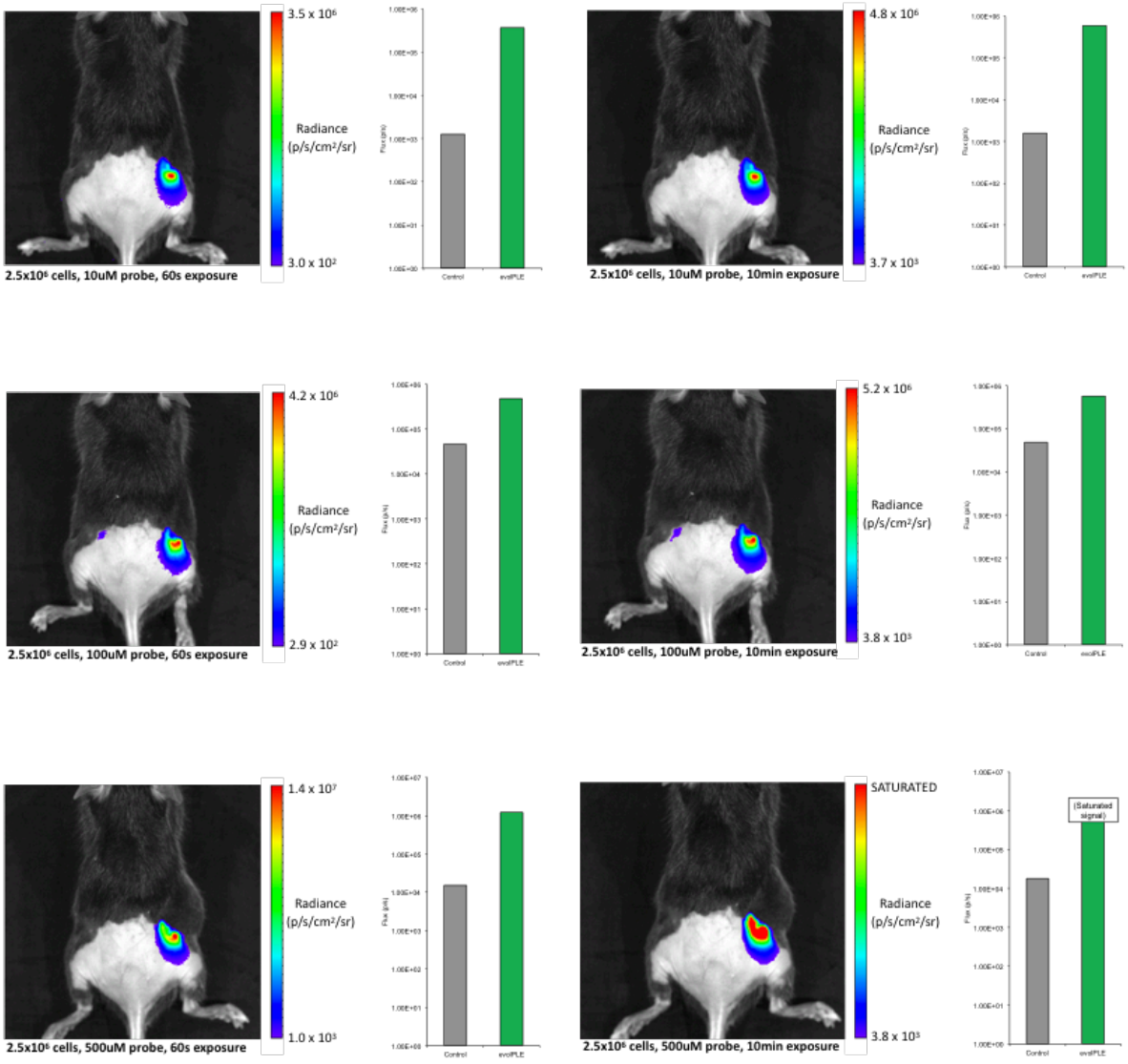


Figure A.6 evolPLE dose response in vivo. C57BL/6J mice were subcutaneously injected with ~2.5 x 10^6 MDA-MB-231 cells either expressing evolPLE or control. 24h later, an IP injection of 10, 100, or 500 uM Chemilum-CM was administered. The mice were then imaged in an IVIS at exposure times of 60s or 10min. Images, radiance, and flux values were obtained from IVIS software. **This signal determined to correspond to liver uncaging and not evolPLE tumor-mediated uncaging as evident by the consistent signal shape regardless of specific tumor location.**



REFERENCES

1. Buxbaum, A.R.; Haimovich, G.; Singer, R. H. In the right place at the right time: visualizing and understanding mRNA localization. *Nat Rev Mol Cell Biol.* **2015**, *16*, 95-109.

2. Bashirullah, A.; Cooperstock, R. L.; Lipshitz, H. D. RNA localization in development. *Annu Rev Biochem.* **1998**, *67*, 335-394.

3. Lipshitz, H. D.; Smibert, C. A. Mechanisms of RNA localization and translational regulation. *Curr Opin Cenet Dev.* **2000**, *10*, 476-488.

4. Ryan, D. P.; Matthews, J. M. Protein-protein interactions in human disease. *Curr Opin Struct Biol.* **2005**, *15*(4), 441-446.

5. Kolch, W. Meaningful relationships: The regulation of the Ras/Raf/MEK/ERK pathway by protein interactions. *Biochem J* **2000**, *351*, 289−305.

6. Sudhof, T. C. The synaptic cycle - A cascade of protein-protein interactions. *Nature* **1995**, *375* (6533), 645−653.

7. Ghosh, I.; Shekhawat, S. S. Split-protein systems: beyond binary protein-protein interactions. *Curr Opin Chem Biol*. **2011**, *15*(6), 789-797.

8. Jones, S.; Thornton, J. M. Principles of protein-protein interactions. *Proc Natl Acad Sci USA* **1996**, *93* (1), 13−20.

9. Kerppola, T. K. Visualization of molecular interactions using bimolecular fluorescence complementation analysis: Characteristics of protein fragment complementation. *Chem Soc Rev* **2009**, 38 (10), 2876−2886.

10. Kodama, Y.; Hu, C. D. Bimolecular fluorescence complementation (BiFC): A 5-year update and future perspectives. *BioTechniques* **2012**, 53 (5), 285−298.

11. Martell, J. D.; Yamagata, M.; Deerinck, T. J.; Phan, S.; Kwa, C. G.; Ellisman, M. H.; Sanes, J. R.; Ting, A. Y. A split horseradish peroxidase for the detection of intercellular protein-protein interactions and sensitive visualization of synapses. *Nat Biotechnol* **2016**, 34 (7), 774−780.

12. Kim, C. K.; Cho, K. F.; Kim, M. W.; Ting, A. Y. Luciferase- LOV BRET enables versatile and specific transcriptional readout of cellular protein-protein interactions. *eLife* **2019**, 8, 21.

13. Azad, T.; Tashakor, A.; Hosseinkhani, S. Split-luciferase complementary assay: Applications, recent developments, and future perspectives. *Anal Bioanal Chem* **2014**, 406 (23), 5541−5560.

14. Pu, J.; Zinkus-Boltz, J.; Dickinson, B. C. Evolution of a split RNA polymerase as a versatile biosensor platform. *Nat Chem Biol*. **2017**, *13*(4), 432-438.

15. Pu, J.; Dewey, J. A.; Hadji, A.; LaBelle, J. L.; Dickinson, B. C. RNA polymerase tags to monitor multidimensional protein-protein interactions reveal pharmacological engagement of Bcl-2 proteins. *J Am Chem Soc*. **2017**, *139*(34), 11964-11972.

16. Pu, J.; Kentala, K.; Dickinson, B. C. Multidimensional control of Cas9 by evolved RNA polymerase-based biosensors. *ACS Chem Biol*. **2018**, *13*(2), 431-437.

17. Daringer, N. M.; Dudek, R. M.; Schwarz, K. A.; Leonard, J. N. Modular extracellular sensor architecture for engineering mammalian cell-based devices. *ACS Synth Biol* **2014**, 3 (12), 892−902.

18. Fink, T.; Lonzaric, J.; Praznik, A.; Plaper, T.; Merljak, E.; Leben, K.; Jerala, N.; Lebar, T.; Strmsek, Z.; Lapenta, F.; Bencina, M.; Jerala, R. Design of fast proteolysis-based signaling and logic circuits in mammalian cells. *Nat Chem Biol* **2019**, 15 (2), 115−122.

19. Zetsche, B.; Volz, S. E.; Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol* **2015**, 33 (2), 139−142.

20. Seath, C. P.; Trowbridge, A. D.; Muir, T. W.; MacMillian, D. W. C. Reactive intermediates for interactome mapping. *Chem Soc Rev.* **2021**, *50*, 2911-2926.

21. Qin, W.; Cho, K. F.; Cavanagh, P. E.; Ting, A. Y. Deciphering molecular interactions by proximity labeling. *Nat Methods*. **2021**, *18*(2), 133-143.

22. Kwon, K.; Beckett, D. Function of a conserved sequence motif in biotin haloenzyme synthetases. *Protein Sci.* **2000**, *9*(8), 1530-1539.

23. Roux, K. J.; Kim, D. I.; Raida, M.; Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol*. **2012**, *196*(6), 801-810.

24. Branon, T. C.; Bosch, J. A.; Sanchez, A. D.; Udeshi, N. D.; Svinkina, T.; Carr, S. A.; Feldman, J. L.; Perrimon, N.; Ting, A. Y. Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol.* **2018**, *36*(9), 880-887.

25. Rhee, H.-W.; Zou, P.; Udeshi, N. D.; Martell, J. D.; Mootha, V. K.; Carr, S. A.; Ting, A. Y. Proteomic mapping of mitochondria in living cells via spatially restricted enzyme tagging. *Science.* **2013**, *339*(6125), 1328-1331.

26. Martell, J. D.; Deerinck, T. J.; Sancak, Y.; Poulos, T. L.; Mootha, V. K.; Sosinsky, G. E.; Ellisman, M. H.; Ting, A. Y. Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat Biotechnol.* **2012**, *30*(11), 1143-1148.

27. Lam, S. S.; Martell, J. D.; Kamer, K. J.; Deerinck, T. J.; Ellisman, M. H.; Mootha, V. K.; Ting, A. Y. Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods.* **2015**, *12*(1), 51-54.

28. Geri, J. B.; Oakley, J. V.; Reyes-Robles, T.; Wang, T.; McCarver, S. J.; White, C. H.; Rodriguez-Rivera, F. P.; Parker Jr., D. L.; Hett, E. C.; Fadeyi, O. O.; Oslund, R. C.; MacMillan, D. W. C. Microenvironment mapping via dexter energy transfer on immune cells. *Science.* **2020**, *367*(6482), 1091-1097.

29. Kallat, M.; Resetca, D.; Lourenco, C.; Chan, P.-K.; Wei, Y.; Shiah, Y.-J.; Vitkin, N.; Tong, Y.; Sunnerhagen, M.; Done, S. J.; Boutros, P. C.; Raught, B.; Penn, L. Z. MYC protein interactome profiling reveals functionally distinct regions that cooperate to drive tumorigenesis. *Mol Cell*. **2018**, *72*(5), 836-848.

30. Cho, K. F.; Branon, T. C.; Rajeev, S.; Svinkina, T.; Udeshi, N. D.; Thoudam, T.; Kwak, C.; Rhee, H.-W.; Lee, I.-K.; Carr, S. A.; Ting, A. Y. Split-TurboID enables contact-dependent proximity labeling in cells. *Proc Natl Acad Sci U S A.* **2020**, *117*(22), 12143-12154.

31. Paek, J.; Kalocsay, M.; Staus, D. P.; Wingler, L.; Pascolutti, R.; Paulo, J. A.; Gygi, S. P.; Kruse, A. C. Multidimensional tracking of GPCR signaling via peroxidase-catalyzed proximity labeling. *Cell.* **2017**, *169*(2), 338-349.

32. Lobingier, B. T.; Hüttenhain, R.; Eichel, K.; Miller, K. B.; Ting, A. Y.; von Zastrow, M.; Krogan, N. J. An approach to spatiotemporally resolve protein interaction networks in living cells. *Cell.* **2017**, *169*(2), 350-360.

33. Myers, S. A.; Wright, J.; Peckner, R.; Kallish, B. T.; Zhang, F.; Carr, S. A. Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. *Nat Methods.* **2018**, *15*(6), 437-439.

34. Gao, X. D.; Tu, L.-C.; Mir, A.; Rodriguez, T.; Ding, Y.; Leszyk, J.; Dekker, J.; Shaffer, S. A.; Zhu, L. J.; Wolfe, S. A.; Sontheimer, E. J. C-BERST: Defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. *Nat Methods.* **2018**, *15*(6), 433-436.

35. Dixon, A. S.; Schwinn, M. K.; Hall, M. P.; Zimmerman, K.; Otto, P.; Lubben, T. H.; Butler, B. L.; Binkowski, B. F.; Machleidt, T.; Kirkland, T. A.; Wood, M. G.; Eggers, C. T.; Encell, L. P.; Wood, K. V. NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells. *ACS Cehm Biol.* **2016**, *11*(2), 400-408.

36. Rathbun, C. M.; Porterfield, W. B.; Jones, K. A.; Sagoe, M.; Reyes, M. R.; Hua, C. T.; Prescher, J. A. Parallel screening for rapid indentification of orthogonal bioluminescent tools. *ACS Cent Sci.* **2017**, *3*(12), 1254-1261.

37. Fazal, F. M.; Han, S.; Parker, K. R.; Kaewsapsak, P.; Xu, J.; Boettiger, A. N.; Chang, H. Y.; Ting, A. Y. Atlas of subcellular RNA localization revealed by APEX-Seq. *Cell*. **2019**, *178*, 473-490.

38. Wang, P.; Tang, W.; Li, Z.; Zou, Z.; Zhou, Y.; Li, R.; Xiong, T.; Wang, J.; Zou, P. Mapping spatial transcriptome with light-activated proximity-dependent RNA labeling. *Nat Chem Biol.* **2019**, *15*(11), 1110-1119.

39. Engel, K. L.; Lo, H.-Y. G.; Goering, R.; Li, Y.; Spitale, R. C.; Taliaferro, J. M. Analysis of subcellular transcriptomes by RNA proximity labeling with Halo-seq. *bioRxiv*, doi: 10.1101/2021.06.08.447604.

40. Doudna, J. A.; Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **2014**, *346*, 1258096.

41. Hsu, P. D.; Lander, E. S.; Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **2014**, *157*, 1262−1278.

42. Dominguez, A. A.; Lim, W. A.; Qi, L. S. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat Rev Mol Cell Biol* **2015**, *17*, 5−15.

43. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012**, *337*, 816−821.

44. Nunez, J. K.; Harrington, L. B.; Doudna, J. A. Chemical and Biophysical Modulation of Cas9 for Tunable Genome Engineering. *ACS Chem Biol* **2016**, *11*, 681−688.

45. Davis, K. M.; Pattanayak, V.; Thompson, D. B.; Zuris, J. A.; Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat Chem Biol* **2015**, *11*, 316−318.

46. Wright, A. V.; Sternberg, S. H.; Taylor, D. W.; Staahl, B. T.; Bardales, J. A.; Kornfeld, J. E.; Doudna, J. A. Rational design of a split-Cas9 enzyme complex. *Proc Natl Acad Sci USA* **2015**, *112*, 2984−2989.

47. Zetsche, B.; Volz, S. E.; Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol* **2015**, *33*, 139−142.

48. Nguyen, D. P.; Miyaoka, Y.; Gilbert, L. A.; Mayerl, S. J.; Lee, B. H.; Weissman, J. S.; Conklin, B. R.; Wells, J. A. Ligand- binding domains of nuclear receptors facilitate tight control of split CRISPR activity. *Nat Commun* **2016**, *7*, 12009.

49. Lee, M.; Li, J.; Liang, Y.; Ma, G.; Zhang, J.; He, L.; Liu, Y.; Li, Q.; Li, M.; Sun, D.; Zhou, Y.; Huang, Y. Engineered Split- TET2 Enzyme for Inducible Epigenetic Remodeling. *J Am Chem Soc* **2017**, *139*, 4659−4662.

50. Oakes, B. L.; Nadler, D. C.; Flamholz, A.; Fellmann, C.; Staahl, B. T.; Doudna, J. A.; Savage, D. F. Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat Biotechnol* **2016**, *34*, 646−651.

51. Richter, F.; Fonfara, I.; Bouazza, B.; Schumacher, C. H.; Bratovic, M.; Charpentier, E.; Moglich, A. Engineering of temperature- and light-switchable Cas9 variants. *Nucleic Acids* Res **2016**, *44*, 10003−10014.

52. Maji, B.; Moore, C. L.; Zetsche, B.; Volz, S. E.; Zhang, F.; Shoulders, M. D.; Choudhary, A. Multidimensional chemical control of CRISPR-Cas9. *Nat Chem Biol* **2016**, *13*, 9−11.

53. Luo, J.; Liu, Q. Y.; Morihiro, K.; Deiters, A. Small- molecule control of protein function through Staudinger reduction. *Nat Chem* **2016**, *8*, 1027−1034.

54. Burstein, D.; Harrington, L. B.; Strutt, S. C.; Probst, A. J.; Anantharaman, K.; Thomas, B. C.; Doudna, J. A.; Banfield, J. F. New CRISPR-Cas systems from uncultivated microbes. *Nature* **2016**, *542*, 237−241.

55. Shmakov, S.; Abudayyeh, O. O.; Makarova, K. S.; Wolf, Y. I.; Gootenberg, J. S.; Semenova, E.; Minakhin, L.; Joung, J.; Konermann, S.; Severinov, K.; Zhang, F.; Koonin, E. V. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell* **2015**, *60*, 385−397.

56. Liu, Y.; Zhan, Y.; Chen, Z.; He, A.; Li, J.; Wu, H.; Liu, L.; Zhuang, C.; Lin, J.; Guo, X.; Zhang, Q.; Huang, W.; Cai, Z. Directing cellular information flow via CRISPR signal conductors. *Nat Methods* **2016**, *13*, 938−944.

57. Ferry, Q. R.; Lyutova, R.; Fulga, T. A. Rational design of inducible CRISPR guide RNAs for de novo assembly of transcriptional programs. *Nat Commun* **2017**, *8*, 14633.

58. Pu, J.; Chronis, I.; Ahn, D.; Dickinson, B. C. A Panel of Protease-Responsive RNA Polymerases Respond to Biochemical Signals by Production of Defined RNA Outputs in Live Cells. *J Am Chem Soc* **2015**, *137*, 15996−15999.

59. Pu, J.; Zinkus-Boltz, J.; Dickinson, B. C. Evolution of a split RNA polymerase as a versatile biosensor platform. *Nat Chem Biol* **2017**, *13*, 432−438.

60. Liang, F. S.; Ho, W. Q.; Crabtree, G. R. Engineering the ABA plant stress pathway for regulation of induced proximity. *Sci Signaling* **2011**, *4*, rs2.

61. Ran, F. A.; Cong, L.; Yan, W. X.; Scott, D. A.; Gootenberg, J. S.; Kriz, A. J.; Zetsche, B.; Shalem, O.; Wu, X.; Makarova, K. S.; Koonin, E. V.; Sharp, P. A.; Zhang, F. In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **2015**, *520*, 186−191.

62. Esvelt, K. M.; Carlson, J. C.; Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **2011**, *472*, 499− 503.

63. Chipuk; J. E.; Moldoveanu, T.; Llambi, F.; Parsons, M. J.; Green, D. R. The BCL-2 family reunion. *Mol Cell* **2010**, *37*, 299− 310.

64. Pu, J.; Dewey, J.; Hadji, A.; LaBelle, J. L.; Dickinson, B. C. RNA Polymerase tags to monitor multidimensional protein- protein interactions reveal pharmacological engagement of Bcl-2 proteins. *J Am Chem Soc* **2017**, *139*, 11964-11972.

65. Souers, A. J.; Leverson, J. D.; Boghaert, E. R.; Ackler, S. L.; Catron, N. D.; Chen, J.; Dayton, B. D.; Ding, H.; Enschede, S. H.; Fairbrother, W. J.; Huang, D. C.; Hymowitz, S. G.; Jin, S.; Khaw, S. L.; Kovar, P. J.; Lam, L. T.; Lee, J.; Maecker, H. L.; Marsh, K. C.; Mason, K. D.; Mitten, M. J.; Nimmer, P. M.; Oleksijew, A.; Park, C. H.; Park, C. M.; Phillips, D. C.; Roberts, A. W.; Sampath, D.; Seymour, J. F.; Smith, M. L.; Sullivan, G. M.; Tahir, S. K.; Tse, C.; Wendt, M. D.; Xiao, Y.; Xue, J. C.; Zhang, H.; Humerickhouse, R. A.; Rosenberg, S. H.; Elmore, S. W. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat Med* **2013**, *19*, 202−208.

66. Petris, G.; Casini, A.; Montagna, C.; Lorenzin, F.; Prandi, D.; Romanel, A.; Zasso, J.; Conti, L.; Demichelis, F.; Cereseto, A. Hit and go CAS9 delivered through a lentiviral based self- limiting circuit. *Nat Commun* **2017**, *8*, 15334.

67. Frommer, W. B.; Davidson, M. W.; Campbell, R. E. Genetically encoded biosensors based on engineered fluorescent proteins. *Chem Soc Rev* **2009**, *38*, 2833−2841.

68. Oldach, L.; Zhang, J. Genetically encoded fluorescent biosensors for live-cell visualization of protein phosphorylation. *Chem Biol* **2014**, *21*, 186−197.

69. Perez-Pinera, P.; Kocak, D. D.; Vockley, C. M.; Adler, A. F.; Kabadi, A. M.; Polstein, L. R.; Thakore, P. I.; Glass, K. A.; Ousterout, D. G.; Leong, K. W.; Guilak, F.; Crawford, G. E.; Reddy, T. E.; Gersbach, C. A. RNA-guided gene activation by CRISPR-Cas9- based transcription factors. *Nat Methods* **2013**, *10*, 973−976.

70. Ringquist, S.; Shinedling, S.; Barrick, D.; Green, L.; Binkley, J.; Stormo, G. D.; Gold, L. Translation initiation in Escherichia coli: sequences within the ribosome-binding site, *Mol Microbiol* **1992**, *6*, 1219-1229.

71. Cheetham, G. M.; Steitz, T. A.Structure of a transcribing T7 RNA polymerase initiation complex, *Science* **1999**, *286*, 2305-2309.

72. Braun, P. Interactome mapping for analysis of complex phenotypes: Insights from benchmarking binary interaction assays. *Proteomics* **2012**, *12* (10), 1499−1518.

73. Jones, S.; Thornton, J. M. Principles of protein-protein interactions. *Proc Natl Acad Sci USA* **1996**, *93* (1), 13−20.

74. Sudhof, T. C. The synaptic cycle - A cascade of protein-protein interactions. *Nature* **1995**, *375* (6533), 645−653.

75. Kolch, W. Meaningful relationships: The regulation of the Ras/Raf/MEK/ERK pathway by protein interactions. *Biochem J* **2000**, *351*, 289−305.

76. Arkin, M. R.; Tang, Y. Y.; Wells, J. A. Small-molecule inhibitors of protein-protein interactions: Progressing toward the reality. *Chem Biol* **2014**, *21* (9), 1102−1114.

77. Ran, X.; Gestwicki, J. E. Inhibitors of protein-protein interactions (PPIs): An analysis of scaffold choices and buried surface area. *Curr Opin Chem Biol* **2018**, 44, 75−86.

78. Wells, J. A.; McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **2007**, 450 (7172), 1001−1009.

79. Daringer, N. M.; Dudek, R. M.; Schwarz, K. A.; Leonard, J. N. Modular extracellular sensor architecture for engineering mammalian cell-based devices. *ACS Synth Biol* **2014**, 3 (12), 892−902.

80. Fink, T.; Lonzaric, J.; Praznik, A.; Plaper, T.; Merljak, E.; Leben, K.; Jerala, N.; Lebar, T.; Strmsek, Z.; Lapenta, F.; Bencina, M.; Jerala, R. Design of fast proteolysis-based signaling and logic circuits in mammalian cells. *Nat Chem Biol* **2019**, 15 (2), 115−122.

81. Zetsche, B.; Volz, S. E.; Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol* **2015**, 33 (2), 139−142.

82. Pu, J. Y.; Zinkus-Boltz, J.; Dickinson, B. C. Evolution of a split RNA polymerase as a versatile biosensor platform. *Nat Chem Biol* **2017**, 13 (4), 432−438.

83. Pu, J. Y.; Dewey, J. A.; Hadji, A.; LaBelle, J. L.; Dickinson, B. C. RNA polymerase tags to monitor multidimensional protein-protein interactions reveal pharmacological engagement of Bcl-2 proteins. *J Am Chem Soc* **2017**, 139 (34), 11964−11972.

84. Pu, J. Y.; Kentala, K.; Dickinson, B. C. Multidimensional control of Cas9 by evolved RNA polymerase-based biosensors. *ACS Chem Biol* **2018**, 13 (2), 431−437.

85. Pu, J.; Disare, M.; Dickinson, B. C. Evolution of c-terminal modification tolerance in full-length and split T7 RNA polymerase biosensors. *ChemBioChem* **2019**, 20 (12), 1547−1553.

86. Kim, C. K.; Cho, K. F.; Kim, M. W.; Ting, A. Y. Luciferase- LOV BRET enables versatile and specific transcriptional readout of cellular protein-protein interactions. *eLife* **2019**, 8, 21.

87. Piehler, J. New methodologies for measuring protein interactions in vivo and in vitro. *Curr Opin Struct Biol* **2005**, 15 (1), 4−14.

88. Shekhawat, S. S.; Ghosh, I. Split-protein systems: beyond binary protein-protein interactions. *Curr Opin Chem Biol* **2011**, 15 (6), 789−797.

89. Kerppola, T. K. Visualization of molecular interactions using bimolecular fluorescence complementation analysis: Characteristics of protein fragment complementation. *Chem Soc Rev* **2009**, 38 (10), 2876−2886.

90. Remy, I.; Michnick, S. W. Mapping biochemical networks with protein fragment complementation assays. *Methods Mol Biol* **2015**, 1278, 467−481.

91. Azad, T.; Tashakor, A.; Hosseinkhani, S. Split-luciferase complementary assay: Applications, recent developments, and future perspectives. *Anal Bioanal Chem* **2014**, 406 (23), 5541−5560.

92. Kodama, Y.; Hu, C. D. Bimolecular fluorescence complementation (BiFC): A 5-year update and future perspectives. *BioTechniques* **2012**, 53 (5), 285−298.

93. Martell, J. D.; Yamagata, M.; Deerinck, T. J.; Phan, S.; Kwa, C. G.; Ellisman, M. H.; Sanes, J. R.; Ting, A. Y. A split horseradish peroxidase for the detection of intercellular protein-protein interactions and sensitive visualization of synapses. *Nat Biotechnol* **2016**, 34 (7), 774−780.

94. Tian, L.; Yang, Y. L.; Wysocki, L. M.; Arnold, A. C.; Hu, A.; Ravichandran, B.; Sternson, S. M.; Looger, L. L.; Lavis, L. D. Selective esterase-ester pair for targeting small molecules with cellular specificity. *Proc Natl Acad Sci USA* **2012**, 109 (13), 4756−4761.

95. Hay, M. P.; Anderson, R. F.; Ferry, D. M.; Wilson, W. R.; Denny, W. A. Synthesis and evaluation of nitroheterocyclic carbamate prodrugs for use with nitroreductase-mediated gene-directed enzyme prodrug therapy. *J Med Chem* **2003**, 46 (25), 5533−5545.

96. Gruber, T. D.; Krishnamurthy, C.; Grimm, J. B.; Tadross, M. R.; Wysocki, L. M.; Gartner, Z. J.; Lavis, L. D. Cell-specific chemical delivery using a selective nitroreductase-nitroaryl pair. *ACS Chem Biol* **2018**, 13 (10), 2888−2896.

97. Hay, M. P.; Wilson, W. R.; Denny, W. A. Nitroarylmethylcar- bamate prodrugs of doxorubicin for use with nitroreductase gene- directed enzyme prodrug therapy. *Bioorg Med Chem* **2005**, 13 (12), 4043−4055.

98. Porterfield, W. B.; Jones, K. A.; McCutcheon, D. C.; Prescher, J. A. A “caged” luciferin for imaging cell-cell contacts. *J Am Chem Soc* **2015**, 137 (27), 8656−8659.

99. Sellmyer, M. A.; Bronsart, L.; Imoto, H.; Contag, C. H.; Wandless, T. J.; Prescher, J. A. Visualizing cellular interactions with a generalized proximity reporter. *Proc Natl Acad Sci USA* **2013**, 110 (21), 8567−8572.

100. Galarneau, A.; Primeau, M.; Trudeau, L. E.; Michnick, S. W. Beta-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein-protein interactions. *Nat Biotechnol* **2002**, 20 (6), 619−622.

101. Phelan, R. M.; Ostermeier, M.; Townsend, C. A. Design and synthesis of a beta-lactamase activated 5-fluorouracil prodrug. *Bioorg Med Chem Lett* **2009**, 19 (4), 1261−1263.

102. Rossi, F.; Charlton, C. A.; Blau, H. M. Monitoring protein- protein interactions in intact eukaryotic cells by beta-galactosidase complementation. *Proc Natl Acad Sci USA* **1997**, 94 (16), 8405− 8410.

103. Kamal, A.; Tekumalla, V.; Krishnan, A.; Pal-Bhadra, M.; Bhadra, U. Development of pyrrolo 2,1-c 1,4 -benzodiazepine beta- galactoside prodrugs for selective therapy of cancer by ADEPT and PMT. *ChemMedChem* **2008**, 3 (5), 794−802.

104. Leenders, R. G. G.; Damen, E. W. P.; Bijsterveld, E. J. A.; Scheeren, H. W.; Houba, P. H. J.; van der Meulen-Muileman, I. H.; Boven, E.; Haisma, H. J. Novel anthracycline-spacer-beta-glucuronide, -beta-glucoside, and -beta-galactoside prodrugs for application in selective chemotherapy. *Bioorg Med Chem* **1999**, 7 (8), 1597−1610.

105. Wierdl, M.; Morton, C. L.; Weeks, J. K.; Danks, M. K.; Harris, L. C.; Potter, P. M. Sensitization of human tumor cells to CPT-11 via adenoviral-mediated delivery of a rabbit liver carboxylesterase. *Cancer Res* **2001**, 61 (13), 5078−5082.

106. Rehberg, M.; Lepier, A.; Solchenberger, B.; Osten, P.; Blum, R. A new non-disruptive strategy to target calcium indicator dyes to the endoplasmic reticulum. *Cell Calcium* **2008**, 44 (4), 386−399.

107. Yang, Y. L.; Lee, P.; Sternson, S. M. Cell type-specific pharmacology of NMDA receptors using masked MK801. *eLife* **2015**, 4, 12.

108. Beaumont, K.; Webster, R.; Gardner, I.; Dack, K. Design of ester prodrugs to enhance oral absorption of poorly permeable compounds: Challenges to the discovery scientist. *Curr Drug Metab* **2003**, 4 (6), 461−485.

109. Lavis, L. D. Ester bonds in prodrugs. *ACS Chem Biol* **2008**, 3 (4), 203−206.

110. Azzolini, M.; Mattarei, A.; La Spina, M.; Marotta, E.; Zoratti, M.; Paradisi, C.; Biasutto, L. Synthesis and evaluation as prodrugs of hydrophilic carbamate ester analogues of resveratrol. *Mol Pharmaceutics* **2015**, 12 (9), 3441−3454.

111. Landowski, C. P.; Song, X. Q.; Lorenzi, P. L.; Hilfinger, J. M.; Amidon, G. L. Floxuridine amino acid ester prodrugs: Enhancing Caco-2 permeability and resistance to glycosidic bond metabolism. *Pharm Res* **2005**, 22 (9), 1510−1518.

112. Schmidt, M.; Henke, E.; Heinze, B.; Kourist, R.; Hidalgo, A.; Bornscheuer, U. T. A versatile esterase from Bacillus subtilis: Cloning, expression, characterization, and its application in biocatalysis. *Biotechnol J* **2007**, 2 (2), 249−253.

113. Choi, J. W.; Chen, J.; Schreiber, S. L.; Clardy, J. Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP. *Science* **1996**, 273 (5272), 239−242.

114. Liang, F. S.; Ho, W. Q.; Crabtree, G. R. Engineering the ABA plant stress pathway for regulation of induced proximity. *Sci Signaling* **2011**, 4 (164), No. rs2.

115. Hodgman, C. E.; Jewett, M. C. Cell-free synthetic biology: Thinking outside the cell. *Metab Eng* **2012**, 14 (3), 261−269.

116. Ghosh, I.; Hamilton, A. D.; Regan, L. Antiparallel leucine zipper-directed protein reassembly: Application to the green fluorescent protein. *J Am Chem Soc* **2000**, 122 (23), 5658−5659.

117. Czabotar, P. E.; Lessene, G.; Strasser, A.; Adams, J. M. Control of apoptosis by the BCL-2 protein family: Implications for physiology and therapy. *Nat Rev Mol Cell Biol* **2014**, 15 (1), 49−63.

118. Cang, S. D.; Iragavarapu, C.; Savooji, J.; Song, Y. P.; Liu, D. L. ABT-199 (venetoclax) and BCL-2 inhibitors in clinical development. *J Hematol Oncol* **2015**, 8, 129.

119. Dixon, A. S.; Schwinn, M. K.; Hall, M. P.; Zimmerman, K.; Otto, P.; Lubben, T. H.; Butler, B. L.; Binkowski, B. F.; Machleidt, T.; Kirkland, T. A.; Wood, M. G.; Eggers, C. T.; Encell, L. P.; Wood, K. V. NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells. *ACS Chem Biol* **2016**, 11 (2), 400−408.

120. Lippert, A. R. Unlocking the potential of chemiluminescence imaging. *ACS Cent Sci* **2017**, 3 (4), 269−271.

121. Green, O.; Eilon, T.; Hananya, N.; Gutkin, S.; Bauer, C. R.; Shabat, D. Opening a gateway for chemiluminescence cell imaging: Distinctive methodology for design of bright chemiluminescent dioxetane probes. *ACS Cent Sci* **2017**, 3 (4), 349−358.

122. Liu, P.; Grenier, V.; Hong, W.; Muller, V. R.; Miller, E. W. Fluorogenic targeting of voltage-sensitive dyes to neurons. *J Am Chem Soc* **2017**, 139 (48), 17334−17340.

123. Gutova, M.; Goldstein, L.; Metz, M.; Hovsepyan, A.; Tsurkan, L. G.; Tirughana, R.; Tsaturyan, L.; Annala, A. J.; Synold, T. W.; Wan, Z. S.; Seeger, R.; Anderson, C.; Moats, R. A.; Potter, P. M.; Aboody, K. S. Optimization of a neural stem-cell-mediated carboxylesterase/irinotecan gene therapy for metastatic neuroblastoma. *Mol Ther Oncolytics* **2017**, 4, 67−76.

124. Wierdl, M.; Tsurkan, L.; Hatfield, M. J.; Potter, P. M. Tumour- selective targeting of drug metabolizing enzymes to treat metastatic cancer. *Br J Pharmacol* **2016**, 173 (19), 2811−2818

125. DuBois, S. G.; Mosse, Y. P.; Fox, E.; Kudgus, R. A.; Reid, J. M.; McGovern, R.; Groshen, S.; Bagatell, R.; Maris, J. M.; Twist, C. J.; Goldsmith, K.; Granger, M. M.; Weiss, B.; Park, J. R.; Macy, M. E.; Cohn, S. L.; Yanik, G.; Wagner, L. M.; Hawkins,

R.; Courtier, J.; Lai, H.; Goodarzian, F.; Shimada, H.; Boucher, N.; Czarnecki, S.; Luo, C. Q.; Tsao-Wei, D.; Matthay, K. K.; Marachelian, A. Phase II trial of alisertib in combination with irinotecan and Temozolomide for patients with relapsed or refractory neuroblastoma. *Clin Cancer Res* **2018**, 24 (24), 6142−6149.

126. Nagata, N.; Maeda, H.; Ishibashi, K.; Hirata, K.; Makiyama, A.; Iwamoto, S.; Takemoto, H.; Imasato, M.; Yoshida, Y.; Munemoto, Y.; Tanaka, C.; Morita, Y.; Hotta, Y.; Toyofuku, A.; Nagasaka, T.; Morita, S.; Sakamoto, J.; Mishima, H. Multicenter open-label randomized phase II study of second-line panitumumab and irinotecan with or without fluoropyrimidines in patients with KRAS wild-type metastatic colorectal cancer (PACIFIC study). *Med Oncol* **2019**, 36 (6), 46.

127. Nogami, N.; Hotta, K.; Segawa, Y.; Takigawa, N.; Hosokawa, S.; Oze, I.; Fujii, M.; Ichihara, E.; Shibayama, T.; Tada, A.; Hamada, N.; Uno, M.; Tamaoki, A.; Kuyama, S.; Ikeda, G.; Osawa, M.; Takata, S.; Tabata, M.; Tanimoto, M.; Kiura, K. Phase II study of irinotecan and amrubicin in patients with relapsed non-small cell lung cancer: Okayama Lung Cancer Study Group Trial 0402. *Acta Oncol* **2012**, 51 (6), 768−773.

128. Murakami, S.; Saito, H.; Kondo, T.; Ito, H.; Oshita, F.; Yamada, K. Phase II study of nedaplatin and irinotecan as adjuvant chemotherapy for completely resected non-small cell lung cancer. *Cancer Chemother Pharmacol* **2018**, 81 (1), 81−87.

129. Wills, B.; Cardona, A. F.; Rojas, L.; Ruiz-Patino, A.; Arrieta, O.; Reguart, N.; Carranza, H.; Vargas, C.; Otero, J.; Corrales, L.; Martin, C.; Cuello, M.; Pino, L. E.; Rolfo, C.; Rosell, R.; Zatarain-Barron, Z. L.; Latin-Amer Consortium Invest, L. Survival outcomes according to TIMP1 and EGFR expression in heavily treated patients with advanced non-small cell lung cancer who received biweekly irinotecan plus bevacizumab. *Anticancer Res* **2017**, 37 (11), 6429−6436.

130. Vo, K. T.; Michlitsch, J. G.; Shah, A. T.; Long-Boyle, J.; Kim, M. O.; Gustafson, C. Phase 1 multicenter trial to assess the maximum tolerated dose, safety, pharmacokinetics, and pharmacodynamics of pazopanib in combination with irinotecan and Temozolomide (PAZIT) for children and young adults with advanced sarcoma. *J Clin Oncol* **2018**, 36 (15), 2.

131. Palmerini, E.; Jones, R. L.; Setola, E.; Picci, P.; Marchesi, E.; Luksch, R.; Grignani, G.; Cesari, M.; Longhi, A.; Abate, M. E.; Paioli, A.; Szucs, Z.; D'Ambrosio, L.; Scotlandi, K.; Fagioli, F.; Asaftei, S.; Ferrari, S. Irinotecan and Temozolomide in recurrent Ewing sarcoma: An analysis in 51 adult and pediatric patients. *Acta Oncol* **2018**, 57 (7), 958−964.

132. Ciomei, M.; Croci, V.; Stellari, F.; Amboldi, N.; Giavarini, R.; Pesenti, E. Antitumor activity of edotecarin in breast carcinoma models. *Cancer Chemother Pharmacol* **2007**, 60 (2), 229−235.

133. Cabantous, S.; Waldo, G. S. In vivo and in vitro protein solubility assays using split GFP. *Nat Methods* **2006**, 3 (10), 845− 854.

134. Wang, T. N.; Badran, A. H.; Huang, T. P.; Liu, D. R. Continuous directed evolution of proteins with improved soluble expression. *Nat Chem Biol* **2018**, 14 (10), 972−980.

135. Evans, M. S.; Chaurette, J. P.; Adams, S. T.; Reddy, G. R.; Paley, M. A.; Aronin, N.; Prescher, J. A.; Miller, S. C. A synthetic luciferin improves bioluminescence imaging in live mice. *Nat Methods* **2014**, 11 (4), 393−395.

136. Kuchimaru, T.; Iwano, S.; Kiyama, M.; Mitsumata, S.; Kadonosono, T.; Niwa, H.; Maki, S.; Kizaka-Kondoh, S. A luciferin analogue generating near-infrared bioluminescence achieves highly sensitive deep-tissue imaging. *Nat Commun* **2016**, 7, 11856.

137. Hananya, N.; Shabat, D. Recent advances and challenges in luminescent imaging: Bright outlook for chemiluminescence of dioxetanes in water. *ACS Cent Sci* **2019**, 5 (6), 949−959.

138. Ryan, L. S.; Gerberich, J.; Cao, J.; An, W. W.; Jenkins, B. A.; Mason, R. P.; Lippert, A. R. Kinetics-based measurement of hypoxia in living cells and animals using an acetoxymethyl ester chemiluminescent probe. *ACS Sens* **2019**, 4 (5), 1391−1398.

139. Grimm, J. B.; Muthusamy, A. K.; Liang, Y. J.; Brown, T. A.; Lemon, W. C.; Patel, R.; Lu, R. W.; Macklin, J. J.; Keller, P. J.; Ji, N.; Lavis, L. D. A general method to fine-tune fluorophores for live-cell and in vivo imaging. *Nat Methods* **2017**, 14 (10), 987−994.

140. Lavis, L. D.; Raines, R. T. Bright building blocks for chemical biology. *ACS Chem Biol* **2014**, 9 (4), 855−866.

141. Jun, J. V.; Petersson, E. J.; Chenoweth, D. M. Rational design and facile synthesis of a highly tunable quinoline-based fluorescent small-molecule scaffold for live cell imaging. *J Am Chem Soc* **2018**, 140 (30), 9486−9493.

142. Jandu, H.; Aluzaite, K.; Fogh, L.; Thrane, S. W.; Noer, J. B.; Proszek, J.; Do, K. N.; Hansen, S. N.; Damsgaard, B.; Nielsen, S. L.; Stougaard, M.; Knudsen, B. R.; Moreira, J.; Hamerlik, P.; Gajjar, M.; Smid, M.; Martens, J.; Foekens, J.; Pommier, Y.; Brunner, N.; Schrohl, A. S.; Stenvang, J. Molecular characterization of irinotecan (SN-38) resistant human breast cancer cell lines. *BMC Cancer* **2016**, 16, 34.

143. Kawabata, S.; Oka, M.; Shiozawa, K.; Tsukamoto, K.; Nakatomi, K.; Soda, H.; Fukuda, M.; Ikegami, Y.; Sugahara, K.; Yamada, Y.; Kamihira, S.; Doyle, L. A.; Ross, D. D.; Kohno, S. Breast cancer resistance protein directly confers SN-38 resistance of lung cancer cells. *Biochem Biophys Res Commun* **2001**, 280 (5), 1216− 1223.

144. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **2012**, *9*(7), 671-675.

145. Buxbaum, A. R.; Haimovich, G.; Singer, R. H. In the right place at the right time: visualizing and understanding mRNA localization. *Nat Rev Mol Cell Biol*, **2015**, *16*, 95-109.

146. Bashirullah, A.; Cooperstock, R. L.; Lipshitz, H. D. RNA localization in development. *Annu Rev Biochem*, **1998**, *67*, 335-394.

147. Lipshitz, H. D.; Smibert, C. A. Mechanisms of RNA localization and translational regulation. *Curr Opin Cenet Dev*, **2000**, *10*, 476-488.

148. Seath, C. P.; Trowbridge, A. D.; Muir, T. W.; MacMillian, D. W. C. Reactive intermediates for interactome mapping. *Chem Soc Rev*, **2021**, *50*, 2911-2926.

149. Fazal, F. M.; Han, S.; Parker, K. R.; Kaewsapsak, P.; Xu, J.; Boettiger, A. N.; Chang, H. Y.; Ting, A. Y. Atlas of subcellular RNA localization revealed by APEX-Seq. *Cell*, **2019**, *178*, 473-490.

150. Han, S.; Zhao, B. S.; Myers, S. A.; Carr, S. A.; He, C.; Ting, A. Y. RNA-protein interaction mapping via MS2- or Cas13-based APEX targeting. *Proc Natl Acad Sci U S A*, **2020**, *117*, 22068-22079.

151. Padrón, A.; Iwasaki, S.; Ingolia, N. T. Proximity RNA labeling by APEX-Seq reveals organization of translation initiation complexes and repressive RNA granules. *Mol Cell*, **2019**, *75*, 875-887.

152. Wu, E.; Guo, X.; Teng, X.; Zhang, R.; Li, F.; Cui, Y.; Zhang, D.; Liu, Q.; Luo, J.; Wang, J.; Chen, R. Discovery of plasma membrane-associated RNAs through APEX-Seq. *Cell Biochem Biophys*, doi: 10.1007/s12013-021-00991-0.

153. Wang, P.; Tang, W.; Li, Z.; Zou, Z.; Zhou, Y.; Li, R.; Xiong, T.; Wang, J.; Zou, P. Mapping the spatial transcriptome with light-activated proximity-dependent RNA labeling. *Nat Chem Biol*. **2019**, *15*(11), 1110-1119.

154. Engel, K. L.; Lo, H.-Y. G.; Goering, R.; Li, Y.; Spitale, R. C.; Taliaferro, J. M. Analysis of subcellular transcriptomes by RNA proximity labeling with Halo-seq. *bioRxiv*, doi: 10.1101/2021.06.08.447604.

155. Zhou, Y.; Wang, G.; Wang, P.; Li, Z.; Yue, T.; Wang, J.; Zou, P. Expanding APEX2 substrates for proximity-dependent labeling of nucleic acids and proteins in living cells. *Angew Chem Int Ed Engl*. **2019**, *58*(34), 11763-11767.

156. Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc*. **2005**, *127*(12), 4223-4231.

157. Wilkinson, K. A.; Merino, E. J.; Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension 9SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc*. **2006**, *1*(3), 1610-1616.

158. Jones, K. A.; Kentala, K.; Beck, M. W.; An, W.; Lippert, A. R.; Lewis, J. C.; Dickinson, B. C. Development of a split esterase for protein-protein interaction-dependent small-molecule activation. *ACS Cent Sci*. **2019**, *5*(11), 1768-1776.

159. Spitale, R. C.; Flynn, R. A.; Zhang, Q. C.; Crisalli, P.; Lee, B.; Jung, W.-W.; Kuchelmeister, H. Y.; Batista, P. J.; Tore, E. A.; Kool, E. T.; Chang, H. Y. Structural imprirnts in vivo decode RNA regulatory mechanisms. *Nature*. **2015**, *519*(7544), 486-490.

160. Sun, L.; Fazal, F. M.; Li, P.; Broughton, J. P.; Lee, B.; Tang, L.; Huang, W.; Kool, E. T.; Chang, H. Y.; Zhang, Q. C. RNA structure maps across mammalian cellular compartments. *Nat Struct Mol Biol*. **2019**, *26*(4), 322-330.

161. Liu, J.; Dou, X.; Chen, C.; Chen, C.; Liu, C.; Xu, M. M.; Zhao, S.; Shen, B.; Gao, Y.; Han, D.; He, C. N6-methyladenosine of chromosome-associated regulatory RNA regulates chromatin state and transcription. *Science*. **2020**, *367*(6477), 580-586.

162. Kaewsapsak, P.; Shechner, D. M.; Mallard, W.; Rinn, J. L.; Ting, A. Y. Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. *Elife*. **2017**, e29224, doi: 10.7554/elife.29224

163. Hung, V.; Lam, S. S.; Udeshi, N. D.; Svinkina, T.; Guzman, G.; Mootha, V. K.; Carr, S. A.; Ting, A. Y. Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. *Elife*. **2017**, e24463, doi: 10.7554/elife.24463.

164. Reid, D. W.; Nicchitta, C. V. Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. *J Biol Chem*. **2021**, *287*(8), 5518-5527.

165. Jan, C. H.; Williams, C. C.; Weissman, J. S. Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science*. **2014**, *346*(6210), 1257521.

166. Wang, X.; Zhao, B. S.; Roundtree, I. A.; Lu, Z.; Han, D.; Ma, H.; Weng, X.; Chen, K.; Shi, H.; He, C. N(6)-methyladenosine modulates messenger RNA translation efficiency. *Cell*. **2015**, *161*(6), 1388-1399.

167. Tang, Y.; Chen, K.; Song, B.; Ma, J.; Wu, X.; Xu, Q.; Wei, Z.; Su, J.; Liu, G.; Rong, R.; Lu, Z.; Pedro de Magalhaes, J.; Rigden, D. J.; Meng, J. m6A-Atlas: a comprehensive knowledgebase unraveling the N6-methyladenosine (m6a) epitranscriptome. *Nucleic Acids Res.* **2021**, *49*(D1), D134-D143.

168. Howe, K. L., et al. Ensembl 2021. *Nucleic Acids Res.* **2021**, *49*(1), 884-891.

169. Mortimer, S. A.; Trapnell, C.; Aviran, S.; Pachter, L.; Lucks, J. N. SHAPE-seq: high-throughput RNA structure analysis. *Curr Protoc Chem Biol*. **2012**, *4*(4), 275-297.

170. Loughrey, D.; Watters, K. E.; Settle, A. H.; Lucks, J. B. SHAPE-seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res*. **2014**, *42*(21), e165.

171. Kleinpenning, F.; Eising, S.; Berkenbosch, T.; Garzero, V.; Schaart, J. M.; Bonger, K. M. Subcellular protein labeling by a spatially restricted arylamine n-acetyltransferase. *ACS Chem Biol.* **2018**. *13*(8), 1932-1937.

172. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods.* **2012**. *9*(7), 671-675.

173. Fazal, F. M.; Han, S.; Parker, K. R.; Kaewsapsak, P.; Xu, J.; Boettiger, A. N.; Chang, H. Y.; Ting, A. Y. *Cell*. **2019**. *178*(2), 473-490.; protocolexchange doi: 10.21203/rs.2.1857/v1.

174. Bray, N. L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **2016**. *34*(5), 525-527.

175. Xia, C.; Fan, J.; Emanuel, G.; Hao, J.; Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc Natl Acad Sci U S A.* **2019**. *116*(39), 19490-19499.

176. Liu, N.; Dai, Q.; Zheng, G.; He, C.; Parisien, M.; Pan, T. N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature*, **2015**. *518*(7540), 560-564.

177. Rathburn, C. M.; Prescher, J. A. Bioluminescent probes for imaging biology beyond the culture dish. *Biochemistry*. **2017**, *56*(39), 5178-5184.

178. Mollard, S.; Fanciullino, R.; Giacometti, S.; Serdjebi, C.; Benzekry, S.; Ciccolini, J. In vivo bioluminescence tomography for monitoring breast tumor growth and metastatic spreading: comparative study and mathematical modeling. *Sci Rep*. **2016**, *6*, 36173.

179. Jones, K. A.; Porterfield, W. B.; Rathbun, C. M.; McCutcheon, D. C.; Paley, M. A.; Prescher, J. A. Orthogonal luciferase-luciferin pairs for bioluminescence imaging. *J Am Chem Soc*. **2017**, *139*(6), 2351-2358.

180. Liu, M. D.; Warner, E. A.; Morrissey, C. E.; Fick, C. W.; Wu, T. S.; Ornelas, M. Y.; Ochoa, G. V.; Zhang, B. S.; Rathbun, C. M.; Porterfield, W. B.; Prescher, J. A.; Leconte, A. M. Statistical coupling analysis-guided library design for the discovery of mutant luciferases. *Biochemistry*. **2018**, *57*(5), 663-671.

181. Iwano, S.; Sugiyama, M.; Hama, H.; Watakabe, A.; Hasegawa, N.; Kuchimaru, T.; Tanaka, K. Z.; Takahashi, M.; Ishida, Y.; Hata, J.; Shimozono, S.; Namiki, K.; Fukano, T.; Kiyama, M.; Okano, H.; Kizaka-Kondoh, S.; McHugh, T. J.; Yamamori, T.; Hioki, H.; Maki, S.; Miyawaki, A. Single-cell bioluminescence imaging of deep tissue in freely moving animals. *Science*. **2018**, *359*(6378), 935-939.

182. Rathbun, C. M.; Porterfield, W. B.; Jones, K. A.; Sagoe, M. J.; Reyes, M. R.; Hua, C. T.; Prescher, J. A. Parallel screening for rapid identification of orthogonal bioluminescent tools. *ACS Cent Sci*. **2017**, *3*(12), 1254-1261.

183. Yeh, H.-W.; Karmach, O.; Ji, A.; Martins-Green, M. M.; Ai, H-W. Red-shifted luciferase-luciferin pairs for enhanced bioluminescence imaging. *Nat Methods*. **2017**, *14*(10), 971-974.

184. Vassel, N.; Cox, C. D.; Naseem, R.; Morse, V.; Evans, R. T.; Power, R. L.; Brancale, A.; Wann, K. T.; Campbell, A. K. Enzymatic activity of albumin shown by coelenterazine chemiluminescence. *Luminescence*. **2021**, *27*(3), 234-241.

185. Pichler, A.; Prior, J. L.; Piwnica-Worms, D. Imaging reversal of multidrug resistance in living mice with bioluminescence: MDR1P-glycoprotein transports coelenterazine. *Proc Natl Acad Sci U S A*. **2004**, *101*(6), 1702-1707.

186. Bruemmer, K. J.; Green, O.; Su, T. A.; Shabat, D., Chang, C. J. Chemiluminescent probs for activity-based sensing of formaldehyde released from folate degradation in living mice. *Angew Chem Int Ed Engl*. **2018**, *57*(25), 7508-7512.

187. Cao, J.; An, W.; Reeves, A. G.; Lippert, A. R. A chemiluminescent probe for cellular peroxynitrite using a self-immolative oxidative decarbonylation reaction. *Chem Sci*. **2018**, *9*(9), 2552-2558.

188. Tian, L.; Yang, Y.; Wysocki, L. M.; Arnold, A. C.; Hu, A.; Ravichandran, B.; Sternson, S. M.; Looger, L. L.; Lavis, L. D. Selective esterase-ester pair for targeting small molecules with cellular specificity. *Proc Natl Acad Sci U S A*. **2012**, *109*(13), 4756-4761.

189. Hananya, N.; Boock, A. E.; Bauer, C. R.; Satchi-Fainaro, R.; Shabat, D. Remarkable enhancement of chemiluminescent signal by dioxetane-fluorophore conjugates: turn-on chemiluminescence probes with color modulation of sensing and imaging. *J Am Chem Soc*. **2016**, *138*(40), 13438-13446.

190. Green, O.; Eilon, T.; Hananya, N.; Gutkin, S.; Bauer, C. R.; Shabat, D. Opening a gateway for chemiluminescence cell imaging: a distinctive methodology for design of bright chemiluminescent dioxetane probes. *ACS Cent Sci*. **2017**, *3*(4), 349-358.

191. Giver, L.; Gershenson, A.; Freskgard, P. O.; Arnold, F. H. Directed evolution of a thermostable esterase. *Proc Natl Acad Sci USA*. **1998**, *95*(22), 12809-12813.

192. Eilon-Shaffer, T.; Roth-Konforti, M.; Eldar-Boock, A.; Satchi-Fainaro, R.; Shabat, D. ortho-chlorination of phenoxy 1,2-dioxetane yields superior chemiluminescent probes for in vitro and in vivo imaging. *Org Biomol Chem*. **2018**, *16*(10), 1708-1712.

193. Liu, L.; Mason, R. P. Imaging beta-galactosidase activity in human tumor xenografts and transgenic mice using a chemiluminescent substrate. *PLoS One*. **2010**, *5*(8), e12024.

194. Hemmert A. C.; Otto, T. C.; Wierdl, M.; Edwards, C. C.; Fleming, C. D.; MacDonald, M.; Cashman, J. R.; Potter, P. M.; Cerasoli, D. M.; Redinbo, M. R. Human carboxylesterase 1 stereoselectively binds the nerve agent cyclosarin and spontaneously hydrolyzes the nerve agent sarin. *Mol Pharmacol*. **2010**, *77*(4), 508-516.

195. Lippert, A. R. Unlocking the potential of chemiluminescence imaging. *ACS Cent Sci*. **2017**, *3*(4), 269-271.

196. Gibson, D. G.; Young, L; Chuang, R.-Y.; Venter, J. C.; Hutchinson III, C. A.; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. **2009**, *6*(5), 343-345.

197. Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. **2010**, *31*(2), 455-461.

198. Qiu, T.; Kathayat, R. S. Cao, Y.; Beck, M. W.; Dickinson, B. C. A fluorescent probe with improved water solubility permits the analysis of protein s-depalmitoylation activity in live cells. *Biochemistry*. **2018**, *57*(2), 221-225.

199. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchinson, G. R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminform*. **2012**, *4*(1), 17.

200. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem*. **2009**, *30*(16), 2785-2791.

201. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchinson, G. R. Open Babel: an open chemical toolbox. *J Cheminform*. **2011**, *3*, 33.

202. Chen, V. B.; Arendall III, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. **2010**, *66*(1), 12-21.

203. Jones, K. A.; Snodgrass, H. M.; Belsare, K.; Dickinson, B. C.; Lewis, J. C. Phage-assisted continuous evolution and selection of enzymes for chemical synthesis. *ChemRxiv*. **2021**, doi: 10.33774/chemrxiv-2021-n18xc.