The Circuit as the Therapeutic

Thesis by
Lucy Shin Chong

In Partial Fulfillment of the Requirements for the
Degree of
Systems and Synthetic Biology, PhD

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2021
Defended May 18, 2021

ACKNOWLEDGEMENTS

This thesis is a condensed scientific account of my journey through graduate school. This thesis does not capture, in the slightest, the human aspect of that experience. Along the way I made wonderful friends, met beautiful people, and learned from brilliant scientists.

None of this would've been possible without the guidance of my advisor Michael Elowitz. No one could ask for a kinder advisor. He gave me the introduction, independence, and inspiration necessary for me to be the scientist I am today. He is also living proof that science and artistry go hand-in-hand. My committee (Bruce Hay, Mikhail Shapiro, and Viviana Gradinaru) has guided me throughout this process providing invaluable advice. I am grateful for my close collaborator throughout graduate school Ronghui Zhu. Without his suggestions, iterations and evolution of ideas my journey would've been waylaid. The exceptional students of CHOMP, 'wee' Maeve Morgan Phoa and Xiaofei Ge, showed me that the next generation of scientists are exceptional and I look forward to seeing how their own journeys turn out. The work described here would not be possible without the foundational support of the CHOMP squad Matthew Kim and Michaela Ince. Please forgive me for all the unnecessary cloning and CAPS-locking in texts.

Mark Budde and Helen McBride introduced me to the world of startups. Without them and their continuous support, the next chapter in my journey would have started in a dramatically different way. Leah Santat, Jo Leonardo and James Linton ran the ship that would undoubtedly sink without them. Grace Chow, Christina Su, and Heidi Klumpe, my fellow sgsgloungewithchampagne members (or was it sgsgsg?), gave me all the support to navigate the quagmire that is the graduating process. Zibo Chen, Alejandro Granados, and Rachael Kuintzle took on my neurosis and anxieties and their calming presence, perspective and wonderful advice. To my baymates: Jan Gregrowicz, Yitong Ma, Alex Ng and Laurent Potvin-Trottier. Sitting and working next to each of you was a privilege and a pleasure. Laurent, without your commiseration about figures and your expansive knowledge about coffee I would've never made it through the CHOMP writing process. Alex, our bike rides killed me and thanks for providing me with the next move in my career. Yitong, thanks for being a kind and patient baymate always willing to help me with code. And to Jan, you beautiful philosophical kurwo, our conversations, never about science, were always a reminder that life does exist beyond the lab. To all of Labowitz, thank you

for being wonderful and fueling my imposter syndrome.

My Caltech friends, in particular Joseph Ruan, David Basta, and Angel Galven Merchan were invaluable scientific comrades. Thank you for listening to me moan about 'publishing too soon'; and, especially to Joe, entertaining me as I flailed while coding and keeping me company during late figure-making nights. To the Del Mars members (Marica Raguso, Phil Walk, Markus Dick and Sharan Prakash) thank you for being excellent housemates. To my previous mentors and collaborators Dorothea Fiedler, Adam Resnick, Ronda Rolfes, Mingxuan Wu, Nate Brown and Elizabeth Steidle, thank you for kicking off my scientific career. My steadfast friends Jenny Liang, Jenny Ma, Mishal Awadah, David Lee and the "Hot Pot Baby" crew, thanks for being with me over the last ten years. To Cesar Perez, the man with the most flow, without you I would've never met the most charismatic being I know Sissy Waimatha, your beautiful partner Aria Safar, and of course, my partner in crime Andrew Erskine. I'm so lucky to have you as a friend.

To my wonderful partner Andrew Erskine (and his sidekick Ariel), thank you for being with me when the world fundamentally changed and what could've been the scariest and worst year of my life became one of the happiest. And, because I couldn't put your suggested last sentence, I'm going to put it here: "Comrades, do not fear the march of synthetic biology, together let us happily build a future for ALL circuit-based therapies."

Without the support of my family, I would've never made it through graduate school. My dad Yon Sam Chong, my mom Myong Nan Chong, and my brother Thomas In Chong nourished my scientific curiosity. Undoubtedly, their willingness to overlook the various homemade concoctions and somewhat gross body observations around the house and instead redirect my scientific curiosity with illustrated pocketbook encyclopedias and a toy microscope, undoubtedly sparked and nourished my interest in science. They epitomize the meaning of family, having supported me and asked nothing in return.

Finally, I owe the majority of my success to my mentor, collaborator, and friend Xiaojing Gao. He is a guiding light both scientifically and personally and I am lucky that his brilliance, for a brief moment, shined on me. Thank you Xiaojing, I could've never done this without you.

I came into graduate school confused and left even more confused and I have all of you to thank for it.

ABSTRACT

Cell-specific targeting of therapeutics is a fundamental challenge in biomedicine. The use of engineered proteins that interact with one another as designed, synthetic circuits represents a promising solution to this challenge. These circuits can be constructed to directly sense endogenous cell signals, act on these signals to classify cellular state, and produce a specific response such as conditional triggering of cell death or targeted expression of a reporter. Synthetic protein circuits can also be delivered in mRNA vectors transiently to avoid permanent gene modification.

We recently showed viral proteases can be engineered to regulate one another in a composable manner, permitting the construction of diverse protein-level circuits (Circuits of Hacked Orthogonal Modular Proteases). CHOMP could perform a wide range of computations including Boolean logic, analogue signal processing, and dynamic signal processing. Using this system we were also able to directly sense key cellular pathways and conditionally respond to trigger apoptosis in cancer cells. Further expansion of synthetic protein circuits to include nonlinear signal processing enables new system-level behaviors.

Protein-based circuits are compatible with innovative delivery methods including mRNA encapsulated in lipid-nanoparticle formulations and engineered viruses. We were able to develop a controllable, transient RNA-virus delivery system that allowed for targeted delivery to defined cell populations. This RNA-virus platform will enable synthetic protein circuit delivery and is compatible with cell-based therapies.

CHOMP and its delivery system is a holistic platform that establishes the role of post-translational circuits as future therapeutic devices.

PUBLISHED CONTENT AND CONTRIBUTIONS

Gao XJ, Chong LS, Kim MS, Elowitz MB (2018). "Programmable protein circuits in living cells". In: Science, vol.. 361, issue 6408, pp. 1252-1258. https://doi.org/10.1126/science.aat5062

Gao, Xiaojing J et al. (May 2020). "Engineering multiple levels of specificity in an RNA viral vector". In: BioRxiv. https://doi.org/10.1101/2020.05.27.119909

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Number   *Page*

Chapter 1

INTRODUCTION: THE CIRCUIT AS THE THERAPEUTIC

1.1 Synthetic biology-based platforms will transform medicine

The world of pharmacology has undergone a biological revolution. Over the last twenty years pharmacology has fundamentally changed. Historically, treatments have been limited by our capacity to synthesize certain chemicals. Within the last ten years the introduction of biologics has dramatically changed our ability to target disease and we are now witnessing another biomedical revolution that utilizes living cells as a therapeutic modality (Fischbach, Bluestone, and Lim, 2013) Despite these great advances in the areas of biological therapeutics and medical devices, these platforms typically target single-marker disease indicators which remain insufficient in treating more complex diseases.

Certain diseases require more complex therapeutic programs. Careful and laborious work has established fundamental knowledge in delineating pathway components and their relationship to the mechanisms of human disease. These reductionist approaches have resulted in single-marker targets for treatment. However, we often see that treatments for some diseases, like cancer, that are based on small molecule drugs and biologics can often fail due to development of mutations(Gerlinger, 2019). The use of the immune system against disease is a phenomenally exciting and successful idea that has leveraged the natural effector functions of immune cells to target disease, but is generally limited to inspecting one or two surface proteins (Dannenfelser et al., 2020). These proteins are typically not specific and do not reliably discriminate between healthy and diseased cells (Dannenfelser et al., 2020). Even in instances where replacement of an individual gene is sufficient for treatment, as in gene therapy, gene therapies typically do not provide designed regulation for gene expression which can result in severe complications(Prelich, 2012; Goverdhana et al., 2005). What underlies the success or failure of these drug designs is our understanding of the disease, but our reductive approach has underestimated the complexity of the living cell and its ability to evolve.

Living cells are highly dynamic systems. Cells contain networks of interacting components capable of generating complex behaviors. These biological circuits provide a natural mechanism for specifically sensing and responding to aberrant

states. For example, the p53 pathway senses different levels and types of DNA damage and other stresses, and then conditionally triggers cell cycle arrest or cell death (Purvis et al., 2012). However, when dysregulated, these same control circuits result in disease. How can we design specific treatments for disease in the context of this cell complexity? One approach is to leverage that very complexity to engineer synthetic strategies and develop multifaceted therapies.

Synthetic biology enables rational programming of new cellular functions. Synthetic biology provides a framework for deciphering the underlying principles of living systems by "building to understand" . By obeying general circuit design principles, synthetic biologists have used biological parts to engineer robust, controllable behaviour in living cells. While the first wave of synthetic biologists aimed to demonstrate electronic circuit-like behavior in bacteria, including the engineering of robust oscillations (Elowitz and Leibler, 2000; Stricker et al., 2008; Potvin-Trottier et al., 2016) and a toggle switch (Gardner, Cantor, and Collins, 2000) synthetic biology is quickly approaching systems-level control in mammalian cells. Mammalian synthetic biology provides a new framework for biomedicine and holds vast promise for developing new therapeutics.

Even implementations of basic computational paradigms such as logic and dynamic signal processing in mammalian cells have implications for biomedicine. For instance, several key members of the RTK/Ras/MAPK and PI3K/Akt signaling pathways play a critical role in the progression of cancer (Sanchez-Vega et al., 2018). Co-current mutations in these pathways could function as specific input signals for AND-like integration to enable selective targeting of cancer cells. Developing a synthetic biological circuit that effectively senses the disease state, processes the input information, and actuates a program would enable 'smart' therapeutic devices.

What this introduction will cover. The aim for this introduction is to provide an overview for the field of mammalian synthetic biology, in particular, the foundation that has been built to facilitate the building of circuit-based therapeutics. Cell behavior requires at least three simple modules, a module for sensing input, a processing module for conditioning and transmitting that signal, and an output module that determines the physiological and behavioural change. We reasoned that rewiring cellular behavior with therapeutic synthetic circuits necessitates the development of all of these modules; and therefore, our review will cover these topics accordingly. First, we will survey synthetic components used to sense DNA, RNA,

and protein. Second, we will then examine circuit architectures that will be beneficial for biomedicine in processing cell signals. Third, we will identify therapeutically relevant outputs including control of key pathways and cell behavior. Finally, we discuss how circuit-based therapies can be improved and provide a roadmap for the field going forward.

1.2 Sense: Synthetic systems now allow detection of diverse input signals

Nucleic acid-based components enable a wide variety of circuit functions. A growing library of parts has been built to facilitate circuit construction for transcriptional, translational, and post-translational control. However, most efforts at creating synthetic molecular circuits use genes and gene regulatory elements due to the predictability of nucleic acid base-pairing and the relative ease of interchanging gene-protein elements. Seminal work was performed in genetic control using nucleic acids including DNA circuits that self-assemble to perform various computations (Qian, Winfree, and Bruck, 2011); RNA switches that regulate gene expression (Green et al., 2014); and RNA classifying circuits (Xie et al., 2011). Expansion of these nucleic acid-focused systems to include gene regulatory elements such as synthetic promoters (Schlabach et al., 2010; Saxena, Bojar, and Fussenegger, 2017), transcription factors, Cas-family proteins (DiAndreth et al., 2019), integrases (Bonnet et al., 2013), orthogonal ribosome-mRNA pairs(Rackham and Chin, 2005), and programmable orthogonal zinc fingers(Khalil et al., 2012) have permitted a wide variety of circuit functions. Further work has also identified methods of linking these individual elements to generate powerful and flexible approaches for engineered cell control. However, such gene regulatory circuits can require potentially mutagenic genome integration procedures, are limited in operational speed and stability, and interact only indirectly with key protein-level cellular activities.

Proteins are essential for every biological process. Cellular functions are mediated by the interaction of systems of proteins, forming circuits which drive changes in the cell. The inherent modularity of some protein domains enables their potential use as flexible components for synthetic circuit design. For example, the native Notch receptor has been re-engineered to be a fully customizable receptor. Typically, engagement of the extracellular Notch domain will trigger proteolysis of a transmembrane region and release the intracellular transcription factor to induce gene expression (Kopan and Ilagan, 2009). SynNotch retained the transmembrane region and permitted user-defined swapping of the extracellular targeting domain

and intracellular effectors resulting in arbitrary gene expression following detection (Morsut et al., 2016; Roybal et al., 2016). The remarkable flexibility of SynNotch demonstrated modular protein interaction domains can be used to re-wire endogenous protein circuits, and couple their activities to non-natural inputs (Huang et al., 2017; Barnea et al., 2008; Kipniss et al., 2017; Daringer et al., 2014; Yeh et al., 2007).

However, from a design standpoint, orthogonal protein domains might be preferred to limit their interaction with endogenous pathways (Z. Chen and Elowitz, 2021). Excitingly, *de novo* protein design have shown great success and may provide insulated, flexible and customizable scaffolds to mediate signal transmission (Z. Chen, Boyken, et al., 2019; Z. Chen, Kibler, et al., 2020; Ng et al., 2019; Lajoie et al., 2020). Despite this enormous potential, the development of synthetic mammalian protein circuits has been severely limited. The previous work described here has developed ad hoc, non-scalable designs. The key problem has been the lack of composability.

An ideal biological sensor module is both modular and composable to permit construction of a wide variety of circuit architectures in order to generate sophisticated cellular behavior. Just as a few electronic components can make a variety of electronic circuits, composable protein components would permit a similar level of flexibility in implementing protein circuits. *In this thesis, we show how viral proteases can be engineered to regulate one another as composable protein components. (Gao XJ, Chong LS, Kim MS, Elowitz MB, 2018).* The scheme we introduce is based on a simple 'zip and clip' design principle: proteases dock with target proteases using modular leucine zippers and clip the target protease at a specific site to inhibit its function. We demonstrated that the design can be transferred from one protease to another, making the scheme extensible to orthogonal proteases. Using these components, we implemented a diverse set of core functions in mammalian cells.

1.3 Process: Diverse circuit architectures enable new computational paradigms for biomedicine.

Logic gates increase the number of inputs and specificity: Specificity is crucial for biomedicine. The ability to interrogate several pathways in a cell could serve to

enhance selectivity and restrict therapeutic action. Previously, 'classifier' circuits were developed using miRNA and transcriptional control within a Boolean framework (Xie et al., 2011). Low or high miRNA expression levels were binarized as inputs and used in an AND-like logic function to identify cancer cells and selectively kill them. Multi-input configurations that implement logic such as AND, OR, IF/THEN can integrate several endogenous signals and actuate a programmed response with enhanced specificity compared to single-target inputs (Gerlinger, 2019).

Analogue signal processing enables circuits to function in desired input regimes: Cellular inputs are rarely completely on or completely off. The ability to selectively respond to certain input concentrations or adapting to the change in input signal would permit the circuit to function only in desired regimes. In particular, imbuing a system with ultrasensitivity – the ability to have switch-like, large changes in output in response to small changes in input – at specified input levels is key for signal processing and target discrimination (Ferrell and Ha, 2014c). In breast cancer, HER2 amplification is associated with an aggressive phenotype and seen in up to 20% of cases (Gajria and Chandarlapaty, 2011). In order to differentiate between HER2 overexpressing and wildtype cells, an ultrasensitive positive feedback circuit capable of sensing antigen density was built using a two-step mechanism in which engagement of a low affinity HER2 SynNotch receptor induced the expression for a high affinity chimeric antigen receptor (CAR) (Hernandez-Lopez et al., 2021). Natural circuits use a variety of mechanisms to generate ultrasensitive responses including cooperativity, multisite phosphorylation, cascades, and molecular titration (Ferrell and Ha, 2014c; Ferrell and Ha, 2014a; Ferrell and Ha, 2014b; Buchler and Louis, 2008). Inspiration can be drawn from these natural mechanisms and guide circuit designs.

Dynamic signal processing permits adaptive dosing. Signalling dynamics can offer an additional layer of information encoding. Dynamic properties such as signaling frequency and duration can alter gene expression and determine cell fate. Cells that undergo DNA damage may either pulse p53 or sustain a p53 response (Purvis et al., 2012; Geva-Zatorsky et al., 2006; Batchelor et al., 2011). Cells that sustain p53 typically undergo senescence and those that oscillate their p53 signaling will recover from DNA damage (Purvis et al., 2012; S.-H. Chen, Forrester, and Lahav, 2016). Additionally, circuits should incorporate dynamic responses in order to function properly in cells. Current cell therapies use sustained expression of chimeric antigen receptors which result in dysfunctional 'exhausted' cell pheno-

types. Previous results showed that pulsing the antigen for immune cell activation boosted treatment efficacy and functional longevity. These results show that having a controllable temporal factor in drug administration to address pathway dynamics may be crucial for effective treatment.

Controllers give added resilience to noise and genetic variability. Negative feedback increases robustness to variations in circuit output and provides resilience to environmental perturbations. Simple design principles applied from control theory will provide the foundation for designing circuits robust to variations in expression. For example, circuits may be optimized through iterative designs expressed from plasmids; however, there is major variability in the copy number of the plasmid that gets delivered to cells (Bleris et al., 2011). Implementing negative feedback and feed-forward control on the output products would regulate the production of the output.

1.4   Respond: Output pathways and cell therapy will be useful effectors for therapeutics

Control of key therapeutic output pathways. The spectrum of cellular behavior is vast and diverse. However, controlling a few key pathways relevant for therapeutic output could establish a foundation for engineering circuit outputs. For example, cell death pathways can eliminate target cell populations in cancer or senescence. Engineering executioner caspases to induce apoptosis in cells can function as both a therapeutic output in cells that express oncogenes (Gao XJ, Chong LS, Kim MS, Elowitz MB, 2018) and function as a kill-switch for engineered cells (Gargett and Brown, 2014). Controlling the secretion of immune effectors and cytokines can recruit immune cells to a local site or enhance cellular processes (Hu et al., 2017; Chmielewski et al., 2011). Control of transcriptional regulators could induce changes in chromatin (Keung et al., 2015) or cellular state (Zhu et al., 2021).

Circuits combined with cell therapy provide powerful new therapeutic capabilities. Circuits can engineer desired behaviours into therapeutic cells combining all the benefits of cell therapy with a synthetic therapeutic program. The recent advent of cellular immunotherapy has revolutionized cancer treatment. Of particular note, CAR T-cell therapy, in which a patient's T-cells are engineered to express a chimeric antigen receptor (CAR) that specifically targets an antigen of interest, is an elegant sense-and-respond platform which uses immune cells as a 'living drug' (Jackson,

Rafiq, and Brentjens, 2016). Recent efforts have been focused on improving the efficacy and safety of these cells as well as diversifying their utility (Rafiq, Hackett, and Brentjens, 2019). Synthetic biology is already being used to address some of these concerns including the addition of chemically inducible kill switches to safeguard cells (Gargett and Brown, 2014; Giordano-Attianese et al., 2020), logic gated CARs to improve tumor specificity (Lajoie et al., 2020; Hamburger et al., 2020), and conditional synthetic gene expression to enhance tumor targeting (Greenshpan et al., 2021; Martinez and Moon, 2019).

Circuits can allow intercellular communication. Cell therapy has focused on engineering a single cell type and optimizing its natural function; however, many applications require spatial control and communication across cell types to make population-based decisions. Engineering programmable cellular communication modules into sub-populations of cells could permit community-level control. For example, immune cells coordinate immune responses by sensing their population density through local cytokine concentrations that are autonomously produced (Hart, Antebi, et al., 2012; Hart, Reich-Zeliger, et al., 2014). A synthetic quorum sensing system for mammalian cells would permit an orthogonal channel for intercellular communication that would enable cells to sense and control their own populations (Hennig, Rödel, and Ostermann, 2015). Recent work established auxin, a plant-hormone, as an orthogonal signal in mammalian (Holland, 2012; Ma et al., 2020). When hooked up to an auxin-inducible degron that controlled blasticidin resistance, the authors were able to build a quorum sensing circuit capable of controlling cell death (Ma et al., 2020). Incorporation of population control circuits with orthogonal communication signals would complement current cell therapy approaches by permitting activation at certain cell densities.

1.5 Delivery is a key consideration in turning the circuit into a therapeutic.

Considerable work goes into effective and safe pharmaceutical administration. Circuit-based therapeutics face similar challenges and will encounter new ones. Should the circuit be delivered as DNA, mRNA, protein? By which vector should the circuit be delivered? While each case will be unique, recent advances in the biological delivery space hold promise for several routes of administration. mRNA-based therapies could allow expression of circuits without permanent modification of host cells (Sahin, Karikó, and Türeci, 2014). Viruses and replicons (Ying et al., 1999) could be used for longer term operation of a circuit. New formulations of lipid nanoparticles permit delivery of RNA and DNA to specific tissue and cell type

(Dahlman et al., 2017; Mitchell et al., 2021). Exosome and liposome encapsulation is a promising field which uses natural cellular processes to deliver biological payloads (Luan et al., 2017). The cell could also function as a circuit delivery device Lutz et al., 2019. Circuits functioning to improve cell therapy could also improve the cell's delivery capabilities.

In this thesis we show we can engineer the rabies virus to allow for specific control of viral secretion, targeting, and replication by cell surface proteins as well as intracellular proteins(Gao et al., 2020). To achieve these distinct levels of specificity, we combined controlled glycoprotein expression in viral "sender" cells with pseudotyping and engineered bispecific bridge proteins, as well as a degradation-prone nanobody stablized by its target protein. The resulting systems allow one to externally control viral secretion to generate viruses that conditionally infect target cells based on surface proteins, and conditionally replicate in those cells depending on the presence of an intracellular protein target. We also introduce a new mechanism to control viral replication. Critically, the system is designed in a "fail safe" manner such that in the intended use case presents little selection pressure for individual mutations that bypass the regulatory system. Ultimately, we envision these engineered viruses to safely deliver post-translational circuits to specific cells.

What this thesis will cover. The thesis scratches the surface of protein-based synthetic circuit design and post-translational circuit delivery. What we have developed here, as detailed in Chapter II, is a composable programmable protein synthetic biology toolkit with therapeutic implications. We further describe in Chapter III the expansion of these protein components and detail their versatility. Finally in Chapter IV, we describe an engineered viral-based delivery system which further enables synthetic protein circuit delivery.

References

Barnea, Gilad et al. (Jan. 2008). "The genetic design of signaling cascades to record receptor activation". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 105.1, pp. 64–69.

Batchelor, Eric et al. (May 2011). "Stimulus-dependent dynamics of p53 in single cells". en. In: *Mol. Syst. Biol.* 7, p. 488.

Bleris, Leonidas et al. (Aug. 2011). "Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template". en. In: *Mol. Syst. Biol.* 7, p. 519.

Bonnet, Jerome et al. (May 2013). "Amplifying genetic logic gates". en. In: *Science* 340.6132, pp. 599–603.

Buchler, Nicolas E and Matthieu Louis (Dec. 2008). "Molecular titration and ultrasensitivity in regulatory networks". en. In: *J. Mol. Biol.* 384.5, pp. 1106–1119.

Chen, Sheng-Hong, William Forrester, and Galit Lahav (Mar. 2016). "Schedule-dependent interaction between anticancer treatments". en. In: *Science* 351.6278, pp. 1204–1208.

Chen, Zibo, Scott E Boyken, et al. (Jan. 2019). "Programmable design of orthogonal protein heterodimers". en. In: *Nature* 565.7737, pp. 106–111.

Chen, Zibo and Michael B Elowitz (Apr. 2021). "Programmable protein circuit design". en. In: *Cell* 184.9, pp. 2284–2301.

Chen, Zibo, Ryan D Kibler, et al. (Apr. 2020). "De novo design of protein logic gates". en. In: *Science* 368.6486, pp. 78–84.

Chmielewski, Markus et al. (Sept. 2011). "IL-12 release by engineered T cells expressing chimeric antigen receptors can effectively Muster an antigen-independent macrophage response on tumor cells that have shut down tumor antigen expression". en. In: *Cancer Res.* 71.17, pp. 5697–5706.

Dahlman, James E et al. (Feb. 2017). "Barcoded nanoparticles for high throughput in vivo discovery of targeted therapeutics". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 114.8, pp. 2060–2065.

Dannenfelser, Ruth et al. (Sept. 2020). "Discriminatory Power of Combinatorial Antigen Recognition in Cancer T Cell Therapies". en. In: *Cell Syst* 11.3, 215–228.e5.

Daringer, Nichole M et al. (2014). "Modular extracellular sensor architecture for engineering mammalian cell-based devices". In: *ACS Synth. Biol.* 3.12, pp. 892–902.

DiAndreth, Breanna et al. (Dec. 2019). "PERSIST: A programmable RNA regulation platform using CRISPR endoRNases". en.

Elowitz, Michael B and Stanislas Leibler (Jan. 2000). "A synthetic oscillatory network of transcriptional regulators". In: *Nature* 403, p. 335.

Ferrell Jr, James E and Sang Hoon Ha (Oct. 2014a). "Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity". en. In: *Trends Biochem. Sci.* 39.10, pp. 496–503.

– (Nov. 2014b). "Ultrasensitivity part II: multisite phosphorylation, stoichiometric inhibitors, and positive feedback". en. In: *Trends Biochem. Sci.* 39.11, pp. 556–569.

– (Dec. 2014c). "Ultrasensitivity part III: cascades, bistable switches, and oscillators". en. In: *Trends Biochem. Sci.* 39.12, pp. 612–618.

Fischbach, Michael A, Jeffrey A Bluestone, and Wendell A Lim (Apr. 2013). "Cell-based therapeutics: the next pillar of medicine". en. In: *Sci. Transl. Med.* 5.179, 179ps7.

Gajria, Devika and Sarat Chandarlapaty (Feb. 2011). "HER2-amplified breast cancer: mechanisms of trastuzumab resistance and novel targeted therapies". en. In: *Expert Rev. Anticancer Ther.* 11.2, pp. 263–275.

Gao XJ, Chong LS, Kim MS, Elowitz MB (2018). "Programmable protein circuits in living cells". In: *Science*.

Gao, Xiaojing J et al. (May 2020). "Engineering multiple levels of specificity in an RNA viral vector". en.

Gardner, T S, C R Cantor, and J J Collins (Jan. 2000). "Construction of a genetic toggle switch in Escherichia coli". en. In: *Nature* 403.6767, pp. 339–342.

Gargett, Tessa and Michael P Brown (Oct. 2014). "The inducible caspase-9 suicide gene system as a "safety switch" to limit on-target, off-tumor toxicities of chimeric antigen receptor T cells". en. In: *Front. Pharmacol.* 5, p. 235.

Gerlinger, Marco (Dec. 2019). "Targeted drugs ramp up cancer mutability". en. In: *Science* 366.6472, pp. 1452–1453.

Geva-Zatorsky, Naama et al. (June 2006). "Oscillations and variability in the p53 system". en. In: *Mol. Syst. Biol.* 2, p. 2006.0033.

Giordano-Attianese, Greta et al. (Apr. 2020). "A computationally designed chimeric antigen receptor provides a small-molecule safety switch for T-cell therapy". en. In: *Nat. Biotechnol.* 38.4, pp. 426–432.

Goverdhana, S et al. (Aug. 2005). "Regulatable gene expression systems for gene therapy applications: progress and future challenges". en. In: *Mol. Ther.* 12.2, pp. 189–211.

Green, Alexander A et al. (Nov. 2014). "Toehold switches: de-novo-designed regulators of gene expression". en. In: *Cell* 159.4, pp. 925–939.

Greenshpan, Yariv et al. (Jan. 2021). "Synthetic promoters to induce immune-effectors into the tumor microenvironment". en. In: *Commun Biol* 4.1, p. 143.

Hamburger, Agnes E et al. (Sept. 2020). "Engineered T cells directed at tumors with defined allelic loss". en. In: *Mol. Immunol.*

Hart, Yuval, Yaron E Antebi, et al. (May 2012). "Design principles of cell circuits with paradoxical components". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 109.21, pp. 8346–8351.

Hart, Yuval, Shlomit Reich-Zeliger, et al. (Aug. 2014). "Paradoxical signaling by a secreted molecule leads to homeostasis of cell levels". en. In: *Cell* 158.5, pp. 1022–1032.

Hennig, Stefan, Gerhard Rödel, and Kai Ostermann (Aug. 2015). "Artificial cell-cell communication as an emerging tool in synthetic biology applications". en. In: *J. Biol. Eng.* 9.1, pp. 1–12.

Hernandez-Lopez, Rogelio A et al. (Mar. 2021). "T cell circuits that sense antigen density with an ultrasensitive threshold". en. In: *Science* 371.6534, pp. 1166–1171.

Holland, John J (Dec. 2012). *Genetic Diversity of RNA Viruses*. en. Springer Science & Business Media.

Hu, Biliang et al. (Sept. 2017). "Augmentation of Antitumor Immunity by Human and Mouse CAR T Cells Secreting IL-18". en. In: *Cell Rep.* 20.13, pp. 3025–3033.

Huang, Ting-Hao et al. (Dec. 2017). "Tracing neuronal circuits in transgenic animals by transneuronal control of transcription (TRACT)". In: *Elife* 6, e32027.

Jackson, Hollie J, Sarwish Rafiq, and Renier J Brentjens (June 2016). "Driving CAR T-cells forward". en. In: *Nat. Rev. Clin. Oncol.* 13.6, pp. 370–383.

Keung, Albert J et al. (Mar. 2015). "Chromatin regulation at the frontier of synthetic biology". en. In: *Nat. Rev. Genet.* 16.3, pp. 159–171.

Khalil, Ahmad S et al. (Aug. 2012). "A synthetic biology framework for programming eukaryotic transcription functions". en. In: *Cell* 150.3, pp. 647–658.

Kipniss, Nathan H et al. (Dec. 2017). "Engineering cell sensing and responses using a GPCR-coupled CRISPR-Cas system". en. In: *Nat. Commun.* 8.1, pp. 1–10.

Kopan, Raphael and Maria Xenia G Ilagan (Apr. 2009). "The canonical Notch signaling pathway: unfolding the activation mechanism". en. In: *Cell* 137.2, pp. 216–233.

Lajoie, Marc J et al. (Sept. 2020). "Designed protein logic to target cells with precise combinations of surface antigens". en. In: *Science* 369.6511, pp. 1637–1643.

Luan, Xin et al. (Apr. 2017). "Engineering exosomes as refined biological nanoplatforms for drug delivery". en. In: *Acta Pharmacol. Sin.* 38.6, pp. 754–763.

Lutz, Halle et al. (Sept. 2019). "Cells and cell derivatives as drug carriers for targeted delivery". In: *Medicine in Drug Discovery* 3, p. 100014.

Ma, Yitong et al. (Sept. 2020). "Synthetic mammalian signaling circuits for robust cell population control". en.

Martinez, Marina and Edmund Kyung Moon (2019). "CAR T Cells for Solid Tumors: New Strategies for Finding, Infiltrating, and Surviving in the Tumor Microenvironment". In: *Front. Immunol.* 10, p. 128.

Mitchell, Michael J et al. (Feb. 2021). "Engineering precision nanoparticles for drug delivery". en. In: *Nat. Rev. Drug Discov.* 20.2, pp. 101–124.

Morsut, Leonardo et al. (Feb. 2016). "Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors". en. In: *Cell* 164.4, pp. 780–791.

Ng, Andrew H et al. (Aug. 2019). "Modular and tunable biological feedback control using a de novo protein switch". en. In: *Nature* 572.7768, pp. 265–269.

Potvin-Trottier, Laurent et al. (Oct. 2016). "Synchronous long-term oscillations in a synthetic gene circuit". en. In: *Nature* 538.7626, pp. 514–517.

Prelich, Gregory (Mar. 2012). "Gene overexpression: uses, mechanisms, and interpretation". en. In: *Genetics* 190.3, pp. 841–854.

Purvis, Jeremy E et al. (June 2012). "p53 dynamics control cell fate". en. In: *Science* 336.6087, pp. 1440–1444.

Qian, Lulu, Erik Winfree, and Jehoshua Bruck (July 2011). "Neural network computation with DNA strand displacement cascades". en. In: *Nature* 475.7356, pp. 368–372.

Rackham, Oliver and Jason W Chin (Aug. 2005). "A network of orthogonal ribosome x mRNA pairs". en. In: *Nat. Chem. Biol.* 1.3, pp. 159–166.

Rafiq, Sarwish, Christopher S Hackett, and Renier J Brentjens (Dec. 2019). "Engineering strategies to overcome the current roadblocks in CAR T cell therapy". en. In: *Nat. Rev. Clin. Oncol.* 17.3, pp. 147–167.

Roybal, Kole T et al. (Feb. 2016). "Precision Tumor Recognition by T Cells With Combinatorial Antigen-Sensing Circuits". en. In: *Cell* 164.4, pp. 770–779.

Sahin, Ugur, Katalin Karikó, and Özlem Türeci (Oct. 2014). "mRNA-based therapeutics–developing a new class of drugs". en. In: *Nat. Rev. Drug Discov.* 13.10, pp. 759–780.

Sanchez-Vega, Francisco et al. (Apr. 2018). "Oncogenic Signaling Pathways in The Cancer Genome Atlas". en. In: *Cell* 173.2, 321–337.e10.

Saxena, Pratik, Daniel Bojar, and Martin Fussenegger (2017). "Design of Synthetic Promoters for Gene Circuits in Mammalian Cells". en. In: *Methods Mol. Biol.* 1651, pp. 263–273.

Schlabach, Michael R et al. (Feb. 2010). "Synthetic design of strong promoters". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 107.6, pp. 2538–2543.

Stricker, Jesse et al. (Nov. 2008). "A fast, robust and tunable synthetic gene oscillator". en. In: *Nature* 456.7221, pp. 516–519.

Xie, Zhen et al. (Sept. 2011). "Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells". en. In: *Science* 333.6047, pp. 1307–1311.

Yeh, Brian J et al. (May 2007). "Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors". en. In: *Nature* 447.7144, pp. 596–600.

Ying, H et al. (July 1999). "Cancer therapy using a self-replicating RNA vaccine".
    en. In: *Nat. Med.* 5.7, pp. 823–827.

Zhu, Ronghui et al. (Feb. 2021). "Synthetic multistability in mammalian cells". en.

Chapter 2

ENGINEERING PROGRAMMABLE PROTEIN CIRCUITS IN MAMMALIAN CELLS

This work was published in Science as "Programmable protein circuits in living cells" by Gao XJ, Chong LS, Kim MS, Elowitz MB, 2018 and reproduced with permission. One additional results section, Fig. 2.9 was added, performed in collaboration with Michaela H. Ince and Mark W. Budde

2.1 Abstract

Synthetic protein-level circuits could enable engineering of powerful new cellular behaviors. Rational protein circuit design would be facilitated by a composable protein-protein regulation system, in which individual protein components can regulate one another to create a variety of different circuit architectures. Here, we show that engineered viral proteases can function as composable protein components, which can together implement a broad variety of circuit-level functions in mammalian cells. In this system, termed CHOMP (Circuits of Hacked Orthogonal Modular Proteases), input proteases dock with and cleave target proteases to inhibit their function. These components can be connected to generate regulatory cascades, binary logic gates, and dynamic analog signal-processing functions. To demonstrate the utility of this system, we rationally designed a circuit that induces cell death in response to upstream activators of the Ras oncogene. Because CHOMP circuits can perform complex functions yet be encoded as single transcripts and delivered without genomic integration, they offer a scalable platform to facilitate protein circuit engineering for biotechnological applications.

2.2 Main Text

Synthetic biology seeks to enable rational design of circuits that confer new functions in living cells. Most efforts thus far have centered on gene regulation because of the relative ease with which transcription factors and other nucleic-acid-interacting proteins can be configured to regulate one another's expression (Bonnet et al., 2013; Weinberg et al., 2017; Ausländer et al., 2012; Rinaudo et al., 2007; Wroblewska et al., 2015; Khalil et al., 2012; Roquet et al., 2016; Nielsen et al., 2016; Angelici et al., 2016; Lohmueller, Armel, and Silver, 2012). However, many nat-

ural cellular functions are implemented by protein-level circuits, in which proteins specifically modify each other's activity, localization, or stability. For example, caspase-mediated programmed cell death is regulated by a circuit of proteases that activate one another through cleavage (Budihardjo et al., 1999). Synthetic protein circuits could provide advantages over gene regulation circuits, including faster operation, direct coupling to endogenous pathways, single transcript delivery, and function without genomic integration (Fig. 2.1). The key challenge is designing 'composable' protein components whose inputs and outputs are of the same type, so that they can form a wide variety of protein circuits (Marchisio and Stelling, 2008), much as a few electronic components can be wired to produce a variety of electronic circuits (Fig. 2.1). While natural protein domains have been combined to generate proteins with hybrid functions or to re-wire cellular pathways for research (Yeh et al., 2007; Dueber et al., 2003; Park, Zarrinpar, and Lim, 2003; Howard et al., 2003; Barrett et al., 2014; Morsut et al., 2016) and biomedical applications (Barrett et al., 2014; Roybal et al., 2016), the lack of composability has limited our ability to design protein-level function in living cells.
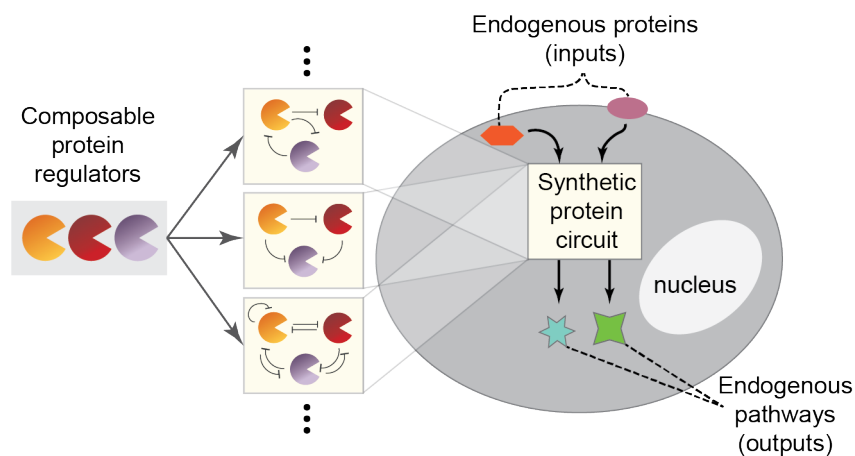
Figure 2.1 | **Composable protein circuits.** Composable protein units (partial circles, left) can regulate one another in arbitrary configurations with diverse functions (middle). Protein-level circuits can interface directly with endogenous protein pathways and operate without modifying the genome or entering the nucleus (right).

Viral proteases provide a promising basis for such a system (Stein and Alexandrov, 2014; Stein, Nabi, and Alexandrov, 2017). Many of them exhibit strong specificity for short cognate target sites, which can be recognized and cleaved in various protein contexts (Carrington and Dougherty, 1988; Tözsér et al., 2005; Bartenschlager, 1999). Natural viral diversity provides multiple proteases with distinct specificities (Adams, Antoniw, and Beaudoin, 2005). Viral proteases can be used with degrons
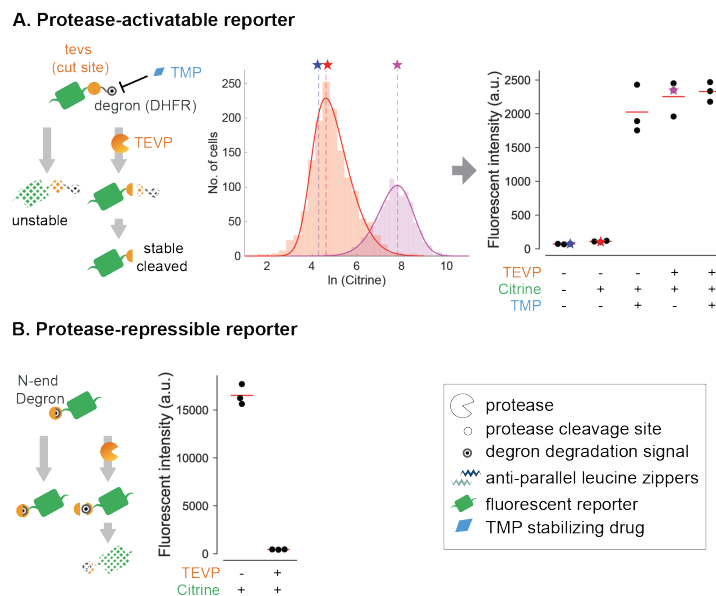
Figure 2.2 | **Design of protease reporters.** (A) (Left) The protease-activatable reporter (green) is stabilized by removal of a DHFR degron (black target) through protease (partial circle) cleavage of a corresponding target site (yellow circle). TMP (blue diamond) inhibits the degron and thus stabilizes the reporter. (Middle) Flow cytometry distributions of reporter fluorescence with (purple) or without (orange) TEVP. Distributions are limited to the gated area in fig. 2.10. Solid curves indicate skew Gaussian fits. Vertical dashed lines and stars indicate distribution modes, which are plotted in subsequent figure panels. (Right) Analysis of reporter response to TMP and/or TEVP. Each dot represents one replicate. Stars indicate data from the middle panel. a.u., arbitrary units. (C) In the protease-repressible reporter, protease cleavage exposes an N-end degron (covered target) to destabilize the reporter.

to control protein stability (Taxis et al., 2009; Butko et al., 2012; Chung et al., 2015). They can also activate transcription factors (Kipniss et al., 2017; Barnea et al., 2008; Daringer et al., 2014), synthetic intein zymogens (Gramespacher et al., 2017), and other proteases in a purified protein system (Stein and Alexandrov, 2014; Stein, Nabi, and Alexandrov, 2017).

We first focused on the well-characterized tobacco etch virus protease (TEVP) (Waugh, 2011). To quantify TEVP activity, we designed a reporter in which a cognate cleavage site (tevs) is inserted between a Citrine fluorescent protein and a dihydrofolate reductase (DHFR) degron, which can be inhibited by trimethoprim (TMP) as a positive control (Iwamoto et al., 2010) (Fig. 2.2A). We transfected human embryonic kidney (HEK293) cells with plasmids expressing different combinations of TEVP, the reporter, and an mCherry co-transfection marker, and analyzed cells by flow cytometry. We used the mCherry signal to select highly transfected cells,

which showed the largest separation of basal reporter fluorescence from cellular aut-ofluorescence to maximize the observable dynamic range of the reporter (Methods, Fig. 2.2, and Fig. 2.10). Treating cells with TEVP strongly increased reporter abundance to levels similar to those obtained by TMP inhibition of the degron (Fig. 2.2A, Fig. 2.11A, left). We also designed a complementary repressible reporter in which TEVP cleavage exposes a destabilizing N-terminal tyrosine residue (Taxis et al., 2009; Varshavsky, 1996) (Fig. 2.2B, Fig. 2.11A, right). These designs generalized in a straightforward way to the related tobacco vein mottling virus protease (TVMVP) (Nallamsetty et al., 2004) and with some modifications to the unrelated hepatitis C virus protease (HCVP) (Bartenschlager, 1999; Taremi et al., 1998) (Fig. 2.11B, Fig. 2.3A; Supplementary Text). Furthermore, measuring activation of each reporter in response to each protease revealed limited cross-activation (Fig. 2.3A). Thus, three viral proteases can be used to orthogonally increase or decrease cognate reporters.

To enable the design of complex circuits, we next sought to achieve protease-protease regulation. The degron strategy used for the reporters failed to produce strong regulation, possibly because proteases can cleave degrons within the same protease molecule with relaxed specificity (Chung et al., 2015). Instead, we designed a scheme that regulates protease activity, rather than abundance. We incorporated antiparallel hetero-dimerizing leucine zipper domains (Ghosh, Hamilton, and Regan, 2000) to each half of a split TEVP (Wehr et al., 2006) to reconstitute its activity (Fig. 2.3B, left). We also inserted HCVP cleavage sites between the leucine zippers and TEVP, to allow HCVP to inhibit TEVP. Finally, we fused a leucine zipper (complementary to one of the zippers on split TEVP) to HCVP, thus enhancing its ability to dock with, and inhibit its TEVP target (Fig. 2.12A, left). This design successfully produced repression of TEVP by HCVP (Fig. 2.3B, left).

To generalize this design, we engineered a similar TEVP variant repressed by TVMVP (Fig. 2.12A, right). Based on its sequence similarity to TEVP (Fig. 2.12B), we also engineered TVMVP variants repressed by either HCVP (Fig. 2.3B, right) or TEVP (Fig. 2.12C). To make these designs more compact, we linked the two halves of each regulated protease with a single leucine zipper flanked by cleavage sites for the input protease, creating single-chain repressible proteases (Fig. 2.3B and Fig. 2.12D,E). Similar approaches enabled us to engineer protease regulation of the unrelated protease HCVP using a different split strategy (Supplementary Text). In these constructs, cleavage by either TEVP or TVMVP strongly reduced HCVP
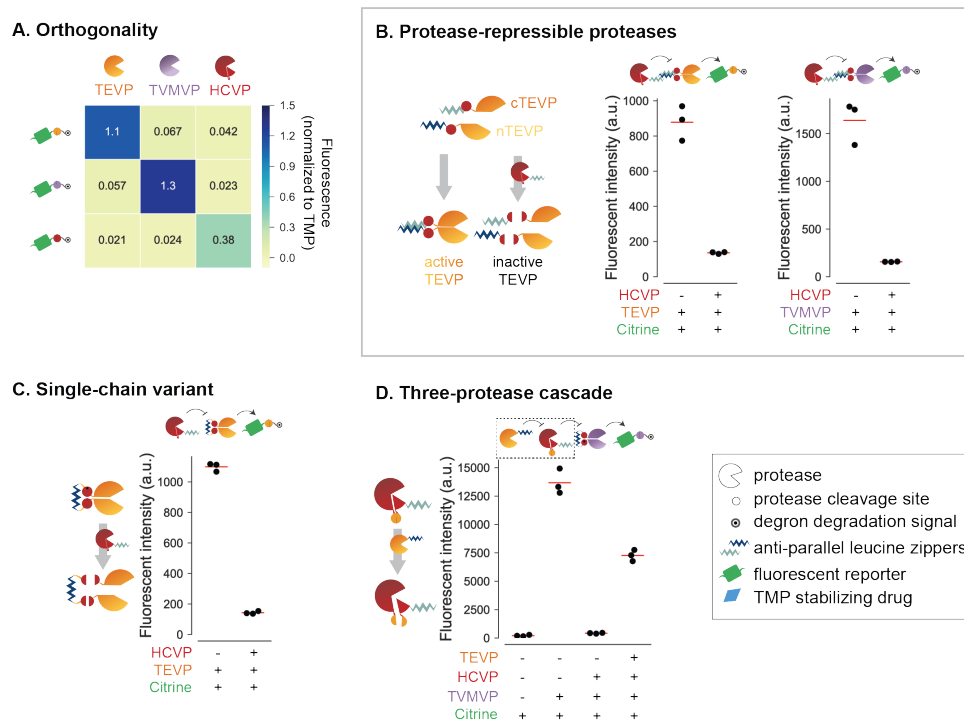
Figure 2.3 | **Design of composable protein circuits.** (A) Three proteases (columns) exhibit orthogonal regulation of three reporters (rows). Mean fluorescence intensity of three independent measurements is normalized to the TMP-stabilized value of the corresponding reporter. (B) Design for protease-repressible proteases. TEVP is split as indicated and then reconstituted through dimerizing leucine zippers (light and dark blue zig-zags). A leucine zipper–tagged HCVP (red partial circle) can dock with the target TEVP and cleave it to remove leucine zippers, effectively repressing TEVP. TVMVP (purple partial circle) can be regulated using the same design. (C) A single-chain variant of the HCV-repressible TEVP allows docking of and repressive cleavage by HCVP. (D) Protease regulation can propagate through a three-stage cascade. Repressible HCVP uses a variant design, in which TEVP cleavage separates core HCVP from its docking leucine zipper and activity-enhancing copeptide (small pie slice). In all panels, red lines indicate triplicate mean. (right).

activity, enabling signal propagation through three-stage protease cascades (Fig. 2.3D and Fig. 2.12F). Together, this strategy established a composable protease regulation system.
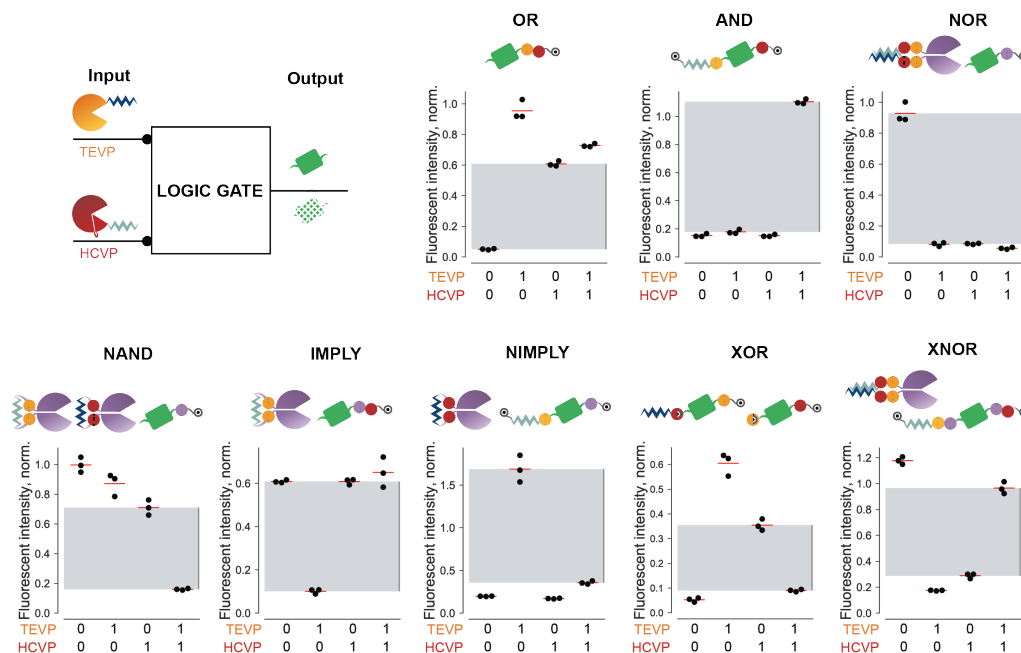
Figure 2.4 | **CHOMP circuits implement binary logic gates.** CHOMP circuits implement binary logic gates. For each indicated gate, TEVP and HCVP serve as binary inputs, which are either included or excluded in transfections. Citrine fluorescence serves as gate output. The design and performance of each nontrivial two-input logic gate is shown for triplicate experiments (black dots). Fluorescence intensity in each panel is normalized to the corresponding reporter stabilized with TMP (for gates containing only C-terminal degrons) or Shield-1 plus TMP (for gates containing degrons at both termini). Gray regions indicate the range from maximum "OFF" value to minimum "ON" value for that gate. (right).

Using this system, we designed core circuit functions, starting with Boolean logic. We identified three design principles that together would be sufficient to enable all 8 two-input gates: First, incorporation of a consecutive pair of distinct cleavage sites between a degron and a target protein can implement OR logic, since cleavage of either site is sufficient to stabilize the protein (Fig. 2.4 and 2.13, OR). Second, to implement AND logic, we flanked the target protein with FKBP (Banaszynski et al., 2006) and DHFR degrons on the N- and C-termini, respectively, each removable with a distinct cleavage site. On the N-terminus, a leucine zipper was necessary to facilitate input protease docking. In this design, removal of both degrons was necessary to stabilize the protein (Fig. 2.4 and 2.13, AND). Third, to implement

negation, we either used the N-end degron strategy (Fig. (2.2B) or propagated signals through an intermediate protease repression step (Fig. 2.3B). Co-transfection of each basic gate (OR, AND, and NOR as a specific case of negation) with varying concentrations of its inputs revealed the expected logic functions (Fig. 2A, 2.14A). Further, varying the concentration of the reporter plasmid enabled tuning of output levels without disrupting the logical computation, facilitating matching of input and output levels in more complex circuits (Fig. 2.14B). Finally, by utilizing the HCVP inhibitor asunaprevir (ASV) (28) and a rapamycin-induced TEVP (40, 42), we found that these gates could also be controlled by small molecule inputs (Fig. 2.15A). These results thus show that three core gates exhibit robust and tunable operation across multiple input methods.

Next, we combined these principles to design and validate the other two-input gates (Fig. 2.4 and 2.13). Furthermore, to test whether output from one gate could be directly used as input to a subsequent gate, we constructed a more complex nested NOR function using additional orthogonal proteases from soybean mosaic virus (SMVP) (Ghabrial et al., 1990) and herpes simplex virus (HSVP) (Weinheimer et al., 1993) (Fig. 2.15B). The output from this system was consistent with that expected from the logical function NOR(TEVP, NOR(SMVP, HSVP)) (Fig. 2.15B).

Beyond Boolean logic, analog signal filtering can enable many cellular functions, such as the ability to selectively respond to specific input concentration ranges (Hart and Alon, 2013; Porcher and Dostatni, 2010). The incoherent feed-forward loop (IFFL) motif, in which an input both activates and inhibits the same target, provides a simple implementation for this function (Basu et al., 2004a; Greber and Fussenegger, 2010). Inspired by the IFFL, we combined an activating arm, in which TEVP removes a C-terminal degron, with a repressing arm, in which TVMVP reveals a destabilizing N-end tyrosine (Fig. 2.5A). To tune the position and sharpness of the bandpass, we also introduced a positive feedback loop based on reciprocal inhibition between HCVP and TVMVP on the repression arm, such that the amount of HCVP expression sets a threshold for TVMVP activity (Fig. 2.5A).

To characterize this bandpass circuit, we considered the abundance of TEVP and TVMVP as input, and varied it through the concentration of transfected DNA, which correlated linearly with protein abundance (Fig. 2.16A). The individual activating and repressing arms of the circuit generated increasing and decreasing responses, respectively, to increasing amounts of TEVP and TVMVP (Fig. 2.5B, C). Addition of HCVP increased both the threshold and the sharpness of the response to TVMVP
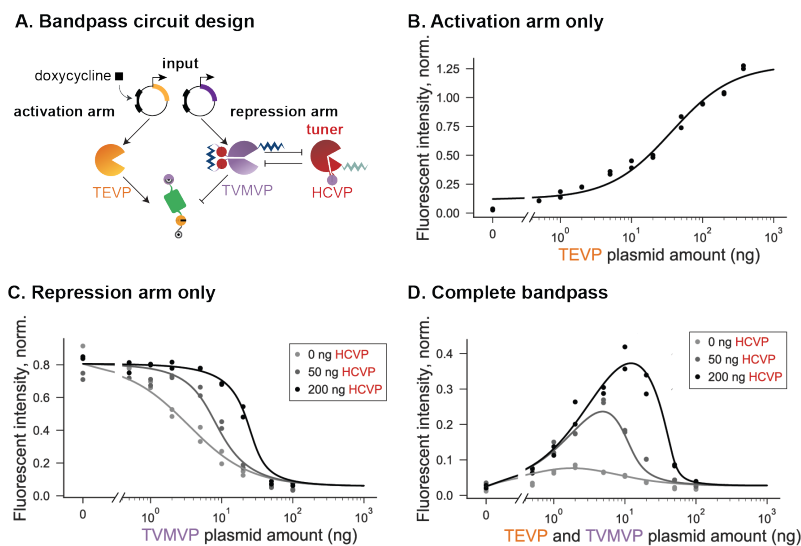
Figure 2.5 | **Design of pulse-generating circuit.** (A) For bandpass filtering, the expression of co-regulated inputs TEVP (yellow) and TVMVP (purple) are controlled by the amount of transfected DNA or by doxycycline (square) induction. The amount of HCVP (red) plasmid can be varied to tune the repression arm. (B) Input-output curve of the activation arm in the absence of TVMVP. Here and in subsequent panels, dots indicate duplicate measurements, and the curve is a model fit (materials and methods). (C) Input-output curve of the repression arm, in the presence of constant TEVP and increasing levels of HCVP (gray shades), which increases the repression threshold and sharpens the response. (D) Bandpass behavior of the complete circuit. Increasing HCVP expression (gray shades) shifts the position and increases the amplitude of the peak response. Data in (B) to (D) are normalized to the TMP-stabilized reporter. (right).

titration (Fig. 2.5C). Combining the two arms into a single circuit generated the anticipated bandpass behavior, when we co-varied TEVP and TVMVP expression through either different amounts of plasmid (Fig. 2.5D) or 4-epitetracycline (4-epi Tc) induction (Fig. 2.16B). Finally, varying the abundance of HCVP tuned the position and amplitude of the bandpass response (Fig. 2.5D and Fig. 2.16B). These results demonstrate rational engineering of tunable analog bandpass filters.

Temporal signal processing, such as adaptation to a change in input, has a critical role in diverse biological systems (Ma et al., 2009), and has been demonstrated synthetically in bacteria at the gene regulation level (Basu et al., 2004b). To engineer adaptation with CHOMP, we designed an IFFL, containing the 3-step cascade (Fig. 2.3D) to introduce a delay in the repressing arm relative to that of the activating

arm (Fig. 2.6A). To enable sudden induction, we adopted the rapamycin-induced TEVP used for the logic gates (Figs. 2.15A, 2.16C). To facilitate dynamic readout of circuit output in individual cells we used a far-red fluorescent protein (IFP) that is synthesized in a non-fluorescent state, but can be post-translationally switched on by TEVP (To et al., 2015) (Fig 2.16D, left). We also added a conditional N-end degron to enable repression by TVMVP (Fig. 2.6A).

We encoded the entire pulse-generation circuit as a single open reading frame, with interleaved 2A "self-cleaving" peptides (Szymczak et al., 2004) to separate distinct protein components (Fig. 2.6B). This gene was then stably incorporated in the genome (Methods). We used flow cytometry to analyze the response of the reporter in a single clone over time after rapamycin addition. Cells exhibited the expected adaptive dynamics, with a rise in fluorescence on a timescale of hours and a subsequent decay to baseline over 1 day (Fig. 2.16D, right). To obtain a direct view of dynamics in individual cells, we also analyzed the same cell line by time-lapse fluorescence microscopy (Fig. 2.6C, Movie S1). Analysis of individual cells revealed similar adaptive dynamics, responding maximally at 269±68 (mean±s.d.) min after rapamycin addition, decaying to half of their peak values over the subsequent 491±170 min, and eventually reaching fluorescence similar to that before induction (Fig. 2.6D). These results demonstrate the design of single-gene multi-component circuits that generate dynamic signal responses.

By coupling directly to endogenous cellular outputs and inputs, protein-level circuits could act as programmable therapeutic devices. As a proof of principle for such a strategy, we designed a circuit to selectively kill cells with elevated activation of Ras, a protein whose activity is increased in many cancers (Cox et al., 2014; Downward, 2003). More specifically, we designed a core circuit that responds to upstream activators of Ras, such as SOS and EGFR, by activating an engineered TEV protease, which in turn activates Caspase-3 (Casp3) to induce cell death (Budihardjo et al., 1999; Gray, Mahrus, and Wells, 2010) (Fig. 2.7, core circuit). We then improved this circuit by incorporating additional proteases and interactions (Fig. 2.7, full circuit).

To enable efficient protease-dependent induction of cell death at the plasma membrane, where Ras activation occurs, we membrane localized a TEVP-activated Casp3 variant (Gray, Mahrus, and Wells, 2010) by incorporating the 20 amino acid membrane-targeting sequence ('mts') from the C-terminus of human H-Ras (56) (Fig. 2.7, box 2). Using flow cytometry, we quantified the effect of this Casp3
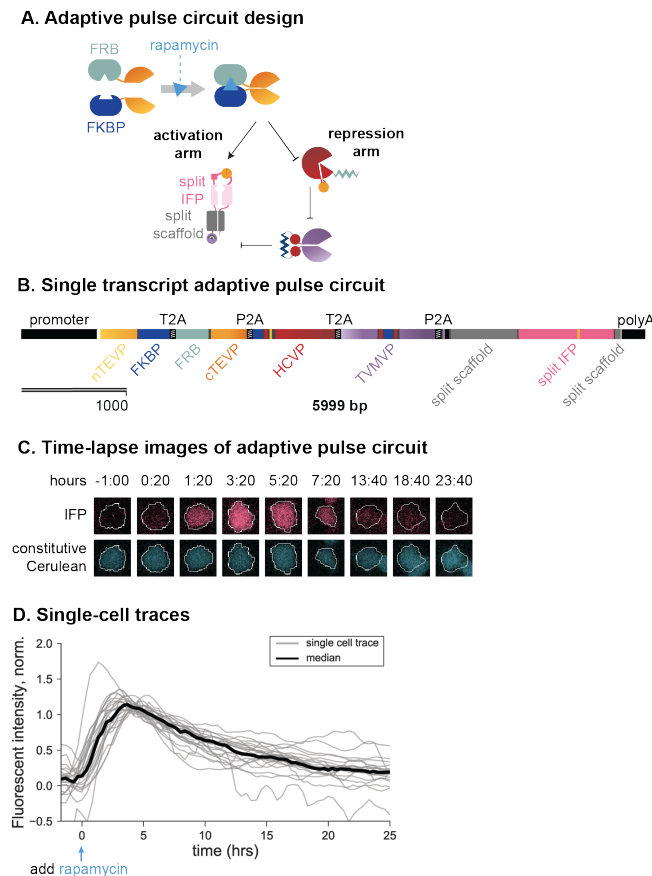
Figure 2.6 | **Design of pulse-generating circuit.** (A) Delayed repression can enable pulse generation. In this design, rapamycin-induced dimerization of FKBP and FRB domains reconstitutes TEVP. Cleavage of the reporter by TEVP allows maturation of far-red fluorescent protein (IFP, pink) (fig. 2.16D). (B) The pulse circuit was completely encoded on a single transcript, with protein components (indicated) separated by self-cleaving sequences (T2A and P2A) (47). (C) Filmstrips of a single cell stably incorporating both the pulse-generation circuit (pink) and a constitutive cerulean segmentation marker (blue). After rapamycin induction (t = 0), the output IFP signal (pink) increases and then decays, whereas the cerulean signal (blue) remains constant. (D) Traces of IFP fluorescence in 24 individual cells (gray lines). This analysis omits cells that exhibited phototoxicity or moved out of the field of view (see materials and methods). The black line indicates median fluorescence over all cells at each time point. (right).
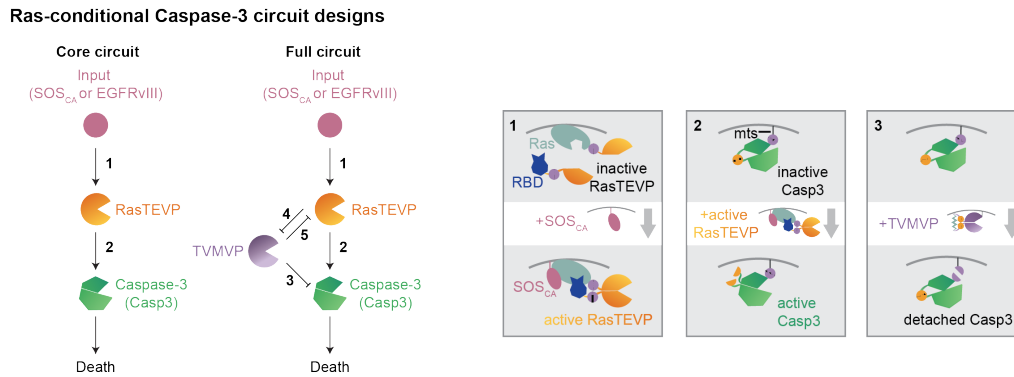
Figure 2.7 | **Design of a Ras-activation detecting circuit.** The core circuit (left) links Ras activation by SOSCA or EGFRvIII to Casp3 activation. The full circuit (right) incorporates an additional TVMVP component to enhance selectivity. New regulatory features introduced in this circuit are explained schematically in the corresponding numbered boxes. Box 1: Input from upstream activators of Ras such as SOSCA and EGFRvIII (pink) activates Ras (light blue), causing it to bind RBD (dark blue), reconstituting RasTEVP. Box 2: Engineered Casp3 (green) tagged with a membrane localization sequence (mts) can be converted from an inactive to an active state by TEVP cleavage. Box 3: TVMVP cleavage detaches Casp3 from the membrane, reducing its ability to be activated by membrane-localized TEVP. (right).

variant on cell numbers in terms of a 'reduction index' whose value measures the relative reduction in cell number compared to a control condition (Methods, Fig. 2.17B). The membrane-targeted Casp3 decreased cell numbers when co-transfected with a similarly membrane-localized TEVP variant (Fig. 2.8A), with higher efficiency than the original cytoplasmic Casp3 variant (Fig. 2.18B). Further, to allow bidirectional regulation by TEVP and TVMVP, we also incorporated a TVMVP cleavage site adjacent to the mts tag (Fig. 2.7, Box 3), enabling membrane-localized TVMVP to remove Casp3 from the membrane and thereby attenuate its activation by TEVP (Fig. 2.8A).

Next, to couple Ras-activating inputs to TEVP, we fused the N-terminal half of TEVP to Ras and its C-terminal half to the Ras-binding domain (RBD) of Raf, which binds to the active form of Ras (Oliveira and Yasuda, 2013; Yasuda et al., 2006). In this design, upstream activators of Ras should reconstitute RasTEVP ((Fig. 2.7), core circuit and Box 1, Fig. 2.18A) and thereby activate Casp3. To validate this design, we constructed a HEK293 cell line stably expressing a constitutively active Son of Sevenless (SOSCA) variant with a membrane-localization myristoylation signal and no inhibitory C-terminal region (Aronheim et al., 1994). Transfection of the core

A. Protease regulation of Caspase-3

TEVP	+	-	+	+
TVMVP	-	-	-	+
Casp3	-	+	+	+

B. Comparison of core and complete circuits

core circuit full circuit

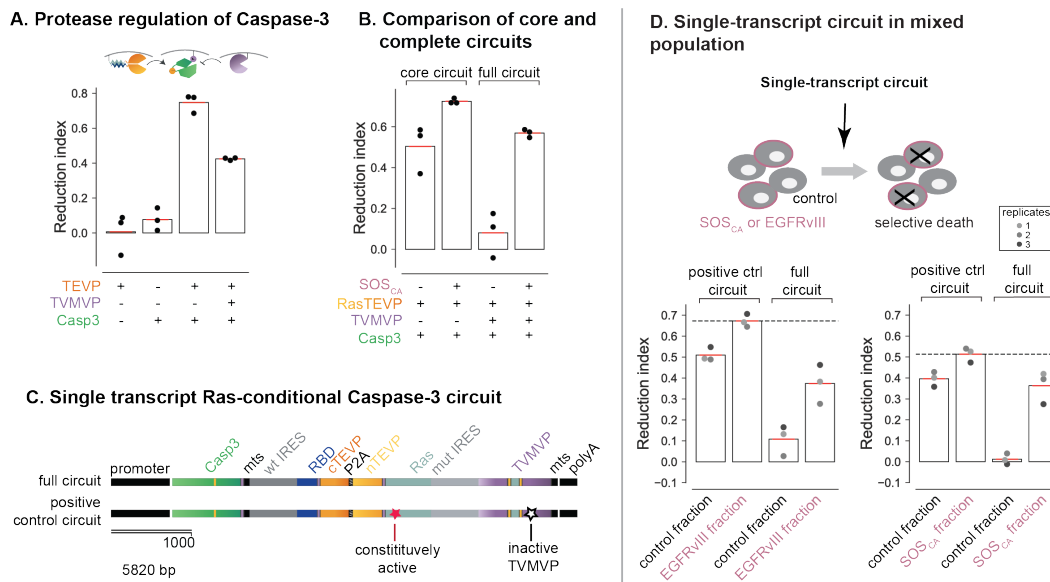|				
SOS$_{CA}$	-	+	-	+
RasTEVP	+	+	+	+
TVMVP	-	-	+	+
Casp3	+	+	+	+

C. Single transcript Ras-conditional Caspase-3 circuit

full circuit
positive control circuit

1000
5820 bp

constitutively active
inactive TVMVP

D. Single-transcript circuit in mixed population

Single-transcript circuit

control → selective death

SOS$_{CA}$ or EGFRvIII

replicates 1 2 3

positive ctrl circuit / full circuit

control fraction / EGFRvIII fraction

positive ctrl circuit / full circuit

control fraction / SOS$_{CA}$ fraction

Figure 2.8 | **CHOMP circuit enables conditional activation of Casp3 in Ras-activating cells.** (A) TEVP activates the engineered Casp3, whereas TVMVP inhibits this activation. Cells transfected with the indicated components were analyzed to determine the reduction index (percentage of cell number reduction compared to cells transfected with only a fluorescent marker; see materials and methods and fig. 2.17B). (B) The core circuit preferentially reduced cell number in the presence of ectopic SOSCA. The full circuit exhibited improved selectivity. (C) The full circuit (top) and a positive-control circuit incorporating a Gly12→Val mutation that makes Ras constitutively active and a Cys152→Ala mutation that abolishes TVMVP activity (bottom) were each encoded as a single transcript. (D) In a mixed population, the single-transcript circuit (D, top) conditionally reduced the number of EGFRvIII cells (left) and SOSCA cells (right) compared with that of cocultured control cells. The positive-control circuit (D, bottom) reduced the number of both fractions. The dashed line indicates the upper limit of the reduction index measured with the positive-control circuit. Dots from the same well are color matched. (right).

circuit reduced cell numbers both in this SOSCA cell line and its parental control line lacking ectopic SOSCA, but preferentially affected the SOSCA cells (Fig. 2.8B, core circuit, and Fig. 2.18D), and required the regulated Ras-RBD interaction for selectivity (Fig. 2.18C, Supplementary Text). However, while this core circuit provided some selectivity, it also exhibited a relatively high background rate of Casp3 activation in the control cells.

To improve the circuit's selectivity, we incorporated a TVMVP-TEVP reciprocal inhibition motif (Fig. 2.17A, boxes 4,5) similar to the one used in the bandpass circuit, as well as feed-forward repression of Casp3 activation by TVMVP ((Fig. 2.7), Box 3). In this "full circuit" design, TVMVP should suppress activation of Casp3

in control cells, both directly and indirectly through TEVP. By contrast, in SOSCA cells, elevated activation of TEVP should override the inhibitory effects of TVMVP. The full circuit indeed improved selectivity (Figs. 2.8B, 2.18E, 2.19, Supplementary Text). More specifically, expressing TVMVP in amounts comparable to, but lower than those of TEVP nearly abolished off-target effects in control cells, while retaining most of the on-target reduction in cell number (Figs. 2.8B, 2.18E).

To simulate a more biomedically relevant context, we encoded the full 4-protein circuit on a single transcript, optimizing the relative abundance of components with internal ribosome entry site (IRES) variants (Koh et al., 2013) (Fig. 2.8C, 2.20A, and Supplementary Text), and transfected it into a mixed population of SOSCA and control cells. At its optimal concentration (Fig. 2.20B), the single-transcript circuit reduced the number of SOSCA cells by 40%, approaching the 50% upper limit achieved by a positive control circuit that constitutively activates Casp3 (Fig. 2.8C, and Fig. 2.8D, right). (The upper limit is constrained by gene delivery and expression efficiency.) Importantly, it exhibited minimal effects on the control population (Fig. 2.8D, right). SOSCA-dependent killing could also be observed using Annexin-V staining as an independent readout of apoptosis (Fig. 2.20C). Finally, to test the generality of the circuit, we considered a distinct, and more biomedically relevant input, EGFRvIII, an oncogenic EGFR mutant found in glioblastoma and other cancer types (Wikstrand et al., 1998). The single-transcript full circuit also selectively killed EGFRvIII cells (Fig. 2.8D, left, and Fig. 2.20C).

This initial circuit design incorporated full-length Ras, limiting its ability to capture activating mutations in Ras itself. Active Ras will co-localize and recruits proteins containing Ras-binding domains (RBDs) from the cytoplasm to the cellular membrane (Downward, 2003; Huang et al., 2019). By fusing each half of split TEVP to an RBD (RBDTEVP), Ras activity can increase local concentration of protease domains, leading to reconstitution of functional proteases (Fig. 2.9A).

The post-translational nature of protein circuits enables mRNA-based delivery. In vitro transcribed (IVT) chemically-modified mRNA has revolutionized medicine as a transient delivery system for genetic material (Sahin, Karikó, and Türeci, 2014). Typically, these mRNA-based therapies encode for single-genes, but has yet to be fully extended to the level of multi-gene circuits (Sahin, Karikó, and Türeci, 2014; Sample et al., 2019). To establish that capability, we asked if CHOMP components could function as effectively when delivered as mRNA. We transcribed RBD-TEVP and Casp3 as chemically modified, N1-methyl-pseudouridinylated mRNA (Svitkin
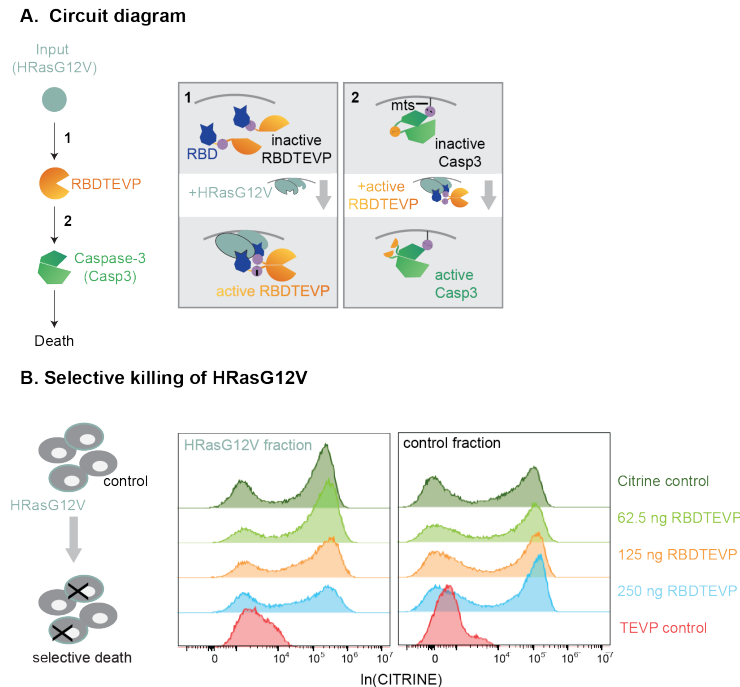
Figure 2.9 | **Design of a Ras-activity detecting circuit.** (A) The core circuit (left) links Ras activity, such as that found with Ras oncogenes, to Casp3 activation. Box 1: Input from active Ras (light blue), will co-localize and recruit RBD (dark blue) to the membrane, reconstituting RBDTEVP. Box 2: Engineered Casp3 (green) tagged with a membrane localization sequence (mts) can be converted from an inactive to an active state by TEVP cleavage. (B) RBDTEVP enables selective killing of HRasG12V cells in a mixed population. mRNA-encoded circuit components conditionally reduced the population of HRasG12V cells (left) compared to the cocultured control cells (right) in a titration-dependent manner. The constitutively active TEVP control reduced cells in both fractions.

et al., 2017). To validate this sensor design in a more clinically relevant context, we integrated HRasG12V into HEK293 cells and used a TEVP-activatable Casp3 as our output. We transfected these mRNA components into a co-culture of HEK293 and HRasG12V cells. HRasG12V cells were selectively and efficiently killed demonstrating the feasibility of an mRNA-based circuit delivery approach (Fig. 2.9B).

We found selective killing in cells overexpressing HRasG12V. Together, these results show that a CHOMP circuit can be engineered to detect and kill in response to both upstream activators of Ras and Ras activity through rational iterative design optimization.

The astonishing diversity of natural cellular behaviors stems from the flexibility with which regulatory components can form distinct circuits. Our results demonstrate

how a set of composable protein regulators and circuit design principles can enable a remarkably broad range of protein-based circuits and functions. The use of a small number of composable components shifts the design problem, in part, from the level of the individual protein to the level of the protein circuit. Because the operation of CHOMP components does not depend on how they are expressed, they can be optimized through transient transfections, accelerating the overall design-build-test cycle. Although powerful, CHOMP could be improved with additional features. Protease-activating proteases would simplify some circuit designs and facilitate signal amplification. Protein design strategies to control the intrinsic nonlinearity (effective cooperativity) of input-output responses could enable the implementation of useful dynamics such as multistability (Gardner, Cantor, and Collins, 2000) and oscillation (Elowitz and Leibler, 2000; Stricker et al., 2008). With one exception, the circuits shown here were created with three proteases, but additional orthogonal proteases would allow larger and more complex circuits (Adams, Antoniw, and Beaudoin, 2005). Finally, future work could expand the range of CHOMP inputs and outputs, enabling direct sensing of the activities of Ras and other oncogenes, and allow for combinatorial sensing of multiple inputs.

CHOMP circuits could provide distinct capabilities compared to transcriptional systems. In terms of speed, proteases can respond rapidly to an increase in input protease activity (Fig. S7, Supplementary Text). CHOMP circuits can also operate in parallel at specific subcellular sites within a cell. Because CHOMP circuits have a relatively compact genetic design and do not require regulatory interactions with DNA, they could be introduced into differentiated and even post-mitotic cells with gene therapy vectors or other viruses and improve the specificity of oncolytic virotherapy (Russell, Peng, and Bell, 2012). Synthetically, hybrid circuits combining transcriptional or translational regulation with engineered proteases could offer the programmability of base-pairing interactions together with protein level operation. For example, existing cancer-detection circuits (Nissim and Bar-Ziv, 2010; Xie et al., 2011) could conditionally express CHOMP components to increase specificity and couple to protein-mediated inputs and outputs. Integrating these capabilities, one can envision smart therapeutics or sentinels based on CHOMP circuits (Kojima, Aubel, and Fussenegger, 2016; Lienert et al., 2014).

References

Adams, Michael J, John F Antoniw, and Frederic Beaudoin (July 2005). "Overview and analysis of the polyprotein cleavage sites in the family Potyviridae". en. In: *Mol. Plant Pathol.* 6.4, pp. 471–487.

Angelici, Bartolomeo et al. (Aug. 2016). "Synthetic Biology Platform for Sensing and Integrating Endogenous Transcriptional Inputs in Mammalian Cells". en. In: *Cell Rep.* 16.9, pp. 2525–2537.

Aronheim, A et al. (Sept. 1994). "Membrane targeting of the nucleotide exchange factor Sos is sufficient for activating the Ras signaling pathway". en. In: *Cell* 78.6, pp. 949–961.

Ausländer, Simon et al. (July 2012). "Programmable single-cell mammalian bio-computers". en. In: *Nature* 487.7405, pp. 123–127.

Banaszynski, Laura A et al. (Sept. 2006). "A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules". en. In: *Cell* 126.5, pp. 995–1004.

Barnea, Gilad et al. (Jan. 2008). "The genetic design of signaling cascades to record receptor activation". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 105.1, pp. 64–69.

Barrett, David M et al. (2014). "Chimeric antigen receptor therapy for cancer". en. In: *Annu. Rev. Med.* 65, pp. 333–347.

Bartenschlager, R (May 1999). "The NS3/4A proteinase of the hepatitis C virus: unravelling structure and function of an unusual enzyme and a prime target for antiviral therapy". en. In: *J. Viral Hepat.* 6.3, pp. 165–181.

Basu, Subhayu et al. (Apr. 2004a). "Spatiotemporal control of gene expression with pulse-generating networks". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 101.17, pp. 6355–6360.

– (Apr. 2004b). "Spatiotemporal control of gene expression with pulse-generating networks". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 101.17, pp. 6355–6360.

Bonnet, Jerome et al. (May 2013). "Amplifying genetic logic gates". en. In: *Science* 340.6132, pp. 599–603.

Budihardjo, I et al. (1999). "Biochemical pathways of caspase activation during apoptosis". en. In: *Annu. Rev. Cell Dev. Biol.* 15, pp. 269–290.

Butko, Margaret T et al. (Dec. 2012). "Fluorescent and photo-oxidizing TimeS-TAMP tags track protein fates in light and electron microscopy". en. In: *Nat. Neurosci.* 15.12, pp. 1742–1751.

Carrington, J C and W G Dougherty (May 1988). "A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 85.10, pp. 3391–3395.

Chung, Hokyung K et al. (Sept. 2015). "Tunable and reversible drug control of protein production via a self-excising degron". en. In: *Nat. Chem. Biol.* 11.9, pp. 713–720.

Cox, Adrienne D et al. (Nov. 2014). "Drugging the undruggable RAS: Mission possible?" en. In: *Nat. Rev. Drug Discov.* 13.11, pp. 828–851.

Daringer, Nichole M et al. (2014). "Modular extracellular sensor architecture for engineering mammalian cell-based devices". In: *ACS Synth. Biol.* 3.12, pp. 892–902.

Downward, Julian (Jan. 2003). "Targeting RAS signalling pathways in cancer therapy". en. In: *Nat. Rev. Cancer* 3.1, pp. 11–22.

Dueber, John E et al. (Sept. 2003). "Reprogramming control of an allosteric signaling switch through modular recombination". en. In: *Science* 301.5641, pp. 1904–1908.

Elowitz, Michael B and Stanislas Leibler (Jan. 2000). "A synthetic oscillatory network of transcriptional regulators". In: *Nature* 403, p. 335.

Gardner, T S, C R Cantor, and J J Collins (Jan. 2000). "Construction of a genetic toggle switch in Escherichia coli". en. In: *Nature* 403.6767, pp. 339–342.

Ghabrial, S A et al. (Sept. 1990). "Molecular genetic analyses of the soybean mosaic virus NIa proteinase". en. In: *J. Gen. Virol.* 71 (Pt 9), pp. 1921–1927.

Ghosh, Indraneel, Andrew D Hamilton, and Lynne Regan (June 2000). "Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein". In: *J. Am. Chem. Soc.* 122.23, pp. 5658–5659.

Gramespacher, Josef A et al. (June 2017). "Intein Zymogens: Conditional Assembly and Splicing of Split Inteins via Targeted Proteolysis". en. In: *J. Am. Chem. Soc.* 139.24, pp. 8074–8077.

Gray, Daniel C, Sami Mahrus, and James A Wells (Aug. 2010). "Activation of specific apoptotic caspases with an engineered small-molecule-activated protease". en. In: *Cell* 142.4, pp. 637–646.

Greber, David and Martin Fussenegger (Oct. 2010). "An engineered mammalian band-pass network". en. In: *Nucleic Acids Res.* 38.18, e174.

Hart, Yuval and Uri Alon (Jan. 2013). "The utility of paradoxical components in biological circuits". en. In: *Mol. Cell* 49.2, pp. 213–221.

Howard, Perry L et al. (Sept. 2003). "Redirecting tyrosine kinase signaling to an apoptotic caspase pathway through chimeric adaptor proteins". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 100.20, pp. 11267–11272.

Huang, William Y C et al. (Mar. 2019). "A molecular assembly phase transition and kinetic proofreading modulate Ras activation by SOS". en. In: *Science* 363.6431, pp. 1098–1103.

Iwamoto, Mari et al. (Sept. 2010). "A general chemical method to regulate protein stability in the mammalian central nervous system". en. In: *Chem. Biol.* 17.9, pp. 981–988.

Khalil, Ahmad S et al. (Aug. 2012). "A synthetic biology framework for programming eukaryotic transcription functions". en. In: *Cell* 150.3, pp. 647–658.

Kipniss, Nathan H et al. (Dec. 2017). "Engineering cell sensing and responses using a GPCR-coupled CRISPR-Cas system". en. In: *Nat. Commun.* 8.1, pp. 1–10.

Koh, Esther Y C et al. (Dec. 2013). "An internal ribosome entry site (IRES) mutant library for tuning expression level of multiple genes in mammalian cells". en. In: *PLoS One* 8.12, e82100.

Kojima, Ryosuke, Dominique Aubel, and Martin Fussenegger (Oct. 2016). "Toward a world of theranostic medication: Programming biological sentinel systems for therapeutic intervention". en. In: *Adv. Drug Deliv. Rev.* 105.Pt A, pp. 66–76.

Lienert, Florian et al. (Feb. 2014). "Synthetic biology in mammalian cells: next generation research tools and therapeutics". en. In: *Nat. Rev. Mol. Cell Biol.* 15.2, pp. 95–107.

Lohmueller, Jason J, Thomas Z Armel, and Pamela A Silver (June 2012). "A tunable zinc finger-based framework for Boolean logic computation in mammalian cells". en. In: *Nucleic Acids Res.* 40.11, pp. 5180–5187.

Ma, Wenzhe et al. (Aug. 2009). "Defining network topologies that can achieve biochemical adaptation". en. In: *Cell* 138.4, pp. 760–773.

Marchisio, M A and J Stelling (Sept. 2008). "Computational design of synthetic gene circuits with composable parts". en. In: *Bioinformatics* 24.17, pp. 1903–1910.

Morsut, Leonardo et al. (Feb. 2016). "Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors". en. In: *Cell* 164.4, pp. 780–791.

Nallamsetty, Sreedevi et al. (Nov. 2004). "Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro". en. In: *Protein Expr. Purif.* 38.1, pp. 108–115.

Nielsen, Alec A K et al. (Apr. 2016). "Genetic circuit design automation". en. In: *Science* 352.6281, aac7341.

Nissim, Lior and Roy H Bar-Ziv (Jan. 2010). "A tunable dual-promoter integrator for targeting of cancer cells". en. In: *Mol. Syst. Biol.* 6.1, p. 444.

Oliveira, Ana F and Ryohei Yasuda (Jan. 2013). "An improved Ras sensor for highly sensitive and quantitative FRET-FLIM imaging". en. In: *PLoS One* 8.1, e52874.

Park, Sang-Hyun, Ali Zarrinpar, and Wendell A Lim (Feb. 2003). "Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms". en. In: *Science* 299.5609, pp. 1061–1064.

Porcher, Aude and Nathalie Dostatni (Mar. 2010). "The bicoid morphogen system". en. In: *Curr. Biol.* 20.5, R249–54.

Rinaudo, Keller et al. (July 2007). "A universal RNAi-based logic evaluator that operates in mammalian cells". en. In: *Nat. Biotechnol.* 25.7, pp. 795–801.

Roquet, Nathaniel et al. (July 2016). "Synthetic recombinase-based state machines in living cells". en. In: *Science* 353.6297, aad8559.

Roybal, Kole T et al. (Feb. 2016). "Precision Tumor Recognition by T Cells With Combinatorial Antigen-Sensing Circuits". en. In: *Cell* 164.4, pp. 770–779.

Russell, Stephen J, Kah-Whye Peng, and John C Bell (July 2012). "Oncolytic virotherapy". en. In: *Nat. Biotechnol.* 30.7, pp. 658–670.

Sahin, Ugur, Katalin Karikó, and Özlem Türeci (Oct. 2014). "mRNA-based therapeutics–developing a new class of drugs". en. In: *Nat. Rev. Drug Discov.* 13.10, pp. 759–780.

Sample, Paul J et al. (July 2019). "Human 5' UTR design and variant effect prediction from a massively parallel translation assay". en. In: *Nat. Biotechnol.* 37.7, pp. 803–809.

Stein, Viktor and Kirill Alexandrov (Nov. 2014). "Protease-based synthetic sensing and signal amplification". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 111.45, pp. 15934–15939.

Stein, Viktor, Masuda Nabi, and Kirill Alexandrov (July 2017). "Ultrasensitive Scaffold-Dependent Protease Sensors with Large Dynamic Range". en. In: *ACS Synth. Biol.* 6.7, pp. 1337–1342.

Stricker, Jesse et al. (Nov. 2008). "A fast, robust and tunable synthetic gene oscillator". en. In: *Nature* 456.7221, pp. 516–519.

Svitkin, Yuri V et al. (June 2017). "N1-methyl-pseudouridine in mRNA enhances translation through eIF2α-dependent and independent mechanisms by increasing ribosome density". en. In: *Nucleic Acids Res.* 45.10, pp. 6023–6036.

Szymczak, Andrea L et al. (Apr. 2004). "Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide–based retroviral vector". In: *Nat. Biotechnol.* 22, p. 589.

Taremi, S S et al. (1998). "Construction, expression, and characterization of a novel fully activated recombinant single-chain hepatitis C virus protease". In: *Proteins*.

Taxis, Christof et al. (Jan. 2009). "Efficient protein depletion by genetically controlled deprotection of a dormant N-degron". en. In: *Mol. Syst. Biol.* 5.1, p. 267.

To, Tsz-Leung et al. (Mar. 2015). "Rationally designed fluorogenic protease reporter visualizes spatiotemporal dynamics of apoptosis in vivo". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 112.11, pp. 3338–3343.

Tözsér, József et al. (Jan. 2005). "Comparison of the substrate specificity of two potyvirus proteases". en. In: *FEBS J.* 272.2, pp. 514–523.

Varshavsky, A (Oct. 1996). "The N-end rule: functions, mysteries, uses". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 93.22, pp. 12142–12149.

Waugh, David S (Dec. 2011). "An overview of enzymatic reagents for the removal of affinity tags". en. In: *Protein Expr. Purif.* 80.2, pp. 283–293.

Wehr, Michael C et al. (Dec. 2006). "Monitoring regulated protein-protein interactions using split TEV". en. In: *Nat. Methods* 3.12, pp. 985–993.

Weinberg, Benjamin H et al. (May 2017). "Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells". en. In: *Nat. Biotechnol.* 35.5, pp. 453–462.

Weinheimer, S P et al. (Oct. 1993). "Autoproteolysis of herpes simplex virus type 1 protease releases an active catalytic domain found in intermediate capsid particles". en. In: *J. Virol.* 67.10, pp. 5813–5822.

Wikstrand, C J et al. (Apr. 1998). "The class III variant of the epidermal growth factor receptor (EGFRvIII): characterization and utilization as an immunotherapeutic target". en. In: *J. Neurovirol.* 4.2, pp. 148–158.

Wroblewska, Liliana et al. (Aug. 2015). "Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery". en. In: *Nat. Biotechnol.* 33.8, pp. 839–841.

Xie, Zhen et al. (Sept. 2011). "Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells". en. In: *Science* 333.6047, pp. 1307–1311.

Yasuda, Ryohei et al. (Feb. 2006). "Supersensitive Ras activation in dendrites and spines revealed by two-photon fluorescence lifetime imaging". en. In: *Nature Neuroscience; New York* 9.2, pp. 283–291.

Yeh, Brian J et al. (May 2007). "Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors". en. In: *Nature* 447.7144, pp. 596–600.

2.3   Materials and Methods

Plasmid construction

All constructs were generated using standard procedures. The backbones were linearized using restriction digestion or PCR, and inserts were generated using PCR or gBlock synthesis (IDT). A list of all plasmids reported in this manuscript is included in Table S1.

Tissue culture

The Flp-In™ T-REx™ 293 Cell Line (Human Embryonic Kidney cells that contain a single stably integrated FRT site at a transcriptionally active genomic locus, and stably expressing the tetracycline repressor protein) was purchased from Thermo Fisher Scientific (R78007). Cells were cultured in a humidity controlled chamber at $37°$ C with 5% CO_2 in media containing DMEM supplemented with 10% FBS, 1 mM sodium pyruvate, 1 unit/ml penicillin, 1 μg/ml streptomycin, 2 mM L-glutamine and 1X MEM non-essential amino acids. 100 ng/mL doxycycline was added whenever expression is needed from a CMV-TO promoter. All stably integrated transgenes were inducible with doxycycline, which was only added one day before characterization. Trimethoprim (TMP) was delivered at 1 μ M. Rapamycin was delivered at 5 nM. Epidermal growth factor (EGF) was delivered at 25 ng/mL. SHIELD1 was delivered at 1 μ M. ASV was delivered at 3 μ M. For bulk measurement of pulsing dynamics, cells were cultured in the presence of 40 μ M biliverdin, and rapamycin was added at different time points before preparation for flow cytometry. For stimulation with EGF, cells were cultured to near 100% confluency before transfection, and, one day after transfection, exposed to 40 μ M biliverdin, 25 ng/mL EGF, and 100 ng/mL doxycycline for 6 hours prior to flow cytometry analysis.

Transient transfection

293 cells were seeded at a density of 0.05 x 10^6 cells per well of a 24-well plate and cultured under standard conditions overnight. The following day, the cells were transfected with plasmid constructs using Lipofectamine 2000 (Thermo Fisher) as per manufacturer's protocol

Transient transfection of mRNA

293 cells were seeded at a density of 0.05 x 10^6 cells per well of a 24-well plate and cultured under standard conditions overnight. The following day, the cells were transfected with mRNA using TransIT-mRNA (Mirus Biology) as per manufacturer's protocol. The maximum mRNA used per transfection was 1 μg. Cells were incubated with mRNA overnight and media was changed the following day. mRNA was synthesized using N1-pseudomethyluridine from Trilink Biotechnologies.

Flow cytometry

Two days after transfection for plasmids, and one day after transfection for mRNA, cells were prepared for flow cytometry by trypsinizing with 30μL of 0.05% trypsin for 1 min at room temperature. Protease activity was neutralized by resuspending the cells in buffer containing 70 μL of HBSS with 2.5mg/ml Bovine Serum Albumin (BSA). For cells stimulated with EGF, cells were resuspended in buffer containing 70μ L of HBSS with 2.5 mg/mL BSA and 1 mM EDTA. Cells were then filtered through a 40 μm cell strainer and analyzed by flow cytometry (MACSQuant VYB, Miltenyi or CytoFLEX, Beckman Coulter). We used the EasyFlow Matlab-based software package developed in-house by Yaron Antebi to process flow cytometry data.

Annexin V staining

Staining was performed using a standard kit (ThermoFisher A13201). One day after transfection, cell culture medium was removed from each well, and replaced with 7.5 μL FITC-conjugated annexin V within 150 μL binding buffer. After incubation in dark at 37C for 15 min, the staining medium was removed, and the cells trypsinized for flow cytometry analysis.

Fluorescent signal quantification from flow cytometric measurements

To maximize the observable reporter dynamic range, we selected and compared cells with the highest expression of the co-transfection marker, which showed the largest separation of basal reporter fluorescence from cellular autofluorescence (Figure 2.10). For each sample in a comparison group (experiments performed in the same batch and data shown on the same plot), we calculated the 98 and 99.5 percentiles of fluorescence of the co-transfection marker (mCherry in most cases). We identified the sample with the lowest 98 percentile value, and used its 98 and 99.5 percentiles

as lower and upper limits to gate on all samples. For all cells within the gate in each sample, we fit the distribution of the logarithm of their signal fluorescence (Citrine in most cases) with skew Gaussian distributions, i.e. N$*$ normcdf(x,m,k)$*$ normpdf(x,m,s) in Matlab using non-linear least-square fitting, and reported the mode (peak position, representing the reporter level that's most likely to be observed) of the resulting fit (Fig. 2.10). Here, the normcdf(x, μ, σ) and normpdf(x, μ, σ) functions are cumulative probability density and probability density functions for a Gaussian distribution respectively, and the parameter n is a normalization factor, m=μ is the mean of the Gaussian function, s=σ is the inverse standard deviation of the Gaussian, and k parameterizes skewness. No gating was performed on monoclonal cells with the genomically integrated pulsing circuit, because, unlike transient transfection, here expression variation is already limited.

Calculating reduction index from flow cytometric measurements

To quantify the reduction of cell numbers due to Caspase-3 activation, we compared the effects of various treatments on cell numbers, comparing each measurement to a negative control transfected with only a fluorescent marker, and using the size of the untransfected cell population for internal normalization. To do this, we proceeded in several steps: First, we fit the distribution of the logarithm of autofluorescence collected in the Citrine channel from mock transfected cells with the MATLAB function N_0* normcdf(x, m_0, k_0) $*$ normpdf(x, m_0, s_0) using non-linear least-square fitting. Here, the parameters n_0, m_0, s_0, and k_0 and functions normcdf() and normpdf() have the same meanings as in the previous section. Reference values for m_0, s_0, k_0, were thus determined from measurement of autofluorescence in untransfected cells and fixed for subsequent two-component model fits. Second, for each transfected well, we fit the distribution of the logarithm of Citrine signal with N_1* normcdf(x,m_0,k_0)$*$ normpdf (x,m_0,s_0) +N_2* normpdf(x,m_2,s_2), where N_1, N_2, m_2, s_2 were free parameters and m_0, s_0, k_0 were fixed to values extracted from autofluorescence fit. The area under the curve N_1* normcdf(x,m_0,k_0)$*$ normpdf(x,m_0,s_0) ("area a_0" and "area a" in Fig. 2.17B) corresponds to the number of untransfected cells, which serves as an internal reference. Third, we subtracted the number of untransfected cells from the total number of cells to get the number of transfected cells that survived ("area b_0" and " area b" in Fig. 2.17B). For each sample, the number of transfected cells that survived was then normalized to the number of untransfected cells, and the ratio between normalized survival number in that condition ((area a)/(area b) in

Fig. 2.17B) and normalized survival number in the Citrine-only control condition ((area a_0)/(area b_0) in Fig. 2.17B) was defined as survival percentage. Finally, the reduction index was defined as 1-survival percentage.

In experiments with SOS+/EGFRvIII+ cells, a small fraction of these cells silenced their transgene expression during cell culture. To make sure that we were only analyzing cells that do express a Ras activator, we gated on mCherry that's co-expressed with SOS/EGFRvIII, and excluded the mCherry– population. This co-expressed mCherry marker was also utilized in co-culture experiments, to distinguish SOS+/EGFRvIII+ cells from control cells, so that we could calculate their reduction index separately.

Bandpass analysis

To analyze the behavior of the bandpass circuit, we built mathematical models considering protein production, first-order degradation, and cleavage by proteases. The model took the general forms of Equations 2.1 and 2.2 below, where A represents the production rate of the substrate protein, $k_{cat}^{Protease}$ represents the catalytic coefficient (approximating a Michaelis-Menten reaction far from saturation with a second-order reaction), and k_{dX} represent the first-order degradation rates by cellular degradation pathways that could take high or low values depending on whether the substrate protein and its cleaved form are unstable or stable, respectively.

$$\frac{d\left[Substrate\right]}{dt} = A - k_{cat}^{Protease}\left[Protease\right]\left[Substrate\right] - k_{dA}\left[Substrate\right] \quad (2.1)$$

$$\frac{d\left[Substrate_{cleaved}\right]}{dt} = k_{cat}^{Protease}\left[Protease\right]\left[Substrate\right] - k_{dB}\left[Substrate_{cleaved}\right]$$
$$(2.2)$$

To simplify the analysis without loss of generality, we nondimensionalized every equation involving a fluorescent reporter, such that the values of $\left[Substrate\right]$ in the nondimensionalized version corresponds to the values of $\left[Substrate\right]/A$ in the original version, and the production rate is 1 in the nondimensionalized version.

We first considered a Cit_{DHFR} reporter, whose DHFR degron can be removed by TEVP with a coefficient k_{cat}^{TEV}. The initial reporter degrades at a rate k_{d1} (Equation

2.3), and its cleaved product Cit degrades at a rate k_{d2} (Equation 2.4). We assume the system is able to reach steady-state once the production rate is equal to the degradation rates $\left(\frac{dCit}{dt} = 0 \text{ and } \frac{dCit_{DHFR}}{dt} = 0\right)$.

$$\frac{dCit_{DHFR}}{dt} = 1 - k_{cat}^{TEV}[Cit_{DHFR}][TEVP] - k_{d1}[Cit_{DHFR}] \tag{2.3}$$

$$\frac{dCit}{dt} = k_{cat}^{TEV}[Cit_{DHFR}][TEVP] - k_{d2}[Cit] \tag{2.4}$$

The steady-state solutions of Equations 2.3 and 2.4 are:

$$Cit_{DHFR} = \frac{1}{k_{cat}^{TEV}[TEVP] + k_{d1}} \tag{2.5}$$

$$Cit = \frac{k_{cat}^{TEV}[TEVP]}{k_{d2}\left(k_{cat}^{TEV}[TEVP] + k_{d1}\right)} \tag{2.6}$$

Experimentally measured reporter fluorescence corresponds to the sum $Cit_{DHFR} + Cit$. The absolute value of the independent variable $[TEVP]$ is not known. However, based on experiments in which protein expression levels correlated linearly with the amount of transfected plasmid (Fig. 2.16A), we substituted the concentration of transfected plasmid, p_{TE}, for $[TEVP]$ in all equations, effectively absorbing the constant of proportionality relating $[TEVP]$ and p_{TE} into the k_{cat}^{TE} values. With these simplifications, measured fluorescence can be written:

$$Cit_{total} = Cit_{DHFR} + Cit = \frac{\frac{k_{cat}^{TE} p_{TE}}{k_{d2}} + 1}{k_{cat}^{TE} p_{TE} + k_{d1}} \tag{2.7}$$

Using Matlab's curve fitting toolbox, we determined best fit values of the parameters k_{cat}^{TE}, k_{d1} and k_{d2} by fitting Eq. 2.7 to the experimentally measured p_{TE} - Cit_{total} curve (Fig. 2.5B).

To model the repression arm of the bandpass circuit, we must take into account the mutual inhibitory activities of TVMVP and HCVP in the circuit. These protease-

protease equations take on the general form outlined in Eqs 2.1, 2.2. However, because reporter and protease concentrations are measured in different units (fluorescence and plasmid concentration, respectively), their production rates cannot both be arbitrarily set to 1. Instead, we denoted the protease production rate B, to account for the different units used for these two species. Specifically, for 1 unit of plasmid input to produce 1 unit of protease at steady-state, B must equal the degradation rate of the protease multiplied by the amount of plasmid input ($p_{Protease}$, as shown below in Equations 2.8 and 2.9.

$$\frac{d[TVMVP]}{dt} = k_{dTV} p_{TV} - k_{cat}^{HC} [HCVP] [TVMVP] - k_{dTV} [TVMVP] \qquad (2.8)$$

$$\frac{d[HCVP]}{dt} = k_{dHC} p_{HC} - k_{cat}^{TV} [TVMVP] [HCVP] - k_{dHC} [HCVP] \qquad (2.9)$$

At steady-state, the concentration of TVMV protease can be expressed as a function of the plasmid inputs of TVMVP and HCVP:

$$[TVMVP] = \frac{W + \left(W^2 + 4k_{cat}^{TV} k_{dTV}^2 k_{dHC} p_{TV}\right)^{\frac{1}{2}}}{2k_{cat}^{TV} k_{dTV}} \qquad (2.10)$$

where $W \equiv k_{dTV} k_{cat}^{TV} p_{TV} - k_{dHC} k_{dTV} - k_{cat}^{HC} k_{dHC} p_{HC}$. The reporter repressed by TVMVP is denoted Cit when not cleaved (first-order degradation rate k_{d3}), and Cit_{Ndeg} when cleaved by TVMVP to expose an N-end degron (first-order degradation rate k_{d4}).We then used a procedure similar to Eqs. 2.3- 2.7 to express reporter expressions in terms of [TVMVP]:

$$Cit = \frac{1}{k_{cat}^{TV} [TVMVP] + k_{d3}} \qquad (2.11)$$

$$Cit_{Ndeg} = \frac{k_{cat}^{TV} [TVMVP]}{k_{d4} \left(k_{cat}^{TV} [TVMVP] + k_{d3}\right)} \qquad (2.12)$$

$$Cit_{total} = \frac{\frac{k_{cat}^{TV}[TVMVP]}{k_{d4}} + 1}{k_{cat}^{TV}[TVMVP] + k_{d3}} \qquad (2.13)$$

$*$ For all equations denoted with "$*$", $[TVMVP]$ takes the value defined in Eq. 2.10.

We estimated the values of parameters, k_{cat}^{HC}, k_{cat}^{TV}, $k_{dHC}, k_{dTV}, k_{d3}, k_{d4}$, by fitting Eq. 2.13 to experimentally measured Cit_{total} , p_{TV} , and p_{HC} (Fig. 2.5C).

To characterize the cooperativity caused by TVMVP-HCVP mutual inhibition, we fit the repression curves in Fig. 2.5C with a sigmoidal function:

$$Cit_{total} = \frac{C}{1 + \left(\frac{p_{TV}}{K}\right)^n}$$

The 95% confidence intervals for the Hill coefficient, n, were $0.95 \pm 0.13, 2.0 \pm 0.4, and 2.4 \pm 0.5$, for p_{HC} values of 0, 50, and 200 ng, respectively.

Finally, for the reporter that's simultaneously regulated by the activation and repression arms, depending on whether the DHFR degron is removed and whether the N-end degron is exposed, there are four possible species Cit_{DHFR} , $Cit_{DHFR+Ndeg}$, Cit , and Cit_{Ndeg} , the first-order degradation rates of which we denote as k_{dA} , k_{dB} , k_{dC} , and k_{dD} , respectively. Similarly, the dynamics of these four species can be expressed as:

$$\frac{dCit_{DHFR}}{dt} = 1 \smallsmile k_{cat}^{TE}[TEVP][Cit_{DHFR}] - k_{cat}^{TV}[TVMVP][Cit_{DHFR}] - k_{dA}[Cit_{DHFR}]$$
$$(2.14)$$
$$\frac{dCit_{DHFR+Ndeg}}{dt} = k_{cat}^{TV}[TVMVP][Cit_{DHFR}] - k_{cat}^{TE}[TEVP][Cit_{DHFR+Ndeg}] - k_{dB}[Cit_{DHFR+Ndeg}]$$
$$(2.15)$$
$$\frac{dCit}{dt} = k_{cat}^{TE}[TEVP][Cit_{DHFR}] - k_{cat}^{TV}[TVMVP][Cit] - k_{dC}[Cit] \qquad (2.16)$$

$$\frac{dCit_{Ndeg}}{dt} = k_{cat}^{TE}[TEVP]\left[Cit_{DHFR+Ndeg}\right] + k_{cat}^{TV}[TVMVP][Cit] - k_{dD}\left[Cit_{Ndeg}\right]$$
$$(2.17)$$

We summed the steady-state solutions of all species from these equations to derive the final input-output equation for the bandpass circuit:

$$Cit_{total} = \frac{1 + X + Y + \frac{k_{cat}^{TE}p_{TE}X + k_{cat}^{TV}[TVMVP]Y}{k_{dD}}}{k_{cat}^{TE}p_{TE} + k_{cat}^{TV}[TVMVP] + k_{dA}} \tag{2.18}$$

Where $X \equiv \frac{k_{cat}^{TV}[TVMVP]}{k_{cat}^{TE}p_{TE}+k_{dB}}$ and $Y \equiv \frac{k_{cat}^{TE}p_{TE}}{k_{cat}^{TV}[TVMVP]+k_{dC}}$

We used this equation to fit the experimentally observed bandpass behavior (Fig 2.5).

Cell line construction

Some of the experiments do require more stable/homogenous transgene expression, for which we used antibiotic selection to generate cell lines with stably integrated transgenes. Two days after transfection in 24-well plates, cells were transferred to 6-well plate and selected with either 50 μ g/mL Hygromycin (Hyg) or 400 μ g/mL Geneticin (Gen). Sos+ cells: CMV-TO-MSos-2A-H2BChe-FlpIn co-transfected with pOG44, Hyg; pulse cells: PB-CMV-TO-rapTEV-teHCV-hcTVMV-tvDiTEV-Neo co-transfected with a plasmid expressing PiggyBac transposase, Gen; EGFRvIII+ cells: PB-CMV-TO-EGFRvIII-IRES-nlsChe co-transfected with a plasmid expressing PiggyBac transposase, Gen. After PiggyBac-based integration, monoclonal cell populations were established through limiting dilution, and preliminary screening was performed to identify clones with highest transgene expression (based on GFP that serves as the scaffold in iTEV, and mCherry that's co-expressed with EGFRvIII), which were used in subsequent experiments. Among the pulse cell clones with highest GFP expression, the one with the least variance was selected. We then subjected this clone to another round of transgenesis (Hyg, CMV-TO-Cer-HO1-FlpIn co-transfected with pOG44) to provide Cerulean as a segmentation marker and heme oxygenase-1 to increase the intracellular concentration of biliverdin that's necessary for enhancing iTEV signal. The final cell line was used in time-lapse imaging.

Time-lapse imaging

For time-lapse imaging of pulse dynamics (Figure 3), monoclonal pulse-generation cells were mixed with parental wild-type HEK293 cells at a 1:10 ratio. Cells were plated on 24-well glass-bottom plates which had been coated with 5 μ g/mL with hamster fibronectin for 1 hour at room temperature. Cells were induced with 100 ng/mL overnight in normal culturing conditions. The following morning, the media was replaced with imaging media containing FluoroBrite DMEM (Thermo Fisher) supplemented with 10% FBS, 1 mM sodium pyruvate, 1 unit/ml penicillin, 1 μg/ml streptomycin, 2 mM L-glutamine and 1X MEM non-essential amino acids and 100 ng/mL doxycycline.

All time-lapse images were acquired on an inverted Olympus IX81 fluorescence microscope with Zero Drift Control (ZDC), an ASI 2000XY automated stage, iKon-M CCD camera (Andor, Belfast, NIR), and a 60x oil objective (1.42 NA). Fluorophores were excited with an X-Cite XLED1 light source (Lumen Dynamics). Cells were kept in a custom-made environmental chamber enclosing the microscope, with humidified 5% CO2 flow at 37° C. Microscope and image acquisition were controlled by Metamorph software (Molecular Devices).

Imaging started approximately 2 hours after changing the media to fluorescent imaging media. 5 nM rapamycin was added after approximately 2 hours of imaging to induce the pulse. Images were acquired every 20 or 25 min, typically for 20-40 hrs. Cells that were in the field of view before rapamycin induction and remained alive and visible in the field of view without death for at least 20 hours were used for initial data analysis.

For analysis, we only included cells that remained alive throughout the duration of the experiment, remained within the field of view, and had detectable signal/background ratio. IFP fluorescence intensity is dependent on the biliverdin chromophore. Addition of exogenous biliverdin increases IFP fluorescence but also produces IFP-independent background fluorescence. For the movies, to minimize background, we omitted biliverdin from the media, relying instead on lower concentrations produced endogenously. Under these conditions, IFP excitation illumination levels caused some phototoxicity, resulting in a subpopulation of ~ 50% of cells that died within ~ 7 hours. The remaining cells continued active division until the end of the movie, or until exit from the field of view. These cells exhibited a range of IFP fluorescence levels overlapping background. 30-60% of these cells in which IFP fluorescence exceeded background. About half of this set had morphologies that

were amenable to image-based segmentation and therefore were analyzed further. Within this group, we verified that the circuit dynamics were independent of expression level, as measured by peak IFP fluorescence, suggesting that circuit dynamics are not influenced by expression level within this range.

Movie analysis

Matlab-based single-cell tracking and image normalization software was developed in-house by Yaron Antebi.

Single-cell tracking and image normalization: Single-cell tracking and image normalization procedures were performed as previously described (*70*) with a few modifications. Briefly, cells constitutively express cytoplasmic Cerulean as a segmentation marker. Due to the diffuse and weak Cerulean signal, manual segmentation was frequently required and cell boundaries were identified in part by phase contrast and GFP fluorescence images (GFP is the protein identified as the "split scaffold" in Fig. 2.6. It serves a structural role in the context of the IFP reporter, but also fluoresces). Two normalization procedures were applied to each fluorescent channel across all images: (1) We performed background subtraction to account for to media auto-fluorescence and camera background (*F_back(x,y)*) with a scaling factor *A(x, y, t)*; (2) We performed a field flattening procedure to account for non-uniformity of illumination within the field of view (*F_illum(x,y)*) with a scaling factor *a(x, y)*. Here *x,y* denote the spatial coordinates within the image and *t* represents the frame number. These values were calculated by fitting a paraboloid to the images.

$$F_{corrected}(x, y, t) = \frac{F_{raw}(x, y, t) - B}{I(x, y)} - A(t)$$

For generating the supplementary movie, mean intensities < 5% were set to zero and mean intensities > 99.5% were set to maximum pixel values to reduce speckling noise.

Quantification of amplitude and pulse decay:

Data processing: The amplitude and pulse decay calculations were based on total levels of fluorescence in the IFP fluorescent channel. To systematically quantify the fluorescent signal in the IFP channel, total IFP signal intensity *(IFP(x, y, t))* was normalized by the total constitutive Cerulean signal *(CFP(x, y, t))* and rescaled with a baseline variable (90th percentile of *IFP(x, y, t)/CFP(x, y, t))*) at all *x* positions. To capture the pulse of IFP signal and avoid distortion of the peak shape, the resulting data was smoothed with a Savitzky-Golay filter using a 3rd order polynomial and a window length of 9. After smoothing, the data was interpolated to equidistant timepoints of 20 minute intervals (Fig. 2.6D).

Fitting: We fit the pulsing dynamics by taking the smoothed and interpolated data and subtracting the minimum value of the normalized signal intensity from each timepoint. Using MATLAB's tfest function, the normalized data was deconvolved with a finite impulse signal and a third-order linear transfer function resulting in the equation:

$$y = a_1 * exp\hat{} \ (p_1 * x(t)) + a_2 * exp\hat{} \ (p_2 * x(t)) + a_3 * exp\hat{} \ (p_3 * x(t))$$

The resulting fit was used to determine: (1) the location at which the maximal value of IFP occurred and (2) the delay time, τ , after peak signal at which the signal intensity decayed to 50% its maximum value. After determining the peak location and τ , the mean and standard deviation were calculated.

Data and code availability

The datasets generated and analyzed and the computer code used during the current study are available from the corresponding author on reasonable request.

2.4 Supplementary Materials and Methods
Supplementary Text

Characterization and optimization of HCVP and its reporter.

For HCV protease (HCVP), we adopted a previously described construct in which the protease and its co-peptide are fused to create a more active single chain protease (Taremi et al., 1998). This HCVP initially showed more modest regulation than the other proteases, especially for the repressible reporter (Fig. 2.11C). We reasoned that increasing the protease affinity to its target could improve its regulatory range. Indeed, incorporating a pair of hetero-dimerizing leucine zippers (Ghosh, Hamilton, and Regan, 2000) in the protease and its target indeed improved regulation (Fig. 2.11C, right).

Characterization and optimization of circuits that selectively reduce Ras-activating cells

To exclude the possibility that SOS+ cells are generally more sensitive to caspase activation, we first analyzed constitutively dimerized split TEVP variants, one using leucine zippers, and the other adopting a RasG12V mutant that binds constitutively to RBD (Herrmann, Martin, and Wittinghofer, 1995) (Fig. 2.18C). When co-transfected with the TEVP-activatable caspase, these control constructs displayed no selectivity for SOS+ cells (Fig. 2.18C), indicating that the regulated Ras-RBD interaction is necessary for the selectivity observed in the main text (Fig. 2.8B).

To assess the contribution of each additional regulatory interaction in the full circuit, we systematically removed them one at a time, and compared their effects on control and SOS+ cells to the full circuit. Removal of caspase inhibition by TVMVP re-introduced substantial reduction in control cells (Fig. 2.19, left), and removal of TVMVP inhibition by RasTEVP increased survival in SOS+ cells (Fig. 2.19, middle). By contrast, removal of RasTEVP inhibition by TVMVP had no effect on survival in either control or SOS+ cells (Fig. 2.19, right). These results indicate that Arms 3,4 (Figs. 2.17, 2.19) are major contributors to full circuit performance.

For single-transcript delivery of the full circuit, we interposed a wild type internal ribosome entry site (IRES) between Caspase-3 and RasTEVP coding sequences, followed by one of several IRES variant sequences (Wikstrand et al., 1998) and then the TVMVP (Fig. 2.20A). Inspired by TVMVP titration results (Fig. 2.18E), we chose variants with ~30% and ~ 70% of wild-type strength for the second IRES

(Gray, Mahrus, and Wells, 2010), and found that the circuit functioned optimally with the $\sim 70\%$ IRES (Fig. 2.20A).

Response of RasTEVP to EGF stimulation

To assess the response of RasTEVP to a physiological ligand that normally activates the Ras pathway, we stimulated cells expressing either RasTEVP or constitutively dimerized and membrane-localized TEVP (negative control TEVP) with epidermal growth factor (EGF). When co-transfected with a membrane-localized iTEV reporter, the control construct TEVP-CAAX exhibited minimal response to EGF stimulation, whereas RasTEVP displayed a modest response to EGF (Fig. 2.18A).

Comparison of protease-protease and transcriptional regulatory dynamics

In this supplementary section, we use a minimal model to address the question of how a simple transcription factor regulatory step differs in dynamics from a simple protease regulation step. To make a controlled comparison between the two kinds of regulation, we assume that shared biochemical parameters, such as protein degradation rates, are similar in the two systems. The main conclusion is that protease regulation can occur more rapidly than transcriptional regulation but with timescales that depend on the direction of regulation. By contrast, transcriptional regulation is expected to be slower but show similar timescales in both directions of regulation. While we have considered typical biochemical parameter values here, we note that additional features of any specific system, including feedback structure, could impact their dynamic behavior. Additionally, the quantitative values of the resulting timescales in general depend on the specific choice of biochemical parameter values.

Protease-protease regulation. We modeled repression of one protease by another through direct cleavage, based on the scheme in Fig. 2.3B. We assume the concentration of the input protease, denoted P_0, is maintained at a constant level, with its activity controlled by a small molecule input, as in the scheme of Fig. 2.6A. The output protease, denoted P, is produced at a constant rate A, and undergoes first-order degradation with rate γ_p. The input protease cleaves the output protease at a single cleavage site, converting it to a cleaved form, whose concentration is denoted P_c, with a cleavage rate constant k. The cleaved protease irreversibly

dissociates at rate k_d, and undergoes first-order degradation with rate γ_p for a total rate of elimination of $\gamma_p k_d$. We assume a single cleavage for simplicity, but the same conclusions hold true for two independent cleavage sites, cleavage of either of which is sufficient to inactivate the output protease.

The reactions in the protease-protease model are as follows, where ϕ denotes 'nothing':

1. Synthesis of the output protease: $\phi \xrightarrow{A} P$

2. Degradation of the output protease: $P \xrightarrow{\gamma_p} \phi$

3. Catalytic cleavage of the output protease: $P_0 + P \xrightarrow{k} P_0 + P_c$

4. Dissociation of the cleaved protease: $P_c \xrightarrow{k_d} \phi$

5. Degradation of the cleaved protease: $P_c \xrightarrow{\gamma_p} \phi$

Assuming protease cleavage functions in a linear regime far from saturation, consistent with published K_m values (Tözsér et al., 2005) and our bandpass modeling, the reaction can be expressed as a set of ordinary differential equations (ODEs):

$$\frac{dP}{dt} = A - kP_0P - \gamma_pP$$

$$\frac{dP_c}{dt} = kP_0P - P_c\left(k_d + \gamma_p\right)$$

Because the absolute value of the production rate A does not affect the dynamics of the system, we arbitrarily set its value to $1\ \mathrm{M}h^{-1}$. For the dissociation rate, we assumed $k_d = 5h^{-1}$ based on indirect measurements (Kim et al., 2017). For the protein degradation rate, we assumed a biologically realistic value of $\gamma_p = 0.1h^{-1}$.

Based on our bandpass fits (Figs. 3B-D, Methods), cleavage by a protease, when the input protease activity is high, occurs at a rate comparable to the rate of degron-mediated degradation ($\sim 5h^{-1}$). We also assumed that the OFF input protease is 20-fold less active than the ON state based on the dynamic range observed in Figure

1. (Note that the value of this regulatory range does not affect our conclusions about the timescales of regulation.) Finally, we assumed the small-molecule-induced ON-OFF switch reaches steady-state much faster than the other reactions, so that the cleavage term can be approximated by a step function, taking one of two possible values:

$$kP_0 = 0.25h^{-1} \text{ (input OFF) or } 5 \ h^{-1} \text{ (input ON)}$$

To simulate output dynamics in response to changes in the input, we first set the input protease to ON, and the output protease to its steady state value of $P + P_c$. At t=10h, we switched the input to OFF and simulated the equations for 70h (10-80h). Finally, we switched the input back to ON and simulated another 70h (80h-150h). In Fig. 2.21, we plot the resulting dynamics of the output protease, normalized to its maximum value. Note the asymmetric response time, which is faster for input OFFON switch than ONOFF ($t_{\frac{1}{2}}$ = 0.32 h^{-1} vs. 2.3 h^{-1}).

Transcriptional regulation. As a comparison to protease regulation, we modeled a logically equivalent transcriptional repression step. The input transcription factor was maintained at a constant concentration of T_0 , with its activity assumed to be controlled by a small molecule, as with the protease. The input transcription factor regulates the output mRNA, T_m , whose production follows a standard rate law: $\frac{K}{K+T_0}A_m$. T_m undergoes first-order degradation with rate γ_m . The output protein T_p is translated from the mRNA at rate A_p , and degraded with rate γ_p . The reactions are as follows:

1. mRNA synthesis: $\phi \xrightarrow{\frac{K}{K+T_0}A_m} T_m$

2. mRNA degradation: $T_m \xrightarrow{\gamma_m} \phi$

3. protein synthesis: $T_m \xrightarrow{A_p} T_m + T_p$

4. protein degradation: $T_p \xrightarrow{\gamma_p} \phi$

These reactions can be converted to ODEs for each of the components:

$$\frac{dT_m}{dt} = \frac{K}{K + T_0} A_m - \gamma_m T_m$$

$$\frac{dT_p}{dt} = A_p T_m - \gamma_p T_p$$

Without loss of generality we set the production rate $A_m = 1Mh^{-1}$ and $A_p = 1h^{-1}$. We used the same protein degradation rate as in the protease regulation case above: $\gamma_p = 0.1h^{-1}$. For mRNA degradation, we simulate two values at opposite extremes of the biological range for mammalian mRNA *(72)*: $\gamma_m = 0.1h^{-1}$ (more stable), and $5h^{-1}$ (less stable). As above, we also assumed that the small-molecule-controlled input ON-OFF switch is much faster than the other reactions. To match the protease conditions, we assumed T_0 also undergoes a 20-fold regulation, from $T_0 = 0.5K$ (input OFF) to $10K$ (input ON), although we note that the exact dynamic range of T_0 or the exact choice of the Hill function does not affect output dynamics.

We simulated this simple model of transcriptional regulation with fast and slow mRNA degradation rates, following the same ON OFF ON input temporal profile used in the protease regulation case. To focus on the timescale of regulation, we normalized each curve to its maximal value. For transcriptional regulation, $t_{\frac{1}{2}} = 7.2\ h^{-1}$ and $17\ h^{-1}$ for fast and slow mRNA decay, respectively, regardless whether the input undergoes ONOFF or OFFON switch. When input switches from ON to OFF, protease and transcriptional regulation occurs on comparable timescales, although their difference is more apparent in the slower mRNA degradation case. When input switches from OFF to ON, however, protease regulation generates a much faster response time compared to transcriptional regulation and the ON to OFF switch in the protease regulation case (Fig. S7). Intuitively, the dynamics of each process is limited by the slowest rate at which a species decays, which is the relatively slow protein degradation rate for transcriptional control (or both protein and mRNA degradation rates when mRNA is more stable); in contrast, the output protease decays at a much faster rate because, in addition to regular protein degradation, it is also cleaved by input protease, and the rate is even higher when the input is switched to its active state.

References

Ghosh, Indraneel, Andrew D Hamilton, and Lynne Regan (June 2000). "Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein". In: *J. Am. Chem. Soc.* 122.23, pp. 5658–5659.

Gray, Daniel C, Sami Mahrus, and James A Wells (Aug. 2010). "Activation of specific apoptotic caspases with an engineered small-molecule-activated protease". en. In: *Cell* 142.4, pp. 637–646.

Herrmann, C, G A Martin, and A Wittinghofer (Feb. 1995). "Quantitative analysis of the complex between p21ras and the Ras-binding domain of the human Raf-1 protein kinase". en. In: *J. Biol. Chem.* 270.7, pp. 2901–2905.

Kim, Min Woo et al. (Nov. 2017). "Time-gated detection of protein-protein interactions with transcriptional readout". en. In: *Elife* 6.

Taremi, S S et al. (1998). "Construction, expression, and characterization of a novel fully activated recombinant single-chain hepatitis C virus protease". In: *Proteins*.

Tözsér, József et al. (Jan. 2005). "Comparison of the substrate specificity of two potyvirus proteases". en. In: *FEBS J.* 272.2, pp. 514–523.

Wikstrand, C J et al. (Apr. 1998). "The class III variant of the epidermal growth factor receptor (EGFRvIII): characterization and utilization as an immunotherapeutic target". en. In: *J. Neurovirol.* 4.2, pp. 148–158.
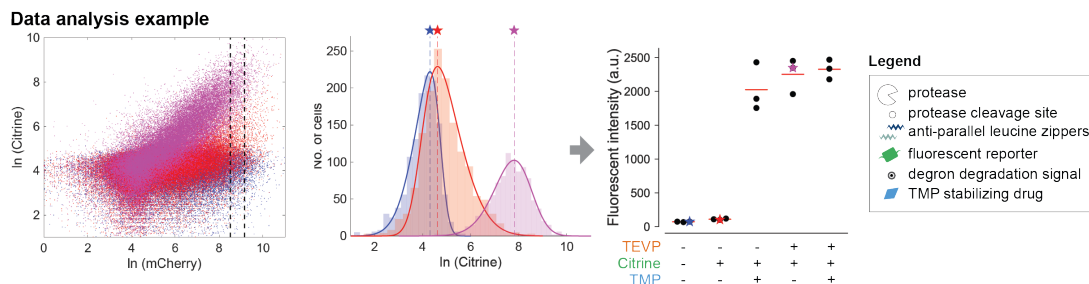
Figure 2.10 | **Data analysis example of CHOMP circuits.** Three representative log-log flow cytometry scatter plots showing autofluorescence (blue) as well as reporter co-transfected with (purple) and without (orange) TEVP. Citrine signal is represented on the y-axis and the co-transfection marker mCherry on the x-axis. Dashed lines indicate the gate on mCherry expression analyzed in Fig. 2.3A. The histograms and data points are the same as in Fig. 2.3A, except for the additionally displayed autofluorescence distribution
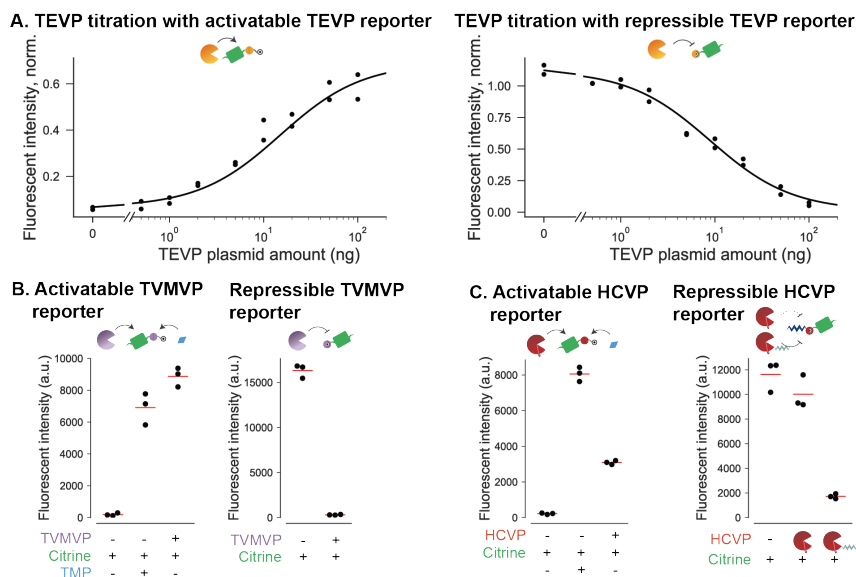
Figure 2.11 | **Characterization and optimization of CHOMP reporters.** A, Dose-response curves for activatable (left) and repressible (right) TEVP reporters (indicated schematically above each plot). The solid lines are fits based on the same equations as those used in bandpass analysis. B,C, Reporters activatable (left) and repressible (right) by TVMVP (B) and HCVP (D). The designs are identical to those of the TEVP reporters with two exceptions: First, the specific cleavage site sequences have been replaced with those of the regulatory protease. Second, the repressible HCVP reporter contains an additional leucine zipper compared to the other constructs, and it exhibits stronger repression when HCVP is tagged with the complementary leucine zipper (both shown in schematic, right-hand side of (D).
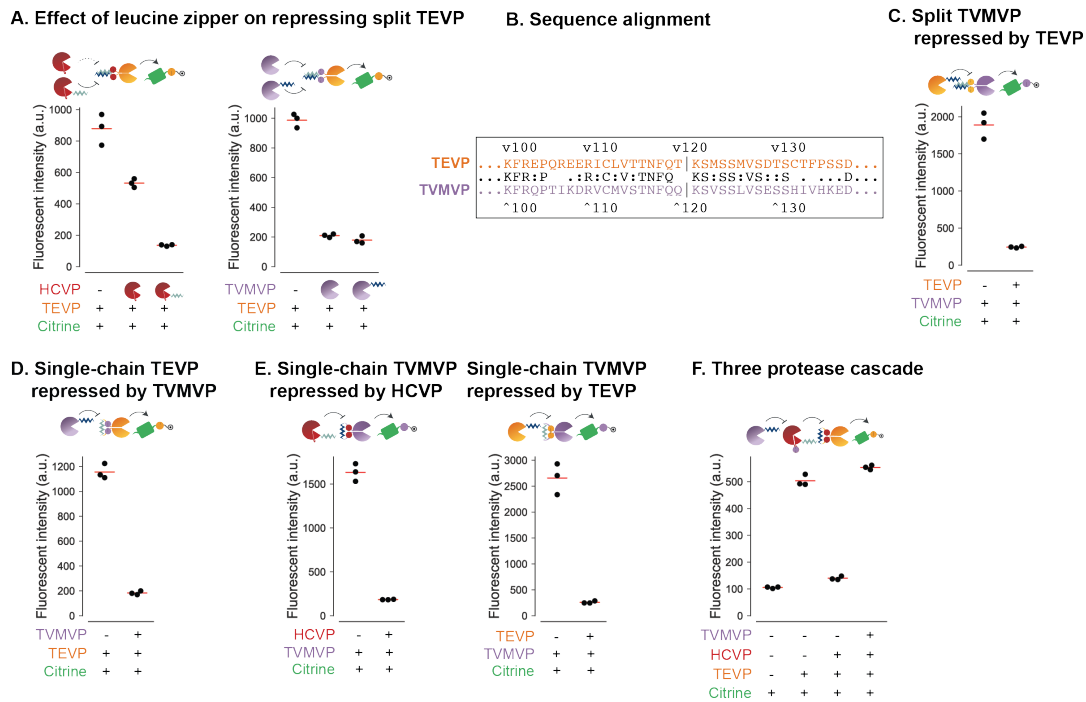
Figure 2.12 | **Characterization and optimization of CHOMP proteases.** E, Incorporating a leucine zipper (zig-zag) on HCVP (left) enhances repression of TEVP but has minimal effects when used on TVMVP (right). F, Alignment of TEVP and TVMVP sequences enables identification of TVMVP split site (vertical bars). G, A similar design enables repression of split TVMVP by TEVP. H, TVMVP can repress a single-chain TEVP. I, The single-chain TVMVP is repressed by HCVP (left) and TEVP (right). J, An alternative three protease cascade, distinct from that in Fig. 1G, can also propagate signals.

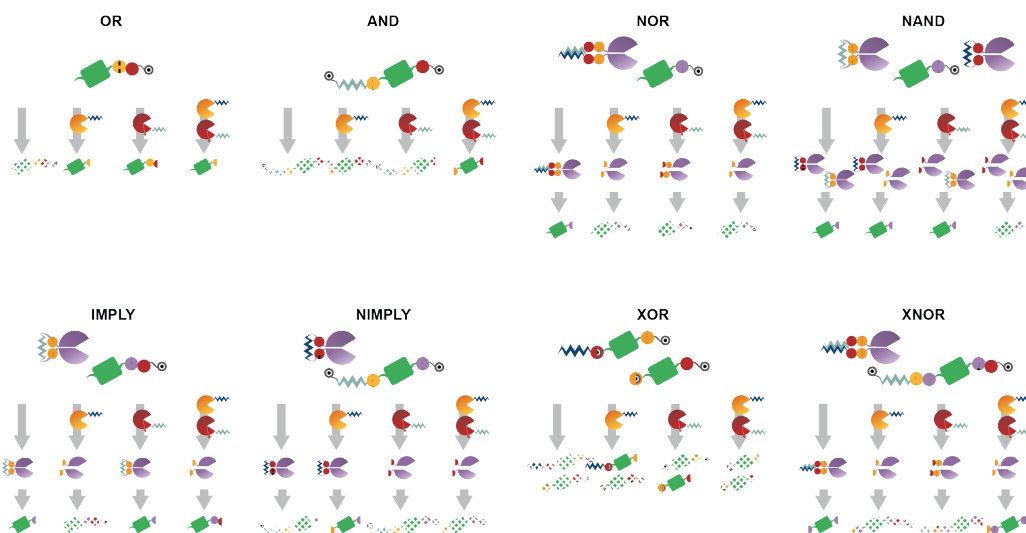**Expanded schematics for logic gates and each input state**

Figure 2.13 | **Expanded schematics for logic gates.** . For each gate, the corresponding diagram is shown on top, followed by the expected behavior in each of the four input states, with or without TEVP (yellow) and HCVP (red). The presence of Citrine (green) indicates the "ON" output state, while degraded Citrine (shown as chopped up reporter) represents the "OFF" state.
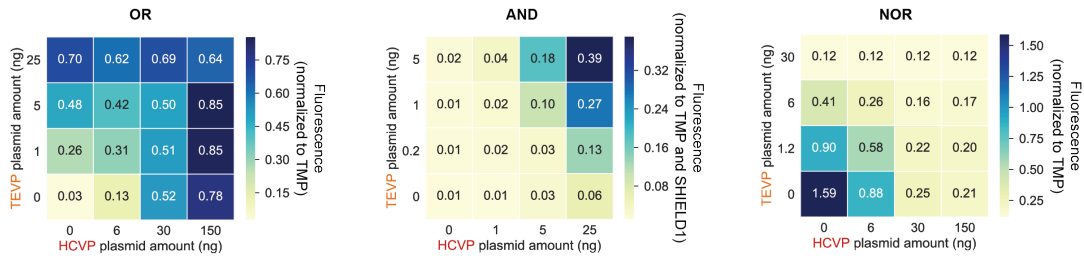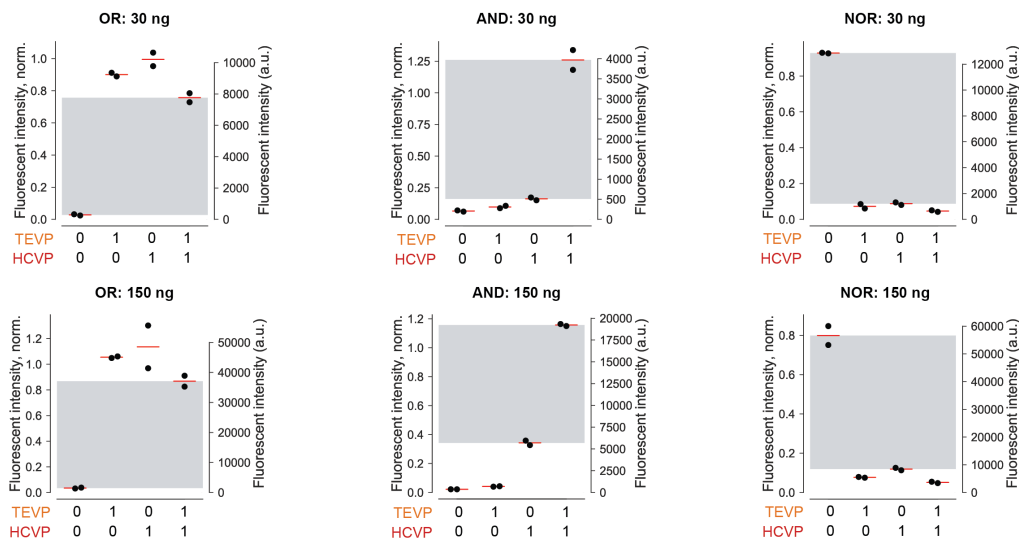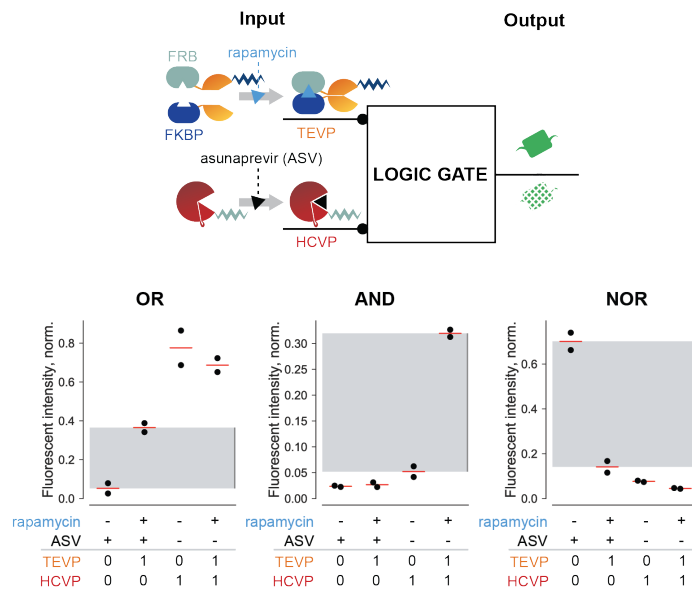
Figure 2.14 | **Expanded input and output characterization for OR, AND, and NOR gates.** B, Responses of logic gates across 16 input concentration combinations for OR, AND, and NOR gates. Fluorescent intensities are normalized to the corresponding reporter stabilized with TMP (OR and NOR) or TMP and SHIELD1 (AND). In each case, reporter was used at a concentration of 150 ng. C, Varying reporter expression levels by transfecting OR, AND, and NOR reporter plasmids at 30 ng and 150 ng. Left axis displays fluorescent intensity values normalized to reporter stabilized with TMP or TMP and SHIELD1. Inputs TEVP and HCVP at 150 ng each. Right axis shows raw fluorescent intensity values.
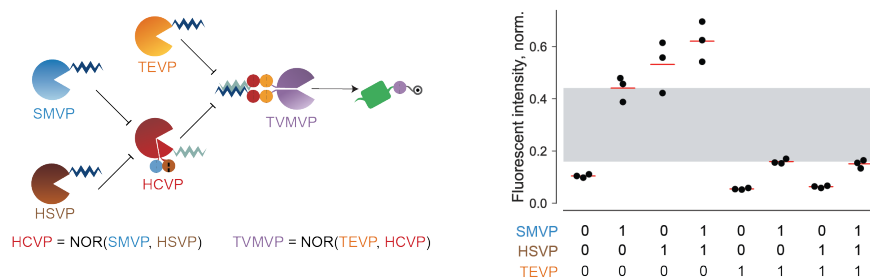
Figure 2.15 | **Expanding the inputs and complexity of logic gates.** A, Characterization of OR, AND, and NOR gates using small molecule inputs. Asunaprevir (ASV), an inhibitor of HCVP and rapamycin, a chemical inducer of dimerization of a FRB/FKBP and thereby an inducer of split TEVP, were used as inputs. Each plot shows the output behavior in the presence or absence of each of the two small molecule inputs. The expected presence or absence of input protease activities is shown below the inducer rows. B, NOR gates can be composed. Left, diagram of nested NOR gate. Soybean mosaic virus protease (SMVP) and herpes simplex virus Protease (HSVP) are inputs to HCVP activity. HCVP and TEVP are, in turn, inputs to TVMVP. Finally, TVMVP stabilizes the Citrine reporter. Right, performance of the nested NOR gate with protease inputs SMBVP, HSVP, and TEVP indicated in graph. SMVP at 80 ng, HSVP at 150 ng, TEVP at 30 ng, HCVP at 100 ng, and TVMVP at 100 ng.
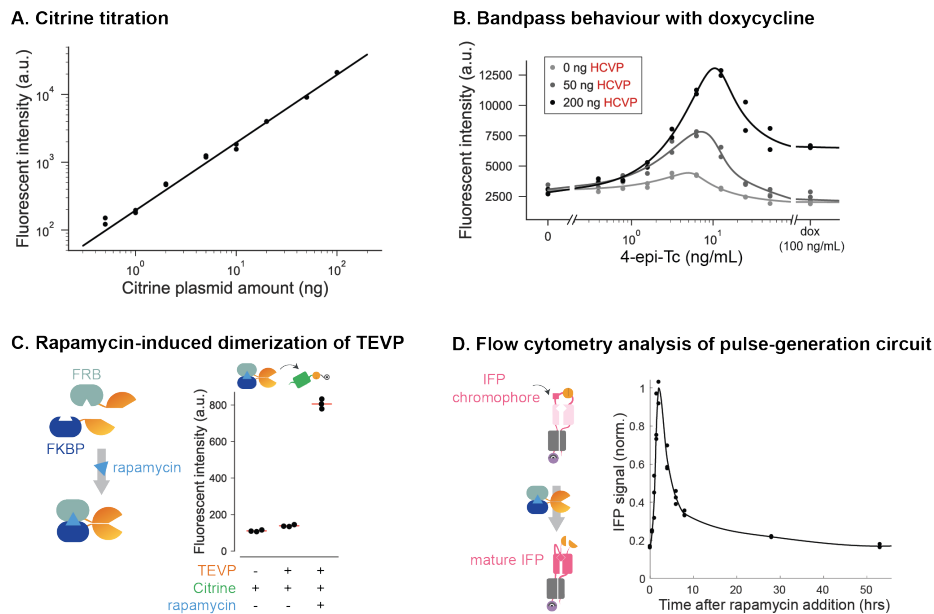
Figure 2.16 | **Characterization of bandpass and pulse-generation circuits.** A, Linear correlation between the amount of transfected DNA and Citrine expression from CMV promoter. B, Bandpass behavior in response to TEVP and TVMVP expressed at constant DNA concentration but with different levels of induction by tetracycline analog 4-epi-Tc, x-axis). C, A TEVP variant activated by rapamycin-mediated dimerization of FKBP and FRB domains exhibits rapamycin-dependent activation. D, Left, diagram for activation of the IFP reporter by TEVP cleavage. Right, flow cytometry analysis of the dynamics of the pulse generation circuit (also see Figs. 3E,F for diagrams). Each dot represents the mode of the reporter fluorescence distribution at each time point. These data were obtained with the same stable cell line as in Fig. 2.6D.
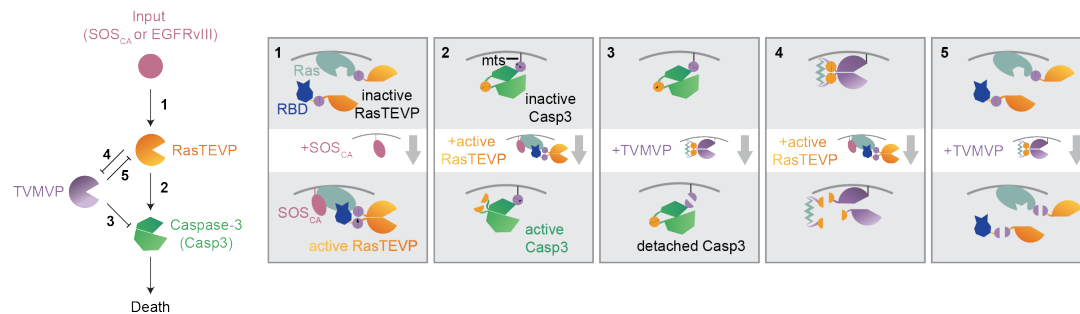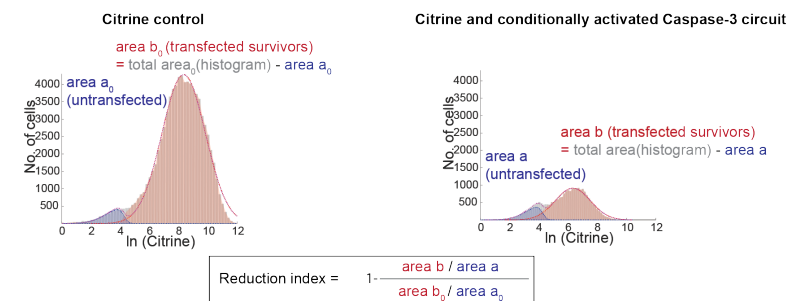
Figure 2.17 | **Expanded schematic diagram of the full circuit and example of reduction index analysis.** A, Expanded schematic diagram of the full circuit and each of its regulatory interactions (numbered arrows and corresponding boxes). B, Example of reduction index analysis. The reduction index is calculated by comparing the number of surviving transfected cells in experimental vs. Citrine-only conditions, normalized to their respective untransfected populations, as shown in the equation. See also Methods for more details. Blue and red dashed lines indicate individual Gaussian distributions in the two-component fit, and purple dashed line indicates their sum. Blue shaded area is the area beneath the blue dashed line, and red shaded area is total area (grey histogram) minus blue shaded area (not the area beneath the red dashed line).

Figure 2.18 | **Characterization and optimization of circuits that selectively activate Casp3 in response to Ras activation.** C, Response of RasTEVP to physiological ligand epidermal growth factor, EGF. Left, diagram for activation of the membrane-localized IFP reporter (same as iTEV used in the pulse circuit (Fig. 2.6AE) but with an additional 12 amino-acid N-terminal signal peptide from Lyn for membrane localization) by RasTEVP cleavage upo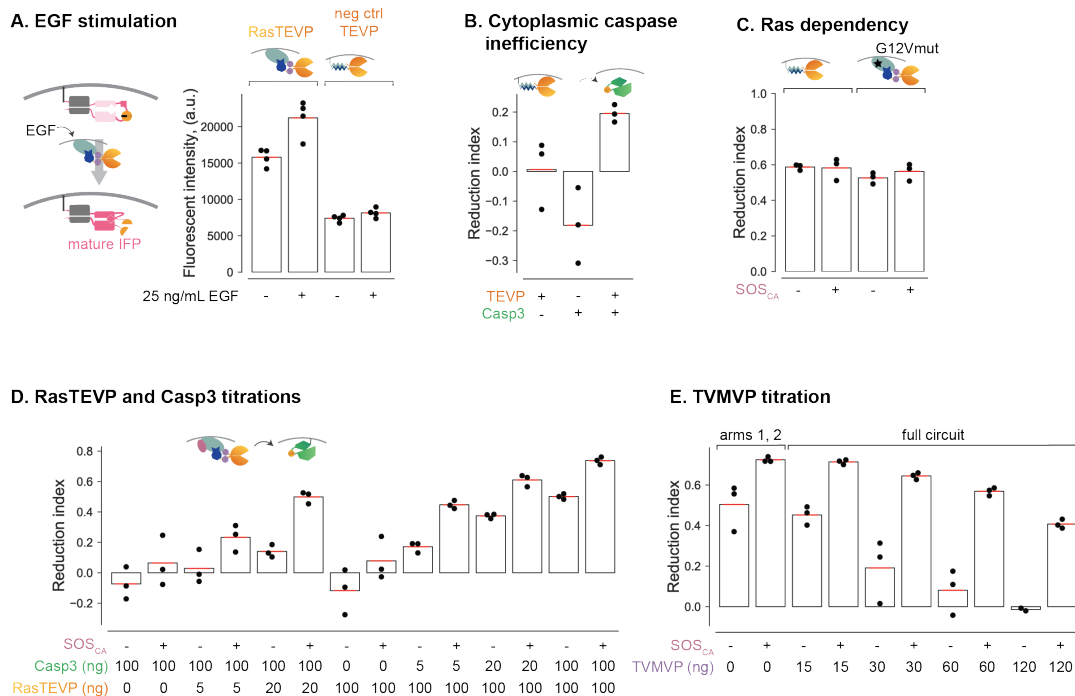n EGF stimulation. Right, cotransfection of iTEV reporter and RasTEVP or constitutively dimerized membrane-localized TEVP ('neg ctrl TEVP'). Left two bars show RasTEVP activation in response to EGF. Right two bars show negative control TEVP's relatively lower response to EGF stimulation. These transfections included 25 ng of RasTEVP and 5 ng each for the negative control TEVP components. EGF was used at 25 ng/mL. D, Cytoplasmic TEVP-activatable Casp3 causes limited reduction of cell number in the presence of membrane-localized TEVP reconstituted through leucine zippers (compare to Fig. 2.8A). E, Reduction index is unaffected by SOSCA status in the presence of constitutive Casp3 activation with no Ras-dependent regulation (Casp3 not depicted). For the left bars, TEVP is constitutively active through the membrane-tethered leucine zippers. The right bars uses a G12V mutation in Ras that renders it constitutively active (Russell, Peng, and Bell, 2012). F, The effects of RasTEVP and Casp3 doses on reduction index. Each bar represents the reduction indices from indicated concentrations of RasTEVP and Casp3 plasmids in control or SOSCA cells. G, Dose of TVMVP tunes the circuit's selectivity for SOSCA cells (the first and fourth pairs of bars also shown in Fig. 2.8B). 90 ng of RasTEVP and Casp3 were transfected in each case.
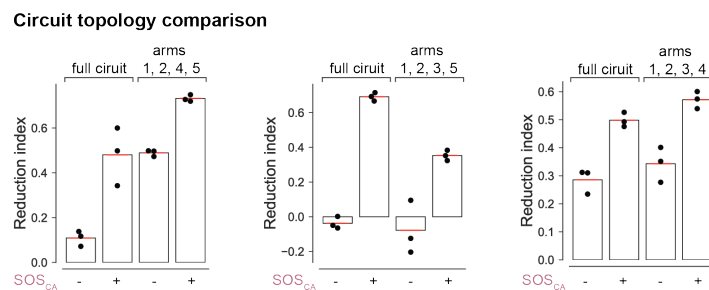
Figure 2.19 | **Further characterization of circuit topology that selectively activate Casp3 in response to Ras activation.** Analysis of contributions of individual regulatory edges in panel 2.17A to overall selectivity. Left, removing TVMVP⊣Casp3 (Arm 3) increases reduction index for both control and SOSCA cells; middle, removing RasTEVP⊣TVMVP (Arm 4) decreases reduction in SOSCA cells; right, removing TVMVP⊣RasTEVP (Arm 5) has no significant effect. We also note that, despite the qualitatively consistent selectivity, there is quantitative day-to-day variability.
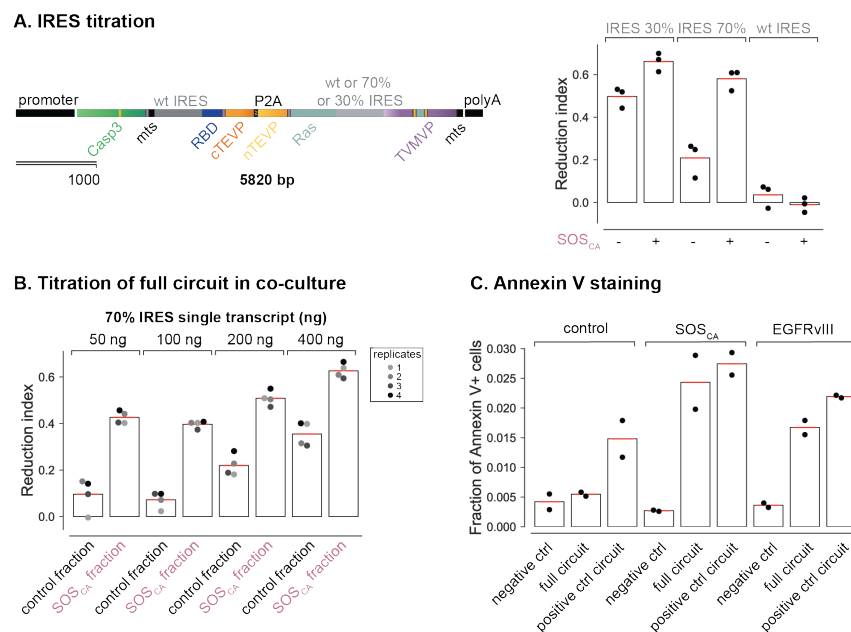
Figure 2.20 | **Further characterization and optimization of circuits that selectively activate Casp3 in response to Ras activation.** B, IRES variants with reported strengths of 30strength can be used to optimize TVMVP expression level in a single transcript. The IRES variant reported to express at 70balances survival of control cells and reduction of SOSCA cells. 200 ng for each singletranscript variant. C, Optimizing transfection dose for full single-transcript circuit with 70% IRES. Each pair of bars represents 4 replicate cocultures (gray dots) of control and SOSCA cells transfected with the indicated amount of the single-transcript circuit. D, Annexin V staining of control, SOSCA and EGFRvIII+ cells. Transfection of a negative control, full circuit and the positive control circuit from Fig 4D into each cell line at 50 ng each. We note that the fraction of apoptotic cells in all conditions are smaller than what would be indicated by reduction index, as expected due to heterogeneity in the timing of initiation of apoptosis and the loss of Annexin-V+ cells due to cell death. The two effects together cause any given time window to capture only a fraction of the cumulative number of Annexin-V+ cells over the whole time-course.

Figure 2.21 | **Simulated protease-protease and TF-TF regulation dynamics.** This plot compares the dynamic response of protease-protease regulation (blue) and transcriptional regulation (pink and red, representing distinct parameter sets) to step changes in an input protease/TF (black line). See Supplementary text for models and parameter values

Chapter 3

EXPANDING THE CAPABILITIES OF CHOMP

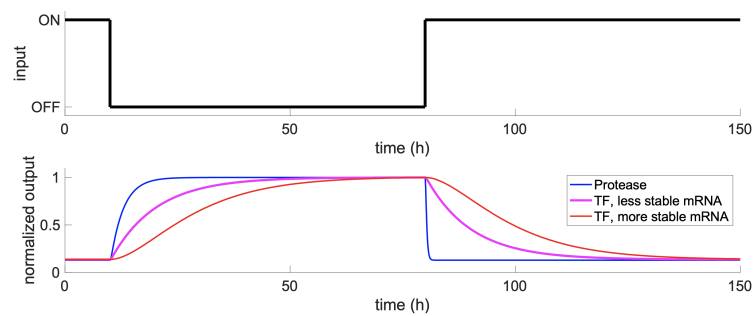**This work was done in collaboration with Ronghui Zhu, Xiaojing J. Gao, Xiaofei Ge, Matthew S. Kim, Michaela H. Ince, Maeve Morgan Phoa, and Tatiana Brailovskaya.**

3.1 Abstract

Synthetic protein circuits, in which proteins can directly regulate the activities of each other, can provide a powerful platform for engineering cellular behaviors. Previously we engineered proteases to be composable building blocks to build a variety of circuit architectures (Circuits Harnessing Orthogonal Modular Proteases). The original CHOMP framework relied on proteases inhibiting protease function. Here we aim to expand the current CHOMP toolkit by building activatable proteases. We use a general design strategy based on uncaging inactive domains through protease cleavage that permits protease-activatable protease regulation. To demonstrate the utility of these additional components, we propose activatable proteases will simplify and expand current circuit designs, provide faster dynamics and enable nonlinear signal processing. We envision activatable proteases will facilitate building complex circuits and have implications for therapeutic applications.

3.2 Main Text

For all of the diverse circuits implemented in the original CHOMP framework, proteases inhibited protease function. The current capabilities of CHOMP are limited because positive regulation requires inhibition of a repressing protease, resulting in slow dynamics that are dependent on transcription. Protease-activating-proteases could simplify circuit designs, have faster dynamics, and amplify signal. Additionally, positive self-regulation may result in ultrasensitive response motifs (Ferrell and Ha, 2014c; Ferrell and Ha, 2014a; Ferrell and Ha, 2014b; Q. Zhang, Bhattacharya, and Andersen, 2013). Controlling the intrinsic non-linearity of protease catalytic activity would facilitate useful dynamic signal processing and behaviors (Stricker et al., 2008; Elowitz and Leibler, 2000). We demonstrate that these activatable proteases can detect the presence of flaviviruses in mammalian cells and demonstrate the wide applicability of protein-based circuits.

Design and optimization of composable protease-activatable proteases

Proteases must be "uncaged" to transition from an inactive to active state. We first expressed the protease with a dihydrofolate reductase (DHFR) degron (Iwamoto et al., 2010), but failed to negatively regulate its expression. Therefore, we sought a 'caging' strategy that controls protease activity by fusing one half-protease to an inactivated (containing mutated catalytic residues) complementary 'cage' half as a single polypeptide. Upstream regulatory protease cleavage can remove the 'caging domain' and permits for the catalytically active half to associate with the 'uncaged' protease, thereby forming a fully functional protease. Here we employed two designs to build protease-activatable-proteases: (1) complementary leucine zippers enables split protease formation (2) split inteins enables protein splicing.

Dimerization domains enable competitive reconstitution of protease activity. For our initial test, we used TVMVP as our starting protease. We designed a construct with cis-caging of TVMVP that can be activated by TEVP-dependent decaging (Fig. 3.1A). We mutated key residues involved in catalytic cleavage in the complementary half of the protease and inserted a TEVP cleavage site between the heterodimerization domains to enable removal of the caging domain. We then placed another TEVP cleavage site between the leucine zipper and inactive half. We reasoned separation of the inactive protease half from the heterodimerization domain would reduce competition with the complementary active half. Expression of the catalytically active TVMVP half with the corresponding heterodimerization domain will displace the inhibitory protease half upon upstream TEVP cleavage. Co-expression of the 'caged' TVMVP with the active of TVMVP revealed caged activity. Cleavage by TEVP resulted in TVMVP activity (Fig. 3.1A).

While this work was being done, a report was published (Fink et al., 2018) in which an activatable TEVP was built using our same initial design with one cleavage site inserted between the leucine zippers. In our hands, we found two cleavage sites permitted wildtype levels of protease activity, but also had a higher background activation (Fig. 3.1B). We therefore aimed to further inactivate the TEVP by caging both halves of the split protease with cognate dead domains (Fig. 3.1B)). We found this design could sufficiently inactivate the TEVP to have minimal baseline activity and be activated by TVMVP (Fig. 3.1B), right two bars).

Additional split orthogonal proteases will enable multiple inputs and more complex circuit designs. We have identified additional proteases (Fig. 3.1D) that will expand the number of orthogonal proteases in CHOMP. While these proteases
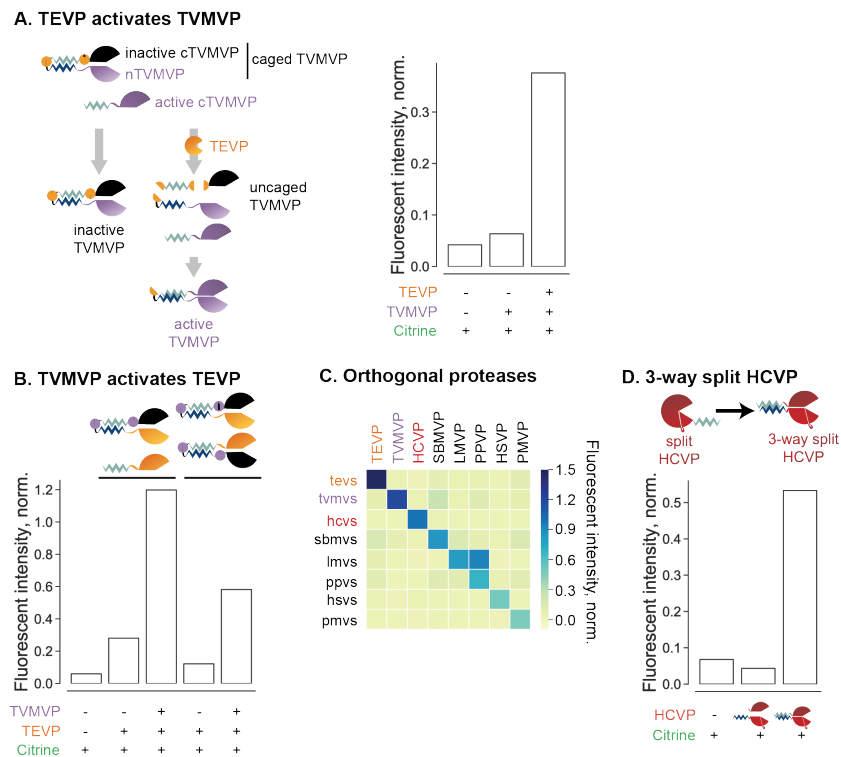
Figure 3.1 | **Design of activatable proteases using leucine zippers.** Design of composable activatable protein components. A, TVMVP is expressed as a single-chain split variant with dimerizing leucine zippers. As indicated the caged TVMVP has an active N-terminal half (nTVMVP) and inactive C-terminal half (cTVMVP). A leucine-zipper tagged active cTVMVP is co- expressed. An active TEVP can cleave the caging inactive cTVMVP away, allowing active nTVMVP to dimerize with active cTVMVP, effectively activating TVMVP. B, The same design can be applied to TEVP activated by TVMVP in which one half can be caged, or both halves can be caged. C, Identification of new orthogonal proteases. The proteases used in CHOMP ((Gao XJ, Chong LS, Kim MS, Elowitz MB, 2018)), are color-coded and do not exhibit cross-reactivity to SBMVP, LMVP, PPVP, HSVP, and PMVP. Of these protease, PPVP exhibits cross-reactivity with LMVP. D, A new tripartite split HCVP where the active HCVP lobe is split in half. The N-terminal half is co-expressed with the activating co-peptide. Dimerization of the two halves with leucine zippers reconsitutes activity.

will require further characterization, we engineered a split site into HCVP. HCVP contains two domains, a core protease domain (NS3 - large pie slice) and activity-enhancing co-peptide (NS4A - small pie slice). We previously inserted a cleavage site between the core and the activity-enhancing peptide to successfully reduce HCVP activity. However, this design is intractable in building an activatable HCVP. Therefore, we split NS3 in half and expressed half of NS3 as a single chain with its co-peptide (Fig. 3.1C). We demonstrated three-way split HCVP activity can be reconstituted with heterodimerizing leucine zippers (Fig. 3.1C).

Split inteins enable engineerable protein splicing. Intein-based activatable proteases are modular and orthogonal. To build circuits that required more than one activatable protease, we required a modular design for protease activation. We found inteins to provide such a basis. Inteins, intervening proteins, are autoprocessing domains which are able to carry out protein splicing (Shah and Muir, 2014). Inteins excise themselves from a polypeptide precursor and ligate the two exteins, external proteins, through a new peptide bond (Shah and Muir, 2014). Split inteins, unlike the contiguous inteins, are translated in two distinct polypeptide sequences of an N-intein and C-intein, each with its own extein (Shah and Muir, 2014). Upon association, the split inteins will perform protein splicing in trans (Fig. 3.2A, adapted from (Gramespacher, Stevens, Nguyen, et al., 2017). Elegant work by the Muir group have developed split intein zymogens in which each split intein pair is caged and activated upon proteolysis (Gramespacher, Stevens, Nguyen, et al., 2017). We decided to use this design to have each extein express one half of a split protease.

We used TEVP as our initial test protease with TVMVP as our activating protease. We caged the extein halves by using the corresponding inactive protease half. We initially saw high background activity and therefore introduced a heterodimerizing leucine zipper pair to each half of the protease on the N-intein in an effort to sterically constrain the ability of the split inteins to associate with each other. We found that addition of the leucine zippers to our overall design enhanced spliced TEVP activity (Fig. 3.2B). We demonstrated this overall design can be transferred from one orthogonal intein pair to another (Npu and NrdJ1) (Gramespacher, Stevens, Nguyen, et al., 2017), and these intein pairs can be used with moderate success with TEVP and TVMVP protease (Fig. 3.2B).

Irreversible intein zymogens and reversible leucine-zipper activatable proteases have different input sensitivities. We have validated two activatable protease systems, each with unique properties. The leucine zipper TVMVP is activated at extremely
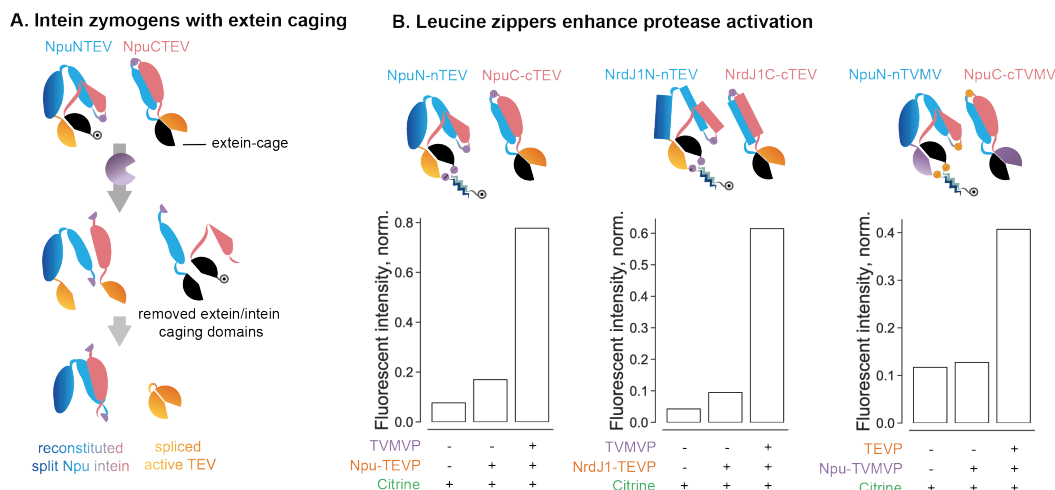
Figure 3.2 | **Adapted intein zymogen design to activate proteases.** A, Intein zymogens along with 'caging' exteins of inactive protease halves decreases basal splicing. B, Leucine zippers along with the extein inactive protease cage and intein zymogen enhance protease activation. Split intein zymogen design is modular and can utilize orthogonal intein pairs, NrdJ1, and Npu, as shown with Npu-TEVP, NrdJ1-TEVP, and Npu-TVMVP.

low TEVP concentrations (Materials and Methods), making it amenable to circuit designs which require high sensitivity to low input concentrations (Fig. 3.1, 3.6). In contrast, the intein zymogen proteases have a higher dynamic range compared to the activatable leucine zipper proteases, but require a significant amount of input to activate the system (Fig. 3.2, Materials and Methods). This thresholding response before full activation of the intein zymogen proteases may be useful for circuit designs which require noise filtering or a delay into a system.

Split inteins enable thresholding via molecular titration. Ultrasensitive responses occur frequently in nature and are required to generate system-level behaviors including signal amplification in cascades, switches, and oscillations (Ferrell and Ha, 2014c; Ferrell and Ha, 2014a; Ferrell and Ha, 2014b). Precise and tunable thresholding is an indispensable process for responding appropriately and specifically to input signals. In particular, ultrasensitive responses to input signals at a certain threshold permits switch-like behavior. Here we describe a protein-level thresholding module that utilizes molecular titration, in which active components are sequestered in an inactive form. We used the Cfa split intein system, made of two proteins: a CfaN-intein and a CfaC-intein, each expressed with its own extein (Stevens, Brown, et al., 2016; Gramespacher, Stevens, R. E. Thompson, et al., 2018). A truncated version of CfaN, CfaN2, lacks an essential domain for irre-
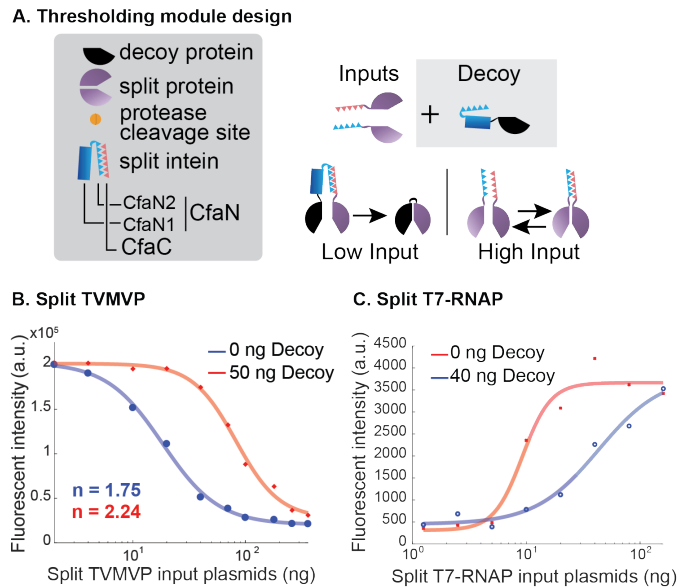
Figure 3.3 | **Molecular titration using split inteins permits tunable thresholds.** (A) Design of thresholding module. The split intein pair CfaC and CfaN is split into three protein domains: CfaC, CfaN, and the dimerization domain of CfaN, CfaN2. Input split proteins can be reconstituted by dimerization of CfaN2 and CfaC. Decoy protein expressed with CfaN can splice with CfaC to molecularly titrate protein inputs. (B) Protease activity can be thresholded. Hill coefficients are indicated by n. (C) Thresholding module is modular. Split T7 RNA polymerase was used as the input and the decoy protein for the thresholding module was an inactive protease half used in (B).

versible protein splicing, but can still reversibly dimerize with CfaC. It can thereby act as a caging domain that inhibits CfaC's ability to productively splice with CfaN (Fig. 3.3).

To engineer thresholding, we will design a system in which productive reconstitution of two half-proteases competes with an unproductive, irreversible event in which an intein on one half-protease splices with a cognate intein on a 'decoy' protein, deemed the 'thresholding module' (TM), mutually inactivating the two proteins (Fig. 3.3). In this system, one half-protease can only accumulate to high levels, and reconstitute full protease activity when its concentration exceeds that of the decoy. This interaction leads to a threshold-linear response (Buchler and Louis, 2008), with the concentration of the decoy protein controlling the threshold.

Using the protease-repressible fluorescent reporter as an output, we co-transfected varying amounts of input plasmids expressing nTVMVP-CfaN2 and CfaC-cTVMVP, with or without the thresholding species, inactivated nTVMVP-CfaN. Compared

with the control group (no thresholding module), protease activity in the experimental group (with thresholding) was fully inhibited under low input concentrations, and increased only when input concentrations crossed a threshold, as expected (Fig. 3.3).

The thresholding module can regulated by an additional protein-level input. The flexibility of the CfaN domains indicated the overall protein architecture may be permissive to protease cleavage site insertion. We introduced a TEV protease cleavage site into the CfaN domain between the CfaN2 and CfaN1 (Fig. 3.3D. This design enables negative regulation of the thresholding module by TEVP cleavage of CfaN. We further optimized this design and and fused heterodimerizing leucine zippers, to the thresholding module and TEVP (Chong et al., 2020). As expected, addition of TEVP strongly reduced the thresholding capability of the system (Chong et al., 2020). These results show proteases provide an additional control module for generating tunable thresholds.

The design presented here is generalizable to control any singular protein or protein complex of interest that can be split and reconstituted. We have successfully applied this thresholding module to split T7-RNA polymerase (Pu, Zinkus-Boltz, and Dickinson, 2017) (Fig. 3.3), split Cas9 (Zetsche, Volz, and F. Zhang, 2015), and a zinc-finger transcription factor (Khalil and James J Collins, 2010). These results validate the principle of molecular titration by split inteins. The design described here should also be compatible with alternative dimerization domains such as other split inteins (Pinto, Thornton, and Wang, 2020; Gramespacher, Stevens, Nguyen, et al., 2017; Gramespacher, Stevens, R. E. Thompson, et al., 2018; Stevens, Sekar, et al., 2017) and heterodimerizing protein domains (Z. Chen et al., 2019; K. E. Thompson et al., 2012).

Activatable proteases in conjunction with repressible proteases enable implementation of diverse dynamic behaviors.

Cells frequently use dynamics to encode multiple signals through the same pathway. This has been demonstrated in multiple systems including p53 (Purvis et al., 2012; Batchelor et al., 2011), growth factor signaling (Marshall, 1995; Santos, Verveer, and Bastiaens, 2007), and Notch (Nandagopal et al., 2018). Converting one type of signal to another, for example a sustained signal to an oscillation, could be particularly beneficial in delineating biological systems and programming endogenous cellular responses. The dynamics of signaling molecules can also guide the design of combinatorial therapies. For example, MDMX is an upstream

supressor of p53 and is frequently overexpressed in cancer. Characterization of p53 signaling dynamics in response to MDMX inhibition showed two distinct phases of p53 accumulation: an initial pulse of p53 and then sustained low-amplitude oscillations (S.-H. Chen, Forrester, and Lahav, 2016). The first pulsatile phase resulted in increased apoptosis while the second phase inhibited cell death. By developing dynamic synthetic post-translational circuits we may couple endogenous inputs into programmed, dynamic outputs based on cellular state.

Protein circuits could perform time derivatives. Many biochemical processes in cells exhibit transient responses to long-lasting changes in environmental and intracellular conditions. Of particular importance is adaptation, in which the cell can respond to a step change in input, and then return to the prestimulus output level (Adler and Alon, 2018). Examples can be found both in single cells and multicellular organisms, such as the adaptation of tumbling probabilities to nutrient levels in bacterial chemotaxis or the response to Wnt stimulation (Goentoro and Kirschner, 2009; Yi et al., 2000). One network motif, the incoherent feed-forward loop (IFFL) in which the input both activates and represses the output, is of particular interest because it can generate diverse dynamic features such as a temporal pulse (Basu et al., 2004), a band-pass filter (Greber and Fussenegger, 2010), a fold-change detector, including exact adaptation behavior (Kim et al., 2014; Adler and Alon, 2018).

Modeling shows that IFFL can produce tunable pulse responses. We used a minimal model to address the question of dynamics between two CHOMP-based pulses. To facilitate dynamic readout of the circuit we used a far-red fluorescent protein (IFP) that is synthesized in a non-fluorescent state, but can be post-translationally switched on by TEVP (To et al., 2015). We also added a conditional N-end degron to enable repression by TVMVP. The first pulse uses an IFFL with a three-step protease cascade to introduce a delay (Gao XJ, Chong LS, Kim MS, Elowitz MB, 2018). The second pulse will use an IFFL with an activatable protease which will repress the output, bypassing the need for a three-step protease cascade (Fig. 3.4), Supplementary Materials). An activatable pulse exhibits higher amplitude with a faster repression response compared to the IFFL with a three-step cascade.

Nonlinearity, activation, and repression enable protein-level oscillators. Periodic fluctuations in protein signaling are widespread and vital to biological functions. Previous work has shown that a three protein system (based on the cyanobacterial circadian clock) can exhibit self-sustaining oscillations of phosphorylation (Nagai,
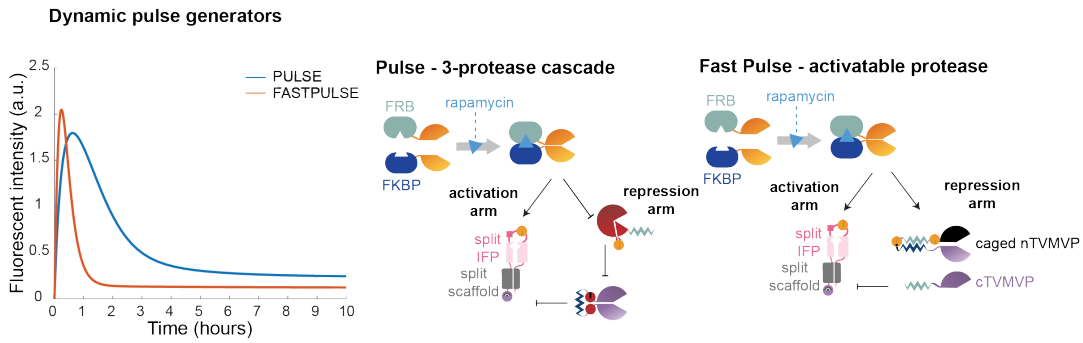
Figure 3.4 | **A tale of two pulses.** Comparison of dynamics between two protease-based pulses. Left, A toy model for two protease-based pulses. Center and right, Schematics for two pulses. Center, delayed repression can enable pulse generation. In this design, rapamycin-induced dimerization of FKBP and FRB domains reconstitutes TEVP. Cleavage of the reporter by TEVP allows maturation of far-red fluorescent protein. TEVP inhibits HCVP activity, thereby permitting TVMVP activity. TVMVP cleavage of the far-red fluorescent protein reveals an N-end degron leading to degradation of the reporter. Right, Similar to the 3-protease cascade pulse, rapamycin reconstitution of TEVP to activates TVMVP. TVMVP activation inhibits the reporter by revealing an N-end degron.

Terada, and Sasai, 2010). The eukaryotic cell cycle is also based, in part, on non-linear sequential protein-level activation steps (Csikász-Nagy et al., 2006). Cells will also induce un-dampened oscillations of proteins in response to stress, as seen with p53 oscillations following gamma irradiation (Geva-Zatorsky et al., 2006). However, it has remained unclear if similar types of protein level oscillations can be rationally designed and engineered. Protein design strategies to control the intrinsic nonlinearity (effective cooperativity) of input-output responses could enable the implementation of useful dynamics such as multistability (Gardner, Cantor, and J J Collins, 2000) and oscillation (Stricker et al., 2008; Elowitz and Leibler, 2000; Potvin-Trottier et al., 2016). The activatable CHOMP components we have designed and validated have the necessary properties required to oscillate.

Specifically, we will create mathematical models to test the parameters required to generate a Hasty oscillator, based on an interacting activator and repressor (Stricker et al., 2008) as well as other oscillatory circuits, including a modified repressilator (Elowitz and Leibler, 2000). The original Hasty oscillator had two key design principles: First, a time delay in the negative feedback loop. Second, a positive feedback loop to increase the robustness of the oscillator (Stricker et al., 2008). Using a simple mathematical model (Appendix A, Fig. 3.5), we have found that self-activating proteases exhibit an ultrasensitive response to input. Signal amplification
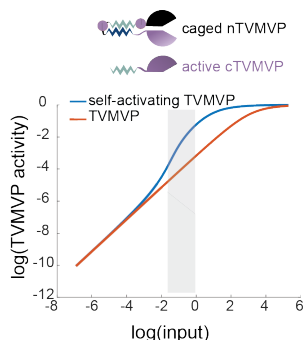
Figure 3.5 | **Self-activating proteases facilitate ultrasensitivity..** A simple mathematical model shows a log-log plot varying the input of leaky caged TVMVP to output activated TVMVP. Grey area highlights the ultrasensitive regime.

(Brandman and Meyer, 2008; Mitrophanov and Groisman, 2008) and bistability (Angeli, Ferrell, and Sontag, 2004; Ajo-Franklin et al., 2007; Becskei, Séraphin, and Serrano, 2001) may emerge as features of ultrasensitive positive feedback loops, enabling dynamic circuits. The protein circuit components we have developed enable nonlinear signal processing necessary for robust oscillations.

CHOMP can detect the presence of flaviviruses. Flaviviruses, single-stranded RNA viruses transmitted from arthropods to humans, cause severe global epidemics including those of dengue virus, West Nile virus, Yellow Fever virus and Zika virus (Pierson and Diamond, 2020). Dengue, alone, can infect up to 400 million people globally per year (Bhatt et al., 2013). Infection can result in hepatitis, encephalitis and fetal death (Pierson and Diamond, 2020). The widespread distribution, spontaneous emergence and re-emergence in human populations, and severe health implications calls for a rapid detection and containment method that functions across a broad spectrum of flaviviruses.

Flaviviruses encode a NS3 protease and its cofactor, NS2B, that proteolytically process the virus polyprotein in the endoplasmic reticulum (ER) in viral production factories (Aktepe and Mackenzie, 2018; Z. Li, J. Zhang, and H. Li, 2017). The NS2B-NS3 protease is necessary for producing infective viral particles and presents as an attractive candidate for antiviral therapeutics; however, previous attempts at targeting the protease had little success (Z. Li, J. Zhang, and H. Li, 2017; Nitsche, 2018). The similarity between the NS2B-NS3 active site to serine proteases present in mammalian cells makes a broad-spectrum antiviral compound likely to cross-react with endogenous proteases (Majerová et al., 2019; Nitsche, 2018).

While each flavivirus encodes a unique NS2B-NS3 protease that preferentially
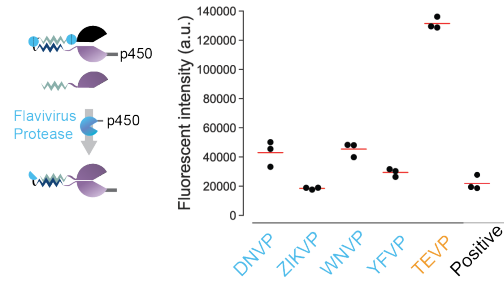
Figure 3.6 | **Design of flavivirus protease detector.** Flavivirus proteases activate split TVMVP. Expression of flavivirus proteases activates TVMVP, resulting in cleavage of an N-end degron repressible reporter.

recognizes its cognate cleavage site, we aimed to develop a flavivirus sense-and-respond circuit capable of detecting multiple flaviviruses. Our initial attempts to detect Dengue virus protease (DNVP) using a repressible CHOMP cascade (Gao XJ, Chong LS, Kim MS, Elowitz MB, 2018) was only able to sense high input concentrations of DNVP. The activatable proteases have high sensitivity to input; therefore, we designed a flavivirus protease-activatable TVMVP. We utilized a promiscuous cleavage site Shiryaev et al., 2007 recognized by both West Nile virus protease (WNVP) and Dengue virus protease (DNVP). To enable efficient flavivirus protease-based activation at the ER, where flavivirus replication occurs, we localized the caged TVMVP half to the ER (Ahn, Szczesna-Skorupa, and Kemper, 1993). We expressed the active half cytoplasmically to reduce background activity. To read out TVMVP activity we integrated a repressible TVMVP fluorescent reporter into HEK293 cells (Gao XJ, Chong LS, Kim MS, Elowitz MB, 2018), Materials and Methods.

We were able to verify the presence of DNVP and WNVP sufficiently activated the caged TVMVP while TEVP could not (Fig. 3.6). Beyond Dengue and West Nile, we validated our protease sensor against Yellow Fever virus protease (YFVP) (Lee et al., 2000; Rice et al., 1985) and Zika virus protease (ZIKVP) (Phoo et al., 2016). This simple design was successfully activated by all four flavivirus proteases (Fig. 3.6). Our results demonstrate how a simple protease-based sensor and signal transmission could function as a broad-spectrum flavivirus detector. We envision replacing the fluorescent reporter with an apoptosis-inducing output such as a protease-activatable Casp3 (Gray, Mahrus, and Wells, 2010) to rapidly kill infected cells. With further development, the CHOMP sensor could be employed as a rapid diagnostic (Myhrvold et al., 2018; Slomovic, Pardee, and James J Collins, 2015) as a way to detect flavivirus emergence or executed as a gene drive (Kyrou

et al., 2018; Champer, Buchman, and Akbari, 2016) to reduce transmission potential within the carrier population.

CHOMP provides a plug-and-play platform to engineer user-defined circuit functionalities into mammalian cells. Incorporation of inactive protease halves and post-translational splicing into the CHOMP toolkit enables positive regulation and nonlinear signal processing. Post-translational circuits that implement ultrasensitive signal processing modules can generate a broad variety of useful dynamic circuit functions including multistability (Gardner, Cantor, and J J Collins, 2000) and oscillations (Elowitz and Leibler, 2000; Stricker et al., 2008). The circuit components described here have diversified the number of ways to approach the same design goal, adding a layer of complexity to circuit design. For example, implementation of an 'AND' gate using CHOMP can either occur on the effector protein level (Gao XJ, Chong LS, Kim MS, Elowitz MB, 2018), or by using an intermediate protease in which both half-proteases are caged and can be reconstituted in the presence of two distinct input proteases. Although powerful in its flexibility, CHOMP will greatly benefit from characterization and standardization of its parts to ideally provide a foundation for automated post-translational circuit design (Nielsen et al., 2016).

Proteolytic processing is an effective mechanism to transmit signal. The ability to cleave protein domains provides a flexible way to regulate protein function. We demonstrate that composable synthetic proteases can control the activity of other synthetic components (Fig. 3.3), can interface with both endogenous proteins (Fig. 2.8) and pathogenic proteins (Fig. 3.6). CHOMP provides a formidable platform for protein-protein regulation, establishing a foundation for circuit-based therapies.

References

Adler, Miri and Uri Alon (Apr. 2018). "Fold-change detection in biological systems". In: *Current Opinion in Systems Biology* 8, pp. 81–89.

Ahn, K, E Szczesna-Skorupa, and B Kemper (Sept. 1993). "The amino-terminal 29 amino acids of cytochrome P450 2C1 are sufficient for retention in the endoplasmic reticulum". en. In: *J. Biol. Chem.* 268.25, pp. 18726–18733.

Ajo-Franklin, Caroline M et al. (Sept. 2007). "Rational design of memory in eukaryotic cells". en. In: *Genes Dev.* 21.18, pp. 2271–2276.

Aktepe, Turgut E and Jason M Mackenzie (Aug. 2018). "Shaping the flavivirus replication complex: It is curvaceous!" en. In: *Cell. Microbiol.* 20.8, e12884.

Angeli, David, James E Ferrell, and Eduardo D Sontag (Feb. 2004). "Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 101.7, pp. 1822–1827.

Basu, Subhayu et al. (Apr. 2004). "Spatiotemporal control of gene expression with pulse-generating networks". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 101.17, pp. 6355–6360.

Batchelor, Eric et al. (May 2011). "Stimulus-dependent dynamics of p53 in single cells". en. In: *Mol. Syst. Biol.* 7, p. 488.

Becskei, A, B Séraphin, and L Serrano (May 2001). "Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion". en. In: *EMBO J.* 20.10, pp. 2528–2535.

Bhatt, Samir et al. (Apr. 2013). "The global distribution and burden of dengue". en. In: *Nature* 496.7446, pp. 504–507.

Brandman, Onn and Tobias Meyer (Oct. 2008). "Feedback loops shape cellular signals in space and time". en. In: *Science* 322.5900, pp. 390–395.

Buchler, Nicolas E and Matthieu Louis (Dec. 2008). "Molecular titration and ultra-sensitivity in regulatory networks". en. In: *J. Mol. Biol.* 384.5, pp. 1106–1119.

Champer, Jackson, Anna Buchman, and Omar S Akbari (Feb. 2016). "Cheating evolution: engineering gene drives to manipulate the fate of wild populations". en. In: *Nat. Rev. Genet.* 17.3, pp. 146–159.

Chen, Sheng-Hong, William Forrester, and Galit Lahav (Mar. 2016). "Schedule-dependent interaction between anticancer treatments". en. In: *Science* 351.6278, pp. 1204–1208.

Chen, Zibo et al. (Jan. 2019). "Programmable design of orthogonal protein heterodimers". en. In: *Nature* 565.7737, pp. 106–111.

Chong, Lucy S et al. (Sept. 2020). "SYNTHETIC SYSTEM FOR TUNABLE THRESHOLDING OF PROTEIN SIGNALS". 20200277333.

Csikász-Nagy, Attila et al. (June 2006). "Analysis of a generic model of eukaryotic cell-cycle regulation". en. In: *Biophys. J.* 90.12, pp. 4361–4379.

Elowitz, Michael B and Stanislas Leibler (Jan. 2000). "A synthetic oscillatory network of transcriptional regulators". In: *Nature* 403, p. 335.

Ferrell Jr, James E and Sang Hoon Ha (Oct. 2014a). "Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity". en. In: *Trends Biochem. Sci.* 39.10, pp. 496–503.

– (Nov. 2014b). "Ultrasensitivity part II: multisite phosphorylation, stoichiometric inhibitors, and positive feedback". en. In: *Trends Biochem. Sci.* 39.11, pp. 556–569.

Ferrell Jr, James E and Sang Hoon Ha (Dec. 2014c). "Ultrasensitivity part III: cascades, bistable switches, and oscillators". en. In: *Trends Biochem. Sci.* 39.12, pp. 612–618.

Fink, Tina et al. (Dec. 2018). "Design of fast proteolysis-based signaling and logic circuits in mammalian cells". en. In: *Nat. Chem. Biol.*

Gao XJ, Chong LS, Kim MS, Elowitz MB (2018). "Programmable protein circuits in living cells". In: *Science*.

Gardner, T S, C R Cantor, and J J Collins (Jan. 2000). "Construction of a genetic toggle switch in Escherichia coli". en. In: *Nature* 403.6767, pp. 339–342.

Geva-Zatorsky, Naama et al. (June 2006). "Oscillations and variability in the p53 system". en. In: *Mol. Syst. Biol.* 2, p. 2006.0033.

Goentoro, Lea and Marc W Kirschner (Dec. 2009). "Evidence that Fold-Change, and Not Absolute Level, of β-Catenin Dictates Wnt Signaling". en. In: *Mol. Cell* 36.5, pp. 872–884.

Gramespacher, Josef A, Adam J Stevens, Duy P Nguyen, et al. (June 2017). "Intein Zymogens: Conditional Assembly and Splicing of Split Inteins via Targeted Proteolysis". en. In: *J. Am. Chem. Soc.* 139.24, pp. 8074–8077.

Gramespacher, Josef A, Adam J Stevens, Robert E Thompson, et al. (Mar. 2018). "Improved protein splicing using embedded split inteins". en. In: *Protein Sci.* 27.3, pp. 614–619.

Gray, Daniel C, Sami Mahrus, and James A Wells (Aug. 2010). "Activation of specific apoptotic caspases with an engineered small-molecule-activated protease". en. In: *Cell* 142.4, pp. 637–646.

Greber, David and Martin Fussenegger (Oct. 2010). "An engineered mammalian band-pass network". en. In: *Nucleic Acids Res.* 38.18, e174.

Iwamoto, Mari et al. (Sept. 2010). "A general chemical method to regulate protein stability in the mammalian central nervous system". en. In: *Chem. Biol.* 17.9, pp. 981–988.

Khalil, Ahmad S and James J Collins (May 2010). "Synthetic biology: applications come of age". en. In: *Nat. Rev. Genet.* 11.5, pp. 367–379.

Kim, Jongmin et al. (May 2014). "Synthetic circuit for exact adaptation and fold-change detection". en. In: *Nucleic Acids Res.* 42.9, pp. 6078–6089.

Kyrou, Kyros et al. (Dec. 2018). "A CRISPR-Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes". en. In: *Nat. Biotechnol.* 36.11, pp. 1062–1066.

Lee, E et al. (Jan. 2000). "Mutagenesis of the signal sequence of yellow fever virus prM protein: enhancement of signalase cleavage In vitro is lethal for virus production". en. In: *J. Virol.* 74.1, pp. 24–32.

Li, Zhong, Jing Zhang, and Hongmin Li (Jan. 2017). "Chapter 7 - Flavivirus NS2B/NS3 Protease: Structure, Function, and Inhibition". In: *Viral Proteases and Their Inhibitors*. Ed. by Satya P Gupta. Academic Press, pp. 163–188.

Majerová, Taťána et al. (Nov. 2019). "Exploiting the unique features of Zika and Dengue proteases for inhibitor design". en. In: *Biochimie* 166, pp. 132–141.

Marshall, C J (Jan. 1995). "Specificity of receptor tyrosine kinase signaling: Transient versus sustained extracellular signal-regulated kinase activation". en. In: *Cell* 80.2, pp. 179–185.

Mitrophanov, Alexander Y and Eduardo A Groisman (June 2008). "Positive feedback in cellular control systems". en. In: *Bioessays* 30.6, pp. 542–555.

Myhrvold, Cameron et al. (Apr. 2018). "Field-deployable viral diagnostics using CRISPR-Cas13". en. In: *Science* 360.6387, pp. 444–448.

Nagai, Tetsuro, Tomoki P Terada, and Masaki Sasai (June 2010). "Synchronization of circadian oscillation of phosphorylation level of KaiC in vitro". en. In: *Biophys. J.* 98.11, pp. 2469–2477.

Nandagopal, Nagarajan et al. (Feb. 2018). "Dynamic Ligand Discrimination in the Notch Signaling Pathway". en. In: *Cell* 172.4, 869–880.e19.

Nielsen, Alec A K et al. (Apr. 2016). "Genetic circuit design automation". en. In: *Science* 352.6281, aac7341.

Nitsche, Christoph (2018). "Strategies Towards Protease Inhibitors for Emerging Flaviviruses". en. In: *Adv. Exp. Med. Biol.* 1062, pp. 175–186.

Phoo, Wint Wint et al. (Nov. 2016). "Structure of the NS2B-NS3 protease from Zika virus after self-cleavage". en. In: *Nat. Commun.* 7.1, pp. 1–8.

Pierson, Theodore C and Michael S Diamond (June 2020). "The continued threat of emerging flaviviruses". en. In: *Nat Microbiol* 5.6, pp. 796–812.

Pinto, Filipe, Ella Lucille Thornton, and Baojun Wang (Mar. 2020). "An expanded library of orthogonal split inteins enables modular multi-peptide assemblies". en. In: *Nat. Commun.* 11.1, pp. 1–15.

Potvin-Trottier, Laurent et al. (Oct. 2016). "Synchronous long-term oscillations in a synthetic gene circuit". en. In: *Nature* 538.7626, pp. 514–517.

Pu, Jinyue, Julia Zinkus-Boltz, and Bryan C Dickinson (Apr. 2017). "Evolution of a split RNA polymerase as a versatile biosensor platform". en. In: *Nat. Chem. Biol.* 13.4, pp. 432–438.

Purvis, Jeremy E et al. (June 2012). "p53 dynamics control cell fate". en. In: *Science* 336.6087, pp. 1440–1444.

Rice, C M et al. (Aug. 1985). "Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution". en. In: *Science* 229.4715, pp. 726–733.

Santos, Silvia D M, Peter J Verveer, and Philippe I H Bastiaens (Feb. 2007). "Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate". In: *Nat. Cell Biol.* 9, p. 324.

Shah, Neel H and Tom W Muir (2014). "Inteins: Nature's Gift to Protein Chemists". en. In: *Chem. Sci.* 5.1, pp. 446–461.

Shiryaev, Sergey A et al. (Feb. 2007). "Cleavage preference distinguishes the two-component NS2B-NS3 serine proteinases of Dengue and West Nile viruses". en. In: *Biochem. J* 401.3, pp. 743–752.

Slomovic, Shimyn, Keith Pardee, and James J Collins (Nov. 2015). "Synthetic biology devices for in vitro and in vivo diagnostics". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 112.47, pp. 14429–14435.

Stevens, Adam J, Zachary Z Brown, et al. (Feb. 2016). "Design of a Split Intein with Exceptional Protein Splicing Activity". en. In: *J. Am. Chem. Soc.* 138.7, pp. 2162–2165.

Stevens, Adam J, Giridhar Sekar, et al. (Aug. 2017). "A promiscuous split intein with expanded protein engineering applications". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 114.32, pp. 8538–8543.

Stricker, Jesse et al. (Nov. 2008). "A fast, robust and tunable synthetic gene oscillator". en. In: *Nature* 456.7221, pp. 516–519.

Thompson, Kenneth Evan et al. (Apr. 2012). "SYNZIP protein interaction toolbox: in vitro and in vivo specifications of heterospecific coiled-coil interaction domains". en. In: *ACS Synth. Biol.* 1.4, pp. 118–129.

To, Tsz-Leung et al. (Mar. 2015). "Rationally designed fluorogenic protease reporter visualizes spatiotemporal dynamics of apoptosis in vivo". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 112.11, pp. 3338–3343.

Yi, Tau-Mu et al. (Apr. 2000). "Robust perfect adaptation in bacterial chemotaxis through integral feedback control". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 97.9, pp. 4649–4653.

Zetsche, Bernd, Sara E Volz, and Feng Zhang (Feb. 2015). "A split-Cas9 architecture for inducible genome editing and transcription modulation". en. In: *Nat. Biotechnol.* 33.2, pp. 139–142.

Zhang, Qiang, Sudin Bhattacharya, and Melvin E Andersen (Apr. 2013). "Ultrasensitive response motifs: basic amplifiers in molecular signalling networks". en. In: *Open Biol.* 3.4, p. 130031.

3.3   Materials and Methods

Plasmid construction

All constructs were generated using standard procedures. The backbones were linearized using restriction digestion or PCR, and inserts were generated using PCR or gBlock synthesis (IDT).

Tissue culture

The Flp-In™ T-REx™ 293 Cell Line (Human Embryonic Kidney cells that contain a single stably integrated FRT site at a transcriptionally active genomic locus, and stably expressing the tetracycline repressor protein) was purchased from Thermo Fisher Scientific (R78007). Cells were cultured in a humidity controlled chamber at $37°$ C with 5% CO_2 in media containing DMEM supplemented with 10% FBS, 1 mM sodium pyruvate, 1 unit/ml penicillin, 1 μg/ml streptomycin, 2 mM L-glutamine and 1X MEM non-essential amino acids. 100 ng/mL doxycycline was added whenever expression is needed from a CMV-TO promoter. Protease-repressible reporter transgenes were integrated with PiggyBac-based integration and a polycloncal cell population was established through FACS.

Transient transfection

293 cells were seeded at a density of 0.05 x 10^6 cells per well of a 24-well plate and cultured under standard conditions overnight. The following day, the cells were transfected with plasmid constructs using Lipofectamine 2000 (Thermo Fisher) as per manufacturer's protocol.

Flow cytometry

Two days after transfection, cells were prepared for flow cytometry by trypsinizing with 30μL of 0.05% trypsin for 1 min at room temperature. Protease activity was neutralized by resuspending the cells in buffer containing 70 μL of HBSS with 2.5mg/ml Bovine Serum Albumin (BSA). For cells stimulated with EGF, cells were resuspended in buffer containing 70μ L of HBSS with 2.5 mg/mL BSA and 1 mM EDTA. Cells were then filtered through a 40 μm cell strainer and analyzed by flow cytometry (MACSQuant VYB, Miltenyi or CytoFLEX, Beckman Coulter). We used the EasyFlow Matlab-based software package developed in-house by Yaron Antebi to process flow cytometry data.

Fluorescent signal quantification from flow cytometric measurements

To maximize the observable reporter dynamic range, we selected and compared cells with the highest expression of the co-transfection marker, which showed the largest separation of basal reporter fluorescence from cellular autofluorescence (Figure 2.10). For each sample in a comparison group (experiments performed in the same batch and data shown on the same plot), we calculated the 98 and 99.5 percentiles of fluorescence of the co-transfection marker (mCherry in most cases). We identified the sample with the lowest 98 percentile value, and used its 98 and 99.5 percentiles as lower and upper limits to gate on all samples. For all cells within the gate in each sample, we fit the distribution of the logarithm of their signal fluorescence (Citrine in most cases) with skew Gaussian distributions, i.e. N∗ normcdf(x,m,k)∗ normpdf(x,m,s) in Matlab using non-linear least-square fitting, and reported the mode (peak position, representing the reporter level that's most likely to be observed) of the resulting fit (Fig. 2.10). Here, the normcdf(x, μ, σ) and normpdf(x, μ, σ) functions are cumulative probability density and probability density functions for a Gaussian distribution respectively, and the parameter n is a normalization factor, m=μ is the mean of the Gaussian function, s=σ is the inverse standard deviation of the Gaussian, and k parameterizes skewness. Fluorescent intensity was normalized to the positive control, the cognate full-length wildtype protease and is reported as normalized fluorescent intensity in Figs. 3.1 and 3.2

3.4   Supplementary Materials and Methods

MATHEMATICAL MODELING

For the mathematical models we have considered typical biochemical parameter values. We note that additional features of any specific system, including feedback structure, could impact their dynamic behavior. Additionally, the quantitative values of the resulting timescales in general depend on the specific choice of biochemical parameter values. We assume that protease cleavage functions in a linear regime far from saturation, consistent with published K_m values (Tözsér et al., 2005)and our bandpass modeling [(Gao XJ, Chong LS, Kim MS, Elowitz MB, 2018), Figures 3B-D, Methods], the reaction can be expressed as a set of ordinary differential equations (ODEs).

Ultrasensitive responsive motif through positive auto-regulation In this supplementary section, we use a minimal model to address the question of how self-

activation of a protease permits ultrasensitivity. We modeled self-activation of one protease by itself through direct cleavage based on the scheme in Fig. 3.5. We assumed the caged protease, P_c, is produced at a constant rate α. Based on the reversibility of the leucine zipper system would allow for temporary uncaging of the caged protease, produced at a rate α. The leaky open confirmation of P_c would contribute to the uncaged state of the protease, P_{unc}, at a rate of k_1. The active protease half is produced at a constant rate of β. The dimerization mediated by the leucine zippers of P_{unc} with its active protease half, P_{half}, proceeds at a rate of k_{on}. This dimerization event results in the active protease P. The activated protease P, can dissociate at a rate of k_{off}, resulting in the production of P_{unc} and P_{half}. P can activate P_{unc} at a rate of k_2. Each of these substrates, P_c, P_{unc}, and P have a first-order degradation with rate γ_p. These reactions may be written as a set of ordinary differential equations (ODEs):

$$\frac{dP_c}{dt} = \alpha - k_2 P P_c - k_1 P_c - \gamma_p P_c$$

$$\frac{dP_{half}}{dt} = \beta - k_{on} P_{unc} P_{half} + k_{off} P - \gamma_p P_{half}$$

$$\frac{dP_{unc}}{dt} = k_2 P_c P - k_{on} P_{unc} P_{half} + k_{off} P - \gamma_p P_{unc}$$

$$\frac{dP}{dt} = k_{on} P_{unc} P_{half} - k_{off} P - \gamma_p P$$

Because the absolute value of the production rate β does not affect the dynamics of the system, we arbitrarily set its value to $1\,\mathrm{M}h^{-1}$. For the association rate, we assumed $k_{on} = 10 h^{-1}$. For the dissociation rate, we assumed $k_{off} = 5 h^{-1}$ based on indirect measurements (Kim et al., 2017). For the protein degradation rate, we assumed a biologically realistic value of $\gamma_p = 0.1 h^{-1}$. Based on our bandpass fits ((Gao XJ, Chong LS, Kim MS, Elowitz MB, 2018) Figs. 3B-D, Methods), cleavage by a protease, k_2, when the input protease activity is high, occurs at a rate comparable to the rate of degron-mediated degradation ($\sim 5 h^{-1}$). We assumed that the leaky activation by uncaged P_c would have be functionally similar to the leucine zipper association rate of $10 h^{-1}$.

We varied the production rate, α , of the leaky open confirmation of P_c and measured the output of P . In Figure 3.5, we have plotted the input-output response on a

log-log scale to determine the response coefficient (Ferrell and Ha, 2014). We found that in a certain regime, the response coefficient is greater than 1, indicating that self-activating proteases exhibit an ultrasensitive response.

A Tale of Two Pulses In this supplementary section, we use a minimal model to address the question of how a pulse using a classic IFFL with an activatable protease differs in dynamics from an IFFL with a delayed repression-arm using repressible proteases. To make a controlled comparison between the two kinds of pulses, we assume that shared biochemical parameters, such as protein degradation rates, are similar in the two systems. The main conclusion is that the regulation from an activatable proteases can occur more rapidly than the regulation from a three-protease cascade.

3-protease cascade pulse: Please refer to (Gao XJ, Chong LS, Kim MS, Elowitz MB, 2018) which is reproduced here for ease of comparison between the two circuit designs. We modeled repression of one protease by another through direct cleavage, based on the scheme in Fig 1E. We assume the concentration of the input protease, denoted P_1 , is maintained at a constant level, with its activity controlled by a small molecule input, a, as in the scheme of Fig 3E. The second protease, denoted P_2, , is produced at a constant rate β , and undergoes first-order degradation with rate γ_p . The input protease cleaves the second protease at a single cleavage site, converting it to an inactive cleaved form, with a cleavage rate constant k_1 . The cleaved protease irreversibly dissociates at rate k_d , and undergoes first-order degradation with rate γ_p for a total rate of elimination of $\gamma_p k_d$. The third protease is produced at a constant rate β . The second protease, P_2, cleaves the third protease, denoted P_3 , at a cleavage rate constant k_2 . Similarly, the third cleaved protease irreversibly dissociates at rate k_d , and undergoes first-order degradation with rate γ_p for a total rate of elimination of $\gamma_p k_d$.

These reactions can be expressed as a set of ordinary differential equations (ODEs):

$$\frac{dP_1}{dt} = \beta - \gamma_p P_1$$

$$\frac{dP_2}{dt} = \beta - a k_1 P_1 P_2 - \gamma_p P_2$$

$$\frac{dP_3}{dt} = \beta - k_2 P_2 P_3 - \gamma_p P_3$$

The reporter is produced at a constant rate β. The input protease, P_1, when turned on by the small molecule input a, will cleave the inactive reporter IFP_{in} at a rate of k_1. The inactive reporter undergoes first-order degradation at a rate of γ_p. The active reporter, denoted IFP_{act}, is formed from the cleavage by the input protease, at a rate of k_1. The active reporter undergoes first-order degradation at a rate of γ_p. The third protease C cleaves both the active reporter and the inactive reporter at a rate of k_3. Cleavage of the active reporter by protease C reveals a N-end degron on the reporter, denoted by IFP_{deg}. Cleavage of the inactive reporter by protease C will form a distinct species that only has the revealed N-end degron, denoted by IFP_{indeg}. Revealing the N-end degron will accelerate the rate of degradation resulting a first-order degradation of γ_N.

$$\frac{dIFP_{in}}{dt} = \beta - a k_1 P_1 IFP_{in} - k_3 P_3 IFP_{in} - \gamma_p IFP_{in}$$

$$\frac{dIFP_{act}}{dt} = a k_1 P_1 IFP_{in} - k_3 P_3 IFP_{act} - \gamma_p IFP_{act}$$

$$\frac{dIFP_{indeg}}{dt} = k_3 P_3 IFP_{in} - a k_1 P_1 IFP_{indeg} - \gamma_N IFP_{indeg}$$

$$\frac{dIFP_{\text{deg}}}{dt} = a k_1 P_1 IFP_{indeg} + k_3 P_3 IFP_{act} - \gamma_N IFP_{\text{deg}}$$

Activatable protease pulse: We modeled repression of one protease by another through direct cleavage, based on the scheme in Fig 1E. We assume the concentration of the input protease, denoted P_1, is maintained at a constant level, with its activity controlled by a small molecule input, a, as in the scheme of Fig 3E. The activatable protease, denoted P_{2c}, is produced at a constant rate β, and undergoes first-order degradation with rate γ_p. The input protease cleaves the caged protease at a single cleavage site, converting it to a cleaved, uncaged form, whose concentration is denoted P_{2unc}, with a cleavage rate constant k_1. The uncaged protease undergoes a first-order degradation with rate γ_p. The uncaged protease reversibly associates with the active half, denoted P_{2half}, at a rate of k_{on} to form an active protease P_2. The active protease may dissociate at a rate of k_{off}. We assume a single cleavage

for simplicity, but the same conclusions hold true for two independent cleavage sites, cleavage of either of which is sufficient to activate the output protease.

These reactions can be expressed as a set of ordinary differential equations (ODEs):

$$\frac{dP_1}{dt} = \beta - \gamma_p P_1$$

$$\frac{dP_{2c}}{dt} = \beta - a k_1 P_1 P_{2c} - \gamma_p P_{2c}$$

$$\frac{dP_{2unc}}{dt} = a k_1 P_1 P_{2c} - P_{2unc} P_{2half} k_{on} + P_{2c} k_{off} - \gamma_p P_{2unc}$$

$$\frac{dP_{2half}}{dt} = \beta - P_{2unc} P_{2half} k_{on} - P_{2c} k_{off} - \gamma_p P_{2half}$$

$$\frac{dP_2}{dt} = P_{2unc} P_{2half} k_{on} - P_{2c} k_{off} - \gamma_p P_2$$

We modeled the activation and repression of the reporter through direct cleavage by the input protease, P_0 , and the activated protease, P_2 . The input protease, when turned on by the small molecule input a, will cleave the inactive reporter IFP_{in} at a rate of k_1 . The inactive reporter undergoes first-order degradation at a rate of γ_p . The activated protease can also cleave the inactive reporter at a rate of k_2 . The active reporter, denoted IFP_{act} , is formed from the cleavage by the input protease, at a rate of k_1 . The active reporter undergoes first-order degradation at a rate of γ_p . Cleavage of the active reporter by the activated protease occurs at a rate of k_2 to form a reporter with a revealed N-end degron, denoted by $IFP_{\deg}$. Cleavage of the inactive reporter by protease P_2 at a rate of k_2 will form a distinct species that only has the revealed N-end degron, denoted by IFP_{indeg} . Species of the reporter with a revealed N-end degron will accelerate the rate of first-order degradation, denoted by γ_N .

The reaction can be expressed as a set of ordinary differential equations (ODEs):

$$\frac{dIFP_{in}}{dt} = \beta - a k_1 P_1 IFP_{in} - k_2 P_2 IFP_{in} - \gamma_p IFP_{in}$$

$$\frac{dIFP_{act}}{dt} = ak_1P_1IFP_{in} - k_2P_2IFP_{act} - \gamma_p IFP_{act}$$

$$\frac{dIFP_{indeg}}{dt} = k_2P_2IFP_{in} - ak_1P_1IFP_{indeg} - \gamma_N IFP_{indeg}$$

$$\frac{dIFP_{deg}}{dt} = ak_1P_1IFP_{indeg} + k_2P_2IFP_{act} - \gamma_N IFP_{deg}$$

Parameters for the pulses: Because the absolute value of the production rate A does not affect the dynamics of the system, we arbitrarily set its value to $1 \text{ M}h^{-1}$. For the protein degradation rate, we assumed a biologically realistic value of $\gamma_p = 0.1h^{-1}$. For the protein degradation rate with a revealed N-end degron, we assumed a faster degradation rate of $\gamma_p = 5h^{-1}$.

Based on our bandpass fits (Figs. 3B-D, Methods), cleavage by a protease, when the input protease activity is high, occurs at a rate comparable to the rate of degron-mediated degradation ($\sim 5h^{-1}$). We also assumed that the OFF input protease is 20-fold less active than the ON state based on the dynamic range observed in Figure 1. (Note that the value of this regulatory range does not affect our conclusions about the timescales of regulation.) Finally, we assumed the small-molecule-induced ON-OFF switch reaches steady-state much faster than the other reactions, so that the cleavage term can be approximated by a step function, taking one of two possible values:

$k_1P_1 = 0.25h^{-1}$ (input OFF) or $5 \ h^{-1}$ (input ON)

To simulate output dynamics in response to changes in the input, we first set the input protease to OFF and allowed the system to reach its steady-state values. We then switched the input to ON and simulated the equations for 10 hours. In Fig. 3.4, we plot the resulting dynamics of the active fluorescent reporter IFP_{act}. We found that the activatable pulse had a larger amplitude and faster repression than the three-protease cascade pulse.

Chapter 4

ENGINEERING MULTIPLE LEVELS OF SPECIFICITY INTO A RNA VIRAL VECTOR

This work was submitted to bioRxiv as "Engineering multiple levels of specificty into a RNA viral vector" in 2020, (Gao et al., 2020).

4.1 Abstract

Synthetic molecular circuits could provide powerful therapeutic capabilities, but delivering them to specific cell types and controlling them remains challenging. An ideal "smart" viral delivery system would enable controlled release of viral vectors from "sender" cells, conditional entry into target cells based on cell-surface proteins, conditional replication specifically in target cells based on their intracellular protein content, and an evolutionarily robust system that allows viral elimination with drugs. Here, combining diverse technologies and components, including pseudotyping, engineered bridge proteins, degrons, and proteases, we demonstrate each of these control modes in a model system based on the rabies virus. This work shows how viral and protein engineering can enable delivery systems with multiple levels of control to maximize therapeutic specificity.

4.2 Main Text

The ability to deliver a designed nucleic acid to target cell types would open up powerful possibilities for basic research and therapeutic applications. Improved delivery capabilities could enable new generations of gene therapy that take advantage of advances in synthetic biology to provide increased specificity and control. In fact, synthetic biologists have now developed a broad range of biological circuit designs that sense and respond to endogenous cellular states (Fischbach, Bluestone, and Lim, 2013; Kitada et al., 2018; Weber and Fussenegger, 2011). These include systems based on protein-DNA interaction(Keung et al., 2015), regulatory RNAs (Benenson, 2012), and protein-level circuits (Gao XJ, Chong LS, Kim MS, Elowitz MB, 2018; Fink et al., 2018; Siciliano et al., 2018), as well as combinations of these modalities (Cella et al., 2018; Xie et al., 2011). By operating intracellularly, they would not be limited to sensing surface proteins, but could instead directly interrogate the core pathways that drive cellular behaviors, allowing conditional responses

to specific cellular states. While circuit engineering has progressed enormously, our limited ability to deliver circuits to cells in vivo has prevented their therapeutic use. A specific, effective, and controllable method to transfer circuits into target cell types could provide a foundation for the future development of circuits as therapeutics.

Viruses possess many powerful capabilities as delivery systems. Viruses can preferentially infect specific cell types, and then replicate intracellularly to high levels, enabling strong expression of virally encoded proteins, potentially including engineered "cargo" genes. Researchers have therefore engineered diverse classes of DNA and RNA viruses for gene therapy (Kotterman, Chalberg, and Schaffer, 2015; Thomas, Ehrhardt, and Kay, 2003) and cancer therapeutic (Stanford et al., 2008) applications. Among these, RNA riboviruses (Kotterman, Chalberg, and Schaffer, 2015; Thomas, Ehrhardt, and Kay, 2003; John J Holland, 2012) (RNA viruses, excluding retroviruses) offer unique advantages, since they remain at the RNA level, thereby avoiding integration into the host genome and potential mutagenesis of the host. More specifically, viruses in the order Mononegavirales have compact, well-studied genomes that offer multiple avenues for engineering control, can support high level expression of protein components, and exhibit relatively high genomic stability (Drake and J J Holland, 1999). Notable examples of engineered viruses from this order include vesicular stomatitis virus (VSV) for oncolytic viral therapies (Felt and Grdzelishvili, 2017), Sendai virus for stem cell reprogramming (Nakanishi and Otsu, 2012), and rabies virus for synaptic tracing (Callaway and Luo, 2015).

Rabies virus could provide an ideal platform to demonstrate multiple levels of external and internal control with minimal risk of host genome integration. Rabies virus has been well-characterized and extensively used in neurobiology contexts. Further, its ability to spread can be limited and controlled by deleting the glycoprotein, G, from the viral genome and supplying it in the host host cell, in trans (Wickersham, Finke, et al., 2007). Such G-deleted rabies viruses can also be pseudotyped to target specific cell surface proteins (Wickersham, Finke, et al., 2007). Most recently, researchers have explored the possibility of making rabies infection transient through modification of an essential viral protein (Ciabatti et al., 2017).

Nevertheless, transforming rabies virus into a practical circuit delivery system requires integrating existing control mechanisms into a single platform, and developing additional capabilities that together allow independent control of multiple stages of the viral life cycle. To limit viral production in space and time, this system should enable control of viral exit from virus-producing cells based on inducers or envi-

ronmental signals. To restrict the virus to specific targeted cell types, it should also limit viral entry to cells expressing selected cell-surface antigens and make viral replication conditional on the expression of intracellular proteins. For external control, the overall system should be controllable with a well characterized and safe small-molecule drug. To maximize safety and minimize unintended toxicity from the virus itself, viral infection should be reversible through a control system that is robust to mutations.

Here, we set out to develop an integrated system that enables control of viral release, entry, and replication by combining existing rabies virus control mechanisms with new mechanisms for control of viral replication (Fig. 4.1). These results demonstrate and integrate multiple levels of control, and provide a framework for engineering a broadly useful RNA viral delivery platform for emerging applications in therapeutic synthetic biology and conventional gene therapy.
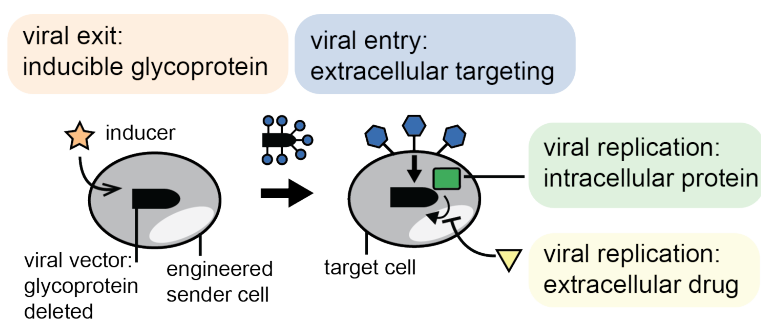

Figure 4.1 | **An ideal viral circuit delivery system with four distinct levels of control.** Sender cells (left) would be engineered to secrete the virus under the control of an inducer (orange square). Produced viruses would be packaged using a heterologous glycoprotein ("pseudotyped", blue circles) to selectively infect target cells (right) expressing a desired surface antigen (blue hexagons). Once inside the target cell, viral replication would be conditional on the presence of an intracellular protein (green square). For safety, it would also be possible to suppress viral replication and eliminate the virus using a drug (yellow triangle).

Doxycycline-inducible expression of glycoprotein controls viral exit from sender cells. We first engineered a controllable 'sender' cell line that releases viral particles in response to external induction. The system takes advantage of the well-characterized paradigm of glycoprotein (G) trans-complementation, in which the G gene can be removed from the viral genome and supplied instead in the host cell line, in order to permit single step infection 18 (Fig. 4.2a). We deleted the native viral G gene, replacing it with mCherry for visualization (RVdG). We then incorporated a single copy of the G transgene in the host genome using the Flp-In

system in HEK293 cells (Fig. 4.2a, b, Methods). The ectopic G gene was controlled by a CMV-TO promoter that was normally repressed by constitutively expressed TetR, but could be readily induced by addition of doxycycline (dox) (Fig. 4.2a, b, Methods). We also stably incorporated a nuclear-localized Citrine (H2B-Citrine) as a marker. This design can be adapted to allow conditional regulation of G protein expression by other transcriptional inputs, including natural or synthetic signaling pathways (Morsut et al., 2016; Chung, Zou, et al., 2019; Daringer et al., 2014).

To validate the inducibility of viral release, we co-cultured a minority of RVdG-infected sender cells with a majority of HEK293 target cells (HEKwt), in the presence or absence of doxycycline. After 3 days, we measured the level of mCherry signal in target cells using flow cytometry. Target cell mCherry intensities exhibited a bimodal distribution, consistent with the presence of uninfected and infected cell populations (Fig. 4.2c). Further, the infected (high mCherry intensity) sub-population showed relatively uniform (coefficient of variation = 0.23 ± 0.01 for triplicates in the "27,000 sender" condition) mCherry expression, suggesting that viral replication drives its concentration to a well-defined concentration within the cell.

Using this system, we next asked how the rate of target cell infection depended on doxycycline induction and sender cell density. The clean separation between the mCherry distributions allowed quantification of infected and uninfected population sizes in different conditions (Fig. 4.2b right, Supplementary Methods). As expected, infection rate depended strongly, though not absolutely, on the presence of doxycycline, with low basal levels of infection likely due to leaky expression of G from its CMV-based TetR-repressible promoter (Fig. 4.2b, right). Infection also exhibited a dose-dependent, but sub-linear, increase with sender cell fraction (Fig. 4.2b, right).

This approach can be extended to enable controlled secretion of pseudotyped viruses, in which the native viral glycoprotein is replaced with a distinct viral glycoprotein conferring different tropism 25. For example, the well-characterized EnvA glyco-protein from the avian sarcoma leukosis virus binds specifically to the avian TVA protein, and abolishes rabies' original tropism for mammalian cells (Wickersham, Finke, et al., 2007; J. A. Young, Bates, and Varmus, 1993). As a demonstration, we engineered an EnvA-pseudotyped sender cell line and a cognate target cell line expressing TVA (Fig. 4.2c). As expected, viral infection was strongly dox-dependent and increased in a dose-dependent manner with sender cell number (Fig. 4.2c, right). Together, these results indicate that inducible glycoprotein expression enables exter-
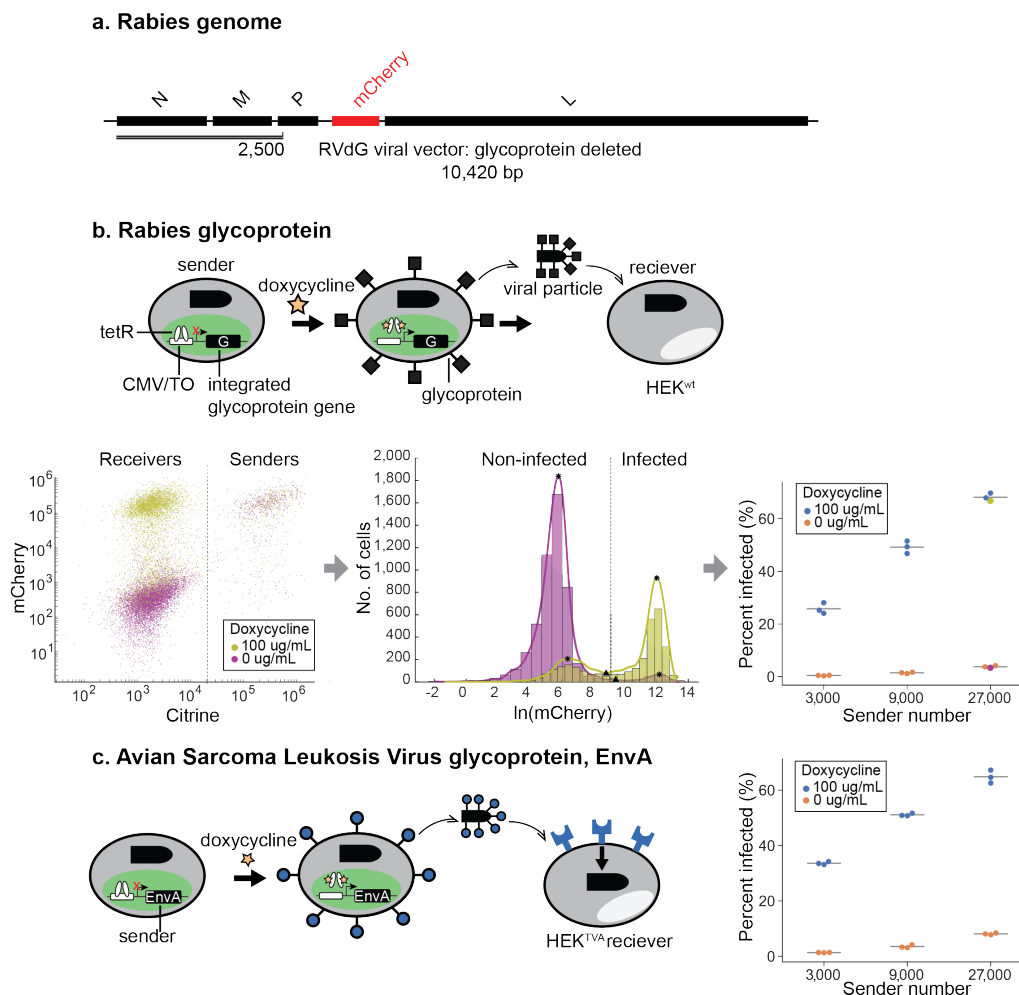
Figure 4.2 | **Doxycycline-inducible expression of glycoprotein controls viral exit from sender cells.** (a) Of the five proteins encoded by the rabies viral genome, the glycoprotein (G) gene is replaced with mCherry (RVdG). The RVdG virus is reconstituted and propagated in producer cells that provide G in trans from the cellular genome. (b) Top, sender cell lines with a single copy of genomically integrated, doxycycline-inducible G for the controlled secretion of RVdG. The senders are also labeled with constitutively expressed Citrine. Wildtype HEK293 (HEKwt) target cells are co-cultured with the senders to quantify the release of infection competent RVdG. Left, flow cytometry of sender and receiver co-culture in the presence or absence of doxycycline. Vertical line separates receiver and sender subpopulations. Middle, distribution of mCherry signal in HEKwt in the presence or absence of doxycycline (after gating out Citrine positive sender cells) measured with flow cytometry. Vertical line indicates mean local minima used to threshold between infected and weakly infected/non-infected cells. The "percent infected" metric reflects the fraction of cells with signal above the threshold (see Methods). The data points in the swarmplot corresponding to the examples are indicated with stars. Similar swarmplots are used in subsequent panels. Right, percent of target cell infected in the presence or absence of doxycycline under varying numbers of sender cells. (c) Replacing the G protein in b with a sarcoma leukosis virus glycoprotein gene (EnvA) under a doxycycline-inducible promoter. Right, percent of TVA-displaying target cell (HEKTVA) infected in the presence or absence of doxycycline under varying numbers of sender cells.

nal regulation of viral secretion for viruses both with wildtype G and pseudotyped with EnvA.

Pseudotyping and bridge proteins control viral entry. Pseudotyping opens up the possibility of targeting engineered rabies viruses to specific mammalian cell types based on their expression of surface proteins. To achieve this, we took inspiration from previous efforts of targeting viral infection to specific cells based on their surface proteins (Snitkovsky and J. A. Young, 1998; Snitkovsky, Niederman, et al., 2000; Boerger, Snitkovsky, and J. A. Young, 1999; Choi, J. A. T. Young, and Callaway, 2010). We engineered a bivalent "bridge" protein, consisting of the TVA extracellular domain fused to a nanobody that recognizes GFP (Kirchhofer et al., 2010) (Fig. 4.3a). In this modular scheme, viral entry requires both the bridge protein and expression of GFP on the surface of the target cell.

We purchased EnvA-pseudotyped rabies virus (RVdG-EnvA), engineered HEK293 derivatives inducibly expressing GFP or Turq2 (which exhibits shifted spectra but is still recognized by the GFP nanobody) on the cell surface as target cells (HEK-memGFP, HEKmemTurq2 Fig. 4.3a, left), and also engineered the bridge protein, termed bridgeGFP. In order to characterize the specificity of the resulting system, we added RVdG-EnvA to co-cultured parental HEKwt cells and target HEKmemGFP cells, either with or without the bridgeGFP. As expected, infection was strongly enhanced by the combination of surface GFP/Turq2 expression and the addition of bridgeGFP (Fig. 4.3a, right, Supplementary Fig. 4.7a). Infection of HEKmemGFP cells was not detectable without bridgeGFP at a multiplicity of infection (MOI) of 1. At much higher MOIs of 5, some non-specific infection of HEKwt and HEK-memGFP cells did occur without bridge protein (Fig. 4.3b). However, the infection rate of HEKmemGFP was 8-fold higher with the bridge protein than without it, even at this high MOI value. Furthermore, even at a more modest MOI of 1, infection remained dependent on the bridge protein (Supplementary Fig. 4.7b). The pseudo-typing bridge protein strategy thus provided strong infection specificity contingent upon surface proteins. In principle, the same design could be adapted to target natural cell types based on cell surface markers by replacing the GFP-targeting nanobody with a corresponding binding domain.

We next sought to combine inducible secretion with bridge-dependent infection. We co-cultured RVdG infected EnvA sender cells together with both target HEK-memTurq2 and non-target HEKCit cells. We then analyzed the relative infection rates of target and non-target cells across different viral secretion rates and numbers
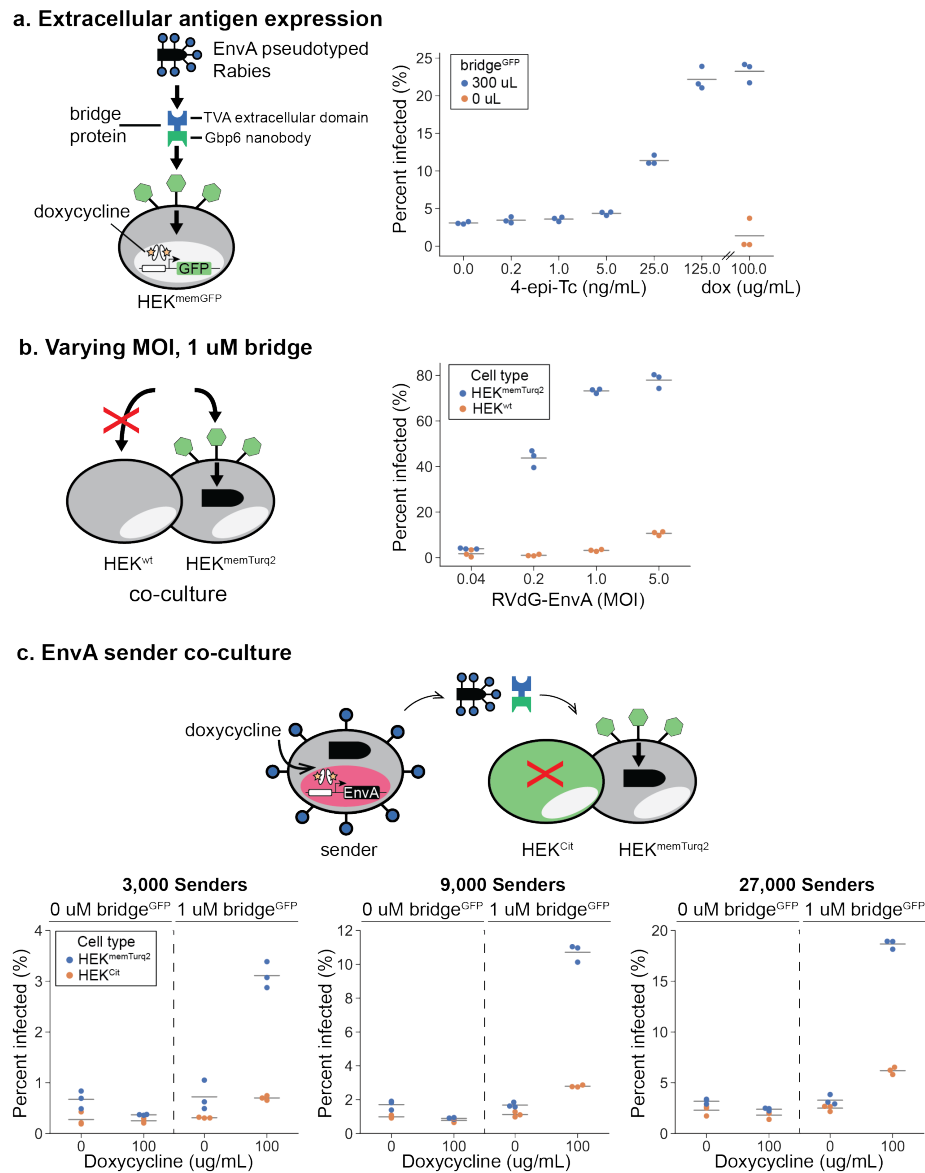
Figure 4.3 | **Pseudotyping and bridge proteins control viral entry.** (A) Left, schematic of the bivalent bridge protein directing EnvA-pseudotyped rabies (RVdG-EnvA) to infect GFP-displaying target cells (HEKmemGFP). Recombinantly expressed bridge protein contains the TVA receptor extracellular domain and Gbp6 nanobody (bridgeGFP). Right, percent of HEKmemGFP infected under increasing concentrations of 4-epi-Tc using 1 MOI RVdG-EnvA and 300 uL of bridgeGFP-conditioned media. (B) Constant bridgeGFP and varying concentrations of RVdG-EnvA administered to a mixed population of HEKwt and HEKmemTurq2 preferentially infected HEKmemTurq2 cells. (C) Top, co-culture of doxycycline-inducible EnvA sender cells co-expressing IFP with target cells HEKmemGFP and non-target cells HEKCit. Bottom, percent of target and non-target cells infected in the presence or absence of doxycycline and bridge protein under varying sender numbers.

of sender cells with or without the bridge protein. As expected, infection strongly depended on the level of induction of the EnvA senders, the presence of bridgeGFP, and the expression of surface antigen, making infection simultaneously dependent on all three factors (Fig. 4.3c).

Viral replication is controlled by an intracellular protein. The ability to control or condition viral replication on intracellular proteins and external inducers would allow more cell type specific control of viral infection. We therefore sought to engineer a replication control system that could be sensitive to the presence of specific intracellular proteins, such as cell type markers or proteins expressed in specific states (e.g. active proliferation), and external small molecule inducers. To connect protein sensing to viral replication in a modular fashion, one needs to design a protein whose activity is required for viral replication but dependent on the presence of an inducer and/or intracellular target protein.

To achieve this, we identified essential viral proteins that can be regulated by the attachment of conditional degradation domains (degrons). The DHFR degradation domain (degron) destabilizes attached proteins, but can be inhibited by the trimethoprim (TMP) (Iwamoto et al., 2010). To identify sensitive sites, we fused DHFR with each essential protein and screened for locations that permitted viral replication in the presence of TMP. The C-terminus of the viral P protein proved to be an ideal site for regulation. It was sensitive to DHFR incorporation, but this effect could be blocked by TMP. We introduced a TEV cleavage site between the P protein and the degron (P-DHFR), so that the virus (RVdG-P-DHFR) can be recovered from a cell line stably expressing the TEV protease without the need for TMP (Supplementary Fig. 4.8a).

To enable regulation by endogenous proteins, we took advantage of an engineered unstable nanobody that is stabilized by binding to its target antigen (Tang et al., 2016). We replaced the DHFR in RVdG-P-DHFR with the destabilized GFP nanobody, denoted GBP (RVdG-P-GBP, Fig. 4.4a, 4b top). To test whether viral replication was indeed conditional on expression of intracellular GFP, or more precisely its yellow fluorescent variant, Citrine, we used RVdG-P-GBP to infect a co-culture of Citrine-expressing HEKCit cells and non-expressing parental HEKwt cells (Fig. 4.4b, bottom left). As measured by mCherry expression, the citrine-positive cells were infected at high rates (Fig. 4.4a, bottom right), while infection levels in citrine-negative cells were indistinguishable from background (Fig. 4.4a, bottom right). This result indicates that rabies virus replication can be made depen-
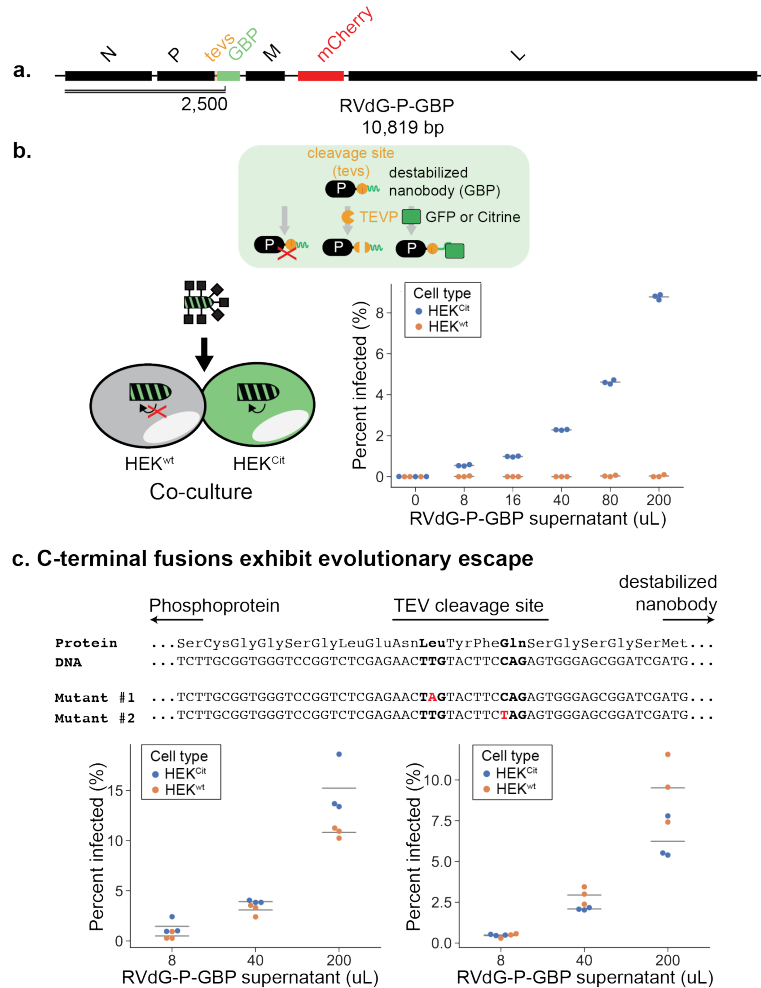
Figure 4.4 | **Viral replication is controlled by an intracellular protein..** (a) Recombinant rabies genome, RVdG-P-GBP. (b) Tagging the phosphoprotein (P) with a degron controls viral replication. Top, Schematic of a P protein (black rectangle) tagged with a destabilized nanobody, GBP (green wavy line), and an intervening TEV protease cleavage site (orange circle). P is stabilized when GFP is removed by TEV protease or stabilized when bound to GFP or Citrine, thus permitting viral replication. Bottom left, in a mixed population of HEKwt and HEK293 with cytoplasmic expression of Citrine (HEKCit, see Methods), RVdG-P-GBP will preferentially replicate in HEKCit. Right, increasing concentration of RVdG-P-GBP infects HEKCit, but not HEKwt. (c) Top, the P-GBP design exhibits evolutionary escape. Sanger sequencing of the junction region between P and GBP of two escape mutants identified two distinct nonsense mutations. Red indicates single nucleotide mutations and bold indicates mutated codons. Bottom, infection of a mixed population of HEKwt and HEKCit using mutant RVdG-PGBP showed diminished discrimination.

dent on the presence of an unrelated intracellular protein (GFP).

Despite the success of the P-degron strategy, conditionality was diminished after 1 month of continuous viral passaging in the producer cells (Fig. 4.4c, bottom, Supplementary Table 2, see Methods). Given the error rate of rabies RNA-dependent RNA polymerase34, we reasoned that this loss of function could arise from mutations that rescue P function in the absence of the target protein. Indeed, when we sequenced the P-degron junction in the RVdG-P-GBP viral genome, we consistently found nonsense mutations that truncate the protein before the degron is translated (Fig. 4.4c, top, Supplementary Fig. 4.8c), bypassing regulation. Similar "escape mutants" have been independently reported in other work (Matsuyama et al., 2019), indicating that engineering evolutionary robustness is essential.

An external drug inhibits viral replication and permits viral removal.

To reduce selection pressure for escape mutants and improve evolutionary robustness in our intended use case (drug-controlled reversal of established infection) we inserted an HCV protease (HCVP) flanked by its cognate cleavage sites (Taremi et al., 1998; Lin, Glenn, and Tsien, 2008; Tague et al., 2018; Jacobs, Badiee, and Lin, 2018) between P and L (RVdG-P-HCVP-L, Fig. 4.5a). In this design, HCVP normally acts to cleave both sites, separating the P and L proteins and allowing them to function (Fig. 4.5b, left). Critically, however, the drug asunaprevir (ASV) blocks HCV protease activity (Chung, Jacobs, et al., 2015), leaving P and L tethered unproductively together, disrupting viral replication (Fig. 4.5b, left). This strategy disfavors selection for escaper mutants in two ways. First, contrary to the previous strategy, here ASV is absent when the virus is being passaged or infection is being established (see the "curing" experiment below), relieving the selection pressure for HCVP to lose sensitivity to ASV inhibition. Second, point mutations or deletions, the most common modes of mutations in rabies virus, should not permit the transcription and translation of L as a protein separate from P. We passaged the RVdG-P-HCVP-L virus in producer cell lines for 8 months in the absence of ASV, periodically assaying viral sensitivity to ASV. This long term passaging did not detectably diminish ASV sensitivity (Fig. 4.5c), suggesting that this "cut-out" design can curb the emergence of escape mutants and maintain pharmaceutical regulation of the viral vector over a substantially more extended period of time than the previous design.

So far, we have focused on controlling the establishment of infection. However, for many applications, it will be crucial to terminate a productive infection after the viral vector and the encoded circuit perform their function. Therefore, we tested the
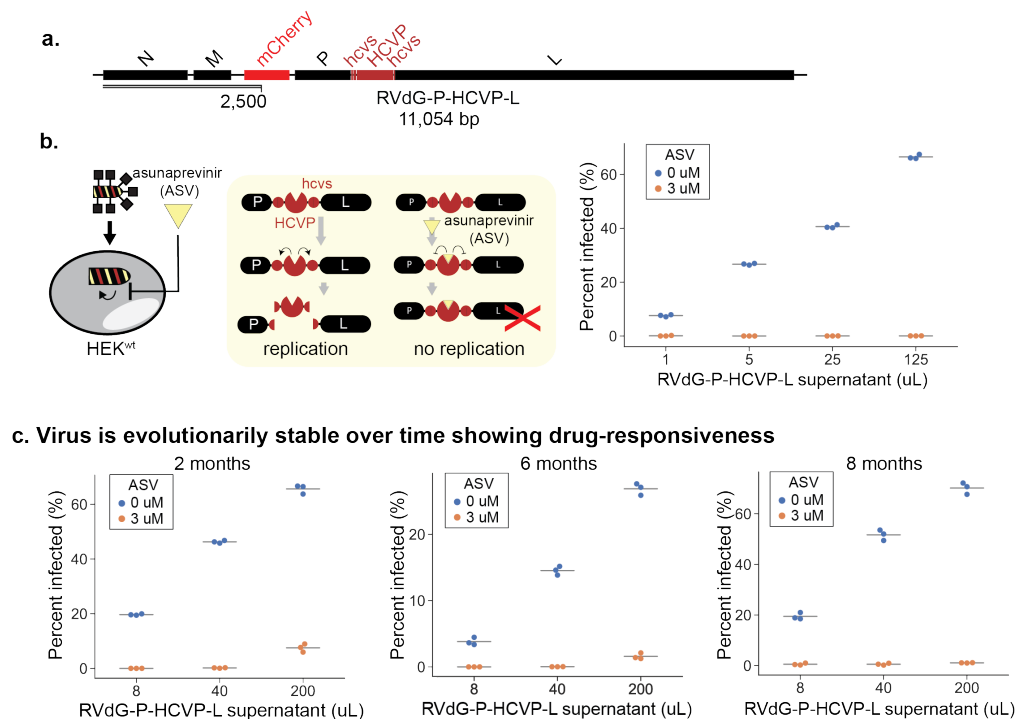
Figure 4.5 | **An external drug inhibits viral replication and permits viral removal..** (a) Recombinant rabies genome sensitive to drug inhibition, RVdG-P-HCVP-L. (b) Asunaprevir (ASV) inhibits viral replication. Left, a Hepatitis C Virus protease (HCVP, red pac-man) flanked by its cleavage sites (hcvs, red circles) is inserted between the P and L proteins to create a P-HCVP-L fusion. HCVP cleavage of the flanking cleavage sites will separate P and L to permit viral replication. Addition of ASV, an HCVP inhibitor, will result in non-functional P-HCVP-L fusion proteins and inhibit RVdG-P-HCVP-L viral replication. (c) RVdG-P-HCVP-L exhibits evolutionary stability. RVdG-P-HCVP-L collected from 2, 6, and 8 months of continuous passage were tested on HEKwt. Regulation by ASV was maintained.

ability of ASV to "cure" an established infection by the engineered virus. We used time-lapse microscopy to follow the dynamics of infection in the same population of cells over time before and after ASV addition (Fig. 4.6a).

We first incubated HEK293 cells that express a constitutive H2B-Citrine (HEKH2B-Citrine) with viral particles, and quantified mCherry signal as a surrogate for the level of infection using two similarly behaving metrics (Fig. 4.6a, see Methods for details). After one day, we observed increased mCherry signal (Fig. 4.6b-d), consistent with viral infection and replication. This signal remained high for more than 13 days (Fig. 4.6b, Supplementary Movie 1, left). In a parallel experiment, we added ASV to the media after one day (Fig. 4.6c). This treatment led to a decay in the mCherry signal over the following 7 days, until no signal could be detected

above background (Fig. 4.6c, Supplementary Movie 1, center). Thus, ASV addition successfully terminated viral protein expression.

Nevertheless, in principle low levels of virus could remain after ASV treatment and resume replication after ASV removal. We therefore asked whether transient addition of ASV could permanently cure an established infection of the engineered virus. First, in a preliminary experiment, we treated infected cells with two concentrations of ASV for 6.2 days (see Methods) and then removed the drug for 3 days. We observed no viral re-emergence in all but one field of view. This result suggested the potential for further improvement through repeated and prolonged ASV dosing (Supplementary Fig. 4.10a and 4b, Supplementary Movie 2). We therefore extended the period of ASV exposure to 6.9 days, during which we carried out daily media changes. We then washed out the drug, and continued to monitor the culture for 5 more days (Fig. 4.6d, Supplementary Movie 1, right), which is more than sufficient time for intracellular amplification. We observed no re-emergence of infection during this time period across all ten fields of cells (Fig. 4.6d). These results suggest that infections of the engineered virus could be successfully reversed by transient ASV addition in the overwhelming majority of cells.

Because they remain outside the nucleus and thereby avoid the potential risk for insertional mutagenesis, RNA viruses are attractive candidates for future synthetic circuit delivery vectors. In fact, while DNA vectors currently remain more prevalent in therapeutic applications, RNA viruses have received growing attention40, including from synthetic biologists seeking to improve their specificity (Wagner et al., 2018; Wroblewska et al., 2015; Takahashi and Yokobayashi, 2019). Here we sought to address key challenges required to make RNA viruses into a more engineerable, and safer, alternative to DNA vectors. Rabies virus, with its extensive history of engineering (Mebatsion et al., 1996) and applications in neuroscience (Wickersham, Finke, et al., 2007; Wickersham, Lyon, et al., 2007), provides an ideal model system. As a step towards increasing the engineerabiity of rabies virus, this work had the dual purposes of integrating well-known mechanisms such as pseudotyping into a single platform, while also designing new mechanisms to address outstanding challenges such as evolutionarily stable drug control of replication. The modules analyzed here are for the most part not rabies-specific and likely to be transferable to other RNA viruses. We can now achieve multiple levels of specificity and control in rabies virus (Fig. 4.1). These levels include controlling viral secretion from "sender" cells (Fig. 4.2, 4.3c); achieving selective infection of target cells based on
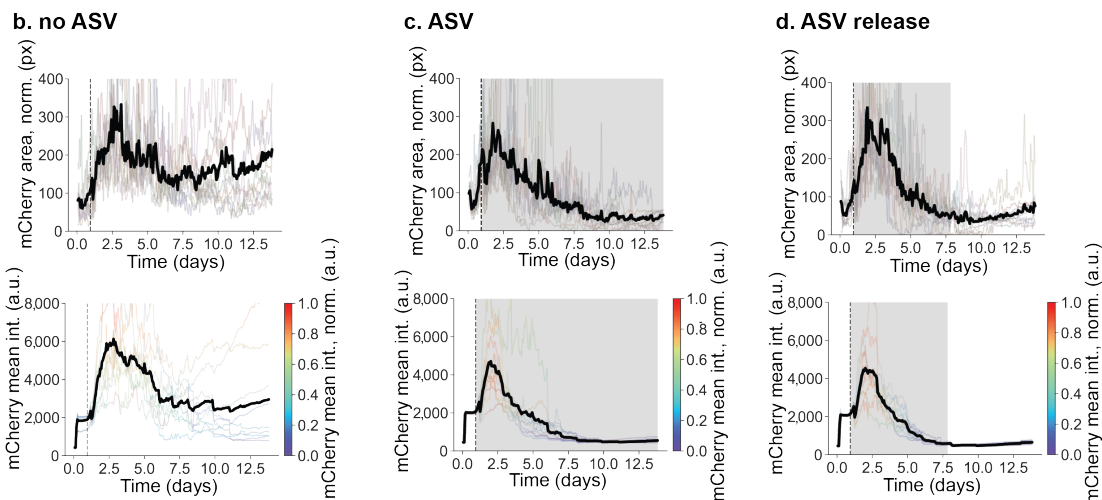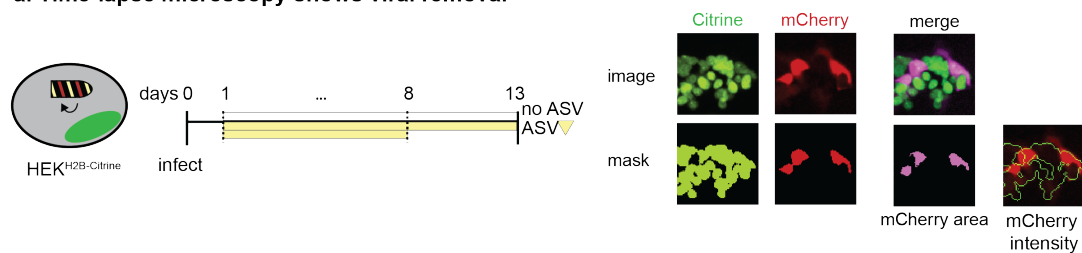
Figure 4.6 | **Time-lapse microscopy shows viral removal from established infections.** (a) Top: Timeline of infection with no ASV (top blank bar), continuous ASV (middle yellow bar), or ASV-then-release (bottom shorter yellow bar). Bottom: Example of image processing. A binary mask was created based on the signal in the Citrine channel. The mCherry area shows the overlay between the mCherry mask and Citrine mask. Note that not all mCherry+ areas are Citrine+, because the tagged H2B-Citrine is localized in the nucleus while mCherry doesn't have a localization tag. The mean intensity of mCherry signal in the Citrine mask was also quantified. HEK target cells expressing H2B-Citrine were infected for 24 hours with RVdG-P-HCVP-L (dotted line) and then cultured in media containing no ASV (b, no ASV), continuous ASV (c, ASV), or ASV for 6.9 days and then for 5 days in media containing no ASV (d, ASV release). Grey shading indicates the presence of ASV in media. Transparent and black lines respectively represent traces from individual movies and the mean of those traces. Fraction of infected cells as indicated by "mCherry area, norm (px)" (b, c, d, left column) was calculated as the fraction of mCherry+ pixels within the H2B-Citrine+ mask. Individual traces for mCherry intensity are color-scaled from purple to red, where mCherry intensity values are normalized by the maximum mCherry intensity within each trace. The slight increases at late times in the mean mCherry intensity traces represent cellular autofluorescence. In order to display all mean traces on the same y-scale, the tops of some individual traces are cut off. For complete traces, see Supplementary Fig. 4.9.

surface proteins, with external control through bispecific bridge proteins (Fig. 4.3); implementing conditional replication in target cells based on an intracellular protein (Fig. 4); and designing evolutionarily robust control of viral replication with the drug ASV (Fig. 4.5, Fig. 4.6). Because RNA viruses have elevated mutation rates compared to many DNA viruses (Matsuyama et al., 2019; J. Holland et al., 1982), this last feature addresses a critical hurdle that will be necessary for RNA viral vectors to compete with DNA vectors in biomedical applications.

As a tool for basic research in neurobiology, the engineered rabies vectors introduced here could be immediately applicable for tracing synaptic connections with better control and higher specificity. Upon further optimization, these vectors might help realize a central biomedical promise of synthetic biology: they could be delivered in sender cells that naturally home to a disease tissue (J. Holland et al., 1982), exit sender cells only in the correct microenvironment, selectively enter target cells that display certain surface markers, replicate specifically in cells positive for specific markers and pathway activities, deliver genetic cargo to change cellular behavior as a therapy, and finally be eliminated with a small-molecule drug. The self-replication of such vectors guarantees that their cargo will be highly expressed so as to effectively perform their functions, and the RNA nature of their genome and the ability to eliminate the virus from infected cells minimize the risk of a synthetic circuit leaving "scars" in the host genome or intracellular signaling pathways. While many challenges to this vision undoubtedly remain, the results here provide proof of principle demonstration of core capabilities.

Beyond the artificial signals such as dox and GFP that control the current version of rabies vectors, one next step is to apply a similar design principle to engineering vectors whose replication is conditioned upon the activity of endogenous pathways and antigens, starting with those that drive oncogenesis (Hanahan and R A Weinberg, 2000; Douglas Hanahan and Robert A Weinberg, 2011). For example, a virus that could conditionally replicate in cells with elevated Myc or Ras activity could provide enhanced selectivity for tumor cells.

In addition, to realize the potential of RNA viral vectors, we will still need to address several challenges. First, overly high vector replication and cargo expression in host cells after delivery could generate toxicity by competing for cellular resources, leading to detrimental effects on the patient. To address this, synthetic negative feedback circuits could be added to maintain vector/cargo levels below a tolerance threshold. Second, although rabies virus has evolved to counter detection and clearance by the

immune system (Schnell et al., 2010), it, like most other RNA viruses, is nevertheless immunogenic. One solution would be to encode additional immunomodulating molecules in the vector, so that innate immunity can be temporarily suppressed while the vector and its cargo carry out their function. Third, although rabies vectors with inserts up to 3.7 kb long have been successfully produced and used for in vivo tracing (Osakada et al., 2011), it will be critical to systematically determine how delivery efficiency varies with the size and sequence identity of genetic cargo. If size limits prove to be inadequate for common applications, one could seek to identify related viruses with larger capacity or develop strategies to split cargo among two or more viruses, each of whose replication depends on the other.

Future optimization of RNA viral vectors would benefit from a plethora of novel methods to regulate RNAs and proteins, including the use of synthetic riboswitches (Takahashi and Yokobayashi, 2019). These methods would provide additional means for controlling vector replication. On the other hand, the ability to bypass transcriptional control and host genome integration means that synthetic circuits made of RNAs and proteins are ideal cargo for RNA vectors, especially protein circuits that can be compactly encoded on a single transcript. We anticipate that engineering rabies viral vectors and their cargo could drive the development of other homologous engineered RNA viruses, each adapted to specific applications.

References

Benenson, Yaakov (Aug. 2012). "Synthetic biology with RNA: progress report". en. In: *Curr. Opin. Chem. Biol.* 16.3-4, pp. 278–284.

Boerger, A L, S Snitkovsky, and J A Young (Aug. 1999). "Retroviral vectors preloaded with a viral receptor-ligand bridge protein are targeted to specific cell types". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 96.17, pp. 9867–9872.

Callaway, Edward M and Liqun Luo (June 2015). "Monosynaptic Circuit Tracing with Glycoprotein-Deleted Rabies Viruses". en. In: *J. Neurosci.* 35.24, pp. 8979–8985.

Cella, Federica et al. (Oct. 2018). "Engineering protein-protein devices for multilayered regulation of mRNA translation using orthogonal proteases in mammalian cells". en. In: *Nat. Commun.* 9.1, p. 4392.

Choi, Jiwon, John A T Young, and Edward M Callaway (Sept. 2010). "Selective viral vector transduction of ErbB4 expressing cortical interneurons in vivo with a viral receptor-ligand bridge protein". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 107.38, pp. 16703–16708.

Chung, Hokyung K, Conor L Jacobs, et al. (Sept. 2015). "Tunable and reversible drug control of protein production via a self-excising degron". en. In: *Nat. Chem. Biol.* 11.9, pp. 713–720.

Chung, Hokyung K, Xinzhi Zou, et al. (May 2019). "A compact synthetic pathway rewires cancer signaling to therapeutic effector release". en. In: *Science* 364.6439.

Ciabatti, Ernesto et al. (July 2017). "Life-Long Genetic and Functional Access to Neural Circuits Using Self-Inactivating Rabies Virus". en. In: *Cell* 170.2, 382–392.e14.

Daringer, Nichole M et al. (2014). "Modular extracellular sensor architecture for engineering mammalian cell-based devices". In: *ACS Synth. Biol.* 3.12, pp. 892–902.

Drake, J W and J J Holland (Nov. 1999). "Mutation rates among RNA viruses". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 96.24, pp. 13910–13913.

Felt, Sébastien A and Valery Z Grdzelishvili (Dec. 2017). "Recent advances in vesicular stomatitis virus-based oncolytic virotherapy: a 5-year update". en. In: *J. Gen. Virol.* 98.12, pp. 2895–2911.

Fink, Tina et al. (Dec. 2018). "Design of fast proteolysis-based signaling and logic circuits in mammalian cells". en. In: *Nat. Chem. Biol.*

Fischbach, Michael A, Jeffrey A Bluestone, and Wendell A Lim (Apr. 2013). "Cell-based therapeutics: the next pillar of medicine". en. In: *Sci. Transl. Med.* 5.179, 179ps7.

Gao XJ, Chong LS, Kim MS, Elowitz MB (2018). "Programmable protein circuits in living cells". In: *Science*.

Hanahan, D and R A Weinberg (Jan. 2000). "The hallmarks of cancer". en. In: *Cell* 100.1, pp. 57–70.

Hanahan, Douglas and Robert A Weinberg (Mar. 2011). "Hallmarks of cancer: the next generation". en. In: *Cell* 144.5, pp. 646–674.

Holland, John J (Dec. 2012). *Genetic Diversity of RNA Viruses*. en. Springer Science & Business Media.

Holland, J et al. (Mar. 1982). "Rapid evolution of RNA genomes". en. In: *Science* 215.4540, pp. 1577–1585.

Iwamoto, Mari et al. (Sept. 2010). "A general chemical method to regulate protein stability in the mammalian central nervous system". en. In: *Chem. Biol.* 17.9, pp. 981–988.

Jacobs, Conor L, Ryan K Badiee, and Michael Z Lin (July 2018). "StaPLs: versatile genetically encoded modules for engineering drug-inducible proteins". en. In: *Nat. Methods* 15.7, pp. 523–526.

Keung, Albert J et al. (Mar. 2015). "Chromatin regulation at the frontier of synthetic biology". en. In: *Nat. Rev. Genet.* 16.3, pp. 159–171.

Kirchhofer, Axel et al. (Jan. 2010). "Modulation of protein properties in living cells using nanobodies". en. In: *Nat. Struct. Mol. Biol.* 17.1, pp. 133–138.

Kitada, Tasuku et al. (Feb. 2018). "Programming gene and engineered-cell therapies with synthetic biology". en. In: *Science* 359.6376.

Kotterman, Melissa A, Thomas W Chalberg, and David V Schaffer (2015). "Viral Vectors for Gene Therapy: Translational and Clinical Outlook". en. In: *Annu. Rev. Biomed. Eng.* 17, pp. 63–89.

Lin, Michael Z, Jeffrey S Glenn, and Roger Y Tsien (June 2008). "A drug-controllable tag for visualizing newly synthesized proteins in cells and whole animals". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 105.22, pp. 7744–7749.

Matsuyama, M et al. (2019). "" Self-inactivating" rabies viruses are just first-generation, ΔG rabies viruses". In: *bioRxiv*.

Mebatsion, T et al. (July 1996). "Highly stable expression of a foreign gene from rabies virus vectors". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 93.14, pp. 7310–7314.

Morsut, Leonardo et al. (Feb. 2016). "Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors". en. In: *Cell* 164.4, pp. 780–791.

Nakanishi, Mahito and Makoto Otsu (Oct. 2012). "Development of Sendai virus vectors and their potential applications in gene therapy and regenerative medicine". en. In: *Curr. Gene Ther.* 12.5, pp. 410–416.

Osakada, Fumitaka et al. (Aug. 2011). "New rabies virus variants for monitoring and manipulating activity and gene expression in defined neural circuits". en. In: *Neuron* 71.4, pp. 617–631.

Schnell, Matthias J et al. (Jan. 2010). "The cell biology of rabies virus: using stealth to reach the brain". en. In: *Nat. Rev. Microbiol.* 8.1, pp. 51–61.

Siciliano, Velia et al. (May 2018). "Engineering modular intracellular protein sensor-actuator devices". en. In: *Nat. Commun.* 9.1, p. 1881.

Snitkovsky, S, T M Niederman, et al. (Oct. 2000). "A TVA-single-chain antibody fusion protein mediates specific targeting of a subgroup A avian leukosis virus vector to cells expressing a tumor-specific form of epidermal growth factor receptor". en. In: *J. Virol.* 74.20, pp. 9540–9545.

Snitkovsky, S and J A Young (June 1998). "Cell-specific viral targeting mediated by a soluble retroviral receptor-ligand fusion protein". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 95.12, pp. 7063–7068.

Stanford, Marianne M et al. (Feb. 2008). "Innate immunity, tumor microenvironment and oncolytic virus therapy: friends or foes?" en. In: *Curr. Opin. Mol. Ther.* 10.1, pp. 32–37.

Tague, Elliot P et al. (July 2018). "Chemogenetic control of gene expression and cell signaling with antiviral drugs". en. In: *Nat. Methods* 15.7, pp. 519–522.

Takahashi, Kei and Yohei Yokobayashi (Sept. 2019). "Reversible Gene Regulation in Mammalian Cells Using Riboswitch-Engineered Vesicular Stomatitis Virus Vector". en. In: *ACS Synth. Biol.* 8.9, pp. 1976–1982.

Tang, Jonathan Cy et al. (May 2016). "Detection and manipulation of live antigen-expressing cells using conditionally stable nanobodies". en. In: *Elife* 5.

Taremi, S S et al. (1998). "Construction, expression, and characterization of a novel fully activated recombinant single-chain hepatitis C virus protease". In: *Proteins*.

Thomas, Clare E, Anja Ehrhardt, and Mark A Kay (May 2003). "Progress and problems with the use of viral vectors for gene therapy". en. In: *Nat. Rev. Genet.* 4.5, pp. 346–358.

Wagner, Tyler E et al. (Nov. 2018). "Small-molecule-based regulation of RNA-delivered circuits in mammalian cells". In: *Nat. Chem. Biol.* 14.11, pp. 1043–1050.

Weber, Wilfried and Martin Fussenegger (Nov. 2011). "Emerging biomedical applications of synthetic biology". en. In: *Nat. Rev. Genet.* 13.1, pp. 21–35.

Wickersham, Ian R, Stefan Finke, et al. (Jan. 2007). "Retrograde neuronal tracing with a deletion-mutant rabies virus". en. In: *Nat. Methods* 4.1, pp. 47–49.

Wickersham, Ian R, David C Lyon, et al. (Mar. 2007). "Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons". en. In: *Neuron* 53.5, pp. 639–647.

Wroblewska, Liliana et al. (Aug. 2015). "Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery". en. In: *Nat. Biotechnol.* 33.8, pp. 839–841.

Xie, Zhen et al. (Sept. 2011). "Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells". en. In: *Science* 333.6047, pp. 1307–1311.

Young, J A, P Bates, and H E Varmus (Apr. 1993). "Isolation of a chicken gene that confers susceptibility to infection by subgroup A avian leukosis and sarcoma viruses". en. In: *J. Virol.* 67.4, pp. 1811–1816.

4.3 Materials and Methods

Plasmid construction

All constructs were generated using standard procedures. The backbones were linearized using restriction digestion or PCR, and inserts were generated using PCR or gBlock synthesis (IDT). A list of all plasmids reported in this manuscript is included in Supplementary Table 1, and all sequences were deposited to Addgene.

Tissue culture

Flp-InTM T-RExTM 293 Cell Line (Human Embryonic Kidney cells that contain a single stably integrated FRT site at a transcriptionally active genomic locus, and stably expressing the tetracycline repressor protein) were purchased from Thermo Fisher Scientific (R78007). B7GG cell line (Baby Hamster Kidney cells that contain a stably integrated T7 RNA polymerase and rabies glycoprotein) and HEK-TVA cell line (Human Embryonic Kidney cells that contain stably integrated TVA receptor) were kindly gifted from Dr. Lindsay Schwarz at St. Jude Children's Research Hospital. HEK293 derived cells were cultured in a humidity controlled chamber at 37°C with 5% CO_2 in media containing DMEM (Thermo Fisher Scientific, cat no. 11960-069) supplemented with 10% FBS (VWR, 76308-946), 1 mM sodium pyruvate (Thermo Fisher Scientific, cat no. 11360-070), 1 unit/ml penicillin, 1 g/ml streptomycin, 2 mM L-glutamine (Thermo Fisher Scientific, cat no. 10378-016) and 1X MEM non-essential amino acids (Thermo Fisher Scientific, cat no. 11140-050) (293 media). B7GG derived cells producing rabies virus were cultured in a humidity controlled chamber at 37°C with 5% CO_2 in media containing DMEM supplemented with 2.5-5%FBS, 1 mM sodium pyruvate, 1 unit/ml penicillin, 1 g/ml streptomycin, 2 mM L-glutamine and 1X MEM non-essential amino acids (BHK media). 100 ng/mL doxycycline was added whenever expression is needed from a CMV/TO promoter. Trimethoprim (TMP) was delivered at 1 M. Asunaprevinir (ASV) was delivered at 3 M.

Cell line construction

To generate cell lines with stably integrated transgenes antibiotic selection was performed. Flp-InTM T-RExTM 293 Cell Line or BHK21 cells were transfected in 24-well plates and transferred two days later into a 6-well plate containing selection media (Supplementary Table 2). After PiggyBac-based integration, monoclonal cell populations were established through limiting dilution, and preliminary screening was performed to identify clones with highest transgene expression using flow cytometry (Supplementary Table 2). Transgenesis using pOG44 Flp-recombinase into Flp-InTM T-RExTM 293 Cell Line resulted in singly-integrated cell lines. Cell lines will be available upon request.

Production of Packaging Cell Lines

B7GG was integrated with a transgene that encodes for constitutive co-expression

of TEVP and Cerulean (B7GG-TCer). This packaging line was generated to permit efficient production of modified rabies genomes containing an essential protein tagged with a destabilizing domain, in which a TEVP cleavage site was inserted between the essential protein and the degron. TEVP cleavage of the degron stabilizes the essential protein allowing viral replication.

Production of Rabies Viruses

Viruses were initially established using the protocol as described in Osakada, Nat Pro. 2013 and scaled for 24-well plates. RVdG-P-DHFR and RVdG-P-GBP were produced in B7GG-TCer producer lines while RVdG and RVdG-P-HCVP-L viruses were produced in either cell line (Supplementary Table 2). Briefly, B7GG and B7GG-TCer cells were plated in 24-well plates at 0.05106 cells per well in 293 media.Cells were transfected with a DNA mixture containing 240 ng rabies genome, 120 ng pcDNA-SADB19N, 60 ng pcDNA-SADB19P, 60 ng pcDNA-SADB19L and 40 ng pcDNA-SADB19G. Media was changed one day after transfection to BHK media. Cells were transferred from a 24-well plate to a 6-well plate three days after transfection and maintained in BHK media. Fresh BHK media was added every day to facilitate viral spread. Viral spread was checked every day using fluorescent microscopy. Supernatant collection only proceeded when the virus had visibly infected the majority of the population.

Amplified rabies viruses were concentrated and purified as described in (Osakada, Nat Prot. 2013) to generate starter viral stocks. Briefly, infected producer lines were expanded to three 15 cm dishes and exchanged with fresh BHK media the following day. After three days, conditioned media was collected and filtered through at 0.45 um filter. The supernatant was concentrated in an ultracentrifuge at 70,000 g for 2 hr at 4 o C on a 20% sucrose pad. Viral pellets were resuspended in 1X Hank's Balanced Salt Solution (HBSS) and stored at -80 o C for future use.

Pseudotyped EnvA rabies virus was purchased from Janelia Farms.

Production of Bridge Protein Cell Lines

In vivo production of bridge proteins:

The extracellular domain of ASLV-A envelope protein from the plasmid pAAV-TRE-HTG (Addgene) and the targeting domain encoding the Gbp6 nanobody3 3 were combined by PCR and cloned into a Piggybac transfer vector under a synthetic

constitutive promoter (Supplementary Table 1). Flp-InTM T-RExTM 293 cells were stably integrated to create a polyclonal bridge protein secretion line. To collect produced bridge proteins, the stable line was seeded at a density of 0.1106 cells per well of a 6-well plate and cultured under standard conditions overnight. The media was exchanged the following day for BHK21 media and cells were cultured over two days with additional fresh media applied the second day. Conditioned media was collected and centrifuged at 13,000 rpm for 20 minutes to remove cellular debris. 300 uL of conditioned media was applied in each condition unless otherwise noted.

In vitro production of bridge proteins:

Bridge proteins were purified at the Caltech Protein Expression Center. Briefly, 6XHIS-tagged bridge proteins encoding plasmids (Supplementary Table 1) were transfected into Expi293FTM cells and conditioned media was harvested four days post-transfection. The supernatant was spun at 2,000 rpm for 10 minutes and proteins were purified with Ni-NTA affinity chromatography (Cytiva, 5 mL HisTrap). Expression and purification were checked by SDS-PAGE and Coomassie Blue staining. Protein concentration was quantified using PierceTM 660 nm (Thermo Fischer, cat no. 22662) assay.

Transient Transfection

293 cells were seeded at a density of 0.05106 cells per well of a 24-well plate and cultured under standard conditions overnight. The following day, the cells were transfected with plasmid constructs using Lipofectamine LTX (Thermo Fisher) as per manufacturer's protocol.

Infection with Rabies

Infected producer cell lines B7GG and B7GGTCer were plated to 60% confluency in 6-well plates. The following day, the media was changed to 2 mLs of fresh BHK media. Two days later an additional 1 mL of BHK media was added. Virus containing media was collected on the third day and supernatant was spun down for 5 minutes at 15,000 rpm to remove cellular debris. Supernatant from these freshly prepared stocks were used directly in RVdG-P-GBP (Fig. 4) and RVdG-P-HCVP-L (Fig. 4.5, Fig 6, Supplementary Fig. 4.9, Supplementary Fig. 4.10) experiments. Viral amount was determined by volume, as indicated on figures. Infection using purified RVdG-EnvA was performed by calculating MOI based on the reported

viral titer provided by Janelia Farms. For experiments containing sender cells (Fig. 4.2, Fig. 4.3c), sender cells were plated at 30% confluency and the following day media was exchanged with 1 mL BHK media and 1 mL of freshly prepared RVdG supernatant, as described above.

All experiments were conducted in 24-well plates. In co-culture experiments, 30,000-45,000 cells for each target population were counted and pre-mixed prior to plating. Specifically, for the co-culture performed in Fig. 4.4b, the expression of Citrine was induced with doxycycline overnight prior to plating and was present throughout the experiment (Supplementary Table 2, refer to HEKdoxCit). For single target cell population experiments, the cells were seeded at 75,000 cells per well. Infected sender cells, virus-containing supernatant, purified virus, bridge protein, and compounds were added to each well immediately after target cells were plated unless otherwise noted.

Flow cytometry

Flow cytometry samples were performed in biological triplicates. Three days after infection, cells were prepared for flow cytometry by trypsinizing with 40 uL of 0.05% trypsin for 1 min at room temperature. Protease activity was neutralized by resuspending the cells in buffer containing 100 uL of HBSS with 2.5mg/ml Bovine Serum Albumin (BSA) and 1 mM EDTA (Thermo Fischer Scientific, cat no. 15575020). Cells were then filtered through 40 m FalconT M Cell Strainers (Thermo Fischer Scientific, cat no. 08-771-1) and analyzed by flow cytometry (MACSQuant VYB, Miltenyi or CytoFLEX, Beckman Coulter). Matlab-based software for processing flow cytometry data was developed in-house by Yaron Antebi.

Fluorescent signal quantification from flow cytometric measurements Figures 4.2, 4.3, 4.4, 4.5 and Supplementary Figure 4.7 present quantitative analysis of flow cytometry measurements. For each sample in a comparison group (experiments performed in the same batch and data shown on the same plot), we reported the percent infected as identified by mCherry+ cells. Viable cell populations were determined as events that registered above 105 in both the FSC and SSC channels. A manual gate was drawn around a single contiguous population that excluded both cellular debris and diagonal streaking events.

Co-culture gating:

The fluorescence values of each cell line were determined in the constitutive fluorescent signal channel (Citrine or IFP in most cases). Each cell line was analyzed independently. For example, for a co-culture containing HEKw t and HEKC it, monoculture samples of HEKw t and HEKC it were each analyzed independently in the Citrine fluorescent channel. For analysis, we fit the fluorescence log distribution with skew Gaussian distributions, i.e. n normcdf(x,m,k) normpdf(x,m,s) in Matlab using non-linear least-square fitting, and reported the mode of the resulting fit. Here, the normcdf(x,m,k) and normpdf(x,m,s) functions are cumulative probability density and probability density functions for Gaussian distributions, respectively. The parameter n is a normalization factor, the parameter m is the mean of each Gaussian distribution, s is the standard deviation of the probability density function that parameterizes the width of the distribution, and k is the standard deviation of the cumulative probability density function that parameterizes the skewness of the distribution.

Percent infected calculation:

The mode of the resulting fit is used to gate different cell types from the co-cultured samples. For monoculture experiments, gating was not performed. For all cells within the gate in each sample, we fit a cubic spline and identified the local minima between the two peaks of mCherry+ and mCherry- cells. An average local minimum was calculated across all bimodal samples. We use this average local minima as a threshold between non-infected/weakly infected cells (mCherry-) and highly infected cells (mCherry+). A percent infected metric is calculated from the resulting histogram by calculating the area both above and below the threshold:

Percent infected = (Area above threshold) / (Total area) 100

Sequencing virus genomes P-DHFR and P-GBP

Three independent viral cultures were passaged for 2-4 months prior to genome sequencing. Reverse transcription of viral RNA was performed on samples using the Thermo Fisher Scientific protocol for SuperScript IV. PCR amplification of the genome using

forward primer (5'-ACCCTCCAGGAAAGTCTTC-3')

reverse primer (5'-AATAGGGTCATCATAGACCTCTC-3')

were gel purified and submitted for Sanger sequencing with Laragen Sequencing.

Time-lapse microscopy

For time-lapse imaging of rabies dilution (Fig. 4.6, Supplementary Fig. 4.9, Supplementary Fig. 4.10, Supplementary Movies 1 and 2) 5,000 H2B-Citrine were plated on 24-well glass- bottom plates (CellVis). Cells were induced with 100 ng/mL doxcyline overnight in normal culturing conditions. The following morning, the media was replaced with imaging media containing FluoroBrite DMEM (Thermo Fisher) supplemented with 10% FBS, 1 mM sodium pyruvate, 1 unit/ml penicillin, 1 g/ml streptomycin, 2 mM L-glutamine and 1X MEM non-essential amino acids and 100 ng/mL doxycycline. All time-lapse images were acquired on an inverted Olympus IX81 fluorescence microscope with Zero Drift Control (ZDC), an ASI 2000XY automated stage, iKon-M CCD camera (Andor, Belfast, NIR), and a 20X PAN-FL objective (1.42 NA). Fluorophores were excited with an X-Cite XLED1 light source (Lumen Dynamics). Cells were kept in an environmental stage-top chamber enclosing the microscope, with humidified 5% CO_2 flowing at 37°C (Okolab H301-K stage top with an O2- CO2 UNIT-BL mixer). Microscope and image acquisition were controlled by Metamorph software version 7.10 (Molecular Devices).

Imaging started approximately 2 hours after changing the media to fluorescent imaging media. B7GG lines producing virus were cultured for three days in imaging media. Conditioned media was collected and centrifuged at 13,000 rpm to pellet cellular debris. 250 uL of the centrifuged supernatant containing virus was added after approximately 2 hours of imaging to infect cells. The initial infection was established for 24 hours. To observe the removal of the rabies genome with prolonged ASV treatment, cells were washed with imaging media three times to remove residual virus and then incubated with imaging media containing 3 M of ASV. Media was changed every 24 hours. To confirm that the virus does not re-emerge after ASV removal, the ASV-containing imaging media was replaced with normal imaging media after 7 days.

Images were acquired every 90 minutes throughout the duration of the movie. Cells that were in the field of view before infection and remained alive and visible in the field of view without death for at least five days were used for initial data analysis through manual inspection.

In the preliminary experiment (Supplementary Fig. 4.10, Supplementary Movie 2), cells were infected for 17 hours and media was changed every 3 days over 6 days. Images were acquired every 60 minutes. For analysis, a similar selection criteria was performed in which cells that were infected and remained within the field of

view for five days were used for data analysis. Additionally, we excluded from analysis fields of view that were out of focus, near the edge of the well, or exhibited clumping or apoptosis within the last day of imaging.

Movie analysis

Data processing: mCherry mean intensity values were calculated based on total levels of fluorescence in the mCherry fluorescent channel as identified by the position of cells in the Citrine fluorescent channel. To systematically identify the position of cells, total constitutive Citrine signal was used to segment each image. The resulting segmentation mask was used to calculate the number of mCherry+ pixels within each region. To capture the magnitude of rabies expression within infected cells, the mean intensity was calculated for each mCherry+ region and averaged for each timepoint. The fraction infected metric was calculated by identifying the mCherry+ areas within the Citrine+ areas.

We first estimated the fraction of cells infected by rabies virus. We used the constitutive Citrine signal to generate a mask for all cells, where a pixel is considered to belong to a cell if its Citrine intensity is >400, as determined by an Otsu threshold of the first Citrine containing image. Within the Citrine mask, pixels are considered rabies-positive when mCherry intensity is greater than a moving Otsu threshold to account for fluctuations in background signal. As such rabies-positive area divided by the area of the Citrine mask (mCherry area norm. in Fig. 4.6, Supplementary Fig. 4.9 and Supplementary Fig. 4.10) is reported as a proxy for the fraction of rabies-infected cells in each image. The constitutive Citrine signal is localized to the nucleus and may underestimate the rabies-positive area, especially towards the beginning of the imaging when rabies-encoded mCherry form puncta in the cytoplasm.

In addition to thresholding mCherry signal, we also estimated the quantitative level of rabies infection by calculating the total mCherry intensity within each Citrine positive region of each image and averaged across these regions (mCherry mean int. in Fig. 4.6, Supplementary Fig. 4.9 and Supplementary Fig. 4.10) as a proxy for the level of rabies infection in that image.

Movie generation: For visualization purposes, the timepoint at which maximum mCherry intensity occurs for each movie was determined and used to rescale all images between the 5th and 99.5th percentile intensity values.

Data and code availability

The datasets generated and analyzed and the computer code used during the current study are available at data.caltech.edu, DOI 10.22002/D1.1438. Flow cytometry data analysis software used for this study is available at https://antebilab.github.io/easyflow/.

All authors have seen and approved the manuscript. The manuscript has been submitted to bioRxiv for preprint. The manuscript has not been accepted or published elsewhere.
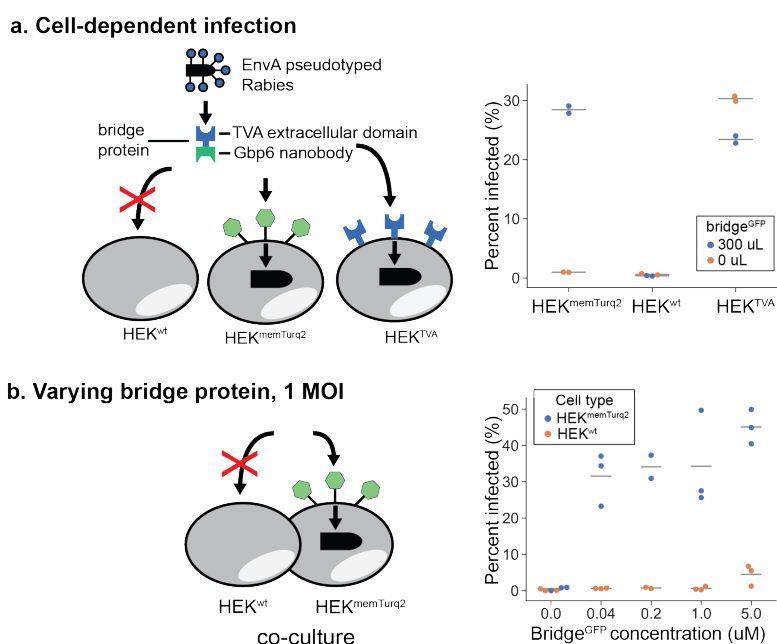
4.4 Supplementary Material

Figure 4.7 | **Pseudotyped rabies viral entry is dependent on presence of bridge and cell-surface marker.** (a) RVdG-EnvA virus infection of independently cultured HEKmemTurq2, HEKwt, and HEKTVA with or without bridgeGFP. (b) RVdG-EnvA and varied bridgeGFP concentrations infection of a co-culture of HEKmemTurq2 and HEKwt.
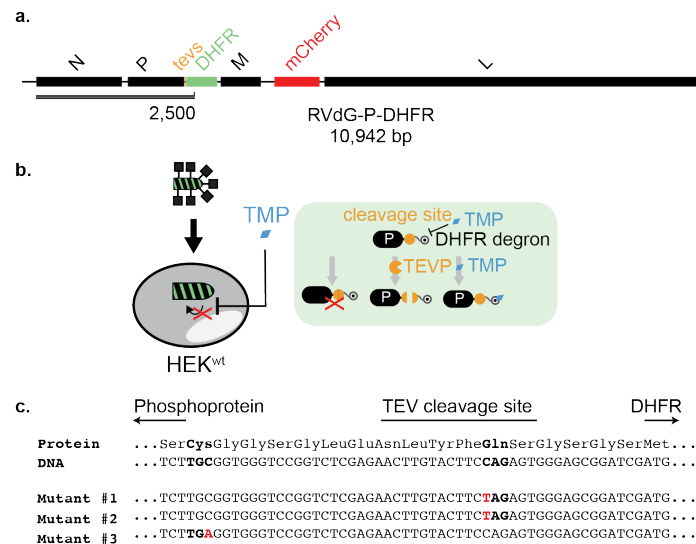
Figure 4.8 | **C-terminal tagging of Phosphoprotein exhibits evolutionary escape.**
(a) Recombinant rabies genome, RVdG-P-DHFR. (b) Schematic detailing phospho-protein regulation through C-terminal DHFR degron tagging. The phosphoprotein is stabilized by removal of the DHFR degron either through TEVP cleavage (orange pac-man) of the corresponding cleavage site (orange circle). Trimethroprim (TMP, blue diamond) inhibits the degron and stabilizes the reporter. (c) Sequencing of three escape mutants identified two distinct missense mutations. Coding sequences and corresponding amino acid sequences of Phosphoprotein, TEVP cleavage site, and DHFR are indicated above. Red indicates single nucleotide mutations and bold indicates mutated codons.
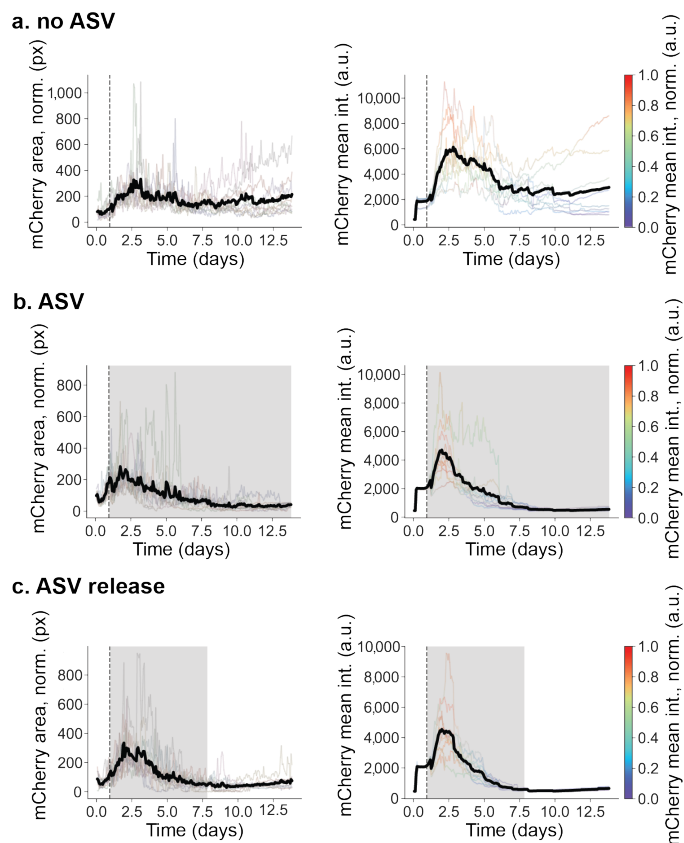
a. no ASV

b. ASV

c. ASV release

Figure 4.9 | **Time-lapse microscopy shows viral removal of non-infected cells.**
Plots shown from Figure 5d with full y-axis. HEK target cells expressing H2B-
Citrine were infected for 24 hours with RVdG-P-HCVP-L (dotted line) and then
cultured in media containing no ASV (b, no ASV), continuous ASV (c, ASV), or
ASV for 6.9 days and then for 5 days in media containing no ASV (d, ASV release).
Grey shading indicates the presence of ASV in media. Transparent and black lines
respectively represent traces from individual movies and the mean of those traces.
Fraction of infected cells as indicated by "mCherry area, norm (px)" (b, c, d, left
column) was calculated as the fraction of mCherry+ pixels within the H2B-Citrine+
mask. Individual traces for mCherry intensity are color-scaled from purple to red,
where mCherry intensity values are normalized by the maximum mCherry intensity
within each trace. The slight increases at late times in the mean mCherry intensity
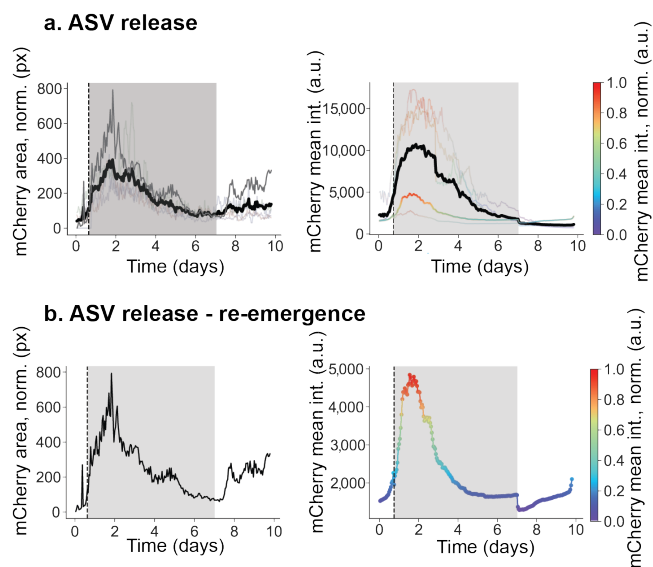traces represent cellular autofluorescence.

Figure 4.10 | **Time-lapse microscopy shows viral reemergence in a preliminary experiment.** (a) HEK target cells expressing H2B-Citrine were infected for one day with RVdG-P-HCVP-L (dotted line) and then cultured in media containing ASV for 6.2 days with media changed every three days and then for 3 days in media containing no ASV. Grey shading indicates the presence of ASV in media. Transparent and black lines respectively represent traces from individual movies and the mean of those traces. Opaque lines indicate viral re-emergence, shown independently in (b).

Chapter 5

CONCLUSION: THE CIRCUIT AS THE THERAPEUTIC

5.1 Key challenges remain to realize this paradigm – a roadmap for the future

Picking the right toolkit will be essential for success. The vast catalogue of system components and the lack of a fully generalized system can be overwhelming. How do synthetic biologists standardize parts so circuits can reliably be 'made to order'? Most efforts focus on building and expanding a small library of biological parts that function reliably with each other and use similar input/output biological activities. Characterization of these components are often limited in scope. Crucially, hybrid circuits which use varying biological components should be developed to integrate across biological processes. For example, combining a DNA sensor, like Cas9, with protease-based signal processing modules would permit a DNA-to-protein output.

Sense-process-respond components can be more sophisticated and diverse. Although sense-process-respond circuits hold the promise of wide-ranging applications in biomedicine and therapeutics, greater flexibility and sophistication in circuit design will be achieved by developing a more extensive suite of components which we exiguously highlight here:

- **Sensing challenges:** To fully recapitulate the systems of a living cell, we will need composable components that incorporate post-translational modifications. Protein modules that enable control through phosphorylation and epigenetic regulation would enable reversible signal processing, signal restoration, signal propagation, stabilization of gene expression programs, and oligomerization.

- **Processing challenges:** The progression of genetic engineering has far outpaced the incorporation of control theory in synthetic biology. The theoretical systems and control theory field can contribute to closed-loop circuit designs in mammalian cells to hedge trade-offs in robustness and fragility for real-world applications. New tools will also be required to design 'cooperative control' at the population level to enable therapies based on cellular consortia. Additionally, developing models that take advantage of eukaryotic subcellu-

lar compartments would permit circuits to run in parallel and allow for an additional control module.

- **Effector / response challenges:** To easily gauge circuit functionality, synthetic biologists have typically used detectable outputs like fluorescent proteins. While some effectors simply require gene expression and could be placed under transcriptional control, post-translational effector development poses a substantial engineering hurdle. Computational tools may ease ease the development of post-translational proteins.

Standardization of biological parts and computational approaches will ease designing circuits. Behavior of circuits designed for one environment may not exactly replicate to another environment. In particular, therapeutic circuits will need to be optimized when transitioning between the testing environment to its operational context. Here we advocate for a two-pronged approach: (1) standardization and characterization of existing biological parts for mammalian cells and (2) implementation of modeling and computational approaches to analyze and predict biological designs. A hardware predictive circuit designing tool was developed for the ease of genetic circuit engineering standardized for the bacterial cell context (Nielsen et al., 2016). A similar effort should be made for mammalian cells. Similarly, mathematical models are typically used to describe the behavior of a circuit within a system of interest (Chung et al., 2019; Okamoto and Savageau, 1984; Torres and Santos, 2015). With further biochemical characterization of these components to form a base dataset combined with the predictive power of mathematical models and machine-learning, it may be feasible to systematically analyze current biological designs and predict new ones.

Physiological limitations to synthetic recapitulations of pathways Are there principles of natural regulatory systems (such as using organ growth control for regulation) that must be respected to replicate physiological regulation? Or, are there other designs for these systems that could be implemented in completely different ways? Synthetic biology, in combination with **systems medicine** (Kirschner, 2016), may explain why particular regulatory systems use the designs they do. It will be important to understand the limitations of natural regulatory functions as we try to design synthetic replacements.

The complexity of circuits will ease manufacturing and processing. The cost in developing a new pharmacological drug was estimated to be approximately >$ 2 billion dollars and takes, on average, 14 years to develop (Wouters, McKee, and Luyten, 2020). Circuit-based therapies are composed of multiple parts and, by design, are more complex than small-molecule therapies. These factors may pose as a substantial obstacle for therapeutic approval. However, the very benefit of using circuits as a biomedical platform – specificity, selectivity, rational design, programmed behavior, and cellular control – may also apply to the developmental pipeline process. Additionally, the rapid design-build-test cycle permits rational iteration using modular parts. Therefore, modifying the circuit to optimize efficacy or incorporating additional safety features could prove easier to fix than traditional small-molecule drugs or biologics. The nature of circuit therapeutic delivery could itself be an effective safety measure by being transient. Circuit-based cell therapy could also contain similar, and much-desired safety features.

5.2 Circuit-based therapeutics – an exciting future lies ahead

Synthetic biology has transformed the way society interacts with biology-derived components: burgers that bleed (Rubio, Xiang, and Kaplan, 2020), fabric originating from fungi and insects (Jones et al., 2020), enhanced bacteria for agricultural uses(Khalil and Collins, 2010) and engineered living cells as biomedical products (Fischbach, Bluestone, and Lim, 2013). As we look to the future, synthetic biology will become a disruptive force in most, if not all, realms of technology. Parallel developments in cell therapy, genome engineering, and delivery methods can be leveraged for circuit-based therapies. As targeted, rationally-designed therapies continue to revolutionize medicine, systematic efforts to standardize and enrich the foundational aspects of circuit-based therapeutics are imperative in order to fully realize the promise of synthetic biology for medicine and beyond.

References

Chung, Hokyung K et al. (May 2019). "A compact synthetic pathway rewires cancer signaling to therapeutic effector release". en. In: *Science* 364.6439.

Fischbach, Michael A, Jeffrey A Bluestone, and Wendell A Lim (Apr. 2013). "Cell-based therapeutics: the next pillar of medicine". en. In: *Sci. Transl. Med.* 5.179, 179ps7.

Jones, Mitchell et al. (Sept. 2020). "Leather-like material biofabrication using fungi". en. In: *Nature Sustainability* 4.1, pp. 9–16.

Khalil, Ahmad S and James J Collins (May 2010). "Synthetic biology: applications come of age". en. In: *Nat. Rev. Genet.* 11.5, pp. 367–379.

Kirschner, Marc (2016). "Systems Medicine: Sketching the Landscape". In: *Systems Medicine*. Ed. by Ulf Schmitz and Olaf Wolkenhauer. New York, NY: Springer New York, pp. 3–15.

Nielsen, Alec A K et al. (Apr. 2016). "Genetic circuit design automation". en. In: *Science* 352.6281, aac7341.

Okamoto, M and M A Savageau (Apr. 1984). "Integrated function of a kinetic proofreading mechanism: steady-state analysis testing internal consistency of data obtained in vivo and in vitro and predicting parameter values". en. In: *Biochemistry* 23.8, pp. 1701–1709.

Rubio, Natalie R, Ning Xiang, and David L Kaplan (Dec. 2020). "Plant-based and cell-based approaches to meat production". en. In: *Nat. Commun.* 11.1, p. 6276.

Torres, Nestor V and Guido Santos (Dec. 2015). "The (Mathematical) Modeling Process in Biosciences". en. In: *Front. Genet.* 6, p. 354.

Wouters, Olivier J, Martin McKee, and Jeroen Luyten (Mar. 2020). "Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018". en. In: *JAMA* 323.9, pp. 844–853.