Dynamic control of intracellular signaling and genome engineering in space and time

John C. Rose

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2017

Reading Committee
Dustin Maly, Chair
Douglas Fowler
Paul Nghiem

Program Authorized to Offer Degree:
Molecular and Cellular Biology

University of Washington

## Abstract

John C. Rose

Chair of the Supervisory Committee:
Professor and Raymon E. and Rosellen M. Lawton Distinguished Scholar in Chemistry
Dustin J. Maly
Department of Chemistry

Cells continuously sense both their external and internal environments, integrate diverse and often conflicting information, and respond. The potential responses are remarkably varied: it could be to kill an invading pathogen, repair damaged DNA, proliferate to heal a wound, or undergo programmed cell death. The process of sensing, integrating, and responding to information is carried about by complex biochemical networks within the cell. These cellular networks dictate a cell's behavior, function, and identity. Moreover, dysregulation of these networks is a hallmark of diseases as prevalent and damaging as cancer and diabetes. It is imperative that we advance our understanding of these cellular networks in order to improve treatments for when these networks go awry. Unfortunately, our ability to adequately investigate these networks is hampered by their sheer complexity. Inherent to cellular networks are multiple layers of complexity: they are complex in time, in space, and in their architecture (i.e. the arrangement of the interactions that comprise the network).

Recently, high-throughput methods have greatly improved our ability to observe these networks, allowing the characterization of thousands of genes or proteins in a single experiment. Yet, our ability to interrogate cellular networks has not kept pace with our

ability to observe them. As a result, studies often yield conclusions which are largely phenomenological with limited mechanistic insight. Mechanistic insights are critical to developing novel therapies or better utilizing existing ones. To gain such insights would require tools that confer tunable control of individual network components with spatial and temporal precision, allowing systematic dissection of a network. Up to this point, the development of such tools, generally engineered proteins that are controlled by small molecules or light, has been hampered by the reliance on empirical protein engineering strategies that are inefficient, arduous, and costly.

To address this outstanding problem, we developed a computational framework for the systematic design of small-molecule controlled proteins. This framework utilizes protein design tools developed by David Baker's lab, and we show it greatly expedites the development process. Using this framework, we engineered Chemically Inducible Activator of RAS (CIAR), which enables tunable, spatiotemporally precise control of RAS activation. RAS is frequently hyperactivated in human cancers, and using CIAR we characterized dynamic features or RAS biology which were inaccessible to previous methods. For instance, we found that RAS signaling kinetics differ between cell lines, which may reflect the divergent propensities of different cell types to develop RAS-driven cancers. We also show that RAS signaling can be rewired by small molecule inhibitors currently used to treat melanoma, which sheds light on a phenomenon where this drug actually promotes growth of secondary cancers. More recently, we have demonstrated that we can control RAS activation at different subcellular locations, such as the Golgi. It has been suggested that the subcellular location of RAS activation can dictate phenotypic outcomes, including whether a developing T-cell proliferates or

undergoes apoptosis. CIAR will enable unprecedented examinations of such phenomena.

In parallel, we took a similar approach to study the processes involved in genome engineering. The development of CRISPR/Cas9 for genome editing has led to intense interest in both its potential research and therapeutic applications. Yet, little is known regarding the dynamics of Cas9-mediated DNA cleavage and subsequent DNA repair. Using an approach analogous to that used to develop CIAR, we engineered a rapidly inducible Cas9 variant, chemically-inducible Cas9. Additionally, we developed the first assay for quantitative, temporally-resolved monitoring of double-strand breaks (DSBs), DSB-ddPCR. Using these two technologies, we conducted a first-ever examination of Cas9-mediated DNA cleavage and repair dynamics. We found that Cas9 cleavage is rapid, and that both cleavage and repair kinetics differ between loci. We envision these technologies will enable in-depth examinations of this heretofore unexplored region of CRISPR/Cas9 biology, with potential implications for its applications in research and the clinic.

Taken together, the technologies developed in the course of this dissertation demonstrate the utility of precision tools for the study of complex cellular networks. CIAR and ciCas9 have already begun to yield insights into intracellular signaling and genome engineering and are the foundation for several ongoing investigations. Beyond these specific tools, the design methods and engineering strategies we have devised may have greater impact. It is our hope that they will aid the development of future technologies to probe the inner workings of our cells, deepening our understanding of the networks which drive human health and disease.

Table of Contents

**Acknowledgments**

The work contained herein would not have been possible without the professional and personal support of my family, friends, colleagues, and mentors. It would be impossible to adequately express my gratitude to them all. What follows is a wholly inadequate attempt to do so.

I'd first like to acknowledge my family. My parents have supported my love of science for approximately three decades, ever since they surrounded me with piles books on biology, physics, and chemistry back when I was starting elementary school. My siblings—Sarah, Jenna, and Michael—are simply and irrefutably the best siblings. Science isn't needed to prove that fact (but it could). Similarly, my newest family members, the niblings—Cora, Dylan, and Elsa—are simply and irrefutably the best humans. A quick call with them, or better, yet, a visit, and it was hard to be too upset about a failed experiment. My family was always there (even if "there" is Chicago) to temper my PhD's lowest moments. I'm also lucky to have a wonderful extended family. Special attention should be paid to my cousin Max, who joined me in the Pacific Northwest several years ago. Whenever I needed a break from science, I could always count on Max, the consummate outdoorsman, to take me into the woods

Throughout my PhD, I've been lucky to be surrounded by tremendous colleagues. In particular, Jason Stephany, Possu-Huang, William Valente, and Dan Cunningham-Bryant were particularly helpful. Of note, these four friends and colleagues represent four labs at UW, testament to the collaborative nature of this university and the importance of collaboration to my work and development as a scientist. Members of the

MSTP—particularly Michael Duyzend and Kyle MacQuarrie (as well as honorary MSTP members Drew Miller and William MacQuarrie-Miller)—have become dear friends. And friendships with those outside of academia, especially Dylan Keil, John Freedlund, Charlie Henschen, and Alex Cooper, have kept me (reasonably) sane.

None of this work would have been possible without the support of the UW MSTP, UW MCB, or ARCS. I was Camille and Jim Uhlir's ARCS fellow, and it has been such a pleasure to get to know them over the past five years, and their support has meant a great deal to me.

I'd also like to thank my thesis committee. Doug Fowler has essentially been a co-mentor along with Dusty these past few years, and has always been quick to challenge my assumptions in a way that has made me a better a better scientist. David Baker and Paul Nghiem never failed to ask insightful questions to help focus my research, and also provided invaluable guidance as I considered my career going forward. And from the moment I interviewed at UW to this very day, Mary-Claire King has been a remarkable source of support, knowledge, and inspiration.

Lastly, I'd like to thank my mentor, Dusty Maly. Joining Dusty's lab was easily one of the best decisions I've made in my academic and professional career. In the interest of brevity, I will simply state that I truly appreciate the freedom he gave me to explore my ideas, both the good and those that aspired to mediocrity… And as I explored those ideas, our conversations changed the way I think about science, and, in truth, the world more broadly.

## Chapter 1: Introduction

At its core, the purpose of biomedical research could be considered as two non-mutually exclusive pursuits: (1) to understand and to (2) control biological processes. In the most traditional sense, we seek to understand biological processes, such that we can modulate that process and improve human health. But the relationship between these two pursuits is far from unidirectional. In fact, the ability to control or manipulate biology is integral to efforts to examine and understand its underpinnings.

Until relatively recently, examinations of biological systems were limited to reductionist approaches (Fig. 1.1). In these approaches, one to several nodes (e.g. genes, proteins, phosphosites) are observed before and after a perturbation. Because so few nodes were observed in each experiment, a great deal of information about the behavior of the system is lost in each experiment. Furthermore, the perturbations available to researchers have come with significant limitations. Canonical stimuli, e.g. treatment to growth factors or exposure to environmental stresses, have numerous pleiotropic effects, complicating or even precluding precise elucidation of network architecture. Genetic methods, such as gene knockout/knockdown (e.g. via CRISPR or RNAi) or overexpression, have a high degree of target specificity, but little to no temporal resolution. Consequently, these methods are of limited utility when studying highly dynamic processes, and it can be difficult to parse out direct effects from perturbation of the desired target from compensatory changes made by the system over time. Together, the limitations inherent to both the observations and perturbations traditionally

used in reductionist biological studies hampers their ability to yield deep, quantitative insights into complex networks.

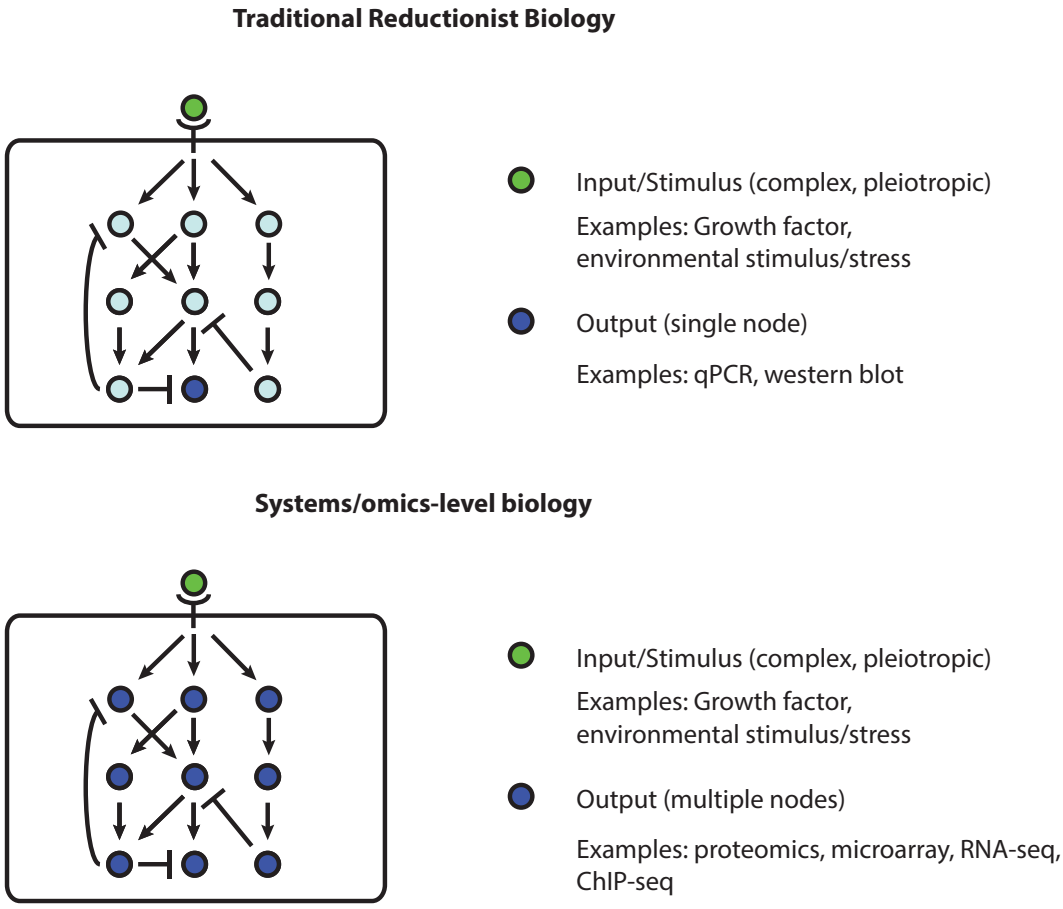



**Figure 1.1 Predominant approaches to the study of biological networks.**

More recently, our ability to observe the complex systems that drive cellular behavior has expanded tremendously. In particular, proteomics and sequencing-based technologies have enabled the simultaneous characterization of tens of thousands of nodes—whether genes, proteins, phosphorylation events, et al. (Fig. 1.1). Through these system-level observations, researchers have made advances in inferring the architecture of signaling networks and other cellular processes[1-8]. Yet, our ability to observe these complex biological systems has far outstripped our ability to precisely

perturb them. Thus, mechanistic understanding of these networks generally remains elusive even as high-throughput experiments reveal their general contours.

To gain quantitative, mechanistic understanding of complex networks, tools that would enable us to control specific nodes with spatiotemporal precision would be of enormous utility. Ideally, these tools—which I refer to as molecular rheostats or simply rheostats—would allow tunable and rapid control of a node of interest (Fig. 1.2). Such rheostats would allow us to interrogate the mechanistic links between individual nodes. Combined with high-throughput observations, rheostats could also yield profound insight into the emergent properties of these networks (Fig. 1.3). While some such tools have been developed[9-12], reliance on empirical design methods has precluded expansion of this approach to many nodes and systems.

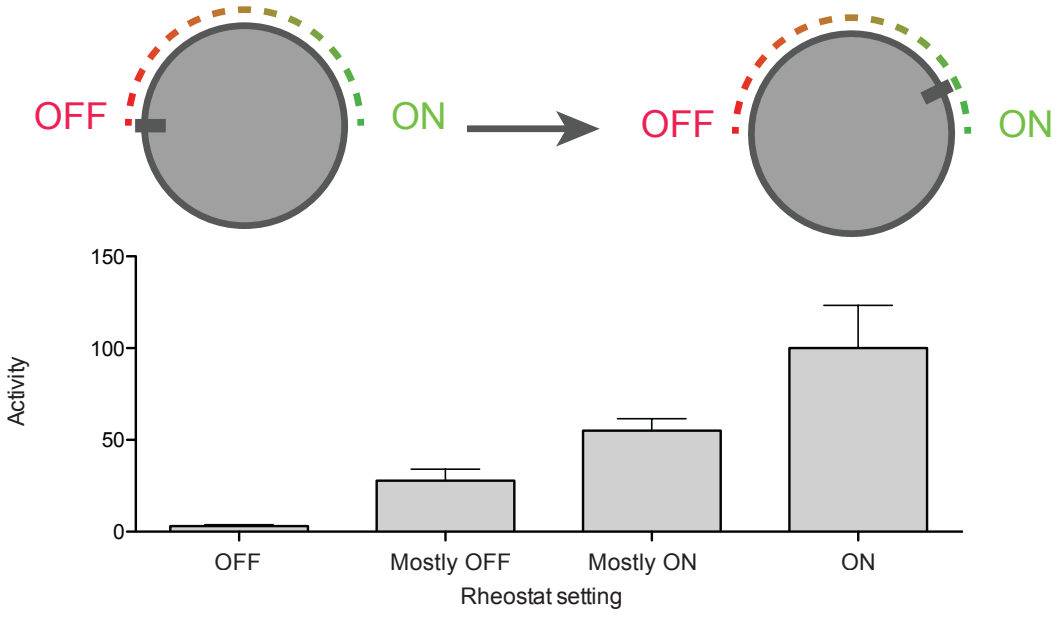



**Figure 1.2 Molecular rheostats confer precise tunable control of a biological node of interest**

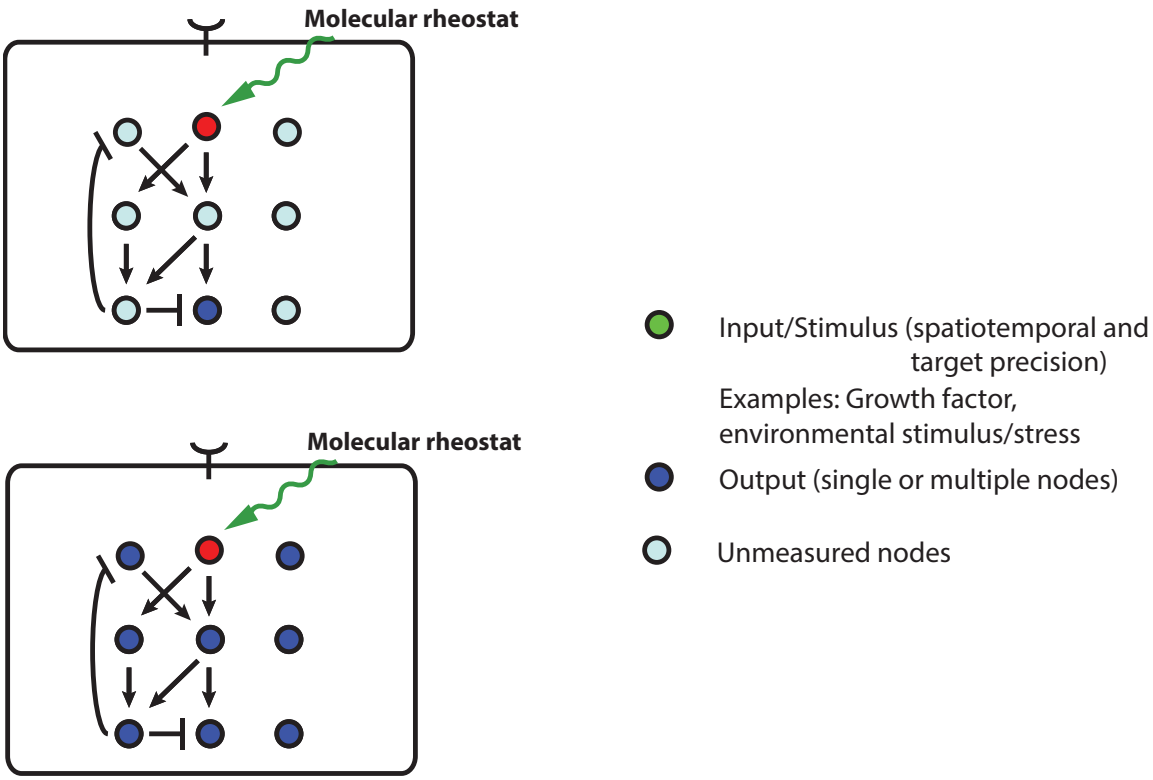



**Figure 1.3. Molecular rheostats allow precise interrogation of complex biological networks and can be combined with reductionist or systems-level observation methods.**

In this dissertation, I will describe two related efforts to develop tools for controlling dynamic cellular processes. **Chapter 2** will describe the development of a computational framework for design of single-component molecular rheostats. This framework facilitated the development of a molecular RAS rheostat and its subsequent application to studies of intracellular signaling. **Chapter 3** will describe the development of tools—including a Cas9 rheostat—for controlling and monitoring genome editing and engineering.

## Chapter 2: A computationally engineered RAS rheostat reveals RAS/ERK signaling dynamics

A version of this chapter has been previously published as:

Rose, J.C., Huang, P.-S., Camp, N.D., Ye, J., Leidal, A.M., Goreshnik, I., Trevillian, B.M., Dickinson, M.S., Cunningham-Bryant, D., Debnath, J., Baker, D., Wolf-Yadlin, A., Maly, D.J. (2017). A computationally engineered RAS rheostat reveals RAS-ERK signaling dynamics. Nat. Chem. Biol. *13*, 119–126.

J.C.R. and D.J.M. conceived the project and designed the study. J.C.R. and P-S.H. developed computational methods and performed the design calculations with guidance from D.B. J.C.R., I.G, and D.C.B. performed biochemical experiments. J.C.R, J.Y., and M.S.D. performed cell biology experiments under the supervision of J.D. and D.J.M. J.C.R, A.M.L., M.S.D., and B.M.T. generated reagents and cell lines. J.C.R. generated samples for mass spectrometry experiments. N.D.C. prepared peptide samples, performed MS analyses, and processed MS data under the supervision of A.W-Y. J.C.R. analyzed the MS data. J.C.R., P-S.H., and D.J.M. wrote the manuscript. All authors read and approved the manuscript.

## 2.1 Abstract

Synthetic protein switches controlled with user-defined inputs are powerful tools for studying and controlling dynamic cellular processes. To date, these approaches have relied primarily on intermolecular regulation. Here, we report a computationally-guided framework for engineering intramolecular regulation of protein function. We utilize this framework to develop Chemically Inducible Activator of RAS (CIAR), a single-component RAS rheostat that directly activates endogenous RAS in response to a small molecule. Using CIAR, we show that direct RAS activation elicits markedly different RAS/ERK signaling dynamics compared to growth factor stimulation, and that these dynamics differ between cell types. We also found that the clinically-approved RAF inhibitor vemurafenib potently primes cells to respond to direct wild-type RAS activation. These results demonstrate the utility of CIAR for quantitatively interrogating RAS signaling. Finally, we demonstrate the general utility of our approach to design intramolecularly-regulated protein tools by applying this methodology to the Rho Family GEFs.

## 2.2 Introduction

Engineered regulatory systems have emerged as useful tools for both studying and manipulating dynamic cellular processes[13,14]. These methods allow temporally precise control of a protein of interest with a researcher-defined input. Such perturbations are ideal for interrogating signaling networks, which are dynamic and act on rapid timescales. With few exceptions[11,12,15,16], methods for chemical genetic and optogenetic activation of signaling networks have relied on multi-protein, intermolecular regulatory

systems[9,14,17-19] that are only applicable to proteins that can be sequestered from their sites of function. Despite several advantages, intramolecular regulatory systems have been far less utilized. As only a single protein component is used, intramolecular regulatory systems can be rapidly transported to diverse cellular systems while obviating issues of stoichiometry. In addition, because intramolecularly regulated proteins do not rely on redistribution to control activity, they can be localized to a site of action in the autoinhibited state, allowing application to signaling systems independent of component location. Furthermore, basal localization to sites of function may permit more rapid responses to inputs. Finally, inter- and intramolecular systems can be integrated, allowing researchers to encode more complex responses with multiple layers of regulation. In large part, the dearth of synthetic intramolecularly regulated protein systems is due to the difficulty inherent in engineering allostery[20].

Here, we describe a computationally-guided framework for intramolecular regulatory design. This approach was applied to an activator of the RAS GTPases, which couple transmembrane receptors to intracellular signaling pathways and regulate diverse cellular processes[21]. Guided by our computational approach, we generated a genetically-encoded RAS rheostat, which we termed Chemically Inducible Activator of RAS (CIAR), capable of tunably controlling endogenous RAS activation state with high temporal precision. Using CIAR, we demonstrate that direct RAS activation drives sustained ERK phosphorylation, whereas epidermal growth factor (EGF) stimulation yields a transient response. Furthermore, direct activation of RAS elicits distinct phosphorylation kinetics in the RAS/ERK module in two different cell lines. CIAR was used with genetic/pharmacological perturbations and global phosphoproteomics to

provide insight into the kinetics of RAS-driven signaling cascades. Finally, we demonstrate the generality of our computational approach in guiding the design of intramolecularly-regulated systems through application to Rho Family GEFs.

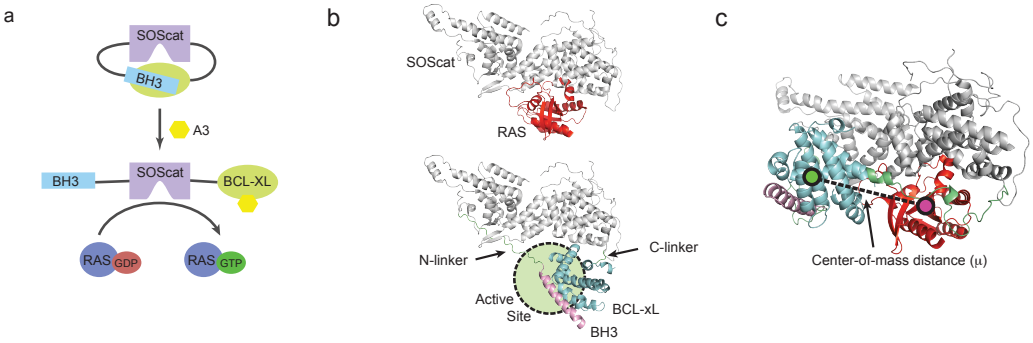
## 2.3 Results

### 2.3.1 Computational design of an autoinhibited RAS activator

In developing a RAS rheostat, we integrated a synthetic regulatory switch with the RAS activator Son of sevenless (SOS)—a RAS guanine nucleotide exchange factor (Fig. 2.1a). SOS can be minimized to a constitutively-active catalytic unit (SOScat), consisting only of the Cdc25 and Rem domains[22], which we predicted could be autoinhibited by constraining a protein-protein interaction complex over the active site (Fig. 2.1b). We selected the interaction between BCL-xL and BH3 peptides as a synthetic regulatory switch because cell-permeable, small-molecule disruptors of this well-characterized protein-protein interaction, including A-385358 ("A3")[6,23], are available. Critical for computational design, the BH3 peptide ("BH3") and BCL-xL complex forms a rigid body that can be modeled as a single protein domain.

For synthetic regulatory system design, one of the most critical parameters was the selection of linkers connecting BH3 and BCL-xL to SOScat's termini. Optimal linkers must satisfy three Design Criteria: (1) be of sufficient length to allow intramolecular BH3/BCL-xL complex formation; (2) the energetic local minimum for this complex must reside over SOScat's active site; and (3) the energetic landscape surrounding this minimum must be sufficiently steep-walled to ensure active site occlusion. Because potential permutations of N- and C-terminal linker lengths—"linker-space"—are too

expansive to screen experimentally, *RosettaRemodel* was employed[24]. To simplify simulations and avoid exhaustive sampling, the BH3/BCL-xL complex was treated as a rigid body within a loop that bridges SOScat's termini. Modeling is, thus, reduced to a loop closure problem based on linker geometry, and the simulated low-energy ensemble–representing all the viable solutions that allow the loop to connect on both ends without strain–is assumed to correlate with the localization of the BH3/BCL-xl complex. To quantitate the localization parameter, we defined a metric, $\mu$, as the distance between the center-of-mass (CoM) of RAS bound to the active site of SOScat and the CoM of the BH3/BCL-xL complex in each output model (Fig. 2.1c). $\mu$ serves as a metric for guiding empirical screening by measuring the deviation of the BH3/BCL-xL complex from SOScat's active site.




1

**Figure 2.1 | Strategy for engineering a Chemically Inducible Activator of RAS (CIAR). (a)** Schematic depiction of the overall approach for engineering small-molecule controlled autoinhibition of a RAS activator. BCL-xL and a BH3 peptide are appended via flexible linkers to the termini of a constitutively active SOScat mutant (T968L). In the absence of A-385358 ("A3"), a disruptor of the BCL-xL/BH3 interaction, the BCL-xL/BH3 complex occludes the SOScat active site. Upon addition of A3, the BCL-xL/BH3 complex is disrupted and the active site liberated, allowing SOScat to activate RAS. **(b)** Comparison of the structure of SOScat bound to RAS and the desired conformation of CIAR, in which the BCL-xL/BH3 complex occludes the SOScat active site. **(c)** Depiction of $\mu$, the metric used to evaluate output structures from computational modeling of CIAR designs.

Preliminary modeling suggested that BH3 be placed at the N-terminus of SOScat and BCL-xL at the C-terminus. For each trajectory, an arbitrary break in the linkers was introduced, and subsequent chain closure only occurred in geometrically allowed models. Models without successful loop closure were discarded. Following chain closure, torsional angles within the linkers were allowed to vary in order to sample the conformations of the BH3/BCL-xL complex relative to SOScat. To gain insight into the effects of linker lengths on BH3/BCL-xL complex localization, we performed 1,000 independent sampling runs for all permutations of N- and C-terminal linkers of length {0, 5, 10, 15, 20, 25, 30} residues. The 49,000 trajectories produced 18,473 geometrically allowed models, for an overall closure rate of 37.7% (Design Criterion 1; Supplementary Fig. A.1).

Next, the localization of the complex relative to SOScat's active site was evaluated. The ensemble average $\mu$ reports on the BH3/BCL-xL complex localization from SOScat's active site (Design Criterion 2). Plot of the average $\mu$ for each permutation indicates that linker length has a pronounced impact on the proximity of the BH3/BCL-xL complex to the active site (Fig. 2.2a). The lowest average $\mu$ (11.79 Å) resulted from N-terminal linker (NL) = 15 residues and C-terminal linker (CL) = 5 residues (termed S.15.5). Our design strategy relies on localizing the BH3/BCL-xL complex over SOScat's active site to obtain autoinhibition. Therefore, we predicted that shorter linkers would provide better autoinhibition than longer linkers by tightly constraining the BH3/BCL-xL complex over the region of RAS binding (Design Criterion 3). To estimate the mobility of the BH3/BCL-xL complex–which is a proxy for the steepness of the gradient surrounding the energetic minimum–the standard deviation (SD) of $\mu$ was calculated. As expected,

shorter linker pairs yielded smaller SDs, with S.15.5 exhibiting the second smallest SD (2.53 Å) (Fig. 2.2b). Indeed, visualizing the BH3/BCL-xL CoM for each trajectory reveals that models of the BH3/BCL-xL complex cluster tightly over SOScat's active site for S.15.5, but are widely dispersed for longer linkers (Fig. 2.2c-d, Supplementary Fig. A.2); suggesting that the linker space centered around (NL = 15, CL = 5) would be a promising starting point for our experimental efforts.

### 2.3.2 Biochemical and cellular refinement of CIAR design

To sample the optimal region of linker space identified through our computational approach, three candidate CIAR constructs–S.21.6, S.17.6, and S.15.3–were tested in an *in vitro* nucleotide exchange assay. All three constructs demonstrated a dose-dependent increase in RAS GEF activity in the presence of a BH3 peptide competitor, with minimal activation in the presence of a control BH3 peptide (Fig. 2.3a, Supplementary Fig. A.3). Thus, disruption of the BH3/BCL-xL complex in *trans* leads to *in vitro* RAS activation.

A luciferase-based serum response element reporter assay for RAS/ERK pathway activation was next used to evaluate constructs in cells. We screened constructs within the linker space supported by the preceding *in vitro* exchange assay results, and found that S.17.3 exhibited the most optimal behavior. In the absence of drug, S.17.3 is highly autoinhibited (Supplementary Fig. A.4), but 25 μM A3 leads to an 11.3-fold induction of luciferase reporter signal (Fig. 2.3b). We hypothesized that constructs containing other linkers were less optimal due to lower levels of autoinhibition for one of two reasons: (1) long linkers permit excessive conformational freedom, allowing the intact BH3/BCL-xL

complex to vacate SOScat's active site or (2) short linkers disfavor the intramolecular association between BH3 and BCL-xL. Indeed, constructs with longer or shorter linkers than S.17.3 display higher levels of basal RAS activation (Supplementary Fig. A.4). Together, the low basal activity, coupled with robust RAS activation upon A3 treatment, makes S.17.3 an effective Chemically Inducible Activator of RAS (from here forward referred to as CIAR).

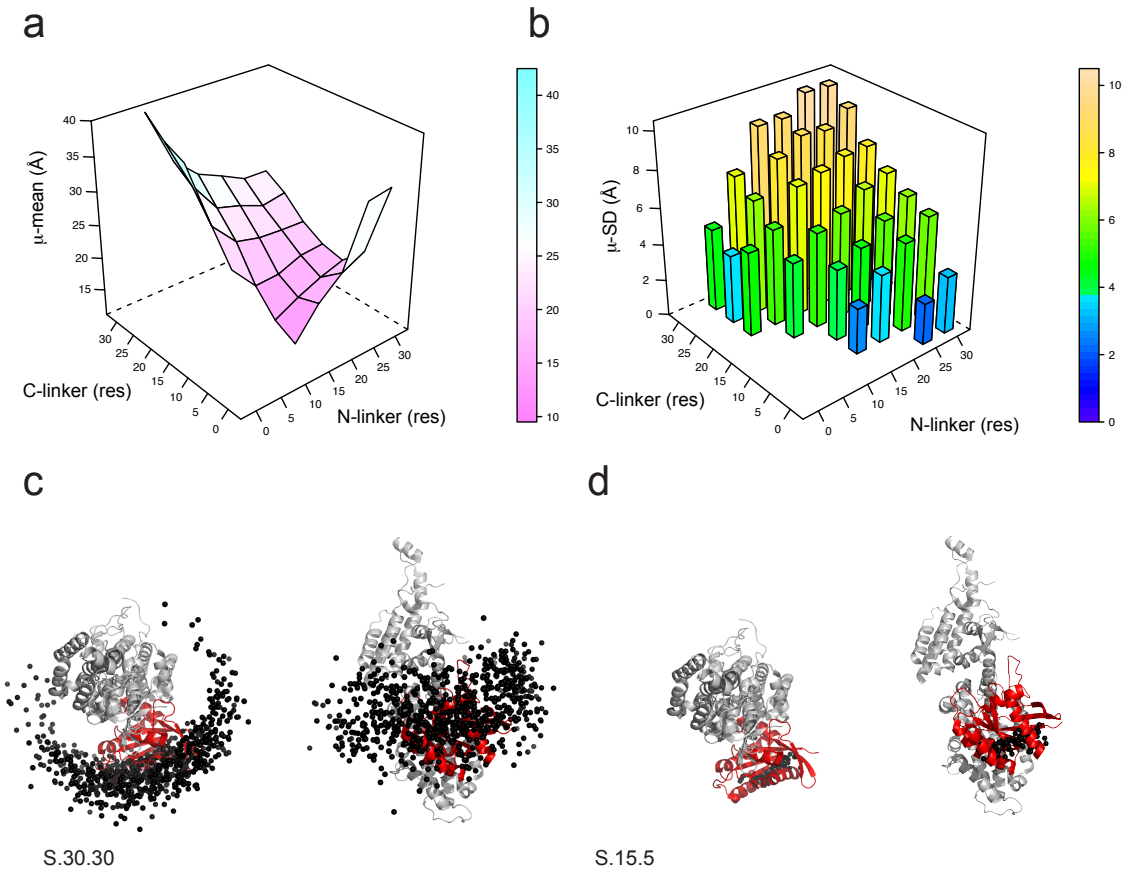


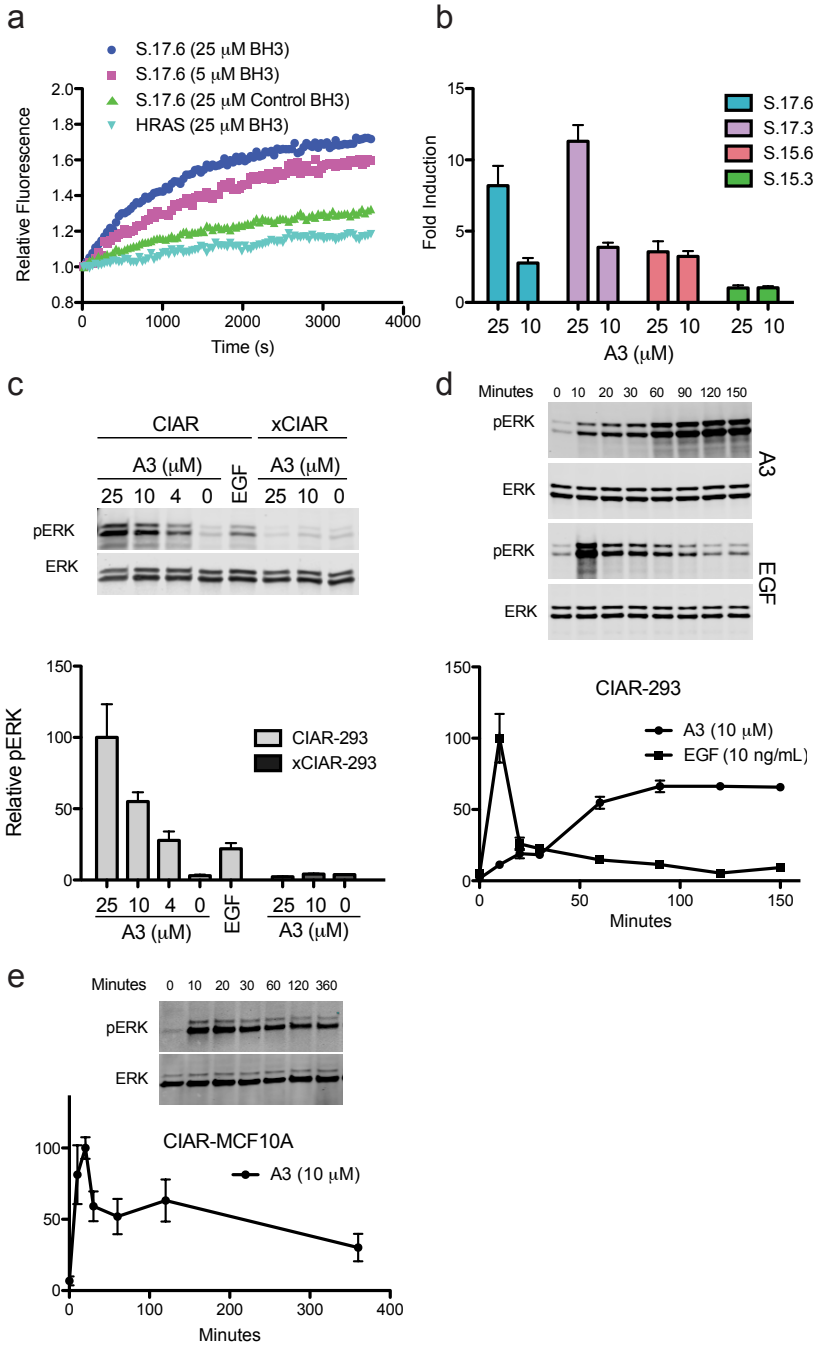

**Figure 2.2 | Computational framework for designing CIAR.** **(a)** Average $\dot{r}$ (in Å) with respect to N- and C-terminal linker length (NL and CL, respectively). **(b)** Standard deviation of $\dot{r}$, as a function of NL and CL. **(c,d)** The BH3/BCL-xL complex mass centers for every trajectory of {NL = 30, CL = 30} **(c)**, and {NL = 15, CL = 5} **(d)**, with BH3/BCL-xL mass centers shown as black spheres, overlaid on the structure of SOScat (cartoon, grey) bound to RAS (red) at its active site. Construct nomenclature: S.(NL).(CL).

### 2.3.3 CIAR functions as a cellular RAS rheostat

Flp-In T-REx 293 cells stably expressing either CIAR or xCIAR–a construct containing a single inactivating mutation (F929A) within SOScat[25]–(CIAR-293s and xCIAR-293s) were generated to precisely characterize CIAR's cellular behavior. Comparison of CIAR-293s and xCIAR-293s allows differentiation between direct RAS activation from potential A3 off-target effects. CIAR-293s or xCIAR-293s were treated for 30 min with A3, EGF, or DMSO to determine the effect of direct RAS activation on RAS/ERK signaling. Consistent with effective CIAR autoinhibition, similar ERK phosphorylation was observed in untreated xCIAR-293s and CIAR-293s (Fig. 2.3c). Provision of A3 resulted in a dose-dependent increase in phospho-ERK in CIAR-293s but no increase in xCIAR-293s, indicating that CIAR functions as a rheostat for RAS activation.

To characterize the temporal features of RAS/ERK signaling in response to direct RAS activation, an A3 and EGF treatment timecourse in CIAR-293s was performed. Consistent with previous studies[26-28], EGF elicited pronounced yet transient ERK phosphorylation. In contrast, direct RAS activation with CIAR led to a linear increase in phospho-ERK for 90 min, reaching a plateau that was sustained until at least 150 min (Fig. 2.3d). Phospho-ERK levels remain elevated 48 hours post-A3 treatment in an extended timecourse (Supplementary Fig. A.5). To investigate the effects of direct RAS activation in another cellular context, we generated MCF10A cells–non-transformed mammary epithelial cells–stably expressing CIAR (CIAR-MCF10As). Interestingly, similar to EGF stimulation, phospho-ERK rapidly reached a transient maximum in

response to direct RAS activation, (Fig. 2.3e, Supplementary Fig. A.6)[28]. However, unlike EGF, direct RAS activation in CIAR-MCF10As resulted in elevated ERK phosphorylation for at least 6 hours. These results indicate that ERK phosphorylation dynamics are not simply a function of RAS activation but are influenced by other factors downstream of EGFR, and by the basal wiring of the RAS/ERK module in different cell types.




3

**Figure 2.3 | CIAR functions as a RAS rheostat. (a)** *In vitro* nucleotide exchange activity on HRAS of a candidate CIAR construct (S.17.6, {NL=17, CL = 6}) in the presence of competitor BH3 or control BH3 peptides. **(b)** Fold induction of the RAS/MAPK pathway in Flp-In T-REx 293 cells relative to DMSO treatment, as measured by a Serum Response Element (SRE) assay. **(c-f)** Phospho-ERK and total ERK immunoblots: **(c)** CIAR-293 and xCIAR-293 cells treated with A3 (4, 10, or 25 μM), EGF (2 ng/mL), or DMSO for 30 minutes. **(d)** Timecourse (0-150 min) of direct RAS activation with A3 (10 ¦ M) or stimulation with EGF (10 ng/mL). **(e)** Timecourse (0-360 min) of direct RAS activation with A3 (10 ¦ M) in CIAR-MCF10A cells. In **(b-e)** error bars = s.e.m (n =3, except in **(e)** n = 4). Uncropped immunoblots available in Supplementary Figs. 18-20.

### 2.3.4 Interrogation of intrinsic and extrinsic RAF regulation

We have demonstrated that direct RAS activation and EGF stimulation lead to markedly different RAS/ERK signaling module behavior in CIAR-293s. To investigate whether this is due to delayed formation of RAS•GTP, RAS activation was determined using Ras Binding Domain pull-downs (Fig. 2.4a). A3 treatment yielded a rapid increase in RAS•GTP in CIAR-293s, with near maximal levels within 10 min. These results suggest that, in addition to activating RAS, EGF stimulation modulates other downstream components that affect flux through the RAS/ERK module. RAF kinases—immediately downstream of RAS—are regulated by various mechanisms, including crosstalk with other signaling pathways. In particular, phosphorylation of several residues by kinases within (intrinsic) and outside (extrinsic) of the RAS/ERK signaling module have been implicated in the regulation of RAF. RAF phosphorylation dynamics were determined by profiling three RAF phosphosites—S29 and S259 in RAF1, and S365 in BRAF—using Parallel Reaction Monitoring Mass Spectrometry (PRM-MS) (Fig 4b). RAF1 S29 phosphorylation by ERK has been shown to be an intrinsic negative feedback loop of RAS/ERK signaling[29]. Inhibitory phosphorylation of RAF1 S259 and a similar site in BRAF–S365–by AKT and other kinases extrinsically regulate the RAS/ERK module[30,31]. PRM-MS revealed that RAF1 S29 phosphorylation and BRAF S365 dephosphorylation kinetics mirror that of phospho-ERK for EGF stimulation and direct RAS activation. In contrast, RAF1 phospho-S259 remained stable in response to both stimuli. Due to the absence of a response at RAF1 S259, we next profiled the status of the PI3K/AKT pathway, which is an extrinsic regulator of the RAS/ERK module, by monitoring phospho-AKT levels. Absent stimuli, AKT was phosphorylated in CIAR-293s and control

T-REx 293s, and phosphorylation levels were unaffected by EGF or A3 treatment (Fig. 2.4c, Supplementary Fig. A.7-8). Basally activated AKT in these cells could contribute to slower RAS/ERK kinetics through suppression of RAF1 activation. The divergent behavior of the extrinsic regulatory sites on RAF1 and BRAF, and the reported differential contributions of these isoforms in various cell lines[32-34], prompted further investigation of RAF regulation in CIAR-293s.

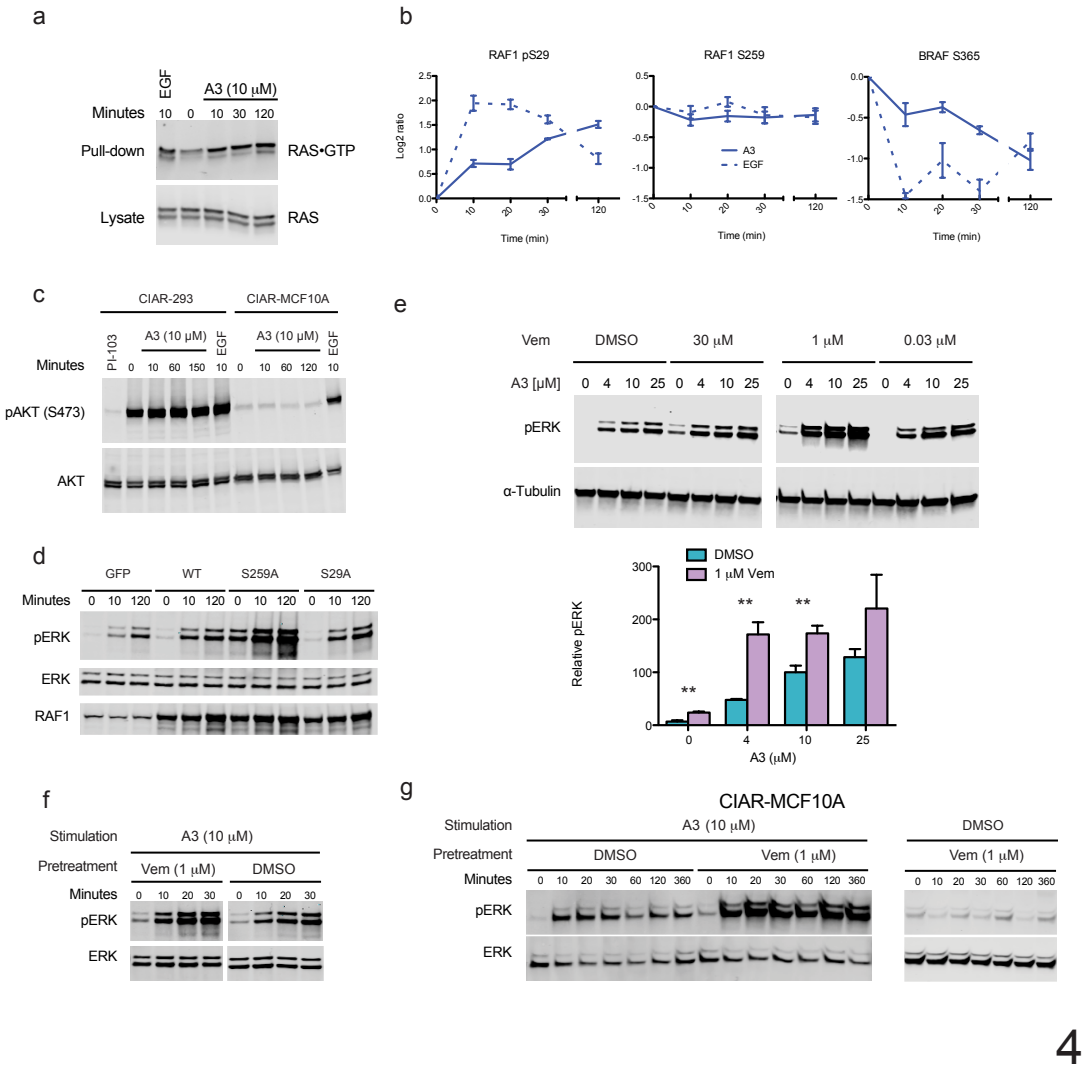
RAF recruitment to the plasma membrane, which is promoted by dephosphorylation of RAF1 S259 and BRAF S365, is critical in the regulation of the RAS/ERK module[35,36]. To examine this regulatory step, CIAR-293s were transfected with BRAF and RAF1 constitutively targeted to the plasma membrane (myrBRAF and myrRAF1). Both myrBRAF and myrRAF1 yielded constitutively active RAS/ERK signaling (Supplementary Fig. A.9). Transfection with wild-type BRAF, but not RAF1, yielded constitutively active RAS/ERK signaling, consistent with the lower levels of inhibitory regulation and higher basal activity of BRAF[35,37]. In fact, wild-type RAF1 had little effect on CIAR-mediated RAS/ERK signaling, indicating the RAS/ERK module is relatively insensitive to RAF1 but not BRAF concentration in 293 cells (Supplementary Fig. A.10). To interrogate the contribution of the RAF1 phosphosites profiled in Fig. 4b, CIAR-293s were transfected with wild-type RAF1, RAF1 S29A, or RAF1 S259A prior to activation of CIAR with A3 (Fig. 2.4d). Abrogation of RAS/ERK intrinsic negative feedback via S29A mutation had minimal effect on RAS/ERK signaling amplitude or kinetics—consistent with two non-exclusive explanations: (1) the majority of RAS/ERK flux does not pass through RAF1 in CIAR-293s, or (2) negative feedback through RAF1 S29 is a minor regulatory mechanism in this context. In contrast, S259A RAF1 led to high basal

RAS/ERK signaling prior to A3 treatment, followed by a pronounced increase in RAS/ERK flux upon CIAR activation. S259 dephosphorylation has been implicated in the relief of inhibitory 14-3-3 binding[35], and these results demonstrate that RAF1 S259 is a critical regulatory node in RAS/ERK signaling—particularly in the context of high basal PI3K/AKT signaling. Consistent with these findings, mutations of RAF1 S259 and surrounding residues, but not BRAF S365, have been observed in patients with Noonan and LEOPARD syndromes[38].

The interaction of RAF with KSR, MEK, and other RAF protomers also plays a role in RAS/ERK module regulation[35,39]. To further probe RAF regulation in response to direct RAS activation, we reasoned it might be feasible to rewire this module by pharmacologically modulating these interactions. Our efforts focused on vemurafenib, a RAF inhibitor that has demonstrated anti-tumor effects in BRAF(V600E)-driven melanomas[40,41]. Despite the clinical utility of RAF inhibitors, outgrowth of cutaneous malignancies expressing oncogenic RAS has been observed[42]. Preclinical evidence indicates vemurafenib and other RAF inhibitors induce proliferation of mutant-RAS tumors via paradoxical activation of RAS/ERK signaling in wild-type BRAF cells[43,44]. We reasoned that pre-treatment with vemurafenib may rewire the RAS/ERK module to sensitize CIAR-293 cells to direct wild-type RAS activation.

To test whether vemurafenib primes the RAS/ERK module for RAS activation, we pre-treated CIAR-293s with a range of vemurafenib concentrations prior to direct RAS activation. Based on phospho-ERK levels, we found that all concentrations of vemurafenib sensitized CIAR-293s to RAS activation (Fig. 2.4e). This effect was most

pronounced when CIAR-293s were incubated with 1 μM vemurafenib prior to treatment with either 4 or 10 μM A3. We next sought to profile the dynamics of the RAS/ERK response in the context of vemurafenib treatment. CIAR-293s were incubated with 1 μM vemurafenib or DMSO, followed by stimulation with A3. As expected, pre-treatment with vemurafenib leads to more pronounced ERK phosphorylation relative to pre-treatment with DMSO (Fig. 2.4f). In CIAR-293s, this effect is strongest at 20 and 30 min after treatment with 10 μM A3. In CIAR-MCF10As, vemurafenib leads to an even greater increase in RAS/ERK signaling, which is sustained until at least 360 min (Fig. 2.4g). The dramatic increase in ERK phosphorylation following direct RAS activation demonstrates that vemurafenib rapidly primes the RAS/ERK module for activation. Our results demonstrate that vemurafenib only weakly stimulates ERK signaling in the absence of RAS activation, but potently primes the RAS/ERK module to respond to direct activation of endogenous, wild-type RAS.



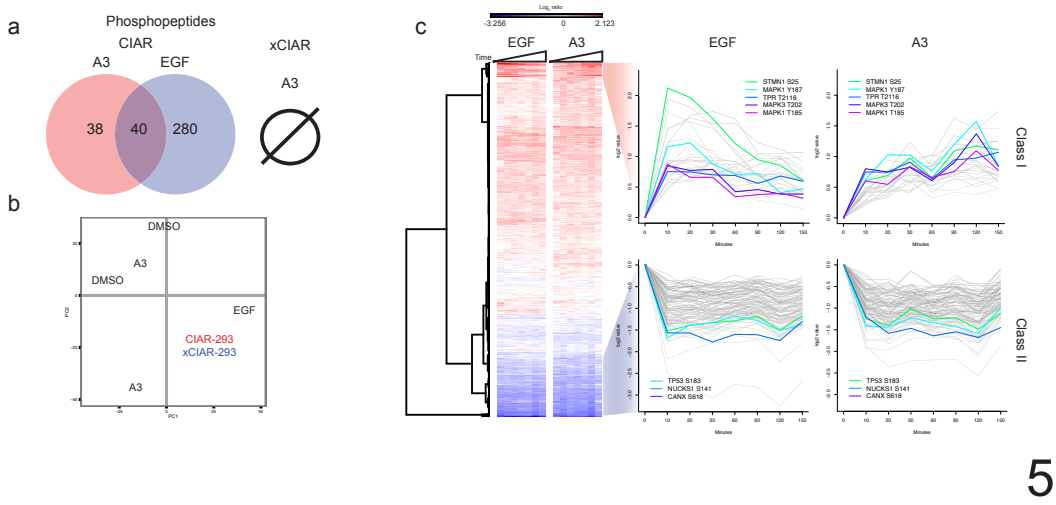

4

**Figure 2.4 | Interrogation of RAF regulation with CIAR. (a)** Endogenous RAS•GTP levels after treatment with A3 or EGF (10 ng/mL) for the indicated incubations. **(b)** PRM-MS of three RAF phosphosites in response to direct RAS activation (A3) or EGF stimulation. **(c)** Phospho-AKT and total AKT levels in CIAR-293 and CIAR-MCF10A cells treated with A3, EGF, or PI-103, a pan-PI3K inhibitor. **(d)** CIAR-293 cells were transfected with GFP, wild-type RAF1, or RAF1 containing the indicating mutations. After overnight serum starvation, cells were treated with A3 for the indicated durations. **(e)** CIAR-293 cells incubated with DMSO or varying concentrations of Vemurafenib for 1 h prior to stimulation with varying concentrations of A3 for 30 min. Quantitation of ERK phosphorylation after direct RAS activation of CIAR-293 cells pretreated with DMSO or Vemurafenib (1 μM) (* $p < 0.05$, ** $p < 0.01$). **(f-g)** Timecourse of phospho-ERK and total ERK immunoblot of direct RAS activation in CIAR-293 cells **(f)** and **(g)** CIAR-MCF10As incubated with DMSO or 1 μM Vemurafenib prior to stimulation with A3 (10 μM) for durations indicated. **(b,e)** Error bars = s.e.m, n = 3. **(d-g)** Representative immunoblots shown, for each n = 3. Uncropped immunoblots available in Supplementary Figs. 21-26.

### 2.3.5 Phosphoproteomic profiling of direct RAS activation

The preceding results demonstrate the utility of CIAR for investigating signal propagation to a single node—ERK—downstream of RAS. We predicted that the differential dynamics of ERK phosphorylation in response to direct RAS activation versus EGF stimulation may be reflected in other phosphosites. Therefore, we used quantitative phosphoproteomics to investigate signal propagation from RAS to thousands of signaling nodes—in this case, serine/threonine phosphosites. CIAR-293s were treated with DMSO, A3, or EGF. As a control, xCIAR-293s were exposed to the same conditions. After 30 min of stimulation, cells were subjected to a modified quantitative phosphoproteomic workflow (Supplementary Fig. A.11a). In CIAR-293s, 78 phosphopeptides showed significant changes in abundance after A3 treatment, while 320 phosphopeptides changed significantly in response to EGF. Of these hits, 40 phosphopeptides were shared across both conditions (Fig. 2.5a, Supplementary Data Set 1). In xCIAR-293s, no phosphopeptides showed significant differences in abundance upon A3 treatment relative to DMSO, indicating A3 is orthogonal to the signaling systems under investigation. Further validating the orthogonality of A3, principal components analysis reveals that A3- and DMSO-treated xCIAR-293s cluster with DMSO-treated CIAR-293s. A3- and EGF-treated CIAR-293s form their own, distinct, single-member clusters, suggesting dissimilarity in response to direct RAS activation and EGF stimulation (Fig. 2.5b). The differences observed between direct RAS activation and EGF stimulation after 30 min can be explained by two mechanisms: (1) phosphosites exhibiting significant differences represent signaling components

uniquely modulated by either direct RAS activation or EGF stimulation, or (2) these phosphosites are modulated by both stimuli but differ in dynamics—as observed with phospho-ERK (Fig 3d).




5

**Figure 2.5 | Phosphoproteomic comparison of direct RAS activation and EGF stimulation in CIAR-293 cells. (a-b)** CIAR-293 and xCIAR-293 cells were serum starved overnight before direct RAS activation with A3 (10 μM) or stimulation with EGF (10 ng/mL, CIAR-293 cells only) for 30 min (Supplementary Fig. A.11a). **(a)** In total, 9,996 phosphopeptides were identified and quantified. Of these, 2,329 phosphopeptides derived from 1,224 proteins were quantified in all biological replicates and used for further analysis. The Venn diagram depicts the number of phosphopeptides showing greater than $|\log_2(0.5)|$ fold-change for each condition relative to treatment with DMSO, with $p < 0.05$. **(b)** Principal components analysis of phosphoproteomic data for CIAR-293 (DMSO, A3 (10 μM), and EGF (10 ng/mL) and xCIAR-293 (DMSO and A3 (10 μM)) samples. **(c)** Timecourse comparison of direct RAS activation and EGF stimulation (Supplementary Fig. A.11b.). While 10,767 phosphopeptides were quantified, we analyzed the 657 phosphosites quantified in all replicates at all timepoints for both EGF and A3 treated samples to allow direct in-depth quantitative comparisons. Hierarchical clustering of 657 phosphopeptides present in all replicates reveals two prominent classes (Classes I & II) of peptides with distinct response patterns. Class I and Class II phosphopeptides are plotted with selected peptides highlighted. Mean peptide intensities from triplicate samples were normalized to 0 min intensities then $\log_2$ transformed. Three representative peptides from each class were validated using PRM-MS (Supplementary Fig. A.12).

To investigate differences in phosphorylation dynamics between direct RAS activation and EGF stimulation, a phosphoproteomic timecourse in CIAR-293s was performed (Supplementary Fig. A.11b). Hierarchical clustering of the 657 phosphosites quantified in all replicates at all timepoints for both EGF and A3 treatment revealed two prominent, distinct classes of responses (Fig. 2.5c). Representative phosphopeptides from each class were validated via PRM-MS (Supplementary Fig. A.12). Class I phosphopeptides display similar dynamics to those observed for phospho-ERK (Fig. 2.3d, Supplementary Fig. A.12a), undergoing a rapid and transient increase after EGF stimulation but a more gradual and sustained increase in response to direct RAS activation. Of the 25 phosphosites in this class, 19 (76%) either contain an ERK substrate motif ([pS/pT]P) or derive from the ERK1/2 activation loops (Supplementary Data Sets 2 & 3). Several of these targets—including STMN1, NUP50, and TPR—are validated ERK substrates[45-47], while others—including SRSF11, HNRNPC, and MARCKS—are candidate ERK substrates[48], indicating this class is largely defined by the RAS/ERK signaling module.

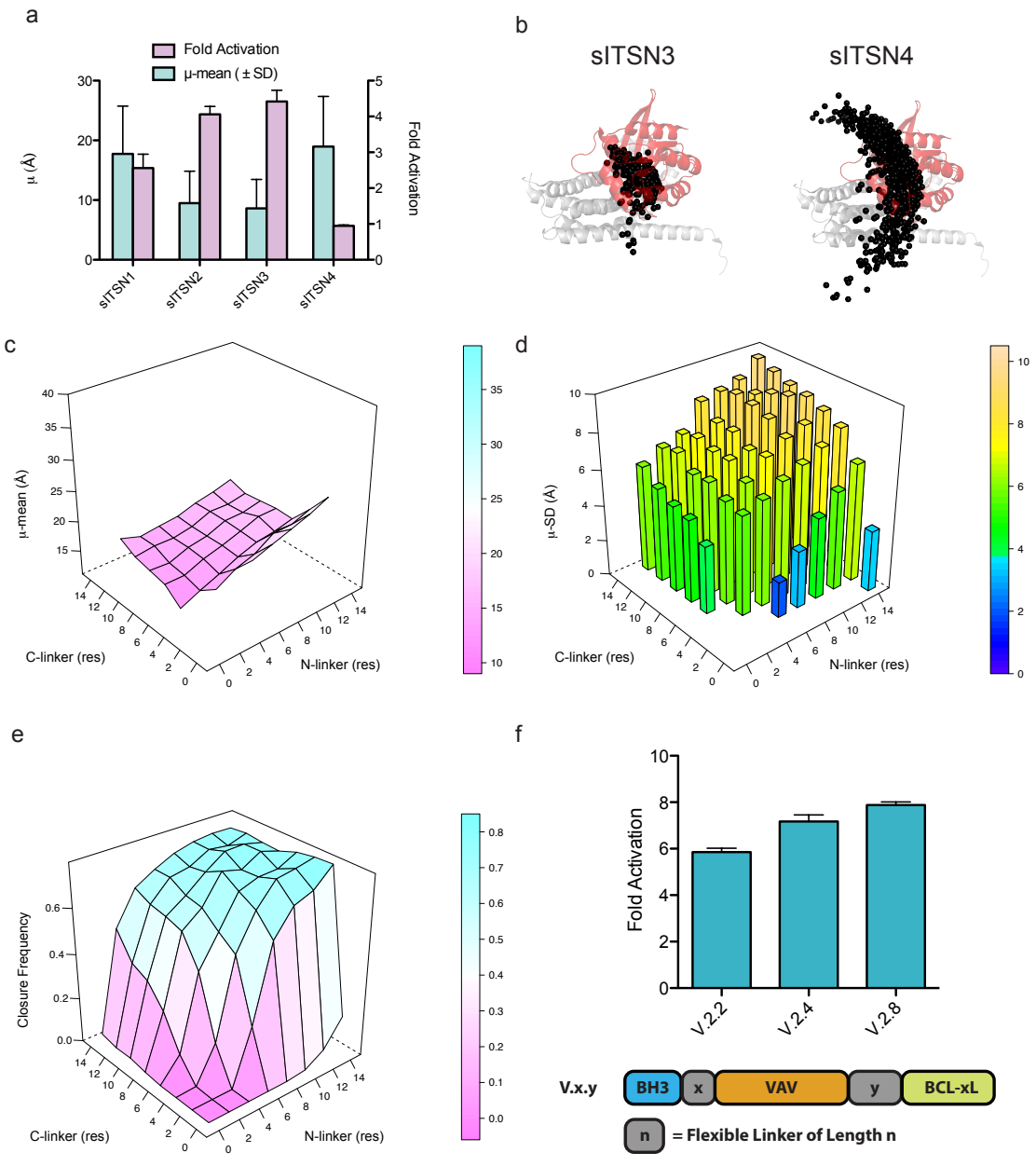
In contrast to Class I, the 106 phosphopeptides that comprise Class II exhibit a rapid, pronounced, and sustained decrease in abundance in response to both EGF stimulation and direct RAS activation (Supplementary Fig. A.12b). Members of this class form a highly interconnected interaction network (Supplementary Fig. A.13). Intriguingly, phosphopeptides from components of the spliceosome are significantly enriched ($p = 0.0306$, Supplementary Fig. A.14). This enrichment is consistent with previous studies showing that MAPKs regulate spliceosome component phosphorylation[48], and the large number of RNA-binding protein phosphosites that have been reported to decrease in abundance within 10 minutes of EGF stimulation[27]. Phosphopeptides in this class lack

appreciable differences in their responses to direct RAS activation and EGF stimulation, suggesting that RAS activation alone is sufficient to drive rapid downregulation of these phosphosites. Moreover, the rapid timescale on which these sites are downregulated suggests that the behavior of this class of peptides is not dictated by the RAS/ERK module.

### 2.3.6 Extension of computational rheostat design

To demonstrate the generality of our computational approach, the computational framework developed for SOS was retrospectively applied to a chemically inducible Rho Family GEF called Intersectin (ITSN) that we previously developed empirically[12]. We found that $\mu$-mean and $\mu$-SD accurately predicted the *in vitro* performance of each synthetic ITSN constructed tested, in that lower $\mu$-mean and $\mu$-SD values correlated with higher fold activation (Fig. 2.6a,b, Supplementary Fig. A.15). Next, we applied this framework to the design of a chemically inducible construct of the Rho Family GEF VAV. Notably, the $\mu$-mean surface for VAV was considerably flatter than that estimated for SOScat, indicating that Rho Family GEFs are more tolerant of linker length variations in the development of intramolecular regulation (Fig. 2.6c). The $\mu$-mean and $\mu$-SD minima (Fig. 2.6d, Supplementary Fig. A.16) suggested constucts with NL = 2 and CL ≤ 8 (V.2.2, V.2.4, V.2.8) were promising candidates. Indeed, all three constructs tested displayed inducible GEF activity *in vitro* and showed correlation between the measured fold activation and closure frequencies (Fig. 2.6e-f). Simulations with closure frequency <10% generally are cases with limited solutions to satisfy the bond geometries in the linkers. The observed reduction in fold activation of constructs with low closure frequencies is reflective of modeling results suggesting that shorter linkers

do not readily allow intramolecular BH3/BCL-xL interaction–do not fully satisfy Design Criterion 1. These results demonstrate that our computational framework can be applied to structurally diverse target proteins.




6

**Figure 2.6 | Computational design of inducible Rho Family GEFs. (a-b)** Retrospective evaluation of inducible ITSN (sITSN) constructs. **(a)** The relationship of μ and fold-activation of inducible synthetic ITSN

(sITSN) designs. **(b)** The BH3/BCL-xL complex mass centers for every trajectory for the indicated designs with BH3/BCL-xL mass centers shown as black spheres, overlaid on the structure of ITSN (cartoon, grey) bound to CDC42 (red) at its active site. **(c-f)** Computational design of an inducible VAV. **(c)** Average $\dot{\underline{r}}$ (in Å) with respect to N- and C-terminal linker length (NL and CL, respectively). **(d)** Standard deviation of $\dot{\underline{r}}$, as a function of NL and CL. **(e)** Closure frequency as a function of NL and CL. **(f)** Fold-activation of inducible VAV constructs. VAV onstruct nomenclature: V.(NL).(CL). **(a,f)** Fold-activation was determined using *in vitro* guanine nucleotide. Constructs were assayed in the presence of 5 μM competitor or 5 μM control BH3 peptides. Fold activation is the ratio of catalytic activity in the presence of a competitor divided by the catalytic activity in the presence of a control. Error bars = s.e.m, n = 3, except for $\dot{\underline{r}}$-mean in **(a)**.

## 2.4 Discussion

Synthetic intramolecular regulation was first demonstrated for N-WASP in cell-free systems[49,50]. Subsequently, empirically engineered Rho Family GEFs governed by synthetic intramolecular regulatory systems have been described[12,15,51]. Here, we report the design and development of CIAR, an intramolecularly regulated RAS rheostat. While SOScat and Rho Family GEFs have similar enzymatic activities, SOScat is larger, more complex, and structurally unrelated to the Rho Family GEFs (Supplementary Fig. A.16)[52], and a synthetic intramolecular regulatory system for SOScat has not been reported to date. To accomplish this, we developed a systematic computational framework for screening potential constructs for autoinhibition that effectively narrowed our design space. To the best of our knowledge, this is the first computational method for guiding the design of intramolecular regulation. The application of our computational framework to the design of inducible Rho Family GEFs, demonstrates that this methodology can be used to guide the engineering of intramolecular regulatory systems for topologically diverse proteins. While a single round of design and testing was sufficient for generating intramolecularly-regulated SOScat and VAV GEFs, our computational method is amenable to further optimization, including introduction of flexibility into the target protein and model refinement based upon empirical testing.

Such a computational approach can guide the engineering of synthetic regulatory regimes for a wide range of structurally and functionally diverse proteins.

In this study, our regulatory system co-opts the BCL-xL/BH3 peptide interaction to gate protein activity. From an engineering perspective, this protein-protein interaction is attractive because minimized constructs of each component are available that allow the BH3/BCL-xL complex to be treated as a single rigid body. Furthermore, the affinity of the BH3 peptide for BCL-xL can be tuned, allowing optimization of interaction strength[12]. Importantly, the small molecule (A-385358) used to disrupt the BH3/BCL-xL interaction appears to have minimal effects on endogenous signaling pathways (Fig. 2.5a). Furthermore, A-385358 is well-tolerated in mice[23], facilitating *in vivo* applications.

Our systematic approach for engineering autoinhibited protein rheostats was applied to RAS signaling. RAS functions as a binary switch, yet it is capable of controlling diverse and often conflicting cellular processes and phenotypes[21]. Bypassing the pleiotropic effects of growth factor stimulation and the caveats inherent to RAS overexpression methods, CIAR enables elucidation of such emergent phenomena through direct activation of endogenous RAS. Chemical genetic approaches, as with any cellular perturbation, may produce artifacts that confound interpretation of results. For this reason, we generated an inactive control, xCIAR, containing a single mutation within the SOScat active site. Comparison of results obtained with CIAR to xCIAR enabled confirmation that observed phenomena were indeed due to RAS activation rather than artifacts of small molecule treatment or localization of a chimeric transgene to the plasma membrane.

We found that the dynamics of ERK phosphorylation in CIAR-293s differ substantially between direct RAS activation and EGF stimulation, and the response to direct RAS activation also differs between cell lines. Direct RAS activation in CIAR-MCF10As elicits a rapid increase in phospho-ERK, whereas, in CIAR-293s, this increase is more gradual. Importantly, direct RAS activation via CIAR had no effect on AKT phosphorylation and did not elicit feedback activation of EGFR (Fig. 2.4c, Supplementary Fig. A.8). RAS•GTP is rapidly formed in CIAR-293s upon A3 addition (Fig. 2.4a), which indicates that the observed differences between EGF and direct RAS activation, as well as between cell lines, are due to factors acting on nodes downstream of RAS. Notably, RAF regulation has been implicated as a major factor governing RAS/ERK signaling dynamics. We reasoned that the ability of CIAR to divorce RAS activation state from the pleiotropic effects of growth factors would permit the interrogation of RAF regulation dynamics. By focusing our efforts on two RAF1 inhibitory phosphorylation sites, we found that disruption of a RAS/ERK extrinsic regulatory mechanism—mediated by factors not considered components of the canonical RAS/ERK module—had a dramatic impact on RAS/ERK signaling. These results demonstrate that the wiring of signaling components extrinsic to the RAS/ERK module can strongly influence RAS/ERK kinetics, even in the absence of pleiotropic signaling stimuli. Differences in RAS/ERK extrinsic basal wiring across cell types may contribute to the great disparities in their predisposition to oncogenic RAS-mediated transformation[53]. Furthermore, our PRM-MS results suggest RAS/ERK flux is transmitted preferentially through BRAF as compared to RAF1 in CIAR-293s in response to both EGF and direct RAS activation (Fig. 2.4b). Based on the rate of S365

dephosphorylation, it also appears that EGF stimulation is capable of more rapidly activating BRAF than direct RAS activation in these cells. The exact mechanistic explanation for this phenomenon remains unclear, but could involve differential modulation of phosphatases, kinases, and scaffolds[39,54,55].

We next utilized CIAR to investigate how the pharmacologic modulation of RAF affects RAS signaling. Despite clinical utility in BRAF(V600E)-driven melanomas, vemurafenib promotes the proliferation of tumors harboring oncogenic RAS and wild-type BRAF. This paradoxical activation is believed to be mediated by modulation of RAF homo- and heterodimerization and interactions with KSR and MEK[43,56,57]. In CIAR-293s and CIAR-MCF10As, we demonstrate that brief incubation with vemurafenib is sufficient to prime RAS/ERK signaling. This finding confirms that the paradoxical activation of RAS/ERK signaling by vemurafenib does not require transcriptional or translational reprogramming. Furthermore, we have shown for the first time that direct activation of wild-type RAS is sufficient to drive paradoxical activation. Previously, paradoxical activation in cells has been demonstrated in the context of oncogenic mutant RAS or upon upstream receptor activation[43,58,59]. We believe CIAR will enable further elucidation of the complex phenomenon of paradoxical activation by RAF inhibitors.

A major advantage of CIAR is the ability to evaluate the propagation of signal from RAS outward through phosphorylation networks. Quantitative phosphoproteomics revealed a class of phosphosites exhibiting dynamic behavior similar to ERK phosphorylation, while another class decreased in abundance rapidly in response to both direct RAS activation and EGF stimulation. Taken together with the investigation of RAF1 regulation, the

observed differences between these classes of phosphosites suggests that the RAS/ERK module is subject to greater extrinsic regulation than other RAS driven signaling processes, at least in the context of the HEK-293-derivative line used in this study. Notably, we did not identify any prominent peptide classes, which were perturbed only by direct RAS activation and not EGF, though the dynamic behaviors of some sites differed markedly between stimuli.

In summary, we report the computationally-guided development of CIAR, a RAS rheostat. We have demonstrated the utility of CIAR in interrogating the effects of growth factor stimulation, cell type, and pharmacological rewiring on RAS signaling dynamics. Moreover, our computational framework for engineering intramolecular regulation should be applicable to a wide range of signaling proteins, guiding the development of new chemical genetic and optogenetic technologies.

Acknowledgements: This research was supported by the NIH (R01GM086858 (D.J.M), R01CA126792 (J.D.), and F30CA189793 (J.C.R.)), the National Science Foundation CAREER Award (CHE- 0954242), the DOD BRCP (W81XWH-11-1-0130), and HHMI (P-S.H and D.B.).

## Chapter 3: DSB-ddPCR and a rapidly inducible Cas9 variant enable exploration of editing kinetics

A version of this chapter is currently in review at *Nature Methods.*

John C. Rose, Jason J. Stephany, William J. Valente, Bridget M. Trevillian, Ha V. Dang, Miles S. Dickinson, Jason H. Bielas, Dustin J. Maly, Douglas M. Fowler

J.C.R., D.M.F, and D.J.M conceived the study and designed the experiments. B.M.T assisted in design of ciCas9. J.C.R. and J.J.S. performed the experiments. M.S.D. provided assistance with experiments. J.J.S. prepared samples for high-throughput sequencing and ddPCR. H.V.D. performed fluorescence BH3 competition experiments under the supervision of J.C.R. and D.J.M. W.J.V helped develop the DSB-ddPCR assay and analyze data under the supervision of J.H.B. J.C.R and D.M.F prepared custom Python scripts. J.C.R., D.M.F and D.J.M wrote and edited the manuscript. All authors approved the final manuscript.

## 3.1 Abstract

To investigate the kinetics of Cas9-mediated double strand break generation and repair *in vivo*, we developed two new tools. The first, chemically inducible Cas9 (ciCas9), is a rapidly-activated, single-component Cas9 variant engineered using a novel domain replacement strategy. ciCas9 can be activated in a matter of minutes, and the level of ciCas9 specificity and activity can be tuned. The second tool, DSB-ddPCR, is a droplet digital PCR-based assay for double strand breaks. DSB-ddPCR is the first assay to demonstrate time-resolved, highly quantitative and targeted measurement of DSBs. Combining these tools facilitated an unprecedented exploration of the kinetics of Cas9-mediated DNA cleavage and repair. We find that sgRNAs targeting different sites generally produce cleavage within minutes and repair within an hour or two. However, we observe distinct kinetic profiles, even for proximal sites, suggesting that target sequence and chromatin state modulate cleavage and repair kinetics.

## 3.2 Introduction

The RNA-guided endonuclease Cas9 is capable of programmable, targeted genome editing. Cas9 can yield insertion or deletion mutations by first generating a double strand break, which is subsequently repaired by the endogenous DNA repair machinery of the cell (Fig. 3.1a). Insights into the processes of cleavage and repair have enabled the manipulation of Cas9, sgRNA, or experimental conditions to produce improvements in editing efficiency, specificity and repair[60-64]. Nevertheless, large gaps remain in our understanding of the Cas9 editing process. In particular, little is known about the kinetics of Cas9-mediated DSB generation and repair *in vivo*. *In vitro,* the Cas9/sgRNA

complex remains bound to both ends of cleaved DNA for approximately 6 hours, which may delay surveillance and repair[60,65]. In cells, progress has been made in characterizing the target search dynamics of nuclease-null Cas9[66]. How these findings relate to cleavage and repair kinetics of nuclease-competent Cas9 *in vivo* remains unclear.

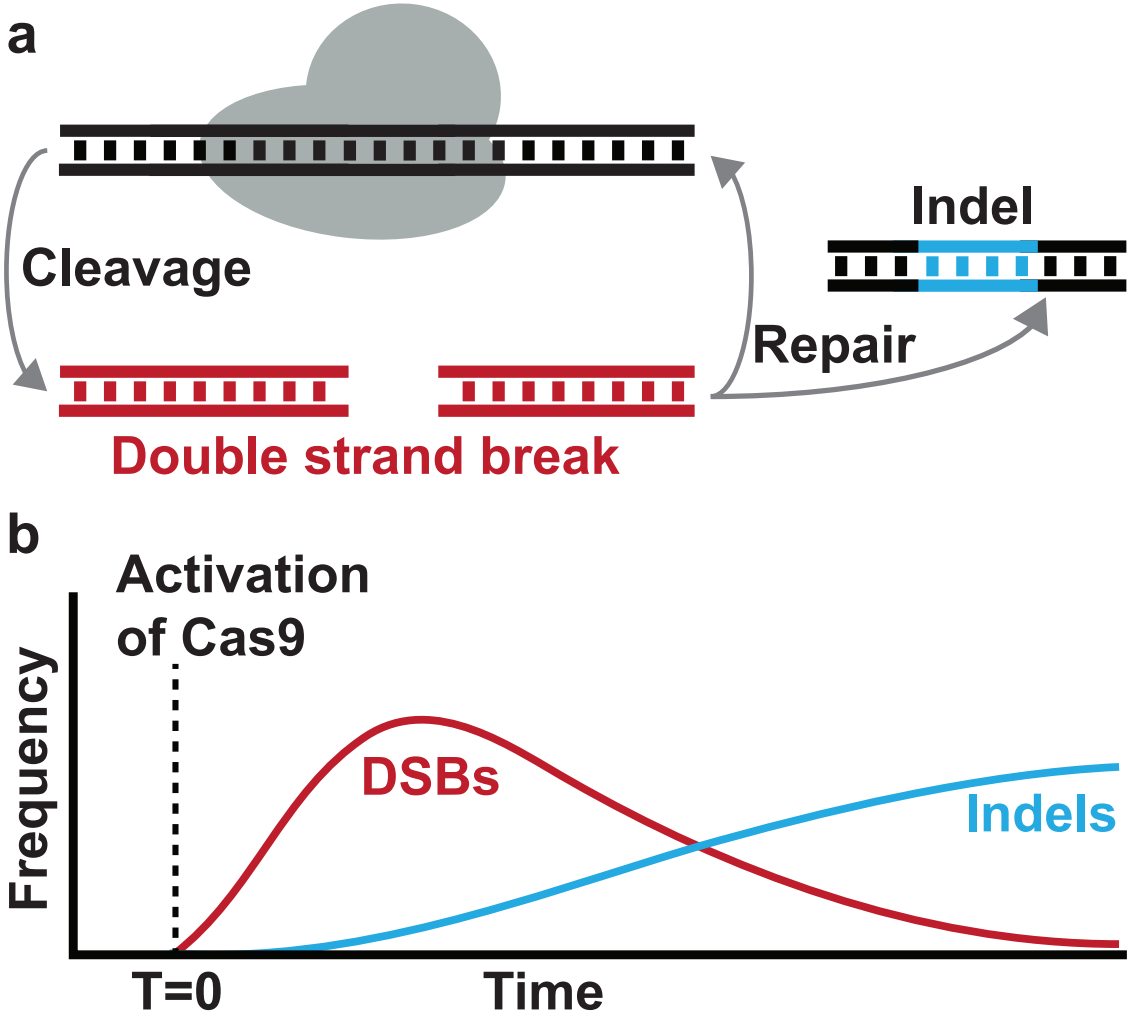



**Figure 3.1 | Schematic of Cas9-mediated DNA cleavage and repair. (a)** Cas9 (gray) is an RNA-guided endonuclease that cleaves DNA to produce double strand breaks (DSBs). DSBs are recognized and repaired by endogenous DNA repair machinery. Repair by non-homologous end joining can create insertions or deletions, but can also be error-free. **(b)** The tools we developed enable a time course experiment, where Cas9 activity is activated at a defined time point. Then, the frequency of both DSBs and indels are measured over time.

Investigating Cas9 cleavage and repair kinetics in cells requires two methodological advances. First, rapid and precise temporal control of Cas9 activity is required, so that cleavage can be activated at a defined point in time (Fig. 3.1b). With reported activation times ranging from two to many hours, it is unclear whether existing inducible Cas9 variants become active rapidly enough (Supplementary Table 1)[67,68]. Second, a precise and temporally-resolved method for quantitation of DSBs is required, so that DNA cleavage can be measured in parallel with DNA editing. Existing methods either qualitatively or semi-quantitatively evaluate DSBs on a locus-specific or genome-wide basis, but none of these methods are suitable for tracking Cas9 cleavage kinetics (Supplementary Table 2).

To overcome these limitations, we developed two new tools. The first, chemically inducible Cas9 (ciCas9), is a rapidly-inducible, single-component Cas9 variant engineered using a novel domain replacement strategy. We show that ciCas9's level of basal activity is low, and that it can be activated within minutes in a dose-dependent fashion. We also demonstrate that the architecture of ciCas9 allows the generation of variants with increased specificity for on-target sites or even lower basal activity. The second tool, DSB-ddPCR, is a droplet digital PCR-based assay for double strand breaks. DSB-ddPCR is the first assay to demonstrate time-resolved, highly quantitative and targeted measurement of DSBs. The combined application of these tools facilitated an unprecedented exploration of the kinetics of Cas9-mediated DSB generation and repair. We find that DSBs are generated rapidly, within ten minutes for some sgRNAs, and that indels generally appear within an hour or two. sgRNAs targeting different sites

produce distinct DNA cleavage and repair kinetics, even when the sites are proximal. These findings suggest that target sequence and chromatin state modulate cleavage and repair kinetics.

## 3.3 Results

### 3.3.1 Generation of a rapidly inducible Cas9 variant

To generate a single-component, chemically inducible Cas9 (ciCas9) with rapid activation kinetics, we used an intramolecular autoinhibition approach. In this approach, the protein-protein interaction between BCL-xL and a BH3 peptide functions as an autoinhibitory switch that blocks Cas9 activity[69]. Disruption of the BCL-xL/BH3 interaction by addition of a small molecule, A-385358 ("A3"), results in release of autoinhibition and activation of Cas9 (Fig. 3.2a). Fusion of BCL-xL and BH3 to Cas9's termini to achieve autoinhibition seemed unpromising because both termini are close together (Fig. 3.2b). Instead, modeling suggested the non-essential Cas9 REC2 domain could be replaced by BCL-xL, which is similar in size (Fig. 3.2b, Supplementary Fig. B.1). Fusion of BH3 to either terminus would likely result in formation of a BCL-xL/BH3 complex that prevents binding to guide RNA or DNA. Thus, we replaced the REC2 domain with BCL-xL, creating Cas9.BCL (Fig. 3.2c). Assessment of editing by high-throughput sequencing in HEK-293T cells revealed that Cas9.BCL was expressed and retained activity (Supplementary Fig. B.2), as has been reported for Cas9 containing a REC2 domain deletion[70].

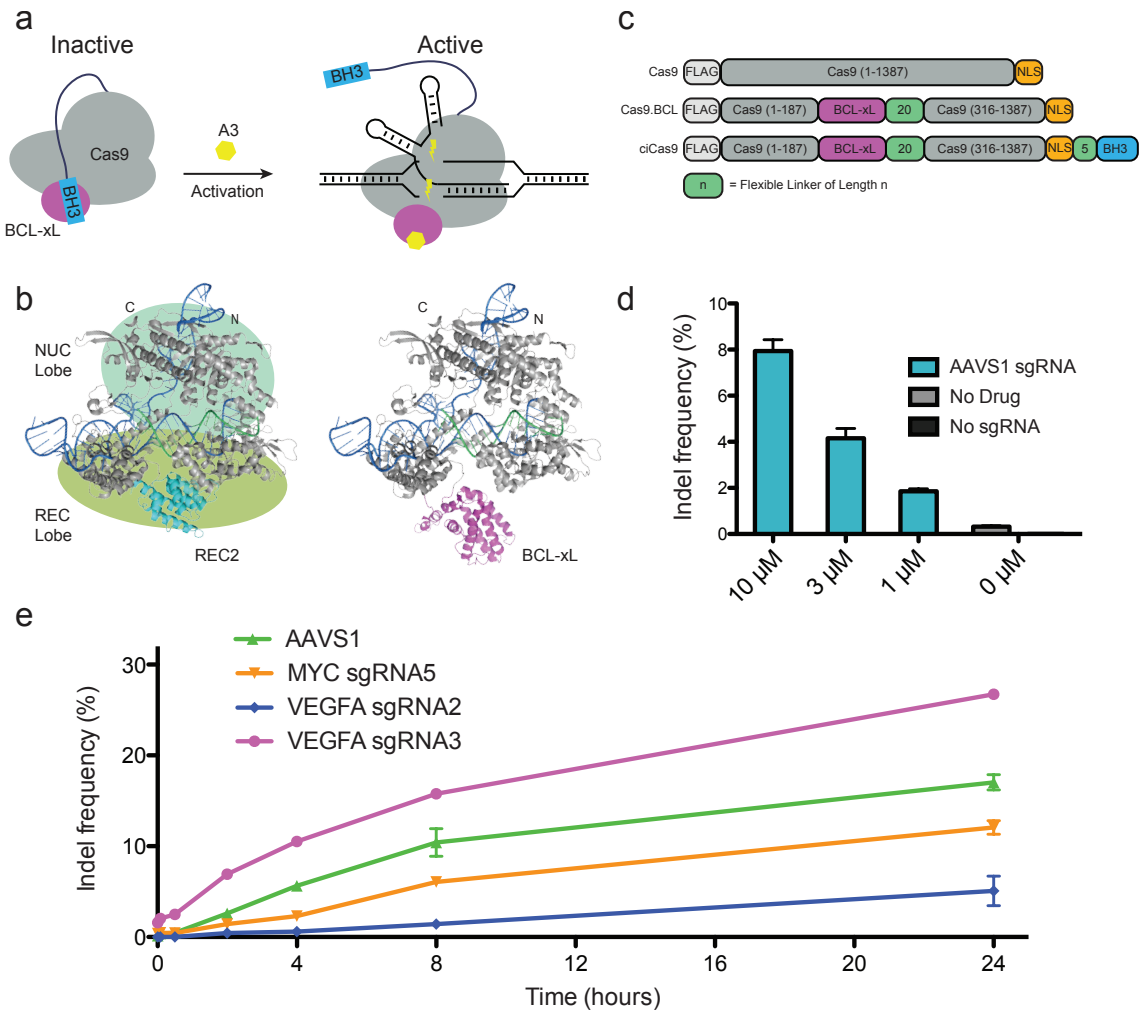



**Figure 2 | Development of a chemically inducible Cas9 (ciCas9). (a)** A schematic depiction of the strategy to engineer a single-component, chemically inducible *Streptococcus pyogenes* Cas9 variant is shown. **(b)** The REC2 domain was replaced with BCL-xL (left panel, PDB: 4OO8, right panel, model based on structure), and **(c)** a BH3 peptide was appended to the C-terminus via flexible linkers of varying lengths. **(d)** Indel frequency at the AAVS1 locus 24 hours after activation of ciCas9 activity is shown for different concentrations of A3. Error bars depict s.e.m. (n = 3 biological replicates). **(e)** Indel frequency at different times following activation of ciCas9 with A3 is shown for four sgRNAs at three different loci. Error bars depict s.e.m. (n = 3 biological replicates).

Having confirmed BCL-xL can replace the REC2 domain, we appended a BH3 peptide to either terminus of Cas9 via linkers of 5 to 30 residues (Supplementary Fig. B.3). The activity of each construct was assessed at the AAVS1 locus in the absence or presence of A3. We selected the AAVS1 locus because it is a “safe harbor” where alterations do not affect cell growth or other phenotypes. High-throughput sequencing revealed that C-

terminal BH3 fusions resulted in A3-activated editing, (Supplementary Fig. B.4) while N-terminal fusions were active in the absence of drug (data not shown). A range of shorter linker lengths yielded comparable fold activation, so we selected the C-terminal BH3 fusion with the shortest, five-residue linker for further study. This construct is subsequently referred to as ciCas9.

Expression of ciCas9 in the absence A3 resulted in minimal editing (0.32%, s.e.m. = 0.039%), compared to a no sgRNA control (0.0033%, s.e.m. = 0.0010%) (Fig. 3.2d). Addition of A3 generated a dose-dependent increase in editing, meaning activity can be tuned by varying drug concentration. At the highest concentration of A3 tested, 10 μM, we observed a 24.7-fold (s.e.m. = 3.34) increase in editing compared to the no drug control. This degree of activation is considerably higher than reported for other inducible Cas9 systems[67,71]. To demonstrate the generality of ciCas9, we assessed editing in HCT116 and U2OS cells. We observed A3-dependent editing, with minimal activity in the absence of drug (Supplementary Fig. B.5). Editing via transient transfection is inefficient in these other cell lines[72], as has been previously reported for U2OS and HCT116 cells using the same sgRNA[73].

### 3.3.2 ciCas9 activation rapidly produces indels

We next profiled ciCas9 indel kinetics with four sgRNAs at three distinct loci. (Fig. 3.2e). AAVS1 and VEGFA sgRNA3 showed significant increases in indels within 30 minutes of A3 addition, and a significant increase in indels was present at all sites by two hours (Supplementary Tables 3, 4). These results indicate that editing kinetics differ between

sgRNA target sites at distinct loci. Kinetics also differ between sgRNA target sites at the same locus. The two VEGFA cleavage sites are separated by only 1,091 nucleotides but have markedly different editing rates and total edits at 24 hours. These differences could be due to sequence and chromatin features that are known to influence editing[74-76]. For example, VEGFA sgRNA3 has higher guanine content than VEGFA sgRNA2 (55% vs. 10%) and a higher predicted efficacy score based on the target and surrounding sequence (0.49 vs. 0.29, Azimuth 2.0 on-target scoring model)[77].

We next compared ciCas9 kinetics to other inducible systems (Supplementary Table 1). Some of these systems are much slower. For example, indels were reported after nine hours following activation of photoactivatable split Cas9 (paCas9)[68]. After activation of ciCas9, indels were detectable by two hours using the same sgRNA and cell line (Supplementary Fig. B.6). To facilitate a comparison between ciCas9 and a recently described 4-hydroxytamoxifen-activated Cas9 (iCas9)[67] that is reported to be faster than intein-Cas9[71] and split-Cas9[78], we measured ciCas9 editing kinetics using the same sgRNA. We found that ciCas9 generated a detectable increase in indels within 30 minutes (one-sided t-test, $n = 3$, $p = 0.016$) of A3 addition. The first reported time point for iCas9 was two hours, but it is unclear if the reported increase in indels is significant because no statistical analyses were performed (Supplementary Fig. B.7).

### 3.3.3 DSB-ddPCR enables precise, site-specific quantitation of DSBs

*In vivo,* Cas9-mediated cleavage of DNA target sites is generally measured indirectly, by quantifying indels. To directly and quantitatively measure DSBs at a target site, we

developed a droplet digital PCR assay, DSB-ddPCR. DSB-ddPCR uses a primer pair that spans the sgRNA target site and another proximal pair that does not (Fig. 3.3a). Amplification is detected using primer pair-specific probes (FAM-labeled probe: uncleaved target present; VIC-labeled probe: template present). Droplets containing an uncleaved or repaired template yield products for both primer pairs (target+, template+). Droplets containing a template that has been cleaved by ciCas9 but not yet repaired are negative for the target amplicon (target-, template+).

We evaluated the performance of DSB-ddPCR by generating cleaved control DNA using restriction endonucleases. For each of three loci, we employed endonucleases that cleaved the template within the first primer pair, simulating a Cas9-mediated DSB. As expected, uncleaved control DNA yields probe signal for both amplicons, whereas fully digested control DNA yields signal only for the template+ amplicon (Fig. 3.3b). To investigate the accuracy of DSB-ddPCR, we mixed cleaved and uncleaved control DNA in defined ratios. The relationship between observed and expected DSB frequency is linear (Pearson's $r^2 > 0.99$; Fig. 3c). DSB-ddPCR is also precise: technical replicates are essentially indistinguishable (Fig. 3.3c). Thus, DSB-ddPCR enables accurate, precise quantitation of DSBs.

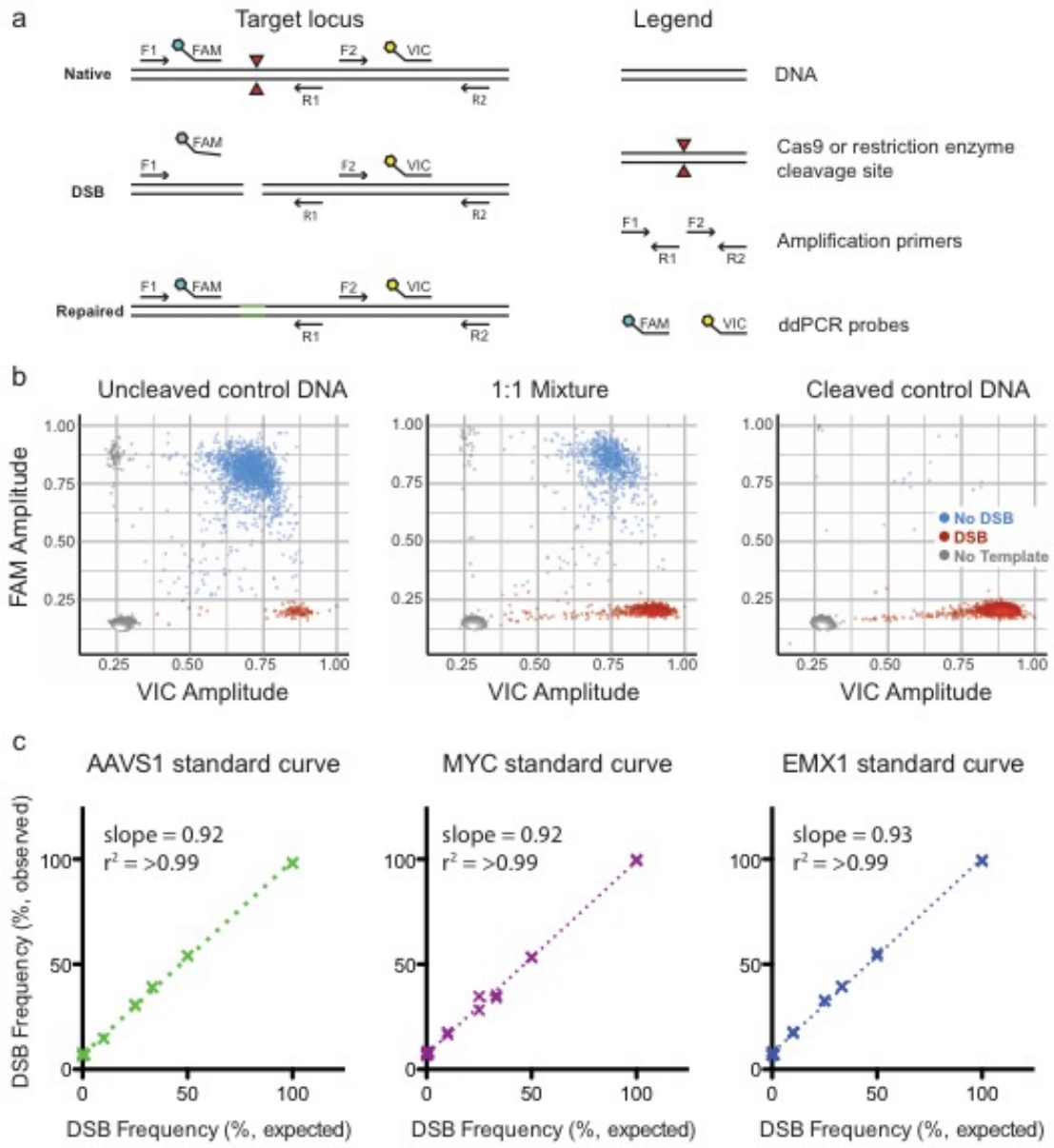



**Figure 3.3 | DSB-ddPCR is an accurate and precise way to quantify DSBs. (a)** A schematic illustrates the DSB-ddPCR assay. The assay relies on two locus-specific PCR amplicons, the first of which spans the cleavage junction (primers F1, R1) and the second of which is adjacent (primers F2, R2). Amplicons are detected by the activation of amplicon-specific fluorescent hydrolysis probes. Template that has been cleaved by Cas9 or another nuclease does not generate the first amplicon or activate the first probe. Uncleaved control DNA was digested with a restriction enzyme to create cleaved control DNA. Uncleaved and cleaved control DNA was mixed in specific proportions. **(b)** Representative droplet FAM and VIC probe intensities are shown for uncleaved control MYC locus template DNA (left panel), a 1:1 mixture of uncleaved:cleaved control DNA (middle panel), and cleaved control DNA (right panel). Colors indicate droplets with no template (gray), intact template (blue), and cleaved template (red). Droplet populations in the dot plots are: uncleaved control DNA (DSB: 120, no DSB: 1,724, no template: 12,360), 1:1 Mixture (DSB: 799, no DSB: 914, no template: 12,305), cleaved control DNA (DSB: 1,757, no DSB: 14, no template: 13,247). **(c)** Standard curves show the relationship between DSB frequency as measured by the DSB-ddPCR assay and expected DSB frequency at the AAVS1, MYC and EMX1 loci. Technical duplicates are shown and, in most cases, overlap.

We next compared DSB-ddPCR to other methods for detecting DSBs (Supplementary Table 2). Immunodetection methods and the comet assay quantify overall DSB levels, but do not reveal where DSBs occur[79,80]. Sequencing-based methods like BLESS, DSB-capture and GUIDE-Seq identify DSB locations genome-wide, but are complex, expensive, and semi-quantitative[81-83]. qPCR can reveal the prevalence of DSBs at a locus of interest relative to a distal control locus[84]. However, absolute quantitation of DSBs, as we show for DSB-ddPCR (Fig. 3.3c), has not been demonstrated. Furthermore, ddPCR offers several inherent advantages over qPCR, including absolute quantification without references, less sensitivity to PCR efficiency, greater precision and improved reproducibility[85]. Thus, our DSB-ddPCR assay represents a new and more quantitative way to detect DSBs at specific loci.

### 3.3.4 Characterization of DNA cleavage and repair kinetics

Cas9-mediated editing depends on DSB generation and subsequent repair, yet the temporal relationship between these two processes is unclear. To explore this relationship, we conducted time course experiments using ciCas9 to generate DSBs, DSB-ddPCR to quantify DSBs and sequencing to quantify indels (Fig. 3.4a). To ensure that protein-DNA complexes, including ciCas9-bound DNA, did not interfere with DSB quantitation, we implemented an hour-long proteinase K digestion (Supplementary Fig. B.8). Similar digestions are often used to remove Cas9 from DNA for sequencing, genome-wide DSB detection, or T7E1, SURVEYOR, or *in vitro* Cas9 cleavage assays[67,75,81,86]. We first examined DSB kinetics at the AAVS1, EMX1 and MYC loci (Fig. 3.4b, Supplementary Fig. B.9). In this first-ever look at the kinetics of Cas9-

mediated DSB generation *in vivo*, we observe some features that are common to all loci. For example, DSBs appear relatively rapidly, and reach a maximum well before the final 24 hour time point. In addition, at all three loci, a large number of unresolved DSBs are present at 24 hours. However, each locus has a distinct DSB kinetic profile. For example, at the AAVS1 locus, a maximum of 19.2% (s.e.m. = 1.8%) of templates have DSBs after two hours, whereas at the EMX1 locus, 26.2% (s.e.m. = 4.9%) of templates have DSBs after eight hours.

We next examined the relationship between DSB and indel appearance at these three loci (Fig. 3.4c). In all cases, indels are generated more slowly than DSBs, appearing in an approximately linear fashion over the first eight hours after ciCas9 activation. However, the rate of indel appearance is comparatively slow at the AAVS1 locus, which might be explained by the lower frequency of DSBs. The MYC and EMX1 loci share similar indel kinetics, but exhibit distinct DSB kinetics. For example, most indels have appeared at both loci by eight hours. However, during this period, DSB frequency declines appreciably at the MYC locus whereas it continues to increase at the EMX1 locus. The precise meaning of these different kinetic profiles remains unclear, but, at least for these three loci, Cas9 cleavage and DNA repair rates appear to differ.

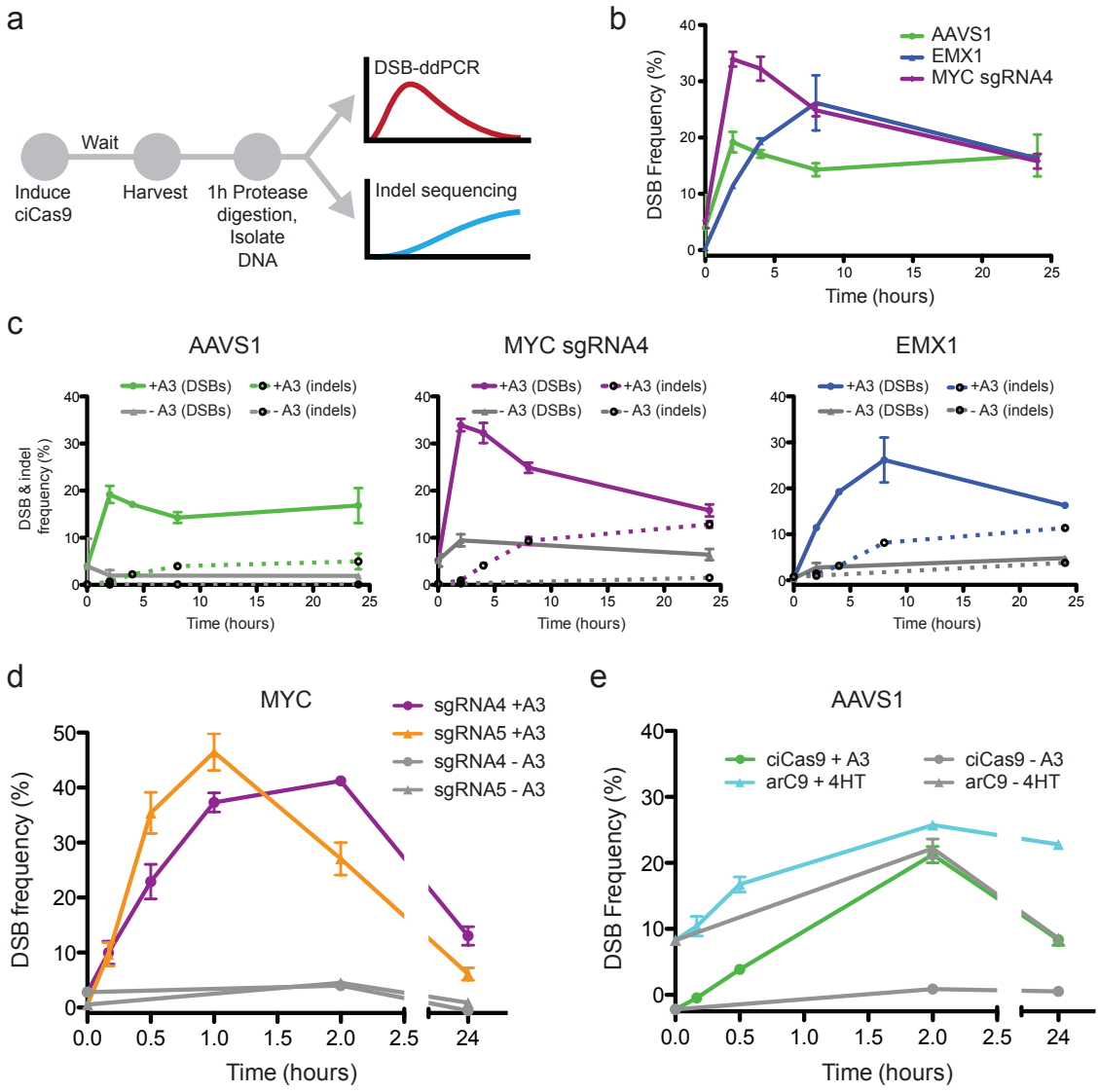



**Figure 3.4 | Investigation of DSB and indel kinetics with ciCas9 and DSB-ddPCR. (a)** ciCas9 was activated by incubating cells with A3 for varying amounts of time. Cells were harvested, lysed, and lysates were treated for one hour with proteinase K at 56ºC. Genomic DNA was isolated; DSBs were quantified using DSB-ddPCR and indels were quantified by high-throughput sequencing. **(b)** A time course of the frequency of DSBs at three loci following activation of ciCas9 is shown. Error bars depict s.e.m. (n = 3 biological replicates). **(c)** Time courses of the frequency of DSBs and indels at the three loci following activation of ciCas9 are shown; the (+A3) DSB data is the same shown in panel b. Solid lines indicate DSB frequency, dashed lines indicate indel frequency. Error bars depict s.e.m. (n = 3 biological replicates). **(d)** A DSB time course focused on the first two hours following ciCas9 activation is shown for two sgRNAs targeting the MYC locus. Error bars depict s.e.m. (n = 3 biological replicates). **(e)** A DSB time course comparing ciCas9 to a 4-hydroxytamoxifen-activated Cas9 variant, arC9, is shown. Error bars depict s.e.m. (n = 3 biological replicates).

To investigate events occurring shortly after activation of ciCas9, we performed high-resolution time courses using two adjacent sgRNAs targeting the MYC locus (Fig. 3.4d).

Here, we observe appreciable generation of DSBs for both sgRNAs after ten minutes, the earliest time point we evaluated (one-sided t-test, n = 3; MYC sgRNA4 p = 0.014; MYC sgRNA5 p = 0.007). Interestingly, although MYC sgRNA5 cleaves only ten nucleotides downstream of sgRNA4 on the same strand, we observe distinct kinetic profiles at each target site. Over the first 30 minutes, cleavage with MYC sgRNA5 is approximately twice as fast as with MYC sgRNA4 (sgRNA5 cleavage = 1.2 ± 0.1 percent per minute; sgRNA5 = 0.66 ± 0.09; Supplementary Fig. B.10). DSB frequency subsequently peaks after one hour at the sgRNA5 site. DSB frequency peaks after two hours at the sgRNA4 site, by which time DSB frequency at the sgRNA5 site has already declined considerably. Despite the greater initial DSB generation rate, higher maximum and earlier decline in DSB frequency, sgRNA5 yields fewer indels after 24 hours than does MYC sgRNA4 (Supplementary Fig. B.11). One possible explanation for this difference between sgRNAs is that repair kinetics are distinct. Another possibility is that error-free repair is more prevalent for DSBs generated by sgRNA5, yielding fewer edits. Given the proximity of these guides, these differences in apparent kinetics and repair outcomes might inhere to differences in the guide sequences themselves, or the immediately adjacent sequence. Finally, loss of MYC impairs cell proliferation[87], and both sgRNAs target the 5' UTR of this gene. Thus, a third possibility is that the differences may relate to proliferation, although any such effects are unlikely to appreciably impact early time points.

We next compared the DSB generation kinetics of ciCas9 to another inducible Cas9 variant, arC9[45-47]. We posited that arC9 might also be rapidly activatable because, like

ciCas9, it is a single-component system activated by a small molecule, 4-hydroxytamoxifen. We directly compared the kinetic profile of ciCas9- and arC9-mediated DSB generation at the AAVS1 locus (Fig. 3.4e). We observe that ciCas9, but not arC9, yields a small but significant increase in DSBs after ten minutes (one-sided t-test, n = 3; ciCas9, p = 0.011; arC9, p = 0.13). arC9 also suffered from much higher basal activity and lower fold-activation. This high basal activity complicates kinetic analysis, since considerable arC9-mediated cleavage occurs prior to addition of 4-hydroxytamoxifen. Despite the differences in basal activity, the DSB profiles generated by the two systems share some common features. For example, both arC9 and ciCas9 have similar absolute DSB frequencies after two hours, although the increase relative to the unactivated control is much larger for ciCas9 than for arC9. Additionally, in both cases, a large number of DSBs remain unresolved after 24 hours.

### 3.3.5 Tuning ciCas9 specificity, basal activity and chemical activation

To increase the utility of ciCas9 for different applications, we explored the possibility of further tuning its performance. We first focused on editing specificity. To determine the specificity of ciCas9 relative to wild type Cas9, we measured specificity ratios (i.e. on-target:off-target indel frequency ratios) for four sgRNAs at a total of five off-target sites. We found that ciCas9 had a higher specificity ratio than wild type Cas9 at four of five on-target:off-target pairs (Figure 5a, Supplementary Fig. B.12). Despite the improved specificity of ciCas9 over wild type Cas9, off-target activity was still detectable at all off-target sites tested. To further improve on-target specificity, we introduced previously-identified, specificity-enhancing mutations[48] to yield enhanced-specificity ciCas9 (e-

ciCas9). e-ciCas9 had no detectable off-target activity for two of three on-target:off-target pairs, dramatically improving the specificity ratios (Fig. 3.5a, Supplementary Figure B.13). For the third on-target:off-target pair, off-target editing was reduced but still detectable. Thus, at these sites, ciCas9 is more specific than wild type Cas9, and e-ciCas9 improves specificity even further.

Next, we focused on minimizing the basal activity of ciCas9 by tuning the strength of the BCL-xL/BH3 peptide interaction. The BH3 peptide we incorporated into ciCas9 has a binding affinity ($K_i$) for BCL-xL of ~50 nM, and we reasoned that strengthening this autoinhibitory interaction would reduce basal activity. We generated a panel of BH3 variants with higher affinity for BCL-xL, with $K_i$s ranging from 40 nM to <0.5 nM, by mutating the G22 residue of the BH3 peptide (Fig. 3.5b, Supplementary Fig. B.14, Supplementary Table 5). Two of the highest-affinity ciCas9 BH3 variants, ciCas9(L22) and ciCas9(F22), demonstrated significantly lower basal editing at the AAVS1 locus relative to ciCas9 (Fig. 3.5c; one-sided t-test; n = 3; L22, p = 0.044; F22, p = 0.039). For both variants, basal editing in the absence of A3 was indistinguishable from the no sgRNA control (one-sided t-test; n = 3; L22, p =0.35; F22, p = 0.29). However, the level of A3-activated editing for both the ciCas9(L22) and ciCas9(F22) was reduced relative to ciCas9. Thus, strengthening the BCL-xL/BH3 interaction increases ciCas9 autoinhibition and reduces basal activity, which comes at the cost of reduced activation by A3.

One concern regarding the use of small molecules that target BCL-xL, a member of the anti-apoptotic BCL-2 family, is the potential for toxicity through the release of pro-apoptotic factors. However, due to functional redundancy among BCL-2 family members, effects on cell viability and proliferation are generally observed only when anti-apoptotic BCL-2 function is disrupted globally[48]**.** For example, ABT-737 is a well-characterized disruptor that potently targets BCL-xL, BCL-2, and BCL-w, and is well-tolerated in mice for extended durations. Furthermore, ABT-737 is generally only toxic as a single agent in cell lines that are deficient in the BCL-2 family member MCL-1, which is highly expressed in most widely-used cell lines[27]. We chose A3 for this study because it is even more selective for BCL-xL over BCL-2 and MCL-1 than ABT-737. BCL-xL-selective disruptors like A3 are well tolerated in mice, with reversible thrombocytopenia being the sole observed phenotypic effect[12]. Furthermore, we confirmed that A3 does not sensitize cells to the added stress of Cas9-mediated cleavage, and has no effect on the level of DSBs or indels generated by wild type Cas9 (Supplementary Fig. B.15). Thus, ciCas9 and A3 could be used to investigate Cas9-mediated cleavage and repair in most cell lines or in mice.

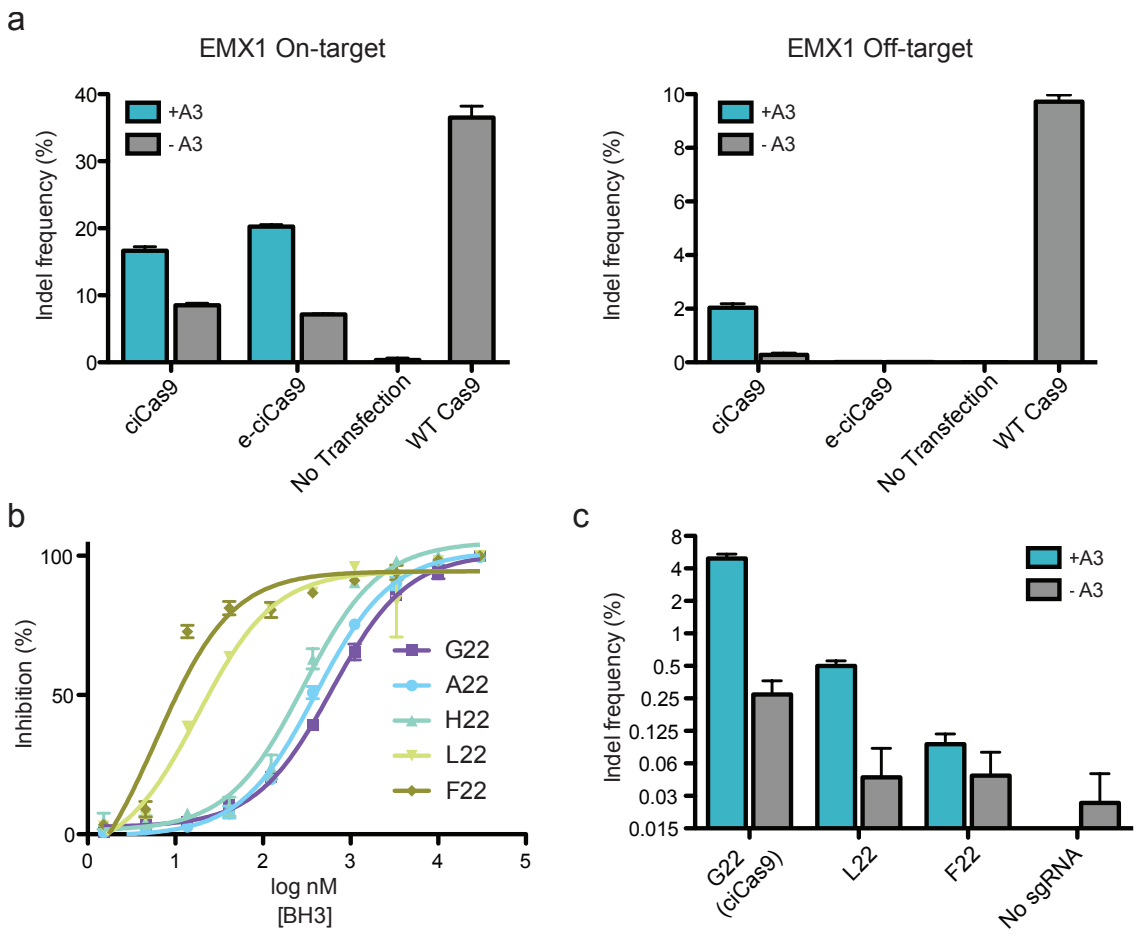



**Figure 3.5 | ciCas9 specificity and basal activity can be tuned. (a)** Editing is shown after 24 hours at an EMX1 on-target site (left panel) and an off-target site (right panel). ciCas9 and an enhanced-specificity variant (e-ciCas9) in the presence and absence of A3 are compared to wild type Cas9. Off-target editing with e-ciCas9 was not significantly increased relative to no transfection control (one-sided t-test, n = 3, p = 0.21). Error bars depict s.e.m. (n = 3 biological replicates). **(b)** Fluorescence polarization competitions between BH3 peptide variants and BODIPY-labeled BAK peptide for binding to BCL-xL. Data shown as inhibition of BODIPY-BAK binding. Error bars depict s.e.m. (n = 3 technical replicates). **(c)** Editing at the AAVS1 locus is shown for ciCas9 and two variants, L22 and F22, after 24 hours in the presence and absence of A3. To highlight low-frequency editing in the absence of drug, the indel frequency is shown on a logarithmic scale. Error bars depict s.e.m. (n = 3 biological replicates).

Finally, we investigated the effectiveness of different BH3/BCL-xL disruptors to activate ciCas9-mediated editing. Because disruptors of BCL-2 family member interactions are being developed for use in the clinic, compounds with improved selectivity and drug-like properties continue to appear[49,50]. We evaluated the ability of two other BCL-2 family member disruptors, ABT-737 and WEHI-539[12,15,51], to activate ciCas9-mediated editing

at the AAVS1 locus (Fig. 3.6a). Both ABT-737 and WEHI-539 resulted in robust editing at lower concentrations than A3, which is consistent with their increased cellular potency against BCL-xL (Fig. 3.6b). The enhanced potency of WEHI-539 also facilitated editing with the most tightly autoinhibited ciCas9 variant, ciCas9(F22), which was not appreciably activated by A3 (compare Fig. 3.6c, Fig. 3.5c). WEHI-539's greater selectivity for BCL-xL and efficacy in activating ciCas9 makes it an attractive candidate for use with our more tightly autoinhibited ciCas9 variants in cell culture models where low basal activity is required. While unfavorable pharmacokinetic properties preclude the use of WEHI-539 in animal models, derivatives like A-1155463 and A-1331852 have comparable BCL-xL affinity and specificity, and can be used in mice[52].

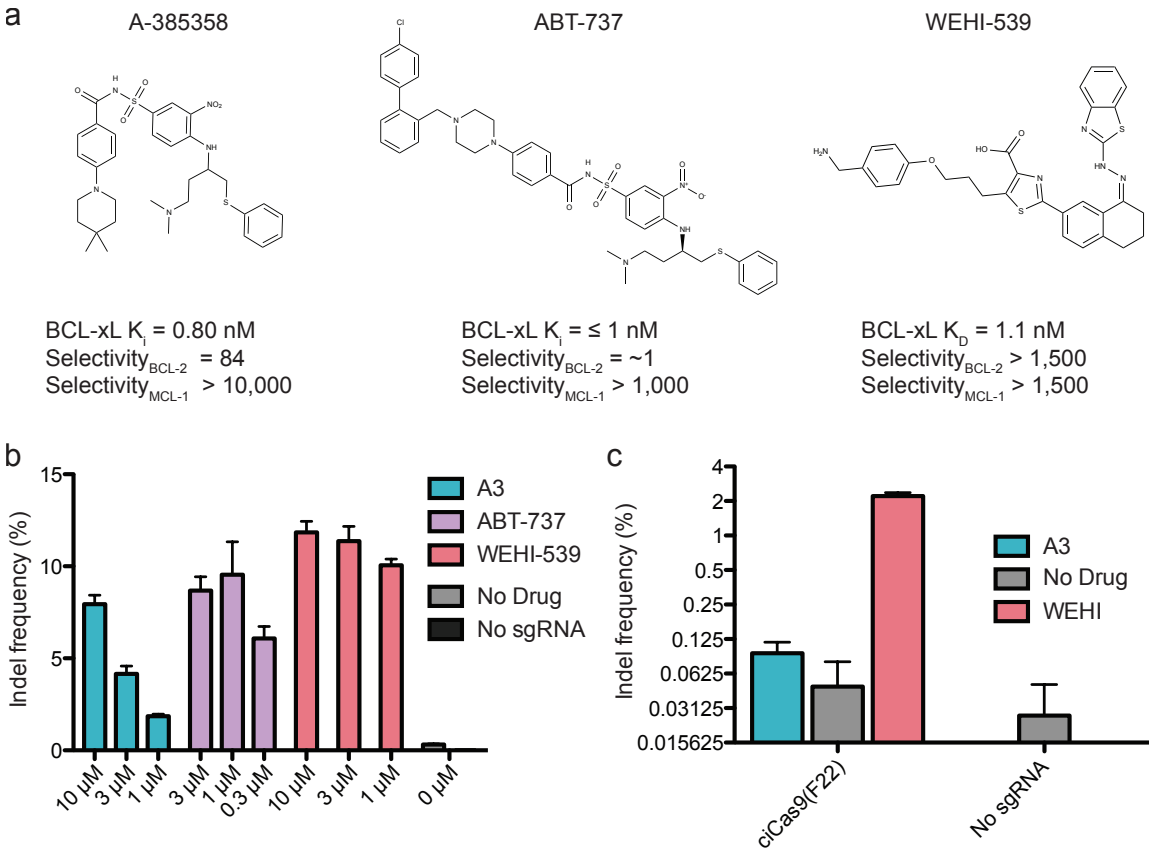



**Figure 3.6 | ciCas9 can be activated by a variety of BCL-xL disruptors. (a)** Structures of the three BCL-xL disruptors used in this study are shown with their reported $K_d$ or $K_i$ for BCL-xL, and their selectivity for BCL-xL over BCL-2 and BCL-xL over MCL-1[43,56,57]. **(b)** Editing at the AAVS1 locus 24 hours

after ciCas9 activation with different concentrations of the three disruptors is shown. The A3 data are also shown in Fig. 2d. Error bars depict s.e.m. (n = 3 biological replicates). **(c)** Editing is shown at the AAVS1 locus 24 hours after activation of the ciCas9(F22) variant with 10 μM of each disruptor. To highlight low-frequency editing in the absence of drug, the indel frequency is shown on a logarithmic scale. Error bars depict s.e.m. (n = 3 biological replicates).

## 3.4 Discussion

We describe the development of two complementary tools: a rapidly chemically inducible Cas9 variant and a quantitative, site-specific assay for DSBs. To develop ciCas9, we employed a novel Cas9 engineering strategy, intramolecular autoinhibition facilitated through REC2 domain replacement. We anticipate that this strategy will complement other Cas9 engineering approaches like insertion scanning[12], and will greatly expand our ability to control Cas9 activity. ciCas9 demonstrated faster activation kinetics than other inducible Cas9s, enabling us to create double strand breaks in a temporally resolved manner. ciCas9 can also be rationally engineered to increase on-target specificity or reduce basal activity. These properties, along with other advantages like small size and high specificity, suggest that ciCas9 will be broadly useful.

We also developed DSB-ddPCR, an assay to precisely quantify DSBs at targeted sites with high temporal resolution. DSB-ddPCR is the first application of ddPCR to measure DNA cleavage, and could be used to monitor other causes of strand disruption. By combining rapid, ciCas9-mediated DNA cleavage and DSB-ddPCR, we were able to investigate the previously unexplored kinetic relationship between Cas9-mediated DNA cleavage and repair. We found that DSBs are generated in a matter of minutes, and indels appear in as little as half an hour. While further work is needed to precisely quantify DSB repair rates, our results suggest that repair occurs more rapidly than the

long half-life of the Cas9-DSB complex measured *in vitro* might imply[23]. We also observe an excess of DSBs relative to indels, which might be due to error-free DSB repair[21].

Our tools give a tantalizing first look at the kinetics of Cas9-mediated DSB generation and repair. However, many challenges remain. In addition to the action of ciCas9 and DNA repair machinery, other processes influence the frequency of DSBs and indels we observed. For example, cell proliferation during these experiments may be impacted by transfection or increasing density, which in turn could affect cleavage and repair kinetics. Disruption of some loci, such as MYC, might influence cell proliferation and thus the frequency of DSBs and indels. These additional processes could be accounted for either experimentally or by modeling, which would enable collection and interpretation of longer time courses. A Cas9 variant that could be activated and inactivated rapidly would enable pulsed DNA cleavage. Pulsed cleavage would facilitate precise estimation of DSB repair rates, and could reveal the prevalence of error-free repair. Furthermore, chemical or genetic perturbation of specific DNA-repair pathways could yield mechanistic insight into cleavage and repair kinetics. For example, TP53 is inhibited by SV40T in HEK-293T cells[53]; examination of repair kinetics in the presence of normal TP53 activity could add to our understanding of this important protein. Ultimately, we envision using ciCas9 and DSB-ddPCR to dissect the influence of chromatin structure, target sequence, and cellular state on the kinetics of Cas9-mediated DSB generation and subsequent repair.

**Acknowledgements**

This work was supported by the NIH (R01GM086858 (D.J.M), R01GM109110 (D.M.F), F30CA189793 (J.C.R)), NSF (0954242 (D.J.M.)), and the Alzheimer's Association (NIRG304150) (D.M.F).

## Chapter 4: Conclusion

In this dissertation, I have described the development of precision tools for the study of dynamic processes. These tools confer researchers exquisite control over processes which are complex in time, space, and in their architecture. We focused on using such tools to studying intracellular signaling and genome engineering, but the advances in this work have far-reaching implications for the study of complex biological processes.

### Computational framework expedites protein rheostat development

Several groups have demonstrated the utility of protein rheostats and related tools for studying dynamic cellular processes[13,14]. The majority of these tools have relied on synthetic intermolecular regulatory systems[9,14,17-19]. Synthetic intermolecular regulatory systems are more straight-forward to engineer, but come with several caveats: (1) they cannot be applied to proteins which cannot be effectively sequestered from locations of activity, (2) the inability to have the protein localized to the site of function in the basal state may impair the speed of response and temporal resolution, (3) function is complicated by issues of stoichiometry of multiple protein components, (4) and the necessity for multiple protein components can complicate transporting the system to new cell lines or model systems. Synthetic intermolecular regulatory systems possess distinct advantages in all of these areas, yet are more difficult to engineer due to the difficulty in engineering allostery. Furthermore, synthetic intramolecular regulatory systems could in principled be combined with synethetic intermolecular regulatory

systems, as occurs in nature—many endogenous proteins are regulated by both intra- and intermolecular mechanisms, e.g. SOS[22,88-100]. Integrating intra- and intermolecular synthetic regulatory systems may permit researchers to encode complex behaviors and functions into engineered cell systems, which could be of great utility in endeavours such as engineered T-cell immunotherapy[101-103].

Our groups and others[11,12,15,16] have had some success engineering protein tools utilizing synthetic intramolecular regulation. However, these efforts have relied on empirical screening of constructs, which makes the process inefficient, costly, and often yields constructs with suboptimal behavior, for instance they may be leaky in the absence of the activator. To improve this process—both in terms of efficiency and the quality of the end-products—we developed a computational framework for the design of synthetic intramolecular regulation. This framework utilizes the protein design and modeling tools of *RosettaRemodel*[24], and efficiently screens candidate designs. We used this computational framework to engineer Chemically Inducible Activator of RAS (CIAR), a precise single-component RAS rheostat. CIAR has many favorable properties, including no detectable activity in the absence of the small molecule activator ("A3"), rapid response, and rapid transportability into other cell systems. The framework has broad utility, as it was able to correctly identify the best inducible construct among those tested in a previous study[12], as well as guide the design of an inducible VAV. VAV is a RHO family guanine exchange factor but is structurally unrelated to SOScat.

Due to its efficiency and general applicability in the design of protein rheostats, we envision this computational framework will expedite the development of many varied rheostats for studying and controlling diverse processes. The spread of such tools may greatly expand our knowledge of the complex cellular networks governing cell behavior. They may also find use in synthetic biology applications, such as engineering therapeutic T-cells or microbes for the production of useful compounds.

**CIAR enables a dissection of RAS signaling networks**

We utilized our computational framework to design a RAS rheostat for two primary reasons: (1) RAS biology is central to numerous malignancies, and (2) there is increasing appreciation that we need to better understand the role the magnitude, location, and timing of RAS activation play in cellular physiology and pathology. After validating that CIAR functions as a RAS rheostat in mammalian cells, we used it to interrogate several features of RAS signaling.

First, we examined RAS/ERK signaling kinetics. We found that while EGF stimulation elicited a rapid yet transient increase in ERK phosphorylation in CIAR-293 cells—a HEK-293 derivative line stably expressing CIAR—activation of RAS via CIAR lead to a more gradual and sustained response. In fact, ERK phosphorylation remained elevated at 48 hours after addition of A3. Intriguingly, the RAS/ERK signaling kinetics differed in

a non-transformed mammary epithelial line, where ERK phosphorylation kinetics after CIAR activation more closely mirrored EGF stimulation than they did in CIAR-293s.

We next examined the role of RAF regulation in controlling flux through the RAS/ERK signaling module. RAF1 (aka CRAF) and BRAF are known to be subject to multiple layers of regulation[35], both intrinsic and extrinsic to the RAS/ERK signaling module. Using genetic and pharmacological methods in concert with CIAR, we probed various aspects of RAF regulation and their impact on RAS/ERK signaling magnitude and kinetics. Perhaps most notably, we demonstrated for the first time that activation of endogenous wild-type RAS is sufficient to drive paradoxical activation by the RAF inhibitor vemurafenib. While not included in this dissertation, we have recently used CIAR to profile the effects of various RAF inhibitors on RAS/ERK signaling kinetics. Interestingly, RAF inhibitors which bind RAF in different modes appear to have differential effects on RAS/ERK signaling kinetics (data not shown).

Having used CIAR to look at signal propagation to a single node (ERK) downstream of RAS, we sought to profile signal propagation to thousands of nodes simultaneous. Using global serine/threonine phosphoproteomics, we profiled phosphorylation kinetics in response to both EGF stimulation and direct RAS activation via CIAR. We identified two classes of phosphopeptide response. One appeared to be dominated by the RAS/ERK signaling module and exhibited similar behavior to that observed for ERK phosphorylation. The second class exhibited a rapid decrease in phosphorylation

whether stimulated with EGF or A3. The significance of this response class, and the underlying mechanisms, remain unclear.

**A novel strategy for Cas9 engineering yields a rapidly inducible Cas9 construct, ciCas9**

In engineering a rapidly inducible Cas9, we replaced the non-essential REC2 domain with BCL-xL. This domain-replacement strategy is a novel approach to Cas9 engineering, and may have applications beyond what has been presented here. For instance, the REC2-replacement could serve as another site for fusion to transcriptional modulation domains. Perhaps more interestingly, one could envision replacing the REC2 domain with a DNA modifying domain, such as a nuclease or deaminase. Introduction of such enzymatic activity in proximity to the native catalytic sites of Cas9 may alter editing outcomes. For our purposes, REC2-replacement allowed us to generate a rapidly inducible Cas9 variant, ciCas9, which yields indels in targeted DNA more rapidly than has been reported for any other inducible Cas9 system. Importantly, ciCas9 activation is tunable and it exhibits low editing activity in the absence of A3, although uninduced editing does vary across sgRNAs. While we were primarily interested in using ciCas9 to study editing kinetics, we also demonstrate that its architecture permits tuning of its performance parameters to suit various applications. Introductions of previously describe mutations can increase its specificity, increasing the affinity of the BH3 peptide for BCL-xL decreases leakiness, and numerous drugs with varying potencies and specificities are available to activate ciCas9. The fact that ciCas9

is customizable, single-component inducible systems with demonstrated advantages over existing systems suggest it may gain wide adoption.

**DSB-ddPCR is the first assay for highly-quantitative, temporally resolved, and targeted double strand breaks**

In attempting to investigate editing kinetics, it quickly became apparent that there was no existing assay capable of monitoring DSBs in a targeted, highly-quantitative, and temporally resolved manner. Thus, we developed a droplet digital PCR-based assay for this purpose. This assay, DSB-ddPCR, accurately quantitated DNA digested with restriction enzymes *ex vivo* with a high degree of reproducibility. While we used DSB-ddPCR to examined Cas9 editing kinetics, in principle it could be used to characterize DSBs generated in a site-specific manner.

**ciCas9 and DSB-ddPCR enable examination of Cas9 cleavage and repair kinetics**

Despite the intense interest in CRISPR/Cas9, little is known regarding the dynamics of Cas9-mediated DNA cleavage and subsequent DNA repair. Together, ciCas9 and DSB-ddPCR enable the first-ever examination of the kinetics of these processes. We found that Cas9 cleavage is rapid, while indels accrue more slowly. Both cleavage and repair kinetics differ between loci, even between sgRNAs separated by only 10bp. We envision these technologies will enable in-depth examinations of this heretofore unexplored region of CRISPR/Cas9 biology. In particular, we are interested in elucidating the impact of sequence, chromatin state, and various DNA repair pathways on cleavage and repair kinetics.

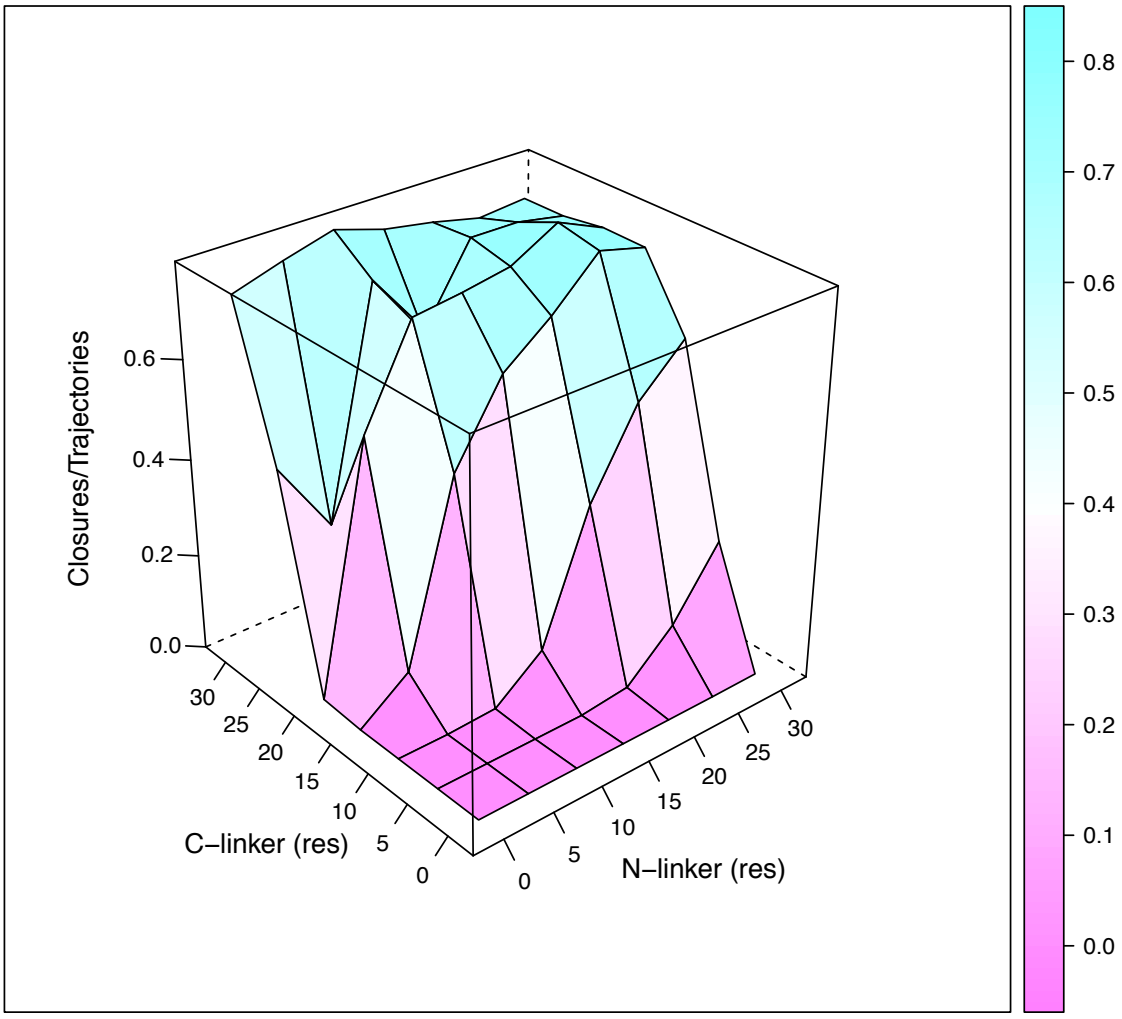
**Summation**

While the work described here has produced several insights into the biological systems studied, they serve largely as proof-of-principle. The tools developed in this work are capable of precisely perturbing and controlling complex and dynamic biological systems, and this precise control can produce previously unobtainable insights.

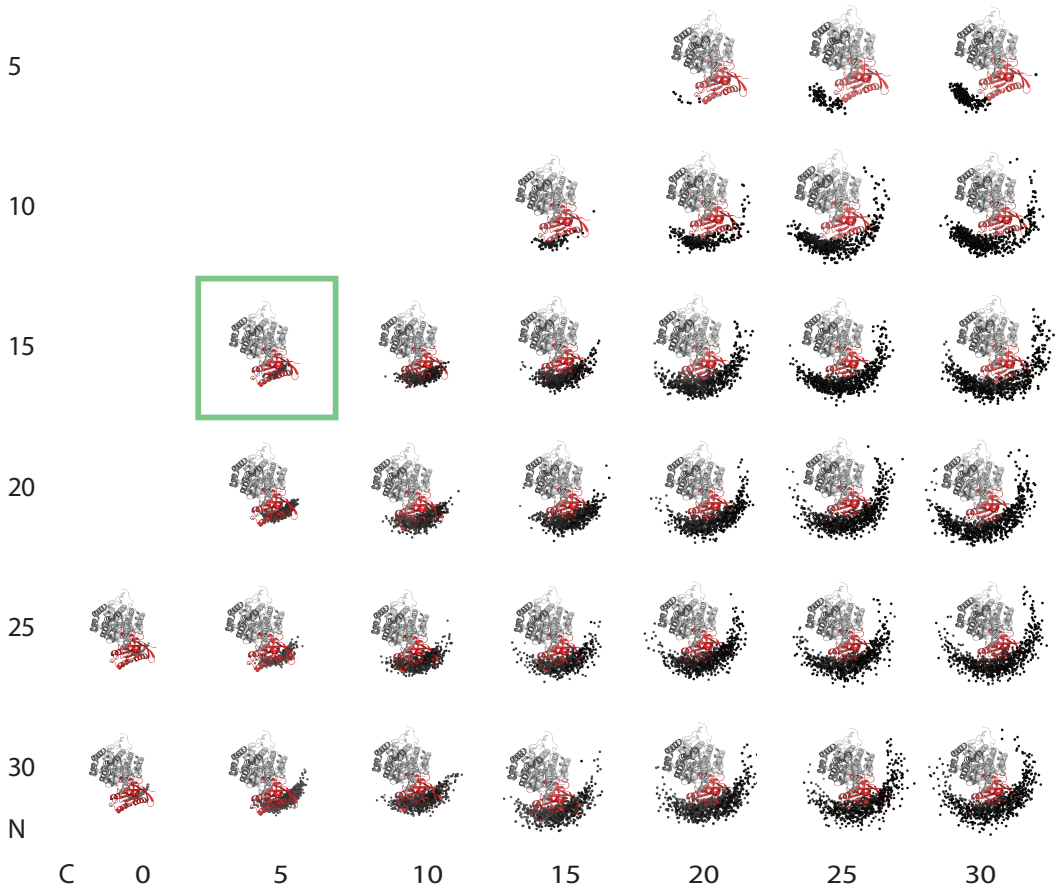
## Appendix A: Supplementary Material for Chapter 2

**Supplementary Figure A.1 |** Closure frequency as a function of N- and C-linker lengths
**Supplementary Figure A.2 |** BH3/BCL-xL center-of-mass (CoM) distributions for CIAR linker permutations
**Supplementary Figure A.3** | Fold activation of CIAR candidate constructs in an *in vitro* guanine nucleotide exchange assay
**Supplementary Figure A.4 |** S.17.3 exhibits minimal activation of the RAS/ERK pathway in the absence of A3 as measured by SRE assay
**Supplementary Figure A.5 |** An extended timecourse of direct RAS activation
**Supplementary Figure A.6 |** EGF timecourse in CIAR-MCF10A cells
**Supplementary Figure A.7 |** Unstimulated AKT phosphorylation in Control-293 and CIAR-293 cells
**Supplementary Figure A.8 |** Direct RAS activation in CIAR-MCF10A cells does not elicit AKT or EGFR phosphorylation
**Supplementary Figure A.9 |** Effect of wild-type and membrane targeted BRAF and RAF1 in CIAR-293 cells
**Supplementary Figure A.10 |** Overexpression of wild-type RAF1 has minimal effect on the response to direct RAS activation in CIAR-293 cells
**Supplementary Figure A.11 |** Overview of quantitative phosphoproteomics workflow
**Supplementary Figure A.12 |** Parallel Reaction Monitoring Mass Spectrometry (PRM-MS) validation of representative phosphosites
**Supplementary Figure A.13 |** Class II phosphosites derive from a highly interconnected protein interaction network
**Supplementary Figure A.14 |** Spliceosome components are enriched in proteins represented by class II phosphosites
**Supplementary Figure A.15 |** Tighter BH3/BCL-xL trajectories distributions are associated with greater fold activation for inducible ITSN constructs
**Supplementary Figure A.16 |** Comparison of the minimal catalytic domains of VAV and SOScat
**Supplementary Figure A.17 |** Termini identification for modeling BH3/BCL-XL, ITSN, and SOScat as rigid bodies
**Supplementary Figures 18-32 |** Uncropped immunoblots
**Supplementary Data Set 1 |** 30 min stimulation phosphoproteomic data set
**Supplementary Data Set 2 |** Timecourse phosphoproteomic data set
**Supplementary Data Set 3 |** Class I and Class II phosphopeptides

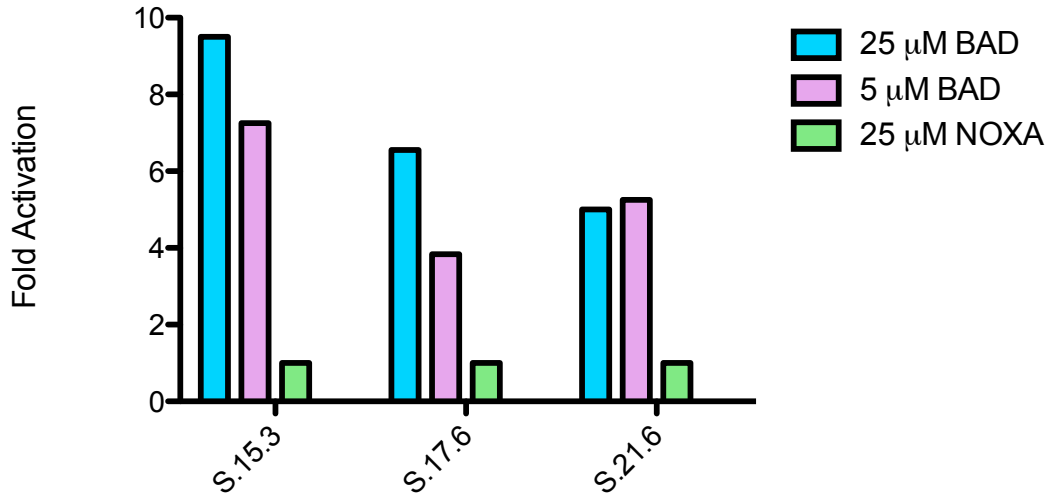
## Supplementary Figures



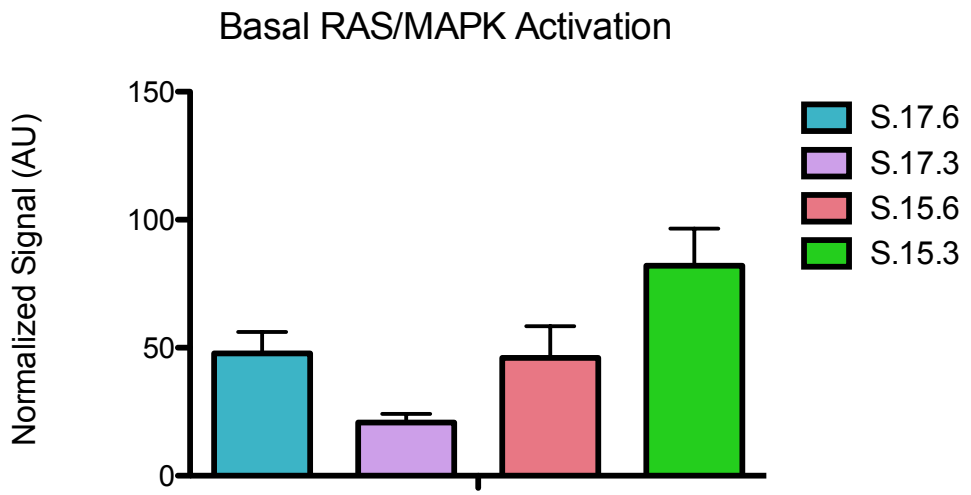

**Supplementary Figure A.1 | Closure frequency as a function of N- and C-linker lengths.** The number of successfully closed models was divided by the total number of trajectories (1000) at each linker permutation .



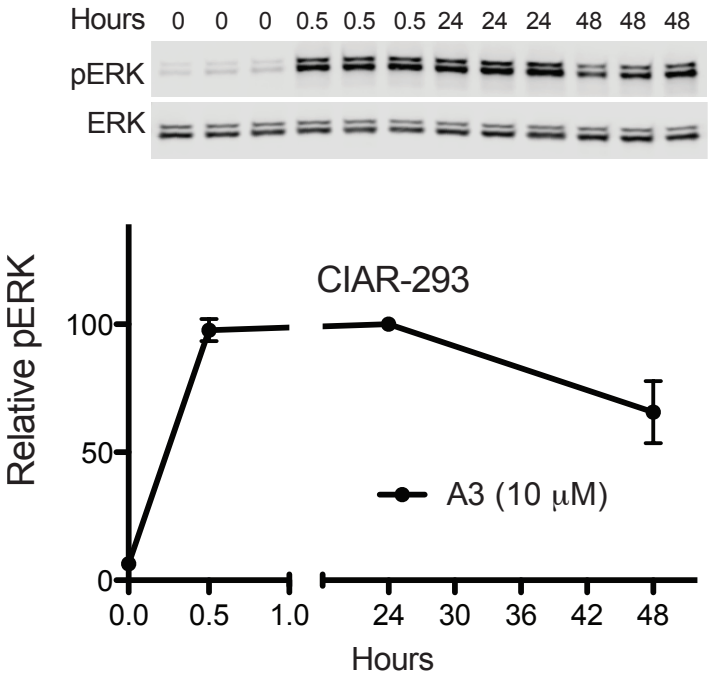

**Supplementary Figure A.2 | BH3/BCL-xL center-of-mass (CoM) distributions for CIAR linker permutations.** The CoM for each trajectory is shown (black spheres) overlaid on the structure of SOScat (grey) bound to RAS (red), with C-terminal linker length on the horizontal axis, and N-terminal linker length on the vertical axis. S.15.5 is indicated by a green box.



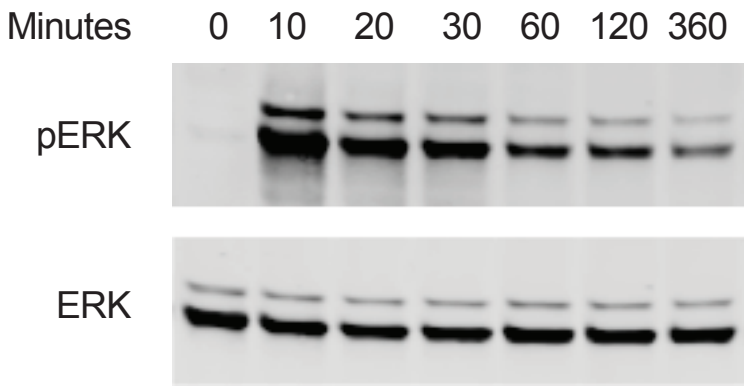

**Supplementary Figure A.3 | Fold activation of CIAR candidate constructs in an *in vitro* guanine nucleotide exchange assay.** HRAS-GDP (10 ¦ M) was incubated with 250 nM of indicated constructs, and 200 nM MANT-GTP. The rate of nucleotide exchange was determined in the presence of BH3 peptide competitor (5 or 25 ¦ M BAD) relative to control peptide (25 ¦ M NOXA). Exchange rates were determined by linear regression of mean values for triplicate experiments (single triplicate experiment per condition). The rate of exchange in the presence of HRAS-GDP alone was used for background subtraction. Fold activation was calculated as the rate for the given condition divided by the rate in the presence of control peptide (25 μM NOXA).



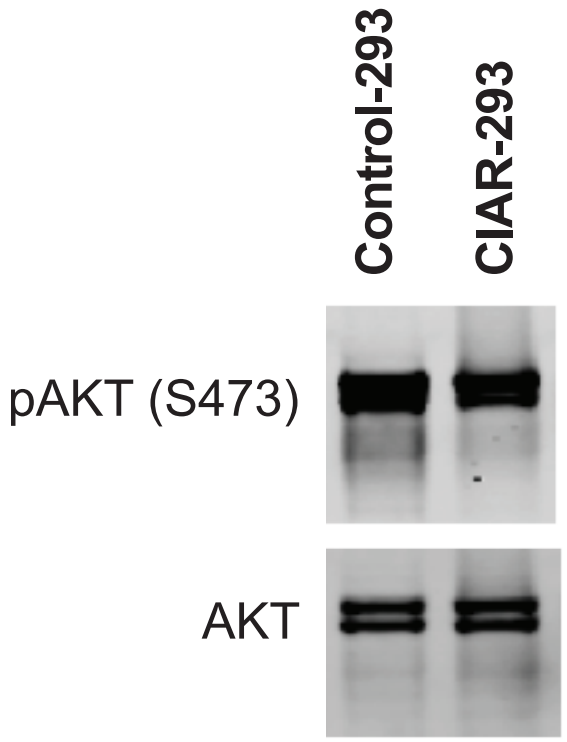

**Supplementary Figure A.4 | S.17.3 exhibits minimal activation of the RAS/ERK pathway in the absence of A3 as measured by SRE assay**. TREx-293 cells were transfected with the indicated constructs along with a SRE luciferase reporter and normalization control luciferase plasmids. For each construct, cells were treated with DMSO or the MEK inhibitor AZD6244 (40 μM). Mean reporter signal in AZD6244 treated wells served as a background signal estimate and was subtracted from wells treated with DMSO. Error bars = s.e.m, n = 3.



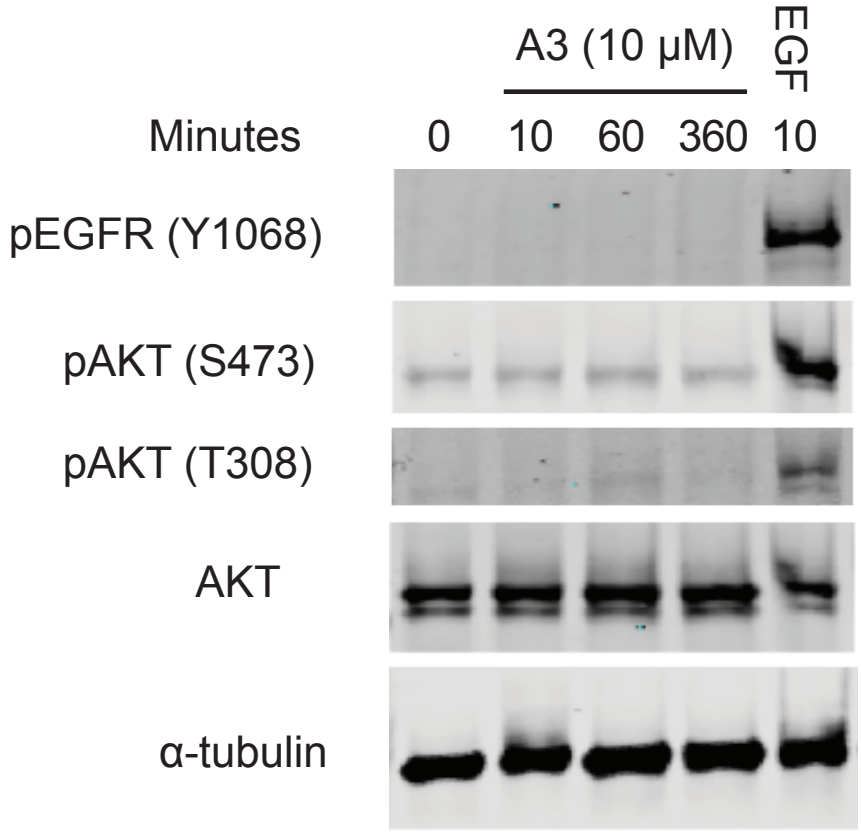

**Supplementary Figure A.5 | An extended timecourse of direct RAS activation.** Timecourse (0-48h) of direct RAS activation with A3 (10 μM) in CIAR-293 cells. Error bars = s.e.m, n = 3. Uncropped immunoblot available in Supplementary Fig. A.27.



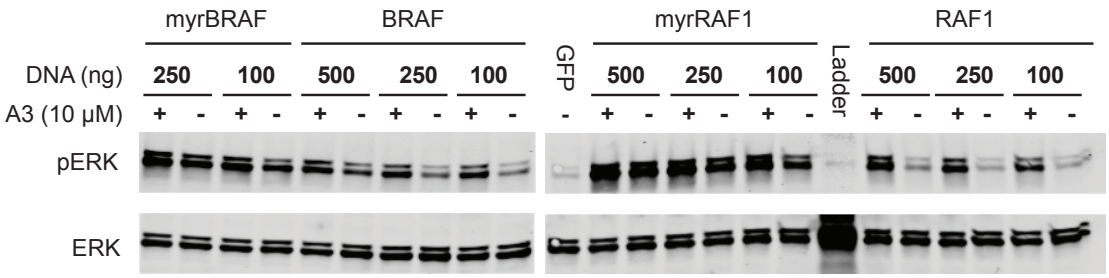

**Supplementary Figure A.6 | EGF timecourse in CIAR-MCF10A cells.** Cells were serum starved overnight, prior to stimulation with EGF (20 ng/mL) for the indicated durations. Cells were then lysed and evaluated by immunoblot. Representative immunoblot, n = 3. Uncropped immunoblot available in Supplementary Fig. A.28.



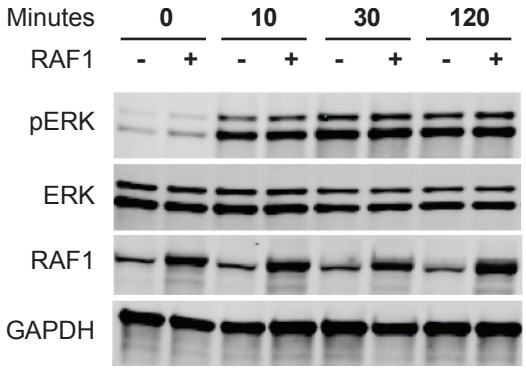

**Supplementary Figure A.7 | Unstimulated AKT phosphorylation in Control-293 and CIAR-293 cells**. Immunoblot analysis of CIAR-293 and Control-293 cells. Cells were serum starved overnight before lysis. Transgene expression was induced via inclusion of doxycycline (1 µg/mL). Control-293 cells express mCherry and were generated via the same selection process as CIAR-293 cells. Uncropped immunoblot available in Supplementary Fig. A.29.




**Supplementary Figure A.8 | Direct RAS activation in CIAR-MCF10A cells does not elicit AKT or EGFR phosphorylation.** Samples from Fig. 3f and Supplementary Fig. A.4 were probed for phospho-EGFR, phospho-AKT, AKT, and α-tubulin. Uncropped immunoblots available in Supplementary Fig. A.30.

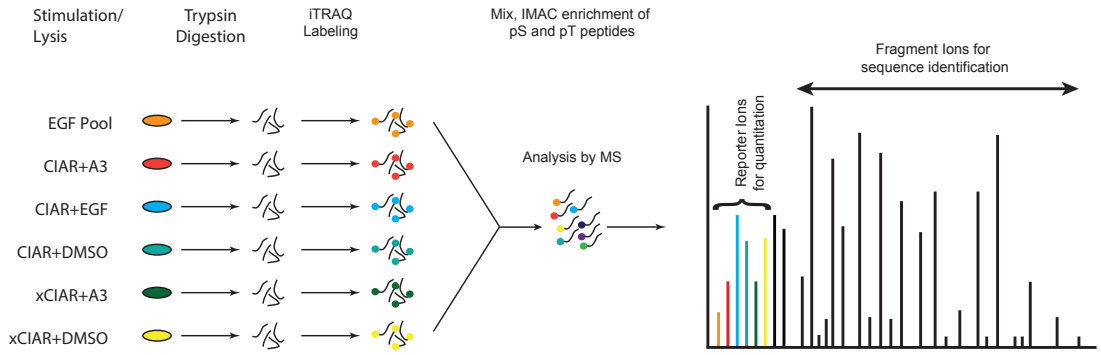



**Supplementary Figure A.9 | Effect of wild-type and membrane targeted BRAF and RAF1 in ClAR-293 cells.** Cells were transfected with wild-type BRAF, RAF1, or versions constitutively targeted to the plasma membrane via an N-terminal myristoylation sequence. 1 μg total DNA was transfected per well in 12-well plates The indicated amounts of RAF plasmid were co-transfected with GFP. Cells were serum starved overnight, incubated with A3 or DMSO for 10 minutes then lysed. Uncropped immunoblot available in Supplementary Fig. A.31.

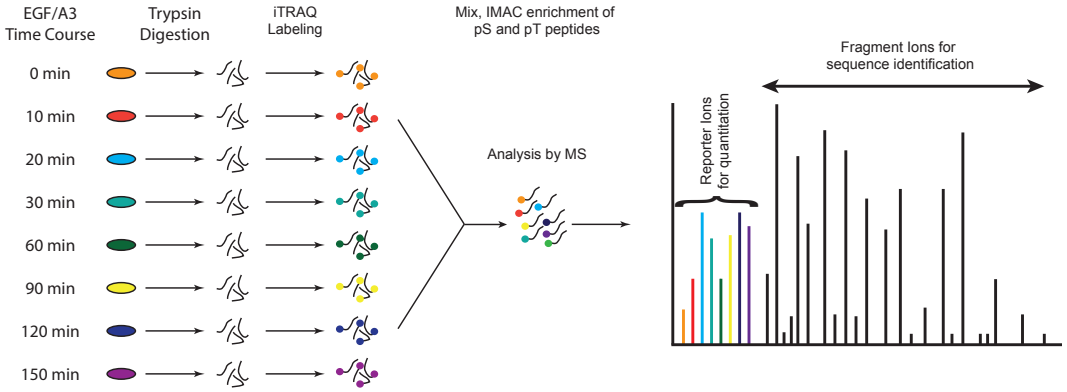



**Supplementary Figure A.10 | Overexpression of wild-type RAF1 has minimal effect on the response to direct RAS activation in CIAR-293 cells.** Cells were transfected with 1ug GFP plasmid alone or 250 ng wild-type RAF1 plasmid and 750ng µg GFP per well in 12-well plates. After overnight serum starvation, cells were incubated with A3 for the indicated durations before lysis. Uncropped immunoblots available in Supplementary Fig. A.32.

a

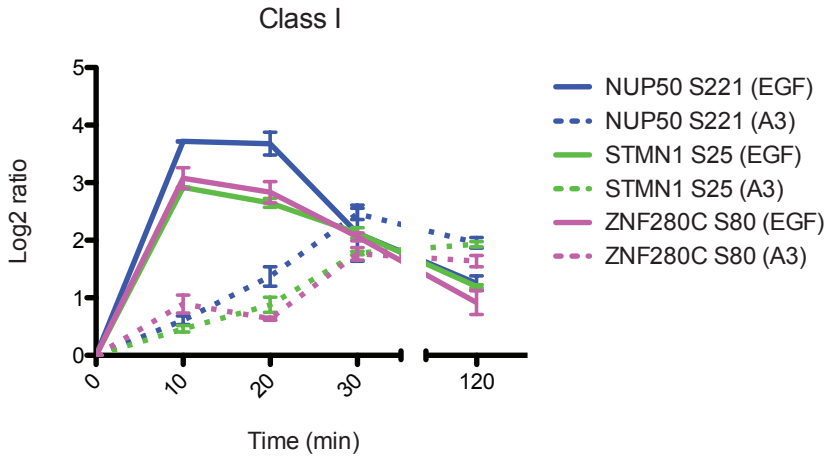



b

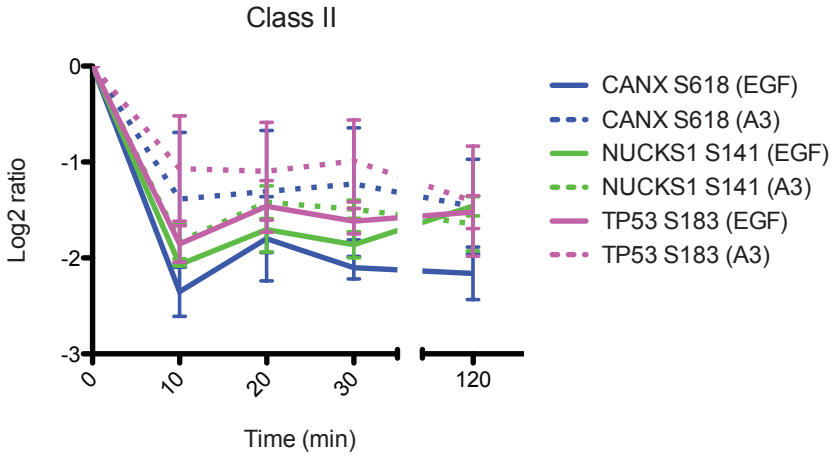



**Supplementary Figure A.11 | Overview of quantitative phosphoproteomics workflow.** Schematic representation of the workflow for **(a)** the CIAR- and xCIAR-293 comparison and **(b)** the timecourse experiment.

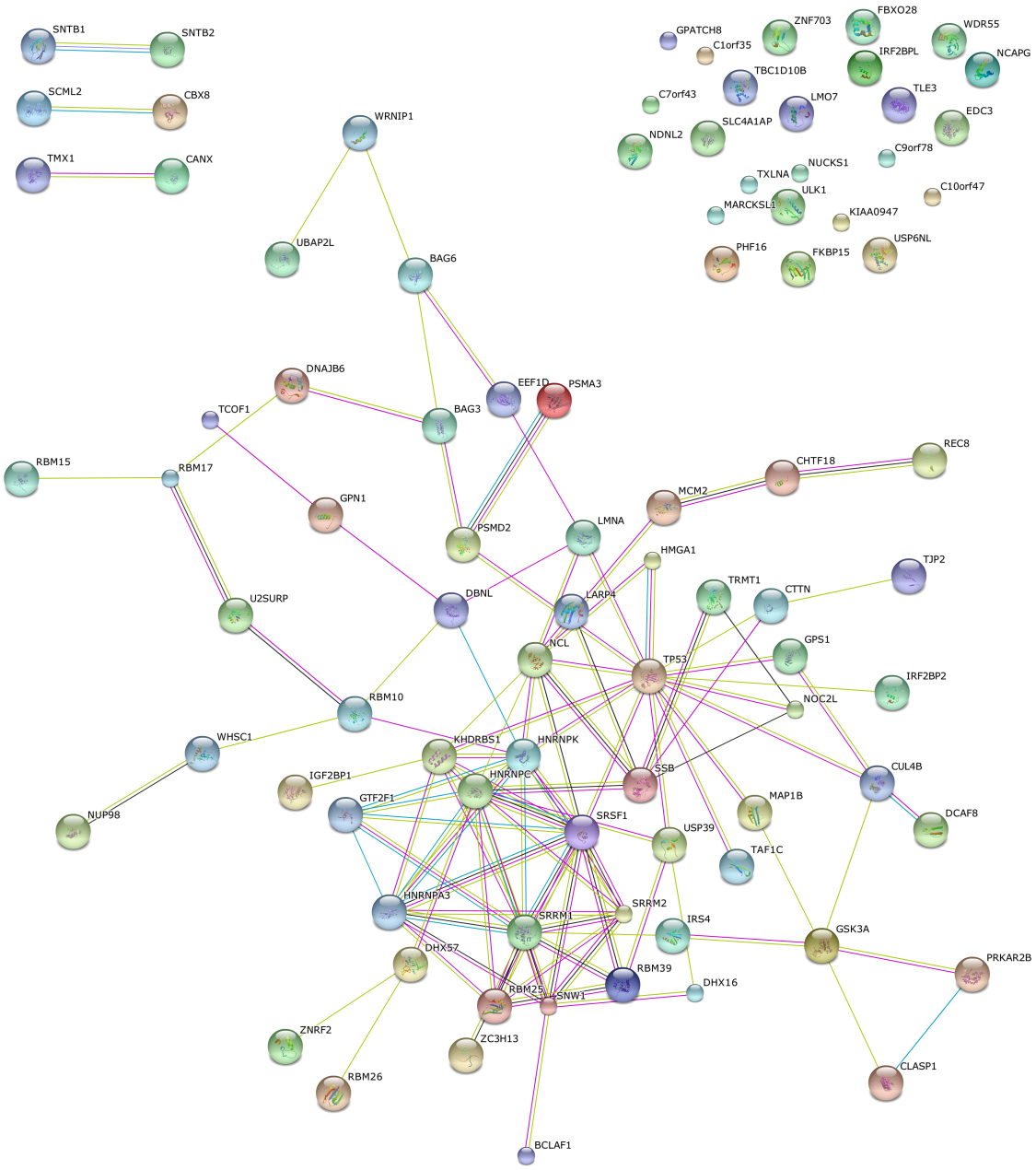
a



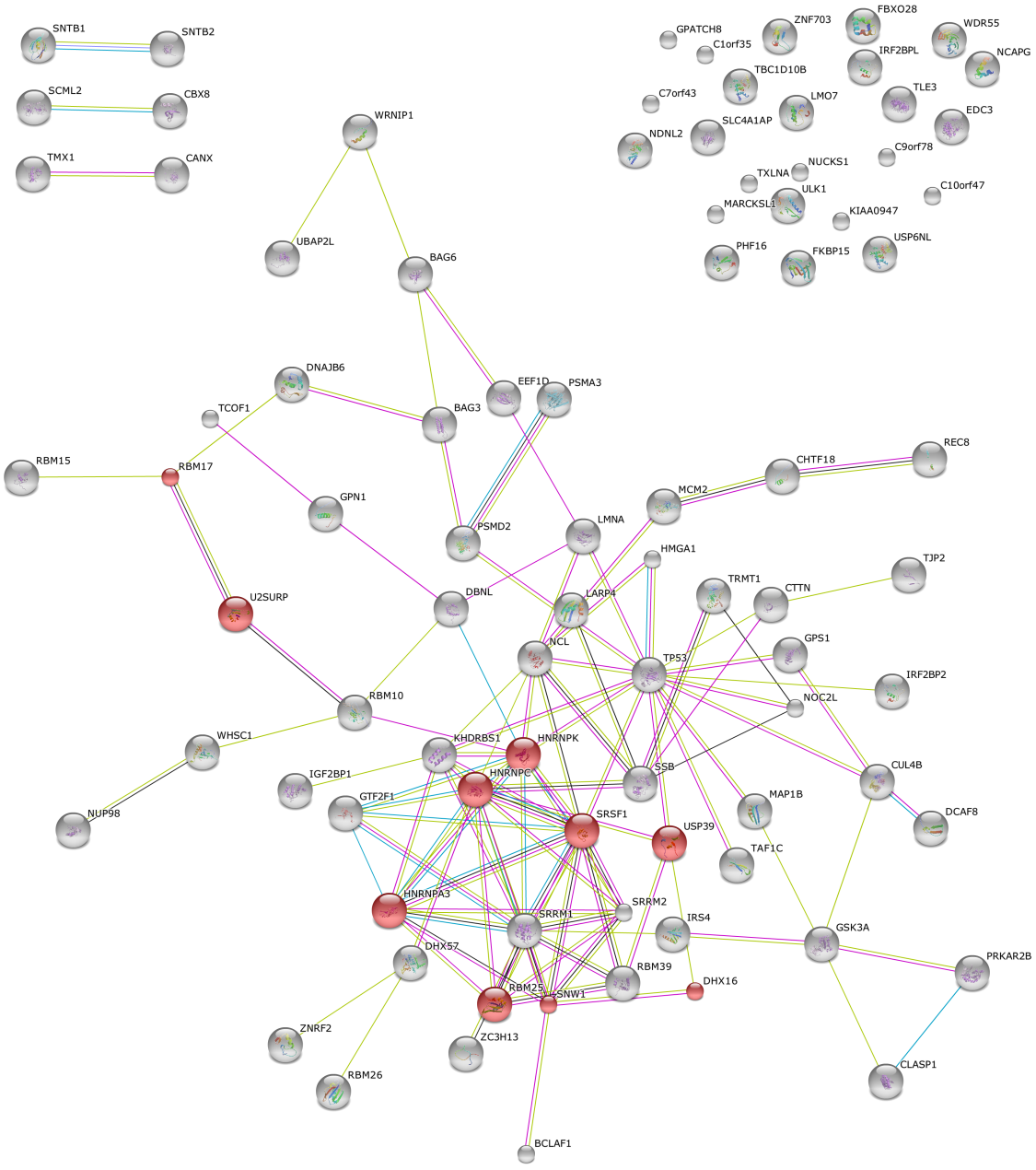

b



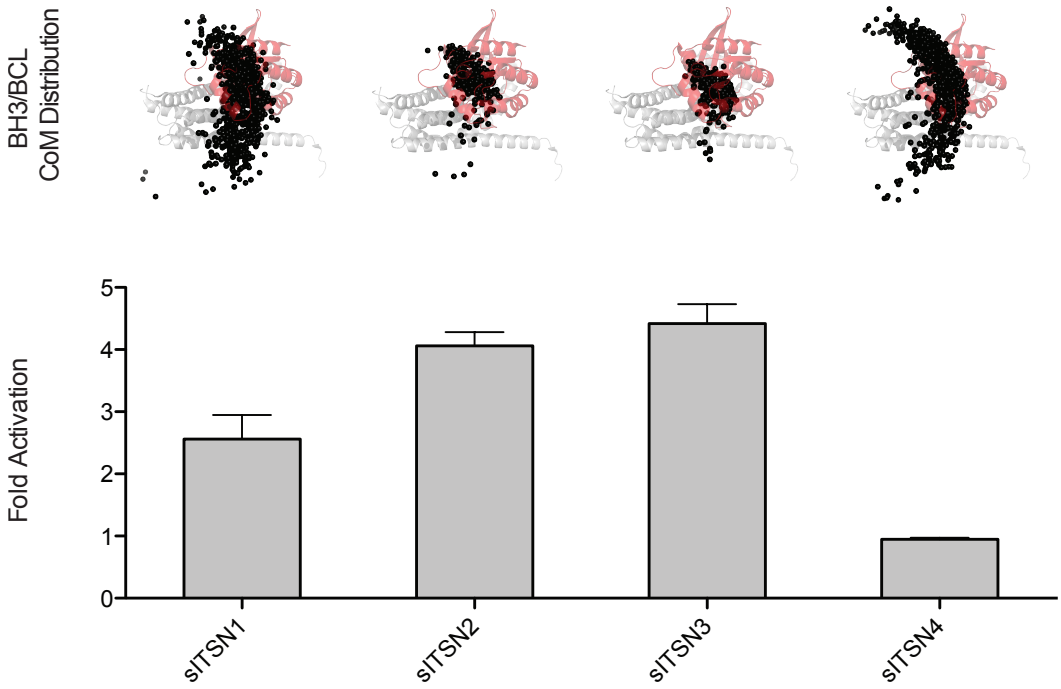

**Supplementary Figure A.12 | Parallel Reaction Monitoring Mass Spectrometry (PRM-MS) validation of representative phosphosites.** PRM-MS assessment of (a) Class I phosphosites (NUP50 S221, STMN1 S25, ZNF280C S80) and (b) Class II phosphosites (CANX S618, NUCKS1 S141, TP53 S183) at five time points. Error bars = s.e.m, n = 3.



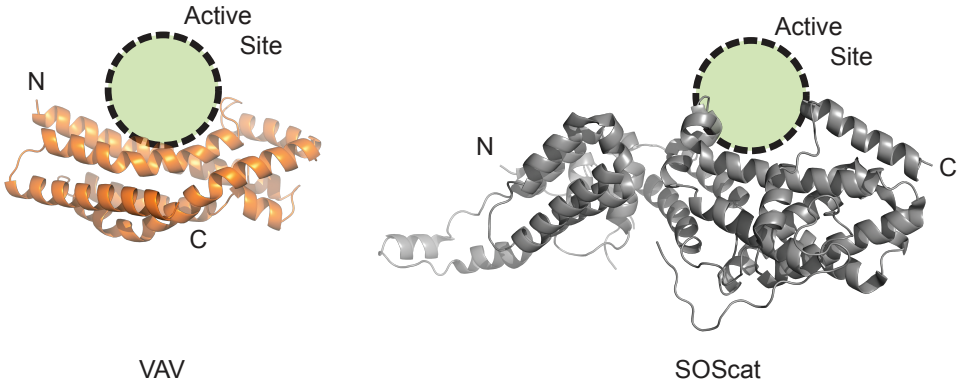

**Supplementary Figure A.13 | Class II phosphosites derive from a highly interconnected protein interaction network.** STRING interaction network of proteins represented by Class II phosphosites. Colored edges between protein nodes represent different types of interaction evidence: neighborhood (green), gene fusion (red), concurrence (blue), coexpression (black), experiments (magenta), databases (cyan), textmining (yellow), and homology (grey)[104].



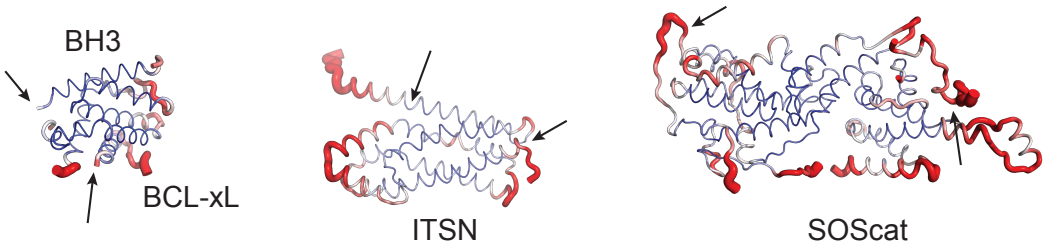

**Supplementary Figure A.14 | Spliceosome components are enriched in proteins represented by class II phosphosites.** Spliceosome components (red) were found to be significantly enriched (adjusted p = 0.0306) by evaluating KEGG pathway enrichment with the WEB-based GEne SeT AnaLysis Toolkit[105]. A hypergeometric test with BH multiple-test correction was used[106]. The 485 proteins quantified in all three biological replicates served as the reference set.



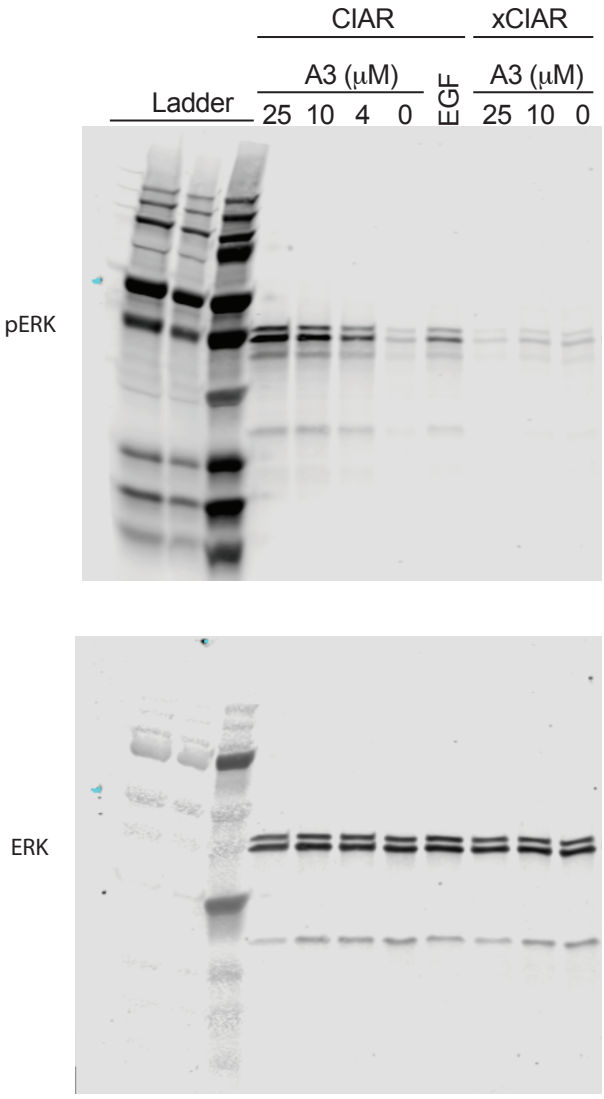

**Supplementary Figure A.15 | Tighter BH3/BCL-xL trajectories distributions are associated with greater fold activation for inducible ITSN constructs.** The center-of-masses (CoMs) for every trajectory for a given inducible synthetic ITSN (sITSN) design are show overlaid on the structure of ITSN (grey) bound to CDC42 (red). Fold-activation was determined using in an *in vitro* nucleotide assay (Online Methods). Error bars = s.e.m, n = 3.



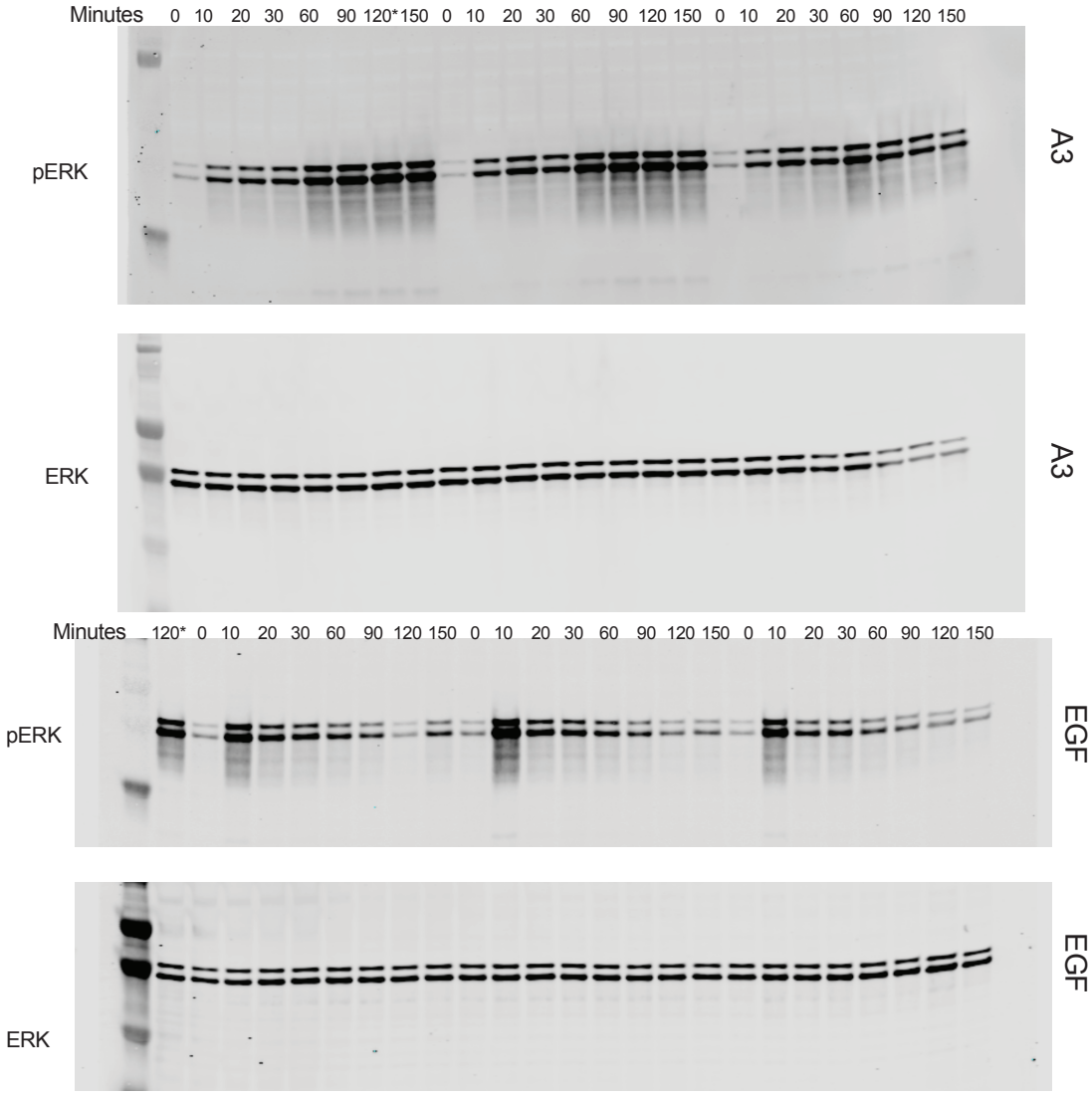

**Supplementary Figure A.16 | Comparison of the minimal catalytic domains of VAV and SOScat.**



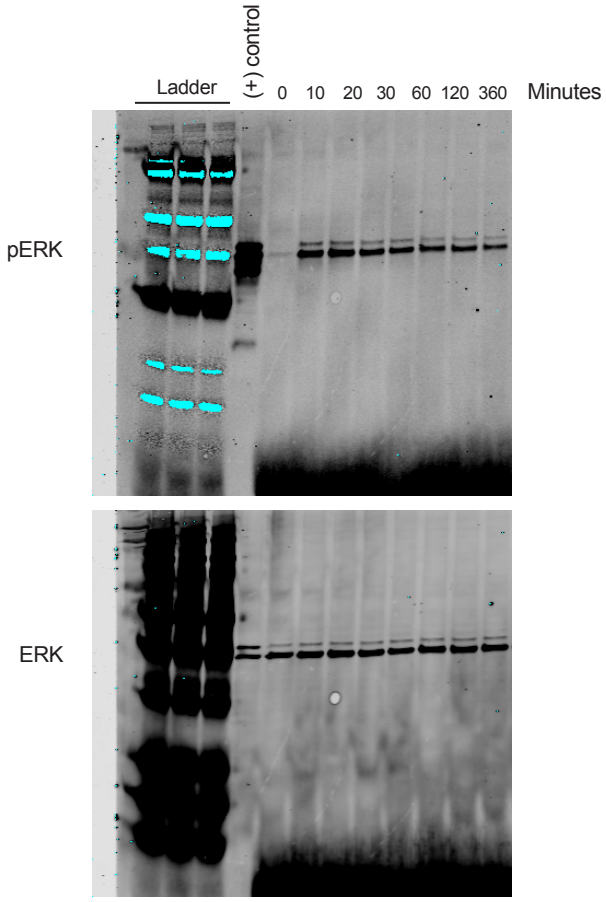

**Supplementary Figure A.17 | Termini identification for modeling BH3/BCL-XL, ITSN, and SOScat as rigid bodies.** Termini of rigid bodies used for computational modeling are indicated by arrows. B-factors are depicted using the B-factor option in pymol: low b-factors (blue, b-factor <30) are indicated by a narrow backbone and high b-factors (red, b-factor > 40) are indicated by a wide backbone. Similar termini were used for ITSN and VAV (not shown).




**Supplementary Figure A.18 | Uncropped immunoblot from Figure 3c.** Blot was visualized on the LI-COR Odyssey scanner using dual color detection (700 nm: pERK, 800 nm: ERK).

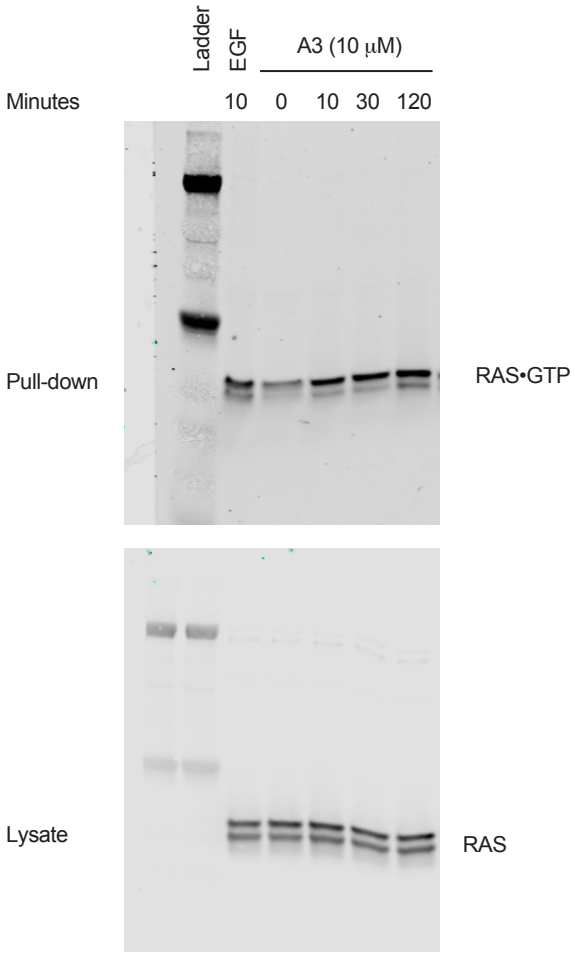



**Supplementary Figure A.19 | Uncropped immunoblots from Figure 3d.** Blots were visualized on the LI-COR Odyssey scanner using dual color detection (700 nm: ERK, 800 nm: pERK). A3 stimulated 120 minute sample denoted with "*" was rerun with EGF samples to allow for inter-blot normalization and comparison.

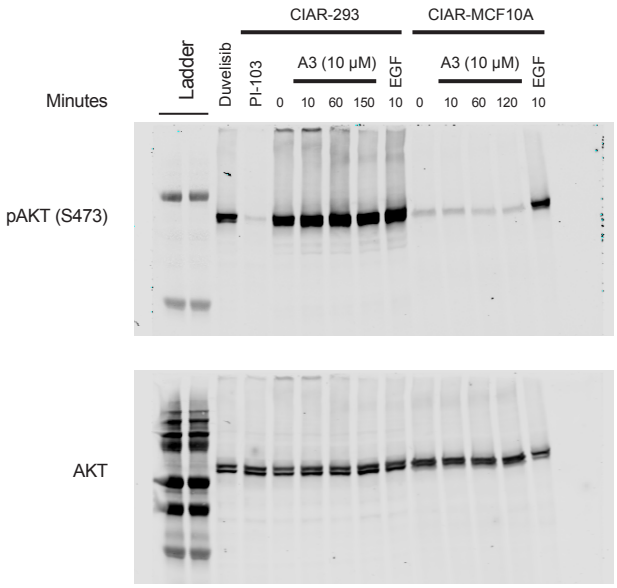



(+) control: EGF stimulated CIAR-293 lysate

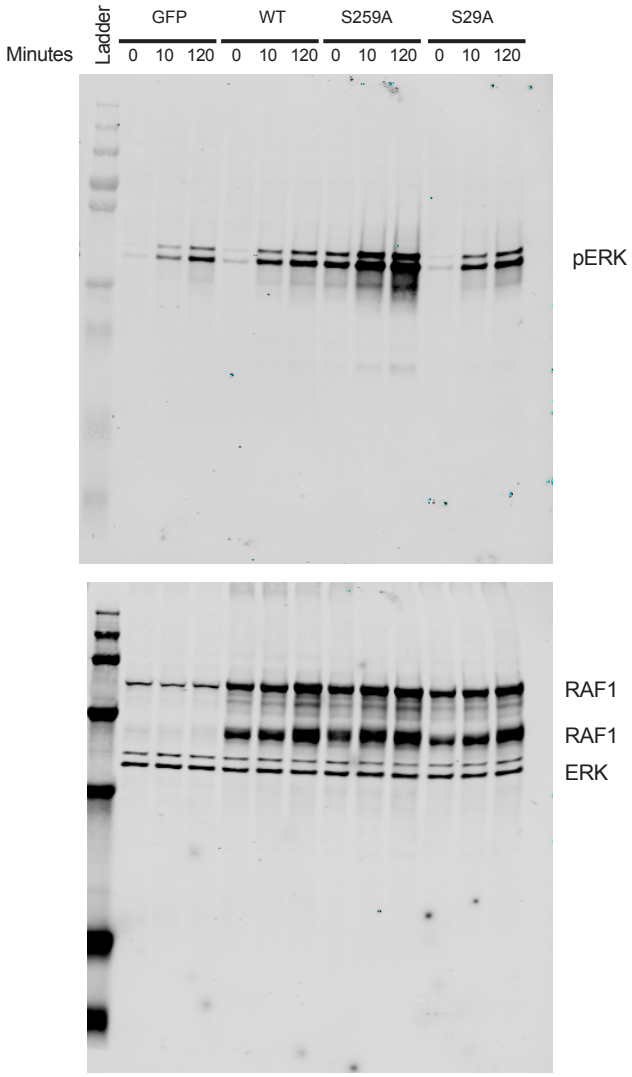
**Supplementary Figure A.20 | Uncropped immunoblot from Figure 3e.** Blot was visualized on the LI-COR Odyssey scanner using dual color detection (700 nm: ERK, 800 nm: pERK).



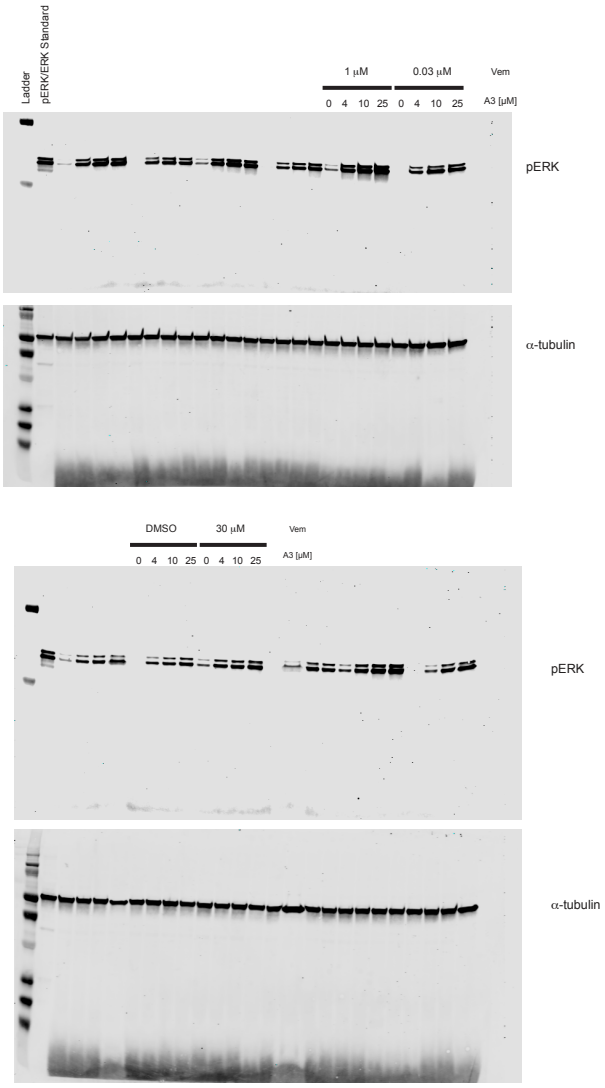

**Supplementary Figure A.21 | Uncropped immunoblots from Figure 4a.** Blots were visualized on the LI-COR Odyssey scanner (800 nm: RAS).



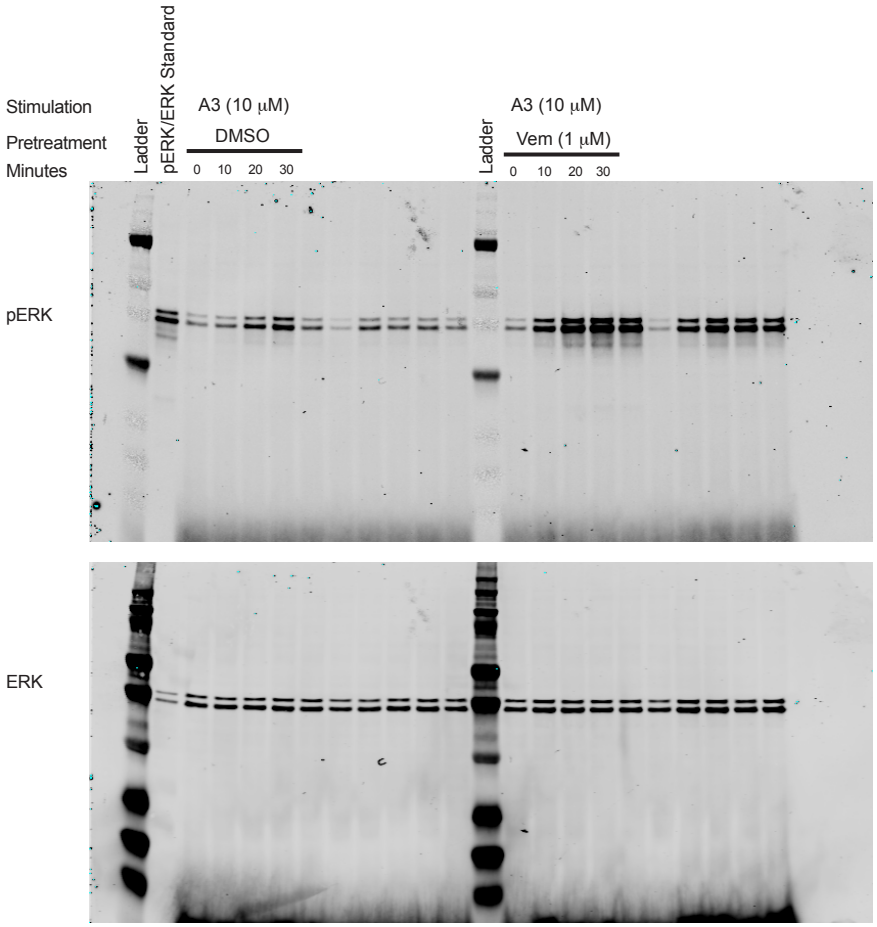

**Supplementary Figure A.22 | Uncropped immunoblot from Figure 4c.** Blot was visualized on the LI-COR Odyssey scanner using dual color detection (700 nm: pAKT (S473), 800 nm: AKT). In contrast to PI-103 (a pan-PI3K inhibitor), Duvelisib (a p110● and p110■inhibitor) resulted in incomplete inhibition of AKT phosphorylation.



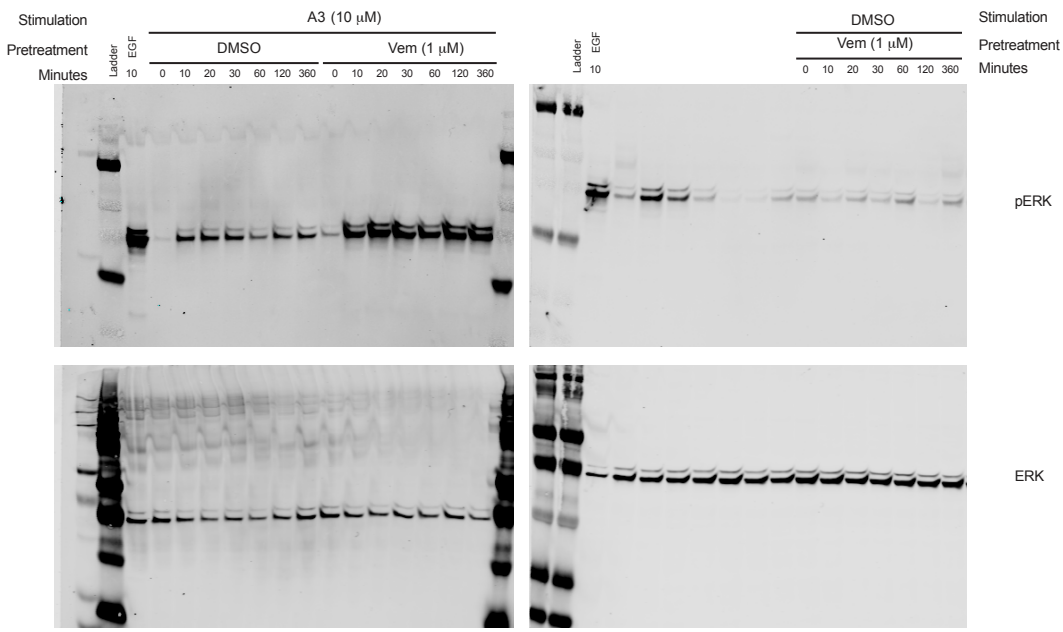

**Supplementary Figure A.23 | Uncropped immunoblots from Figure 4d.** Blot was visualized on the LI-COR Odyssey scanner using dual color detection (700 nm: ERK and RAF1, 800 nm: pERK).



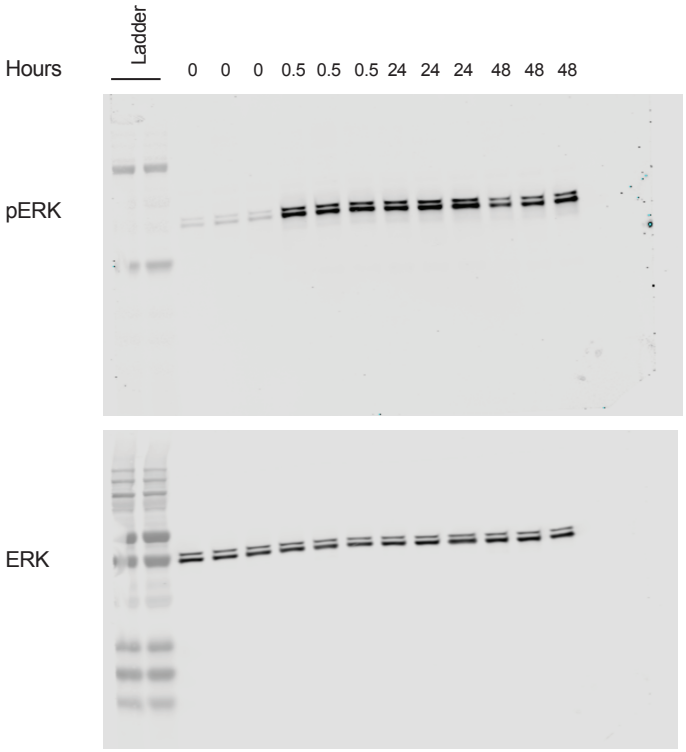

**Supplementary Figure A.24 | Uncropped immunoblots from Figure 4e.** Blots were visualized on the LI-COR Odyssey scanner using dual color detection (700 nm: ✔-tubulin, 800 nm: pERK). A pERK/ERK standard sample (EGF-stimulated CIAR-293 lysate) was run adjacent to the ladder in both gels to allow for inter-blot normalization and comparison.



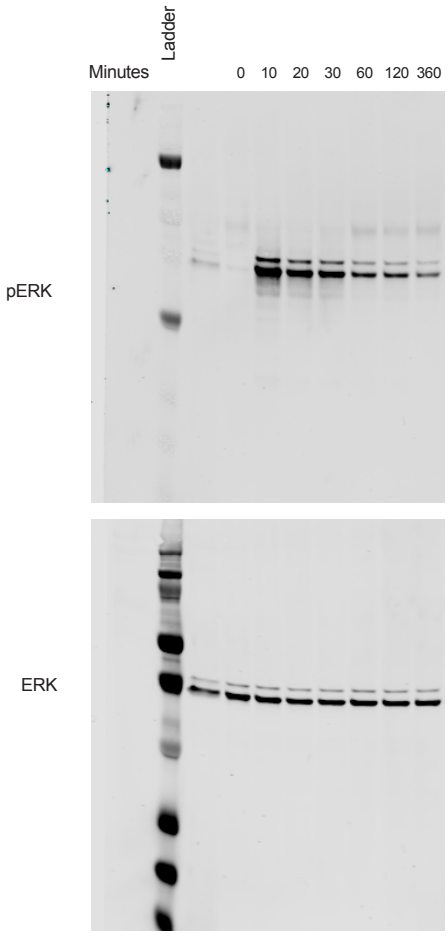

**Supplementary Figure A.25 | Uncropped immunoblot from Figure 4f.** Blot was visualized on the LI-COR Odyssey scanner using dual color detection (700 nm: ERK, 800 nm: pERK).



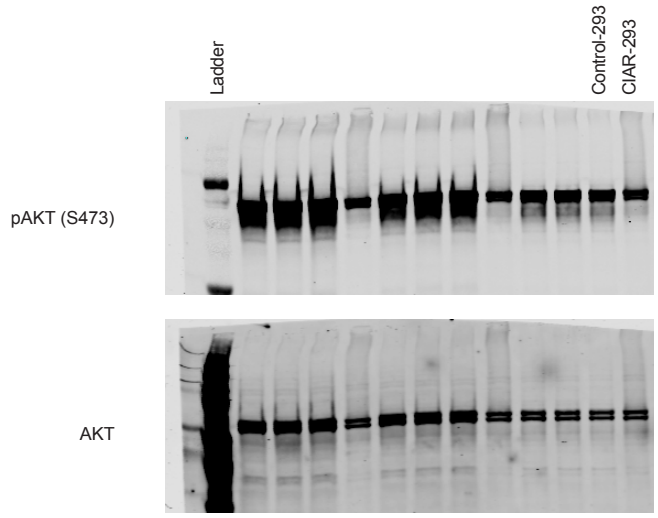

**Supplementary Figure A.26 | Uncropped immunoblot from Figure 4g.** Blots were visualized on the LI-COR Odyssey scanner using dual color detection (700 nm: ERK, 800 nm: pERK). Sample from CIAR-MCF10As stimulated with EGF for 10 min was used as a standard for inter-blot normalization and comparison.



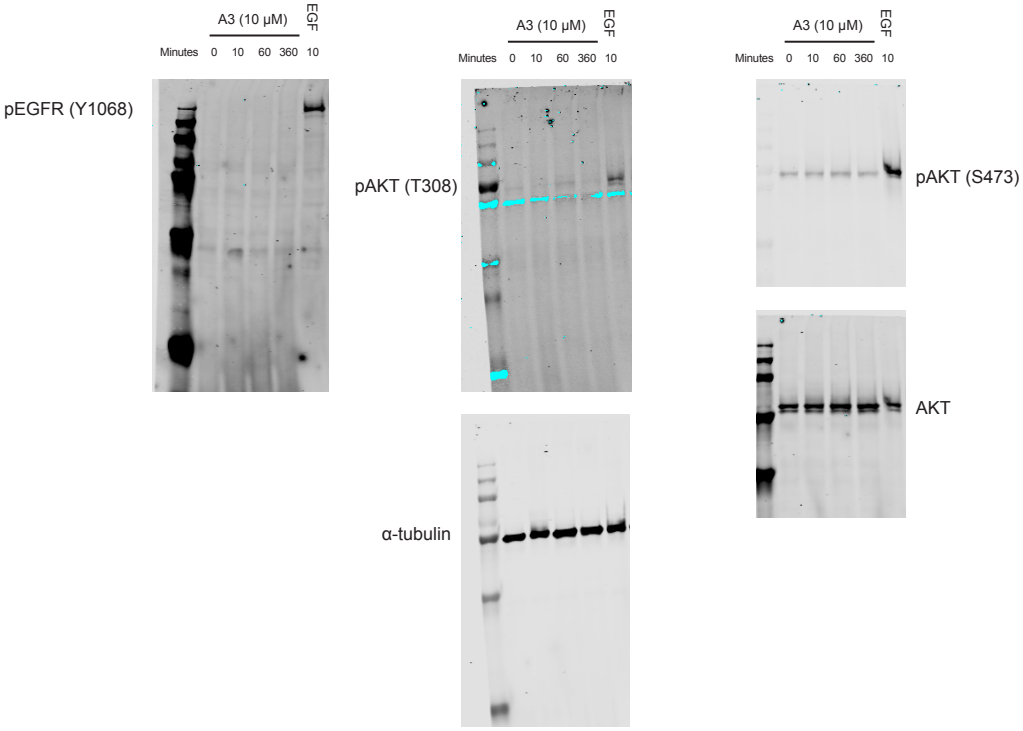

**Supplementary Figure A.27 | Uncropped immunoblot from Supplementary Figure A.5.** Blot was visualized on the LI-COR Odyssey scanner using dual color detection (700 nm: ERK, 800 nm: pERK).



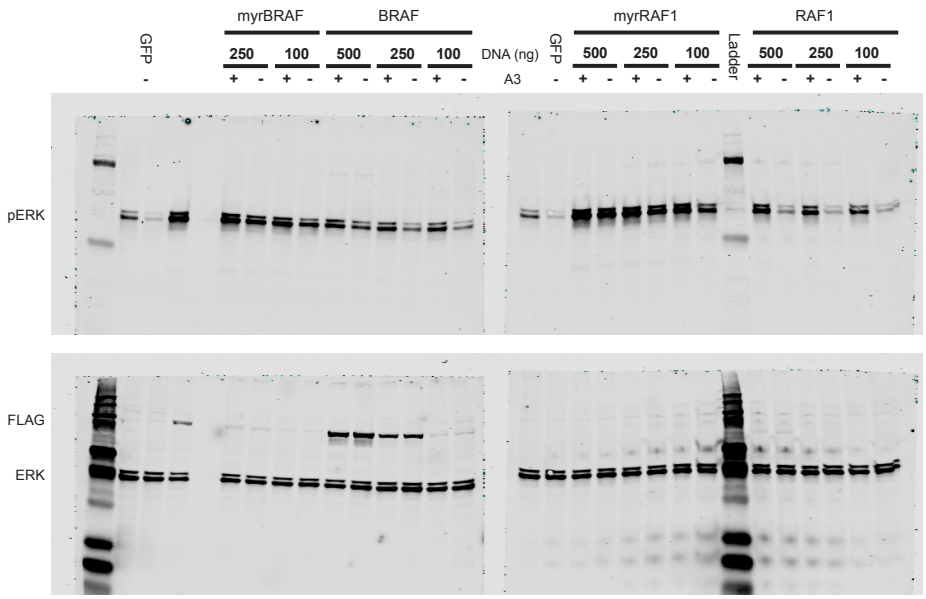

**Supplementary Figure A.28 | Uncropped immunoblot from Supplementary Figure A.6.** Blot was visualized on the LI-COR Odyssey scanner using dual color detection (700 nm: ERK, 800 nm: pERK).



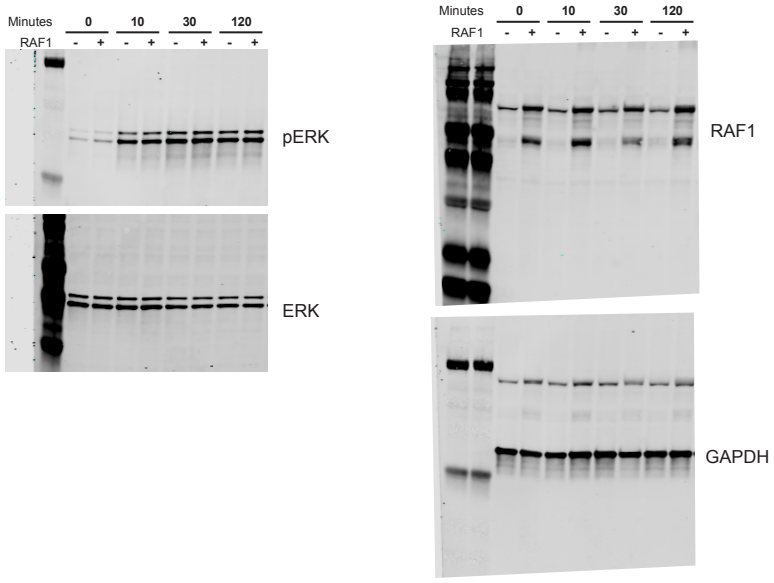

**Supplementary Figure A.29 | Uncropped immunoblots from Supplementary Figure A.7.** Blot was visualized on the LI-COR Odyssey scanner using dual color detection (700 nm: AKT, 800 nm: pAKT (S473).




**Supplementary Figure A.30 | Uncropped immunoblots from Supplementary Figure A.8.** Saturation in pAKT (T308) image is due to bleed through signal from ✔-tubulin channel on same blot. Blots were visualized on the LI-COR Odyssey scanner using dual color detection (left blot – 800 nm: pEGFR; center blot – 700 nm: ✔-tubulin, 800 nm: pAKT (T308); right blot – 700 nm: AKT, 800 nm: pAKT (S473)).




**Supplementary Figure A.31 | Uncropped immunoblots from Supplementary Figure A.9.** Blot was visualized on the LI-COR Odyssey scanner using dual color detection (700 nm: ERK and FLAG, 800 nm: pERK). Lysate from cells transfected with GFP alone (no A3 added) was run in both blots to allow inter-blot comparison.




**Supplementary Figure A.32 | Uncropped immunoblots from Supplementary Figure A.10.** Blots were visualized on the LI-COR Odyssey scanner using dual color detection (left blot – 700 nm: ERK, 800 nm: pERK; right blot – 700 nm: RAF1, 800 nm: GAPDH.

## Supplementary Data Sets

**Supplementary Data Set 1 | 30 min stimulation phosphoproteomic data set.** Phosphoproteomic data used in Figure 5a-b. Each peptide is listed with its fold-change comparing A3 or EGF to DMSO treatment along with the correspond p-value (calculated in excel using a one-sided two-sample Student's t-test). The workflow for generating these data is outlined in Supplementary Figure A.11a.

**Supplementary Data Set 2 | Timecourse phosphoproteomic data set.** Intensity values for 657 phosphopeptides quantified in all replicates for all conditions and time points. Intensity values are normalized to intensity in DMSO control sample. The order of phosphopeptides in the heatmap in Figure 5c is listed. The workflow for generating these data is outlined in Supplementary Figure A.11b.

**Supplementary Data Set 3 | Class I and Class II phosphopeptides.** List of members of Class I and Class II phosphopeptide response classes.

## Materials and Methods

**Conformational Sampling by RosettaRemodel.** The BH3/BCL-xL complex (PDB ID: 2BZW) was treated as a rigid-body connection between the N- and C-termini of the SOScat model (residues 574-1020, PDB ID: 1XD2). The setup to allow this linkage required that we create a chain break in the SOScat structure between residues *i* and *i*+1. (The position was chosen arbitrarily, away from the termini). This circular permutation scheme allowed treatment of the insertion of BH3/BCl-xL across the termini as a loop closure problem.

We used the circularly permuted residue order on SOScat to generate blueprint files for RosettaRemodel[24]. A blueprint file defines regions to rebuild in a modeling task. The functionality for fusing a structure into the other (which is turned on with the *-insert_segment_from_pdb* flag) uses "0 x l" entries, which denote extensions of the peptide chain from the insertion PDB, in the range that corresponds to the placement of the BH3/BCl-xL complex. We varied the lengths of the linkers on either side of the BH3/BCL-xL insertion with "0 x L" entries, which denote extensions for the peptide chain by randomly selecting fragments that are loops from the fragment database. We tested loop lengths between 0 and 30, at five residue increments for the loops leading into the BH3/BCl-xl complex and for the loop leading out of it; this gives 49 different length combinations. The definition for running lengths 5 and 5 on either side of the insertion is given below: (only the residues with assignments in the blueprint file are given; in the actual blueprint, all the entries listed in the brackets were present.)

[residue entries from 1 – 36]
37 E .
38 P L PIKAA P
0 x L PIKAA G

0 x L PIKAA S
0 x L PIKAA G
0 x L PIKAA T
0 x L PIKAA G
0 x I NATRO
[167 lines corresponding to BH3/BCL-xL complex]
0 x I NATRO
0 x L PIKAA G
0 x L PIKAA S
0 x L PIKAA G
0 x L PIKAA S
0 x L PIKAA G
49 D L PIKAA D
50 V .
[residue entries from 51 and on]

The flexible part of the loops were assigned the identity of repeating glycine-(serine/threonine) residues, while their backbone conformations were sampled by introducing random combinations of torsions from loop fragments. To facilitate the conformational search, a random break was made in the loop outside of the BH3/BCL-xL region to be reconnected via both random fragment moves and chain-closure algorithms guided by the Rosetta energy function; trajectories that properly reconnected the chain were considered successful. The lowest energy model from each successful trajectory was saved as a PDB file.

The flags to run the calculations are as follows:
-database [rosetta database location]
-s [circularly permuted template PDB file]
-remodel:blueprint [blueprint files]
-insert_segment_from_pdb [PDB formatted BH3/BCL-xL complex]
-num_trajectory 20
-save_top 20
-remodel:quick_and_dirty
-use_clusters false
-vall debug1000vall
-overwrite
-max_linear_chainbreak 0.2

1000 independent trajectories were sampled in 50 parallel runs that used the flags above.

Modeling of ITSN and VAV designs was performed as above using the ITSN (PDB: 1ki1) and VAV (PDB: 3BJI) structures. sITSN designs were described previously[6].

The orientation of the terminal residues in all components were considered, and residues that are not integral to domain structures were either truncated or grouped as

part of the linker to be sampled. B-factors were used to inform termini identification of rigid bodies used in modeling for SOScat, ITSN, VAV, and the BH3/BCL-xL complex (Supplementary Figure A.17).

**Closure frequency analysis.** We placed an arbitrary lower bound on the frequency of chain closure at 10%; linker length sets giving fewer closure events than the threshold would likely fail to allow BH3/BCL-xL complex formation, i.e these linkers would fail to meet Design Criterion 1. Therefore, linker length sets yielding a closure rate <10% were excluded from further computational analysis. Additionally, we speculated that those linkers exhibiting the highest closure rates would grant the complex too many degrees of freedom, resulting in poor autoinhibition (failing Design Criterion 3).

**BH3/BCL-xL complex localization analysis.** The metric $\mu$ was calculated for each geometrically allowed model. This was achieved via a custom pymol script (available upon request). Briefly, the SOScat, ITSN, or VAV PDB was loaded in pymol and the Center of Mass (CoM) of RAS, CDC42, or RAC1 was determined using the center_of_mass pymol script (http://www.pymolwiki.org/index.php/Center_of_mass). Each model PDB was iteratively loaded, aligned to the SOScat, ITSN, or VAV PDB, the BH3/BCL-xL CoM determined, and the distance to the RAS, CDC42, or RAC1 CoM calculated ($\mu$).

**Plasmids.** For bacterial expression, a previously described BH3 peptide (APPNLWAAQRYGRELRRMSDEGEGSFK), SOScat (res 574-1020, T968L), BCL-xL (res 2-215), and intervening variable length flexible linkers consisting of glycine-(serine/threonine) repeats were subcloned into pMCSG7 containing an N-terminal hexahistidine tag. The T968L mutation yields a constitutively active SOScat[89]. Wild-type HRAS (res 1-166) was subcloned into pMCSG7 containing an N-terminal hexahistidine tag. For expression in HEK-293T and TREx-293 cells, candidate CIAR constructs were subcloned into pcDNA5/FRT/TO, incorporating the 25 C-terminal residues of HRAS for membrane targeting. A lentiviral vector was generated by subcloning GFP, a P2A ribosomal skipping sequence, and CIAR (S.17.3) into pLenti_CMV_GFP_Puro[107]. pLenti_CMV_GFP_Puro (658-5) was a gift from Eric Campeau (Addgene plasmid # 17448). CIAR constructs used in cells utilized the following BH3 peptide sequence (APPNLWAAQRYGRELRRMADEGEGSFK). The catalytic domain of Vav2 (residues 190-374, synthesized by Genscript) was cloned with BCL-xL, BH3 peptide, and linker sequences into a pDest-527 bacterial protein expression vector using Gateway® recombination cloning technology by Life Technologies Human Rac1 (residues 1-177) was amplified from the original pcDNA3-EGFP-Rac1(wt) plasmid (Addgene plasmid #12980). The construct was cloned into a pDest-527 bacterial protein expression vector using Gateway® recombination cloning technology by Life Technologies. Wild-type, myristoylated, and BRAF and RAF1 constructs, and mutant RAF1 constructs were cloned into pcDNA3.1 via Gibson Assembly Cloning (NEB). Wild-type RAF1 sequence was PCR amplified out of pBABEpuro-CRAF, a gift from Matthew Meyerson (Addgene plasmid # 51124). Plasmids generated for this study will be made available via Addgene.

**Protein Expression and Purification**. CIAR constructs were expressed in *E. coli* Rosetta 2(DE3) cells by inoculating 1 L LB broth containing ampicillin (100 µg/mL). Cultures were grown at 37ºC to an $OD_{600}$ of 0.5-0.8, then induced at 15ºC with IPTG (250 µM) for 18 h. To purify protein, pellets were resuspended in CIAR Resuspension Buffer (100mM HEPES, 20 mM imidazole, 500 mM NaCl, 10% glycerol, 1 mM dithiothreitol (DTT), pH 7.5) containing PMSF (100 µg/mL) then lysed by sonication and cleared by centrifugation at 4ºC. Cleared lysate was purified with Promega HisLink™ resin by rotating for 30-60 min at 4ºC. Resin was washed three times with CIAR Wash Buffer (CIAR Resuspension Buffer + 50mM imidazole), then eluted with 1-3mL CIAR elution Buffer (CIAR Resuspension Buffer + 600 mM imidazole). Collected fractions were dialyzed into Assay Buffer (20mM Tris, 50mM NaCl, 10mM $MgCl_2$, 1% glycerol, 1mM DTT, pH 7.0), analyzed for purity by SDS page, aliquoted, snap frozen, and stored at -80ºC. HRAS was expressed in *E. coli* Rosetta 2(DE3) cells by inoculating 2 L Terrific Broth containing ampicillin (100 µg/mL). Cultures were grown to an $OD_{600}$ of 0.4-0.8 and induced at 18ºC with IPTG (250 µM) for 18 h. HRAS protein was purified as above, except wash and elutions buffers contained no glycerol and were supplemented with GDP (5 µM) and $MgCl_2$ (500 µM).

Inducible VAV constructs were expressed as *N*-terminal hexahistidine fusions in *E. coli* Rosetta 2 (DE3) cells by inoculating LB broth containing ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL). Cultures were grown at 37°C to an $OD_{600}$ of 0.6-0.8 and induced at 20°C for 3-4 h with 1 mM IPTG. To purify the protein, cell pellets were re-suspended in His6 wash buffer (50 mM HEPES, 10 mM imidazole, pH 7.5) containing PMSF (100 µg/mL), followed by sonication and centrifugation at 4°C to clear the lysate. The cleared lysate was purified using Promega HisLink™ Protein Purification resin by rotating at 4°C for 30 min. Subsequently, the resin was washed 3x and the protein eluted with His6 elution buffer (100 mM HEPES, 400 mM imidazole, pH 7.5). Collected fractions were analyzed for purity by SDS-PAGE and dialyzed into a buffer containing 20 mM Tris, 50 mM NaCl, 2 mM DTT, pH 7.5. Constructs that were >95% pure were aliquoted, snap-frozen and stored at -80°C.

RAC1 was expressed in *E. coli* Rosetta 2(DE3) cells by inoculating 1 L of Terrific Broth containing ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL). Cultures were grown at 37°C to an $OD_{600}$ of 0.6-0.8 and induced at 25°C for 4 h with 1 mM IPTG. Bacterial pellets were re-suspended in wash buffer (10 mM HEPES, 2 mM imidazole, 1.0 mM DTT, 1.0 mM $MgCl_2$, 10 µM GDP, pH 7.5) containing PMSF (100 µg/mL), followed by sonication and centrifugation at 4°C to clear the lysate. The cleared lysate was purified using Promega HisLink™ Protein Purification resin by rotating at 4°C for 30 min. Subsequently, the resin was washed 3x and the protein eluted with elution buffer (100 mM HEPES, 400 mM imidazole, 100 µM GDP, pH 7.5). Collected fractions were analyzed for purity by SDS-PAGE and dialyzed in a buffer containing 20 mM Tris, 50 mM NaCl, 5 mM EDTA, 2 mM DTT, pH 7.5. GDP•Rac1 was generated by incubating free Rac1 with GDP (10-fold excess) at 30°C for 1 h. Mant-GDP•Rac1 was generated by incubating free Rac1 with methylanthraniloyl-GDP (5-fold excess mant-GDP, Life Technologies) at 30°C for 1 h. Nucleotide exchange was quenched by addition of 250-

fold excess of $MgCl_2$ and excess nucleotide was removed by dialysis into assay buffer (20 mM Tris, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 1% glycerol, pH 7.5). GTPase loaded with mant-GDP was stored at 4°C and used within 1 week. Rac1 loaded with GDP was aliquoted, snap-frozen and stored at -80°C.

**Guanine nucleotide exchange assays.** To evaluate inducible nucleotide exchange, 250 nM of a given CIAR construct was incubated with 200 nM MANT-GDP (Life Technologies, M-12414), and 25 μM or 5 μM of free BH3 peptide (APPNLWAAQRYGRELRRMSDEFEGSFK) or NOXA control peptide (PAELEVECATQLRRFGDKLNFRQKLL) in black 96-well plates in triplicate wells. Free peptides were synthesized by GenScript. 10μM HRAS-GDP was added and the change in fluorescence was monitored with a Victor$^3$ V Multilabel Plate Reader (Perkin Elmer, excitation at 355 nm and emission at 460 nm). Background exchange for HRAS•GDP was also determined in the absence of CIAR constructs. Final reaction volume for each assay well was 125 μL. Triplicate values for each time point were averaged. The rate of fluorescence change—a proxy for MANT-GDP association rate—was determined via linear regression of the linear phase for each assay. For determination of fold change, the following equation was used: $\frac{slope\ (BH3) - slope(HRAS\ alone)}{slope\ (control\ BH3) - slope(HRAS\ alone)}$.

VAV constructs were assayed by mixing the construct of interest (800 nM final concentration), mant-GDP•RAC1 (2 μM final concentration), and excess GDP (400 μM final concentration) in assay buffer (20 mM Tris, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 1% glycerol, pH 7.5), and 5 μM competitor or control BH3 peptide. The change in fluorescence was measured using a Victor$^3$ V Multilabel Plate Reader (Perkin Elmer, excitation at 355 nm and emission at 460 nm). Control experiments were performed that measured the rate of spontaneous nucleotide exchange in the absence of the sVAV construct; mixtures contained mant-GDP•RAC1 (2 μM final concentration) and excess GDP (400 μM final concentration). The final volume of each assay well was 125 μL. All assays were performed in triplicate in 96-well black plates. The rate of fluorescence change—a proxy for MANT-GDP dissociation rate—was determined via linear regression of the linear phase for each assay. For determination of fold change, the following equation was used: $\frac{slope\ BH3 - slope(RAC1\ alone)}{slope\ (control\ BH3) - slope(RAC1\ alone)}$.

**Cell culture and stable cell line generation.** Flp-In T-Rex 293 cells were maintained in high glucose DMEM, 10% FBS, 4 mM L-glutamine (Life Technologies). MCF10A lines were grown as previously described[108]. Doxycycline-inducible CIAR-293 and xCIAR-293 cells were generated according to manufacturer's instructions. Briefly, Flp-In T-REx 293 cells were transfected with S.17.3_pcDNA5/FRT/TO (CIAR) or xS.17.3_pcDNA5/FRT/TO (xCIAR), followed by selection with hygromycin for 7-10 days. After induction with 1 μg/mL doxycycline, expression of the desired transgenes was verified via immunoblot. CIAR-293 and xCIAR-293 cells were maintained in the same media as above supplemented with 50 μg/mL hygromycin and 10 μg/mL blasticidin.

Lentiviral expression vectors were packaged and target MCF10A cells (American Type Culture Collection (ATCC)) transduced according to standard protocols[109]. Briefly, HEK-293T cells were seeded and co-transfected with packaging vectors psPAX2 and pMD2.G, and the indicated pLenti transfer vector. Virus-containing conditioned media was collected 2 days after transfection and clarified using a 0.45 μM filter. Prior to infection, virus-containing media was diluted 1:4 in normal MCF10A growth media and the mix was supplemented with polybrene to a final concentration of 8 μg/mL. Subsequently, the viral transduction mix (5 mL) was added to target MCF10A cells, seeded to 30% confluence, and incubated for 24 hours. Virus-transduced MCF10A cells were given one day to recover and then selected in puromycin for 2 days. Expression of the desired transgenes was verified via immunoblot. Cell lines were confirmed to be free of mycoplasma.

**Luciferase-based Serum Response Element (SRE) Assay**. Flp-In T-REx 293 cells were plated in 96-well plates and transfected the following day with a given candidate CIAR construct, reporter luciferase plasmid (luc2P), and control luciferase plasmid (Renilla). Transfections were carried out using XTremeGENE-HP according to the manufacturer's instructions (Roche). After 4 h at 37ºC, cells were changed to serum free media supplemented with 1 μg/mL doxycycline. 24 h later, cells were changed to serum free media containing either A3 (25 μM or 10 μM), DMSO, or the MEK inhibitor AZD6244 (40 μM, Selleck Chemicals) and incubated at 37ºC. After 6 h, luc2P luciferase and Renilla luciferase activity was quantified using the Dual-Glo™ Assay System (Promega). Data was processed and analyzed as follows: (1) background signal present in non-transfected wells was subtracted from all wells, (2) intra-well normalization was performed by dividing luc2P signal by Renilla signal to control for variation in transfection efficiency, (3) residual signal present in the context of MEK-inhibition by AZD6244 was defined as zero RAS/MAPK activation and subtracted for each transfection. Finally (4), fold induction was calculated by dividing signal after A3 treatment by signal after DMSO treatment. All experiments were performed in triplicate.

**Immunoblotting.** CIAR-293 or xCIAR-293 cells were plated in 12- or 24-well plates in high glucose DMEM (10% FBS, 4 mM L-glutamine, 1 μg/mL doxycycline). The following day, cells were switched to serum-free media and serum starved overnight—except in **Fig 3e**, in which cells were not serum starved. Cells were then treated with A3, EGF, or DMSO for the indicated duration. For experiments incorporating vemurafenib, cells were treated with vemurafenib or DMSO for 1 h prior to addition of A3 or DMSO. MCF10A-CIAR cells were plated in 6-well plates in full MCF10A media conditions as previously described[54]. Cells were starved overnight in 950ml serum-free and EGF-free MCF10A media prior to stimulation.

Lysates were collected by placement of the plate onto ice, washed with ice cold DPBS, and lysed in a modified RIPA buffer (50 mM Tris-HCL, pH 7.8, 1% IGEPAL CA-630, 150 mM NaCl, 1mM EDTA, 2 mM $Na_3VO_4$, 30mM NaF, Pierce Protease Inhibitor Tablet). Cleared lysates were subjected to SDS-PAGE and transferred to nitrocellulose. Blocking and antibody incubation were done in TBS with 0.1% Tween-20 (v/v) and

blocking buffer (Odyssey). Primary antibodies were diluted: total ERK (1:4,000; #9107), phoshpho-ERK (1:2,000; #4370), phospho-AKT (pS473, 1:2,000; #4060), phospho-AKT (T303, 1:2,000; #13038), AKT (1:2,000; #2920), phospho-EGFR (Y1068, 1:2,000; #2236), and α-tubulin (1:20,000; #3873) (all Cell Signaling Technology); RAF1 (1:2,000, BD Transduction Labs #610151). Blots were washed in TBS with 0.1% Tween-20. Antibody binding was detected by using near-infrared-dye-conjugated secondary antibodies and visualized on the LI-COR Odyssey scanner. Blots were quantified via densitometry with Image Studio™ (LI-COR). In cases where quantification across multiple blots was required, a standard was included for intergel normalization. These blots were incubated with the same antibody dilutions in parallel and imaged with identical Odyssey scanner settings.

**RAS binding domain pulldown assays**
Cells were plated in 10 cm dishes and serum starved as above.
Following treatment with A3 (10 μM), EGF (10 ng/mL), or DMSO (0 min sample), cells were placed on ice at the appropriate time points, washed with ice cold DPBS, and lysed. RAS binding domain (RBD) pulldown assays were performed using the Pan-RAS activation kit (Cell Biolabs) according the manufacturer's instructions. A representative experiment is shown in Fig. 4a, n = 3 (except n = 2 for 0 min and 120 min).

**Sample preparation for phosphoproteomic analysis.** Cells were plated in 10 cm or 15 cm dishes and serum starved as above. Following A3 or EGF treatment, plates were placed on ice, washed with ice cold DPBS, and lysed in ice cold 8M Urea. Protein concentration was determined by the BCA assay (Pierce). Samples were reduced with 5 mM DTT at 56 $^{0}$C for 1 hour, then alkylated with 15 mM iodoacetamide for 1 hour at room temperature in the dark. Samples were diluted 4-fold with 100 mM ammonium acetate, pH 8.9, and digested with Sequencing Grade Modified Trypsin (Promega) at a ratio of 1:100 (trypsin to total protein), overnight at room temperature. Following digestion, peptides were desalted and concentrated using Sep-Pak Plus C18 cartridges (Waters) according to the manufacturer's recommendations. Samples were then dried by vacuum centrifugation, lyophilized, and stored at -80 $^{0}$C until further processing.

**Isobaric labeling of peptide samples.** Tryptic peptide samples were labeled with 8-plex iTRAQ reagents (AB Sciex). Lyophilized peptides derived from approximately 400 ug of starting material were resuspended in 30 μL of dissolution buffer (0.5 M $N(Et)_3HCO_3$, pH 8.5-9). iTRAQ labels were resuspended in 70 μL of isopropanol and added to the peptide mixture. Samples were incubated at room temperature for 2 hours, combined, and dried overnight by vacuum centrifugation. The following day, samples were desalted and concentrated using Sep-Pak Vac 1cc (50 mg) cartridges (Waters) according to the manufacturer's recommendations. Samples were then dried by vacuum centrifugation, lyophilized, and stored at -80 $^{0}$C until further processing.

**IMAC.** Approximately 100 μL of packed Ni beads (Ni-NTA Superflow beads, Qiagen) were washed three times in water and stripped with 100 mM EDTA, pH 8.9, for 30 min. Beads were then washed three times with water and incubated with 100 mM $FeCl_3$ for 30 min. Beads were then washed three times with water and once with 80% acetonitrile

(ACN) in 0.1% trifluoroacetic acid (TFA). Lyophilized iTRAQ samples were resuspended in 1.5 mL of 80% ACN in 0.2% TFA and incubated with prepared beads for 1 hour at room temperature. Beads were then washed three times with 80% ACN in 0.1% TFA, and phosphopeptides were eluted from beads twice with 75 µL of 1.4% ammonia. Samples were then vacuum centrifuged down to approximately 20 µL. 2 µL of 200 mM ammonium formate, pH 10, was added and samples were directly analyzed by mass spectrometry.

**Mass spectrometry.** Peptide samples were loaded onto a first dimension trap column (Waters Xbridge, C18, 10 um particle size, 100 Å pore size, 4 cm packing length 150 µm column inner diameter). Online peptide separation coupled to MS/MS was performed with a 2D-nanoLC system (nanoAcquity UPLC system, Waters) and a Velos-Pro/Orbitrap-Elite hybrid mass spectrometer (ThermoFisher Scientific). Six discrete elutions were performed at 1.5 µL/min with 5 mM ammonium formate, pH 10, using increasing concentrations of ACN (1%, 3%, 6%, 15%, 25% and 44%) and diluted with 6 µL/min 0.1% formic acid (FA) prior to loading onto a second dimension trap column (Dr. Maisch ReproSil-Pur, C18, 5 um particle size, 120 Å pore size, 4 cm packing length 150 µM column inner diameter) connected to an analytical column (Orochem Reliasil, C18, 3 µM particle size, 90 Å pore size, 20-25 cm packing length 50 µM column inner diameter) with an incorporated electrospray emitter. Peptide separation was achieved using a gradient from 3 to 80% (V/V) of ACN in 0.1% FA over 115 minutes at a flow rate of 200 nL/min. The mass spectrometer was operated in data-dependent mode using a Top 10 method. Full MS scans (m/z 300-2000) were acquired in the Orbitrap analyzer (resolution = 120,000), followed by high energy collision induced dissociation (HCD) MS/MS (fm/z 100-2000, resolution = 15,000) at a normalized collision energy of 35%.

**Data Processing and Analysis.** MS data files were searched using the COMET algorithm[110] and the output was imported into the Trans-Proteomic Pipeline[111] with the following parameters: variable oxidation of methionine, variable phosphorylation of serine, threonine, or tyrosine, up to 4 variable modifications per peptide, fixed oxidation of cysteine, and fixed iTRAQ labeling of lysines and the N-terminus, two missed cleavages, maximum charge of 7. Peptide false discovery rate (FDR) was set to 2.5%. Peptide quantification based on the iTRAQ labels was determined using the LIBRA software embedded in the Trans-Proteomic Pipeline[112].

For Principal Components Analysis (PCA), abundance data was log transformed and scaled to have unit variance. PCA was performed using the prcomp function in the R Stats package. Hierarchical clustering and heatmap generation of timecourse data (Fig. 3.3c) was performed in GENE-E software (Broad Institute) using Euclidean distance and complete linkage clustering. Protein interaction network analysis was carried out using the STRING interactions database (http://string-db.org/).

**Parallel Reaction Monitoring.** Unlabeled phosphopeptides were first enriched by IMAC. Samples were then loaded onto a trap column (Dr. Maisch ReproSil-Pur, C18, 5 µm particle size, 120 Å pore size, 4 cm packing length 150 µm column inner diameter)

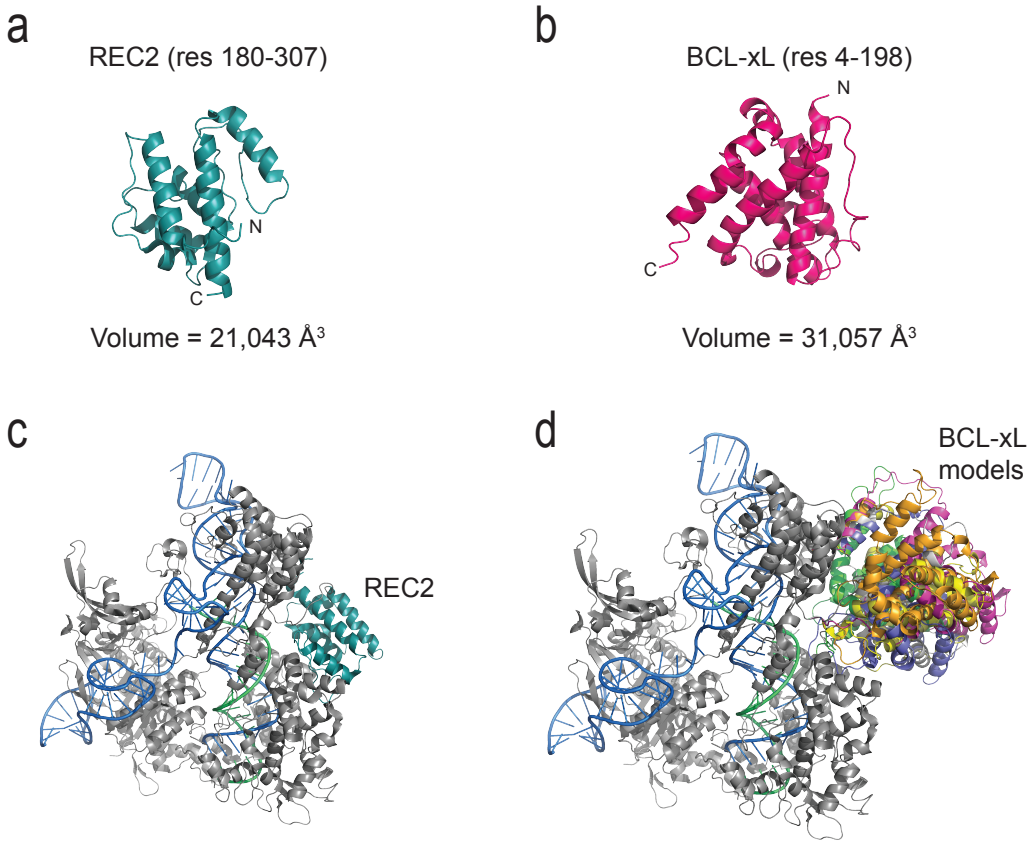
connected to an analytical column (Orochem Reliasil, C18, 3 um particle size, 90 Å pore size, 20-25 cm packing length 50 um column inner diameter) with an incorporated electrospray emitter. Peptide separation was achieved using a gradient from 3 to 80% (V/V) of acetonitrile in 0.1% FA over 60 minutes at a flow rate of 200 nL/min. Parallel Reaction Monitoring (PRM) analyses were performed on a Q-exactive Plus (Thermo). The acquisition method combined 5 sequential PRM events followed by a full scan event. Precursor ions of the peptides were targeted in preliminary experiments without time-scheduling and confirmed precursors were targeted +/- 3 minutes of the observed elution times in the actual experimental runs. PRM events employed an orbitrap resolution of 17,500, a target automatic gain control (AGC) value of 2e5, and maximum fill times of 100 ms. The precursor ion of each targeted peptide was isolated using a 4-m/z window. Fragmentation was performed with a normalized collision energy of 35 eV. The full scan event employed a 200-1750 m/z range, an orbitrap resolution of 70,000, a target AGC value of 1e6, and maximum fill times of 200 ms. Data analysis integration was performed with Skyline software. Each peptide was confirmed using at least 3 transitions and 3 isotopic peaks for each peptide were summed for quantification. Each peptide was normalized to a peptide containing TJP1 S912, which was found to be unresponsive to A3 and EGF in initial analyses.

**Statistical Analysis.** Statistical tests of immunoblot densitometry were performed in excel using a one-sided two-sample Student's t-test. For phosphoproteomic analysis, phosphopeptides present in all replicates were determined to have significant changes (Fig. 3.4a) in abundance for a given condition, *X*, if $\left|\log_2 \frac{abundance\ (X)}{abundance\ (DMSO)}\right|$ >0.5 and p <0.05, where p was calculated with a two-tail heteroscedastic Student's t-Test. KEGG pathway enrichment of the phosphoproteomic timecourse experiment was assessed with the WEB-based GEne SeT AnaLysis Toolkit using a hypergeometric test with Benjamini-Hochberg multiple-test correction[105,106]. The 485 proteins quantified in all three biological replicates served as the reference set.

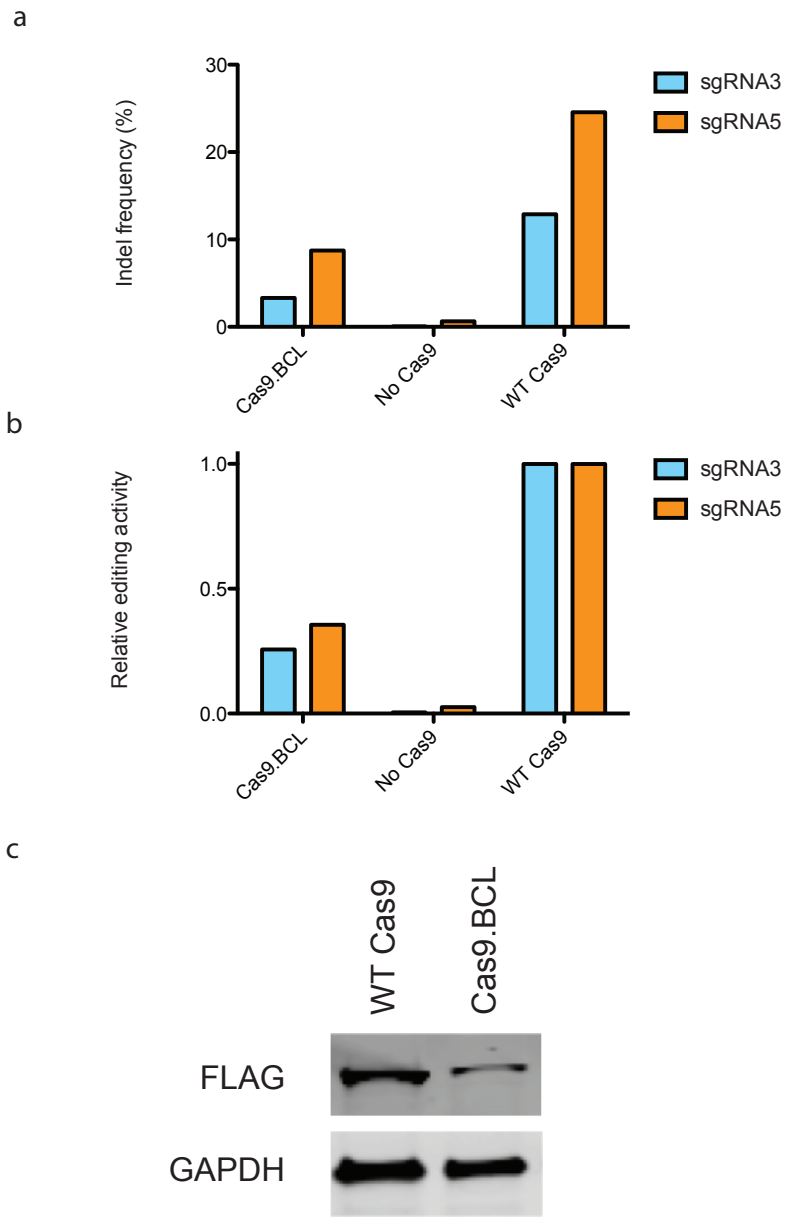
## Appendix B: Supplementary Material for Chapter 3

**Supplementary Figure B.1 |** Comparison of REC2 and BCL-xL domains and modeling domain replacement.
**Supplementary Figure B.2 |** A Cas9 variant in which the REC2 domain was replaced with BCL-xL retains activity
**Supplementary Figure B.3 |** Schematic depiction of Cas9 constructs generated and tested in this study
**Supplementary Figure B.4 |** Indel frequency at the AAVS1 locus as a function of ciCas9 linker length after 24 hours in the presence or absence of A3
**Supplementary Figure B.5 |** ciCas9 enables inducible genome editing in U2OS and HCT116 cells
**Supplementary Figure B.6 |** Comparison of indel kinetics at the EMX1 locus with ciCas9 and paCas9
**Supplementary Figure B.7 |** Comparison of indel kinetics at the EMX1 locus with ciCas9 and iCas9
**Supplementary Figure B.8 |** Proteinase K effectively digests ciCas9 prior to DNA isolation for DSB-ddPCR and high-throughput sequencing
**Supplementary Figure B.9 |** Evaluation of background signal throughout DSB-ddPCR timecourse experiments
**Supplementary Figure B.10 |** Determination of initial DSB generation rate for MYC sgRNA4 and sgRNA5
**Supplementary Figure B.11 |** High-resolution indel timecourse corresponding to DSB timecourses in (Fig 4d)
**Supplementary Figure B.12 |** ciCas9 exhibits improved editing specificity relative to wild-type Cas9
**Supplementary Figure B.13 |** A ciCas9 variant (e-ciCas9) with specificity enhancing mutations to minimize off-target activity
**Supplementary Figure B.14 |** Alterations in BH3/BCL-xL affinity do not affect ciCas9 variant expression
**Supplementary Figure B.15 |** A3 does not affect DSBs, indels, or cell viability in the context of wild type Cas9-mediated editing
**Supplementary Table 1 |** Comparison of inducible Cas9 variants
**Supplementary Table 2 |** Comparison of methods for detecting DSBs
**Supplementary Table 3 |** Indel kinetics for multiple sgRNAs
**Supplementary Table 4 |** *P*-values for comparisons to respective zero hour time point
**Supplementary Table 5 |** BH3 variant peptide sequences and BCL-xL $K_i$'s
**Supplementary Note |** Peptide Sequences for constructs and domains
**Supplementary Data Set |** Primer, probe, and sgRNA sequences

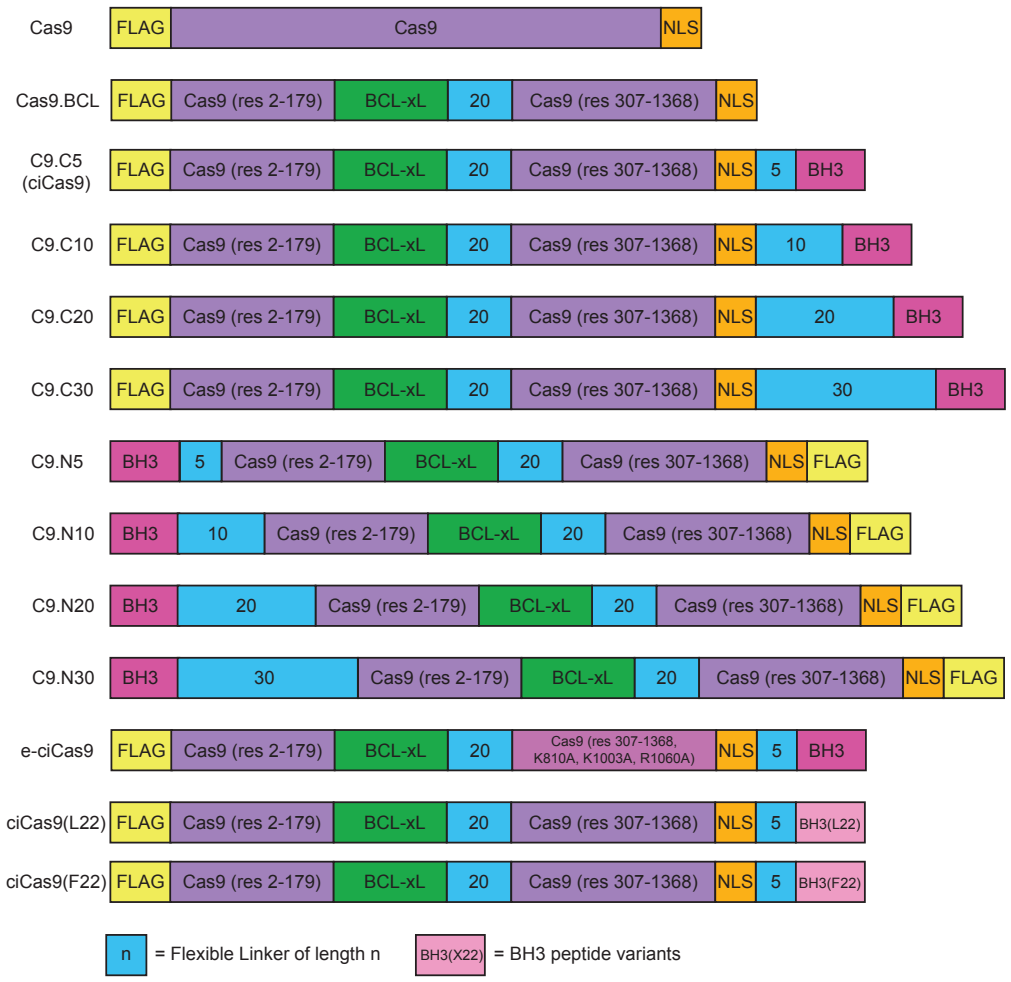
## Supplementary Figures



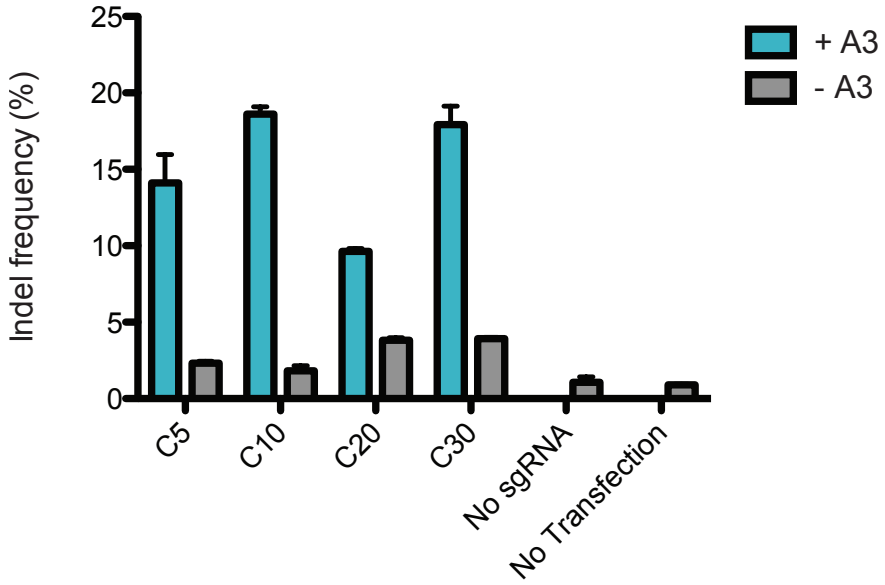

**Supplementary Figure B.1 | Comparison of REC2 and BCL-xL domains and modeling domain replacement. (a)** The REC2 domain of Cas9 and **(b)** BCL-xL display similar globular structures and volumes. **(c, d)** Modeling the replacement of the REC2 domain with BCL-xL. Six BCL-xL models are shown (multiple colors). In order to accommodate differences in relative termini orientation between BCL-xL and the REC2 domain, a flexible linker (length = 15 residues) was used to connect residue 307 of Cas9 to BCL-xL, while the N-terminus of BCL-xL was connected directly to residue 179 of Cas9. Cas9.BCL constructs containing 15 versus 20 residue linkers exhibited no appreciable difference in activity (data not shown).



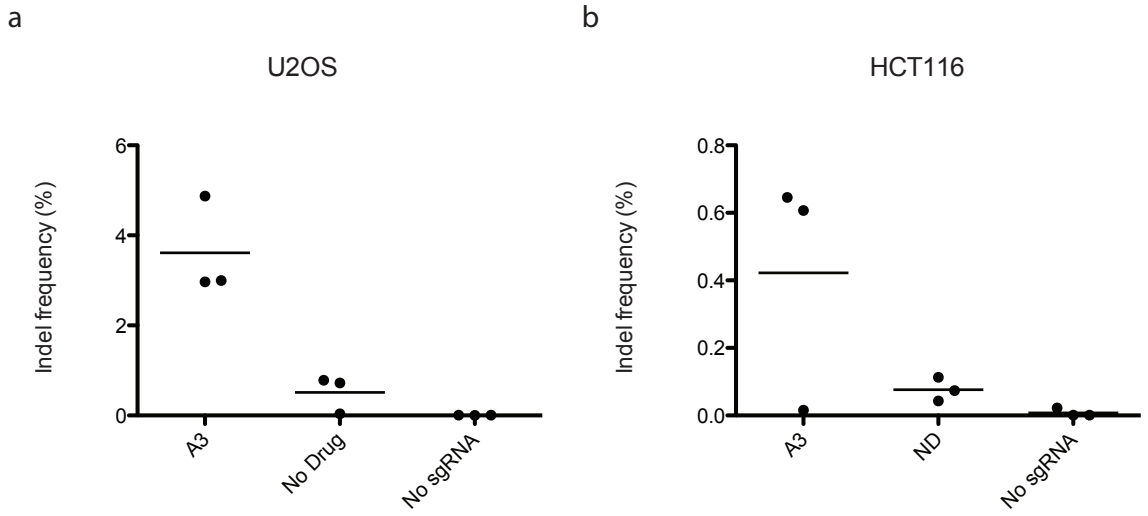

**Supplementary Figure B.2 | A Cas9 variant in which the REC2 domain was replaced with BCL-xL retains activity.** (**a, b**) HEK-293T cells were transfected with Cas9.BCL, wild type Cas9, or no Cas9, in addition to one of two sgRNAs targeting the MYC locus (sgRNA3 or sgRNA5) (n =1). Indel quantification was performed using an analysis method similar to that previously described (Online Methods)[45]. These data indicated Cas9 variants in which the REC2 domain has been replaced by BCL-xL retain editing activity. (**b**) HEK-293T cells were transfected with plasmids encoding Cas9 or Cas9.BCL, each bearing a FLAG-tag, and cells were lysed after 24 hours. Expression was detected using an anti-FLAG antibody (Cell Signaling), and an anti-GAPDH (loading control) antibody.



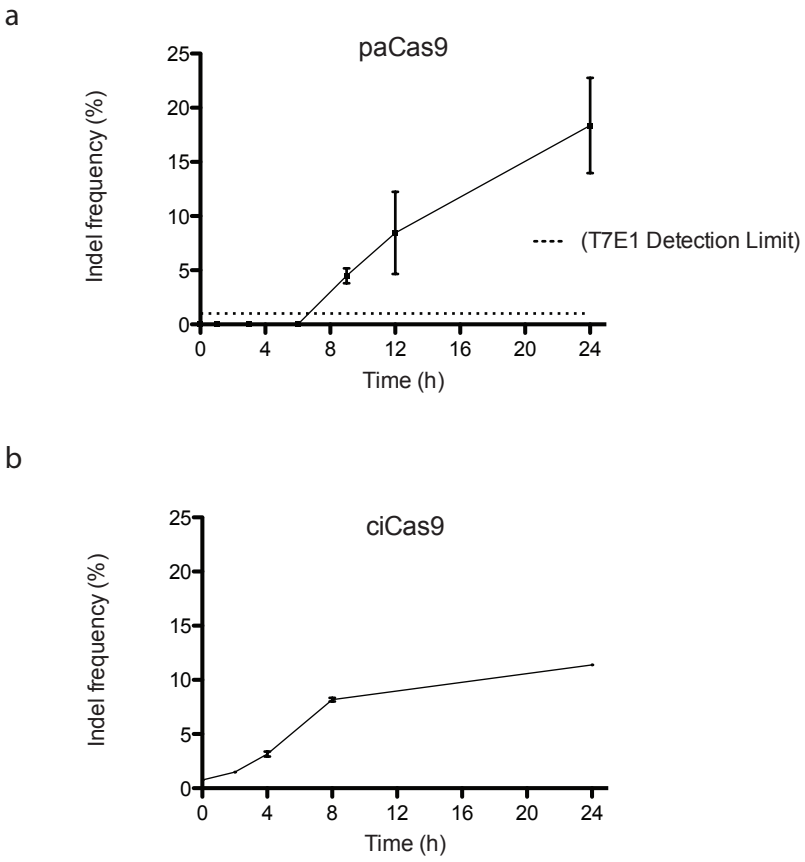

**Supplementary Figure B.3 | Schematic depiction of Cas9 constructs generated and tested in this study.** Constructs contained residues 4-198 of BCL-xL (green); the BH3 peptide sequence (magenta); a nuclear localization signal (NLS, orange); FLAG tag (yellow); and human codon optimized wild type *S. pyogenes* Cas9 (purple). Flexible linkers ranging in length from 5 to 30 amino acids (blue) are composed of glycine, serine, threonine, and alanine. NLS = nuclear localization sequence.



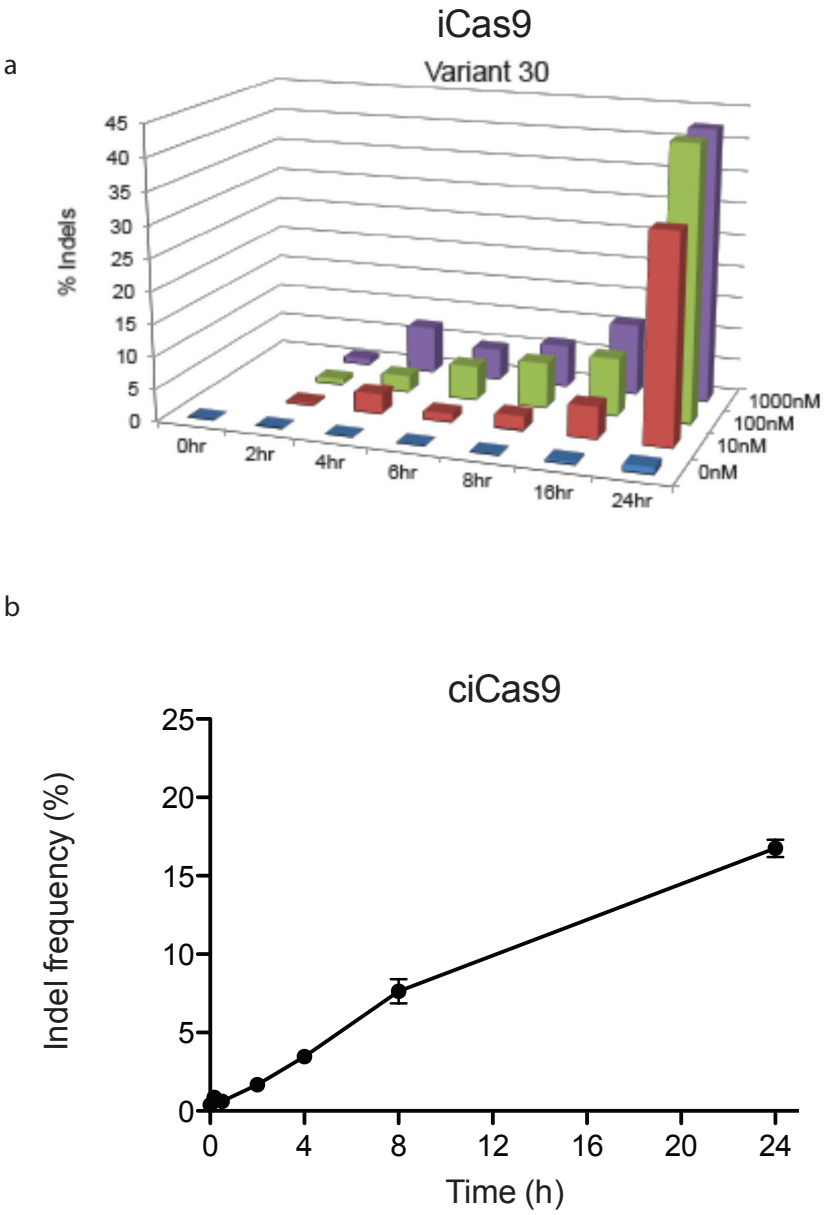

**Supplementary Figure B.4 | Indel frequency at the AAVS1 locus as a function of ciCas9 linker length after 24 hours in the presence or absence of A3.** HEK-293T cells were transfected using Turbofectin 8.0 with (450 ng Cas9 construct, 450 ng AAVS1 sgRNA, and 100 ng pMAX-GFP) or (450 ng ciCas9 and 550 ng pMAX-GFP), or left untransfected. 24 hours after transfection, cells were treated with A3 (10 μM) or left untreated. 24 hours after drug treatment, cells were harvested, and indel were quantified as described in the Online Methods. Error bars depict the s.e.m. (n = 3 biological replicates).



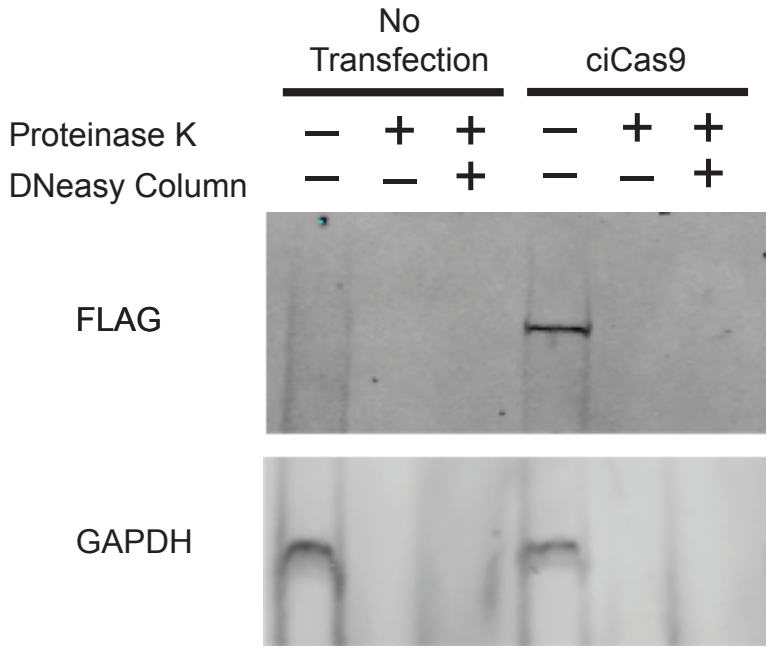

**Supplementary Figure B.5 | ciCas9 enables inducible genome editing in U2OS and HCT116 cells.** U2OS and HCT116 cells were transfected using Turbofectin 8.0 with (450 ng ciCas9, 450 ng AAVS1 sgRNA, and 100 ng pMAX-GFP) or (450 ng ciCas9 and 550 ng pMAX-GFP) as a no sgRNA control. 24 hours after transfection, cells were treated with A3 (10 μM final concentration) or left untreated. Cells were harvested 24 hours after drug addition and indels quantified as described in the Online Methods. Indel frequencies (%) for individual biological replicates (n = 3) are shown (solid line = mean).



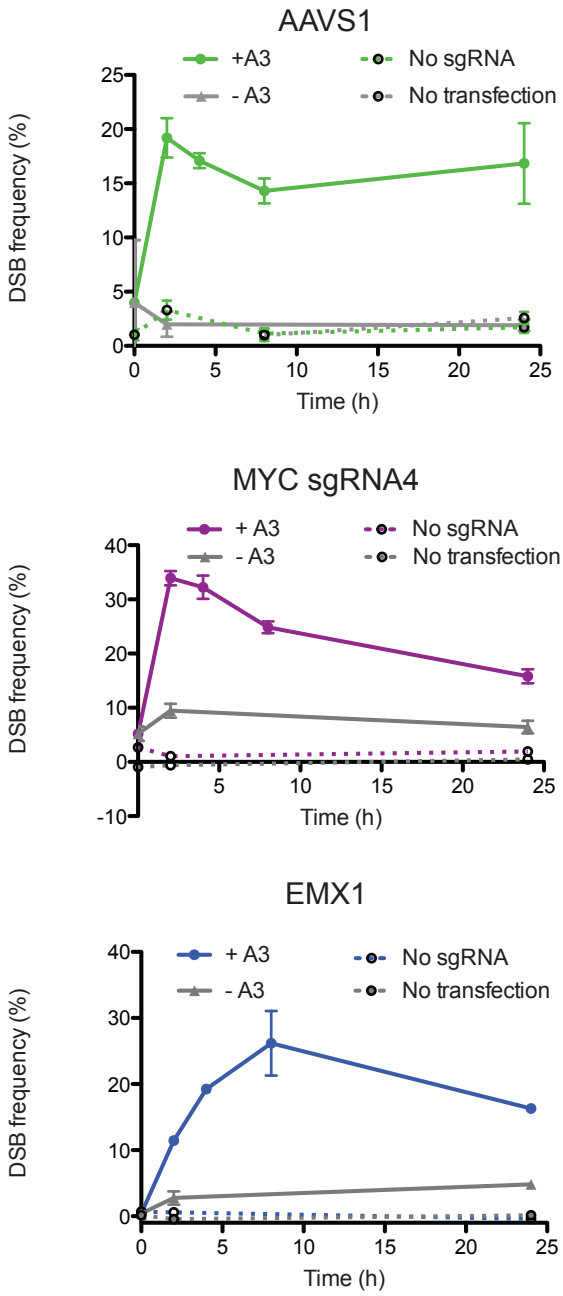

**Supplementary Figure B.6. Comparison of indel kinetics at the EMX1 locus with ciCas9 and paCas9. (a)** Indel kinetics are shown as reported for paCas9 co-transfected with EMX1 sgRNA in HEK-293T cells[68]. Indel frequencies were determined using T7E1 assay, which the authors note has a detection limit of 1%. Error bars depict the s.e.m. (n=4 biological replicates) **(b)** HEK-293T cells were transfected with ciCas9 and an identical EMX1 sgRNA. After 24 hours, cells were treated with A3 and harvested at the indicated time points. Indel frequencies were quantified using high-throughput sequencing, as described in the Online Methods, and are also plotted in Fig. 4c. Indels were detectable by 2 hours, the first time point tested, and this increase in indel frequency was significant relative to both the zero minute and two hour no drug time points (one-sided t-tests: n = 3, p = 0.00045 and p = 0.00056, respectively). Error bars depict the s.e.m (n = 3 biological replicates).



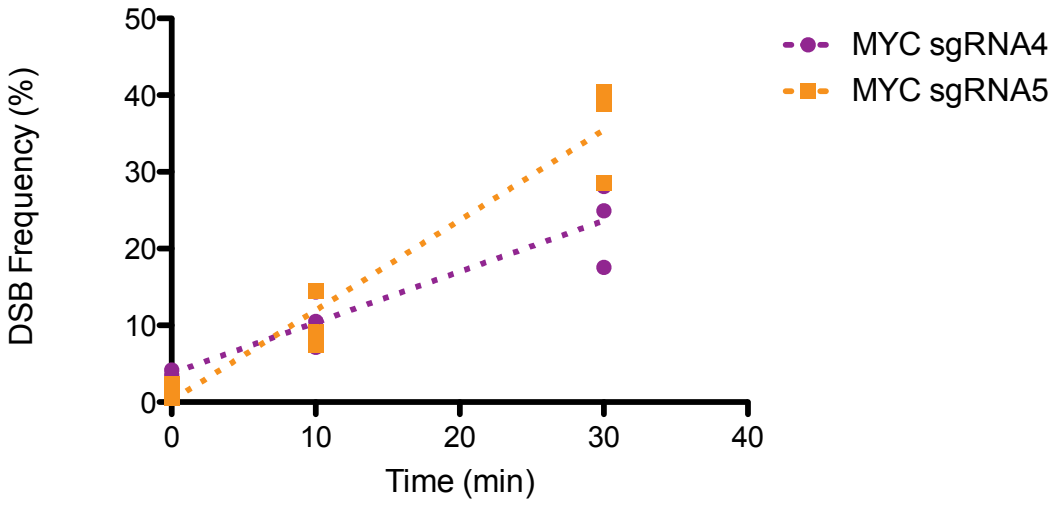

**Supplementary Figure B.7 | Comparison of indel kinetics at the EMX1 locus with ciCas9 and iCas9. (a)** Reproduction of previously reported EMX1* sgRNA kinetics with iCas9 in HEK-293 cells[67]. No statistics were reported. **(b)** HEK-293T cells were transfected with ciCas9 and an identical EMX1* sgRNA. After 24 hours, cells were treated with A3 and harvested at the indicated time points. Indel frequencies were determined using high-throughput sequencing, as described in the Online Methods. Indels were detectable with ciCas9 at 30 min and significant relative to the zero minute time point (one-sided t-test, n = 3, p = 0.016). Error bars depict the s.e.m. (n = 3 biological replicates). The *EMX1 sgRNA used in these experiments has an additional transcribed G preceding the 20 bp target sequence, but it is otherwise identical to the EMX1 sgRNA used throughout this study.



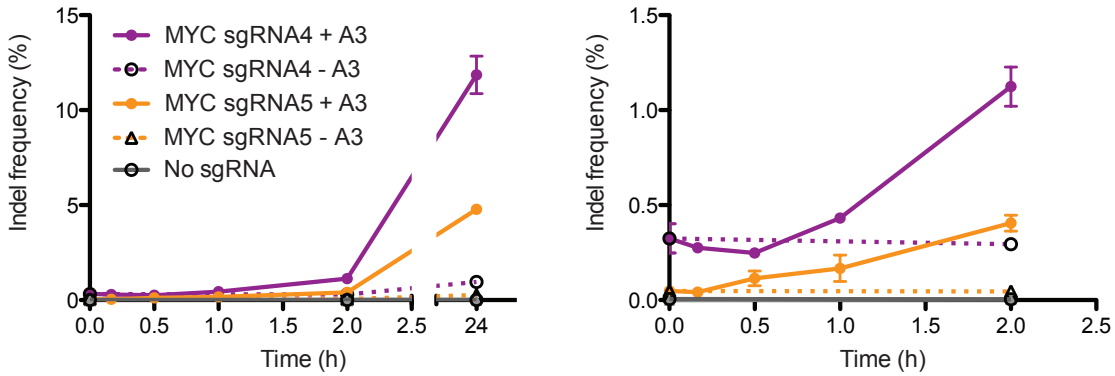

**Supplementary Figure B.8 | Proteinase K effectively digests ciCas9 prior to DNA isolation for DSB-ddPCR and high-throughput sequencing.** HEK-293T cells were transfected with 450 ng ciCas9, 450 ng EMX1 sgRNA, and 100 ng pMAX-GFP per well in 12-well plates. Twenty four hours after transfection, cells were incubated with A3 (10 μM) for two hours, then harvested. Non-transfected wells were harvested as a control. After resuspension in 1:1 mixture of PBS and Buffer AL (DNeasy Blood & Tissue Kit, Qiagen), samples were split into thirds. One third was processed according to the DNA isolation protocol as described in Online Methods, including 1 hour proteinase K digestion at 56ºC and DNA purification via spin-column (Proteinase K +, DNeasy Column +). One third was digested with proteinase K for 1 hour at 56ºC only (Proteinase K +, DNeasy Column -), and one third of the cells underwent only the first step in processing—resuspension in PBS and Buffer AL (DNeasy kit)—but were not treated with Proteinase K or run on DNeasy column. Protein was then isolated from each sample by chloroform/methanol precipitation, resuspended in SDS-PAGE denaturing sample buffer, boiled, and western blotted. ciCas9 expression was evaluated using anti-FLAG antibody (Sigma), and anti-GAPDH used as a loading control. ciCas9 expression was undetectable after proteinase K digestion. n = 3 biological replicates for ciCas9, n = 1 for no transfection control.



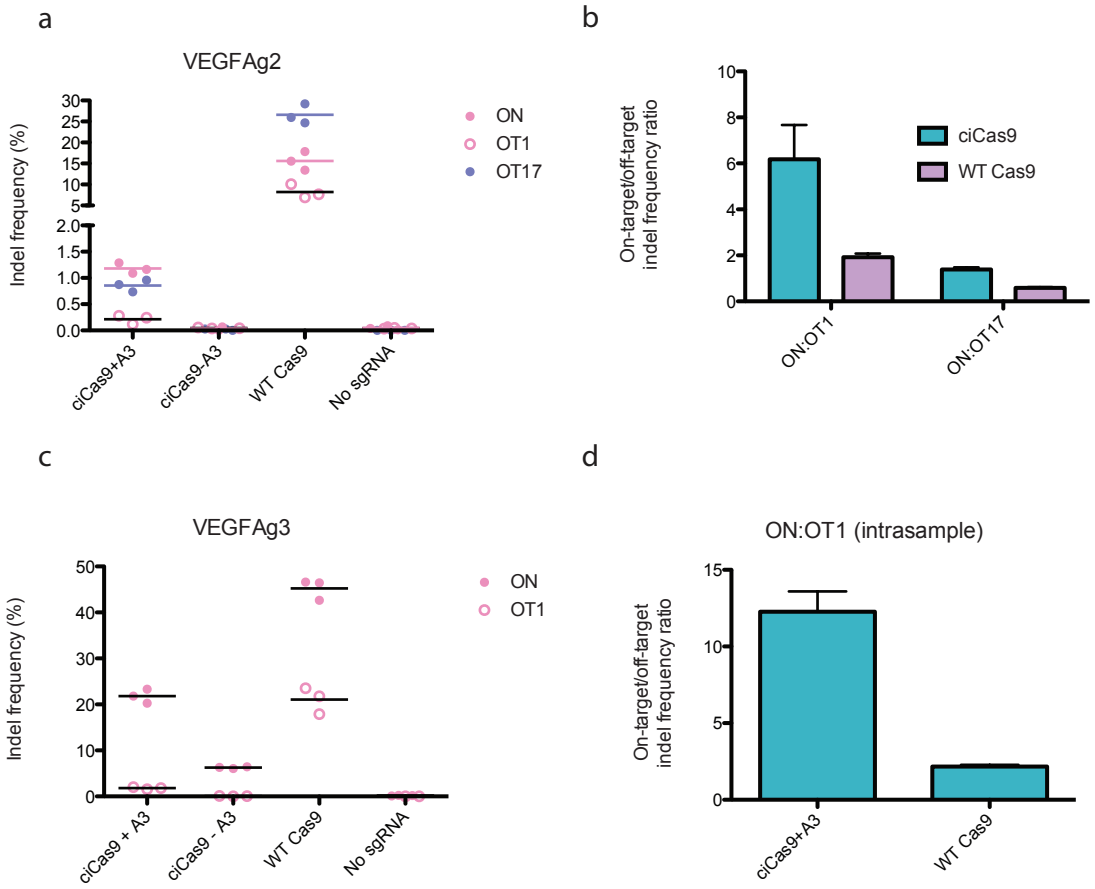

**Supplementary Figure B.9 | Evaluation of background signal throughout DSB-ddPCR timecourse experiments.** DSB frequency curves from Fig. 4c are plotted with corresponding no sgRNA and no transfection controls. Error bars depict the s.e.m. (n = 3 biological replicates).



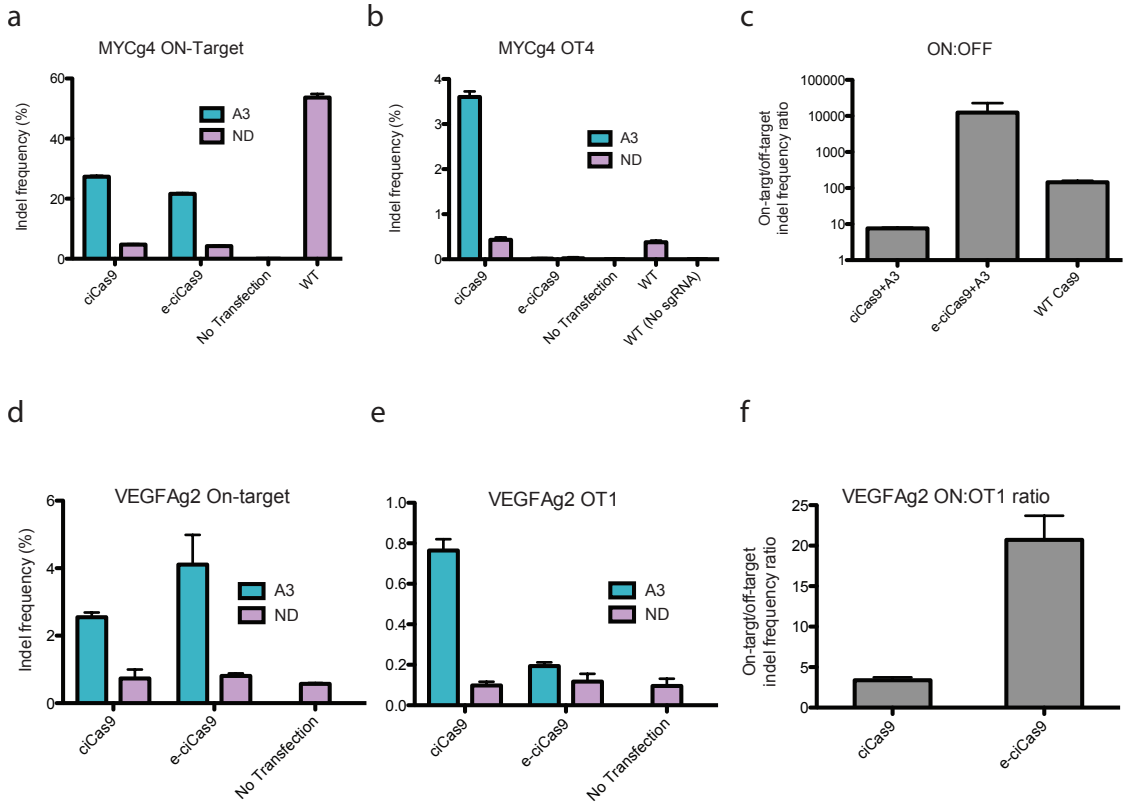

**Supplementary Figure B.10 | Determination of initial DSB generation rate for MYC sgRNA4 and sgRNA5.** Initial DSB generation rates were calculated by performing a linear regression over early time points (0-30 min, Fig.4d) MYC sgRNA4 (slope = 0.66, standard error = 0.094; $r^2$ = 0.88) and sgRNA5 (slope 1.2, standard error = 0.11; $r^2$ = 0.94). The difference in initial rates is significant (analysis of covariance (ANCOVA), p = 0.0035). Linear regressions and statistical calculations were performed in Prism software. Individual biological replicates are plotted at each time point for each sgRNA, n = 3. Dashed lines show the fitted regression lines.



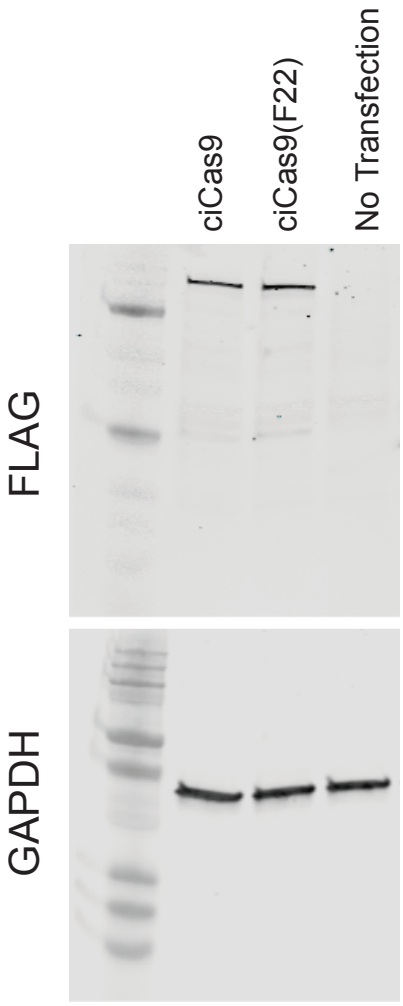

**Supplementary Figure B.11 | High-resolution indel timecourse corresponding to DSB timecourses in (Fig 4d).** Statistically significant indels were detectable at 2 hours for both MYC sgRNA4 ($p = 0.0017$) and MYC sgRNA5 ($p=0.00079$) relative to t = 0min. The increase in indel frequency was also significant relative to no drug ("- A3") at 2 hours (MYC sgRNA4 $p = 0.00071$; MYC sgRNA5 $p = 0.00062$). All p-values were calculated with one-sided t-tests, n = 3, error bars depict the s.e.m.



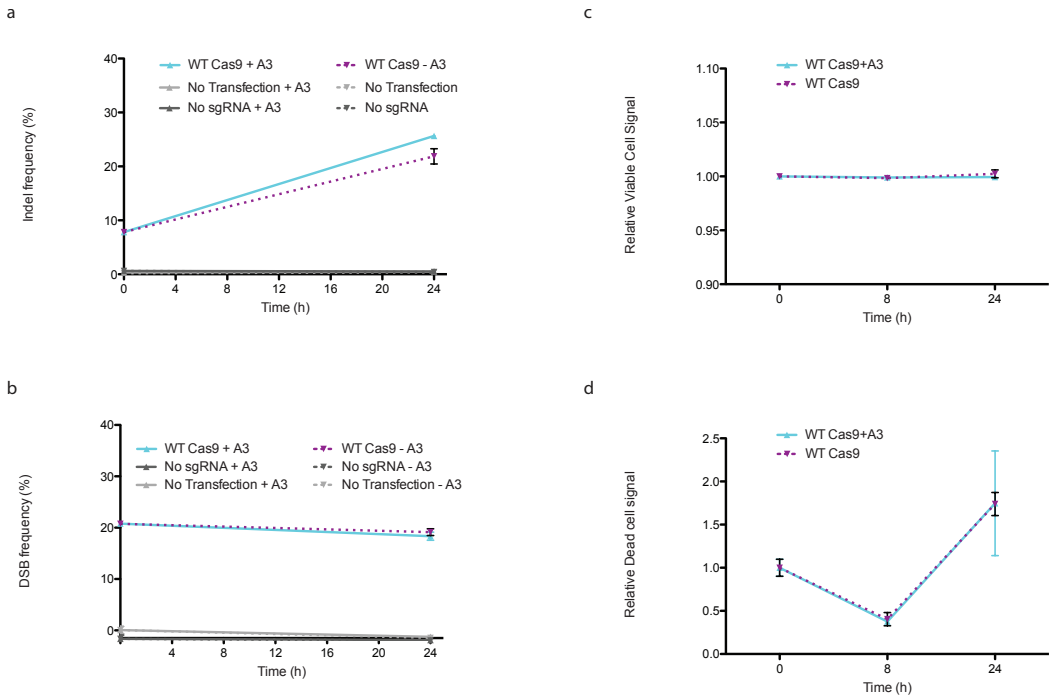

**Supplementary Figure B.12 | ciCas9 exhibits improved editing specificity relative to wild-type Cas9.** HEK-293T cells were transfected using Turbofectin 8.0 with (450 ng ciCas9, 450 ng VEGFA sgRNA2 or VEGFA sgRNA3, and 100 ng pMAX-GFP) or (450 ng ciCas9 and 550 ng pMAX-GFP) as a no sgRNA control. 24 hours after transfection cells were treated with A3 (10 μM final concentration) or left untreated. Cells were harvested 24 hours after drug addition and indels quantified as described in the Online Methods. Indel frequencies (%) for individual biological replicates (n = 3) are shown (solid line = mean). For specificity ratios (on-target/off-target indel frequency ratios) error bars depict the s.e.m (n = 3 biological replicates).




**Supplementary Figure B.13 | A ciCas9 variant (e-ciCas9) with specificity enhancing mutations to minimize off-target activity.** Three mutations (K810A, K1003A, R1060A)[62] were introduced to generate enhanced specificity ciCas9 (e-ciCas9). HEK-293T cells were co-transfected with MYC sgRNA4 (MYCg4) or VEGFA sgRNA2 (VEGFAg2), and either ciCas9 or e-ciCas9. Cells were then incubated either in the presence or absence of A3 for 24 hours before harvesting. Indel frequencies at the on-target site and a prominent off-target site were determined via high throughput sequencing. **(a)** On-target and **(b)** off-target 4 (OT4) indel frequencies (%) for MYC sgRNA4. **(c)** Specificity ratios for MYC sgRNA4 on-target and OT4 editing. **(d)** On-target and **(e)** off-target 1 (OT1) indel frequencies (%) for VEGFA sgRNA2. **(c)** Specificity ratios for VEGFA sgRNA2 on-target and OT1 editing. Error bars = s.e.m., n = 3 biological replicates.




**Supplementary Figure B.14 | Alterations in BH3/BCL-xL affinity do not affect ciCas9 variant expression.** HEK-293T cells were transfected with 450 ng ciCas9 or ciCas9(F22), 450 ng EMX1 sgRNA, and 100 ng pMAX-GFP per well in 12-well plates using Turbofectin 8.0. Non-transfected wells were included as a control. Cells were harvested and lysed 48 hours after transfection and ciCas9 variant expression was determined by western blot using antibodies targeting FLAG (Sigma) and GAPDH (loading control). Blots representative of two independent experiments.




**Supplementary Figure B.15 | A3 does not affect DSBs, indels, or cell viability in the context of wild type Cas9-mediated editing.** HEK-293T cells were transfected with 225 ng of plasmid encoding wild type Cas9, 225 ng of plasmid encoding AAVS1 sgRNA, and 50 ng of plasmid encoding pMAX-GFP per well in 24-well plates using Turbofectin 8.0. 24 hours after transfection cells were treated with A3 (10 μM final concentration) or left untreated. **(a-b)** Cells were harvested at 0 and 24 hours after drug addition. **(a)** Indels were quantified via high-throughput sequencing, and **(b)** DSB frequency was determined via ddPCR. In parallel, **(c)** cell viability and **(d)** cell death relative to WT Cas9 (– A3) at zero hours were determined using the LIVE/DEAD Viability/Cytotoxicity Kit (ThermoFisher). Error bars depict s.e.m. (n = 3 biological replicates).

## Supplementary Tables

| Construct | Gene size (bp) | Gene size (kDa) | Mechanism of activation | Activator | Editing Kinetics Reported | Earliest Timepoint at which indels detected | Single-component | Dose-dependent editing demonstrated | Specificity relative to WT Cas9 | Reference |
|---|---|---|---|---|---|---|---|---|---|---|
| ciCas9 | 4521 | 173 | Disruption of autoinhibitory BCL-xL/BH3 Protein-peptide interaction | A-385358, WEHI-539, ABT-737, et al. | Indels + DSBs | 30 min (AAVS1, VEGFAg2, EMX1*) 2 h (MYC | Yes | Yes | Improved | This report |
| iCas9 | 8058 | 305 | ERT2-mediated nuclear localization | 4-HT | Indels (no statistics, errors, or clear indication of | 2h (EMX1, no statistics or replicates reported) | Yes | No* | Improved | [67] |
| intein-Cas9 | 5451 | 208 | Excision of intein | 4-HT | Indels (12 - 96h), reported in [67] | Not tested prior to 12 h | Yes | No | Improved | [71] |
| split-Cas9 | 4974 | 189 | Reconstitution of split-Cas9 via induced dimerization | Rapamycin | Indels (12 - 96h), reported in [67] | Not tested prior to 12 h | No | No | Not reported | [78] |
| arC9 | 4905 | 189 | Allosteric effect of 4HT binding | 4-HT | No | NA | Yes | Yes | Not reported | [113] |
| paCas9 | 5469* | 207 | Reconstitution of split-Cas9 via induced dimerization | Blue light | Indels | 9h (EMX1) | No | No | Not reported | [68] |
| split-Cas9-ERT | 6924* | 263 | Reconstitution of split-Cas9 via Rapamyxin induced dimerization + 4- | Rapamycin and/or 4HT | No | NA | No | Demonstrated for transcriptional control, but not | Not reported | [114] |
| DIG-Cas9, PRO-Cas9 | 4566 | 176 | Small-molecule mediated stabilization of destabilized ligand binding | Digoxin, Progesterone | No | NA | Yes | No | Not reported | [115] |
| DD-Cas9 | Not reported | Not reported | Small-molecule mediated stabilization of destabilized ligand binding | Shield-1 (rapalog) | No | NA | Yes | No | Not reported | [116] |
| DD regulated Cas9 (DD.SpCas9.DD) | 5175 (DHFR), 5697 (ER50) | 198 (DHFR), 218 (ER50) | Small-molecule mediated stabilization of destabilized ligand binding | Trimethoprim (DHFR), 4HT (ER50), | No | NA | Yes | Yes | Improved | [117] |

**Supplementary Table 1 | Comparison of inducible Cas9 variants.** NA – not applicable. *Size (bp) includes 2 stop codons. Indicated split-Cas9 constructs delivered via 2 plasmids.

| Method | Site specific | Multiplex | Absolute quantitation demonstrated | References |
|---|---|---|---|---|
| Immunodetection | No | Yes | No | [79] |
| Comet | No | No | No | [80] |
| Pulse-field gel electrophoresis | No | No | No | [118] |
| BLESS/DSB-capture | No | Yes | No | [83] |
| GUIDE-Seq | No | Yes | No | [82] |
| qPCR | Yes | No | No | [84,119] |
| Enzymatic methods | Yes | Yes | No | [120] |
| DSB-ddPCR | Yes | Yes | Yes | This work |

**Supplementary Table 2 | Comparison of methods for detecting DSBs.**

| Time (h) | 0 | 0.167 | 0.5 | 2 | 4 | 8 | 24 |
|---|---|---|---|---|---|---|---|
| AAVS1 | 0.15±0.09 | 0.38±0.03 | 0.47±0.02 | 2.59±0.14 | 5.64±0.52 | 10.41±1.52 | 17.04±0.84 |
| MYC sgRNA5 | 0.37±0.07 | 0.47±0.11 | 0.44±0.12 | 1.42±0.04 | 2.31±0.31 | 6.05±0.61 | 12.07±0.75 |
| VEGFA sgRNA2 | 0.0012±0.0006 | 0.0069±0.0031 | 0.009±0.008 | 0.44±0.12 | 0.60±0.05 | 1.43±0.13 | 5.08±1.62 |
| VEGFA sgRNA3 | 1.59±0.11 | 2.05±0.09 | 2.49±0.39 | 6.92±0.47 | 10.52±0.58 | 15.76±0.33 | 26.72±0.58 |

**Supplementary Table 3 | Indel kinetics for multiple sgRNAs.** Indel frequencies (% ± s.e.m., n = 3 biological replicates) plotted in **Fig. 3.2b**.

| Time (h) | 0.167 | 0.5 | 2 | 4 | 8 | 24 |
|---|---|---|---|---|---|---|
| AAVS1 | 0.0348 | 0.0121 | 5.71E-05 | 0.000249 | 0.00126 | 1.81E-05 |
| MYC sgRNA5 | 0.254 | 0.317 | 8.56E-05 | 0.00191 | 0.000373 | 5.091E-05 |
| VEGFA sgRNA2 | 0.0689 | 0.196 | 0.0103 | 0.000148 | 0.00019 | 0.0174 |
| VEGFA sgRNA3 | 0.0154 | 0.0464 | 0.000195 | 5.73E-05 | 1.13E-06 | 9.02E-07 |

**Supplementary Table 4 | *P*-values for comparisons to respective zero hour time point.** *P*-values for comparisons of each time point to the zero hour time point for each sgRNA as plotted in **Fig. 3.2e**. All sgRNAs exhibit statistically significant increase in indels by 2 hours. *P*-values were calculated using a one-sided two-sample Student's t-test (n = 3 biological replicates).

| Variant | BH3 peptide sequence | $K_i$ (nM) |
|---|---|---|
| G22 (ciCas9) | APPNLWAAQRYGRELRRMADE**G**EGSFK | 48 ± 5 |
| A22 | APPNLWAAQRYGRELRRMADE**A**EGSFK | 36 ± 2 |
| V22 | APPNLWAAQRYGRELRRMADE**V**EGSFK | 20 ± 3 |
| L22 | APPNLWAAQRYGRELRRMADE**L**EGSFK | 1.7 ± 0.2 |
| F22 | APPNLWAAQRYGRELRRMADE**F**EGSFK | < 0.5 |

**Supplementary Table 5 | BH3 variant peptide sequences and BCL-xL $K_i$'s.**

**Supplementary Note: Peptide Sequences for constructs and domains**

Cas9:
MDYKDDDDKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED
KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE
GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG
EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF
LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSE
ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY
VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN
ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI
QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
QELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW
RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKY
DENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK
KDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS
QLGGDSRADPKKKRKV

ciCas9:
MDYKDDDDKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED
KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE
GDLNPDNSSNRELVVDFLSYKLSQKGYSWSQFSDVEENRTEAPEGTESEMETPSAIN
GNPSWHLADSPAVNGATGHSSSLDAREVIPMAAVKQALREAGDEFELRYRRAFSDLT
SQLHITPGTAYQSFEQVVNELFRDGVNWGRIVAFFSFGGALCVESVDKEMQVLVSRIA
AWMATYLNDHLEPWIQENGGWDTFVELYGNNGSGTASGTGSGTGSATGSGTVNTEI
TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE
FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF
IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNE
ENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG
SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL

KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER
GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSD
FRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL
FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKVTGS
GTAPPNLWAAQRYGRELRRMADEGEGSFK

BCL-xL (res 2-197):
SNRELVVDFLSYKLSQKGYSWSQFSDVEENRTEAPEGTESEMETPSAINGNPSWHLA
DSPAVNGATGHSSSLDAREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGT
AYQSFEQVVNELFRDGVNWGRIVAFFSFGGALCVESVDKEMQVLVSRIAAWMATYLN
DHLEPWIQENGGWDTFVELYGNN

FLAG:
DYKDDDDK

NLS:
SRADPKKKRKV

enhanced specificity-ciCas9 (e-ciCas9):
MDYKDDDDKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED
KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE
GDLNPDNSSNRELVVDFLSYKLSQKGYSWSQFSDVEENRTEAPEGTESEMETPSAIN
GNPSWHLADSPAVNGATGHSSSLDAREVIPMAAVKQALREAGDEFELRYRRAFSDLT
SQLHITPGTAYQSFEQVVNELFRDGVNWGRIVAFFSFGGALCVESVDKEMQVLVSRIA
AWMATYLNDHLEPWIQENGGWDTFVELYGNNGSGTASGTGSGTGSATGSGTVNTEI
TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE
FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF
IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNE
ENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG
SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
ADDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER
GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSD
FRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPALESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKAPLIETNGETGEIVWDKGRDFATVR
KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS

VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL
FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKVTGS
GTAPPNLWAAQRYGRELRRMADEGEGSFK

**Materials and Methods**

**Modeling Cas9 REC2 domain replacement with BCL-xL**
Volume estimates for the REC2 (residues 180-307) and BCL-xL (residues 4-198) domains were performed using the 3V volume calculator[121]. Computational modeling of REC2 replacement with BCL-xL was performed using RosettaRemodel[24]. Briefly, BCL-xL (PDB ID: 2BZW) was treated as a rigid-body inserted after Cas9 residues 179 and connected to 308 of Cas9 with a flexible linker (PDB ID: 4OO8). A blueprint file defined regions to rebuild in a modeling task. The functionality for fusing one structure into another (which is turned on with the *-insert_segment_from_pdb* flag) uses “0 x l” entries, which denote extensions of the peptide chain from the inserted PDB, in the range that corresponds to the placement of the BCl-xL domain. The flexible linker was composed of repeating glycine-(serine/threonine) residues, and backbone conformations were sampled by introducing random combinations of torsions from loop fragments. To facilitate the conformational search, a random break was made in the flexible linker to be reconnected via both random fragment moves and chain-closure algorithms guided by the Rosetta energy function; trajectories that properly reconnected the chain were considered successful. The lowest energy model from each successful trajectory was saved as a PDB file.

The flags to run the calculations are as follows:
-database [rosetta database location]
-s [template Cas9 PDB file] #4OO8, after deletion of REC2, nucleic acids
-remodel:blueprint [blueprint files]
-insert_segment_from_pdb [PDB formatted BCL-xL domain]
-num_trajectory 10
-save_top 1
-remodel:quick_and_dirty
-use_clusters false
-vall debug1000vall
-overwrite

300 independent trajectories were sampled in 30 parallel runs that used the flags above.

**Cas9, ciCas9, arC9, and sgRNA expression plasmids**
An N-terminal FLAG tag sequence was appended via Gibson Assembly Cloning (New England Biosciences) to a human codon optimized Cas9 (sub-cloned from hCas9, a gift from George Church, Addgene plasmid #41815) with a single C-terminal NLS expressed from a pcDNA3.3-TOPO vector. BCL-xL (residues 4-198, Uniprot Q07817-

1), BH3 peptide (APPNLWAAQRYGRELRRMADEGEGSFK), and linker sequences (Supplementary Fig. B.3) were introduced using gBlocks (Integrated DNA Technologies) and Gibson Assembly Cloning. Enhanced specificity-ciCas9, ciCas9(L22), and ciCas9(F22) were generated via restriction digest and NEBuilder HiFi Gibson Cloning (New England Biosciences). To create ciCas9(L22) and ciCas9(F22), we used higher-affinity BH3 sequences (Supplementary Table 5). The specificity enhancing mutations from eSpCas9(1.1) (gift from Feng Zhang, Addgene plasmid # 71814) were subcloned into ciCas9 to create enhanced-specificity ciCas9 (e-ciCas9). sgRNA target sequences (Supplementary Data)—except for VEGFA sgRNAs—were cloned into the gRNA_Cloning Vector according to the hCRISPR gRNA Synthesis protocol (https://www.addgene.org/static/data/93/40/adf4a4fe-5e77-11e2-9c30-003048dd6500.pdf). gRNA_Cloning Vector and human codon optimized Cas9 were gifts from George Church (Addgene plasmids # 41824 and # 41815). VEGFA site#2 ("VEGFA sgRNA2") and VEGFA Site#3 ("VEGFA sgRNA3) were gifts from Keith Joung (Addgene plasmids # 47506 and # 47507). arC9 (containing NLS) expression plasmid was a gift from David Savage.

**Cell culture**

HEK-293T cells (293T/17, ATCC) were maintained in high glucose DMEM, 10% FBS, 4 mM L-glutamine (Life Technologies). U2OS and HCT116 cells were maintained in McCoy's 5A (Modified) medium, 10% FBS (Life Technologies). Cells were tested and certified free of mycoplasma monthly.

**Editing of genomic loci**

Cells were plated in 12-well plates at the following densities: HEK-293T 3.25 x $10^5$ cells/well; U2OS 7.5 x $10^4$ cells/well; HCT116 3.00 x $10^5$ cells/well. The following day, cells were transfected with Turbofectin 8.0 (Origene), except in Fig. 2e, Supplementary Fig. B.2 and 7, in which cells were transfected with X-tremeGENE HP. All cells were transfected with ◂ µL transfection reagent and 1 µg plasmid DNA per well unless otherwise noted. The plasmid DNA mixture consisted of 450 ng sgRNA plasmid, 450 ng of Cas9 construct plasmid and 100 ng pMAX-GFP as a transfection control. 24 hours after transfection, ciCas9 was activated with the indicated concentrations of A-385358, WEHI-539 (ApexBio Technology), or ABT-737 (Selleck Chemicals). A-385358 was synthesized according to previously reported procedure[122,123]. The identity was confirmed by 1H-NMR and mass spectrometry (Bruker Esquire Ion Trap MS). >95% purity was confirmed by analytical HPLC. arC9 was activated with 1 µM 4-hydroxytamoxifen (Selleck Chemicals). Compounds were stored as 10 mM stocks in DMSO at -20ºC. Final DMSO concentration in drug treated wells was 0.1%.

To harvest HEK-293T and HCT116 cells, wells were washed with 1 mL/well ice cold DPBS. Cells were subsequently harvested via trituration with 600 µL ice cold DPBS, spun down at 1500 x g for 10 min at 4ºC, decanted, and cell pellets stored at -80ºC. To enable temporal precision, plates for time course experiments were placed on ice prior to harvesting. To harvest U2OS cells, wells were washed with 1 mL pre-warmed DPBS, trypsinized (0.25% trypsin–EDTA, Life Technologies), and quenched with 10% FBS supplemented DMEM. Cells were spun down at 1500 x g for 10 min at 4ºC, washed with

ice cold DPBS, spun down at 1500 x g for 10 min at 4ºC, decanted, and cell pellets stored at -80ºC. For DSB-ddPCR experiments (Fig. 3.4b-e, Supplementary Fig. B.9-11) transfections were carried out as above except cells were transfected using 1.5µLTurbofectin 8.0 transfection reagent using 225 ng sgRNA, 225 ng ciCas9, and 50 ng pMAX-GFP plasmids per well in 12-well plates, which was found to result in low background signal in the absence of Cas9 activity (Supplementary Fig. B.9). For cell viability experiments (Supplementary Fig. B.15), cells were transfected using 1.5 µLTurbofectin 8.0 transfection reagent using 225 ng sgRNA, 225 ng ciCas9, and 50 ng pMAX-GFP plasmids per well in 24-well plates (plating density was scaledto 1.625 x $10^5$ cells/well).

**Indel quantification with high-throughput sequencing**

Genomic DNA was isolated using the DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's instructions with the following modification: proteinase K digestion was extended to 1h at 56 ˚C. 20 cycles of primary PCR to amplify the region of interest was performed using 2 µL of DNeasy eluate (~100-300 ng template) in a 5 µL Kapa HiFi HotStart polymerase reaction (Kapa Biosystems, for primers see Supplementary Data). The PCR reaction was diluted with 45 µL DNAse free water (Ambion). Illumina adapters and indexing sequences were added via 25 cycles of secondary PCR with 3 µL of diluted primary PCR product in a 10 µL Taq polymerase reaction (New England Biosciences, for primers see Supplementary Data). The final amplicons were run on a TBE-agarose gel (0.7%), and the product band was excised and extracted using the Freeze and Squeeze Kit according to the manufacturer's instructions (Bio-Rad). Gel purified amplicons were quantified by qPCR using the Illumina Library Quantification Kit (Kapa Biosystems). Then, up to 300 indexed amplicons were pooled and sequenced on a MiSeq (MiSeq 150 V3 kit, Illumina, for primers see Supplementary Data).

After demultiplexing of reads (bcl2fastq, Illumina), indels were quantified with a custom Python script (github link will be made available upon publication). Briefly, 8-mer sequences were identified in the reference sequence located 20 base pairs up- and downstream of the target sequence. Sequence distal to these 8-mers was trimmed. Reads lacking these 8-mers were discarded. The trimmed reads were then evaluated for indels using the Python difflib package. Trimmed reads which differed in length from the trimmed reference and for which an insertion or deletion operation spanning or within 1 bp of the predicted Cas9 cleavage site was present. Indel quantification for the initial Cas9.BCL screen (Supplementary Fig. B.2) was performed using an analysis method similar to that previously described[124]. Briefly, reads were aligned to reference using Bowtie 2[125]. SAMtools[126] was used to generate pileup files, which were used to map and quantify indels.

**DSB-ddPCR**

Two amplicons were designed for digital droplet PCR (ddPCR), one including the sgRNA binding/Cas9 cut site, the other a proximal control uncut site. For the MYC locus, a TaqMan® probe (Thermo Fisher Scientific) was designed for each amplicon using Primer 3 (http://bioinfo.ut.ee/primer3-0.4.0/). Dual-quenched probes were used for the EMX1 and AAVS1 loci (Integrated DNA Technologies). Droplets were created using

Droplet generating oil for probes, DG8 cartridges, DG8 Gaskets and the QX200™ Droplet generator (Bio-Rad), amplification was performed using ddPCR Supermix for Probes (Bio-Rad). The ddPCR Supermix amplification reactions were set up according to the manufacturer's specifications (Bio-Rad). 1µL of DNeasy eluate (~50-100 ng), obtained as described above, was added to a 20 µL amplification reaction with final primer and probe concentrations of 900 nM and 250 nM respectively. The reaction was divided into droplets for amplification following the manufacturer's protocol (Bio-Rad). Droplets were amplified using the following cycling conditions: MYC 95 °C x 10 minutes, 40 cycles (94 °C x 30 seconds, 55 °C x 60 seconds), 98 °C x 10 minutes, all other amplicons 95 °C x 10 minutes, 40 cycles (94 °C x 30 seconds, 60 °C x 60 seconds), 98 °C x 10 minutes. Following thermal cycling, droplets were individually scanned using the QX200™ Droplet Digital™ PCR system (Bio-Rad). Positive and negative droplets in each fluorescent channel (VIC/FAM) were distinguished on the basis of fluorescence amplitude using a global threshold, set by the minimal, intrinsic fluorescence signal resulting from imperfect quenching of the fluorogenic probes (negative droplets) as compared to the strong fluorescence signal from cleaved probes in droplets with amplified template(s). Cleaved control DNA was created by digesting genomic DNA using the restriction enzyme Sfc-I for the MYC locus, DrdI for the EMX1 locus and BSU36I for the AAVS1 locus (New England Biosciences) according to the manufacturer's instructions. The DSB frequency in the control and experimental samples was calculated as:

$$\frac{[target-,template+]}{[target-,template+] + [target+,template+]}$$

To generate standard curves, restriction enzyme digested control DNA was mixed with uncleaved control DNA in defined amounts, DSB frequency was quantified using the ddPCR protocol and a linear regression was performed. For the time course experiments, absolute DSB frequencies were calculated by fitting raw frequencies to standard curves.

**Western blotting**

Cells were harvested as above then lysed in a modified RIPA buffer (50 mM Tris-HCL, pH 7.8, 1% IGEPAL CA-630, 150 mM NaCl, 1mM EDTA, Pierce Protease Inhibitor Tablet). Lysates were cleared by centrifugation at 17,000 x g for 10 minutes at 4ºC. Then, samples were incubated at 98 C for 7 minutes, subjected to SDS-PAGE (MOPS/SDS running buffer, Criterion and Mini-PROTEAN TGX gels, Bio-Rad) and transferred to nitrocellulose. Blocking and antibody incubation were done in TBS with 0.1% Tween-20 (v/v) and blocking buffer (Odyssey). Primary antibodies were diluted: FLAG (1:1,000, Cell Signaling Technology #8146; or 1:1,000 or Sigma M2 #F1804) and GAPDH (1:2,000, Cell Signaling Technology #2118). Blots were washed in TBS with 0.1% Tween-20. Antibody binding was detected by using near-infrared-dye-conjugated secondary antibodies and visualized on the LI-COR Odyssey scanner.

For evaluation of proteinase K digestion efficacy (Supplementary Fig. B.8), cells from a 12-well plate were split into thirds. One third was processed as described above for

DSB-ddPCR samples. One third was processed as described, but was not run on DNeasy column. One third of the cells underwent only the first step in processing—resuspension in PBS and Buffer AL (DNeasy kit)—but not treated with Proteinase K or run on DNeasy column. Protein was isolated from the three samples using methanol chloroform precipitation, resuspended in 30 µL LDS Sample buffer (ThermoFisher Scientific) supplemented with DTT and heated to 98ºC for seven minutes. 10 µL of each sample was run on a Criterion gel (Bio-Rad) and blots were processed and visualized as described above.

**Fluorescence anisotropy BH3 competition assay**

Assays were performed as described previously[12]. The BH3 peptide variants (initial concentration = 10 µM, 3-fold serial dilutions, 10 data points) G22, V22, and A22 (GenScript) were assayed against 38 nM BCL-xL in the presence of 35 nM Bak-BODIPY. The BH3 peptide variants (initial concentration = 10 µM, 3-fold serial dilutions, 10 data points) L22 and F22 (GenScript) were assayed against 12 nM BCL-xL in the presence of 35 nM Bak-BODIPY. Competition assays were incubated at room temperature for 7 hours. The amount of polarization was measured in millipolarization (mP) units with an excitation of 485 nm and emission of 530 nm. $IC_{50}$ values were determined using Graphpad Prism software (non-linear regression analysis). For each BH3 variant titration, three technical replicates were performed. The calculated $IC_{50}$ values were used to determine $K_I$ values (classic model, http://botdb.abcc.ncifcrf.gov/toxin/kiConverter.jsp)[127]. $K_I$ values shown in Supplementary Table 5 are the average of assays performed in triplicate ± s.e.m. (n = 3 technical replicates).

**Statistical Analysis**

Statistical tests of indel and DSB frequency were performed in Excel using a one-sided two-sample Student's t-test. Statistical tests of linear regression slopes were performed in Prism software.

**Bibliography**


1. Humphrey, S. J., Azimifar, S. B. & Mann, M. High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. *Nat. Biotechnol.* (2015). doi:10.1038/nbt.3327
2. Choudhary, C. & Mann, M. Decoding signalling networks by mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell Biol.* **11,** 427–439 (2010).
3. J D'Souza, R. C. *et al.* Time-resolved dissection of early phosphoproteome and ensuing proteome changes in response to TGF. *Sci Signal* **7,** rs5–rs5 (2014).
4. Räschle, M. *et al.* Proteomics reveals dynamic assembly of repair complexes during bypass of DNA cross-links. *Science* **348,** 1253671–1253671 (2015).
5. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537,** 347–355 (2016).
6. Lara-Astiaso, D. *et al.* Chromatin state dynamics during blood formation. *Science* **345,** 943–949 (2014).
7. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31,** 46–53 (2012).
8. Neph, S. *et al.* Circuitry and dynamics of human transcription factor regulatory networks. *Cell* **150,** 1274–1286 (2012).
9. Toettcher, J. E., Weiner, O. D. & Lim, W. A. Using Optogenetics to Interrogate the Dynamic Control of Signal Transmission by the Ras/Erk Module. *Cell* **155,** 1422–1434 (2013).
10. Dagliyan, O. *et al.* Engineering extrinsic disorder to control protein activity in living cells. *Science* **354,** 1441–1444 (2016).
11. Wu, Y. I. *et al.* A genetically encoded photoactivatable Rac controls the motility of living cells. *Nature* **461,** 104–108 (2009).
12. Goreshnik, I. & Maly, D. J. A Small Molecule-Regulated Guanine Nucleotide Exchange Factor. *J. Am. Chem. Soc.* **132,** 938–940 (2010).
13. Buskirk, A. R. & Liu, D. R. Creating Small-Molecule-Dependent Switches to Modulate Biological Functions. *Chem. Biol.* **12,** 151–161 (2005).
14. Inoue, T., Heo, W. D., Grimley, J. S., Wandless, T. J. & Meyer, T. An inducible translocation strategy to rapidly activate and inhibit small GTPase signaling pathways. *Nat. Meth.* **2,** 415–418 (2005).
15. Zhou, X. X., Chung, H. K., Lam, A. J. & Lin, M. Z. Optical control of protein activity by fluorescent protein domains. *Science* **338,** 810–814 (2012).
16. Karginov, A. V., Ding, F., Kota, P., Dokholyan, N. V. & Hahn, K. M. Engineered allosteric activation of kinases in living cells. *Nat. Biotechnol.* **28,** 743–747 (2010).
17. Toettcher, J. E., Gong, D., Lim, W. A. & Weiner, O. D. Light-based feedback for controlling intracellular signaling dynamics. *Nat. Meth.* **8,** 837–839 (2011).
18. Suh, B. C., Inoue, T., Meyer, T. & Hille, B. Rapid Chemically Induced Changes of PtdIns(4,5)P2 Gate KCNQ Ion Channels. *Science* **314,** 1454–1457 (2006).
19. Komatsu, T. *et al.* Organelle-specific, rapid induction of molecular activities and membrane tethering. *Nat. Meth.* **7,** 206–208 (2010).
20. Ambroggio, X. I. & Kuhlman, B. Design of protein conformational switches. *Curr. Opin. Struct. Biol.* **16,** 525–530 (2006).
21. Karnoub, A. E. & Weinberg, R. A. Ras oncogenes: split personalities. *Nat. Rev.*

*Mol. Cell Biol.* **9,** 517–531 (2008).
22. Boriack-Sjodin, P. A., Margarit, S. M., Bar-Sagi, D. & Kuriyan, J. The structural basis of the activation of Ras by Sos. *Nature* **394,** 337–343 (1998).
23. Shoemaker, A. R. *et al.* A small-molecule inhibitor of Bcl-XL potentiates the activity of cytotoxic drugs in vitro and in vivo. *Cancer Res.* **66,** 8731–8739 (2006).
24. Huang, P.-S. *et al.* RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS ONE* **6,** e24109 (2011).
25. Roose, J. P., Mollenauer, M., Ho, M., Kurosaki, T. & Weiss, A. Unusual interplay of two types of Ras activators, RasGRP and SOS, establishes sensitive and robust Ras activation in lymphocytes. *Mol. Cell. Biol.* **27,** 2732–2745 (2007).
26. Santos, S. D. M., Verveer, P. J. & Bastiaens, P. I. H. Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nat. Cell. Biol.* **9,** 324–330 (2007).
27. Olsen, J. V. *et al.* Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* **127,** 635–648 (2006).
28. Tarcic, G. *et al.* EGR1 and the ERK-ERF axis drive mammary cell migration in response to EGF. *FASEB J.* **26,** 1582–1592 (2012).
29. Dougherty, M. K. *et al.* Regulation of Raf-1 by direct feedback phosphorylation. *Mol. Cell* **17,** 215–224 (2005).
30. Dhillon, A. S. *et al.* Cyclic AMP-dependent kinase regulates Raf-1 kinase mainly by phosphorylation of serine 259. *Mol. Cell. Biol.* **22,** 3237–3246 (2002).
31. Zimmermann, S. & Moelling, K. Phosphorylation and regulation of Raf by Akt (protein kinase B). *Science* **286,** 1741–1744 (1999).
32. Marquette, A., André, J., Bagot, M., Bensussan, A. & Dumaz, N. ERK and PDE4 cooperate to induce RAF isoform switching in melanoma. *Nat Struct Mol Biol* **18,** 584–591 (2011).
33. McPhillips, F. *et al.* Raf-1 is the predominant Raf isoform that mediates growth factor-stimulated growth in ovarian cancer cells. *Carcinogenesis* **27,** 729–739 (2006).
34. Wixler, V., Smola, U., Schuler, M. & Rapp, U. Differential regulation of Raf isozymes by growth versus differentiation inducing factors in PC12 pheochromocytoma cells. *FEBS Letters* **385,** 131–137 (1996).
35. Lavoie, H. & Therrien, M. Regulation of RAF protein kinases in ERK signalling. *Nat. Rev. Mol. Cell Biol.* **16,** 281–298 (2015).
36. Kubicek, M. *et al.* Dephosphorylation of Ser-259 regulates Raf-1 membrane association. *J. Biol. Chem.* **277,** 7913–7919 (2002).
37. Mason, C. S. *et al.* Serine and tyrosine phosphorylations cooperate in Raf-1, but not B-Raf activation. *EMBO J.* **18,** 2137–2148 (1999).
38. Pandit, B. *et al.* Gain-of-function RAF1 mutations cause Noonan and LEOPARD syndromes with hypertrophic cardiomyopathy. *Nat. Genet.* **39,** 1007–1012 (2007).
39. Kolch, W. Coordinating ERK/MAPK signalling through scaffolds and inhibitors. *Nat. Rev. Mol. Cell Biol.* **6,** 827–837 (2005).
40. Bollag, G. *et al.* Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **467,** 596–599 (2010).

41. Sosman, J. A. *et al.* Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. *N. Engl. J. Med.* **366,** 707–714 (2012).
42. Oberholzer, P. A. *et al.* RAS mutations are associated with the development of cutaneous squamous cell tumors in patients treated with RAF inhibitors. *J. Clin. Oncol.* **30,** 316–321 (2012).
43. Poulikakos, P. I., Zhang, C., Bollag, G., Shokat, K. M. & Rosen, N. RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. *Nature* **464,** 427–430 (2010).
44. Hatzivassiliou, G. *et al.* RAF inhibitors prime wild-type RAF to activate the MAPK pathway and enhance growth. *Nature* **464,** 431–435 (2010).
45. Kosako, H. *et al.* Phosphoproteomics reveals new ERK MAP kinase targets and links ERK to nucleoporin-mediated nuclear transport. *Nat. Struct Biol.* **16,** 1026–1035 (2009).
46. Vomastek, T. *et al.* Extracellular signal-regulated kinase 2 (ERK2) phosphorylation sites and docking domain on the nuclear pore complex protein Tpr cooperatively regulate ERK2-Tpr interaction. *Mol. Cell. Biol.* **28,** 6954–6966 (2008).
47. Marklund, U., Brattsand, G., Shingler, V. & Gullberg, M. Serine 25 of oncoprotein 18 is a major cytosolic target for the mitogen-activated protein kinase. *J. Biol. Chem.* **268,** 15039–15047 (1993).
48. Courcelles, M. *et al.* Phosphoproteome dynamics reveal novel ERK1/2 MAP kinase substrates with broad spectrum of functions. *Mol. Syst. Biol.* **9,** 669–669 (2013).
49. Dueber, J. E., Yeh, B. J., Chak, K. & Lim, W. A. Reprogramming Control of an Allosteric Signaling Switch Through Modular Recombination. *Science* **301,** 1904–1908 (2003).
50. Dueber, J. E., Mirsky, E. A. & Lim, W. A. Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat. Biotechnol.* **25,** 660–662 (2007).
51. Yeh, B. J., Rutigliano, R. J., Deb, A., Bar-Sagi, D. & Lim, W. A. Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors. *Nature* **447,** 596–600 (2007).
52. Bos, J. L., Rehmann, H. & Wittinghofer, A. GEFs and GAPs: Critical Elements in the Control of Small G Proteins. *Cell* **129,** 865–877 (2007).
53. Prior, I. A., Lewis, P. D. & Mattos, C. A Comprehensive Survey of Ras Mutations in Cancer. *Cancer Res.* **72,** 2457–2467 (2012).
54. Ory, S., Zhou, M., Conrads, T. P., Veenstra, T. D. & Morrison, D. K. Protein Phosphatase 2A Positively Regulates Ras Signaling by Dephosphorylating KSR1 and Raf-1 on Critical 14-3-3 Binding Sites. *Current Biology* **13,** 1356–1364 (2003).
55. Zhang, B. H. *et al.* Serum- and glucocorticoid-inducible kinase SGK phosphorylates and negatively regulates B-Raf. *J. Biol. Chem.* **276,** 31620–31626 (2001).
56. Haling, J. R. *et al.* Structure of the BRAF-MEK complex reveals a kinase activity independent role for BRAF in MAPK signaling. *Cancer Cell* **26,** 402–413 (2014).
57. McKay, M. M., Ritt, D. A. & Morrison, D. K. RAF Inhibitor-Induced KSR1/B-RAF

Binding and Its Effects on ERK Cascade Signaling. *Current Biology* **21,** 563–568 (2011).
58. Heidorn, S. J. *et al.* Kinase-dead BRAF and oncogenic RAS cooperate to drive tumor progression through CRAF. *Cell* **140,** 209–221 (2010).
59. Callahan, M. K. *et al.* Paradoxical Activation of T Cells via Augmented ERK Signaling Mediated by a RAF Inhibitor. *Cancer Immunol. Res.* **2,** 70–79 (2014).
60. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34,** 339–344 (2016).
61. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3,** e04766 (2014).
62. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351,** 84–88 (2016).
63. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* (2016). doi:10.1038/nature16526
64. Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540,** 144–149 (2016).
65. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507,** 62–67 (2014).
66. Knight, S. C. *et al.* Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* **350,** 823–826 (2015).
67. Liu, K. I. *et al.* A chemical-inducible CRISPR-Cas9 system for rapid control of genome editing. *Nat. Chem. Biol.* **12,** 980–987 (2016).
68. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat. Biotechnol.* **33,** 755–760 (2015).
69. Rose, J. C. *et al.* A computationally engineered RAS rheostat reveals RAS-ERK signaling dynamics. *Nat. Chem. Biol.* **13,** 119–126 (2017).
70. Nishimasu, H. *et al.* Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156,** 935–949 (2014).
71. Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* **11,** 316–318 (2015).
72. Liang, X. *et al.* Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J. Biotechnol.* **208,** 44–53 (2015).
73. Liao, S., Tammaro, M. & Yan, H. Enriching CRISPR-Cas9 targeted cells by co-targeting the HPRT gene. *Nucleic Acids Research* **43,** e134–e134 (2015).
74. Moreno-Mateos, M. A. *et al.* CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Meth.* **12,** 982–988 (2015).
75. Isaac, R. S. *et al.* Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *Elife* **5,** e13450 (2016).
76. Chen, X. *et al.* Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Research* **44,** 6482–6492 (2016).
77. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize

off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34,** 184–191 (2016).
78. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33,** 139–142 (2015).
79. Ivashkevich, A., Redon, C. E., Nakamura, A. J., Martin, R. F. & Martin, O. A. Use of the γ-H2AX assay to monitor DNA damage and repair in translational cancer research. *Cancer Lett.* **327,** 123–133 (2012).
80. Olive, P. L. & Banáth, J. P. The comet assay: a method to measure DNA damage in individual cells. *Nat Protoc* **1,** 23–29 (2006).
81. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* (2015). doi:10.1038/nature14299
82. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33,** 187–197 (2015).
83. Lensing, S. V. *et al.* DSBCapture: in situ capture and sequencing of DNA breaks. *Nat. Meth.* **13,** 855–857 (2016).
84. Liang, Z., Sunder, S., Nallasivam, S. & Wilson, T. E. Overhang polarity of chromosomal double-strand breaks impacts kinetics and fidelity of yeast non-homologous end joining. *Nucleic Acids Research* **44,** 2769–2781 (2016).
85. Hindson, C. M. *et al.* Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat. Meth.* **10,** 1003–1005 (2013).
86. Anders, C. & Jinek, M. In vitro enzymology of Cas9. *Meth. Enzymol.* **546,** 1–20 (2014).
87. Hart, T. *et al.* High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163,** 1515–1526 (2015).
88. Groves, J. T. & Kuriyan, J. Molecular mechanisms in signal transduction at the membrane. *Nat Struct Mol Biol* **17,** 659–665 (2010).
89. Freedman, T. S. *et al.* A Ras-induced conformational switch in the Ras activator Son of sevenless. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 16692–16697 (2006).
90. Freedman, T. S. *et al.* Differences in flexibility underlie functional differences in the Ras activators son of sevenless and Ras guanine nucleotide releasing factor 1. *Structure* **17,** 41–53 (2009).
91. Findlay, G. M. G. *et al.* Interaction domains of sos1/grb2 are finely tuned for cooperative control of embryonic stem cell fate. *Cell* **152,** 1008–1020 (2013).
92. Margarit, S. M. *et al.* Structural evidence for feedback activation by Ras.GTP of the Ras-specific nucleotide exchange factor SOS. *Cell* **112,** 685–695 (2003).
93. Boykevisch, S. *et al.* Regulation of Ras Signaling Dynamics by Sos-Mediated Positive Feedback. *Current Biology* **16,** 2173–2179 (2006).
94. Sondermann, H., Nagar, B., Bar-Sagi, D. & Kuriyan, J. Computational docking and solution x-ray scattering predict a membrane-interacting role for the histone domain of the Ras activator son of sevenless. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 16632–16637 (2005).
95. Christensen, S. M. *et al.* One-way membrane trafficking of SOS in receptor-triggered Ras activation. *Nat Struct Mol Biol* **23,** 838–846 (2016).
96. Iwig, J. S. *et al.* Structural analysis of autoinhibition in the Ras-specific exchange factor RasGRP1. *Elife* **2,** e00813 (2013).
97. Yadav, K. K. & Bar-Sagi, D. Allosteric gating of Son of sevenless activity by the

histone domain. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 3436–3440 (2010).
98. Findlay, G. M. & Pawson, T. How is SOS activated? Let us count the ways. *Nat Struct Mol Biol* **15,** 538–540 (2008).
99. Gureasko, J. *et al.* Membrane-dependent signal integration by the Ras activator Son of sevenless. *Nat Struct Mol Biol* **15,** 452–461 (2008).
100. Hall, B. E., Yang, S. S., Boriack-Sjodin, P. A., Kuriyan, J. & Bar-Sagi, D. Structure-based Mutagenesis Reveals Distinct Functions for Ras Switch 1 and Switch 2 in Sos-catalyzed Guanine Nucleotide Exchange. *J. Biol. Chem.* **276,** 27629–27637 (2001).
101. Wu, C.-Y., Roybal, K. T., Puchner, E. M., Onuffer, J. & Lim, W. A. Remote control of therapeutic T cells through a small molecule-gated chimeric receptor. *Science* **350,** aab4077 (2015).
102. Morsut, L. *et al.* Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. *Cell* (2016).
103. Roybal, K. T. *et al.* Engineering T Cells with Customized Therapeutic Response Programs Using Synthetic Notch Receptors. *Cell* **167,** 419–432.e16 (2016).
104. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* **43,** D447–52 (2015).
105. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Research* **41,** W77–83 (2013).
106. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* **57,** 289–300 (1995).
107. Campeau, E. *et al.* A Versatile Viral System for Expression and Depletion of Proteins in Mammalian Cells. *PLoS ONE* **4,** e6529 (2009).
108. Debnath, J., Muthuswamy, S. K. & Brugge, J. S. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* **30,** 256–268 (2003).
109. Tiscornia, G., Singer, O. & Verma, I. M. Production and purification of lentiviral vectors. *Nat Protoc* **1,** 241–245 (2006).
110. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: An open-source MS/MS sequence database search tool. *Proteomics* **13,** 22–24 (2013).
111. Deutsch, E. W. *et al.* A guided tour of the Trans-Proteomic Pipeline. *Proteomics* **10,** 1150–1159 (2010).
112. Pedrioli, P. G. A. *et al.* Automated identification of SUMOylation sites using mass spectrometry and SUMmOn pattern recognition software. *Nat. Meth.* **3,** 533–539 (2006).
113. Oakes, B. L. *et al.* Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat. Biotechnol.* **34,** 646–651 (2016).
114. Nguyen, D. P. *et al.* Ligand-binding domains of nuclear receptors facilitate tight control of split CRISPR activity. *Nat. Commun.* **7,** 12009 (2016).
115. Feng, J. *et al.* A general strategy to construct small molecule biosensors in eukaryotes. *Elife* **4,** 4004 (2015).
116. Senturk, S. *et al.* Rapid and tunable method to temporally control gene editing based on conditional Cas9 stabilization. *Nat. Commun.* **8,** 14370 (2017).
117. Maji, B. *et al.* Multidimensional chemical control of CRISPR-Cas9. *Nat. Chem.*

*Biol.* **13,** 9–11 (2017).
118. Hanada, K. *et al.* The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks. *Nat Struct Mol Biol* **14,** 1096–1104 (2007).
119. Furda, A. M., Bess, A. S., Meyer, J. N. & Van Houten, B. Analysis of DNA damage and repair in nuclear and mitochondrial DNA of animal cells using quantitative PCR. *Methods Mol. Biol.* **920,** 111–132 (2012).
120. Grégoire, M.-C. *et al.* Quantification and genome-wide mapping of DNA double-strand breaks. *DNA Repair (Amst.)* **48,** 63–68 (2016).
121. Voss, N. R. & Gerstein, M. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Research* **38,** W555–W562 (2010).
122. Wendt, M. D. *et al.* Discovery and structure-activity relationship of antagonists of B-cell lymphoma 2 family proteins with chemopotentiation activity in vitro and in vivo. *J. Med. Chem.* **49,** 1165–1181 (2006).
123. Bruncko, M. *et al.* Studies leading to potent, dual inhibitors of Bcl-2 and Bcl-xL. *J. Med. Chem.* **50,** 641–662 (2007).
124. Kiani, S. *et al.* Cas9 gRNA engineering for genome editing, activation and repression. *Nat. Meth.* **12,** 1051–1054 (2015).
125. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Meth.* **9,** 357–359 (2012).
126. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
127. Cer, R. Z., Mudunuri, U., Stephens, R. & Lebeda, F. J. IC50-to-Ki: a web-based tool for converting IC50 to Ki values for inhibitors of enzyme activity and ligand binding. *Nucleic Acids Research* **37,** W441–5 (2009).