Interrogation and Control of Mammalian Transcription
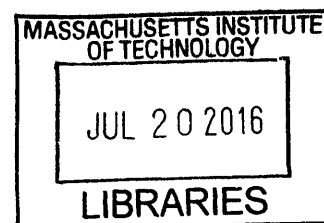
by

Silvana Konermann

B.S., Swiss Federal Institute of Technology (2009)

Submitted to the Department of Brain and Cognitive Sciences
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN NEUROSCIENCE
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2016

Signature of Author: .... **Signature redacted**

Department of Brain and Cognitive Sciences
December 30, 2015

Certified by: .... **Signature redacted**

Fêng Zhang
Professor of Brain and Cognitive Science
Thesis Supervisor

Accepted by: .. **Signature redacted**

Matthew Wilson
Sherman Fairchild Professor of Neuroscience
Director of Graduate Education for Brain and Cognitive Sciences Program

Interrogation and Control of Mammalian Transcription

By

Silvana Konermann

Submitted to the Department of Brain and Cognitive Sciences
on December 30, 2015 in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY IN NEUROSCIENCE

Abstract

Gene expression is dynamic in living systems, enabling environmental adaptation and homeostasis. Transcript levels may change temporarily during distinct phases of biological processes, while longer lasting modifications to their regulatory machinery can lead to specific cell states or disease phenotypes. However, versatile and robust methods to investigate causal relationships between gene expression states and biological phenotypes remain elusive. My thesis work – divided into two main parts – has focused on the development of technologies to enable efficient, generalizable, and precise perturbation of mammalian gene expression.

The first part of my research focused on the development of light-inducible transcriptional effectors (LITEs) to mediate positive and negative regulation of endogenous mammalian gene expression (Konermann et al. Nature 2013). Optical stimulation enables precise spatiotemporal control to closely match endogenous transcriptional dynamics. I engineered the LITE system based on the programmable TALE DNA binding proteins from plant pathogens in combination with the light-inducible dimer cryptochrome 2 – cib1 from *Arabidopsis thaliana*. Light enables fast and reversible recruitment of transcriptional effector domains to the TALE bound to the endogenous target promoter through dimerization of cryptochrome 2 – cib1.

I applied LITEs to control gene expression in primary neurons as well as in the mouse brain *in vivo,* demonstrating their potential to dissect genetic contributions to dynamic behaviors such as learning. Epigenetic regulation of transcriptional state is an additional layer of endogenous control exerted by the cell to store more permanent states such as memories. To interrogate epigenetic in addition to transcriptional dynamics, I next developed TALE-mediated targeting of 32 repressive histone effectors to alter epigenetic states in a locus-specific manner. The LITE system establishes a novel mode of optogenetic control of endogenous cellular processes and enables direct testing of the causal roles of genetic and epigenetic regulation in normal biological processes and disease states.

One major limiting aspect of TALE-based transcriptional activators is the costly and labor-intensive construction of their repetitive DNA binding domains. As a result, the utility of TALEs for higher-throughput gene targeting experiments remains limited. The

CRISPR nuclease Cas9, however, can be easily programmed using a short guide RNA homologous to the target genomic DNA of interest. Additionally, Cas9 can be easily converted into an RNA-guided DNA binding protein (dCas9) via inactivation of its two catalytic domains.

The ease and scalability of the CRISPR-Cas9 system potentially enables systematic, genome-scale perturbation, but the magnitude of transcriptional up-regulation achieved by the current generation of Cas9 transcriptional activators typically ranges from low to ineffective. In order to achieve a system where the majority of Cas9 activators are highly functional, I undertook structure-guided engineering to generate a potent, synergistic Cas9 activation complex (SAM) capable of mediating robust up-regulation with a single sgRNA (Konermann et al., Nature, 2015) which outperforms current systems by more than two orders of magnitude. I demonstrated that these transcriptional effectors are capable of activating up to 10 genes simultaneously, allowing for understanding of complex genetic and regulatory networks.

Genome-scale GOF screening approaches have largely remained limited to the use of cDNA library systems, which are costly and challenging to use in a pooled format. To overcome this limitation, I designed a genome-scale sgRNA library targeting every coding isoform from the RefSeq database (23,430 isoforms) for a final library of 70,290 guides. I next aimed to identify gain-of-function changes that can lead to the development of BRAF inhibitor resistance in $BRAF^{V600}$ mutant melanoma cells. The screen results highlighted a number of gene candidates that both confirm known BRAF inhibitor-resistance pathways and suggest novel mechanisms of action. SAM activators present a highly reliable and generalizable tool for genome-wide interrogation of gene function and interaction in diverse biological processes.

Recently, we have extended the utility of the SAM system to enable bimodal control through the use of modified, truncated deadRNAs (dRNAs) (Dahlman, et al., Nature Biotechnology 2015). These dRNAs prevent nucleolytic activity of an active Cas9 nuclease and transform the wildtype enzyme into an efficient transcriptional activator when combined with the SAM activatorcomponents. This system enables simultaneous knock-out of gene A and activation of gene B in the same cell population, enabling bidirectional interrogation of gene interaction and regulatory networks.

Thesis Supervisor: Feng Zhang,
Title: Professor, Department of Brain and Cognitive Sciences

Acknowledgements

I have many people to thank for this thesis. Fundamentally, all aspects of this research were accomplished through teamwork. When I first started my rotation in Feng Zhang's lab, just two weeks after he had come to MIT and the Broad Institute I was not only stricken by his quick mind and fascinating yet clear ideas but also by the energetic, open and collaborative community he had built so quickly in his lab. I am deeply grateful for the opportunity to learn from him, and to be a part in seeing his lab grow and prosper.

I am grateful to all members of Feng Zhang's lab. Specifically I would like to thank Matthias Heidenreich for helping me with LITEs, Le Cong for introducing me to TALEs and being a great labmate from the start, Randall Platt for kick-starting Cas9-activators, Clea Barcena for spending a great summer screening with me, Naomi Habib for being a great colleague and helping me with RNAseq analysis, Ophir Shalem for fascinating discussions, great humor and screening advice and James Dahlman for working on deadguides.

Much of my work in this thesis was done together with Mark Brigham and I am so thankful to have had him first as my rotation student and then as my teammate and friend. Alex Trevino joined us right out of Harvard undergrad and together we created a great team (and SAM). Patrick Hsu has been an important element in most of my work and I am so grateful for his help and his friendship. Julia Joung,

4

Omar Abbuddayyeh, and Jonathan Gootenberg joined the lab in my last two years and I have enjoyed working with them and mentoring them tremendously.

I would also like to thank Dr. Hiroshi Nishimasu and Dr. Osamu Nureki from the University of Tokyo for the opportunity to collaborate on work on the structure of the Cas9 complex which was the foundation for the development of the SAM activator complex.

Finally I would like to thank Dr. Steve Hyman, Dr. Aviv Regev, Dr. Ann Graybiel and Dr. Phil Sharp for their mentorship and for making time for meeting me in their busy schedules.

This Thesis is dedicated to my mother.

Contents

List of Figures

Chapter 1:

Introduction

The dynamic nature of gene expression enables cellular programming, homeostasis, and environmental adaptation in living systems. Internal and external cellular inputs are functionally integrated and processed by precise regulation within transcriptional networks. To understand the causal role of transcriptional states of individual genes within these networks, tools for precise perturbation of endogenous transcription are required.

TAL Effectors: reprogrammable DNA binding proteins

One key component of any tool that can be programmed to alter endogenous transcription is a reprogrammable DNA binding protein than can be directed to its intended target in the genome. TAL effector proteins are used by the plant pathogen Xanthomonas sp. to control gene expression of target genes in their host to aid bacterial proliferation[1]. Their amino acid sequence can be divided into a highly repetitive central portion flanked by an N- and C-terminal domain. The central part consists of repeats 34aa in length and recently it has been discovered that each of these repeats binds a single DNA base and that the affinity for a given DNA nucleotide is determined by amino acids in position 12 and 13 of each repeat (called RVDs) [2].The code governing this recognition has been unlocked, opening up the possibility of creating engineered TALEs with known DNA binding affinity[2,1]. In naturally occurring TALEs, the number of

repeats ranges from 1.5 to 33.5 and it has been shown that engineered TALEs of variable length (12 repeats to 20 repeats) can efficiently bind target DNA[1,3]. Engineered TAL effectors (eTALEs) have been combined with functional domains fused to their C-terminus through a linker such as a nuclear localization signal sequence (NLS). Transcriptional activator domains such as VP16, VP64 or P65 have been shown to be capable of upregulating a targeted gene in mammalian cells when fused to an eTALE binding within the promoter region of the gene[3-5]. Engineered TAL effectors are therefore highly promising tools for targeted modification of gene expression. The modular architecture of their DNA recognition specificity gives them an advantage over engineered zinc-finger proteins, where contextual effects (neighboring bases/repeats) seem to play an important role and screening of a larger number of ZFs is usually necessary to achieve the desired DNA targeting[6,7].

CRISPR-Cas9: RNA guided DNA targeting

TAL Effectors were an important advance in the field of genome engineering and enabled targeted perturbations of transcription of individual genes. However, as each new effector requires construction of a new protein to target a specific DNA sequence, the utility of TAL Effectors for targeting of multiple genes simultaneously or performing high throughput screening is limited. A tool that would enable multiplexed perturbation of transcription would greatly benefit the

interrogation of gene-gene interactions within regulatory networks, while genome-wide screening with transcriptional effectors would enable the de-novo identification of key genes within a specific biological process or disease phenotype.

With the recent development of the RNA-guided bacterial nuclease CRISPR-Cas9 for use in mammalian cells[8,9], reprogrammable DNA targeting on a larger scale has been greatly simplified. The Cas9 protein is guided to its DNA target in the mammalian genome by its guide RNA. This guide RNA (sgRNA) consists of two elements: 20 variable nucleotides that basepair directly with the DNA targets and a constant tracr sequence whose secondary structure enables binding between sgRNA and Cas9 protein. This Cas9 nuclease complex can be easily converted to a "dead" DNA binding module by mutating its two catalytic domains through point mutations[10,11].

Chapter 2:

Optical control of endogenous transcription and epigenetic states using TAL effectors

Preface
This chapter is adapted from:

Konermann S*, Brigham MD*, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F. Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 500, 472-476 (2013).

to fit the format of this thesis.

Dissecting the contributions of genes to cellular and organismal function requires an approach to enable spatially and temporally controlled modulation of gene expression.

Light provides a fast, reversible, and spatially targeted modality for manipulating a variety of biological processes. Microbial- and plant-derived light-sensitive proteins have been engineered as optogenetic actuators, allowing precise optical control of cellular functions including membrane potential[12-14], intracellular biochemical signaling[15], protein interactions[16-19], and heterologous gene

expression[18-23]. However, the ability to directly modulate endogenous gene

expression using light has remained elusive.

An ideal optogenetic approach for controlling endogenous gene transcription

would be easily generalizable to target any gene locus, would not require

manipulation of the endogenous genomic sequence or the addition of exogenous

co-factors, and would exhibit fast and reversible kinetics. The DNA-binding

domain of transcription activator-like effectors (TALEs)[1,2] from *Xanthamonas sp.*

can be easily customized to bind a variety of DNA sequences in mammalian

cells[3,5,24]. TALE DNA-binding domains are modular and can be fused with a

variety of effector domains, including nucleases, transcriptional activators, and

transcriptional repressors to edit or modulate endogenous mammalian genomic

loci[3,5,24,25].

We sought to use light-inducible heterodimer proteins to mediate the recruitment

of transcriptional effector domains to a TALE targeted to an endogenous locus.

While several plant-based light-sensitive proteins have been developed for

mammalian applications, many suffer from slow or irreversible kinetics and some

depend on the supplementation of exogenous co-factors that are not present in

mammalian cells. The *Arabidopsis thaliana* cryptochrome 2 (CRY2) was

previously shown to employ flavin, an abundant biomolecule in mammalian cells,

as its light-sensing chromophore. Upon photoexcitation with blue light (~450nm)

the flavin chromophore in CRY2 is reduced and triggers a conformational change that allows dimerization with its interacting protein partner CIB1. The dimerization between CRY2 and CIB1 occurs within seconds and is reversible within a few minutes after withdrawal of light illumination.

In this chapter we describe the development of Light-Inducible Transcriptional Effectors (LITEs), a two-component system integrating the customizable TALE DNA-binding domain with the light-sensitive cryptochrome 2 protein and its interacting partner CIB1 from *Arabidopsis thaliana*. LITEs can be engineered to mediate positive and negative regulation of endogenous mammalian gene expression in a reversible manner, and changes in transcript levels occur within minutes after stimulation. Like other optogenetic tools, LITEs can be packaged into viral vectors and genetically targeted to specific cell types to probe gene function within specific cell populations. We demonstrate the application of this system in primary cells as well as in the mouse brain *in vivo*.

Engineering light-inducible transcriptional effectors using programmable

TALE DNA targeting proteins

The LITE system employs a modular design consisting of two independent

components (**Fig. 1a**): The first component is the genomic anchor and includes a

customizable DNA-binding domain, based on transcription activator-like effectors

(TALEs)[1,2] from *Xanthamonas sp.*, fused to the light-sensitive cryptochrome 2

(CRY2) protein from *Arabidopsis thaliana*[19,26] (TALE-CRY2). The second

component includes the CRY2 interacting partner CIB1[19,26] fused to a desired

effector domain (CIB1-effector). In the absence of light (inactive state), TALE-

CRY2 binds the promoter region of the target gene while CIB1-effector remains

free within the nuclear compartment. Illumination with blue light (peak ~450 nm)

triggers a conformational change in CRY2 and subsequently recruits CIB1-

effector (VP64 shown in **Fig. 1a**) to the target locus to mediate transcriptional

modulation. This modular design allows each LITE component to be

independently engineered, allowing the same genomic anchor to be combined

with activating or repressing effectors[25,27] to exert positive and negative

transcriptional control over the same endogenous genomic locus.

To identify the most effective architecture, we assessed the efficacy of different

LITE designs by measuring blue light illumination induced transcriptional changes

of the neural lineage-specifying transcription factor neurogenin 2 (*Neurog2*) (**Fig.**

1b). 3 out of 4 initial LITE pairings produced significant light-induced *Neurog2*

mRNA up-regulation in Neuro 2a cells (p < 0.001, **Fig. 1b**).
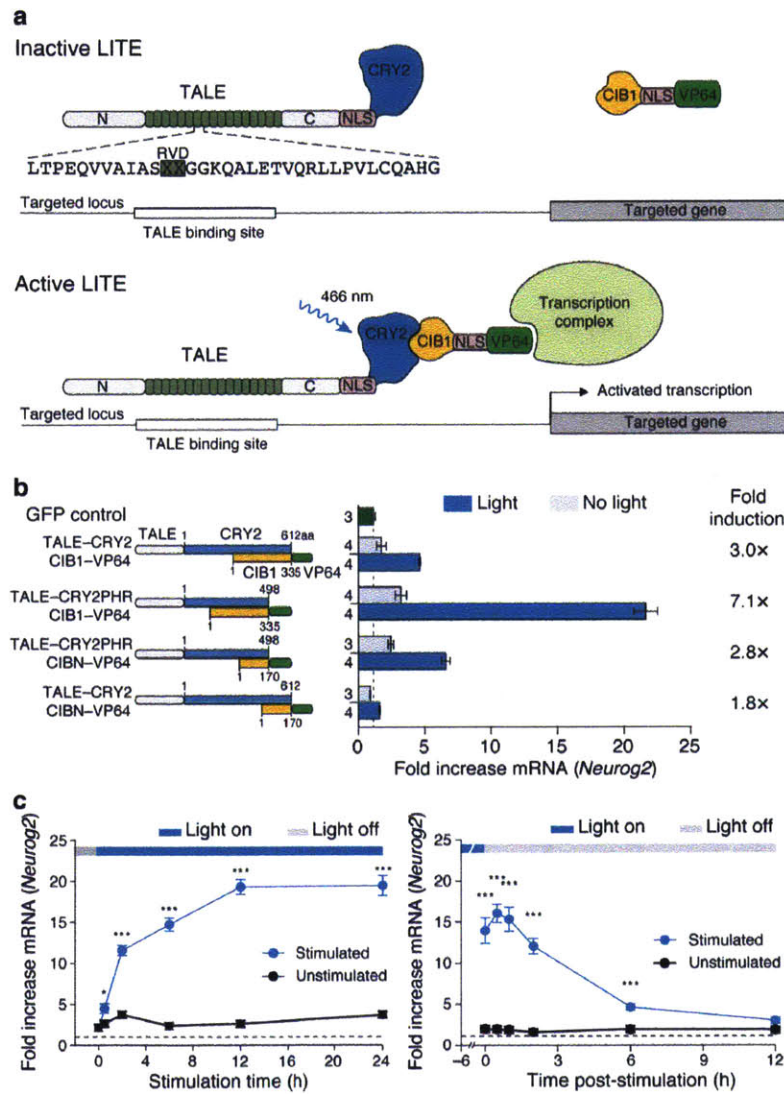
Figure 1 | Design and optimization of the LITE system. a, Schematic of the LITE system. Light stimulation induces dimerization of CRY2 and CIB1, recruiting the effector to the target promoter. **b**, LITE architecture was optimized by fusing TALE and the transcriptional activator VP64[3,27] to different truncations of CRY2 and CIB1[19] (n next to each bar). **c**, Time-course of light-dependent *Neurog2* upregulation and decay post-illumination (n = 4 biological replicates; *p<0.05; ***p<0.001). Cells were stimulated with $5\,mW/cm^2$ light (466 nm, 1 s pulses at 0.067 Hz). Mean±s.e.m. in all panels.

Of these combinations tested, a truncated form of CRY2 missing the photolyase domain (CRY2PHR, amino acids 1–498) combined with the full length CIB (TALE(*Neurog2*)-CRY2PHR::CIB1-VP64) yielded the strongest light-mediated transcription activation as well as the highest induction ratio (light/no light mRNA levels). Therefore TALE-CRY2PHR::CIB1-VP64 was used in subsequent experiments. To ensure optimal function, we also systematically tuned light stimulation parameters and effector domains (wavelength, **Fig. 2a**; duty cycle, **Fig. 2b**; light intensity[28], **Fig. 2c** and **d**; and choice of activation domain, **Fig. 2e**).
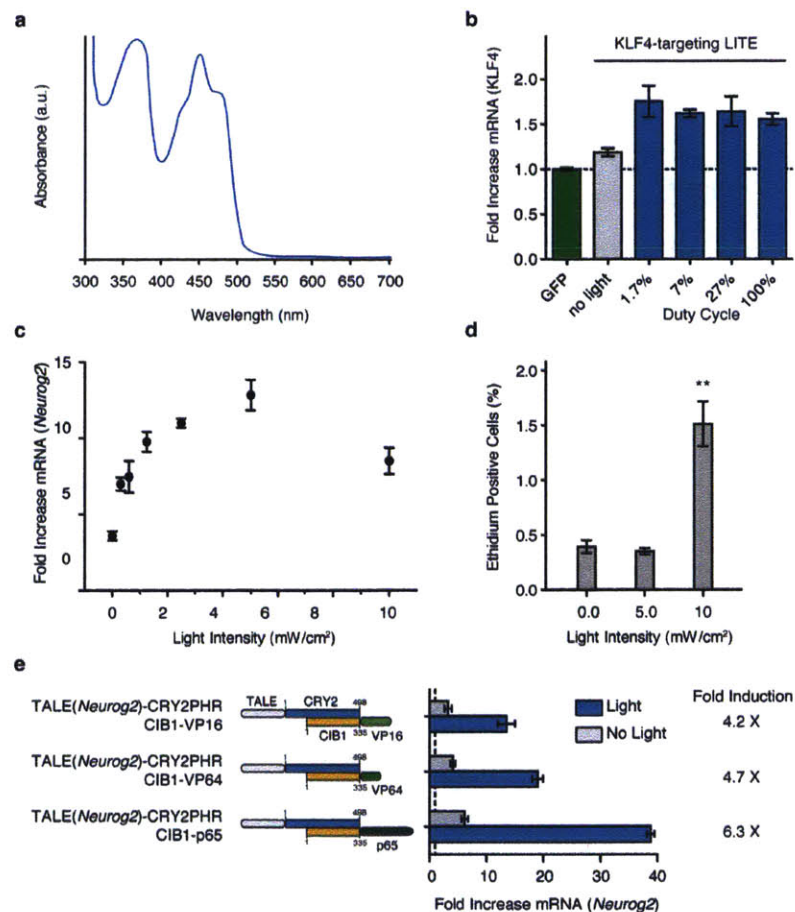
Figure 2 | Engineering of light stimulation parameters and activation domains of LITEs. a, Illustration of the absorption spectrum of CRY2 in vitro. Cryptochrome 2 was optimally activated by 350–475 nm light[28]. A sharp drop in absorption and activation was

seen for wavelengths greater than 480 nm. Spectrum was adapted from Banerjee *et al*[28]. **b**, Impact of illumination duty cycle on LITE-mediated gene expression. Varying duty cycles (illumination as percentage of total time) were used to stimulate 293FT cells expressing LITEs targeting the KLF4 gene. KLF4 expression levels were compared to cells expressing GFP only. Stimulation parameters were: 466 nm, 5 mW/cm^2 for 24 h. Pulses were performed at 0.067 Hz with the following durations: 1.7% = 0.25 s pulse, 7% = 1 s pulse, 27% = 4 s pulse, 100% = constant illumination. (mean ± s.e.m.; n = 3–4 biological replicates.) **c**, The transcriptional activity of CRY2PHR/ CIB1 LITE was found to vary according to the intensity of 466 nm blue light. Neuro 2a cells were stimulated for 12 h at a 7% duty cycle (1 s pulses at 0.067 Hz). All *Neurog2* mRNA levels were measured relative to cells expressing GFP only (mean ± s.e.m.; n = 3–4 biological replicates). **d**, Light-induced toxicity measured as the percentage of cells positive for red fluorescent ethidium homodimer-1 versus calcein-positive cells (mean ± s.e.m.; n = 3 biological replicates; **p<0.01). **e**, We compared the activation domains VP16 and p65 in addition to VP64 to test the modularity of the LITE CIB1–effector component. Neuro 2a cells transfected with LITE were stimulated for 24 h with 466 nm light at an intensity of 5mW/cm^2 and a duty cycle of 7% (1 s pulses at 0.067 Hz). All three domains produced a significant light-dependent *Neurog2* mRNA upregulation (p<0.001). We selected VP64 for subsequent experiments due to its lower basal activity in the absence of light-stimulation (mean ± s.e.m.; n = 3–4 biological replicates).

Although the interaction between CRY2 and CIB1 occurs on a sub-second timescale[19], LITE-mediated transcriptional activation is likely dependent on many factors, including rate of transcription, mRNA processing, and transcript stability[29,30]. We found that LITE-mediated *Neurog2* expression increased considerably as early as 30 min after initial stimulation and rose steadily until saturating at 12 h with approximately 20-fold up-regulation compared to GFP-transfected negative controls (**Fig. 1c**). Interestingly, *Neurog2* transcript levels continued to increase for up to 30 min post-illumination, an effect that may have

resulted from residual CRY2PHR-CIB1 dimerization or from previously recruited RNA polymerases. Thereafter, *Neurog2* mRNA returned to baseline levels with a half-life of ~3 h. In contrast, a small-molecule inducible TALE system based on the plant hormone abcisic acid receptor[31] exhibited slower transcriptional kinetics (**Fig. 3**).
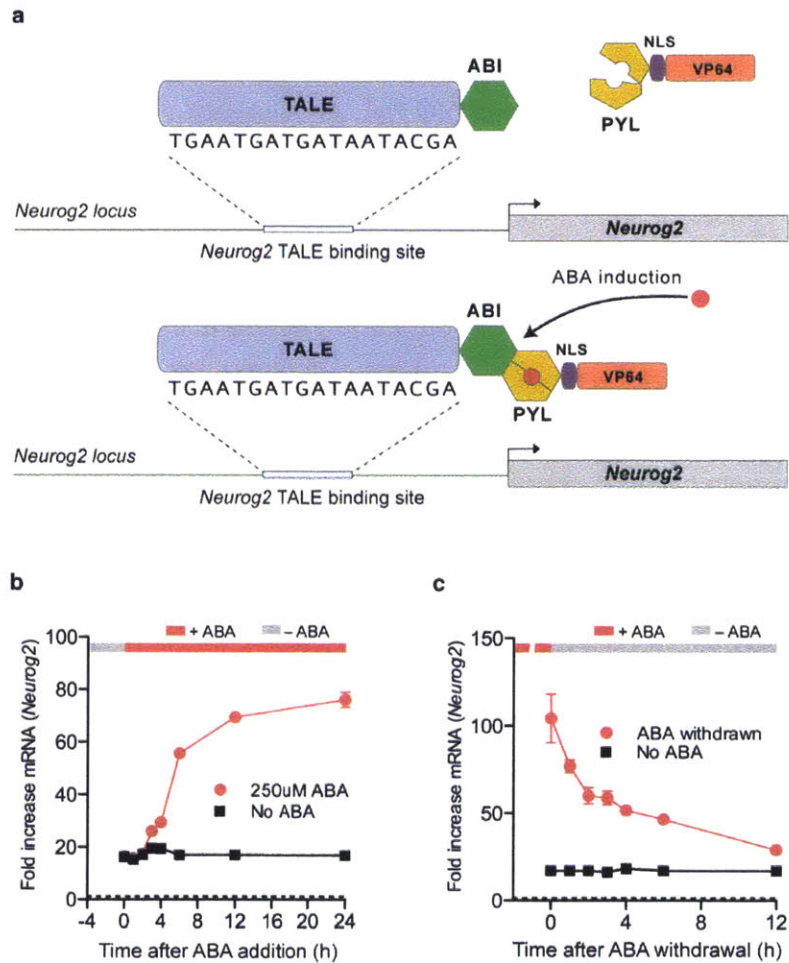
Figure 3 | Chemical induction of endogenous gene transcription. a, Schematic showing the design of a chemically inducible two-hybrid TALE system based on the abscisic acid (ABA) receptor system. ABI and PYL dimerize upon the addition of ABA and dissociate when ABA is withdrawn. **b,** Time-course of ABA-dependent *Neurog2* upregulation. 250 µM of ABA was added to Neuro 2a cells expressing TALE(*Neurog2*)– ABI and PYL–VP64. Fold mRNA increase was measured at the indicated time points

after the addition of ABA. **c**, Decrease of *Neurog2* mRNA levels after 24 h of ABA stimulation. All *Neurog2* mRNA levels were measured relative to GFP-expressing control cells (mean $\pm$ s.e.m.; n = 3–4 biological replicates).

modulating endogenous transcription in neurons

To apply LITE for neuronal applications, we developed a simplified process for adeno-associated virus (AAV) production (**Fig. 4**) and an AAV-based vector (**Fig. 5a** and **b**) for the delivery of TALE genes. The ssDNA-based genome of AAV is less susceptible to recombination, providing an advantage over lentiviral vectors[32]. We evaluated a panel of 28 TALE activators targeting the mouse genome in primary neurons and found that most were able to up-regulate transcription in primary neurons (**Fig. 5c**). Moreover, *in vivo* expression of TALE(*Grm2*)-VP64 in the prefrontal cortex (PFC) (**Fig. 5d** and **e**) induced a 2.5-fold increase in *Grm2* mRNA levels compared to GFP-only controls (**Fig. 5f**).
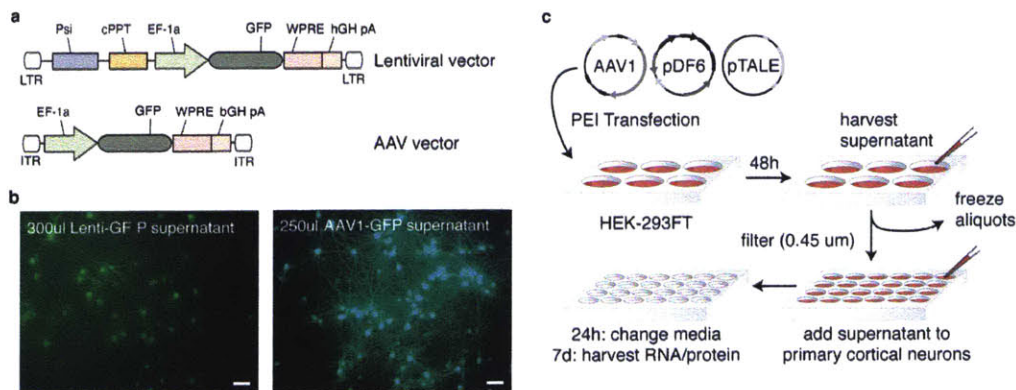
Figure 4 | Efficient AAV production using cell supernatant. a, Lentiviral and AAV vectors carrying GFP were used to test transduction efficiency. **b**, Primary cortical neurons were transduced with supernatant derived from the same number of lentivirus- or AAV-transduced 293FT cells. Representative images of GFP expression were collected at 7 days post infection. Scale bars, 50 µm. **c**, The depicted process was developed for the production of AAV supernatant and subsequent transduction of primary neurons. 293FT cells were transfected with an AAV vector carrying the gene of interest, the AAV1 serotype packaging vector (pAAV1), and helper plasmid (pDF6) using PEI. 48 h later, the supernatant was collected and filtered through a 0.45 µm PVDF membrane. Primary neurons were then transduced with supernatant and remaining

aliquots were stored at −80 °C. Stable levels of AAV construct expression were reached after 5–6 days. AAV supernatant production following this process can be used for production of up to 96 different viral constructs in 96-well format.
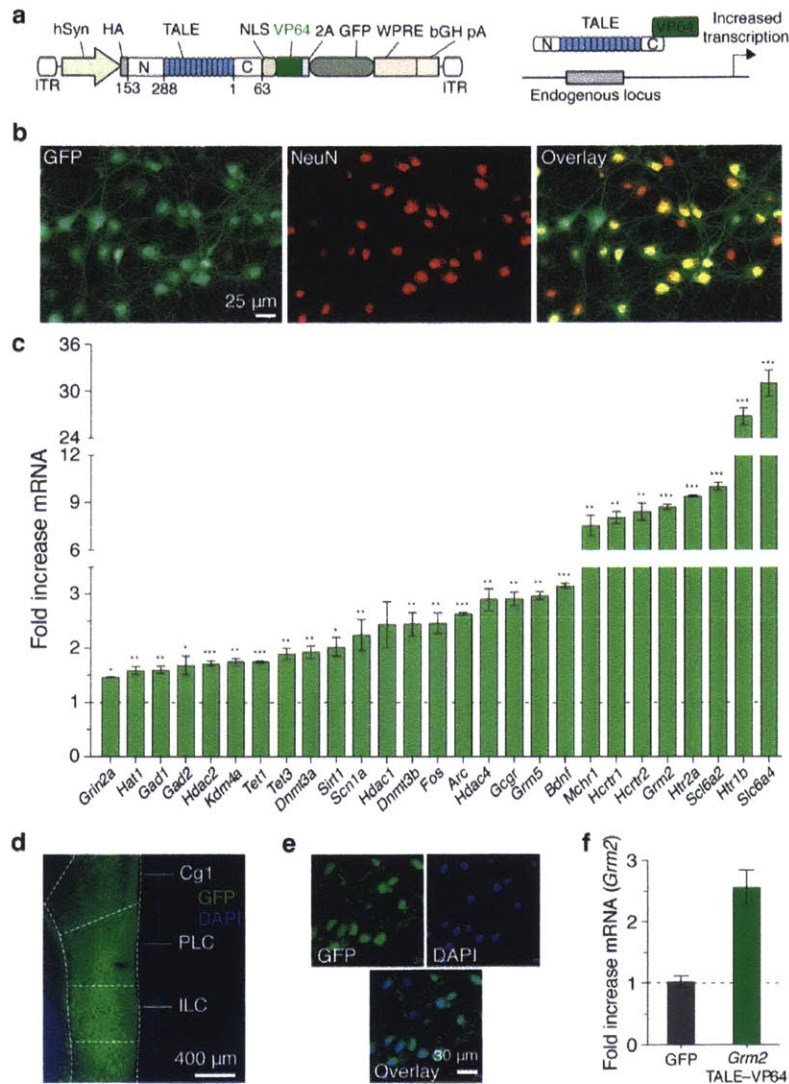
Figure 5 | *In vitro* and *in vivo* AAV-mediated TALE delivery targeting endogenous loci in neurons. a, Schematic of AAV vectors for TALE delivery. **b,** Representative images of primary cortical neurons expressing TALE–VP64. **c,** TALE–VP64 constructs targeting a variety of endogenous genes were screened for transcriptional activation in primary cortical neurons (*p<0.05; **p<0.01; ***p<0.001; $n = 3$ biological replicates). **d,** TALE–VP64 expression in PFC. DAPI, 4′,6-diamidino-2-phenylindole; Cg1, cingulate cortex area 1; PLC, prelimbic cortex; ILC, infralimbic cortex. **e,** Higher magnification

image of TALE–VP64 expressing neurons in PFC. **f**, *Grm2* mRNA upregulation by TALE–VP64 *in vivo* in PFC ($n = 4$ animals). Mean±s.e.m. in all panels.

Similarly, we introduced LITEs into primary cortical neurons via co-delivery of two AAVs (**Fig. 6a** and **b**). We tested a *Grm2*-targeted LITE at 2 light pulsing frequencies with a reduced duty cycle of 0.8% to ensure neuron health (**Fig. 7**). Both stimulation conditions achieved a ~7-fold light-dependent increase in *Grm2* mRNA levels (**Fig. 6c**). Further study verified that substantial target gene expression increases could be attained quickly (4-fold up-regulation of mRNA within 4 h; **Fig. 6d**). In addition, we observed significant up-regulation of mGluR2 protein after stimulation, confirming that LITE-mediated transcriptional changes are translated to the protein level ($p < 0.01$ vs GFP control, $p < 0.05$ vs no-light condition; **Fig. 6e**).
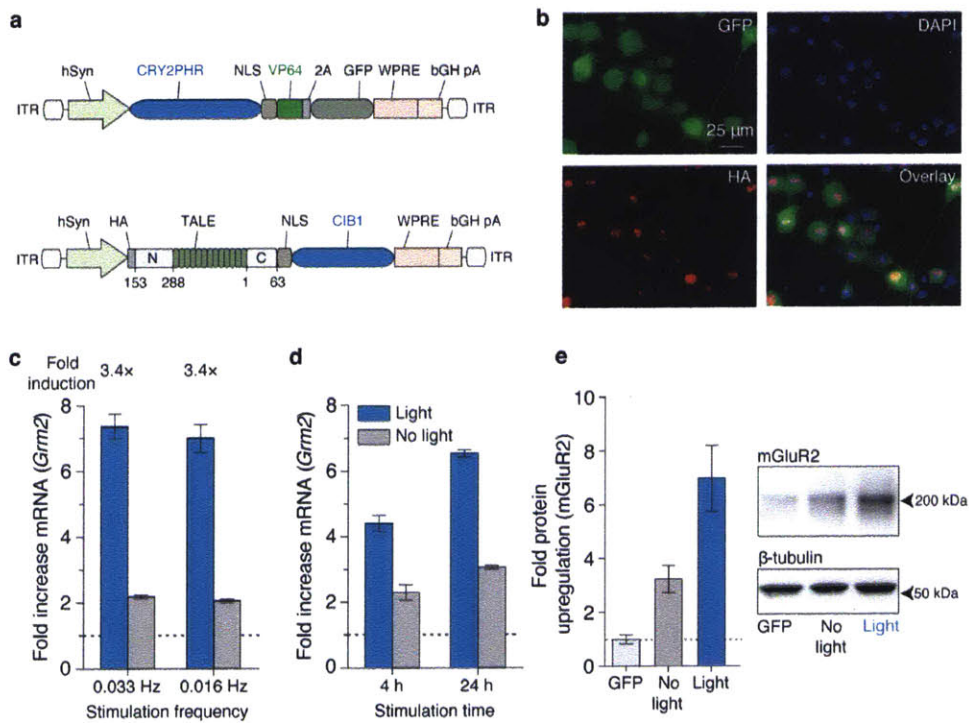
Figure 6 I LITE-mediated optogenetic modulation of endogenous transcription in primary neurons. a, Schematic of AAV LITE constructs. **b**, Images of primary neurons expressing LITE constructs. HA, haemagglutinin tag. **c**, Light-induced activation of *Grm2* in primary neurons after 24 h of stimulation (250 ms pulses at 0.033 Hz or 500 ms pulses at 0.016 Hz; 5 mW/cm^2 ; n = 4 biological replicates). **d**, Upregulation of *Grm2* in primary cortical neurons after 4 h or 24 h of stimulation. Expression levels are shown relative to neurons transduced with GFP only (n=3-4 biological replicates). **e**, Light-mediated changes in mGluR2 protein levels (n = 7 biological replicates).
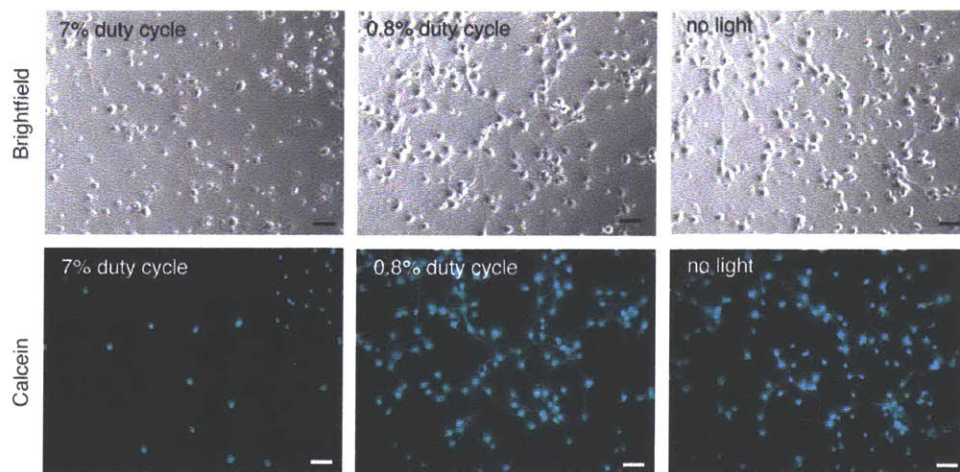
Figure 7 I Impact of light duty cycle on primary neuron health. The effect of light stimulation on primary cortical neuron health was compared for duty cycles of 7%, 0.8%, and no light conditions. Calcein was used to evaluate neuron viability. Brightfield images show cell morphology and integrity. Primary cortical neurons were stimulated with the indicated duty cycle for 24 h with 5 mW/cm^2 of 466 nm light. Representative images, scale bar, 50 μm. Pulses were performed in the following manner: 7% duty cycle = 1 s pulse at 0.067 Hz, 0.8% duty cycle = 0.5 s pulse at 0.0167 Hz.

To test the *in vivo* functionality of the LITE system, we stereotactically delivered a 1:1 mixture of high titer AAV vectors carrying TALE(*Grm2*)-CIB1 and CRY2PHR-VP64 into the PFC. We used a previously established fiber optic cannula system[33] to deliver light to LITE-expressing neurons *in vivo* (**Fig. 8**). After 12 h of

stimulation, we observed a significant increase in *Grm2* mRNA compared with unstimulated PFC (**Fig. 8d**, $p \leq 0.01$). Taken together, these results confirm that LITEs enable optical control of endogenous gene expression in cultured neurons and *in vivo*.
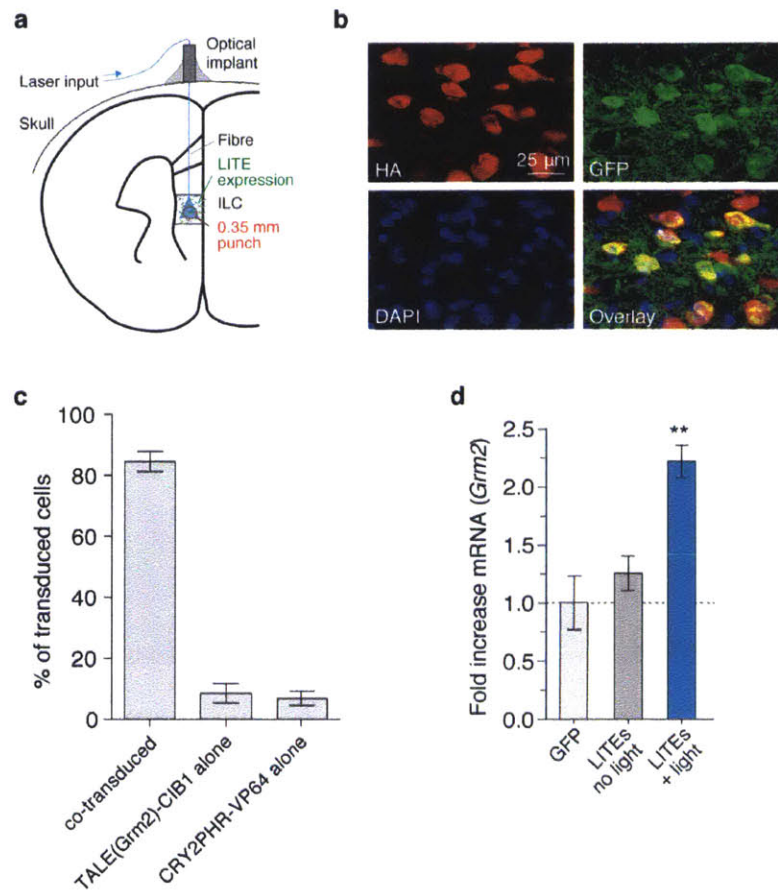
Figure 8 | LITE-mediated optogenetic modulation of endogenous transcription in vivo. a, Schematic of *in vivo* optogenetic stimulation setup. **b**, Representative images of PFC neurons expressing both LITE components. **c**, Co-transduction efficiency of LITE components by AAV1/2 *in vivo* in mouse infralimbic cortex. Cells transduced by TALE(*Grm2*)–CIB1 alone, CRY2PHR–VP64 alone, or co-transduced were calculated as a percentage of all transduced cells (mean ± s.e.m.; $n = 9$ fields from 3 animals). **d**, Light-induced activation of endogenous *Grm2* expression using LITEs transduced into ILC. (**p < 0.01; n = 4–6 animals.)

Improving the Baseline: light-dependent nuclear import and other strategies

Having observed persistent baseline up-regulation *in vivo,* we undertook further rounds of optimization to reduce background activity and improve the gene induction ratio of LITEs. We observed that TALE(*Grm2*)-CIB1 alone produced similar levels of up-regulation as background activation, yet CRY2PHR-VP64 alone did not significantly affect transcription (**Fig. 9**). Therefore we rationalized that LITE-dependent background transcriptional activation arises mainly from TALE-CIB1.
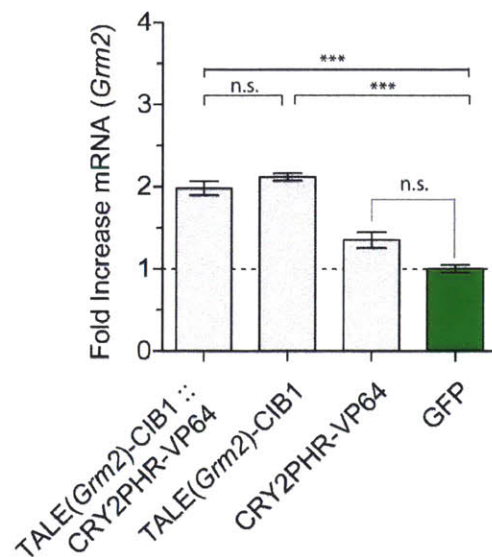
Figure 9 I Basal activation by LITE components in primary neurons. *Grm2* mRNA levels were determined in primary neurons transfected with individual LITE components. Primary neurons expressing TALE(*Grm2*)–CIB1 alone led to a similar increase in *Grm2* mRNA levels as unstimulated cells expressing the complete LITE system (mean±s.e.m.; *n* = 3–4 biological replicates).

The subsequent comprehensive screen to reduce baseline TALE-CIB1-mediated

up-regulation focuses on two strategies: First, although CIB1 is a plant

transcription factor, it may have intrinsic activity in mammalian cells[34]. To address

this, we deleted three CIB1 regions conserved amongst basic helix-loop-helix

transcription factors of higher plants (**Fig. 10**). Second, to prevent TALE-CIB1

from binding the target locus in absence of light, we engineered TALE-CIB1 to

localize in the cytoplasm pending light-induced dimerization with the NLS-

containing CRY2PHR-VP64 (**Fig. 11**). To test both strategies independently or in

combination, we evaluated 73 distinct LITE architectures and identified 12

effector/targeting-domain pairs (denoted by the "+" column in **Fig. 10**) with both

improved light-induction efficiency and reduced background (fold mRNA increase

in the no-light condition compared with the original LITE; $p < 0.05$).
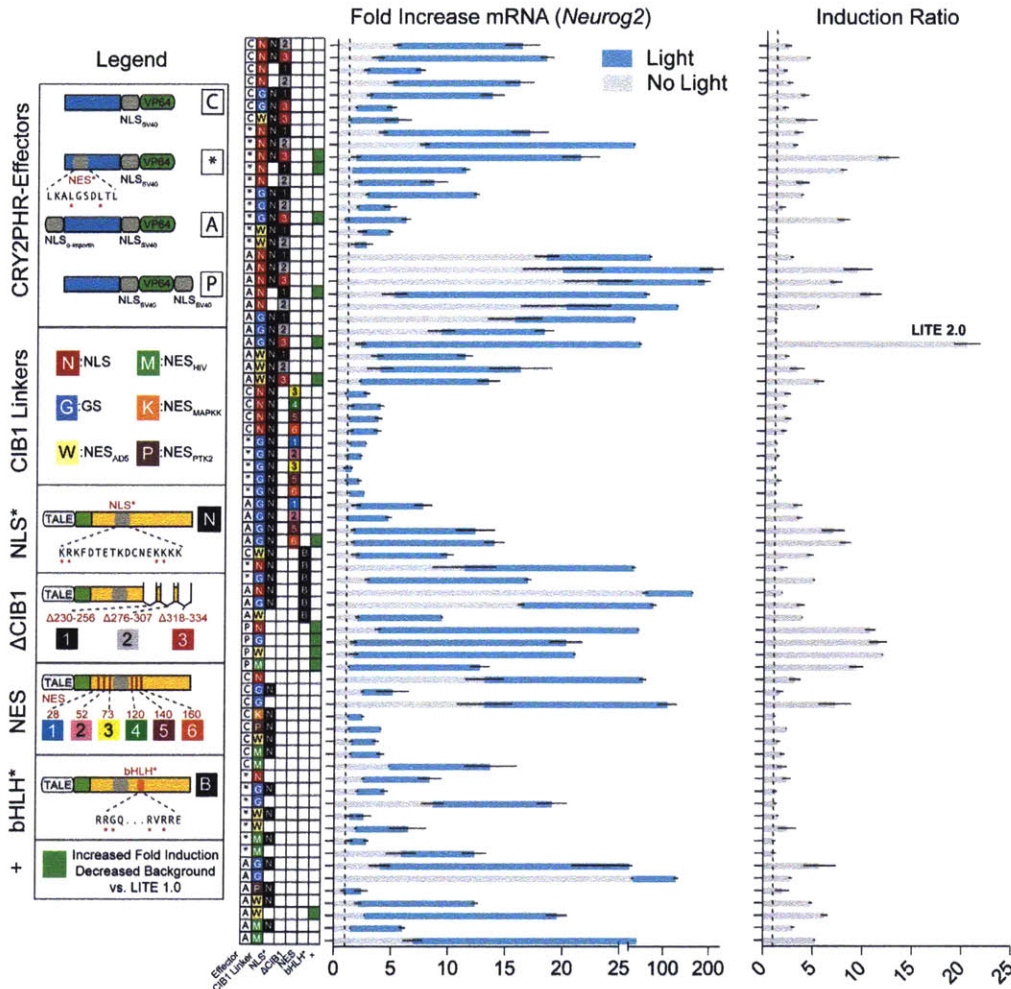
Figure 10 | Effects of LITE component engineering on activation, background signal and fold induction. Protein modifications were used to find LITE components resulting in reduced background transcriptional activation while improving induction ratio by light. In brief, nuclear localization signals and mutations in an endogenous nuclear export signal were used to improve nuclear import of the CRY2PHR–VP64 component. Several variations of CIB1 intended to either reduce nuclear localization or CIB1 transcriptional activation were pursued to reduce the contribution of the TALE–CIB1 component to background activity. The results of all tested combinations of CRY2PHR–VP64 and TALE–CIB1 are shown. The table to the left of the bar graphs indicates the particular combination of domains/mutations used for each condition. Each row of the table and bar graphs contains the component details, light/no light activity, and induction ratio by light for the particular CRY2PHR/CIB1 combination. Combinations that resulted in both decreased background and increased fold induction compared to LITE1.0 are

highlighted in green in the table column marked '+' (*t*-test p<0.05). (mean±s.e.m.; *n* = 2–3 biological replicates).
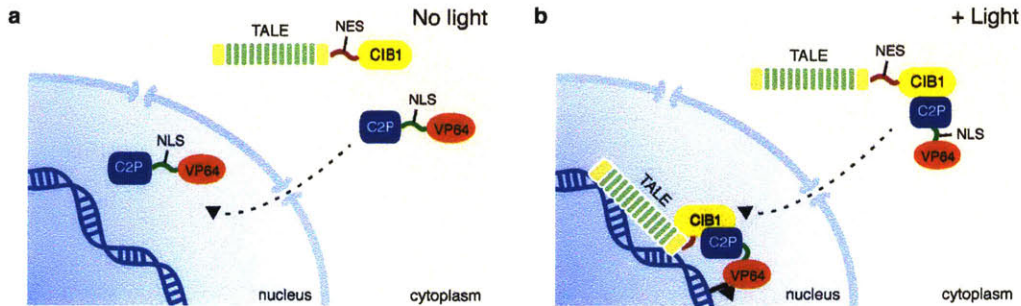
Figure 11 | Inducible nuclear import. a, In the absence of light, the TALE–CIB1 LITE component resides in the cytoplasm due to the absence of a nuclear localization signal, NLS (or the addition of a nuclear export signal, NES). The CRY2PHR–VP64 component containing an NLS on the other hand is actively imported into the nucleus on its own. **b,** In the presence of blue light, TALE–CIB1 binds to CRY2PHR. The NLS present in CRY2PHR–VP64 then mediates nuclear import of the complex comprising both LITE components, enabling them to activate transcription at the targeted locus.

One architecture successfully incorporating both strategies, designated LITE2.0, demonstrated the strongest light induction (light/no-light = 20.4) and resulted in greater than 6-fold reduction of background activation compared with the original design (**Fig. 12a**). Another architecture, LITE1.9.1, produced minimal background activation (1.06) while maintaining four-fold light induction (**Fig. 12b**).
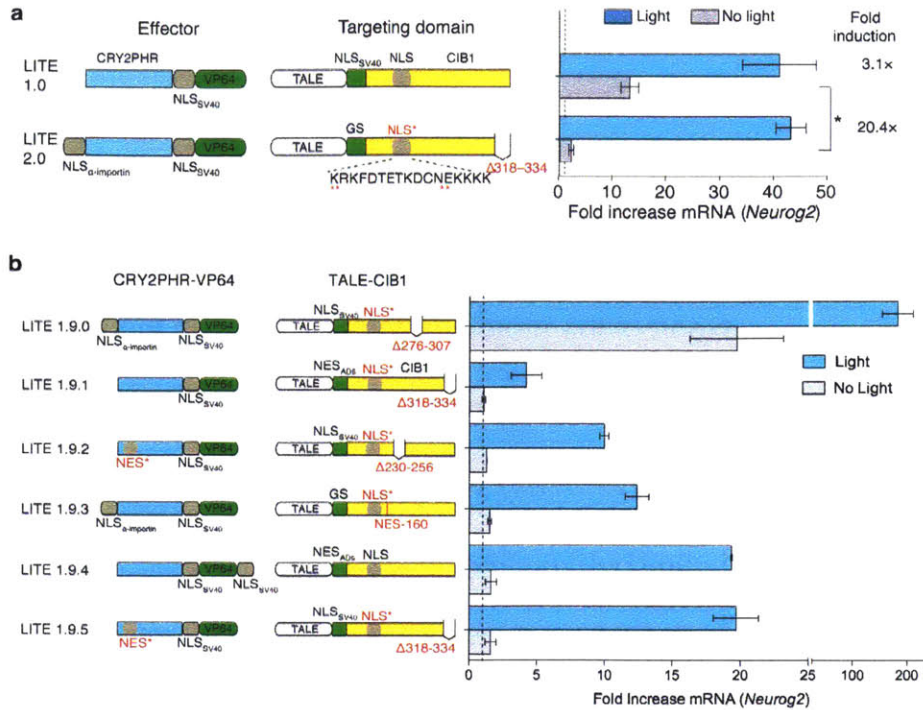
Figure 12 | LITE2.0 and selected LITE1.9 optimizations. a, LITE2.0 significantly reduces the level of background activation in Neuro 2a cells ($n = 3$ biological replicates). Mean ± s.e.m. in all panels. **b**, Several CRY2PHR–VP64/TALE–CIB1 combinations from the engineered LITE component screen were of particular note. LITE1.9.0, which combined the α-importin NLS effector construct with a mutated endogenous NLS and Δ276–307 TALE–CIB1 construct, exhibited an induction ratio greater than 9 and an absolute light activation of more than 180. LITE1.9.1, which combined the unmodified CRY2PHR–VP64 with a mutated NLS, Δ318–334, AD5 NES TALE–CIB1 construct, achieved an induction ratio of 4 with a background activation of 1.06. A selection of other LITE1.9 combinations with background activations lower than 2 and induction ratios ranging from 7 to 12 were also highlighted (mean ± s.e.m.; $n = 2$–3 biological replicates).

Modification strategies for background elimination

CRY2PHR-VP64 Constructs: Three new constructs were designed with the goal of improving CRY2PHR-VP64 nuclear import. First, the mutations L70A and L74A within a predicted endogenous nuclear export sequence of CRY2PHR were created to limit nuclear export of the protein (referred to as '*' in the Effector column of **Fig. 10**). Second, the α-importin nuclear localization sequence was fused to the N-terminus of CRY2PHR-VP64 (referred to as 'A' in the Effector column of **Fig. 10**). Third, the SV40 nuclear localization sequence was fused to the C-terminus of CRY2PHR-VP64 (referred to as 'P' in the Effector column of **Fig. 10**).

TALE-CIB1 Linkers: The SV40 NLS linker between TALE and CIB1 used in LITE 1.0 was replaced with one of several linkers designed to increase nuclear export of the TALE-CIB1 protein (The symbols used in the CIB1 Linker column of **Fig. 10** are shown in parentheses): a flexible glycine-serine linker (G), an adenovirus type 5 E1B nuclear export sequence (W), an HIV nuclear export sequence (M), a MAPKK nuclear export sequence (K), and a PTK2 nuclear export sequence (P).

NLS* Endogenous CIB1 Nuclear Localization Sequence Mutation: A nuclear localization signal exists within the wild type CIB1 sequence. This signal was mutated in NLS* constructs at K92A, R93A, K105A, and K106A in order to

diminish TALE-CIB1 nuclear localization (referred to as 'N' in the NLS* column of **Fig. 10**).

ΔCIB1 Transcription Factor Homology Deletions: In an effort to eliminate possible basal CIB1 transcriptional activation, deletion constructs were designed in which regions of high homology to basic helix-loop-helix transcription factors in higher plants were removed. These deleted regions consisted of Δaa230-256, Δaa276-307, Δaa308-334 (referred to as '1' '2' and '3' in the ΔCIB1 column of **Fig. 10**). In each case, the deleted region was replaced with a 3 residue GGS link.

NES Insertions into CIB1: One strategy to facilitate light-dependent nuclear import of TALE-CIB1 was to insert an NES in CIB1 at its dimerization interface with CRY2PHR such that the signal would be concealed upon binding with CRY2PHR. To this end, an NES was inserted at different positions within the known CRY2 interaction domain CIBN (aa 1-170). The positions are as follows (The symbols used in the NES column of **Fig. 10** are shown in parentheses): aa28 (1), aa52 (2), aa73 (3), aa120 (4), aa140 (5), aa160 (6).

***bHLH basic Helix-Loop-Helix Mutation:** To reduce direct CIB1-DNA interactions, several basic residues of the basic helix-loop-helix region in CIB1 were mutated. The following mutations are present in all *bHLH constructs

(referred to as 'B' in the *bHLH column of **Figure 10**): R175A, G176A, R187A,

and R189A

Repressing gene expression through targeted epigenetic modifications

We sought to further expand the range of transcriptional processes addressable by TALE and LITE. We developed a TALE-repressor by fusing TALEs to a 4X concatenation of the transcriptional repressor mSin3 interaction domain (SID4X). We observed that, in a manner analogous to TALE-activator constructs, TALE-SID4X could repress endogenous transcription when targeted to the promoter of a gene of interest (**Fig. 13**).
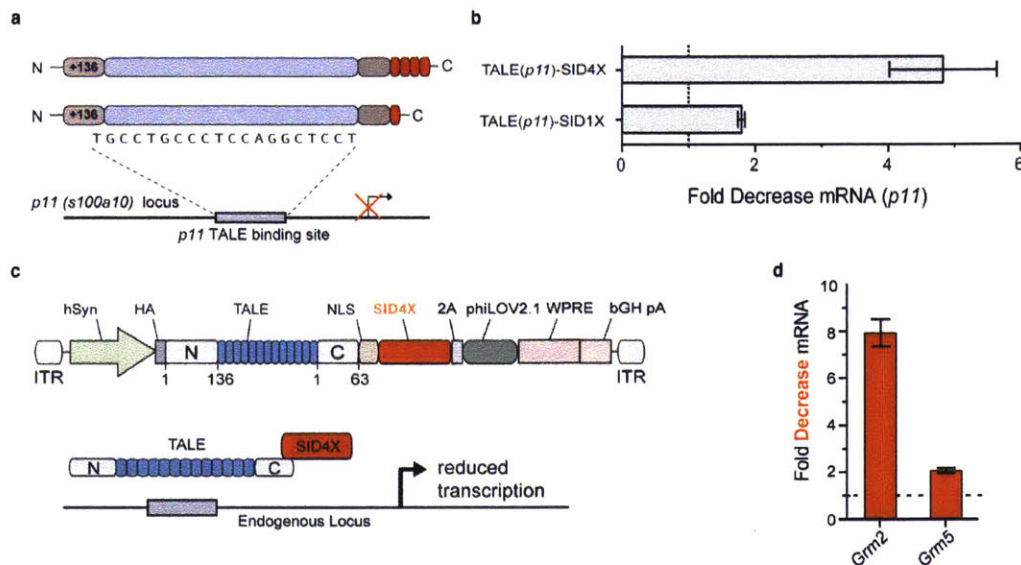
Figure 13 I TALE SID4X repressor characterization and application in neurons. a, A synthetic repressor was constructed by concatenating 4 SID domains (SID4X). To identify the optimal TALE-repressor architecture, SID or SID4X was fused to a TALE designed to target the mouse *p11* (also known as *S100a10*) gene. **b,** Fold decrease in *p11* mRNA was assayed using qRT–PCR (mean±s.e.m.; *n* = 3 biological replicates). **c,** General schematic of constitutive TALE transcriptional repressor packaged into AAV. Effector domain SID4X is highlighted. hSyn, human synapsin promoter; 2A, *Thosea asigna* virus 2A self-cleaving peptide[35]; WPRE, woodchuck hepatitis post-transcriptional response element; bGH pA, bovine growth hormone poly-A signal. phiLOV2.1[36] (330 bp) was chosen as a shorter fluorescent marker to ensure efficient AAV packaging. **d,** A TALE targeting either the endogenous mouse

locus *Grm5* or *Grm2* was fused to SID4X and virally transduced into primary neurons. SID4X-mediated target gene downregulation is shown for each TALE relative to levels in control neurons expressing GFP only (mean±s.e.m.; n = 3–4 biological replicates).

We further hypothesized that TALE-mediated targeting of histone effectors, such as SID, to endogenous loci could induce specific epigenetic modifications, which would enable the interrogation of epigenetic as well as transcriptional dynamics (**Fig. 14a**). We fused CRY2PHR with SID4X (**Fig. 14b**) and observed light-mediated transcription repression of *Grm2* in neurons (**Fig. 14c**) accompanied by ~2-fold reduction in H3K9 acetylation at the targeted *Grm2* promoter (**Fig. 14d**).
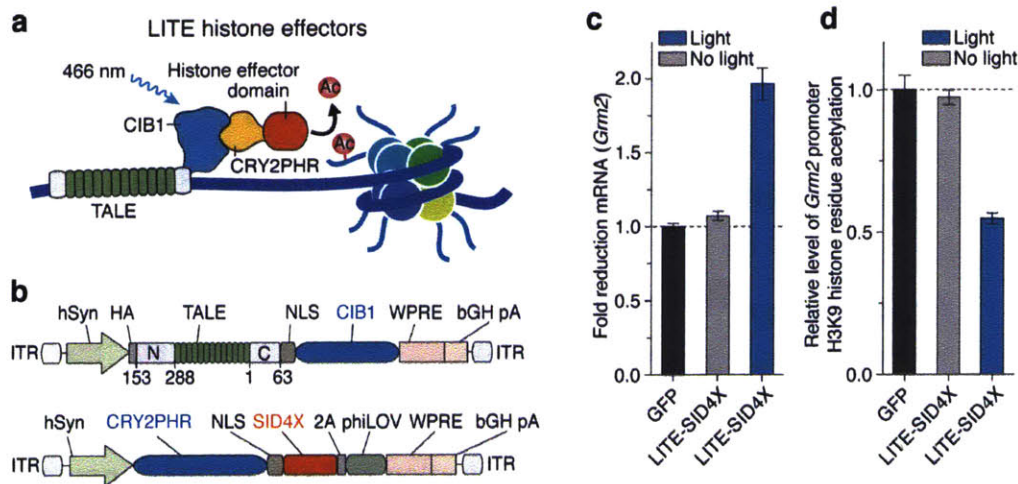
Figure 14 I LITE mediated epigenetic modifications. a, LITE epigenetic modifiers (epiLITE). **b**, epiLITE AAV vectors. **c**, epiLITE-mediated repression of endogenous *Grm2* in neurons (mean±s.e.m; n = 4 biological replicates). **d**, epiLITE-mediated decrease in H3K9 histone acetylation at the *Grm2* promoter (mean±s.e.m; n = 4 biological replicates).

In order to expand the diversity of histone residue targets for locus-specific histone modification, we next derived a set of 32 repressive histone effector

domains. Selected from across a wide phylogenetic spectrum, the domains include histone deacetylases (HDACs), methyltransferases (HMTs), acetyltransferase (HAT) inhibitors, as well as HDAC and HMT recruiting proteins. Preference was given to proteins and functional truncations of small size to facilitate efficient AAV packaging. The resulting epigenetic mark-modifying TALE-histone effector fusion constructs (epiTALEs) were evaluated in primary neurons and Neuro 2a cells for their ability to repress *Grm2* and *Neurog2* transcription, respectively (**Fig. 15** and **Fig. 16**). In primary neurons, 23 out of 24 epiTALES successfully repressed transcription of *Grm2* ($p < 0.05$). Similarly, epiTALE expression in Neuro 2a cells led to decreased *Neurog2* expression for 20 of the 32 histone effector domains tested (**Fig. 16**; $p < 0.05$). We then expressed a subset of promising epiTALEs in primary neurons and Neuro 2a cells and quantified the relative histone residue mark levels at the target locus using ChIP-RT-qPCR (**Fig. 15** and **Fig. 17**). These domains provide a ready source of epigenetic effectors for LITE-mediated control of specific epigenetic modifications.
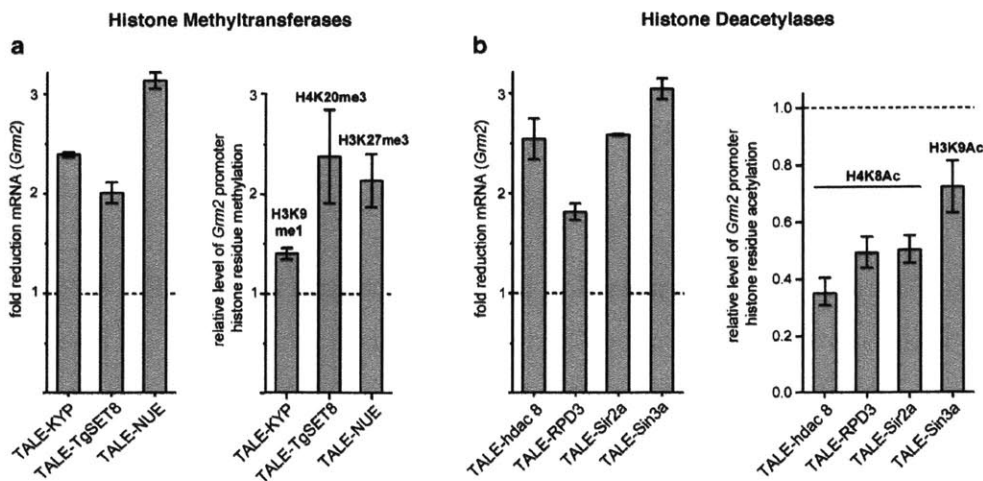
Figure 15 | TALE-mediated epigenetic modifications in primary neurons. a, b, epiTALE methyltransferases mediated decrease in *Grm2* mRNA and corresponding enrichment of H3K9me1, H4K20me3 and H3K27me3 at the *Grm2* promoter ($n = 3$ biological replicates). **c, d,** epiTALE histone deacetylases mediated repression of *Grm2* and corresponding decreases in H4K8Ac and H3K9Ac marks at the *Grm2* promoter ($n=2$–5 biological replicates). Mean ± s.e.m. in all panels.
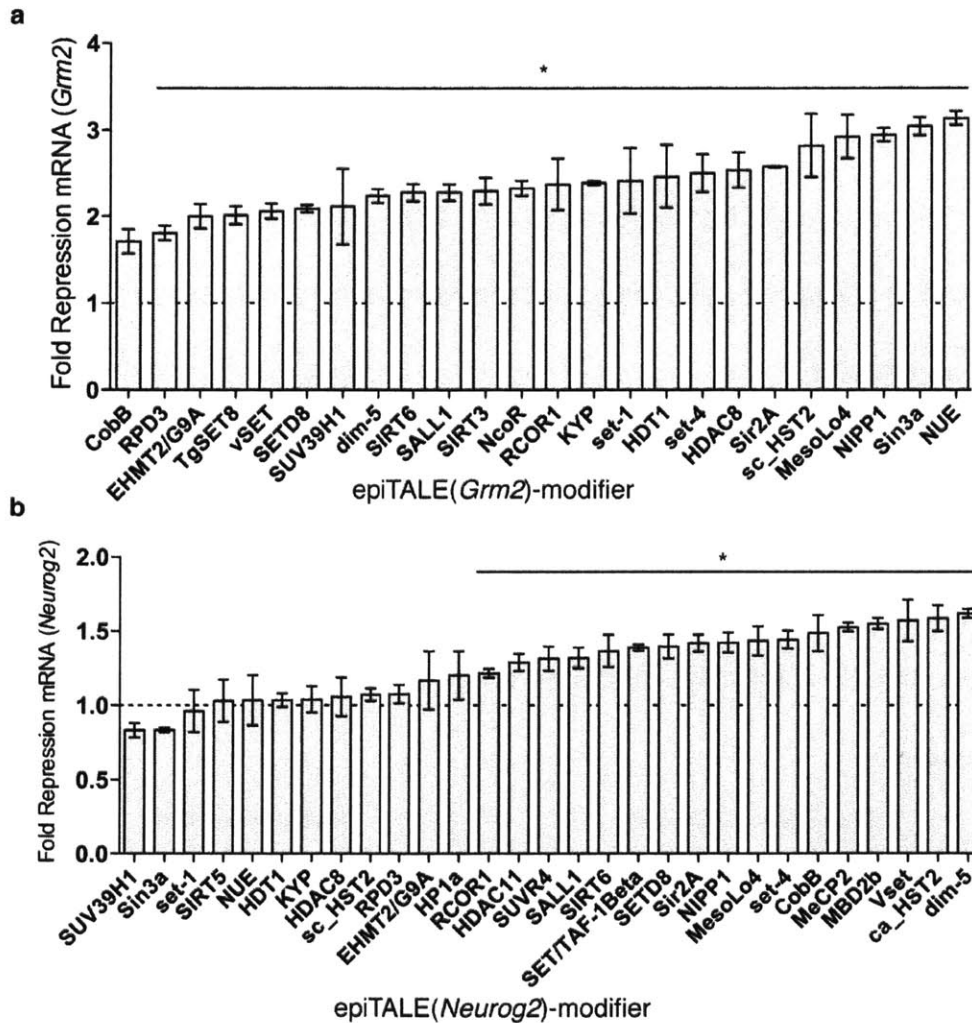
Figure 16 | Transcriptional repression by a diverse set of epiTALEs. a, 24 different histone effector domains were each fused to a *Grm2* targeting TALE. TALE–effector fusions were expressed in primary cortical mouse neurons using AAV transduction. *Grm2* mRNA levels were measured using qRT–PCR relative to neurons

38

transduced with GFP only. (*p<0.05; mean ± s.e.m.; n = 2–3 biological replicates.) **b**, A total of 32 epiTALEs were transfected into Neuro 2a cells. 20 of them mediated significant repression of the targeted *Neurog2* locus (*p<0.05; mean±s.e.m.; n = 2–3 biological replicates).

Histone Deacetylases   **Histone Methyltransferase**
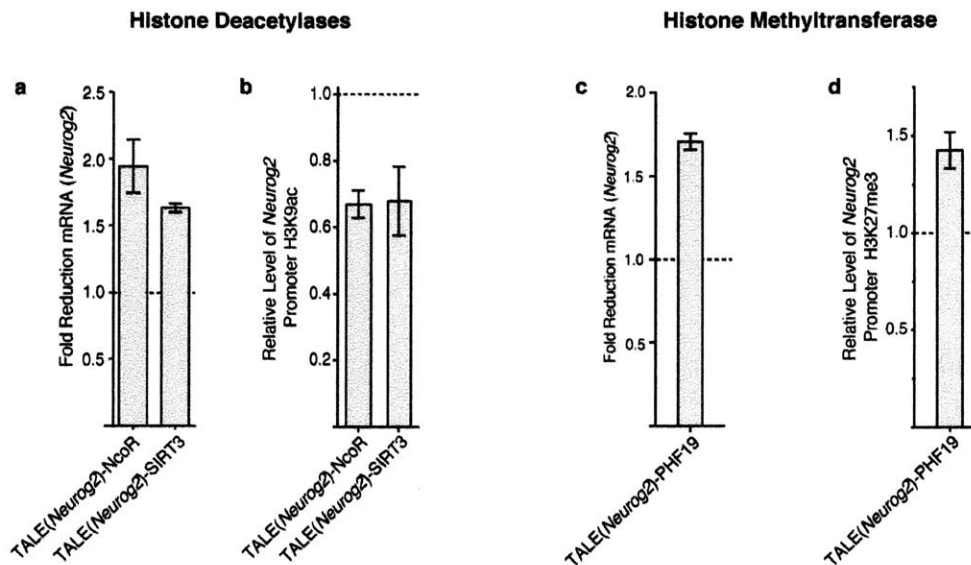
Figure 17 | EpiTALE mediated transcriptional repression and histone modifications in Neuro 2a cells. a, TALEs fused to histone-deacetylating epigenetic effectors NcoR and SIRT3 targeting the murine *Neurog2* locus in Neuro 2a cells were assayed for repressive activity on *Neurog2* transcript levels (mean±s.e.m.; n = 2–3 biological replicates). **b**, ChIP qRT–PCR showing a reduction in H3K9 acetylation at the *Neurog2* promoter for NcoR and SIRT3 epiTALEs (mean±s.e.m.; n = 2–3 biological replicates). **c**, The epigenetic effector PHF19 with known histone methyltransferase binding activity was fused to a TALE targeting *Neurog2*. Repression of *Neurog2* mRNA levels was observed (mean±s.e.m.; n = 2–3 biological replicates). **d**, ChIP qRT–PCR showing an increase in H3K27me3 levels at the *Neurog2* promoter for the PHF19 epiTALE (mean±s.e.m.; n = 2–3 biological replicates

Chapter 3:

Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex

Preface
This chapter is adapted from:

to fit the format of this thesis.

Achieving systematic, genome-scale perturbations within intact biological systems is important for elucidating gene function and epigenetic regulation. Genetic perturbations can be broadly classified as either loss-of-function or gain-of-function (GOF) based on their mode of action. To date, various genome-scale loss-of-function screening methods have been developed, including approaches employing RNA interference[37,38] and the RNA-guided endonuclease Cas9 from the microbial adaptive immune system CRISPR[39-41]. Genome-scale GOF screening approaches have largely remained limited to the use of cDNA library overexpression systems. However, it is difficult to capture the complexity of

transcript isoform variance using these libraries, and large cDNA sequences are often difficult to clone into size-limited viral expression vectors. The cost and complexity of synthesizing and using pooled cDNA libraries have also limited their use. Additionally, the endogenous regulatory contexts of the overexpressed genes cannot be recapitulated with cDNA constructs. Novel technologies that overcome such limitations would enable systematic, genome-scale GOF perturbations at endogenous loci.

Programmable DNA binding proteins have emerged as an exciting platform for engineering synthetic transcription factors for modulating endogenous gene expression[42-50]. Among the established custom DNA binding domains, the CRISPR-associated endonuclease Cas9 is most easily scaled to facilitate genome-scale perturbations[39-41] due to its simplicity of programming relative to zinc finger proteins and transcription activator-like effectors (TALEs). Cas9 nuclease can be easily converted into an RNA-guided DNA binding protein (dCas9) via inactivation of its two catalytic domains[10,11] and can then be fused to transcription activation domains. Targeting dCas9-activator domain fusions to the promoter region of endogenous genes can then achieve modulation of gene expression[45-49]. Although the current generation of dCas9-based transcription effectors are able to achieve activation of some endogenous loci, the magnitude of transcriptional up-regulation achieved by individual single-guide RNAs (sgRNAs) typically ranges from low to ineffective (**Fig. 18**)[46,47,49]. Targeting with a

combination of sgRNAs tiling a given promoter region can result in more robust

transcriptional activation[47-49], but this requirement presents enormous challenges

for scalability, and in particular for establishing pooled, genome-wide GOF

screens.
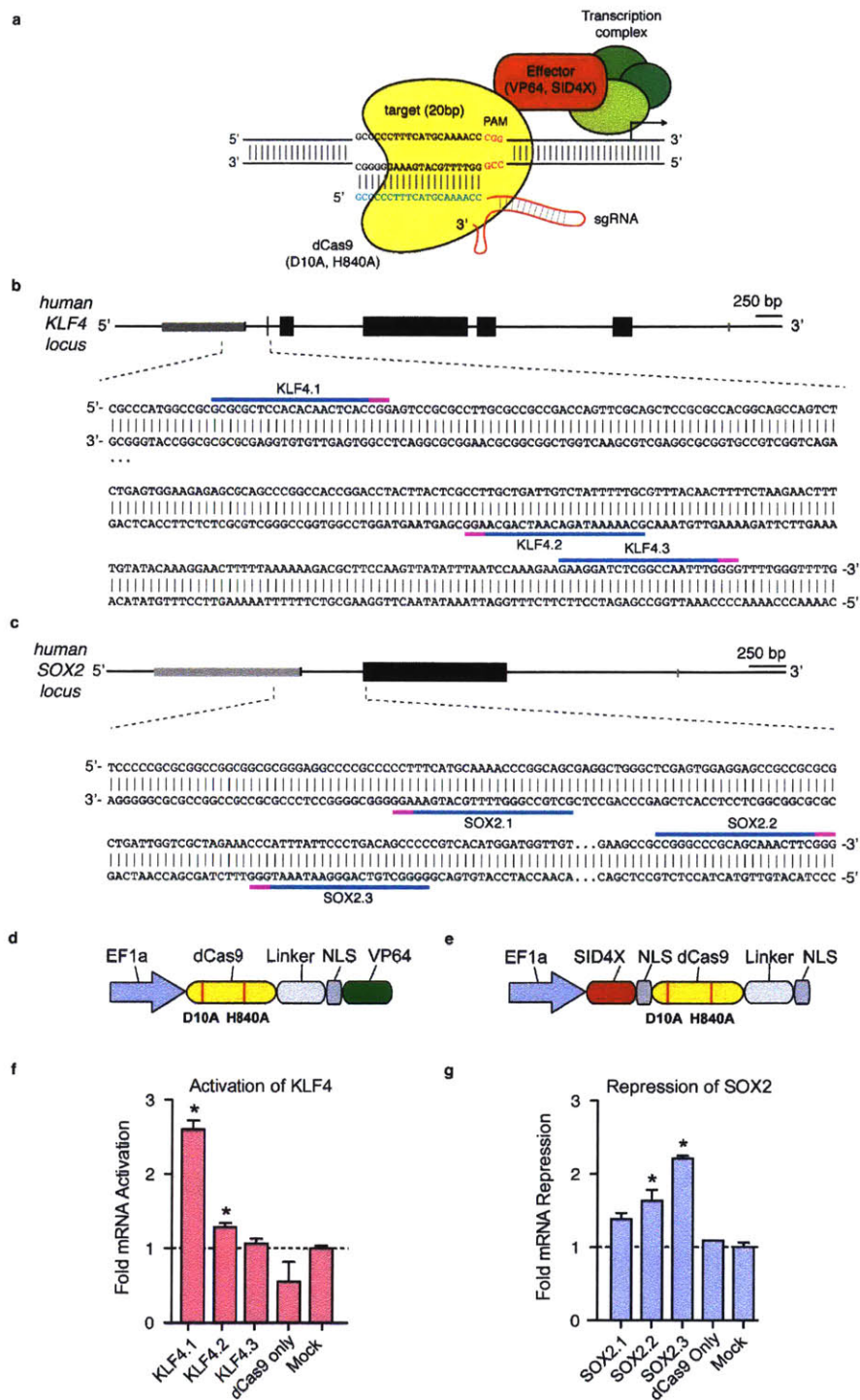
Figure 18 | RNA-guided DNA binding protein Cas9 targeting of transcription effector domains to specific genomic loci. a, The RNA-guided nuclease Cas9 from the type II Streptococcus pyogenes CRISPR/Cas system can be converted into a nucleolytically inactive RNA-guided DNA binding protein (dCas9) by introducing two alanine substitutions (D10A and H840A). Schematic showing that a synthetic guide RNA

(sgRNA) can direct dCas9-effector fusion to a specific locus in the human genome. The sgRNA contains a 20-bp guide sequence at the 5' end, which specifies the target sequence. On the target genomic DNA, the 20-bp target site needs to be followed by a 5'-NGG PAM motif. **b**, **c**, Schematics showing the sgRNA target sites in the human KLF4 and SOX2 loci, respectively. Blue bars indicate target sites. Magenta bars indicate PAM sequences. **d**, **e**, Schematics of the dCas9–VP64 transcription activator and SID4X–dCas9 transcription repressor constructs. **f**, **g**, dCas9–VP64 and SID4X–dCas9 mediated activation of KLF4 and repression of SOX2, respectively. All mRNA levels were measured relative to GFP transfected 293FT cells (mean ± s.e.m.; n = 3 biological replicates).

In order to improve and expand applications of Cas9, we recently undertook crystallographic studies elucidating the atomic structure of the Cas9-sgRNA-target DNA tertiary complex[51], enabling rational engineering of Cas9 and sgRNA. Here we report a series of structure-guided engineering efforts to create a potent transcription activation complex capable of mediating robust up-regulation with a single sgRNA. Using this new activation system, we demonstrate activation of endogenous genes as well as non-coding RNAs, elucidate design rules for effective sgRNA target sites, and establish and apply a genome-wide dCas9-based transcription activation screening system to study drug resistance in a cellular model of melanoma. These results collectively demonstrate the broad applicability of CRISPR-based GOF screening for functional genomics research.

Structure-guided design of a dCas9-based transcription activation complex

A key step in transforming the Cas9-sgRNA complex into an effective transcriptional activator is finding optimal anchoring positions for the activation

domains. Previous designs of dCas9-based transcription activators have relied on fusion of transactivation domains to either the N- or C-terminus of the dCas9 protein. To explore whether alternate anchoring positions would provide improved performance, we examined a crystal structure of the *Streptococcus pyogenes* dCas9 (D10A/H840A) in complex with a single guide RNA (sgRNA) and complementary target DNA[51]. We observed that the tetraloop and stem-loop 2 of the sgRNA protrude outside of the Cas9-sgRNA ribonucleoprotein complex, with the distal 4 bp of each stem completely free of interactions with Cas9 amino acid sidechains. Based on these observations, along with functional data demonstrating that substitutions and deletions in the tetraloop and stem-loop 2 regions of the sgRNA sequence do not affect Cas9 catalytic function[51] we reasoned that the tetraloop and stem-loop 2 could tolerate the addition of protein-interacting RNA aptamers to facilitate the recruitment of effector domains to the Cas9 complex (**Fig. 19**).
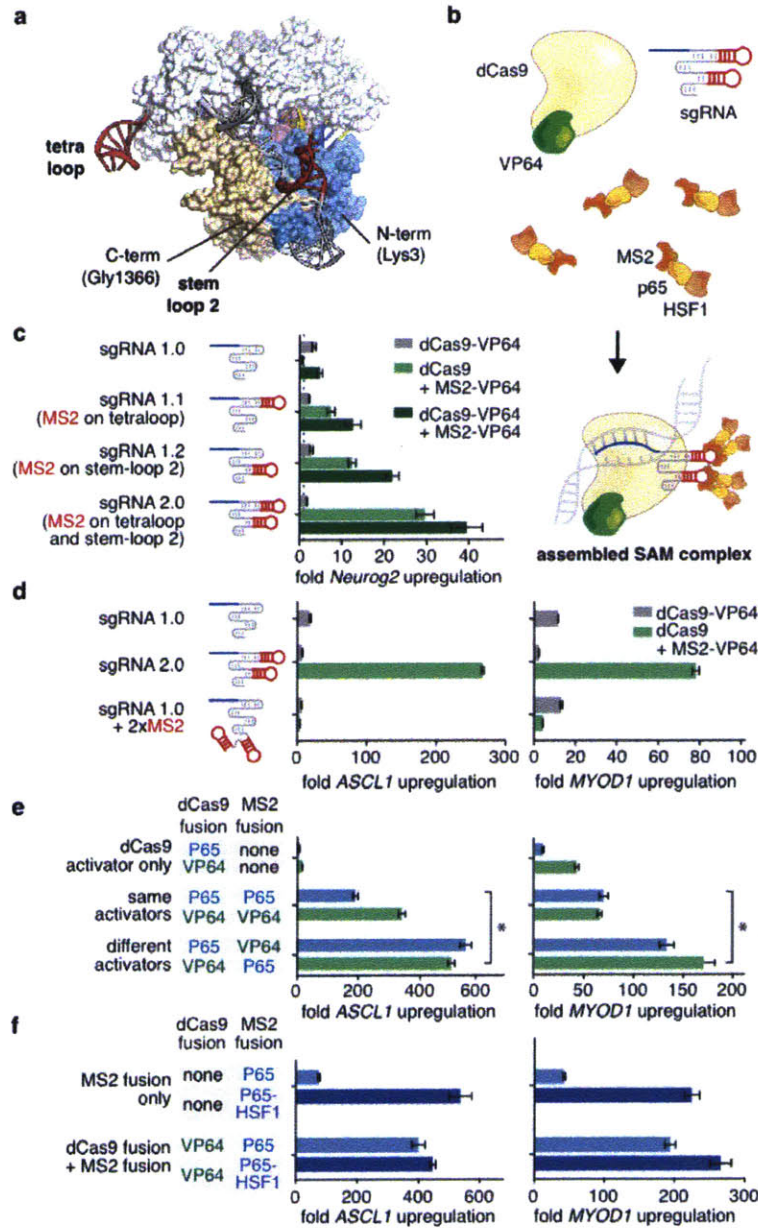
Figure 19 | Structure-guided design and optimization of an RNA-guided transcription activation complex. a, A crystal structure of the Cas9-sgRNA-target DNA tertiary complex (PDB ID: 4OO8)[51] reveals that the sgRNA tetraloop and stem loop 2 largely do not contact Cas9 amino acid residues in this conformation and can be modified without altering existing Cas9-sgRNA interactions. **b,** Diagram of the three-component synergistic activation mediator (SAM) system: sgRNA 2.0, MS2-p65-HSF1, and dCas9-VP64. MS2 stem-loop additions on the sgRNA are highlighted in red. **c,** Design and optimization of sgRNA scaffolds for optimal recruitment of MS2-VP64

transactivators in Neuro-2a cells. **d,** MS2 stem-loop placement within the sgRNA significantly affects transcription activation efficiency. **e,** Combinations of different activation domains act in synergy to further enhance the level of transcription activation. **f,** Addition of the HSF1 transactivation domain to MS2-p65 further increases the efficiency of transcription activation. Experiments for d-f were performed in 293FT cells. All values are mean ± SEM with n = 3. * indicates p <0.05 based on Student's t-test.

We selected a minimal hairpin aptamer, which selectively binds dimerized MS2 bacteriophage coat proteins in mammalian cells, and appended it to the sgRNA tetraloop and stem-loop 2[52,53] (**Fig. 20**). We next tested whether MS2-mediated recruitment of VP64 to the tetraloop and stem-loop 2 could mediate transcriptional up-regulation more efficiently than a dCas9-VP64 fusion alone. As predicted, aptamer-mediated recruitment of MS2-VP64 to either tetraloop (sgRNA1.1) or stem-loop 2 (sgRNA1.2) mediated 3- and 5-fold higher levels of *Neurog2* up-regulation than a dCas9-VP64 fusion (sgRNA 1.0), respectively. Recruitment of VP64 to both positions (sgRNA 2.0) resulted in an additive effect, leading to 12-fold increase over dCas9-VP64 (sgRNA 1.0). Combining sgRNA 2.0 with dCas9-VP64 instead of dCas9 provided an additional 1.3-fold increase in *Neurog2* up-regulation. We further compared sgRNA 2.0 to a previously-described sgRNA bearing two MS2-binding stem-loops at the 3` end (sgRNA + 2xMS2)[48] and found that sgRNA 2.0 drove 14- and 8.5-fold higher levels of transcription activation than sgRNA + 2xMS2 for *ASCL1* and *MYOD1*, respectively (**Fig. 19**). This difference could be due to either improved positioning of MS2 stemloops or to dCas9 protection of internal MS2 stemloops from

47

exonuclease degradation.
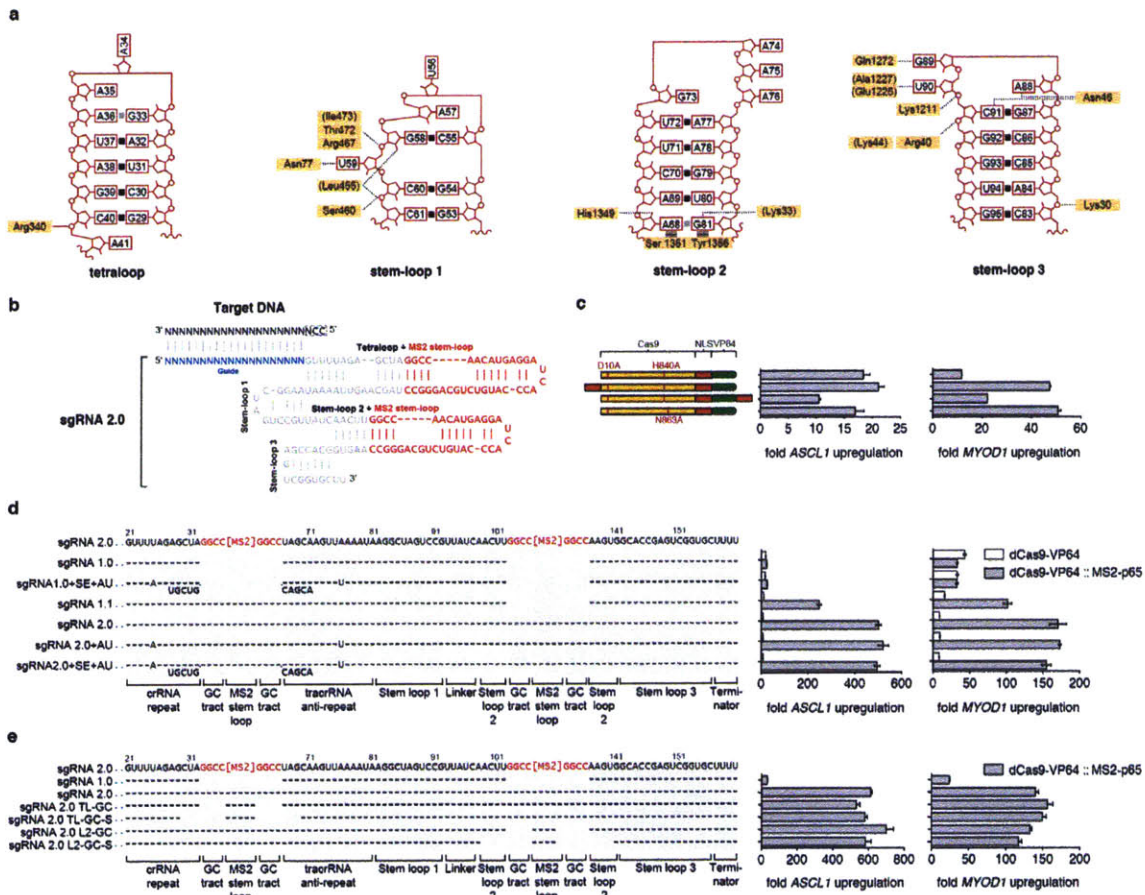
Figure 20 | Structure-guided engineering of Cas9 sgRNA. a, Schematic of the sgRNA stem-loops showing contacts between each stem-loop and Cas9. Contacting amino acid residues are highlighted in yellow. Tetraloop and stem-loop 2 do not make any contacts with Cas9 whereas stem-loops 1 and 3 share extensive contacts with Cas9. **b,** sgRNA 2.0 with MS2 stem-loops inserted into the tetraloop and stem-loop 2. **c,** Addition of a second NLS or an alternative HNH domain inactivating point mutation in Cas9 improve efficiency of transcription activation for *MYOD1* moderately. **d,** dCas9-VP64 activators exhibit improved performance by recruitment of MS2-P65 to the tetraloop and stem-loop 2. Addition of an AU flip or extension in the tetraloop does not increase the effectiveness of dCas9-mediated transcription activation. **e,** Tetraloop and stem-loop 2 are amenable to replacement with MS2 stem-loops. Base changes from the

48

sgRNA 2.0 scaffold are shown at the respective positions, with dashes indicating unaltered bases and bases below dashes indicating insertions. Deletions are indicated by absence of dashes at respective positions. All figures are n = 3 and mean ± SEM.

Effector Domains Act in Synergy to Enhance Transcription Activation

To further improve the potency of Cas9-mediated gene activation, we considered how transcriptional activation is achieved in natural contexts, where endogenous transcription factors generally act in synergy with co-factors to stimulate transcription[54]. We thus hypothesized that combining VP64 with additional, distinct activation domains could improve activation efficiency. We chose the NF-κB trans-activating subunit p65, which, while sharing some common co-factors with VP64, recruits a distinct subset of transcription factors and chromatin remodeling complexes. For example, p65 has been shown to recruit AP-1, ATF/CREB, and SP1[55], whereas VP64 recruits PC4[56], CBP/p300[57], and the SWI/SNF complex[58].

We then varied the effector domain fused to dCas9 or MS2. Hetero-effector pairing of dCas9 and MS2 fusion proteins (e.g. dCas9-VP64 paired with MS2-p65 or dCas9-p65 with MS2-VP64) provided over 2.5-fold higher transcription activation for both *ASCL1* and *MYOD1* than homo-effector pairing (e.g. dCas9-VP64 paired with MS2-VP64 or dCas9-p65 with MS2-p65) (**Fig. 19**). We further explored this concept of domain synergy by introducing the activation domain

from human heat-shock factor 1 (HSF1)[59] as a third activation domain, and found that an MS2-p65-HSF1 fusion protein further improved transcriptional activation of *ASCL1* (12%) and *MYOD1* (37%). Additional modifications to the sgRNA as well as Cas9 protein provided only minor improvements (**Fig. 20**). Based on these collective results, we concluded that the combination of sgRNA 2.0, NLS-dCas9-VP64, and MS2-p65-HSF1 comprises the most effective transcription activation system, and designated it synergistic activation mediator (SAM). For simplicity, we will refer to sgRNA 2.0 as sgRNA in subsequent discussions, unless noted otherwise.

Characterization of SAM efficacy and determination of design rules for efficient sgRNAs

To thoroughly evaluate the effectiveness of SAM for activating endogenous gene transcription, we chose 12 genes that have been previously found by several groups to be difficult to activate using dCas9-VP64 and individual sgRNA 1.0 guides[46,48,49]. For each gene, we selected 8 sgRNA target sites spread across the proximal promoter between -1000bp and the +1 transcription start site (TSS). For 9 out of 12 genes, the maximum level of activation achieved using dCas9-VP64 with any of the 8 sgRNA 1.0 guides was less than 2-fold, while the remaining three genes (ZFP42, KLF4 and IL1b) were maximally activated between 2- and 5-fold (**Fig. 21**). In contrast, SAM stimulated transcription at least

2-fold for all genes and more than 15-fold for 8 out of 12 genes. Consistently, SAM performed better than sgRNA 1.0 + dCas9-VP64 for all 96 guides, with a median gain of 105-fold greater upregulation (activation by SAM divided by activation by sgRNA 1.0 + dCas9-VP64) across all 12 genes.
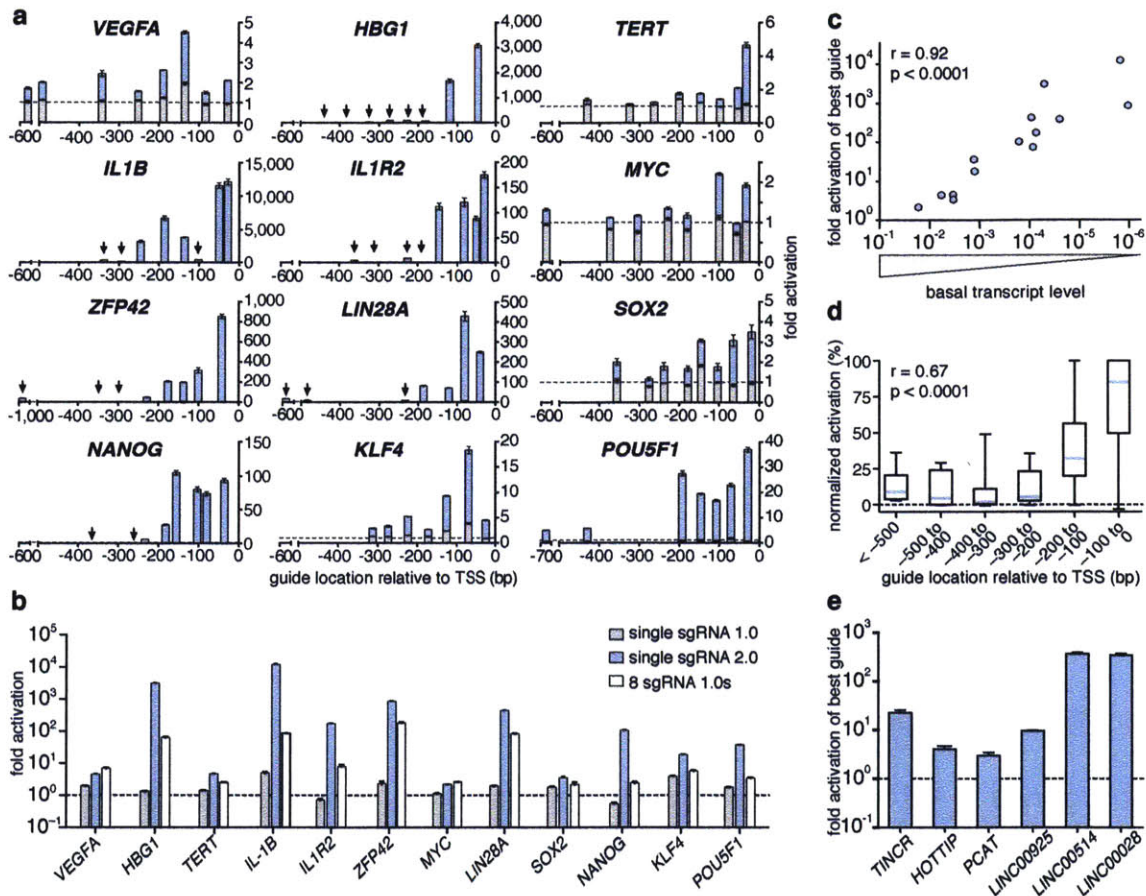
Figure 21 I Characterization of SAM-mediated gene and lincRNA activation and derivation of selection rules for efficient sgRNAs. a, Fold activation of 12 different genes plotted against the location of the sgRNA. Distances are measured in bp relative to the TSS at +1. sgRNA 1.0 with dCas9-VP64 (grey), sgRNA 2.0 with dCas9-VP64 and MS2-p65-HSF1 (blue). Arrows indicate sgRNA target sites with poor transcription activation. All values are mean ± SEM with $n = 3$. **b,** Comparison of activation efficiency of different target genes: dCas9-VP64 and a single sgRNA 1.0; dCas9-VP64 with a single sgRNA 2.0 and MS2-p65-HSF1, and dCas9-VP64 with a mixture of 8 sgRNA

51

1.0s. **c,** Efficiency of target gene transcription activation as a function of their baseline expression levels. Genes with a higher basal level of transcription exhibit a lower fold up-regulation. For each target gene, the baseline expression level is measured using qPCR in the GFP-transfected control cells and expressed as fold lower expression relative to GAPDH on the x-axis. **d,** Correlation of gene activation efficiency with sgRNA targeting position in the proximal promoter region expressed as distance to the TSS. Activation efficiency of each sgRNA for the same target gene is normalized against the highest-activating sgRNA. Proximity to the TSS is positively correlated with target up-regulation. Blue lines indicate median values, boxes indicate 25th and 75th percentiles. **e,** Fold activation of six lincRNA transcripts by dCas9-VP64, MS2-p65-HSF1, and sgRNA (best sgRNA out of 8 tested). SAM activates both characterized and uncharacterized lincRNA transcripts. All experiments were performed in 293FT cells. All values are mean ± SEM with n = 3.

Previous studies have demonstrated that the poor activation efficiency of single sgRNAs can be overcome by combining dCas9-VP64 with a pool of sgRNAs tiling the proximal promoter region of the target gene[47-49]. Therefore we compared the single sgRNA activation efficiency of SAM against dCas9-VP64 combined with a pool of 8 sgRNA 1.0 guides, all targeting the same gene. For most genes, SAM with a single sgRNA performed more robustly than dCas9-VP64 with pools of 8 sgRNA 1.0 guides (**Fig. 2b**). On average, SAM with single sgRNAs achieved 15 times more activation than dCas9-VP64 combined with pools of 8 sgRNA 1.0 guides. For all 12 genes, we also tested MS2-p65 and MS2-p65-MyoD1 where MyoD1 denotes the minimal transactivating subunit of human *MYOD1*[60]. For 9 out of 12 genes, incorporation of either HSF1 or MyoD1 as a third transactivator fusion outperformed MS2-p65 alone by 42% to 196% ($p < 0.01$, Student's t-test with FDR correction; **Fig. 22**).
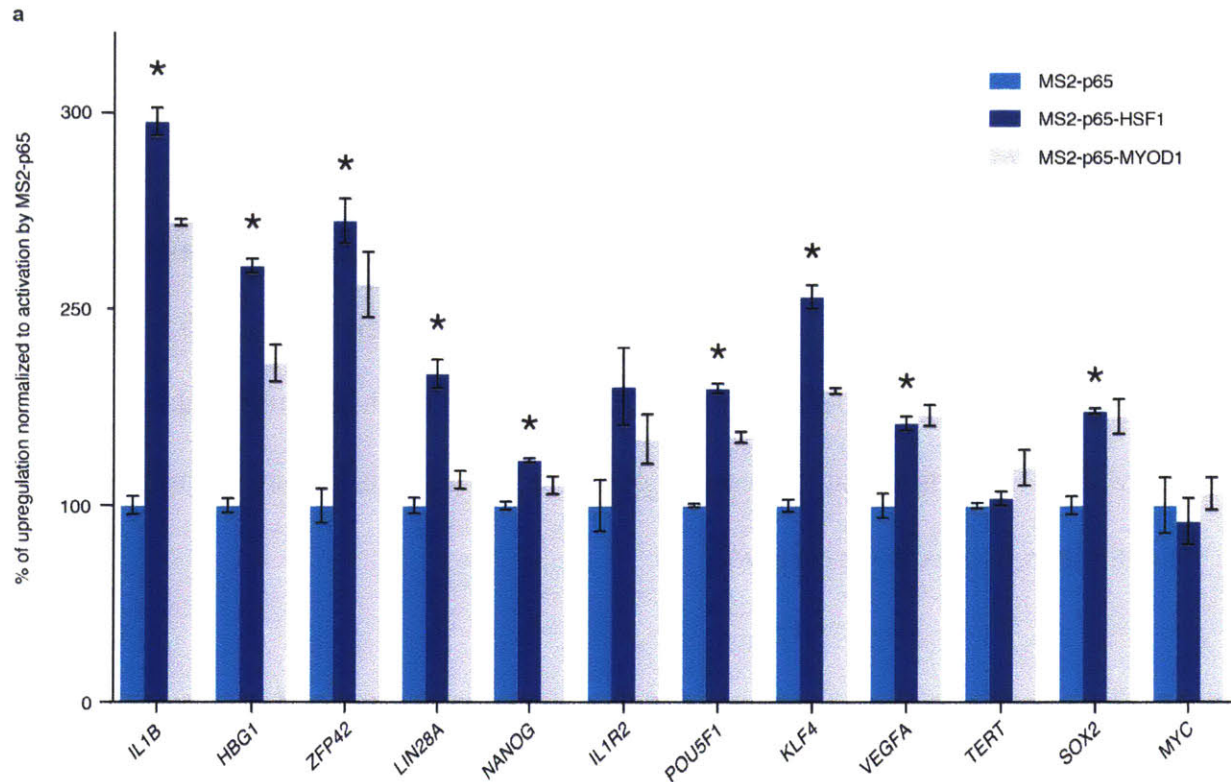
Figure 22 | SAM mediates efficient activation of a panel of 12 coding genes. a, Comparison of the activation levels of 12 genes with dCas9-VP64 in combination with MS2-P65, MS2-P65-HSF1, or MS2-P65-MYOD1. MS2-P65-HSF1 mediated significantly higher levels of activation than MS2-P65 alone for 9 out of 12 genes. The best guide out of 8 tested for each gene (**Fig. 2a**) was used in this experiment. Activation levels for each type of MS2-fusion is presented as a percentage relative to the activation achieved using MS2-P65. All figures are n = 3 and mean ± SEM.

Next, we sought to determine factors that contribute to inter- and intragenic variability of activation efficiency by different sgRNAs. For intergene variability, differences in activation magnitudes could be due to epigenetic factors (e.g. variance in chromatin accessibility, interactions with distal enhancers, and transcription factor competition) and/or from variation in basal transcription levels. We were thus interested correlating basal transcription with the level of

transcription activation achieved using SAM. Using the relative transcriptional levels of target genes in control samples, we observed a highly significant correlation between the inverse of basal transcript level and the fold up-regulation achieved using SAM (**Fig. 2c**; r = 0.94, $p < 0.0001$). Whereas highly expressed genes (e.g. *MYC, VEGFA, TERT, SOX2*) were moderately upregulated, lowly expressed genes (e.g. *HBG1, IL1B, ZFP42*) were more highly upregulated by SAM. This suggests that the level of possible activation is largely determined by the initial basal expression level.

To study the variability of SAM activity across target genes, we aggregated the activation data for all 96 guides and found the distance between the guide RNA target site and the TSS to be the most significant predictor of activation efficiency (**Fig. 2d**; r = 0.67, p < 0.0001). For all genes, the strongest levels of activation were consistently achieved by targeting within the -200 bp to +1 bp window, with 85% of guides within 200 bp upstream of the TSS achieving at least 25% of the maximal activation of a given gene. However, proximity to TSS within the -200 to +1 bp window did not correlate with more potent activation. This simple design guideline can inform the selection of efficient sgRNAs for gene activation.

We also sought to test whether SAM is able to activate non-coding elements in addition to protein-coding genes. We chose 3 lincRNA targets with known functions (TINCR[61], HOTTIP[62], and PCAT[63]) and 3 with unknown functions

(LINC00925, LINC00514 and LINC00028). Similar to our previous mRNA up-regulation experiments, we used RefSeq annotations to select 8 sgRNA target sites from the proximal promoter (-800bp to +1) of each lincRNA. We found that SAM indeed mediated significant up-regulation of lincRNA transcripts from 3-fold up-regulation of PCAT to 360-fold up-regulation of LINC00514 (**Fig. 21**). Similar to our finding with activation of protein coding genes, dCas9-VP64 coupled with MS2-p65-HSF1 or MS2-p65-MyoD1 led to significantly higher maximal activation than dCas9-VP64 with MS2-P65 for all 6 lincRNAs ($p < 0.01$) (**Fig. 23**).
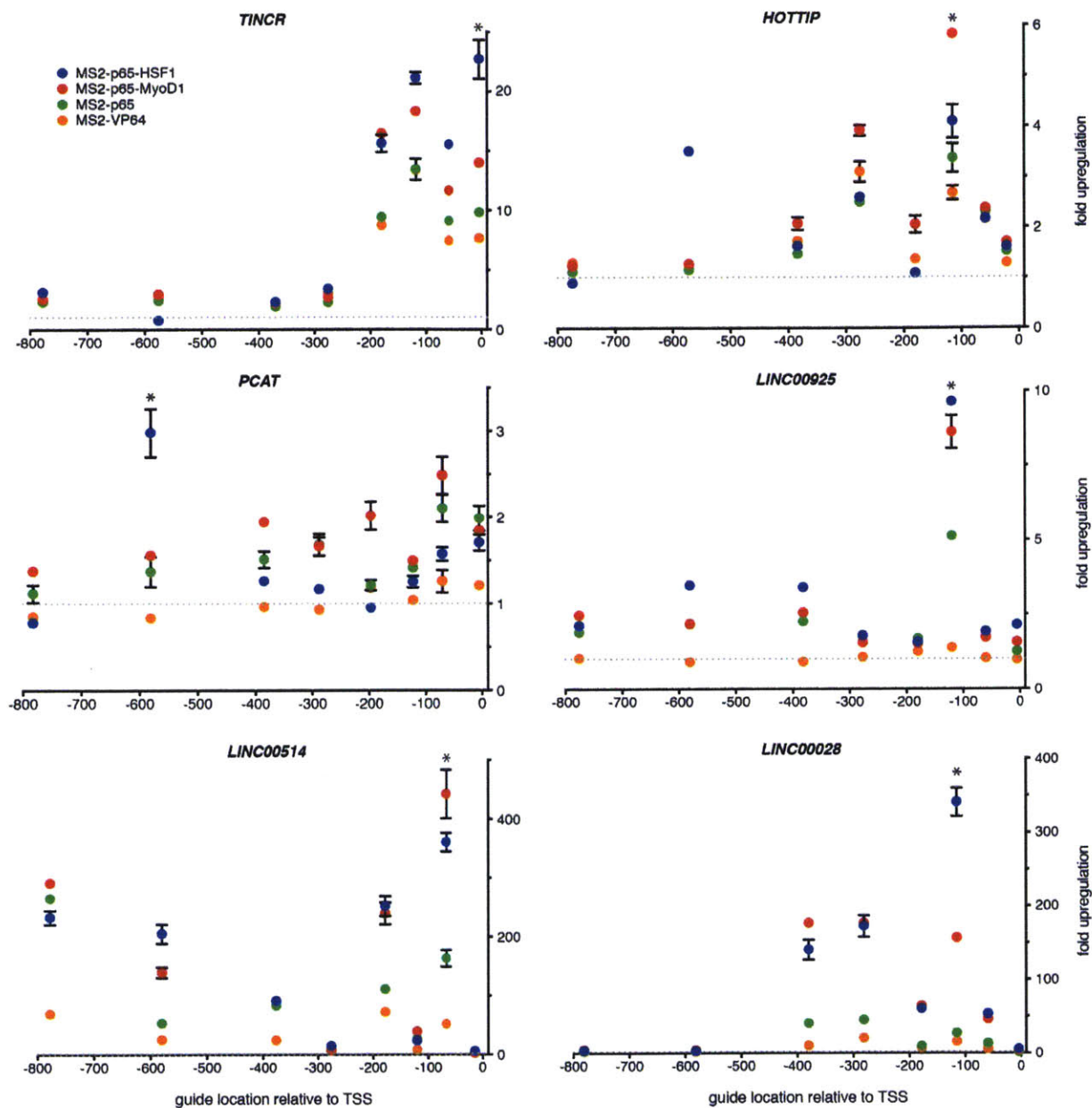
Figure 23 I Activation of lincRNAs by SAM. Six lincRNAs, three characterized and three uncharacterized, were targeted using SAM. For each lincRNA, 8 sgRNAs were designed to target the proximal promoter region (+1 to -800bp from the TSS) with 4 different MS2 activators (MS2-P65-HSF1, MS2-P65-MyoD1, MS2-P65, and MS2-VP64) in combination with dCas9-VP64. MS2 activators with a combination of 2 different domains (MS2-p65-HSF1 or MS2-p65-MyoD1) consistently provided the highest activation for each lincRNA, * denotes $p < 0.01$ for MS2-p65-HSF1 or MS2-p65-MyoD1 vs. MS2-p65. N = 3 and mean ± SEM is shown.

56

SAM mediates simultaneous activation of multiple genes

Understanding the complexity of gene network and transcription regulation

beyond single gene contributions requires the ability to simultaneously modulate

gene expression at multiple loci. This could enable, for example, combinatorial

targeting of multiple elements of a signaling pathway or sets of genes that

coordinate signaling in disease states. We thus sought to test whether SAM can

activate multiple genes simultaneously, and to characterize parameters impacting

multiplexing performance. We tested simultaneous activation of three sets of 2, 4,

6 or 8 genes and one set of 10 genes (**Fig. 24**) by co-expressing combinations of

sgRNAs.
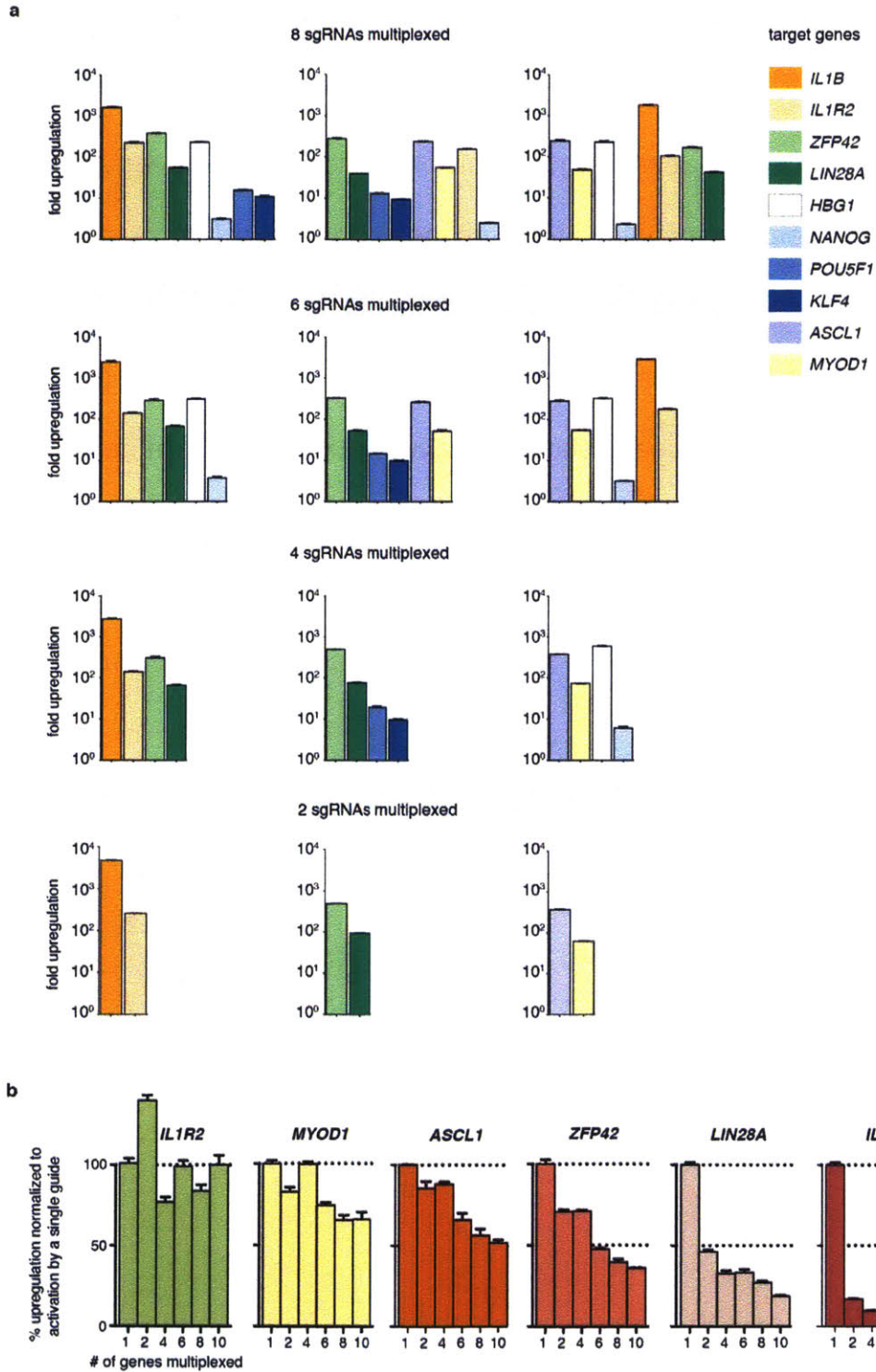
Figure 24 | Multiplexed activation using SAM. a,b Activation of a panel of 10 genes by combinations of 2, 4, 6, or 8 sgRNAs simultaneously. The mean fold up-regulation is shown on a $\log_{10}$ scale. MS2-P65-HSF1 and dCas9-VP64 were used in this experiment.

58

We observed successful activation of all genes (>2-fold) for all gene combinations tested, including simultaneous activation of 10 genes (**Fig. 25**). As expected, most genes (excluding *IL1R2*) exhibited a decrease in the amount of up-regulation achieved when concurrently targeted with 9 other genes.
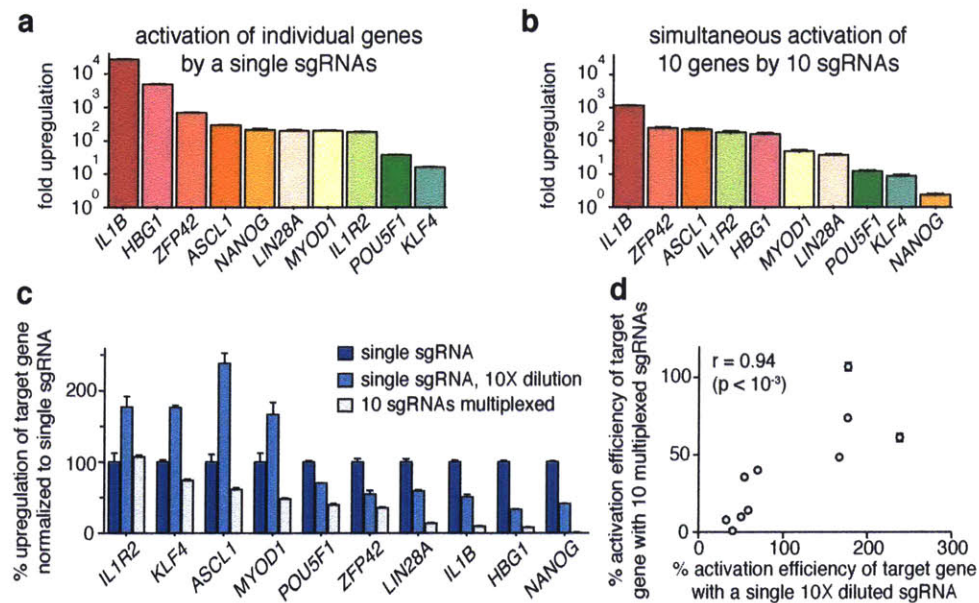
Figure 25 I Simultaneous activation of endogenous genes using multiplexed sgRNA expression. a, Activation of individual genes by single sgRNAs with dCas9-VP64 and MS2-p65-HSF1. **b,** Simultaneous activation of the same ten genes using a mixture of ten sgRNAs each targeting a different gene. **c,** Effect of sgRNA dilution on gene activation efficiency. **d,** Correlation between the activation efficiency of a single 10-fold diluted sgRNA and that of the same sgRNA delivered within a mixture of ten different-gene targeting sgRNAs. All values are mean ± SEM with n = 3.

Interestingly, the relative activation levels of each gene changed between multiplex activation and single-gene activation experiments. For example, whereas *NANOG* ranked 5th among the 10 targeted genes during single-gene

activation, it ranked 10th in the 10-plex activation experiment. Some genes showed no change or only a modest and gradual drop in activation when concurrently targeted alongside an increasing number of genes (e.g. *IL1R2, MYOD1, ASCL1*). Others, however, displayed a steep decrease in upregulation when combined with even a single gene partner (e.g. *LIN28A, IL1B, NANOG*). These distinct behaviours between genes were observed generally, across different gene pairings (**Fig. 26**).
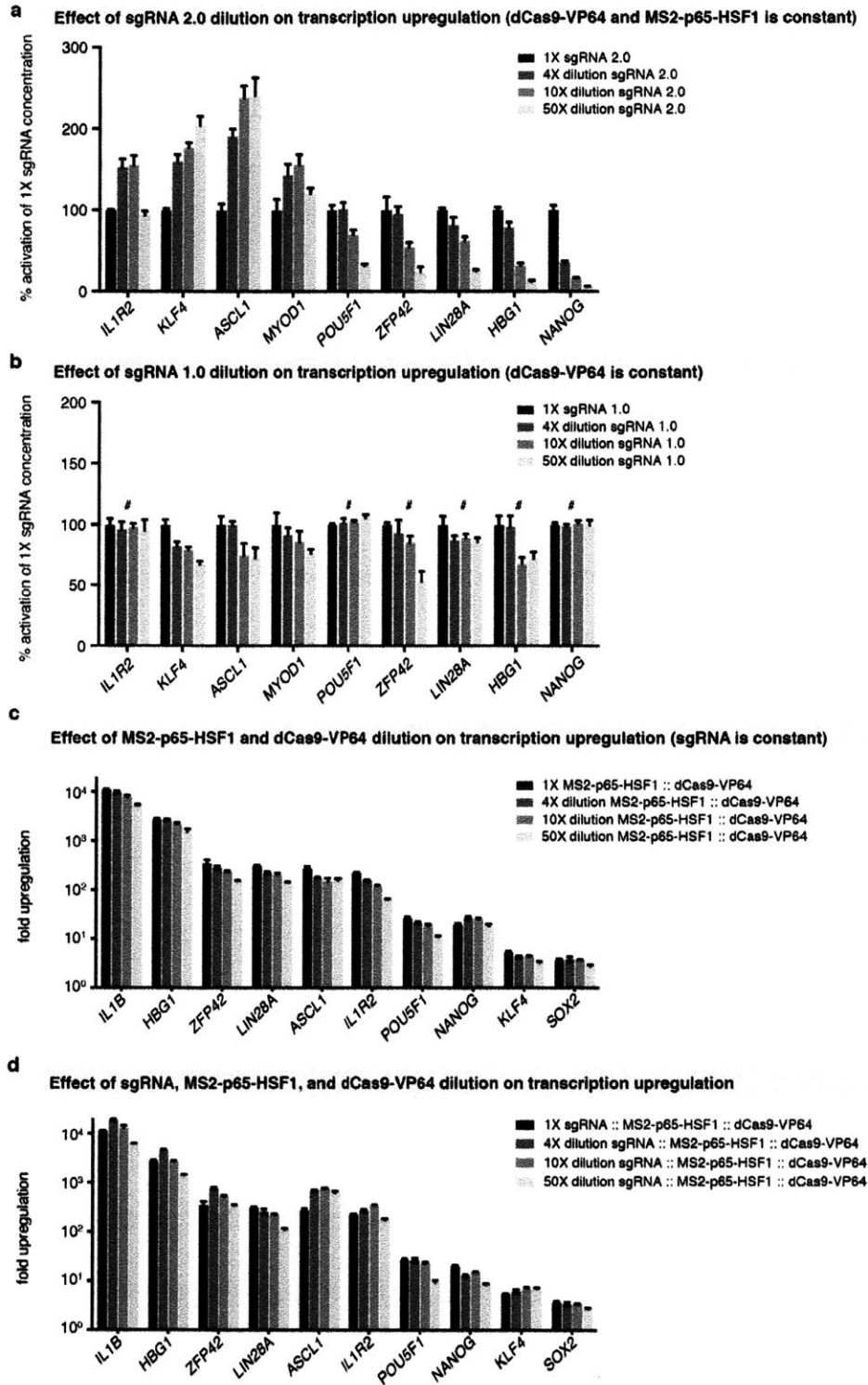
Figure 26 | The effect of guide and SAM-component dilution on target activation.

a, The results for dilution of sgRNA 2.0 on target activation. **b,** The result for dilution of sgRNA 1.0 on target activation. # denotes an activation of < 2-fold at 1X guide dilution. **c,** Effect of MS2-p65-HSF1 and dCas9-VP64 dilution, at 1:1, 1:4, 1:10, and 1:50 of the

original dosage for each component, on the effectiveness of transcription up-regulation. The amount of sgRNA expression plasmid was kept constant. **d,** Effect of diluting all three SAM components (dCas9-VP64, MS2-p65-HSF1, and sgRNA) at 1:4, 1:10, and 1:50 of the original dosage for each component. Fold up-regulation is calculated using GFP-transfected cells as the baseline. Error bars indicate S.E.M. and N = 3 for all figures.

We asked if this reduced activation of targets during multiplexing of 10 genes was due to the reduced amounts of sgRNA or SAM protein components available per gene. Surprisingly, diluting the sgRNA expression plasmid by 10-fold in single-gene activation experiments did not reduce activation for all genes (**Fig. 25**). We found that the genes exhibiting reduced activation as a result of sgRNA dilution were also dampened by multiplexing (**Fig. 25**; r = 0.94, p<0.001). In contrast, the activation efficiency of SAM was generally unperturbed by dilution of its protein components (dCas9-VP64 and MS2-p65-HSF1). Reducing the amount of transfected plasmid for both components by 10-fold led to an average drop of 26% in activation efficiency (**Fig. 26**). Activation efficiency remained stable particularly when all three components (including sgRNA) were diluted, retaining on average 100% activation efficiency across a 50-fold dilution range (**Fig. 26**). The finding that SAM is highly efficient even at low transfection concentrations was particularly promising for application in genome-scale pooled screens, which rely on single-copy lentiviral integration.

Assessing the specificity of SAM-mediated transcription activation

An important consideration for the utility of Cas9 or SAM is its targeting

specificity. Recent analysis of genome-wide dCas9-binding revealed significant

concentration-dependent off-target binding[64], yet the effect of such off-target

binding events on the specificity of Cas9-based transcription activation or

repression systems remains unclear. In order to assess the specificity of SAM-

mediated transcription activation, we chose HBG1/2 as our target gene,

reasoning that globin genes, which primarily function in oxygen transport, would

have few downstream targets that could confound our specificity analysis. Using

RNA-seq, we measured whole transcriptome changes resulting from the

expression of SAM along with a single HBG1/2-targeting sgRNA. We found that

SAM specifically activated both HBG1 and HBG2 isoforms ($p < 0.05$, T-test after

0.01 FDR correction for multiple hypothesis testing), which share the same TSS,

(**Fig. 27**), while no other genes were found to be differentially expressed.
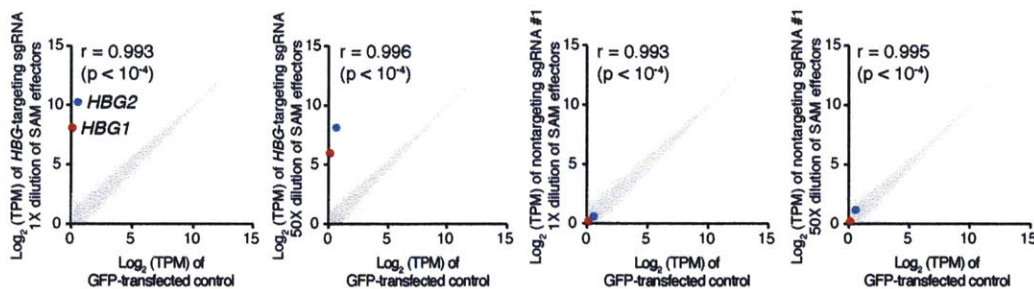
Figure 27 I Evaluation of SAM specificity. Expression levels in log(TPM) values of all detected genes in RNA-seq libraries of GFP-transfected controls (x-axis of all graphs) compared to (from left to right): SAM targeting HBG1/2 genes in 1x dilution and 50x dilution, non-targeting control sgRNAs in 1x dilution and 50x dilution (y-axis). Marked are the two statistically significant differentially expressed genes (T-test q-value < 0.05 with FDR correction): HBG1 (red) and HGB2 (blue). The average from n = 3 is shown

We also tested two additional non-targeting sgRNAs with guide sequences that do not share perfect homology with the human genome. We found only two genes, *S100A1* and *CYB5R2*, to be differentially expressed ($p < 0.05$, T-test after 0.01 FDR correction for multiple hypothesis testing) compared with GFP-expressing control (**Fig. 28**) for both non-targeting guides. These results suggest that SAM-mediated gene activation is specific with minimal off-target activity.
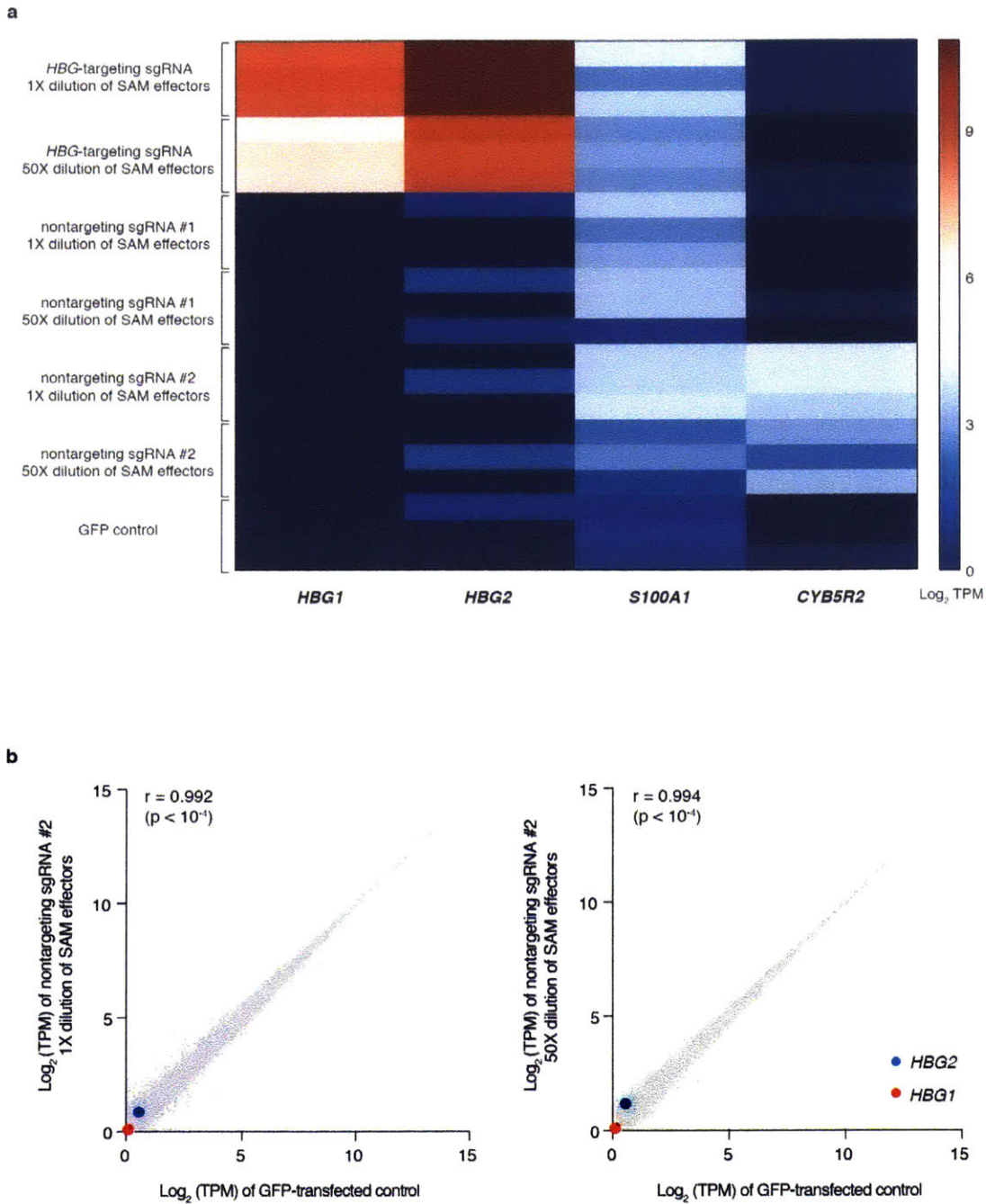
Figure 28 | RNA-seq analysis of transcriptome changes mediated by SAM. a, A heat map of log(TPM) expression values of all statistically significant differentially expressed genes (T-test q-value < 0.05 adjusted with FDR multiple hypothesis correction) found in any of the six experimental conditions compared to the GFP-transfected control. **b,** Expression levels in log(TPM) values of all detected genes in RNA-seq libraries of GFP-transfected controls (x-axis of all graphs) compared to (from left to right): non-targeting control sgRNA #2 in 1x dilution and 50x dilution (y-axis).

65

Marked are *HBG1* (red) and *HGB2* (blue).

Development of a genome-scale pooled transcription activation screen

The ability to activate target genes using a single sgRNA greatly facilitates the development of pooled, genome-scale transcriptional activation screening. As a first step towards developing a SAM-based screen, we cloned all three components into lentiviral vectors (**Fig. 29**). Each vector encodes a unique selection marker (Blast, Hygro, and Zeo or Puro) to enable selection of cells co-expressing all three SAM components. To assess the activity of a SAM system delivered via lentivirus at low multiplicity of infection (MOI), we targeted three genes: *MYC*, which is weakly activated; and *KLF4* and *MYOD1*, which are only moderately activated. We co-transduced HEK293FT cells with lenti-dCas9-VP64 and lenti-MS2-p65-HSF1 at MOI <1 and concurrently selected with Blast and Hygro for 7 days. We then transduced dCas9-VP65- and MS2-p65-HSF1-expressing cells with lentiviral sgRNA vectors (lenti-sgRNA) at low MOI (< 0.2) and selected for successfully transduced cells using either Puromycin or Zeocin. Target gene expression levels were measured four days post-transduction. All three genes were efficiently upregulated to levels comparable (*MYOD1*) or greater than those observed after transient SAM transfection (*MYC* and *KLF4*). Notably, expression levels achieved with Puromycin or Zeocin resistance markers on the sgRNA construct were not equal (**Fig. 29**).
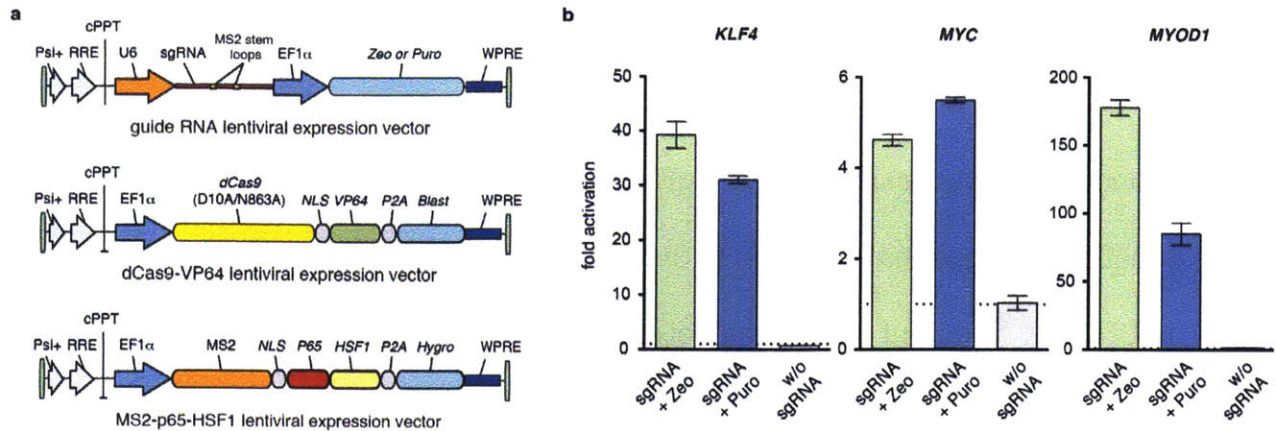
Figure 29 | lentiviral delivery of SAM. a, Design of three lentiviral vectors for expressing sgRNA, dCas9-VP64, and MS2-p65-HSF1. Each vector contains a distinct selection marker to enable co-selection of cells expressing all three vectors. **b,** Lentiviral delivery of SAM components was tested by first generating 293FT cell lines stably integrated with dCas9-VP64 and MS2-p65-HSF1, and subsequently transducing these cells with single-gene targeting lentiviral sgRNAs at MOI <0.2. Transcription activation efficiency is measured 4 days post sgRNA lentivirus transduction and selection with Zeocin or Puromycin. Activation is at least as effective as previously observed with transient transfection in all three cases.

Using genome-scale transcription activation screen to identify genes involved in BRAF inhibitor resistance

To enable genome-scale gain-of-function screening using SAM, we designed a genome-scale sgRNA library targeting every coding isoform from the RefSeq database (23,430 isoforms). For each gene, 3 sgRNAs were chosen to target sites within 200 bp upstream of the TSS, which was previously determined to provide more efficient activation (**Fig. 30**). The final library contained 70,290 guides, and we generated two separate libraries with Zeocin (lenti-sgRNA-Zeo)

or Puromycin (lenti-sgRNA-Puro) resistance.

Using this newly designed SAM sgRNA library, we aimed to identify gain-of-function changes that can lead to the development of BRAF inhibitor resistance in BRAF^{V600E} mutant melanoma cells. This complements our previous study employing a Cas9-mediated genome-scale gene knockout library[39]. The A375 melanoma cell line harbors the BRAF^{V600E} mutation and is naturally sensitive to BRAF inhibitors such as PLX4720 (PLX) and the closely related commercial therapeutic vemurafenib. Cells harboring sgRNAs that activate genes leading to PLX resistance should therefore be enriched after continued culture in the presence of the drug, whereas no such effect should be observed in cells treated with vehicle only.
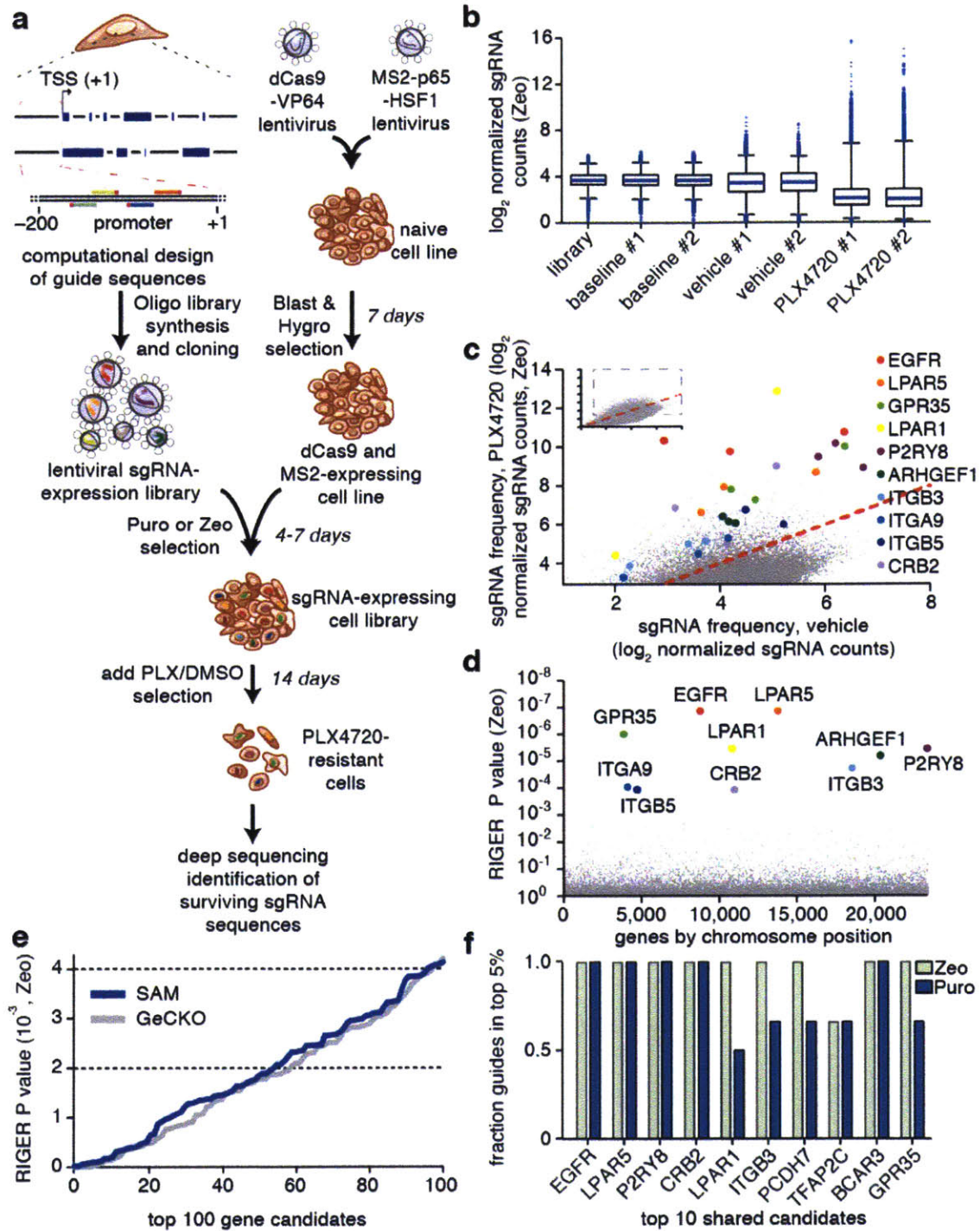
Figure 30 | Genome-scale gene activation screening identifies mediators of BRAF inhibitor resistance. a, Flow chart of transcription activation screening using SAM. **b,** Box plot showing the distribution of sgRNA frequencies post lentiviral transduction for baseline (day 3), vehicle (day 21), and PLX-4720 (day 21) conditions. **c,** Scatterplot

69

showing enrichment of specific sgRNAs after PLX-4720 treatment. **d,** Identification of top candidate genes using the RIGER *P* value analysis based on the average of both infection replicates. **e,** Comparison of RIGER *P* values for the top 100 hits from SAM and GeCKO[39] PLX-4720 resistance screens. **f,** Consistency of sgRNAs for top screening hits. Fraction of unique sgRNAs targeting each gene that are in the top 5% of all sgRNAs is plotted

To conduct the PLX-resistance screen, we generated a polyclonal A375 melanoma cell line constitutively expressing both dCas9-VP64 and MS2-p65-HSF1 components and transduced these cells with our genome-scale lenti-sgRNA-Zeo library at an MOI of 0.2. We first analyzed normalized guide counts for the input sgRNA-Zeo library at the baseline time point (3 days post infection) as well as 14 days post treatment with either PLX or vehicle. The sgRNA distribution was significantly different between cells treated with PLX and with vehicle for two independent infection replicate screens, with the majority of sgRNAs exhibiting a reduced representation and a small set of guides showing high enrichment for PLX treated cells (Wilcoxon rank sum test, $P < 0.0001$, median -1.3 for PLX vs. DMSO) **(Fig. 30)**.

For a number of gene targets, several sgRNAs targeting the same gene were enriched in PLX-treated cells, suggesting the importance of these genes for the formation of PLX resistance. To determine genes exhibiting consistently high enrichment across multiple sgRNAs, we employed the RNAi Gene Enrichment Ranking (RIGER) algorithm **(Fig. 30)**. The 10 most significant hits were

distributed throughout the genome. The significance of the *p*-values of our top

100 RIGER hits was comparable to those observed for genome-scale CRISPR-

mediated knockout (GeCKO) screening[39], indicating that the results obtained

from the SAM gain-of-function activation screen have similar statistical power

compared to Cas9 nuclease-based knockout screening (**Fig. 30**). 50% of our top

20 RIGER hits were replicated in a validation screen using puro selection, rather

than zeo, on the sgRNA library (**Fig. 31**).
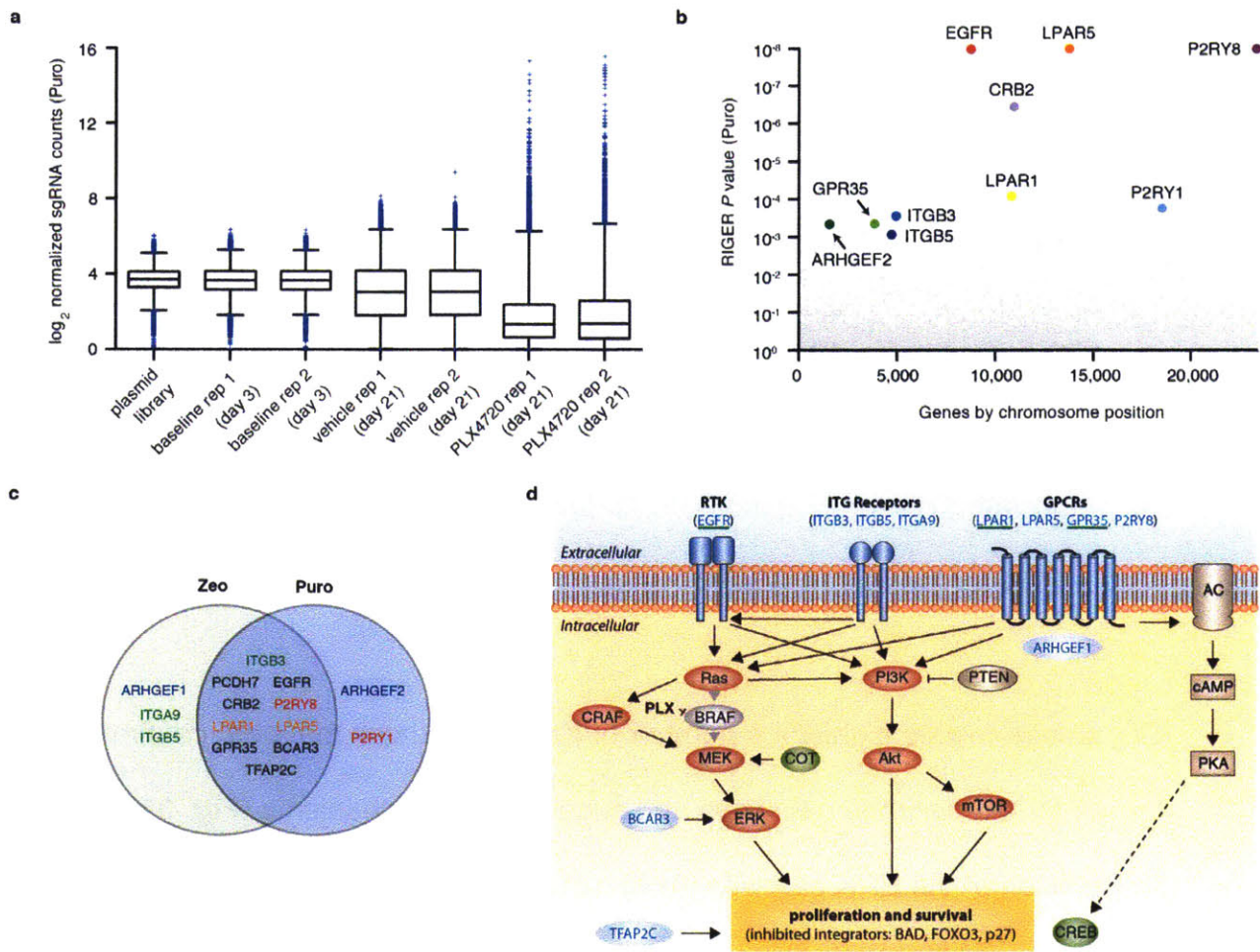
Figure 31 | Genome-scale lentiviral screen using Puromycin-resistant SAM sgRNA library. a, Box plot showing the distribution of sgRNA frequencies at different time points post lentiviral transduction with the Puromycin library, after treatment with DMSO vehicle

71

or PLX-4720. Two infection replicates are shown. **b,** Identification of top candidate genes using the RIGER *P* value analysis (KS method) based on the average of both infection replicates. Genes are organized by positions within chromosomes. **c,** Overlap between the top 20 hits from the Zeo and Puro screens. Genes belonging to the same family are indicated by the same color. There is a 50% overlap between the top hits of each screen as shown in the intersection of the Venn diagram. **d,** Relevant signaling pathways in BRAF inhibitor resistance. Reactivation of the Ras-ERK pathway as well as the parallel PI3K-Akt pathway have previously been implicated as two alternative resistance mechanisms to BRAF inhibitors[65-70]. Both pathways have been described as stimulating proliferation and survival[71]. BAD, FOXO and p27 are common inhibited downstream targets[71]. Recently, stimulation of the cAMP - CREB pathway by GPCRs has been described as a potential additional resistance mechanism[72]. Top candidates from our screen are indicated in blue and putative connections to all three pathways are shown[73-75]. Candidates previously validated to mediate PLX-4720 resistance are underlined in green[67,72]. COT and CREB are independently validated mediators of resistance[65,72].

In addition, for the top 10 shared hits between zeo and puro screens, the fraction of effectively enriched guides per gene (present in the top 5% of all guides) was very high with 97% for zeo and 81% for puro (89% ± 10.7% overall, compared to 78% ± 27% for the top 10 GECKO hits).

Our screen results highlight a number of gene candidates that both confirm known PLX-resistance pathways and suggest new mechanisms (**Fig. 31**). First, reactivation of the ERK pathway is one of the main known resistance mechanisms[65,66], and two of our screening hits, *BCAR3* and *EGFR*, likely modulate downstream and upstream nodes of this pathway, respectively[67,73]. *EGFR* has been previously validated as a mediator of resistance to PLX through

PI3K and AKT, in addition to ERK[67,68]. These two pathways are thought to be alternative routes of PLX resistance[66,69,70]. Furthermore, four out of the top 10 hits from our screen belong to the family of G protein-coupled receptors (GPCRs: GPR35, LPAR1, LPAR5, and P2RY8), which emerged as the top-ranked protein class conferring resistance to multiple MAP kinase inhibitors in melanoma cells in a recent screen using cDNA overexpression[72]. GPCRs signal through multiple downstream pathways including ERK, PI3K, as well as cAMP and PKA[76,77]. The final class of protein candidates from our screen belong to the ITG receptor family, which is thought to interact with RTK and activate both ERK and PI3K pathways[78,79].

Validation of screening hits

To verify the top hits from the PLX-resistance screen, we individually validated 13 genes (top 10 hits from Zeo screen, as well as additional genes belonging to the top 10 shared hits between Zeo and Puro screens). All sgRNAs used to target these 13 genes from the screen conferred PLX resistance when individually expressed in A375 along with SAM (**Fig. 32 and 33**). We also verified that SAM was able to facilitate robust increase in target transcript and protein levels (**Fig. 32**). Since 5 of our top candidates from the pooled SAM screen overlapped with hits from a previously conducted arrayed cDNA screen (**Fig. 33**), we compared the relative efficacy of cDNA overexpression with SAM-mediated transcription

73

activation. Interestingly, for these 5 targets, SAM led to at least similar levels of PLX resistance when compared with corresponding cDNA overexpression conditions, despite cDNA leading to higher transcript levels. Furthermore, we found that, for most genes, the levels of PLX resistance mediated by all three sgRNAs were comparable (**Fig. 33**).
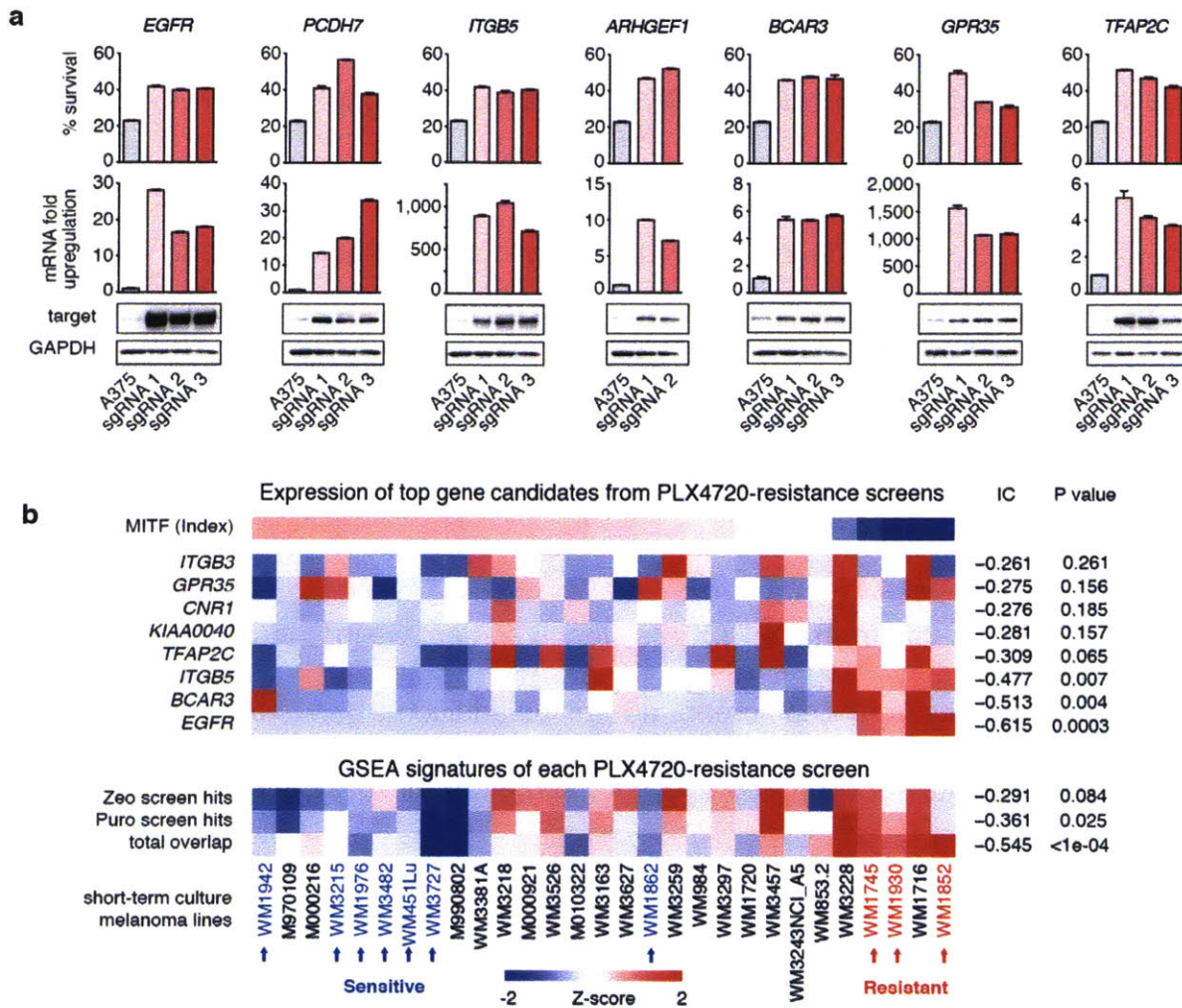
Figure 32 I Validation of top hits from genome-scale gene activation screen for PLX-4720 resistance mediators. a, Comparison of PLX-4720 resistance, transcription activation and protein upregulation in A375 cells for top screening hits. **b,** Expression levels of top hits and screen signatures are elevated in the resistant state of short-term BRAF^{V600} melanoma cultures (see Methods for signature generation). The subset of

samples which were previously tested for PLX-4720 sensitivity and resistance are indicated by blue and red arrows respectively[80]. IC: Information Coefficient. All values are mean ± SEM with n = 3
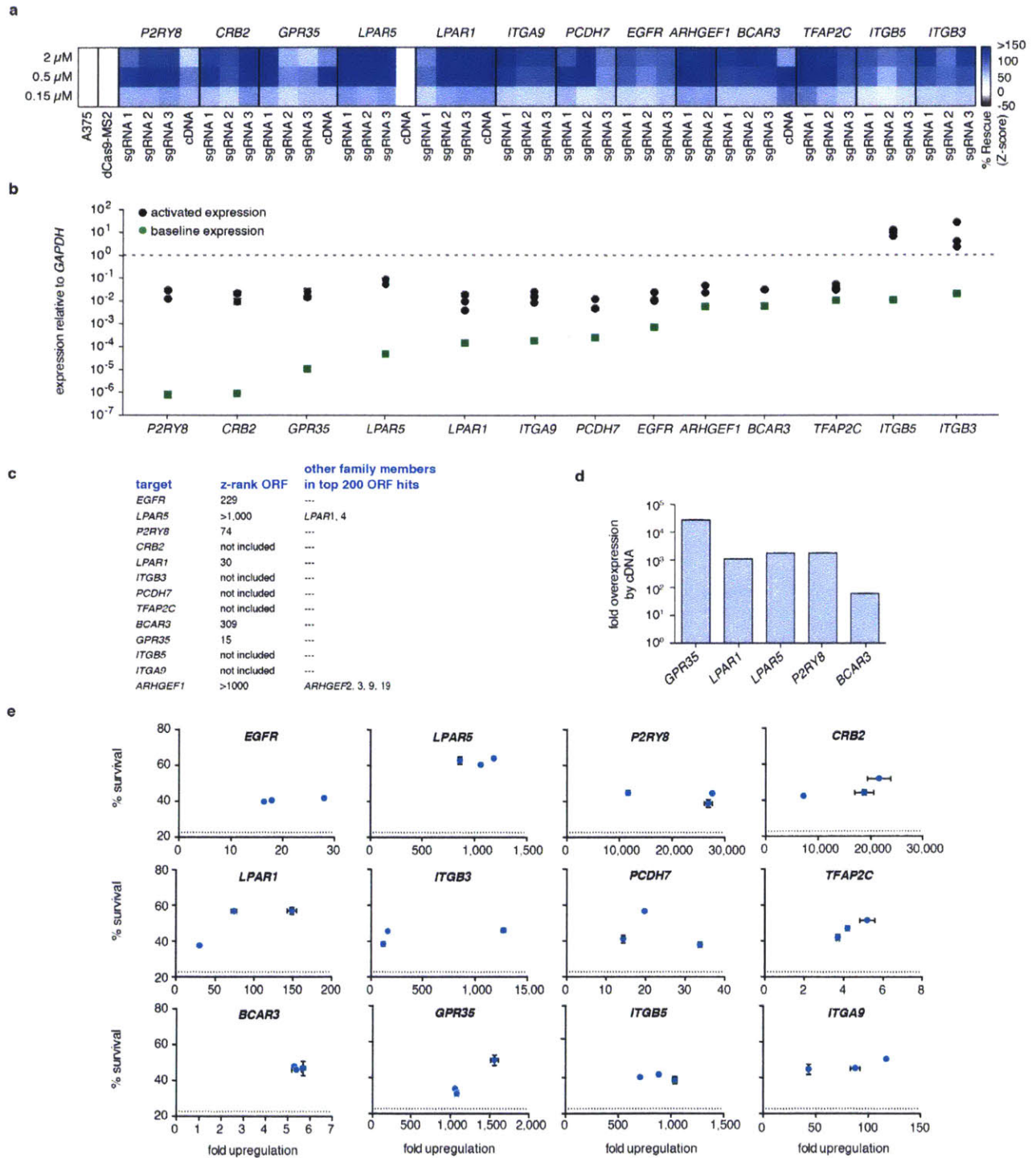
Figure 33 | Individual validation of PLX-4720 resistance mediation by top screen hits. a, Validation of the top 10 Zeo screen hits and the top 10 shared hits (13 genes total). Every gene was independently activated by all three guides from the screen and tested for the ability to increase survival of A375 cells treated with three different

concentrations of PLX-4720 (2µM, 0.5µM and 0.15µM). The z-score based on the % increase in survival relative to control (A375 cells transduced with dCas9-VP64 and MS2-p65-HSF1 alone) is shown for each guide and PLX-4720 concentration. Five cDNAs available from a previous large-scale gain-of-function PLX-4720 resistance screen were also included[72]. Every guide for each top hit mediates significant PLX-4720 resistance. **b,** The same panel of top hits exhibits a large range of basal expression levels and is effectively activated by all guides. The expression level relative to the housekeeping gene *GAPDH* is shown both at baseline as well as after activation by each individual guide. **c,** Ranks of the validated set of genes in the previous ORF screen. Six genes were not part of the cDNA library, five hits are shared (present in the top 3%) and only *LPAR5* and *ARHGEF1* were present but not highly ranked. Both of these genes had highly ranked members of the same family. **d,** Levels of overexpression from the five tested cDNA constructs. Transcript levels were higher for these five cDNAs than those mediated by SAM for the same genes. **e,** Correlation of survival at 2µM PLX-4720 treatment and transcript upregulation achieved by individual guides. For most genes (9 out of 12 shown), the percent survival is very similar across transcript levels achieved by all three guides. Dotted lines indicate control survival.

In addition to validating our top screening hits through individual sgRNA or cDNA overexpression, we analyzed the expression profile of our screening hits using four different collections of gene expression data (CCLE: 27 BRAF^{V600}-mutant melanoma cell lines; TCGA: collection of 113 primary and metastatic BRAF^{V600} melanoma samples from The Cancer Genome Atlas; short-term melanoma: gene expression profiles from 29 melanoma short-term cultures; and pre/post treatment: paired pre- and post-treatment melanoma biopsies from patients). As shown previously[80], a distinct transcriptional state defines BRAF-inhibition sensitivity/resistance where sensitive and resistant states are described by activation of endogenous MITF/associated markers (e.g. PMEL) and NF-κB-

pathway activity/associated markers (e.g. AXL), respectively (**Fig. 32** and **Fig. 34**). Based on short-term melanoma data, we found that expression of top genes from the PLX-resistance screen were significantly enriched in the resistant state. Correspondingly, a gene expression signature derived from the top overlapping zeo and puro screen hits (**Fig. 32**, total overlap) was correlated with BRAF-inhibitor resistance (p < 0.0001).
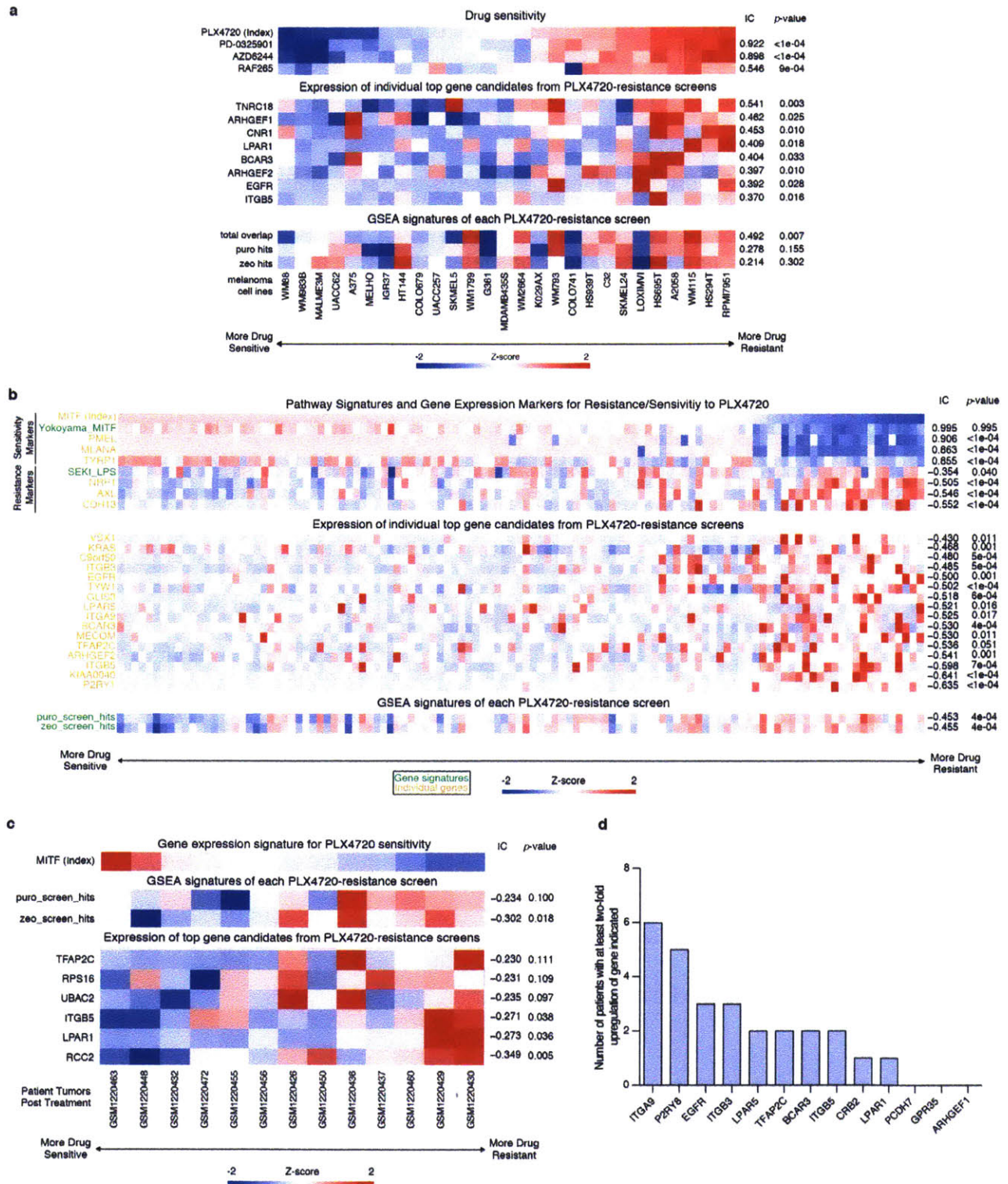
Figure 34 I Expression of top hits and screen signatures are elevated in PLX-4720 **resistant melanoma cell lines and patient samples**. **a,** Heat map showing sensitivity to different drugs (top), expression of SAM top screen hits (middle), and SAM screen

signature scores (bottom; see Online Methods for signature generation) in Cancer Cell Line Encyclopedia cell lines[81]. Drug sensitivities are measured as Activity Areas (AA). The melanoma cell lines are sorted by PLX-4720 drug sensitivity. RAF inhibitors: PLX-4720 and RAF265; MEK inhibitors: AZD6244 and PD-0325901. **b,** Heat map showing expression of gene/signature markers for BRAF-inhibitor sensitivity (top), expression of SAM top screen hits (middle) and screen signature scores (bottom) in different BRAF^{V600} patient melanoma samples (primary or metastatic) from The Cancer Genome Atlas. **c,** Heat map showing MITF expression (top), screen signature scores (middle), and expression of SAM top screen hits (bottom) in different BRAF^{V600E} patient melanoma biopsies post-treatment with BRAF inhibitors[82]. **d,** Bar chart showing the number of patients from (c) with at least a two-fold change (post/pre treatment) in gene expression of the top PLX-4720 screen hits in the post-treatment samples. All associations are measured using the information coefficient (IC) between the index and each of the features and P values are determined using a permutation test. All heat maps show z-scores.

We additionally investigated the expression of the top hits in the CCLE dataset. The gene expression of individual top hits as well as the signature from the activation screen are enriched and significantly associated with resistance to BRAF-inhibition (**Fig. 34**; p = 0.007 for overlapping hits from Zeo and Puro screens). Further analysis performed using the TCGA and pre/post treatment data set also revealed similar correlations (**Fig. 34**).

Discussion

In summary, we have taken a structure-guided approach to design a dCas9-based transcription activation system for achieving robust, single sgRNA-

mediated gene up-regulation. By engineering the sgRNA to incorporate protein-interacting aptamers, we were able to assemble a synthetic transcription activation complex consisting of multiple distinct effector domains modeled after natural transcription activation processes. Here we have shown that the SAM system is robust, specific, and can be combined with a compact pooled library of sgRNAs to facilitate genome-scale gain-of-function screening. SAM-mediated screens can exhibit a high degree of consistency and validation, with the fraction of effectively enriched guides per gene hit being greater than 80%, and validation for the top 10 hits being 100%.

Additional developments of the SAM system may be able to take advantage of the modularity and customizability of the sgRNA scaffold to establish a series of sgRNA scaffolds with different aptamers for recruiting distinct types of effectors in an orthogonal manner. For instance, replacement of the MS2 stem-loops with PP7-interacting stem-loops may be used to recruit repressive elements, potentially enabling multiplexed bidirectional transcriptional control. Although we have taken initial steps in this study toward defining selection rules for potent sgRNAs, future studies will be able to reveal additional selection criteria that are critical for guide efficacy, such as sequence-intrinsic properties.

Future engineering of the Cas9 complex based on crystal structures[16,43] will further expand the Cas9 toolbox[51,83]. Applications of the Cas9 transcription

activation complex, either in the context of individual gene perturbation or genome-scale gene activation libraries, will enable the dissection of many types of genetic elements, ranging from protein-coding genes to non-coding lincRNA elements. Furthermore, combining SAM with genome editing or other dCas9-mediated gene modulation systems[84,85] in both positive as well as negative selection screens will constitute powerful approaches for studying gene interactions in diverse biological processes.

Chapter 4:

Orthogonal gene control with a catalytically active Cas9 nuclease

Preface
This chapter is adapted from:

to fit the format of this thesis.

A CRISPR-based tool to knock out and activate genes in the same cell would provide a powerful method to manipulate biological processes. One system that holds promise as a targeted orthogonal gene regulator is the RNA-guided Cas9 nuclease from the microbial CRISPR (clustered regularly interspaced short palindromic repeats)-Cas system[86,87]. Once bound to its target DNA, the active sgRNA-Cas9 complex can induce a double stranded break (DSB) via its RuvC and HNH domains. By mutating these nuclease domains, Cas9 can be made catalytically inactive[10,11] and repurposed for genetic perturbation beyond DNA editing. For example, this 'dead' Cas9 (dCas9) has been combined with protein

83

domains that suppress or activate gene expression[45-48,88] or change the epigenetic state of a target locus[89].

The simplest CRISPR-based orthogonal gene control system would function with only catalytically active wild type Cas9. However, previous work indicated that it would be unlikely that an active Cas9 could be retargeted for activation[87]. Here we show that by reducing the length of the RNA targeting sequence to 14-15 nucleotides (nt), and by adding MS2 binding loops into the sgRNA backbone[90], sgRNAs can guide catalytically active Cas9 to activate transcription without inducing DSBs **(Fig. 35)**. After systematically studying how the design of these 'dead sgRNAs' (dRNAs) influences transcriptional activation and indel formation, and characterizing off-target transcriptional perturbations, we show that this system can simultaneously knock out and upregulate target genes in melanoma cells. These data demonstrate that sgRNAs can be engineered to exert transcriptional control using active Cas9, and that a single active Cas9 protein can be used for orthogonal gene control in mammalian cells.
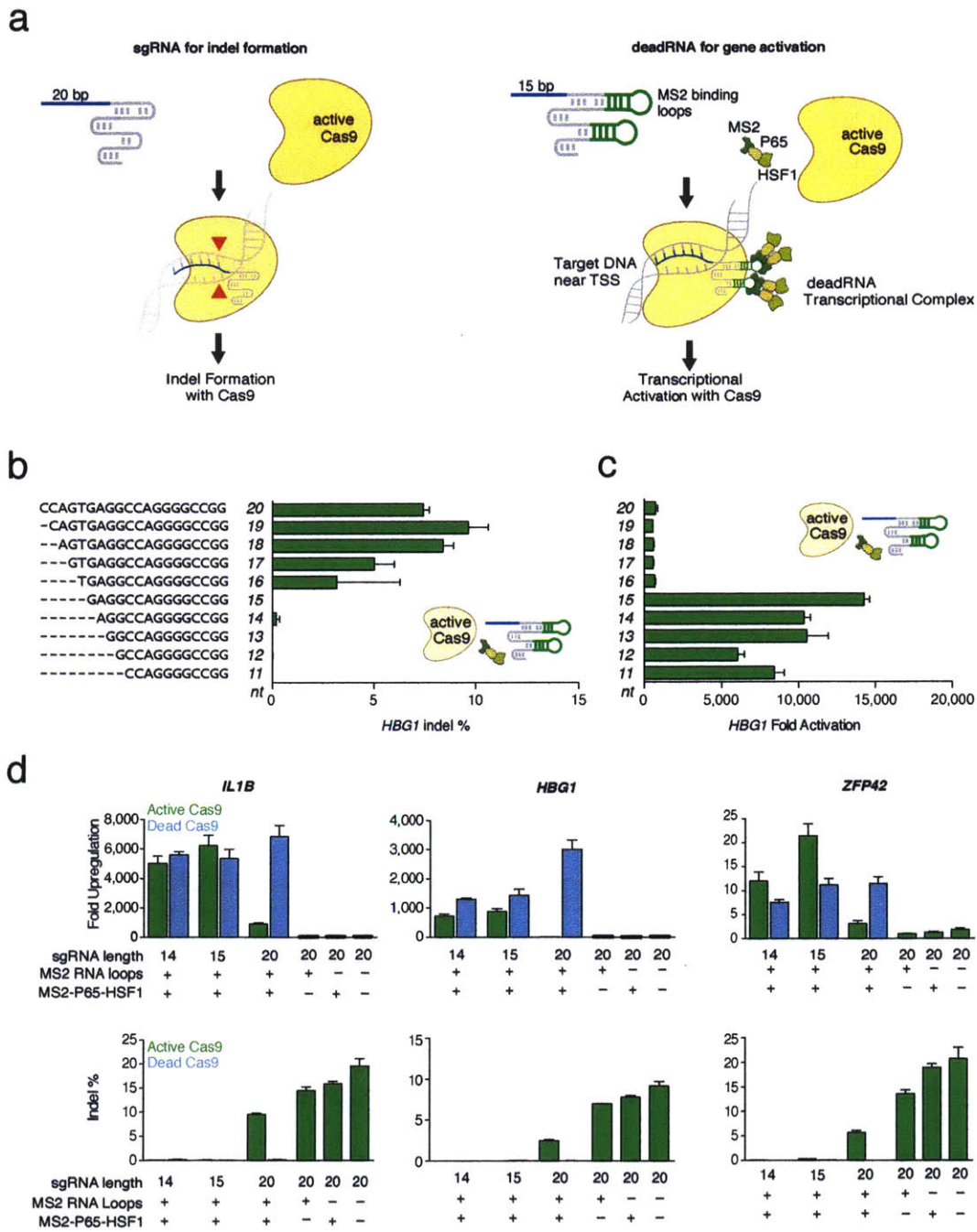
Figure 35 | deadRNAs (dRNAs) can mediate robust gene activation using an active SpCas9. **(a)** dRNA-mediated gene activation. 15bp dRNA with MS2 loops on the scaffold can bind the MS2-P65-HSF1 (MPH) transcriptional activation complex and activate gene expression without inducing Cas9-mediated DNA indel formation. **(b)** The length of the RNA targeting sequence was varied from 11nt to 20nt. *HBG1* indel

85

frequencies were quantified with **(c)** *HBG1* mRNA levels (normalized to *GAPDH*, and compared to cells transfected with GFP plasmid). No indel formation was observed when sgRNAs had less than 16bp of homology to target DNA. In all cases, guides were designed with MS2 binding loops in the tetraloops and stemloop two[90], and were co-transfected with active Cas9 and the MPH transcriptional activation complex. **(d)** Three dRNAs targeting the promoter regions of *IL1B, HBG1, and ZFP42* were tested for activation and indel formation. dRNAs with 14bp or 15bp of homology to target DNA did not induce detectable indel formation. dRNAs co-transfected with Cas9 and MPH activated transcription to a similar extent as 20nt sgRNA-MS2 co-transfected with dCas9 and MPH. (In all cases, mean +/- S.E.M. is plotted. N=2-3 replicates / group).

Changes in sgRNA structure and mismatches between the sgRNA targeting sequence and DNA can prevent Cas9-mediated DNA cleavage[51,91-93]. However, it is unknown whether these modified sgRNAs still allow binding of Cas9 to the DNA target. To test this, we designed sgRNAs with two structural characteristics. First, we added two aptamers that selectively bind dimerized MS2 bacteriophage coat proteins into to the tetraloop and stem loop two of the sgRNA as previously described (sgRNA-MS2)[90]. Second, we shortened the length of the sgRNA guide sequence from 20nt to 11nt. We reasoned that these changes could result in a dRNA that would still enable binding of Cas9 while preventing nuclease activity **(Figure 35)**.

We transfected eighty sgRNA-MS2s targeting four DNA sequences within 200 bp of the transcriptional start site of human hemoglobin 1 (*HBG1*) together with active Cas9 and the MS2-P65-HSF1 (MPH) activation complex. The MPH complex has previously been reported to mediate efficient target upregulation by

binding to MS2 loops in the sgRNA[90]. We observed that guides from 20nt to 16nt

resulted in indel formation, whereas shorter guides (11nt to 15nt) did not show

detectable levels of indel formation in most cases. Notably, guides truncated to

11-15nt of complementarity to the target DNA were able to increase *HBG1*

mRNA expression by as much as 10,000 fold (**Figure 35 and 36**).
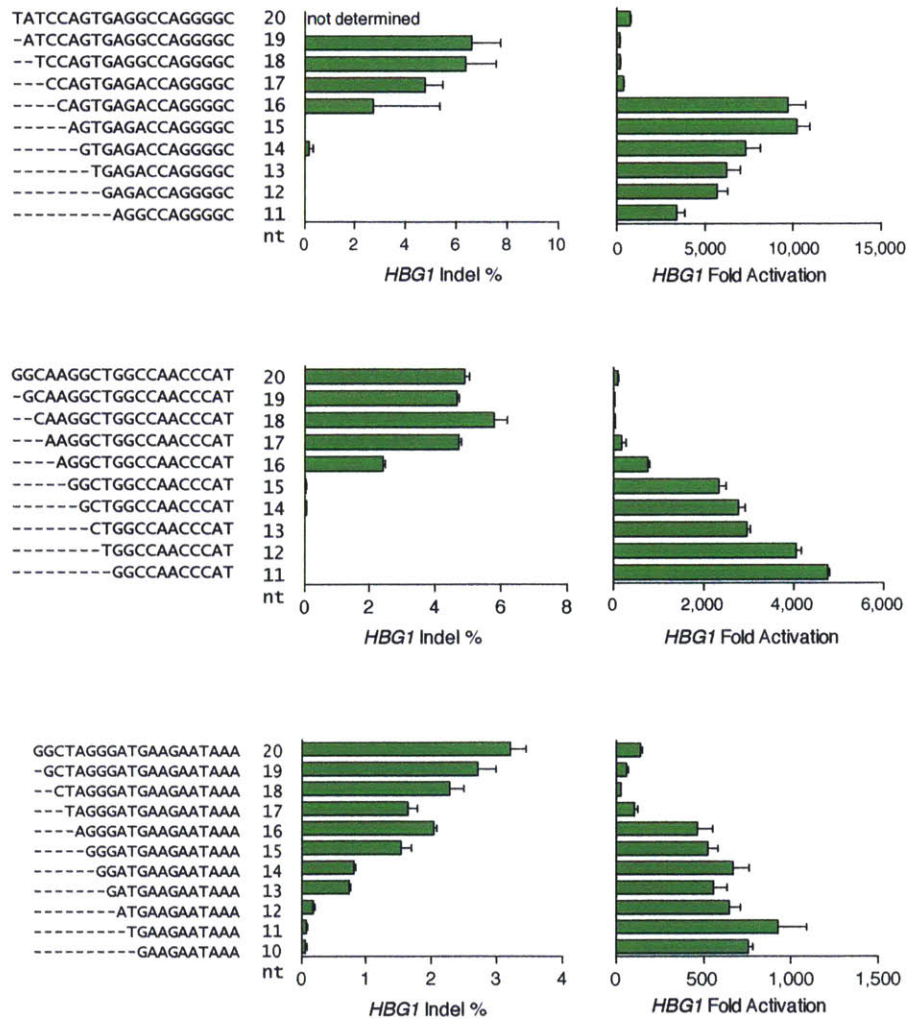
Figure 36 | dRNAs can mediate robust gene activation using an active Cas9. Three
different dRNAs targeting the *HBG1* promoter region were designed. The length of the
RNA targeting sequence was varied from 11nt to 20nt. *HBG1* mRNA (normalized to
GAPDH, and compared to cells transfected with GFP plasmid) was quantified, as well

87

HBG1 indel frequency. In all cases, guides were designed with MS2 binding loops in the tetraloops and stem loop two[90], and were co-transfected with active Cas9 and the MPH transcriptional activation complex. Average +/- SEM is plotted, N=2-3 replicates group.

We proceeded to investigate the efficiency of 14 and 15nt dRNAs at three loci. Fourteen and fifteen nt dRNAs, when co-transfected into HEK293FT cells with active Cas9 and the MPH complex, increased target mRNA expression of all three human genes (*HBG1*, Interleukin 1B (*IL1B*), and Zinc Finger Protease 42 (*ZFP42*)) without inducing significant indel formation (**Fig. 1d**). Notably, dRNA activation was comparable to the recently reported system using dCas9 in combination with a 20nt sgRNA-MS2[90]. At all three loci 20nt sgRNAs cut target DNA and did not activate gene expression when combined with active Cas9. This was true for sgRNAs with and without the MS2 binding loops (**Fig. 35**). Taken together, these data demonstrate that dRNAs can activate gene expression without forming indels at targeted DNA using an active Cas9 with comparable efficiency to the current dCas9 system.

Biological studies utilizing activators will require specific target upregulation. Therefore, it is important to understand the specificity of Cas9 mediated gene activation. In addition, recent work has demonstrated that a single sgRNA can bind many sites in the genome, but the relationship between binding and transcriptional control is not clear[64]. Specificity may also change when the sgRNA is shortened to 15 bp. To assess the difference in specificity between 20nt

sgRNA-MS2 and 15nt dRNAs we compared whole transcriptome mRNA levels in

HEK293FT cells. Cells were co-transfected with dCas9, the MPH complex, and a

20nt activator sgRNA-MS2, or active Cas9, the MPH complex, and 15nt dRNA

targeting the same sequence in the human *HBG1/2* promoter. We previously

determined that *HBG1/2* upregulation induces limited downstream effects that

could confound our analysis in HEK293FT cells. RNA-seq results showed that

both the sgRNA/dCas9 and dRNA systems significantly activated *HBG1/2* only,

demonstrating that dRNAs can specifically upregulate target genes (**Fig. 37**). We

next performed off-target analysis on a second 15nt dRNA and 20nt sgRNA

targeting the same *HBG1/2* promoter. Surprisingly, we found a significant number

of perturbed transcripts for both the 15nt and 20nt guide RNAs (**Fig. 37**).

a

GGCTAGGGATGAAGAATAAA

GGGATGAAGAATAAA

b

CCAGTGAGGCCAGGGGCCGG

GAGGCCAGGGGCCGG

c

Cas9 + MS2-P65-HSF1 A375 cells

dRNA-MS2 zeo lentivirus

sgRNA puro lentivirus

7 days
puro and zeo selection

orthogonal gene control

d

LPAR5 dRNA	+	+	+
MED12 sgRNA	–	+	–
TADA2B sgRNA	–	–	+

LPAR5 dRNA	+	+	+
MED12 sgRNA	–	+	–
TADA2B sgRNA	–	–	+

e

MED12
TADA2B

LPAR5 dRNA	–	+	–	+
MED12 sgRNA	+	+	–	–
TADA2B sgRNA	–	–	+	+
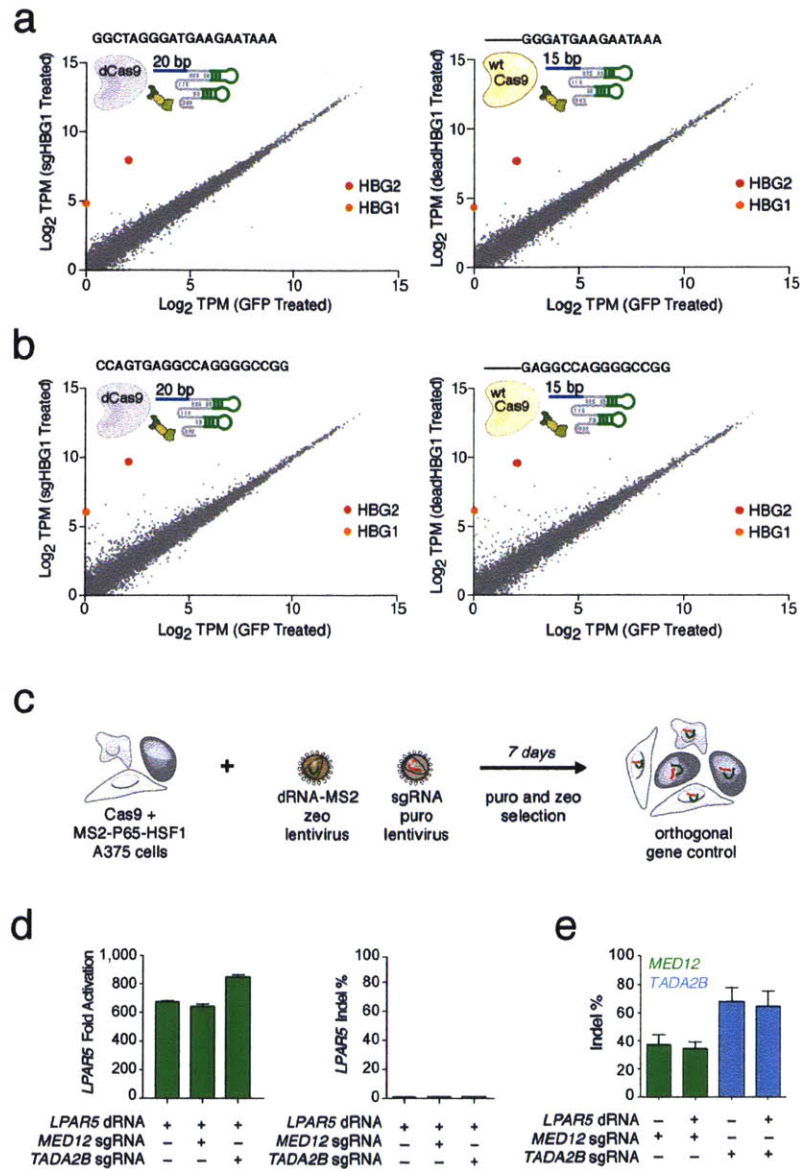
Figure 37 | dRNAs can mediate orthogonal gene control in human cells.
Sequences targeted to the *HBG1/2* promoter were tested for off-target transcriptional activation using RNA-seq. 20nt sgRNAs with MS2 binding loops were co-transfected with dCas9 and the MPH activation complex. These were compared to dRNAs co-delivered with active Cas9 and the MPH activation complex. Both systems showed similar off-target profiles. **(a)** Zero significantly upregulated genes apart from *HBG1/2* were observed for both the 20nt/dCas9 and dRNA/Cas9 treated cells. **(b)** A second

guide showed 55 significantly upregulated genes apart from HBG1/2 for the 20nt/dCas9-treated cells, while 31 significantly upregulated genes were measured for dRNA-treated cells. (In all cases, N=3 replicates / group). **(c)** Orthogonal gene control in melanoma A375 cells expressing an active Cas9 and the MS2-P65-HSF1 fusion protein. Cells were transduced with lentivirus containing a dRNA targeting one gene and an sgRNA targeting a second gene. Selected cells were subsequently treated with BRAF-inhibitor PLX4720 and their survival was quantified. **(d)** Activation and indel % were measured for individually and orthogonally controlled genes. *Left: LPAR5* transcriptional upregulation mediated by dRNA was robust in the presence and absence of sgRNAs targeting *MED12* or *TADA2B. Right: LPAR5* indel formation was 0.85% at the dRNA target site. **(e)** Robust indel formation was detected at DNA sites targeted by *MED12* and *TADA2B* sgRNAs alone and when delivered together with a dRNA targeting *LPAR5*.

We expanded this analysis by studying whole transcriptome analysis on ten additional sgRNAs targeting the proximal promoter of *HBG1/2* (**Fig. 38**). Overall, four out of twelve 20bp guides exhibited high specificity (<3 significant genome-wide off-targets), confirming activators can be specific. Notably, a previously published algorithm that predicts off-target indels did not correlate with the number of non-targeted transcripts that were altered by each guide (R = 0.12, p = 0.7)[91].
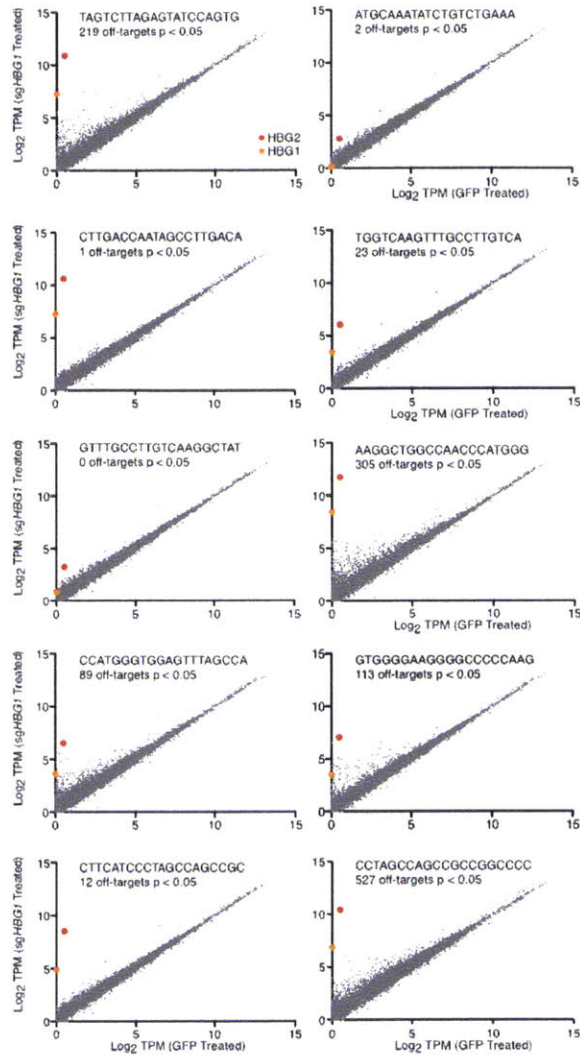
Figure 38 | Transcriptome-wide mRNA profiles for ten different sgRNAs targeting *HBG1/2*. 20nt sgRNAs with MS2 binding loops were co-transfected with dCas9 and the MPH complex. A previously published nuclease OT score[91] did not significantly correlate with guide specificity. (In all cases, N=3 replicates / group

Finally, we aimed to test whether dRNAs in combination with sgRNAs could

mediate orthogonal gene control (activation and knockout) using only active

Cas9. We previously used CRISPR-Cas9 loss-of-function (LOF)[39] and gain-of-

function (GOF)[90] screens to identify genetic modifiers that promote resistance of

A375 melanoma cells to the BRAF inhibitor PLX-4720. We performed orthogonal

gene regulation with hits selected from these screens. We first transduced and

selected A375 cells with two lentiviral constructs encoding active Cas9 and the

MPH complex, respectively (**Fig. 37**). We then transduced these cells with

lentiviral constructs encoding a dRNA targeting *LPAR5* for activation and / or

sgRNAs targeting *MED12* or *TADA2B* for gene knockout. *LPAR5* mRNA

expression increased over 600-fold when cells were treated with dRNA targeting

LPAR5, even when combined with sgRNAs targeting other genes. In all

conditions *LPAR5* indels less or equal to 0.85% were detected (**Fig. 37**). Indel

rates of 0.6% and 0.05% at targeted loci were also measured after cells were

treated with two additional dRNAs targeting *EGFR* and *ITGA9*, respectively. By

contrast, the loci targeted by *MED12* and *TADA2B* showed robust indel formation

even in the orthogonal conditions **(Fig. 37)**.

dCas9 proteins have been engineered for diverse functions including gene

suppression, gene activation, and epigenetic modification[45-48,88,89]. Here we have

demonstrated that guide RNAs can be engineered to bind target DNA and

successfully recruit a transcriptional activation complex, without inducing

measurable indel formation using an active Cas9 nuclease. These dRNAs can be

designed to specifically and potently upregulate gene expression. We anticipate

that dRNAs may also be designed to perform other genetic functions, for

example, epigenetic modification or mRNA suppression.

dRNAs can be used to simplify orthogonal gene control experiments, as

nuclease-mediated gene knockout and transcriptional activation can be achieved

in the same cell population with one Cas9 protein. By activating *LPAR5*, a gene

previously described to promote drug resistance when upregulated[90], and

knocking out known tumor suppressors[39], we showed that orthogonal gene

modifications can be performed in a biologically meaningful context. Streamlining

orthogonal gene control is particularly important for *in vivo* experiments, since

concurrently delivering two distinct Cas9 complexes to the same cell is

challenging. Given the fact that mice have been engineered to constitutively and

inducibly express Cas9[94,95], we anticipate that dRNAs can be utilized for *in vivo*

orthogonal experiments, as the MPH activator complex combined with a dRNA

and sgRNA fits within the 4.7kb packaging limit of AAV vectors.

Finally, our data align with previously reported work[96] that show sgRNAs with

seventeen or more nucleotides reliably and efficiently cut target DNA. It is

interesting that by reducing the guide length down to 15 nucleotides, we abrogate

indel formation, but do not abrogate 'functional binding'. This may point to a more

fundamental mechanism, whereby the interactions between the protein, sgRNA,

and DNA dictate whether the protein cuts the target DNA.

Chapter 5:

Materials and Methods

Neuro 2a culture and experiments

Neuro 2a cells (Sigma-Aldrich) were grown in media containing a 1:1 ratio of

OptiMEM (Life Technologies) to high-glucose DMEM with GlutaMax and Sodium

Pyruvate (Life Technologies) supplemented with 5% HyClone heat-inactivated

FBS (Thermo Scientific), 1% penicillin/streptomycin (Life Technologies), and

passaged at 1:5 every 2 days. 120,000 cells were plated in each well of a 24-well

plate 18-20 h prior to transfection. 1 h before transfection, media was changed to

DMEM supplemented with 5% HyClone heat-inactivated FBS and 1%

penicillin/streptomycin. Cells were transfected with 1.0 µg total of construct DNA

(at equimolar ratios) per well with 1.5 μL of GenJet (SignaGen Laboratories)

transfection reagent according to the manufacturer's instructions. Media was

exchanged 24 h and 44 h post-transfection and light stimulation was started at 48

h. Stimulation parameters were: 5 mW cm^{-2}, 466 nm, 7 % duty cycle (1 s light

pulse 0.067 Hz) for 24 h unless indicated otherwise in figure legends. RNA was

extracted using the Rneasy kit (Qiagen) according to manufacturer's instructions

and 1 µg of RNA per sample was reverse-transcribed using qScript (Quanta

Biosystems). Relative mRNA levels were measured by quantitative real-time

PCR (qRT-PCR) using TaqMan probes specific for the targeted gene as well as

GAPDH as an endogenous control (Life Technologies). $\Delta\Delta$Ct analysis was used

to obtain fold-changes relative to negative controls transduced with GFP only and subjected to light stimulation. Toxicity experiments were conducted using the LIVE/DEAD assay kit (Life Technologies) according to manufacturer's protocol.

AAV vector production

The ssDNA-based genome of AAV is less susceptible to recombination, thus providing an advantage over RNA-based lentiviral vectors for the packaging and delivery of highly repetitive TALE sequences[32]. 293FT cells (Life Technologies) were grown in antibiotic-free D10 media (DMEM high glucose with GlutaMax and Sodium Pyruvate, 10% heat-inactivated Hyclone FBS, and 1% 1M HEPES) and passaged daily at 1:2-2.5. The total number of passages was kept below 10 and cells were never grown beyond 85% confluence. The day before transfection, $1x10^6$ cells in 21.5 mL of D10 media were plated onto 15 cm dishes and incubated for 18-22 hours or until ~80% confluence. For use as a transfection reagent, 1 mg/mL of PEI "Max" (Polysciences) was dissolved in water and the pH of the solution was adjusted to 7.1. For AAV production, 10.4 μg of pDF6 helper plasmid, 8.7 μg of pAAV1 serotype packaging vector, and 5.2 μg of pAAV vector carrying the gene of interest were added to 434 μL of serum-free DMEM and 130 μL of PEI "Max" solution was added to the DMEM-diluted DNA mixture. The DNA/DMEM/PEI cocktail was vortexed briefly and incubated at room temperature for 15 min. After incubation, the transfection mixture was added to 22 mL of complete media, vortexed briefly, and used to replace the media for a 15 cm dish

of 293FT cells. For supernatant production, transfection supernatant was harvested at 48 h, filtered through a 0.45 μm PVDF filter (Millipore), distributed into aliquots, and frozen for storage at -80°C.

Primary cortical neuron culture

Dissociated cortical neurons were prepared from C57BL/6N mouse embryos on E16 (Charles River Labs). Cortical tissue was dissected in ice-cold HBSS – (50 mL 10x HBSS, 435 mL dH$_2$O, 0.3 M HEPES pH 7.3, and 1% penicillin/streptomycin). Cortical tissue was washed 3X with 20 mL of ice-cold HBSS and then digested at 37°C for 20 min in 8 mL of HBSS with 240 μL of 2.5% trypsin (Life Technologies). Cortices were then washed 3 times with 20 mL of warm HBSS containing 1 mL FBS. Cortices were gently triturated in 2 ml of HBSS and plated at 150,000 cells/well in poly-D-lysine coated 24-well plates (BD Biosciences). Neurons were maintained in Neurobasal media (Life Technologies), supplemented with 1X B27 (Life Technologies), GlutaMax (Life Technologies) and 1% penicillin/streptomycin.

Primary neuron transduction and light stimulation experiments

Primary cortical neurons were transduced with 250 μL of AAV1 supernatant on DIV 5. The media and supernatant were replaced with regular complete neurobasal the following day. Neurobasal was exchanged with Minimal Essential Medium (Life Technologies) containing 1X B27, GlutaMax (Life Technologies)

and 1% penicillin/streptomycin 6 days after AAV transduction to prevent

formation of phototoxic products from HEPES and riboflavin contained in

Neurobasal during light stimulation. For co-transduction of primary neurons with

two AAV vectors, the co-delivery efficiency was >80%, with individual

components having transduction efficiencies of 83-92%.

Light stimulation was started 6 days after AAV transduction (DIV 11) with an

intensity of 5 mW/cm^2, duty cycle of 0.8% (250 ms pulses at 0.033Hz or 500 ms

pulses at 0.016Hz), 466 nm blue light for 24 h unless indicated otherwise in figure

legends. RNA extraction and reverse transcription were performed using the

Cells-to-Ct kit according to the manufacturers instructions (Life Technologies).

Relative mRNA levels were measured by quantitative real-time PCR (qRT-PCR)

using TaqMan probes as described above for Neuro 2a cells.

Immunohistochemistry of primary neurons

For immunohistochemistry of primary neurons, cells were plated on poly-D-

lysine/laminin coated coverslips (BD Biosciences) after harvesting. AAV1

transductions were performed as described above. Neurons were fixed 7 days

post-transduction with 4% paraformaldehyde (Sigma Aldrich) for 15 min at room

temperature. Blocking and permeabilization were performed with 10% normal

goat serum (Life Technologies) and 0.5% Triton-X100 (Sigma-Aldrich) in DPBS

(Life Technologies) for 1 h at room temperature. Neurons were incubated with

primary antibodies overnight at 4°C, washed 3X with DPBS and incubated with secondary antibodies for 90 min at RT. Coverslips were finally mounted using Prolong Gold Antifade Reagent with DAPI (Life Technologies) and imaged on an Axio Scope A.1 (Zeiss) with an X-Cite 120Q light source (Lumen Dynamics). Images were acquired using an AxioCam MRm camera and AxioVision 4.8.2.

Western Blots

For preparation of total protein lysates, primary cortical neurons were harvested after light stimulation in ice-cold lysis buffer (RIPA, Cell Signaling; 0.1% SDS, Sigma-Aldrich; and cOmplete ULTRA protease inhibitor mix, Roche Applied Science). Cell lysates were sonicated for 5 min at 'M' setting with the Bioruptor water bath sonicator (Diagenode) and centrifuged at 21,000 x g for 10 min at 4°C. Protein concentration was determined using the RC DC protein assay (Bio-Rad). 30-40 µg of total protein per lane was separated under non-reducing conditions on 4-15% Tris-HCl gels (Bio-Rad) along with Precision Plus Protein Dual Color Standard (Bio-Rad). After wet electrotransfer to polyvinylidene difluoride membranes (Millipore) and membrane blocking for 45 min in 5% BLOT-QuickBlocker (Millipore) in Tris-buffered saline (TBS, Bio-Rad), western blots were probed with anti-mGluR2 (Abcam, 1:1.000) and anti-α-tubulin (Sigma-Aldrich 1:20,000) overnight at 4°C, followed by washing and anti-mouse-IgG HRP antibody incubation (Sigma-Aldrich, 1:5,000 – 1:10,000). Detection was performed via ECL Western blot substrate (SuperSignal West Femto Kit, Thermo

Scientific). Blots were imaged with an AlphaImager system (Innotech), and quantified using ImageJ software 1.46r.

Production of concentrated and purified AAV1/2 vectors

Production of concentrated and purified AAV for stereotactic injection *in vivo* was performed using the same initial steps outlined above for production of AAV1 supernatant. However, for transfection, equal ratios of AAV1 and AAV2 serotype plasmids were used instead of AAV1 alone. Five 15cm plates were transfected per construct and cells were harvested with a cell-scraper 48 h post transfection. Purification of AAV1/2 particles was performed using HiTrap heparin affinity columns (GE Healthcare)[97]. We added a second concentration step down to a final volume of 100 μl per construct using an Amicon 500 μl concentration column (100 kDa cutoff, Millipore) to achieve higher viral titers. Titration of AAV was performed by qRT-PCR using a custom Taqman probe for WPRE (Life Technologies). Prior to qRT-PCR, concentrated AAV was treated with DnaseI (New England Biolabs) to achieve a measurement of DnaseI-resistant particles only. Following DnaseI heat-inactivation, the viral envelope was degraded by Proteinase K digestion (New England Biolabs). Viral titer was calculated based on a standard curve with known WPRE copy numbers.

Stereotactic injection of AAV1/2 and optical implant

All animal procedures were approved by the MIT Committee on Animal Care. Adult (10-14 weeks old) male C57BL/6N mice were anaesthetized by intraperitoneal (i.p.) injection of Ketamine/Xylazine (100 mg/kg Ketamine and 10 mg/kg Xylazine) and pre-emptive analgesia was applied (Buprenex, 1 mg/kg, i.p.). Craniotomy was performed according to approved procedures and 1μl of AAV1/2 was injected into ILC at 0.35/1.94/-2.94 (lateral, anterior and inferior coordinates in mm relative to bregma). During the same surgical procedure, an optical cannula with fiber (Doric Lenses) was implanted into ILC unilaterally with the end of the optical fiber located at 0.35/1.94/-2.64 relative to bregma. The cannula was affixed to the skull using Metabond dental cement (Parkell Inc) and Jet denture repair (Lang Dental) to build a stable, supporting cone. The incision was sutured and proper post-operative analgesics were administered for three days following surgery.

Immunohistochemistry on ILC brain sections

Mice were injected with a lethal dose of Ketamine/Xylazine anaesthetic and transcardially perfused with PBS and 4% paraformaldehyde (PFA). Brains were additionally fixed in 4% PFA at 4°C overnight and then transferred to 30% sucrose for cryoprotection overnight at room temperature. Brains were then transferred into Tissue-Tek Optimal Cutting Temperature (OCT) Compound (Sakura Finetek) and frozen at -80°C. 18 μm sections were cut on a cryostat (Leica Biosystems) and mounted on Superfrost Plus glass slides (Thermo

Fischer). Sections were post-fixed with 4% PFA for 15 min, and immunohistochemistry was performed as described for primary neurons above.

Light stimulation and mRNA level analysis in ILC

Neurons at the injection site were efficiently co-transduced by both viruses, with >80% of transduced cells expressing both TALE(*Grm2*)-CIB1 and CRY2PHR-VP64 (**Fig. 8b** and **c**). 8 days post-surgery, awake and freely moving mice were stimulated using a 473 nm laser source (OEM Laser Systems) connected to the optical implant via fiber patch cables and a rotary joint. Stimulation parameters were the same as used on primary neurons: 5 mW (total output), 0.8% duty cycle (500 ms light pulses at 0.016 Hz) for a total of 12 h. Brain tissue from the fiber optic cannula implantation site was analyzed for changes in *Grm2* mRNA.

After the end of light stimulations, mice were euthanized using CO_2 and the prefrontal cortices (PFC) were quickly dissected on ice and incubated in RNA later (Qiagen) at 4°C overnight. 200 μm sections were cut in RNA later at 4°C on a vibratome (Leica Biosystems). Sections were then frozen on a glass coverslide on dry ice and virally transduced ILC was identified under a fluorescent stereomicroscope (Leica M165 FC). A 0.35 mm diameter punch of ILC, located directly ventrally to the termination of the optical fiber tract, was extracted (Harris uni-core, Ted Pella). The brain punch sample was then homogenized using an Rnase-free pellet-pestle grinder (Kimble Chase) in 50 μl Cells-to-Ct RNA lysis

buffer and RNA extraction, reverse transcription and qRT-PCR was performed as described for primary neuron samples.

Chromatin Immunoprecipitation

Neurons or Neuro 2a cells were cultured and transduced or transfected as described above. ChIP samples were prepared as previously described[98] with minor adjustments for the cell number and cell type. Cells were harvested in 24-well format, washed in 96-well format, and transferred to microcentrifuge tubes for lysis. Sample cells were directly lysed by water bath sonication with the Biorupter sonication device for 21 minutes using 30s on/off cycles (Diagenode). qPCR was used to assess enrichment of histone marks at the targeted locus.

Statistical analysis

All experiments were performed with a minimum of two independent biological replicates. Statistical analysis was performed with Prism (GraphPad) using Student's two-tailed t-test when comparing two conditions, ANOVA with Tukey's post-hoc analysis when comparing multiple samples with each other, and ANOVA with Dunnett's post-hoc analysis when comparing multiple samples to the negative control.

Photostimulation Hardware – *in vitro*

In vitro light stimulation experiments were performed using a custom built LED photostimulation device. All electronic elements were mounted on a custom printed circuit board (ExpressPCB). Blue LEDs with peaks 466 nm (model #: YSL-R542B5C-A11, China Young Sun LED Technology; distributed by SparkFun Electronics as 'LED – Super Bright Blue' COM-00529), were arrayed in groups of three aligned with the wells of a Corning 24-well plate. LED current flow was regulated by a 25 mA DynaOhm driver (LEDdymanics #4006-025). Columns of the LED array were addressed by TTL control (Fairchild Semiconductor PN2222BU-ND) via an Arduino UNO microcontroller board. Light output was modulated via pulse width modulation. Light output was measured from a distance of 80 mm above the array utilizing a Thorlabs PM100D power meter and S120VC photodiode detector. In order to provide space for ventilation and to maximize light field uniformity, an 80 mm tall ventilation spacer was placed between the LED array and the 24-well sample plate. Fans (Evercool EC5015M12CA) were mounted along one wall of the spacer unit, while the opposite wall was fabricated with gaps to allow for increased airflow.

Quantification of LIVE/DEAD® assay using ImageJ software.

Images of LIVE/DEAD (Life Technologies) stained cells were captured by fluorescence microscopy and processed as follows: Background was subtracted (*Process → Subtract Background*). A threshold based on fluorescence area was set to ensure accurate identification of cell state (*Image → Adjust → Threshold*).

A segmentation analysis was performed to enable automated counting of individual cells (*Process → Binary → Watershed*). Finally, debris signals were filtered and cells were counted (*Analyze → Analyze Particles*). Toxicity was determined as the percentage of dead cells.

Chemically-inducible TALEs

Neuro2A cells were grown in a medium containing a 1:1 ratio of OptiMEM (Life Technologies) to high-glucose DMEM with GlutaMax and Sodium Pyruvate (Life Technologies) supplemented with 5% HyClone heat-inactivated FBS (Thermo Scientific), 1% penicillin/streptomycin (Life Technologies) and 25mM HEPES (Sigma Aldrich). 150,000 cells were plated in each well of a 24-well plate 18-24 hours prior to transfection. Cells were transfected with 1 μg total of construct DNA (at equimolar ratios) per well and 2 μL of Lipofectamine 2000 (Life Technologies) according to the manufacturer's recommended protocols. Media was exchanged 12 hours post-transfection. For the kinetics test, chemical induction was started 24 hours post-transfection, when abscisic acid (ABA, Sigma Aldrich) was added to fresh media to a final concentration of 250 μM. RNA was extracted using the Rneasy kit (Qiagen) according to manufacturer's instructions and 1 μg of RNA per sample was reverse-transcribed using qScript (Quanta Biosystems). Relative mRNA levels were measured by quantitative real-time PCR (qRT-PCR) using Taqman probes specific for the targeted gene as well as mouse GAPDH as an endogenous control (Life Technologies probe IDs).

ΔΔCt analysis was used to obtain fold-changes relative to negative controls

where cells were subjected to mock transfection with GFP.

Cas9 transcriptional effectors: SOX2 and KLF4 experiments

HEK 293FT cells were co-transfected with mutant Cas9 fusion protein and a

synthetic guide RNA (sgRNA) using Lipofectamine 2000 (Life Technologies) 24

hours after seeding into a 24 well dish. 72 hours post-transfection, total RNA was

purified (Rneasy Plus, Qiagen). 1 μg of RNA was reverse transcribed into cDNA

(qScript, Quanta BioSciences). Quantitative real-time PCR was done according

to the manufacturer's protocol (Life Technologies) and performed in triplicate

using TaqMan Assays for hKlf4 (Hs00358836_m1), hSox2 (Hs01053049_s1),

and the endogenous control GAPDH (Hs02758991_g1).

The hSpCas9 activator plasmid was cloned into a lentiviral vector under the

expression of the hEF1a promoter (pLenti-EF1a-Cas9-NLS-VP64). The hSpCas9

repressor plasmid was cloned into the same vector (pLenti-EF1a-SID4x-NLS-

Cas9-NLS). Guide sequences (20bp) targeted to the KLF4 locus are: 5'-

GCGCGCTCCACACAACTCAC, 5'-GCAAAAATAGACAATCAGCA,

GAAGGATCTCGGCCAATTTG. Spacer sequences for guide RNAs targeted to

the SOX2 locus are: 5'-GCTGCCGGGTTTTGCATGAA, 5'-

CCGGGCCCGCAGCAAACTTC, 5'-GGGGCTGTCAGGGAATAAAT

Transient transfection experiments for Chapter 3

Neuro-2a cells (Sigma-Aldrich) were grown in media containing 1:1 ratio of OptiMEM (Life Technologies) to high-glucose DMEM with GlutaMax and sodium pyruvate (Life Technologies) supplemented with 5% HyClone heat-inactivated FBS (Thermo Scientific), 1% penicillin/streptomycin (Life Technologies), and passaged at 1:5 every 2 days.

HEK293FT cells (Life Technologies) were maintained in high-glucose DMEM with GlutaMax and sodium pyruvate (Life Technologies) supplemented with 10% heat-inactivated characterized HyClone fetal bovine serum (Thermo Scientific) and 1% penicillin/streptomycin (Life Technologies). Cells were passaged daily at a ratio 1:2 or 1:2.5. For gene activation experiments, 20,000 HEK293FT cells/well were plated in 100 μL media in poly-D-lysine coated 96-well plates (BD BioSciences). 24 hours after plating, cells were transfected with a 1:1:1 mass ratio of:

- sgRNA plasmid with gene-specific targeting sequence or pUC19 control plasmid

- MS2-effector plasmid or pUC19.

- dCas9 plasmid, dCas9-effector plasmid, or pUC19.

A total plasmid mass of 0.3 μg/well was transfected using 0.6 μL/well Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions. Culture medium was changed 5 hours after transfection. 48 hours

after transfection, cell lysis and reverse transcription were performed using a Cells-to-Ct kit (Life Technologies). Relative RNA expression levels were quantified by reverse transcription and quantitative PCR (qPCR) using TaqMan qPCR probes (Life Technologies) and Fast Advanced Master Mix (Life Technologies). qPCR was carried out in 5 μL multiplexed reactions and 384-well format using the LightCycler 480 Instrument II. Data was analyzed by the $\Delta\Delta C_t$ method: target Ct values (FAM dye) were normalized to GAPDH Ct values (VIC dye), and fold changes in target gene expression were determined by comparing to GFP-transfected experimental controls.

Lentivirus production

HEK293T cells (Life Technologies) were cultured as described above for HEK293FT cells. 1 day prior to transfection, cells were seeded at ~40% confluency (12xT225 flasks for library scale production, 1xT25 flask for individual guide production). Cells were transfected the next day at ~80-90% confluency. For each flask, 10 μg of plasmid containing the vector of interest, 10 μg of pMD2.G, and 15 μg of psPAX2 (Addgene) were transfected using 100 μL of Lipofectamine 2000 and 200 μL Plus Reagent (Life Technologies). 5h after transfection the media was changed. Virus supernatant was harvested 48h post-transfection, filtered with a 0.45 μm PVDF filter (Millipore), aliquoted, and stored at −80 °C.

Lentiviral transduction

A375 cells (ATCC) were cultured in RPMI 1640 (Life Technologies) supplemented with 10% FBS (Seradigm) and 1% penicillin/streptomycin (Life Technologies) and passaged every other day at a 1:4 ratio. Cells were transduced with lentivirus via spinfection in 12-well plates. 3×10^6 cells in 2 mL of media supplemented with 8 µg/mL polybrene (Sigma) were added to each well, supplemented with lentiviral supernatant and centrifuged for 2h at 1000g. 24h after spinfection, cells were detached with TrypLE (Life Technologies) and counted. Cells were replated at low density (7.5×10^6 cells per T225 Flask) and a selection agent was added either immediately (Zeocin, Blasticidin and Hygromycin, all Life Technologies) or 3h after plating (Puromycin). Concentrations for selection agents we determined using a kill curve: 0.5 µg/ml Puromycin, 200 µg/mL Zeocin, 10 µg/mL Blasticidin, and 300 µg/mL Hygromycin. Media was refreshed on day 2 and cells were passaged every other day starting on day 4 after replating. The duration of selection was 4 days for Puromycin and 7 days for Zeocin, Hygromycin and Blasticidin. Lentiviral titers were determined by spinfecting cells with 6 different volumes of lentivirus ranging from 0 to 600 µL and counting the number of surviving cells after a complete selection (3-6 days).

Design and Cloning of SAM library

RefSeq coding gene isoforms with a unique TSS (total of 23'430 isoforms) were targeted with three guides each for a total library of 70,290 guides. Guides were

109

designed to target the first 200 bp upstream of each TSS and subsequently filtered for GC content >25% and minimal overlap of the target sequence. After filtering, the remaining guides were scored according to predicted off-target matches as described previously[91], and three guides with the best off-target scores were selected. Cloning of the SAM sgRNA libraries was performed as previously described[39] with a minimum representation of 100 transformed colonies/guide.

Depletion and PLX-4720 Screen

A375 cells stably integrated with SAM Cas9 and effector components were transduced with SAM sgRNA libraries as described above at an MOI of 0.2, with a minimal representation of 500 transduced cells/guide. Cells were maintained at >500 cells/guide during subsequent passaging. At 7 DPI (complete selection, see above), cells were split into vehicle (DMSO) and PLX-4720 conditions (2 µM PLX-4720 dissolved in DMSO, Selleckchem). Cells were passaged every 2 days for a total of 14 days of drug treatment. >500 cells/guide were harvested as a baseline at 3 DPI (4 days before treatment) and at 21 DPI (after 14 days of treatment) for gDNA extraction. Genomic DNA was extracted using the Zymo Quick-gDNA midi kit (Zymo Research). PCR of the virally integrated guides was performed on gDNA at the equivalent of >500 cells/guide in 96 parallel reactions using NEBnext High Fidelity 2X Master Mix (New England Biolabs) in a single-step reaction of 22 cycles. Primers are listed below:

forward primer:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC

CGATCTNNNNNNNN(1-10bp

stagger)GCTTTATATATCTTGTGGAAAGGACGAAACACC

8 bp barcode indicated in red

reverse primer:

CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGT

GTGCTC TTCCGATCTGCCAAGTTGATAACGGACTAGCCTT

8 bp index read barcode indicated in red

PCR products from all 96 reactions were pooled, purified using Zymo-Spin™ V with Reservoir (Zymo research) and gel extracted using the Zymoclean™ Gel DNA Recovery Kit (Zymo research). Resulting libraries were deep-sequenced on Illumina MiSeq and HiSeq platforms with a total coverage of >35 million reads passing filter per library.

NGS and screen hits analysis

NGS data were de-multiplexed using unique index reads. Guide counts were determined based on perfectly matched sequencing reads only. For each condition, guide counts were normalized to the total number of counts per condition, and $\log_2$ counts were calculated based on these values. Ratios of

counts between conditions were calculated as $\log_2((\text{count } 1 + 1)/(\text{count } 2 + 1))$ based on normalized counts.

RIGER analysis was performed using GENE-E based on the normalized day 14 $\log_2$ ratios (PLX-4720/DMSO) averaged over two independent infection replicates. All RIGER analysis used the Kolmogorov-Smirnov method as described previously[99].

Gene expression and Pharmacological Validation Analysis

Gene expression data (CCLE, TCGA, short-term cultures, patient melanoma biopsies) and pharmacological data (CCLE, short-term cultures) were analyzed to better understand the biological relevance of the top gene hits from the SAM screens. In the CCLE dataset[81], gene expression data (RNA-sequencing, GCHub: https://cghub.ucsc.edu/datasets/ccle.html) and pharmacological data (activity area for MAPK pathway inhibitors) from $BRAF^{V600}$ mutant melanoma cell lines were used to compute the association between PLX-4720 resistance and the gene expression of each of the top hits. Additionally, gene expression signatures comprised of the top hits were generated using single-sample Gene Set Enrichment Analysis (ssGSEA)[100], and the associations between PLX-4720 resistance and these signatures were computed.

Gene expression data (Affymetrix GeneChip HT-HGU133) and PLX-4720

pharmacological data (GI_{50}; only for a subset of the samples) from short term melanoma cultures (STC)[101] was also used for plotting the gene expression of top hits and their ssGSEA signature scores. Expression data for the STC samples were collapsed to maximum probe value per gene and preprocessed using robust spline normalization.

Gene expression (RNA-sequencing) and genotyping data were collected from 113 $BRAF^{V600}$-mutant primary and metastatic patient tumors from The Cancer Genome Atlas (https://tcga-data.nci.nih.gov/tcga/) and this data was similarly used for determining the association between resistance and the expression of top hits/ssGSEA signature scores. Because pharmacological data was not available for the STCs (only a subset had PLX-4720 data) and the TCGA melanoma samples, a transcriptional state was plotted using marker genes and signatures[80] in order to identify samples resistant to BRAF-inhibition.

Gene expression data from 13 patients with $BRAF^{V600E}$ melanomas[82] was used for analyzing the relationship between resistance and the expression of our top hits/ssGSEA signature scores. Because all the post-treatment tumors were resistant and not every sample had a paired on-treatment biopsy, we decided to order the samples by MITF expression in the pre-treatment samples to reflect the original PLX-4720 sensitivity state of the tumors. We then used the expression data in the post-treatment resistant tumors to plot the expression of top

hits/ssGSEA signature scores. We also calculated the $\log_2$-fold change between each patient's post/pre paired samples and determined the number of patients that had at least a $\log_2$-fold change of 2 per top screen hit.

Single Sample Gene Set Enrichment Analysis

While there was a significant association between the overexpression of some of our top individual SAM screen hits and resistance in three external cancer datasets, we sought a more robust scoring system independent of any single gene. Gene expression signatures were generated based on the set of top hits from each of the two SAM screens and for the overlap between them. Using single-sample Gene Set Enrichment analysis (ssGSEA), a score was generated for each sample that represents the enrichment of the SAM screen gene expression signature in that sample and the extent to which those genes are coordinately up- or down-regulated. Additionally, signature gene sets from the Molecular Signature Database (MSigDB)[102] were used in order to fully map the transcriptional BRAF-inhibitor resistant/sensitive states in the short-term culture and TCGA datasets as previously described[80].

Information Coefficient for Measuring Associations in External Datasets

To measure correlations between different features (signature scores, gene expression, or drug-resistance data) in the external cancer datasets, an information-theoretic approach (Information Coefficient; IC) was used and

significance was measured using a permutation test (n=10,000), as previously described[80]. The IC was calculated between the feature used to sort the samples (columns) in each dataset and each of the features plotted in the heatmap (pharmacological data, gene expression, and signature scores).

sgRNA sequence analysis

Depletion for each sgRNA was calculated as the ratio of counts (see "NGS and screen hits analysis") between day 3 and day 21. The sgRNAs corresponding to genes with significant depletion ($p < 0.05$ by RIGER analysis) in sgRNA-Puro and sgRNA-Zeo libraries were selected for analyses. These sgRNAs were analyzed for nucleotide occurrence in the sgRNA sequence, distance from TSS, and guide strand relative to transcript orientation. For each variable, the correlation and significance with the sgRNA ratio was calculated by Ordinary Least Squares linear regression.

PLX-4720 Survival Assay

A375 cells stably integrated with dCas9-VP64 and MS2-p65-HSF1 were transduced with individual guides from the top screening hits of the Zeocin and Puromycin screens (13 genes total, 3 sgRNAs per gene) as well as available cDNA at an MOI of <0.2 as described above. Cells were selected for guide expression with Zeocin (Life Technologies) for 5 days and replated at low density (3×10^3 cells per well in a 96-well plate). A375 cells and A375 cells expressing

dCas9-VP64 and MS2-p65-HSF1 were plated as controls. Different concentrations of PLX-4720 (2µM, 0.5µM, 0.15µM) or vehicle (DMSO) were added 3h after plating. Cells were treated with PLX-4720 for 4 days before cell viability was measured using CellTiter-Glo Luminescent Cell Viability Assay (Promega). For qPCR quantification of target gene upregulation, cells were also plated at 5 DPI (3×10^4 cells per well in a 96-well plate) and harvested for mRNA 24h after plating.

Western Blot for Chapter 3

Protein lysates were prepared with RIPA lysis buffer (Cell Signaling Technologies) containing a protease inhibitor cocktail (Roche). Samples standardized for protein with the Pierce BCA protein assay (Thermo Scientific) were boiled at 95°C for 5 mins under reducing conditions (except for GPR35 samples, which were incubated at 37°C for 30 mins). After denaturation, samples for probing proteins with lower or higher molecular weight were separated by 10-20% or 4-15% Criterion Tris-HCl gels (Bio-Rad) and electrotransferred onto a 0.2µm or 0.45µm polyvinylidene difluoride membrane (Millipore) respectively. Blots were blocked with 5% BLOT-QuickBlocker (VWR) and probed with different primary antibodies [anti-EGFR (rabbit polyclonal, SC-03, Santa Cruz Biotechnology, 1:1000 dilution), anti-PCDH7 (rabbit polyclonal, HPA011866, Sigma-Aldrich, 1:1000 dilution), anti-ITGB5 (rabbit polyclonal, SC-14010, Santa Cruz Biotechnology, 1:500 dilution), anti-ARHGEF1 (rabbit polyclonal, 11363-1-

AP, Proteintech, 1:5000 dilution), anti-BCAR3 (rabbit polyclonal, A301-671A, Bethyl Laboratories, 1:2000 dilution), anti-GPR35 (rabbit polyclonal, 10007660, Cayman Chemical, 1:1000 dilution), anti-TFAP2C (rabbit polyclonal, 2320, Cell Signaling Technology, 1:1000 dilution, 2.5% bovine serum albumin, Sigma-Aldrich)] in 2.5% BLOT-QuickBlocker (VWR) unless noted otherwise overnight at 4°C. Blots were then incubated with secondary antibody HRP-conjugated goat anti-rabbit IgG (7074, Cell Signaling Technology, 1:1000 dilution) and HRP-conjugated GAPDH (rabbit monoclonal, 3683, Cell Signaling Technology, 1:2000 dilution) in 2.5% BLOT-QuickBlocker (VWR) for 1hr at room temperature. Proteins with molecular weights similar to GAPDH (GPR35 and TFAP2C) were stripped with Restore Plus Western Blot Stripping Buffer (Thermo Scientific) before probing for GAPDH. SuperSignal West Pico and Femto Chemiluminescent Substrates (Thermo Scientific) were used for detection.

RNA Sequencing and Data Analysis

Samples harvested for RNA sequencing were prepped with TruSeq Stranded mRNA Sample Prep Kit (Illumina) and deep-sequenced on the Illumina MiSeq platform (>9 Mio reads per condition). Bowtie2[103] index was created based on the human hg19 UCSC genome and known gene transcriptome, and paired-end reads were aligned directly to this index using Bowtie2 with command line options "-q --phred33-quals -n 2 -e 99999999 -l 25 -I 1 -X 1000 -a -m 200 –p 4 --chunkmbs 512". Next, RSEM v1.27[104] was run with default parameters on the

alignments created by Bowtie2 to estimate expression levels. RSEM's gene level expression estimates (tau) were multiplied by 1,000,000 to obtain transcript per million (TPM) estimates for each gene, and TPM estimates were transformed to log-space by taking $\log_2(TPM+1)$. The normalization between libraries was tested using an MA plot (mairplot function in Matlab V2013b). Genes were considered detected if their transformed expression level was equal to or above 1 (in $\log_2(TPM+1)$ scale). All genes detected in at least one library (out of three libraries per condition) were used to construct scatter plots comparing each of the six conditions to the control GFP condition, using the average across biological replicates with >80% alignment to the hg19 UCSC known gene transcriptome ($\log_2(mean(TPM)+1)$ value per gene) .

To find differentially expressed genes, we performed Student's t-test on each of the six conditions against the GFP condition. The t-test was run on all genes that had expression levels above $\log_2(TPM+1)>2.5$ in at least two libraries. This threshold was chosen as the minimal threshold for which the number of detected genes across all libraries was constant. Only genes that were significant (p-value pass 0.01 FDR correction) and had at least 1.5 fold change were reported and visualized using a heatmap.

Bibliography

1   Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326**, 1509-1512, doi:10.1126/science.1178811 (2009).

2   Moscou, M. J. & Bogdanove, A. J. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science* **326**, 1501, doi:10.1126/science.1178817 (2009).

3   Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotech* **29**, 149-153, doi:http://www.nature.com/nbt/journal/v29/n2/abs/nbt.1775.html - supplementary-information (2011).

4   Morbitzer, R. *et al.* Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc Natl Acad Sci U S A* **107**, 21617-21622, doi:10.1073/pnas.1013133107 (2010).

5   Geissler, R. *et al.* Transcriptional activators of human genes with programmable DNA-specificity. *PLoS ONE* **6**, e19509, doi:10.1371/journal.pone.0019509 (2011).

6   Deng, D. *et al.* Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335**, 720-723, doi:10.1126/science.1215670 (2012).

7   Klug, A. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q Rev Biophys* **43**, 1-21, doi:10.1017/S0033583510000089 (2010).

8   Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).

9   Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826, doi:10.1126/science.1232033 (2013).

10   Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

11   Gasiunas, G. *et al.* Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E2579-2586, doi:10.1073/pnas.1208507109 (2012).

12   Deisseroth, K. Optogenetics. *Nat Methods* **8**, 26-29, doi:10.1038/nmeth.f.324 (2011).

13   Yizhar, O. *et al.* Optogenetics in neural systems. *Neuron* **71**, 9-34, doi:10.1016/j.neuron.2011.06.004 (2011).

14   Zhang, F. *et al.* The Microbial Opsin Family of Optogenetic Tools. *Cell* **147**, 1446-1457 (2011).

15   Airan, R. D. *et al.* Temporally precise in vivo control of intracellular signalling. *Nature* **458**, 1025-1029, doi:10.1038/nature07926 (2009).

16   Levskaya, A. *et al.* Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature* **461**, 997-1001, doi:http://www.nature.com/nature/journal/v461/n7266/suppinfo/nature08446_S1.html (2009).

17   Strickland, D. *et al.* TULIPs: tunable, light-controlled interacting protein tags for cell biology. *Nat Meth* **9**, 379-384, doi:http://www.nature.com/nmeth/journal/v9/n4/abs/nmeth.1904.html - supplementary-information (2012).

18   Yazawa, M. *et al.* Induction of protein-protein interactions in live cells using light. *Nat Biotech* **27**, 941-945, doi:http://www.nature.com/nbt/journal/v27/n10/suppinfo/nbt.1569_S1.html (2009).

19   Kennedy, M. J. *et al.* Rapid blue-light-mediated induction of protein interactions in living cells. *Nat Meth* **7**, 973-975, doi:http://www.nature.com/nmeth/journal/v7/n12/abs/nmeth.1524.html - supplementary-information (2010).

20   Polstein, L. R. & Gersbach, C. A. Light-inducible spatiotemporal control of gene activation by customizable zinc finger transcription factors. *J Am Chem Soc* **134**, 16480-16483, doi:10.1021/ja3065667 (2012).

21   Shimizu-Sato, S. *et al.* A light-switchable gene promoter system. *Nat Biotech* **20**, 1041-1044 (2002).

22   Wang, X. *et al.* Spatiotemporal control of gene expression by a light-switchable transgene system. *Nat Meth* **9**, 266-269, doi:http://www.nature.com/nmeth/journal/v9/n3/abs/nmeth.1892.html - supplementary-information (2012).

23   Ye, H. *et al.* A Synthetic Optogenetic Transcription Device Enhances Blood-Glucose Homeostasis in Mice. *Science* **332**, 1565-1568, doi:10.1126/science.1203535 (2011).

24   Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology* **29**, 143-148, doi:10.1038/nbt.1755 (2011).

25   Cong, L. *et al.* Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun* **3**, 968, doi:http://www.nature.com/ncomms/journal/v3/n7/suppinfo/ncomms1962_S1.html (2012).

26   Liu, H. *et al.* Photoexcited CRY2 Interacts with CIB1 to Regulate Transcription and Floral Initiation in Arabidopsis. *Science* **322**, 1535-1539, doi:10.1126/science.1163927 (2008).

27   Beerli, R. R. *et al.* Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14628-146331998/12/09 (1998).

28   Banerjee, R. *et al.* The Signaling State of Arabidopsis Cryptochrome 2 Contains Flavin Semiquinone. *Journal of Biological Chemistry* **282**, 14916-14922, doi:10.1074/jbc.M700616200 (2007).

29   Moore, M. J. & Proudfoot, N. J. Pre-mRNA Processing Reaches Back toTranscription and Ahead to Translation. *Cell* **136**, 688-700 (2009).

30   Proudfoot, N. J. *et al.* Integrating mRNA Processing with Transcription. *Cell* **108**, 501-512 (2002).

31   Liang, F. S. *et al.* Engineering the ABA Plant Stress Pathway for Regulation of Induced Proximity. *Sci Signal* **4**, doi:ARTN rs2 DOI 10.1126/scisignal.2001449 (2011).

32   Holkers, M. *et al.* Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Research*, doi:10.1093/nar/gks1446 (2012).

33   Zhang, F. *et al.* Optogenetic interrogation of neural circuits: technology for probing mammalian brain structures. *Nature protocols* **5**, 439-456, doi:10.1038/nprot.2009.226 (2010).

34   de Groote, M. L. *et al.* Epigenetic Editing: targeted rewriting of epigenetic marks to modulate expression of selected target genes. *Nucleic Acids Research* **40**, 10596-10613, doi:10.1093/nar/gks863 (2012).

35   Szymczak, A. L. *et al.* Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nature Biotechnology* **22**, 589-594, doi:10.1038/nbt957 (2004).

36   Christie, J. M. *et al.* Structural tuning of the fluorescent protein iLOV for improved photostability. *J Biol Chem* **287**, 22295-22304, doi:10.1074/jbc.M111.318881 (2012).

37   Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431-437, doi:10.1038/nature02371 (2004).

38   Boutros, M. *et al.* Genome-wide RNAi analysis of growth and viability in Drosophila cells. *Science* **303**, 832-835, doi:10.1126/science.1091266 (2004).

39   Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87, doi:10.1126/science.1247005 (2014).

40   Wang, T. *et al.* Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80-84, doi:10.1126/science.1246981 (2014).

41   Koike-Yusa, H. *et al.* Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature biotechnology* **32**, 267-273, doi:10.1038/nbt.2800 (2014).

42   Beerli, R. R. & Barbas, C. F., 3rd. Engineering polydactyl zinc-finger transcription factors. *Nature biotechnology* **20**, 135-141, doi:10.1038/nbt0202-135 (2002).

43   Beerli, R. R. *et al.* Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14628-14633 (1998).

44   Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature biotechnology* **29**, 149-153, doi:10.1038/nbt.1775 (2011).

45   Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451, doi:10.1016/j.cell.2013.06.044 (2013).

46   Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472-476, doi:10.1038/nature12466 (2013).

47   Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nature methods* **10**, 977-979, doi:10.1038/nmeth.2598 (2013).

48   Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology* **31**, 833-838, doi:10.1038/nbt.2675 (2013).

49   Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods* **10**, 973-976, doi:10.1038/nmeth.2600 (2013).

50   Perez-Pinera, P. *et al.* Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nature methods* **10**, 239-242, doi:10.1038/nmeth.2361 (2013).

51   Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949, doi:10.1016/j.cell.2014.02.001 (2014).

52   Peabody, D. S. The RNA binding site of bacteriophage MS2 coat protein. *The EMBO journal* **12**, 595-600 (1993).

53   Auslander, S. *et al.* Programmable single-cell mammalian biocomputers. *Nature* **487**, 123-127, doi:10.1038/nature11149 (2012).

54   Lemon, B. & Tjian, R. Orchestrated response: a symphony of transcription factors for gene control. *Genes & development* **14**, 2551-2569 (2000).

55   van Essen, D. *et al.* Two modes of transcriptional activation at native promoters by NF-kappaB p65. *PLoS biology* **7**, e73, doi:10.1371/journal.pbio.1000073 (2009).

56   Kretzschmar, M. *et al.* A novel mediator of class II gene transcription with homology to viral immediate-early transcriptional regulators. *Cell* **78**, 525-534 (1994).

57   Ikeda, K. *et al.* The H1 and H2 regions of the activation domain of herpes simplex virion protein 16 stimulate transcription through distinct molecular mechanisms. *Genes to cells : devoted to molecular & cellular mechanisms* **7**, 49-58 (2002).

58   Neely, K. E. *et al.* Activation domain-mediated targeting of the SWI/SNF complex to promoters stimulates transcription from nucleosome arrays. *Molecular cell* **4**, 649-655 (1999).

59   Marinho, H. S. *et al.* Hydrogen peroxide sensing, signaling and regulation of transcription factors. *Redox biology* **2**, 535-562, doi:10.1016/j.redox.2014.02.006 (2014).

60   Weintraub, H. *et al.* Muscle-specific transcriptional activation by MyoD. *Genes & development* **5**, 1377-1386 (1991).

61   Kretz, M. *et al.* Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* **493**, 231-235, doi:10.1038/nature11661 (2013).

62   Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120-124, doi:10.1038/nature09819 (2011).

63   Prensner, J. R. *et al.* Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nature biotechnology* **29**, 742-749, doi:10.1038/nbt.1914 (2011).

64   Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature biotechnology* **32**, 670-676, doi:10.1038/nbt.2889 (2014).

65   Johannessen, C. M. *et al.* COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. *Nature* **468**, 968-972, doi:10.1038/nature09627 (2010).

66   Nazarian, R. *et al.* Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* **468**, 973-977, doi:10.1038/nature09626 (2010).

67   Prahallad, A. *et al.* Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature* **483**, 100-103, doi:10.1038/nature10868 (2012).

68   Corcoran, R. B. *et al.* EGFR-mediated re-activation of MAPK signaling contributes to insensitivity of BRAF mutant colorectal cancers to RAF inhibition with vemurafenib. *Cancer discovery* **2**, 227-235, doi:10.1158/2159-8290.CD-11-0341 (2012).

69   Villanueva, J. *et al.* Acquired resistance to BRAF inhibitors mediated by a RAF kinase switch in melanoma can be overcome by cotargeting MEK and IGF-1R/PI3K. *Cancer cell* **18**, 683-695, doi:10.1016/j.ccr.2010.11.023 (2010).

70   Shi, H. *et al.* Combinatorial treatments that overcome PDGFRbeta-driven resistance of melanoma cells to V600EB-RAF inhibition. *Cancer research* **71**, 5067-5074, doi:10.1158/0008-5472.CAN-11-0140 (2011).

71   Smalley, K. S. Understanding melanoma signaling networks as the basis for molecular targeted therapy. *The Journal of investigative dermatology* **130**, 28-37, doi:10.1038/jid.2009.177 (2010).

72   Johannessen, C. M. *et al.* A melanocyte lineage program confers resistance to MAP kinase pathway inhibition. *Nature* **504**, 138-142, doi:10.1038/nature12688 (2013).

73   Musgrove, E. A. & Sutherland, R. L. Biological determinants of endocrine resistance in breast cancer. *Nature reviews. Cancer* **9**, 631-643, doi:10.1038/nrc2713 (2009).

74   Wong, P. P. *et al.* Histone demethylase KDM5B collaborates with TFAP2C and Myc to repress the cell cycle inhibitor p21(cip) (CDKN1A). *Molecular and cellular biology* **32**, 1633-1644, doi:10.1128/MCB.06373-11 (2012).

75   Hart, M. J. *et al.* Direct stimulation of the guanine nucleotide exchange activity of p115 RhoGEF by Galpha13. *Science* **280**, 2112-2114 (1998).

76   Dorsam, R. T. & Gutkind, J. S. G-protein-coupled receptors and cancer. *Nature reviews. Cancer* **7**, 79-94, doi:10.1038/nrc2069 (2007).

77   Lappano, R. & Maggiolini, M. G protein-coupled receptors: novel targets for drug discovery in cancer. *Nature reviews. Drug discovery* **10**, 47-60, doi:10.1038/nrd3320 (2011).

78   Franke, T. F. PI3K/Akt: getting it right matters. *Oncogene* **27**, 6473-6488, doi:10.1038/onc.2008.313 (2008).

79   Desgrosellier, J. S. & Cheresh, D. A. Integrins in cancer: biological implications and therapeutic opportunities. *Nature reviews. Cancer* **10**, 9-22, doi:10.1038/nrc2748 (2010).

80   Konieczkowski, D. J. *et al.* A melanoma cell state distinction influences sensitivity to MAPK pathway inhibitors. *Cancer discovery* **4**, 816-827, doi:10.1158/2159-8290.CD-13-0424 (2014).

81   Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607, doi:10.1038/nature11003 (2012).

82   Rizos, H. *et al.* BRAF inhibitor resistance mechanisms in metastatic melanoma: spectrum and clinical impact. *Clinical cancer research : an official journal of the American Association for Cancer Research* **20**, 1965-1977, doi:10.1158/1078-0432.CCR-13-3122 (2014).

83   Anders, C. *et al.* Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573, doi:10.1038/nature13579 (2014).

84   Tanenbaum, M. E. *et al.* A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell*, doi:10.1016/j.cell.2014.09.039 (2014).

85   Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, doi:10.1016/j.cell.2014.09.029 (2014).

86   Zalatan, J. G. *et al.* Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell* **160**, 339-350, doi:10.1016/j.cell.2014.11.052 (2015).

87   Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature methods* **10**, 1116-1121, doi:10.1038/nmeth.2681 (2013).

88   Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. **10**, 973-976, doi:10.1038/nmeth.2600 (2013).

89   Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature biotechnology* **33**, 510-517, doi:10.1038/nbt.3199 (2015).

90   Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588, doi:10.1038/nature14136 (2015).

91   Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* **31**, 827-832, doi:10.1038/nbt.2647 (2013).

92   Briner, A. E. *et al.* Guide RNA functional modules direct Cas9 activity and orthogonality. *Molecular cell* **56**, 333-339, doi:10.1016/j.molcel.2014.09.019 (2014).

93   Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* **31**, 822-826, doi:10.1038/nbt.2623 (2013).

94   Platt, R. J. *et al.* CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440-455, doi:10.1016/j.cell.2014.09.014 (2014).

95   Dow, L. E. *et al.* Inducible in vivo genome editing with CRISPR-Cas9. *Nature biotechnology* **33**, 390-394, doi:10.1038/nbt.3155 (2015).

96   Fu, Y. *et al.* Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology* **32**, 279-284, doi:10.1038/nbt.2808 (2014).

97   McClure, C. *et al.* Production and titering of recombinant adeno-associated viral vectors. *J Vis Exp*, e3348, doi:10.3791/3348 (2011).

98   Blecher-Gonen, R. *et al.* High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nature protocols* **8**, 539-554, doi:10.1038/nprot.2013.023 (2013).

99   Luo, B. *et al.* Highly parallel identification of essential genes in cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 20380-20385, doi:10.1073/pnas.0810485105 (2008).

100   Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108-112, doi:10.1038/nature08460 (2009).

101   Lin, W. M. *et al.* Modeling genomic diversity and tumor dependency in malignant melanoma. *Cancer research* **68**, 664-673, doi:10.1158/0008-5472.CAN-07-2615 (2008).

102   Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740, doi:10.1093/bioinformatics/btr260 (2011).

103   Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

104   Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).