Comparative Analysis of Cas9 Activators Across Multiple Species

Published in final edited form as:

Nat Methods. 2016 July ; 13(7): 563–567. doi:10.1038/nmeth.3871.

Comparative Analysis of Cas9 Activators Across Multiple Species

Alejandro Chavez[#1,2,3,12], **Marcelle Tuttle**[#1], **Benjamin W Pruitt**[1], **Ben Ewen-Campen**[3], **Raj Chari**[3], **Dmitry Ter-Ovanesyan**[1,3], **Sabina J Haque**[4,5], **Ryan J Cecchi**[1], **Emma J K Kowal**[1], **Joanna Buchthal**[1], **Benjamin E Housden**[3], **Norbert Perrimon**[3,6], **James J Collins**[1,7,8,9,10], and **George Church**[1,3,12]

[1]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, Massachusetts, USA.

[2]Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts, USA.

[3]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA.

[4]Department of Molecular Biology and Biochemistry, Middlebury College, Middlebury, VT, USA

[5]Department of Mathematics, Middlebury College, Middlebury, VT, USA

[6]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts, USA.

[7]Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

[8]Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

[9]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

[10]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

[#] These authors contributed equally to this work.

Abstract

Several groups have generated programmable transcription factors based on the versatile Cas9 protein, yet their relative potency and effectiveness across various cell types and species remain

[12]To whom correspondence should be addressed alex.chavez@wyss.harvard.edu and gchurch@genetics.med.harvard.edu.

Contributions: A.C. and M.T. conceived of the study. A.C., M.T., B.W.P., and R.C. designed and performed experiments. S.J.H., R.J.C., and J.B. performed experiments.B.E.C, B.E.H and N.P. designed and performed all experiments in *Drosophila Melanogaster.* D.T-O and E.J.K.K. performed RNA-seq experiments and analyzed data. J.J.C and G.M.C. supervised the study. A.C. and M.T. wrote the manuscript with support from all authors.

Conflict of interest: G.M.C. has equity in Editas and Caribou Biosciences.

Accession codes:
Primary accessions:
Gene Expression Omnibus
GSE80611

unexplored. Here, we compare Cas9 activator systems and examine their ability to induce robust gene expression in several human, mouse, and fly cell lines. We also explore the potential for improved activation through the combination of the most potent activator systems and assess the role of cooperativity in maximizing gene expression.

Introduction

Cas9 is an RNA-guided endonuclease that can be directed to cut a selected site[1,2]. This process requires complementarity between the Cas9-associated guide RNA (gRNA) and the target site, along with the presence of a short protospacer adjacent motif[3,4,5,6]. Early efforts at engineering the Cas9 protein uncovered several residues involved in DNA catalysis that, when mutated, generate forms of the protein that are still capable of DNA binding but lack detectable nuclease activity[4,7,8]. These nuclease-null, or "dead," Cas9 (dCas9) variants can then be fused to effector domains, allowing users to precisely direct a given functional activity to any arbitrary locus within the genome[7,9,10,11]. Recently, several groups have generated systems to endow dCas9 with the ability to activate gene expression, with dCas9-VP64 representing the first activator and the standard against which subsequent "second-generation" activators are typically compared[12,13,14,15,16,17,18,19]. Due to differences in the cellular context in which the various second-generation activators were tested and non-uniformity with regard to the particular target genes, guide RNA, transfection conditions, and time to analysis of gene induction, it remains ambiguous which system is the most potent and whether any individual system possesses unique properties not displayed by the others[20,21,22]. Here, we conduct a survey of the various second-generation activators, identifying the most potent systems, which we rigorously characterize across a plethora of target genes and species. These data confer much needed guidance to those wishing to adopt dCas9 activator technology and provide the community with an extensive set of validated reagents to aid the adoption of these tools within labs without previous experience.

Results

The number of second-generation dCas9 activators is too large to be systematically tested across a large panel of target genes and cell lines. Accordingly, we first performed a series of pilot experiments within human embryonic kidney (HEK) 293T cells to compare representative examples of all published dCas9 activators (**Fig 1a**). As expected, for the two target genes tested, the majority of second-generation systems show improved levels of activation as compared to dCas9-VP64, with three activators in particular - VPR, SAM, and Suntag - appearing to be the most potent (**Fig 1b** and activators described in detail in **Supplementary Note 1**)[12,13,15,16]. Based on these initial data, we decided to focus our efforts on VPR, SAM, and Suntag.

When compared across a panel of coding and non-coding genes, VPR, SAM, and Suntag demonstrate the ability to induce potent gene expression. At times, this activation reached levels several orders of magnitude higher than the first-generation dCas9-VP64 activator (**Fig. 2a and Supplementary Fig. 1**). SAM was the most consistent in delivering high levels of gene induction, although it always remained within five-fold of either Suntag or VPR,

neither of which was generally superior to the other (**Fig. 2a**). Previously, a negative correlation between the basal expression state of a given gene and the fold change in gene expression upon targeted upregulation by dCas9-based activators was reported[12,13]. In other words, lowly expressed genes tend to have a higher fold induction than highly expressed genes. We find that this phenomena is a general principle for each of the tested synthetic activators, which suggests that these systems are only capable of inducing gene expression to some static upper limit, dependent upon the activator architecture employed (**Supplementary Figure 2**).

One of the benefits of dCas9-based transcriptional effectors over other synthetic transcription factors is the ease by which multiple loci can be upregulated, only requiring the provision of an extra gRNA for each additional locus one desires to activate. Previous work has shown a general decrease in gene activation upon multiplexing, but whether this is true of all systems is unknown[12,19]. To address this question, we performed a series of multiplex activation experiments (**Supplementary Fig. 3**). When activating three genes at once between VPR, SAM, and Suntag, no system appeared superior to the others with all showing similar levels of activation, within an order of magnitude of each other for the majority of target genes. This is a somewhat surprising finding given that SAM and Suntag are believed to require the assembly of large multimember protein complexes in order to generate a highly competent activator. To further investigate the limitations of multiplexing, we simultaneously directed each of the systems to six target genes at once. Even within this more complex activation scheme all systems showed similar levels of relative gene activation, and in doing so highlight the robustness of these tools to actuate complex transcriptional regulation (**Fig 2b**). Of note, we observe within our control samples inherent variation in the level of basal gene expression upon which we base our calculation of relative fold induction. We are therefore unable to make comparisons with regard to the absolute amount of activation a given system can perform when targeted to a single gene versus multiple genes by comparing expression data across experiments.

An important concern for any Cas9-based technology is the specificity of the desired effect. It has been demonstrated that Cas9 binds promiscuously throughout the genome[23]. This effect could potentially result in aberrant transcription for Cas9-activators if they were to bind to the wrong promoter. To test the specificity of each activator, we performed whole RNA-seq on samples with each of our activators targeting *HBG1* (**Figure 3**). We found that the correlation in gene expression between each activator and our control sample was very similar to the correlation between biological replicates in our dataset ($R \sim 0.98$ in each case), indicating that gene expression is not broadly affected by the presence of any activator (**Figure 3a**). Disregarding noise from genes with low baseline expression (<0.1 TPM), *HBG1* was the most highly upregulated gene in each sample except for VP64, indicating that the activators are highly specific (**Figure 3b**).

To further explore the generality of the trends we observed within HEK293T cells, we targeted a smaller subset of genes within other biomedically relevant human cell lines (Hela, U-2 OS, and MCF7). While Hela cells demonstrated a similar trend of SAM leading to more potent activation, U-2 OS and MCF7 cells showed a deviation from this trend with Suntag and VPR being the more potent activators (**Fig 4a and Supplementary Fig 4a**). The

biological basis for these cell-line specific differences remains to be determined and is not likely due to differences in basal expression state given that *HBG1* and *TTN* are expressed at a similarly low basal state across the tested cell lines (data not shown).

Because human cell lines represent only a single context within which dCas9-based activators might be employed, we performed a series of experiments within mouse N2A and 3T3 cells as well as *Drosophila* S2R+ cells[13]. As a whole, all second-generation activators showed improved activity as compared to the dCas9-VP64 standard. Within mouse and *Drosophila* cells, VPR, SAM, and Suntag showed similar levels of gene induction (within five-fold of each other), with the most potent activator changing depending on the target analyzed (**Fig 4b and Supplementary Fig 4b**).

Given that VPR, SAM, and Suntag represent divergent approaches to Cas9 activator design, we sought to determine if any of the activator components could be combined to generate a hybrid activator with increased potency. First, we naively combined each of the systems (**Supplementary Fig. 5**). For example, we used the Suntag dCas9 component and the SAM-modified gRNA to simultaneously recruit both scFv-VP64 to the end of dCas9 and p65-hsf1 to the gRNA. In addition, we tested if systems such as SAM or Suntag that rely on recruiting several copies of their activator p65-hsf1 or VP64, respectively might be enhanced by nucleating a different activator such as VPR (**Supplementary Fig. 6**). Finally, we investigated if the MS2 hairpins utilized within the SAM system to recruit the MS2 binding protein-p65-hsf1 activator component might be enhanced by combining them in part or in whole with the scaffold method of MS2 binding protein mediated activator recruitment (**Supplementary Fig. 7**). Towards this goal a hybrid SAM+Scaffold gRNA was generated along with a series of chimeras containing select portions of each system combined together. Despite our exhaustive attempts, there was no combination of dCas9 activators, secondary effector component alteration, or gRNA modification that could lead to a chimeric system with enhanced transcriptional regulatory capacity for either single or multiplex gene induction.

Previous work with the first-generation dCas9-VP64 activator uncovered a role for cooperativity in enabling synergistic increases in gene activation[7,10,11,24,25]. Whether this observation is also true for the significantly more potent second-generation activators remains unexplored. It is likely that the extendibility of these newer systems may be limited from further increasing gene expression due to insufficient host resources (such as local RNA polymerase concentrations near a given locus)[26]. To study the effects of recruiting multiple dCas9 activator complexes to the same locus, we generated three gRNAs against a lowly, intermediately, and highly expressed gene target (*ASCL1*, *NEUROD1,* and *CXCR4* respectively) within HEK293T cells and transfected each of the gRNAs alone or all the gRNAs against a given target gene at once. Simultaneous targeting of the same gene with multiple gRNAs led to additive or greater levels of gene activation (**Fig. 4c**).

Discussion

Cas9-based activation represents a powerful method to study gain-of-function phenomena on genomic scales and has already been used to uncover mechanisms of toxin-mediated cell

death and resistance of tumor cells to targeted therapy.[12,15] While it is tempting for those analyzing the initial Cas9 activator publications to make cross-study comparisons, these types of analysis are fraught with confounding variables, such as differences in the cell line used, gRNA selected, amount of Cas9 complex transfected, and time until RNA extraction. In addition, often overlooked variation in the basal expression level of some of the more lowly expressed target genes can lead to large differences in the perceived performance of one system over another. Within our own data, there are instances where the absolute level of target gene activation differs due to these changes, despite these experiments being performed under similar conditions (**Fig 1 and Supplementary Fig 1**). This source of variation emphasizes the need for comparisons across activators to be performed within a single experiment if users want to make statements with regard to absolute levels of gene induction. In contrast, for cases where understanding a relative difference in performance is the primary goal, differences in basal expression state will no longer play a role as each activator is compared to each other thus eliminating variance due to basal expression state.

Our results demonstrate that, across a range of target genes and cellular environments, the VPR, SAM, and Suntag systems are consistently superior to the previous VP64 standard. In addition, while SAM shows a trend for more potent activation in some contexts, VPR, SAM, and Suntag generally fall within an order of magnitude of each other with regard to fold increase in gene expression. It is additionally comforting that each activator was similarly specific. This suggests that constraints outside of the ability to induce gene expression or specificity, such as ease of delivery, familiarity with the system, and access to the necessary reagents, may be more practical concerns to consider[27].

It is assuring that the second-generation activators show robust activation even when employed in organisms such as *Drosophila* that are very distantly related to the human cell lines in which these tools were developed. The demonstrated portability of these dCas9 activators across disparate model systems and species suggests that they hold potential as universal genetic tools. It is likely that these systems will prove effective across a wide range of scientifically interesting organisms with minimal additional engineering.

Our attempts to build an improved chimeric activator by fusing elements from VPR, SAM, and Suntag were unsuccessful (**Supplementary Figs. 5-7**). This is particularly unexpected given the drastically different ways in which the systems generate a functional activator[12–16]. It is difficult to disentangle issues such as steric interference between neighboring activator elements from the possibility that each of these systems[28], while different, may be interacting with a similar subset of the transcriptional regulatory machinery[29,30]. Future efforts to improve dCas9-based activators may benefit from exploring other unique architectures or novel activation domains, along with examining the use of epigenetic modifiers to provide complementary mechanisms to further enhance activation[18,31].

While we did not perform experiments to explicitly address potential differences in gRNA binding site preferences between VPR, SAM, and Suntag, it is telling that for experiments where several gRNAs against the same gene were examined, the activators show the same gRNAs as being either the most potent or the least potent (**Fig. 4c**). These data suggest that

the rules governing activator placement within a given promoter region are shared. It will be interesting to characterize the properties shared by highly potent gRNAs to help distinguish effects of distance from the transcriptional start site from features such as gRNA sequence composition and stability.[12,15,32,33,23] In addition, the fact that we are able to achieve increased levels of programmable gene induction by employing multiple gRNAs against a single target gene suggests that none of the current Cas9 activators have yet to achieve "maximal activation". Since if they were, each activator would not have been aided by the use of multiple guides. These results suggest that continued improvements to Cas9 activators are still possible and should aid in generating a more reliable tool for performing systematic genome-wide screens.

Although not highlighted in any of the original studies for VPR, SAM, or Suntag, our results suggest that the combination of multiple gRNAs represents a viable strategy to enhance gene expression in cases where maximal induction is desired. This strategy may be most valuable for highly expressed genes, which are generally recalcitrant to large amounts of overexpression from the native locus[12,13]. It could also be used to assure that all genes are consistently upregulated given that even for easily upregulated genes there can be over a hundred-fold difference in potency between various gRNAs (**Fig. 4c**).

Materials and methods

Vector design and construction

Vectors used and guide scaffolds will be deposited to Addgene

dCas9-VP64 (Addgene #47319), dCas9-VPR (Addgene #63798), dCas9-VP64 (for SAM activation Addgene # 61425), MCP-p65-hsf1 (Addgene #61426), dCas9-10xGCN4 (Addgene #60903), scFv-VP64 (Addgene #60904), dCas9-p300core (Addgene #61357), dCas9-VP160 (Addgene #48225), Cas9-m4 (Addgene #47316) and MCP-VP64[7] were previously described.

For all systems tested the original vectors deposited to Addgene were employed allowing us to use the presumably optimal expression vector decided upon by the various depositing laboratories. dCas9-VP64, dCas9-VPR, dCas9-p300, dCas9-VP160, and dCas9-m4 constructs were all driven under CMV promoters. dCas9-VP64 (for SAM activation) and MCP-p65-hsf1 used EF1alpha promoters and dCas9-10xGCN4 and scFv-VP64 used SV40 promoters.

Gibson cloning was utilized to make all variants. For SAM and Scaffold gRNA variants, gene blocks were used (Integrated DNA Technologies). See **Supplementary Note 2** for full sequence information.

All dCas9 activator components were cloned into a *Drosophila* pActin vector using Gibson assembly,[34,35] adding a Kozak sequence (GCCACC) immediately upstream of the start codon.

To generate *Drosophila* MS2-containing sgRNA expression vectors for SAM and Scaffold, the pCFD3 plasmid was modified via Gibson assembly to include the indicated sgRNA tails.[36] For all sgRNA plasmids guide oligos were cloned using a BbsI digest.[36]

Mammalian cell culture

All human and mouse cell lines were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS and penicillin-streptomycin (all from ThermoFisher) in an incubator set at 5% CO_2 and 37° celsius. MCF7 (gift of J. Lee, Cold Spring Harbor, Cold Spring Harbor, NY) media was supplemented with 0.01mg/ml human recombinant insulin (EMD Millipore). HEK293Ts, Helas (both gifts of P. Mali, UCSD, San Diego, CA), U-2 OS (ATCC HTB-96), MCF7s, N2As (ATCC CCL-131), and NIH-3T3s (gift of S. Shipman, Harvard Medical School, Boston, MA) were seeded into 24-well plates for transfection.

Cell lines were not tested for authenticity and were tested yearly for mycoplasma.

Mammalian transfections

DNA was aliquoted into individual tubes prior to transfection. 25ng of plasmid DNA containing the Cas9 component and 10ng of plasmid DNA with the guide was transfected into each well. For activators requiring a second component (Suntag, SAM, Scaffold), 100ng of plasmid DNA containing that component was also transfected. For all the other activators, 100ng of an empty vector (Puc19) was substituted for the secondary component in order to ensure equal amounts of DNA were transfected. For experiments involving multiple guides targeted towards one gene, the amount of gRNA plasmid per transfection was scaled so that the total plasmid gRNA in each well was 10ng. For example, for three gRNAs activating one gene, 3.33ng of each guide was transfected. In experiments involving multiple guides targeted towards different genes, 10ng of gRNA plasmid DNA was transfected per target. For all gRNA spacer sequences, please see **Supplementary Table 1**.

293T cells were transfected with Lipofectamine 2000 (ThermoFisher) using the following protocol. For each transfection, a stock solution of 50 µl of Opti-Mem (ThermoFisher) and 2 µl of Lipofectamine 2000 per transfection was made. The solution was then vortexed and incubated for five minutes. The Opti-mem/Lipofectamine solution was then added to the individual aliquots of DNA, vortexed and spun in a centrifuge at 100RPM for 1 minute. The DNA was then incubated for 30 minutes before being added to the cells.

All other cell lines were transfected with Lipofectamine 3000 (ThermoFisher) using the following protocol. 25 µl of Opti-MEM was added to each DNA aliquot along with 1 µl P3000 reagent per transfection. Master mixes of Opti-MEM and Lipofectamine 3000 were prepared with 25 µl of Opti-Mem and 1 µl or 0.5 µl of Lipofectamine 3000 per transfection. MCF7s, N2As, and U-2 OSes were transfected using 1 µl of P3000 reagent and 0.5 µl of Lipofectamine 3000 reagent per transfection. N2As and NIH-3T3s were transfected with 1 µl of P3000 reagent and 1 µl of Lipofectamine 3000 reagent per transfection. The solutions were vortexed and incubated separately for five minutes before 25 µl of the Lipofectamine 3000/Opti-Mem mix per transfection were added to the DNA aliquots. These new solutions

were then vortexed, centrifuged at 100RPM for 1 minute, incubated for 30 minutes and then added to the cells.

To avoid excess toxicity to the cells, media was changed after 24 hours for the 293T, N2A and NIH-3T3 cells. Otherwise, media was changed after 4 hours.

RNA extraction and qPCR analysis for mammalian cell lines

Cells were harvested for RNA 48 hours post-transfection. RNA was extracted using the RNeasy Plus Mini Kit (Qiagen). cDNA was synthesized using the iScript cDNA synthesis kit (Bio-Rad) using 500ng of RNA per cDNA reaction. qPCR reactions were prepared using the KAPA SYBR Fast Universal 2x quantitative PCR kit using 0.5 μl of cDNA per reaction in a 20 μl total reaction volume. Gene expression was normalized to the expression of the gene *ACTB* for human cell lines and the mouse ortholog (*actb*) for mouse cell lines as an internal housekeeping gene. Cycling conditions were as follows: 95° for 1 minute, 95° for 10 seconds and 60° for 30 seconds. The latter two steps cycled for 50 repeats with plate reads taken after the 60° step. If a sample failed to amplify after 50 cycles, an arbitrary Cq count of 50 was substituted (this occurred in a single instance). For qPCR primer sequences, please see **Supplementary Table 2**.

RNA-seq analysis

For each sample, 100 ng of total RNA was DNase treated with Turbo DNase (ThermoFisher Scientific) at 37C for 30 minutes and then cleaned up with Agenocourt RNAClean XP Beads (Beckman Coulter). The RNA samples were polyA selected using Dynabeads mRNA Direct Purification Kit (ThermoFisher Scientific). RNA-Seq Libraries were constructed using Maxima H Minus First Strand Synthesis Kit (ThermoFisher Scientific) with random hexamers and then the NEBNext mRNA Second Strand Synthesis Module (New England Biolabs). The resulting cDNA was cleaned up with Agencourt AMPure XP Beads (Beckman Coulter) and then went into the Nextera XT DNA Library Prep Kit (Illumina). Final Libraries were once again cleaned with Agencourt AMPure XP Beads and analyzed on a BioAnalyzer using a High Sensitivity DNA Analysis Kit (Agilent). Libraries were quantified using a Qubit dsDNA HS Assay Kit (ThermoFisher Scientific), pooled, and and run on one lane of an Illumina HiSeq 2500 using 2×25 bp paired end reads. Reads were aligned to the hg19 UCSC Known Genes annotations using RSEM v1.2.1 and were analyzed in Python[37] Histograms showing the distribution of fold-changes in gene expression (activator/guide control). Genes were filtered to include only those with TPM > 0.000001. Arrows indicate fold change of *HBG1* for the indicated activator. Average of two biological replicates shown. Differential gene-expression analysis was done using the Voom[38] and Limma[39] packages in R for all genes with ≥0.000001. TPM mapped reads in each replicate, where differential expression was defined by a Benjamini-Hochberg adjusted P value of <0.05 and fold change of >10 or <0.1.”

Drosophila melanogaster transfections

S2R+ cells were transfected in 24-well plates using Effectene reagent (Qiagen). Each well was transfected with 50 ng of dCas9 component, 200 ng of any additional activator component, and 15 ng of sgRNA, and empty vector was used to equalize the total amount of

DNA in each transfection to 265 ng. For multiplex reactions, 15 ng of each sgRNA were added.

Total RNA was collected three days after transfection, and qPCR was conducted as previously described.[40]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to thank S. Vora, A. Tung, M.K. Cromer, and all the members of the Church and Collins lab for helpful discussions and technical assistance. G.C. acknowledges support from the US National Institutes of Health National Human Genome Research Institute grant P50 HG005550, and the Wyss Institute for Biologically Inspired Engineering. In addition, A.C.was funded by the National Cancer Institute grant 5T32CA009216-34, R.C. was funded by a Banting postdoctoral fellowship from the Canadian Institutes of Health Research and J.J.C. was supported by the Defense Threat Reduction Agency grant HDTRA1-14-1-0006. B.E.-C. acknowledges funding from the National Institutes of Health (NIH) under the Ruth L. Kirschstein National Research Service Award F32GM113395 from the NIH General Medical Sciences Division. We would also like to thank J. Lee, P. Mali, and S. Shipman for gifting us cell lines.

References

1. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol. 2014; 32:347–355. [PubMed: 24584096]
2. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell. 2014; 157:1262–1278. [PubMed: 24906146]
3. Garneau JE, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature. 2010; 468:67–71. [PubMed: 21048762]
4. Jinek M, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012; 337:816–821. [PubMed: 22745249]
5. Mali P, et al. RNA-guided human genome engineering via Cas9. Science. 2013; 339:823–826. [PubMed: 23287722]
6. Cong L, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013; 339:819–823. [PubMed: 23287718]
7. Mali P, et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 2013; 31:833–838. [PubMed: 23907171]
8. Sapranauskas R, et al. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res. 2011; 39:9275–9282. [PubMed: 21813460]
9. Gilbert LA, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013; 154:442–451. [PubMed: 23849981]
10. Perez-Pinera P, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat. Methods. 2013; 10:973–976. [PubMed: 23892895]
11. Maeder ML, et al. CRISPR RNA-guided activation of endogenous human genes. Nat. Methods. 2013; 10:977–979. [PubMed: 23892898]
12. Konermann S, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature. 2015; 517:583–588. [PubMed: 25494202]
13. Chavez A, et al. Highly efficient Cas9-mediated transcriptional programming. Nat. Methods. 2015; 12:326–328. [PubMed: 25730490]
14. Zalatan JG, et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell. 2015; 160:339–350. [PubMed: 25533786]
15. Gilbert LA, et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell. 2014; 159:647–661. [PubMed: 25307932]

16. Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell. 2014; 159:635–646. [PubMed: 25307933]
17. Chakraborty S, et al. A CRISPR/Cas9-based system for reprogramming cell lineage specification. Stem Cell Rep. 2014; 3:940–947.
18. Hilton IB, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. 2015; 33:510–517. [PubMed: 25849900]
19. Cheng AW, et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. 2013; 23:1163–1171. [PubMed: 23979020]
20. Dominguez AA, Lim WA, Qi LS. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. Nat. Rev. Mol. Cell Biol. 2016; 17:5–15. [PubMed: 26670017]
21. La Russa MF, Qi LS. The New State of the Art: Cas9 for Gene Activation and Repression. Mol. Cell. Biol. 2015; 35:3800–3809. [PubMed: 26370509]
22. Thakore PI, Black JB, Hilton IB, Gersbach CA. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. Nat. Methods. 2016; 13:127–137. [PubMed: 26820547]
23. Wu X, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat. Biotechnol. 2014; 32:670–676. [PubMed: 24752079]
24. Maeder ML, et al. Robust, synergistic regulation of human gene expression using TALE activators. Nat. Methods. 2013; 10:243–245. [PubMed: 23396285]
25. Perez-Pinera P, et al. Synergistic and tunable human gene activation by combinations of synthetic transcription factors. Nat. Methods. 2013; 10:239–242. [PubMed: 23377379]
26. Oehler S, Müller-Hill B. High Local Concentration: A Fundamental Strategy of Life. J. Mol. Biol. 2010; 395:242–253. [PubMed: 19883663]
27. Gori JL, et al. Delivery and Specificity of CRISPR/Cas9 Genome Editing Technologies for Human Gene Therapy. Hum. Gene Ther. 2015; 26:443–451. [PubMed: 26068008]
28. Yu K, Liu C, Kim B-G, Lee D-Y. Synthetic fusion protein design and applications. Biotechnol. Adv. 2015; 33:155–164. [PubMed: 25450191]
29. Poss ZC, Ebmeier CC, Taatjes DJ. The Mediator complex and transcription regulation. Crit. Rev. Biochem. Mol. Biol. 2013; 48:575–608. [PubMed: 24088064]
30. Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. 2012; 13:613–626. [PubMed: 22868264]
31. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. 2003; 33(Suppl):245–254. [PubMed: 12610534]
32. Chari R, Mali P, Moosburner M, Church GM. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. Nat. Methods. 2015; 12:823–826. [PubMed: 26167643]
33. Doench JG, et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat. Biotechnol. 2014; 32:1262–1267. [PubMed: 25184501]
34. Gibson DG. Enzymatic assembly of overlapping DNA fragments. Methods Enzymol. 2011; 498:349–361. [PubMed: 21601685]
35. Gibson DG, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods. 2009; 6:343–345. [PubMed: 19363495]
36. Port F, Chen H-M, Lee T, Bullock SL. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proc. Natl. Acad. Sci. U. S. A. 2014; 111:E2967–2976. [PubMed: 25002478]
37. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12:323. [PubMed: 21816040]
38. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014; 15:R29. [PubMed: 24485249]
39. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. 2004; 3 Article3.

40. Lin S, Ewen-Campen B, Ni X, Housden BE, Perrimon N. In Vivo Transcriptional Activation Using CRISPR/Cas9 in Drosophila. Genetics. 2015; 201:433–442. [PubMed: 26245833]

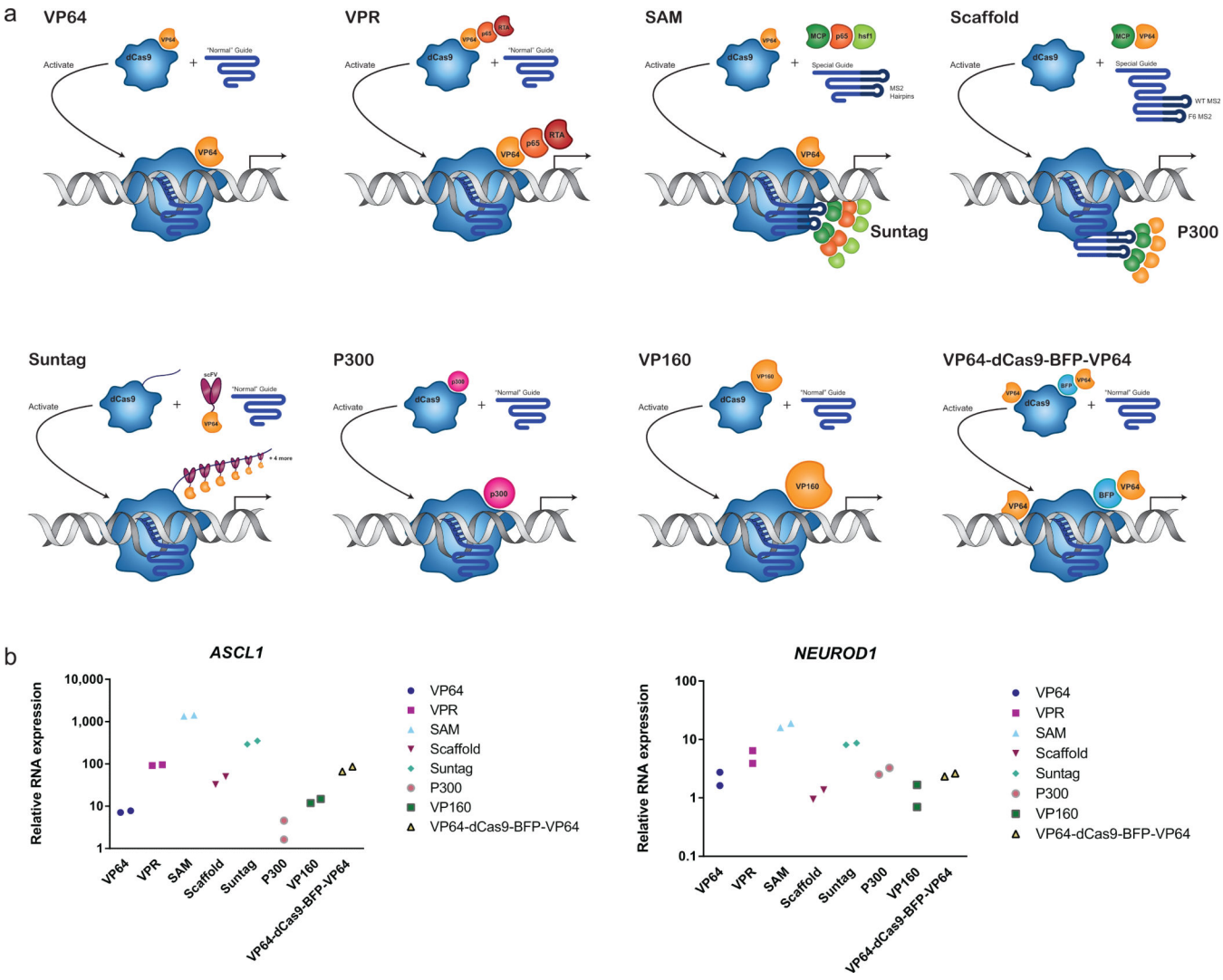

Figure 1. Initial tests of all second-generation activators on endogenous genes in HEK293T cells

(**a**)dCas9-VP64 and dCas9-VPR both work via activation domains fused to the C-terminus of Cas9. SAM uses dCas9-VP64, but recruits more activation domains to the gRNA. Scaffold recruits multiple copies of VP64 to the gRNA. Suntag uses single chain antibodies to recruit multiple copies of VP64 to the peptide tail. P300 uses the catalytic core of the epigenetic modifier fused to dCas9 to modify the chromatin around the promoter to drive transcription. VP160 is the direct fusion of 10 repeats of VP16 protein to dCas9 instead of the usual four that makes up VP64. VP64-dCas9-BFP-VP64 drives transcription via the fusion of VP64 to both the N and C-termini of Cas9 (**b**)Data indicate the mean + s.e.m (n = 2 independent transfections).

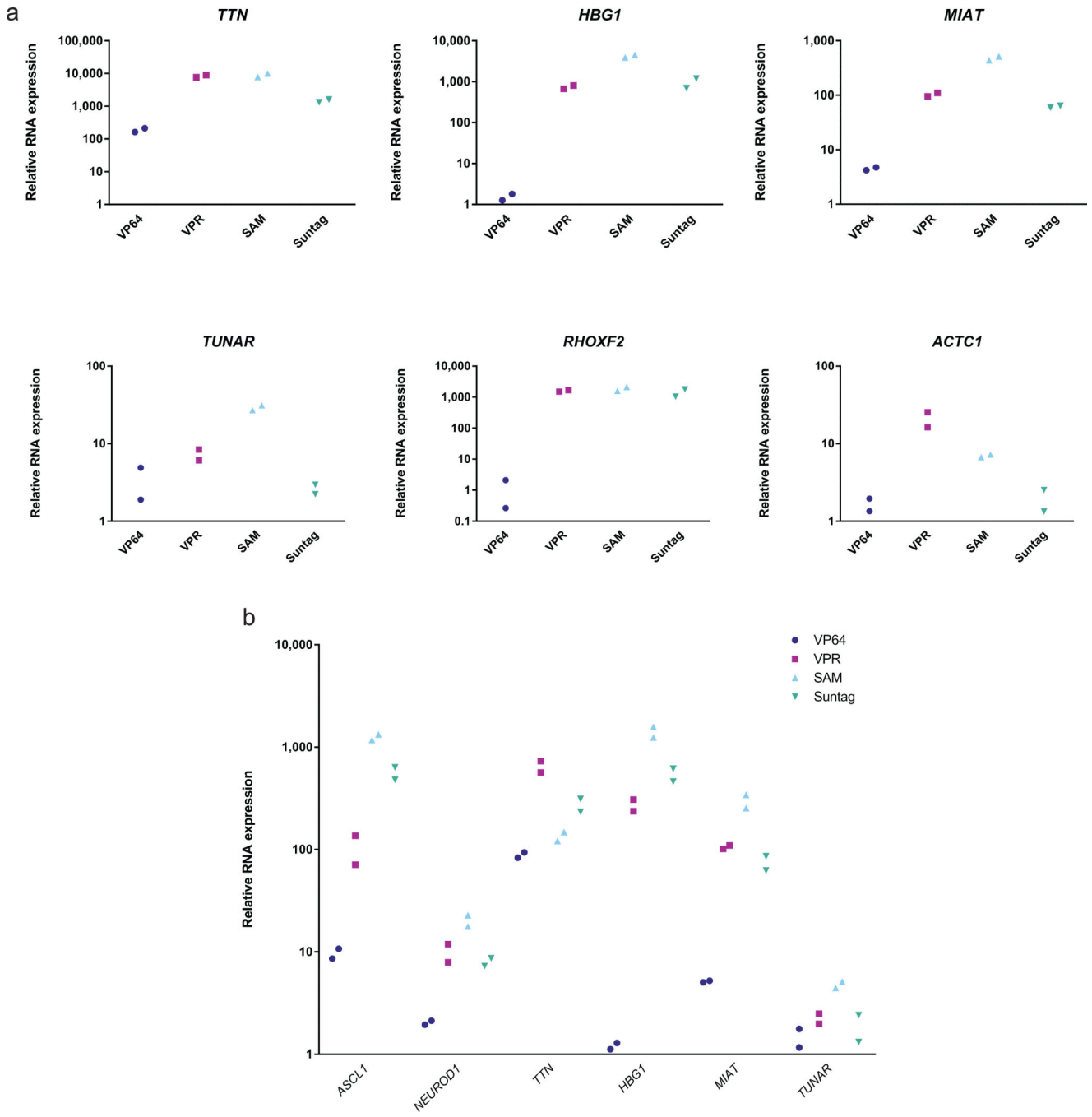

Figure 2. Activation of endogenous genes in HEK 293T cells

(**a**) RNA expression analysis on 6 endogenous human genes. Data indicate the mean + s.e.m (n = 2 independent transfections). (**b**) Multiplexed activation of six endogenous human genes. Data indicate the mean + s.e.m (n = 2 independent transfections).

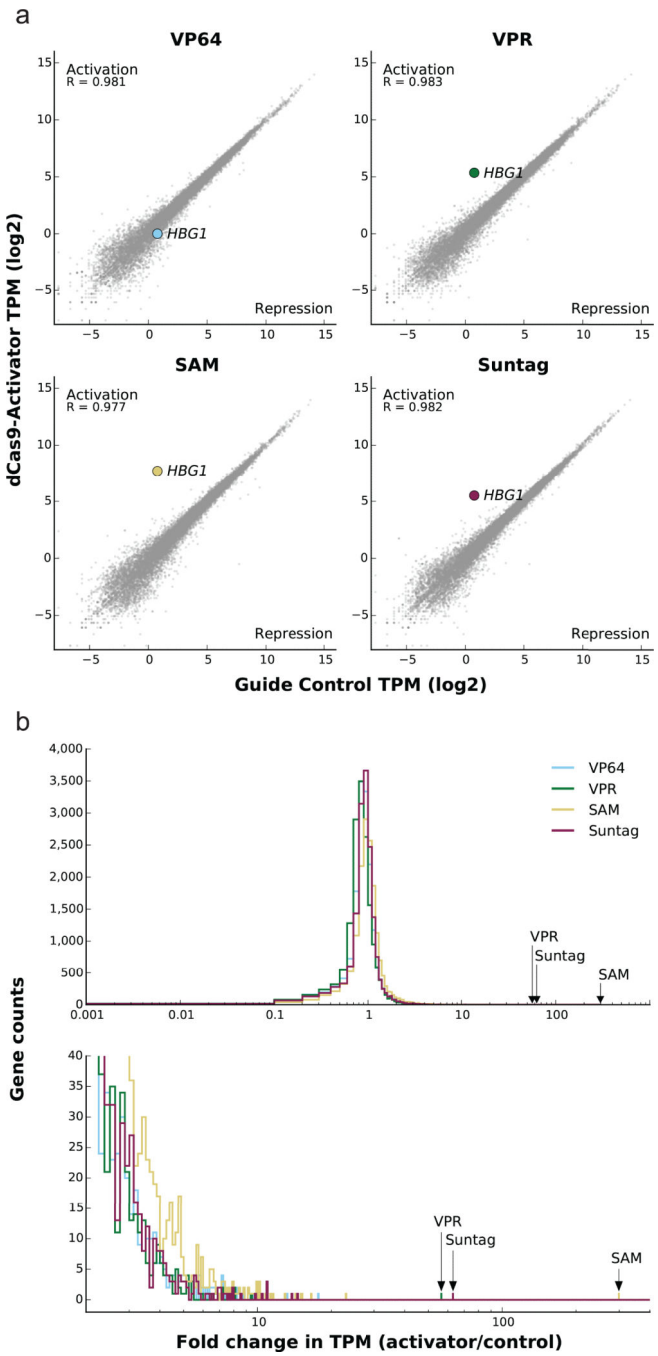

Figure 3. Evaluation of activator specificity by RNA sequencing

(**a**) Gene expression levels (log_2TPM, Transcripts Per Million) in cells transfected with the indicated activator targeting *HBG1* (y axis) vs. expression in cells transfected with guide RNA only (x axis). R indicates Pearson's correlation coefficient, calculated for log-transformed values on all genes except *HBG1*. Genes with 0 TPM in both replicates of either the activator or control were excluded before log transformation. Average of two biological replicates shown. (**b**) Histograms showing the distribution of fold-changes in gene expression (activator/guide control). Genes were filtered to include only those with TPM >

0.1. Bottom panel is zoomed in on the range 2-400. In both panels arrows indicate fold-change of *HBG1* for the indicated activator. Average of two biological replicates shown.

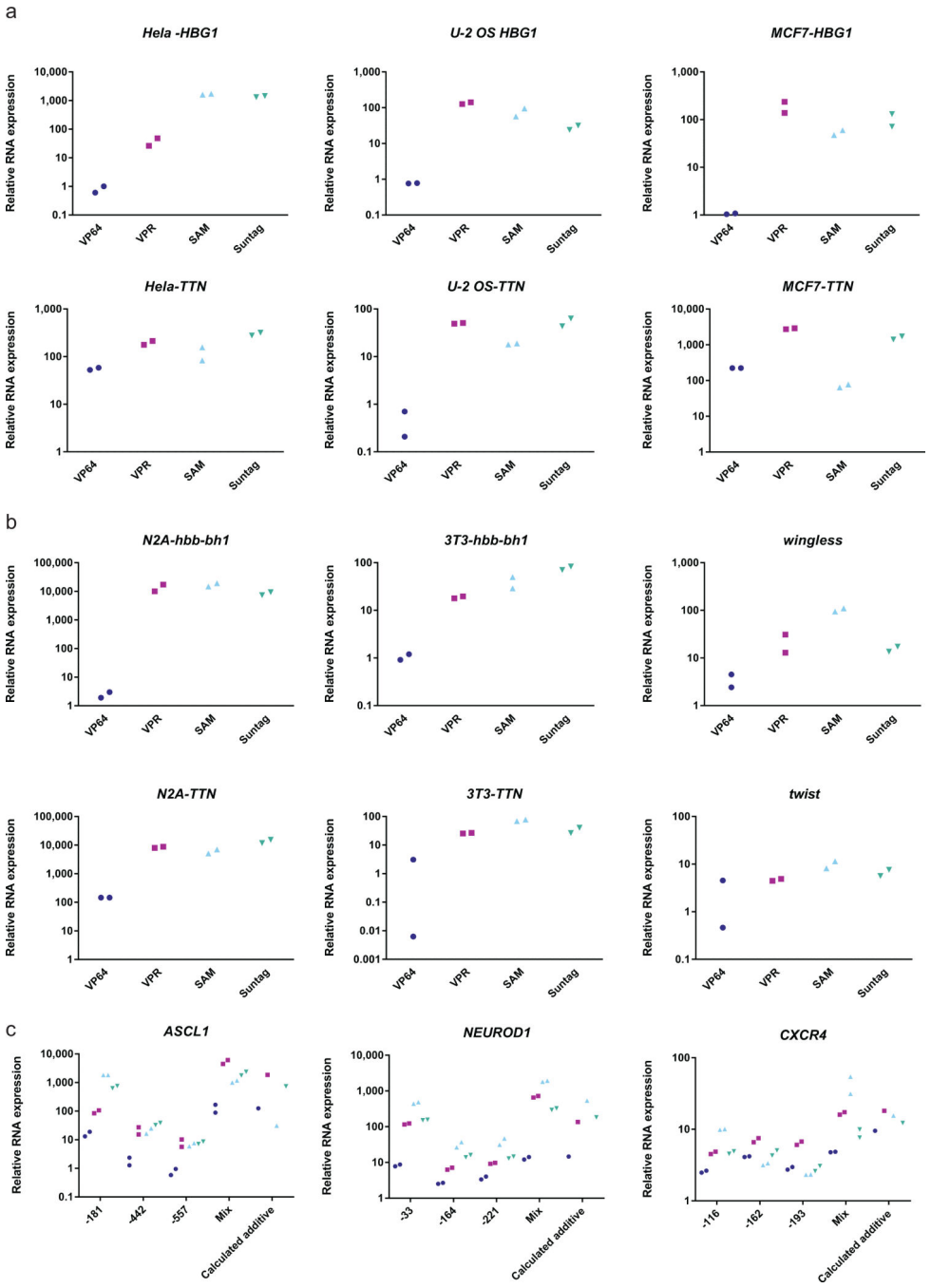

Figure 4. Activation of endogenous genes in alternative human, mouse and fly cell lines and effects of multiple guides

(**a**) Activation of endogenous genes in Hela, U-2 OS and MCF7 cells. Data indicate the mean + s.e.m (n = 2 independent transfections). (**b**) Activation of endogenous genes in mouse and fly. Data indicate the mean + s.e.m (n = 2 independent transfections) (**c**) Samples were transfected with the indicated guide or mixtures of guides. Theoretical sum indicates sum of the relative RNA expression for each activator of the individual gRNAs. Data indicate mean +S.E.M. (n = 2 independent transfections)