広島大学学位請求論文

Development and optimization of CRISPR–Cas9-based artificial transcription activator systems

（CRISPR–Cas9 を基盤とする人工転写活性化システムの開発と最適化）

２０２１年

広島大学大学院理学研究科
数理分子生命理学専攻

國井 厚志

学位論文

國井 厚志

主論文

Contents

General introduction

Genome editing technologies have made a major contribution to the advance of various fields of research. The fundamental platforms for genome editing recognize specific DNA sequences and induce DNA double-strand breaks (DSBs). Chimeric protein-based systems, such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), use repetitive peptide modules and FokI nuclease domains (NDs) for programmed binding to the target DNA sites and the induction of DSBs, respectively (Kim et al., 1996; Miller et al., 2011). Recently, the clustered regularly interspaced short palindromic repeat (CRISPR)–CRISPR-associated protein 9 (Cas9) system, composed of a monomeric protein and a short RNA, was developed (Cong et al., 2013; Mali et al., 2013). In this system, a complex of Cas9 nuclease and single-guide RNA (sgRNA) binds to the target DNA sequence complementary to the sgRNA and induces a DSB via the nuclease activity of Cas9 (Fig. 1a). Because the expression vector of CRISPR–Cas9 is customizable only by the insertion of sgRNA template oligo DNAs, the CRISPR–Cas9 system can be easily constructed and multiplexed (Ran et al., 2013; Sakuma et al., 2014). After the induction of DSBs, simple knockout with small insertions and deletions (indels), or knock-in of exogenous gene cassettes occurs in a manner driven by endogenous DSB repair pathways (Liu et al., 2019).

Genome editing platforms can also be used for various applications other than the induction of DSBs, including transcriptional activation and repression (Beerli et al., 1998; Gilbert et al., 2013; Miller et al., 2011), epigenomic modification (Hilton et al., 2015), chromosome visualization (Chen et al., 2013), and base editing (Komor et al., 2016). In the ZFN- and TALEN-derived approaches, the effector molecules of interest are fused to zinc finger (ZF) arrays and transcription activator-like effector (TALE) modules, instead of the FokI NDs (Beerli et al., 1998; Miller et al., 2011). For CRISPR–Cas9-based tools, in contrast, the effector

molecules are fused to the nuclease-inactive Cas9 (dCas9) and recruited to the target sequences (Gilbert et al., 2013) (Fig. 1b).

Artificial transcription activators, one of the technologies derived from genome editing, are a focus of particular attention due to their practicality for various applications, such as gene function analysis (Gilbert et al., 2014), direct cell reprogramming (Chakraborty et al., 2014), and treating or modeling various diseases (Garcia-Bloj et al., 2016; Li et al., 2015). As the first-generation system, an activation domain such as VP64 was fused to dCas9 (Gilbert et al., 2013) (Figure 2a). However, this simple fusion strategy could reportedly not reactivate some strongly silenced genes (Chavez et al., 2016). To improve the efficacy of transcriptional activation, researchers have developed second-generation systems, in which multiple effector molecules are recruited to the target sites via a variety of approaches. For example, VPR is a chimeric activation domain, in which three types of effector—VP64, p65, and Rta—are tandemly fused to dCas9 (Chavez et al., 2015) (Fig. 2b, top). In contrast, the synergistic activation mediator (SAM) is a *trans*-accumulation system, in which the modified sgRNA (sgRNA2.0) harbors MS2 protein-binding motifs and the MS2 coat proteins (MCPs), fused with activation domains, p65 and HSF1, are recruited to the target sequences via MS2–MCP binding (Konermann et al., 2015) (Fig. 2b, middle). dCas9–SunTag is also a *trans*-accumulation platform, using repeats of GCN4 epitopes and anti-GCN4 minimal antibodies (scFv) fused with activation domains (Tanenbaum et al., 2014) (Fig. 2b, bottom). Although these systems commonly showed higher activity than dCas9–VP64, which system was most effective reportedly varied depending on the tested cell line, target gene, and target sequence (Chavez et al., 2016), suggesting that there is room for further improvement. As described in Chapter 1, I constructed the novel tree-shaped platform, combining the SAM and SunTag systems, and compared it with the previous systems, targeting multiple genes and using multiple human cell lines.

Other than MS2–MCP and GCN4–scFv binding, several RNA–protein binding and protein tagging platforms have been reported. PP7 motif–PP7 coat protein (PCP), boxB motif–

λN22 protein (λN22), and com motif–Com protein (Com) are also RNA–protein binding systems, which have been incorporated into sgRNA-based effector accumulation tools, including for transcriptional modulation and chromosome visualization (Ma et al., 2016; Zalatan et al., 2015). In contrast, sfGFP11–sfGFP1–10, used in one of the strategies involving split fluorescent proteins, and gp41–Nanobody (Nb) binding have been used for dCas9-fused protein tag systems (Boersma et al., 2019; Kamiyama et al., 2016). As described in Chapter 2, I incorporated these additional systems into the novel tools mentioned in Chapter 1, to further optimize the CRISPR–dCas9-based transcriptional activation systems.

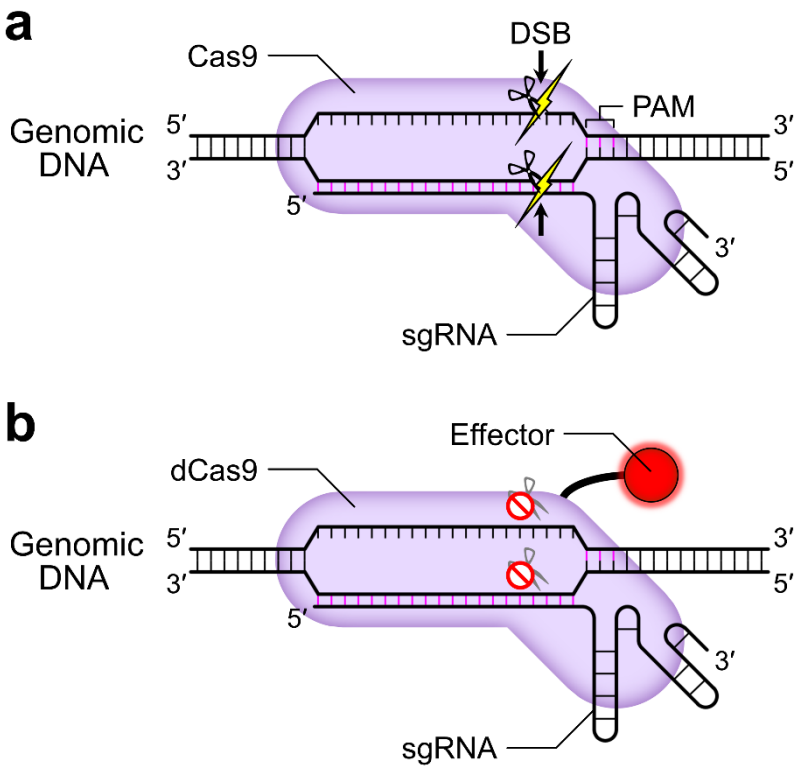

Figure 1. Schematic illustrations of CRISPR–Cas9 and derivative technologies. a Overview of the CRISPR–Cas9 system. A complex of Cas9 and sgRNA binds to the target sequence and induces a DSB. PAM: protospacer adjacent motif (PAM), recognized by Cas9. For the most widely used *Streptococcus pyogenes* (*Sp*) Cas9, the 5′-NGG-3′ PAM is required. Generalized illustration of the CRISPR-based derivative technologies. An effector domain is

fused to dCas9 and recruited to the target sequence. Note that some technologies such as base editing require Cas9 nickase (nCas9) instead of dCas9.

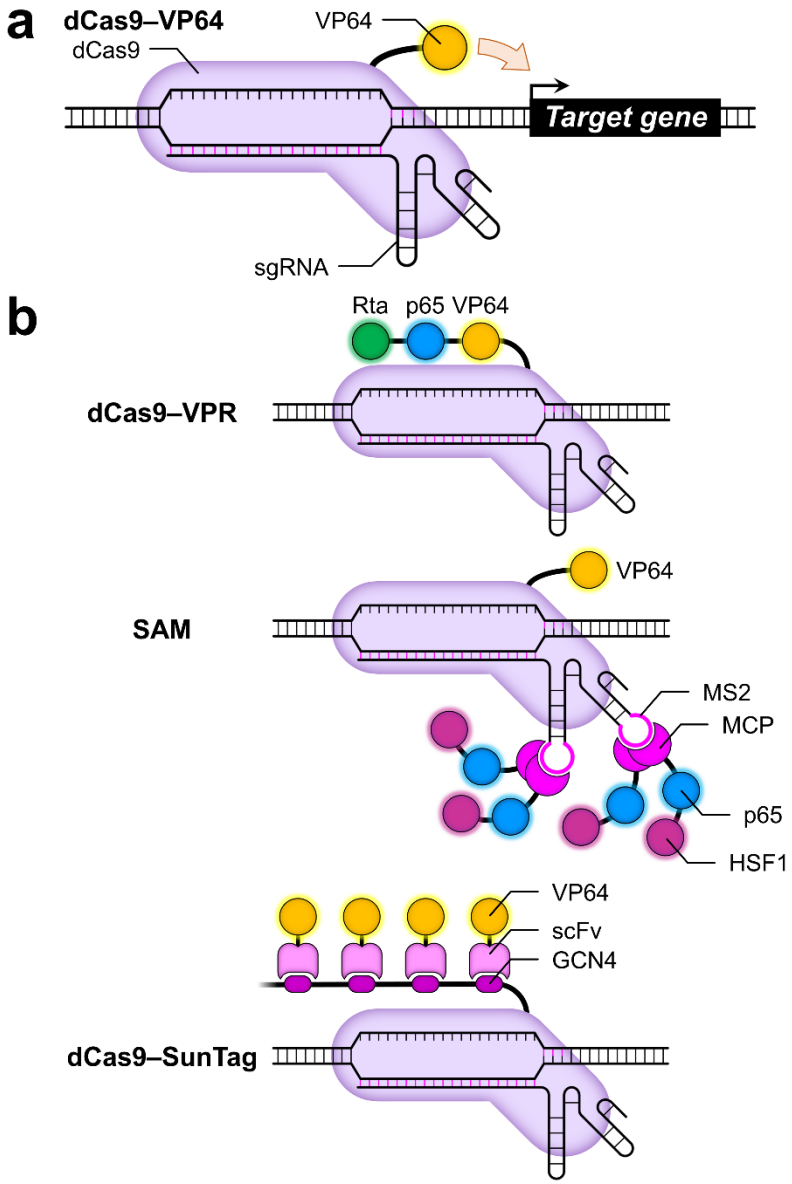

Figure 2. Schematic illustrations of the CRISPR–Cas9-based artificial transcription activators. a The first-generation system, dCas9–VP64. **b** Three types of second-generation system: **top**, dCas9–VPR system; **middle**, SAM system; **bottom**, dCas9–SunTag system.

References

Beerli, R.R., Segal, D.J., Dreier, B., and Barbas, C.F. (1998). 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. Proc. Natl. Acad. Sci. U. S. A. *95*, 14628–14633.

Boersma, S., Khuperkar, D., Verhagen, B.M.P., Sonneveld, S., Grimm, J.B., Lavis, L.D., and Tanenbaum, M.E. (2019). Multi-Color Single-Molecule Imaging Uncovers Extensive Heterogeneity in mRNA Decoding. Cell *178*, 458–472.

Chakraborty, S., Ji, H., Kabadi, A.M., Gersbach, C.A., Christoforou, N., and Leong, K.W. (2014). A CRISPR/Cas9-based system for reprogramming cell lineage specification. Stem Cell Reports *3*, 940–947.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B., Tuttle, M., Iyer, E., Kiani, S., Guzman, C., Wiegand, D., Ter-Ovanesyan, D., et al. (2015). Highly-efficient Cas9-mediated transcriptional programming. Nat. Methods *12*, 326–328.

Chavez, A., Tuttle, M., Pruitt, B.W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S.J., Cecchi, R.J., Kowal, E.J.K., Buchthal, J., et al. (2016). Comparison of Cas9 activators in multiple species. Nat. Methods *13*, 563–567.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479–1491.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Garcia-Bloj, B., Moses, C., Sgro, A., Plani-Lam, J., Arooj, M., Duffy, C., Thiruvengadam, S., Sorolla, A., Rashwan, R., Mancera, R.L., et al. (2016). Waking up dormant tumor suppressor genes with zinc fingers, TALEs and the CRISPR/dCas9 system. Oncotarget *7*, 60535–60554.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442–451.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Kamiyama, D., Sekine, S., Barsi-Rhyne, B., Hu, J., Chen, B., Gilbert, L.A., Ishikawa, H., Leonetti, M.D., Marshall, W.F., Weissman, J.S., et al. (2016). Versatile protein tagging in cells with split fluorescent protein. Nat. Commun. *7*, 11046.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. Proc. Natl. Acad. Sci. U. S. A. *93*, 1156–1160.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420–424.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

Li, K., Pang, J., Cheng, H., Liu, W.P., Di, J.M., Xiao, H.J., Luo, Y., Zhang, H., Huang, W.T., Chen, M.K., et al. (2015). Manipulation of prostate cancer metastasis by locus-specific modification of the CRMP4 promoter region using chimeric TALE DNA methyltransferase and demethylase. Oncotarget *6*, 10030–10044.

Liu, M., Rehman, S., Tang, X., Gu, K., Fan, Q., Chen, D., and Ma, W. (2019). Methodologies for improving HDR efficiency. Front. Genet. *9*, 691.

Ma, H., Tu, L.C., Naseri, A., Huisman, M., Zhang, S., Grunwald, D., and Pederson, T. (2016). Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. Nat. Biotechnol. *34*, 528–530.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., et al. (2011). A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. *29*, 143–150.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., and Yamamoto, T. (2014). Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. Sci. Rep. *4*, 5400.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell *159*, 635–646.

Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi, L.S., et al. (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell *160*, 339–350.

Chapter 1

Development of a novel system for transcriptional activation named TREE

Abstract

In the past few years, several types of artificial transcriptional activator, based on CRISPR–Cas9, have been developed and refined. Of these, in SAM and SunTag systems, the effector proteins, expressed in *trans*, can be recruited to the target sites via the MS2 RNA-binding system and GCN4–scFv antibody system, respectively. Here, I report a strong transcriptional activation system achieved by fusing GCN4 repeat to MCP to accumulate numbers of activators, fused to scFv antibodies. By targeting the *CDH1* gene, I show that the novel system, named "TREE," results in a greater effect of activating exogenous reporter and endogenous gene. Moreover, by targeting another gene, *RANKL*, I consistently show the superiority of the TREE system with fewer sgRNAs compared with the conventional systems. My TREE system is a promising tool for transcriptional activation and can potentially contribute to other dCas9-mediated technologies such as epigenome editing and chromosome visualization.

Introduction

Recently, there has been rapid progress in research on the development and application of genome editing technology. As one of its derivative technologies, programmable regulation of gene expression has also been achieved in a site-specific manner. This system is particularly useful for direct cell reprogramming and for modeling and treating human diseases such as cancers (Chakraborty et al., 2014; Cho et al., 2015; Gao et al., 2013; Garcia-Bloj et al., 2016; Li et al., 2015). As the first-generation tool for activating specific gene expression (i.e., artificial transcriptional activators), an activation domain such as VP64 is fused to specific DNA-binding domains, including zinc-finger array, transcription activator-like effector, and catalytically inactive dCas9 (Beerli et al., 1998; Gilbert et al., 2013; Miller et al., 2011). Among them, the dCas9-based system is especially scalable because multiple sites can be simultaneously targeted, by only expressing multiple single-guide RNAs (sgRNAs).

Based on this, several groups developed second-generation tools, which can mediate stronger activations. For example, the chimeric activator called "VPR" consisting of three types of activation mediator (VP64, p65, and Rta) was shown to have a stronger activation effect than VP64 (Chavez et al., 2015). Alternatively, activators could be provided in *trans* and recruited to the target sequences using RNA-protein or protein-protein interactions. In the SAM (Synergistic Activation Mediator) system, modified sgRNAs harboring MS2 stem loops (sgRNA2.0) were used and activators fused to MCPs were recruited at the dCas9–VP64/sgRNA2.0-binding sites (Konermann et al., 2015). In the "dCas9–SunTag" system, a repeat of GCN4 epitopes (SunTag) was connected to dCas9, which recruited the activators fused to scFv antibodies (Tanenbaum et al., 2014). It was also reported that extending the amino acid linkers connecting the individual GCN4 epitopes enabled a large epigenetic effector to efficiently induce DNA demethylation (Morita et al., 2016).

Both the first- and the second-generation activator systems reportedly present a synergistic activation effect when multiple sgRNAs are used for single gene activation (Chavez

et al., 2016; Perez-Pinera et al., 2013). In addition, in accordance with the previous comparative examination, although the SAM system often showed a superior effect among the second-generation systems, the system inducing the highest expression varied depending on the tested cell type, gene, and target sequence (Chavez et al., 2016). These suggest that none of the second-generation systems has become a definitive one. In this study, to further improve the efficiency of activation by expanding the capacity for activator recruitment, I develop a novel system named TREE (Three-component Repurposed technology for Enhanced Expression) and compare its functionality with that of conventional systems.

Results

Concept of the TREE system

In first-generation CRISPR–Cas9-based artificial transcriptional activators, an activation effector such as VP64 is directly fused to dCas9 (Fig. 1-1a) (Gilbert et al., 2013). In contrast, in the most effective second-generation system, the SAM system, sgRNA2.0 is utilized to recruit transcriptional activators fused to MCPs (Fig. 1-1b) (Konermann et al., 2015). Here, I devised a tree-shaped, multiple-tag system to achieve stronger activation (Fig. 1-1c).

In my TREE system, sgRNA2.0 (root) and modified SunTag fused to MCP (branch) were used as primary RNA tag and secondary peptide tag, respectively. On these branches, leaves (i.e., scFv effectors) were designed to bind, accumulating the transcriptional activation domains at the dCas9-binding sites. To construct the highly tandemized GCN4 repeat arrays, I initially synthesized 4× GCN4 repeat with 22 amino acid linkers, incorporating codon usage variations to avoid completely repeated DNA sequences. Then, each 4× repeat was assembled to create MCP–8× GCN4 (22 a.a.-spaced tag; 22sTag), as well as fusing MCP (Fig. 1-2a). Although I also tried to create MCP–16× and 24× GCN4-expressing vectors and all four vectors were successfully constructed (Fig. 1-2b, c), full-length proteins were not produced from the MCP–

16× and 24× GCN4 vectors (Fig. 1-2d). In the MCP–4× and 8× GCN4 vectors, full-length proteins were successfully expressed along with abundantly produced truncated proteins. Along with the construction of branch vectors, I created the root vector by modifying the previously established all-in-one CRISPR–Cas9 vector system (Sakuma et al., 2014). To optimize the vector for the TREE system, I repurposed the system to express multiple sgRNA2.0s and dCas9–VP64 simultaneously (Fig. 1-2e). As the leaf vector expressing *trans*-activator, I adopted an scFv–sfGFP–effector–GB1 framework, in accordance with a previous report (Fig. 1-2f) (Tanenbaum et al., 2014). Regarding the effector, previously characterized chimeric activators, p65–HSF1 and VPR (VP64–p65–Rta), were used. p65–HSF1 was used in the SAM system with direct fusion to MCP (Konermann et al., 2015). VPR was previously used with direct fusion to dCas9 (dCas9–VPR) (Chavez et al., 2015). Using the 22sTag system, which has longer amino acid linkers than the original SunTag (Tanenbaum et al., 2014) and has more epitope arrays than in the reports by Morita et al. (Morita et al., 2016), large effectors such as VPR are expected to accumulate at high levels and efficiently induce transcriptional activation.

Characterization of the TREE system

To characterize the TREE system, I initially designed sgRNAs to activate the transcription of the human *CDH1* gene encoding the E-cadherin protein. The all-in-one vector expressing dCas9–VP64 and five sgRNAs targeting the promoter region of the *CDH1* gene (Fig, 3a, b) was constructed, and a luciferase reporter vector containing the *CDH1* promoter and 5′ UTR, harboring all of the target sequences, was also constructed for the reporter assay (Fig. 1-3c). Regarding the cell line, I chose MIA-PaCa2 cells, in which the expression level of *CDH1* was reportedly quite low (Tang et al., 2016). I first checked the cytotoxicity of the TREE system with the comparison with the previous systems (Fig. 1-4). No significant cytotoxicity was observed in the TREE-introduced cells or in the dCas9–VP64-, SAM-, and dCas9–VPR-introduced cells. Subsequently, I tested basic mode of action of the TREE system by comparing

the activation efficiency of the full set of TREE components to that of control groups lacking one or two components (Fig. 1-5). As expected, weak and strong activation was observed in the dCas9–VP64/sgRNA2.0-expressing and all three vector-introduced groups, respectively.

Next, I investigated whether the conventional first- and second-generation systems and my TREE system showed stronger activation by simultaneously expressing five sgRNAs (Fig. 1-6). In all three systems, dCas9–VP64, SAM (dCas9–VP64 and MCP–p65–HSF1), and my TREE system, the multiplex vectors showed statistically significant activation compared to the case with one sgRNA expression. In addition, even with only one sgRNA, the TREE system induced relatively high activities compared to the other systems. Subsequently, I checked the activity using the different numbers of GCN4 epitopes (4× and 8×) and different types of *trans*-activators (p65–HSF1 and VPR), as well as the conventional MCP–effector systems (Fig. 1-7a). All of the TREE vectors exhibited higher activity than the conventional second-generation MCP–effector vectors. Of these, especially high activation was observed in the samples in which the 8× 22sTag was used. Similar results were observed at different doses of plasmids (Fig. 1-7b).

Following the results of the reporter assays, I attempted to activate endogenous *CDH1* expression in MIA-PaCa2 cells. qRT-PCR analysis revealed that the first-generation dCas9–VP64 could hardly activate transcription, while the conventional second-generation MCP–effector systems and the TREE systems could upregulate the transcriptional level (Fig. 1-8). Notably, all TREE vectors showed a significantly stronger effect than the MCP–effector vectors, although the relationship of the activity levels among the variations of the TREE systems differed from that observed in the reporter assay (e.g., the activation levels using scFv–p65–HSF1 and scFv–VPR were comparable in the reporter assay, while scFv–p65–HSF1 was better than scFv–VPR in the qRT-PCR analysis). I subsequently quantified the protein level of E-cadherin by immunoblotting (Fig. 1-9). The signals of E-cadherin protein were invisible and slightly visible in the dCas9–VP64- and SAM-introduced samples, respectively, while they

were intense in the TREE-introduced samples. Induced expression of E-cadherin protein was also confirmed by immunostaining (Fig. 1-10). Consistent with the results of immunoblotting analysis, the fluorescence signals were almost invisible in mock- and dCas9–VP64-transfected cells, whereas the E-cadherin-positive cells emerged in the groups transfected with the MCP–effector and TREE systems, with a tendency for stronger fluorescence in the TREE-introduced cells than in the SAM-introduced ones. Notably, in the TREE systems, I could directly observe the TREE component-expressing cells by monitoring the green fluorescence derived from sfGFP. Although some of the E-cadherin-positive cells did not show visible green fluorescence, possibly because of the low intensity of fluorescence of sfGFP fused with various domains, most of the clearly visible sfGFP-positive cells showed highly upregulated E-cadherin signals (Fig. 1-10), suggesting that the upregulation of E-cadherin correctly occurred in the transfected cells.

Target gene- and cell type-independent superiority of the TREE system

To investigate whether the superiority of my TREE system is target gene- or cell line-specific, I targeted another gene, *RANKL* (*TNFSF11*), in another cell line, HEK293T. Previously, the transcriptional activation of *RANKL* using dCas9–TET1 and MCP–TET1 was reported (Xu et al., 2016). I chose two of the sgRNAs shown as the most effective in the corresponding paper, targeting 700 and 200 bp upstream of the transcription start site (TSS; Fig. 1-11a, b), although these designs might be suboptimal for activator-mediated enhancement of expression because the activity range of SAM activators, for example, was shown to be maximally active in the −100 to 0 TSS range (Konermann et al., 2015). Then, I constructed a reporter vector containing both target sequences (Fig. 1-11c) and an all-in-one CRISPR vector expressing two sgRNAs and dCas9–VP64, similar to those for the *CDH1* locus.

Consistent with the results of *CDH1* activation, the *RANKL* reporter assay revealed that my TREE systems showed higher activity than the conventional second-generation MCP–

effector systems (Fig. 1-12a), and the dose-dependent effects were also confirmed (Fig. 1-12b). Critically, endogenous expression of *RANKL* mRNA was not significantly activated using both conventional first- and second-generation systems, while my TREE systems could achieve highly upregulated transcription of the endogenous *RANKL* gene (Fig. 1-13). Note that the *RANKL* sgRNAs were suboptimally designed as described above. However, my TREE systems could act as the strong transcriptional activators, even using such sgRNAs. On the other hand, I found that my TREE system slightly decreased cell viability in HEK293T cells, inconsistent with the results obtained using MIA-PaCa2 cells (Fig. 1-14).

Additionally, I performed a direct comparison of my TREE systems and another second-generation system, dCas9–22sTag fusion, similar to the SunTag, in the *RANKL* reporter assay. I additionally constructed two types of all-in-one CRISPR vector, expressing two sgRNAs without MS2 stem loops targeting the *RANKL* promoter and dCas9–4× or –8× GCN4 (22sTag), instead of dCas9–VP64. Thorough analysis of the expression-enhancing activity revealed the clear superiority of my TREE systems over the dCas9–22sTag systems (Fig. 1-15).

Superiority of the TREE system with only one sgRNA

Finally, I investigated the superiority of my TREE system to the conventional dCas9–VP64, SAM, and dCas9–VPR technologies with only one sgRNA. I first selected three sgRNAs, sgRNA #3–5, for the activation of *CDH1*, based on the activation scores determined in Fig. 1-6. Reporter assays revealed that all the TREE vectors with one sgRNA outperformed SAM and dCas9–VPR systems, with the exception that the activation score of MCP–8× GCN4 (22sTag)-containing TREE system with sgRNA #5 was comparable to that of dCas9–VP64 and SAM system (Fig. 1-16a). One sgRNA-derived transcriptional activation was further characterized by endogenous qPCR analysis. Using sgRNA #5, the overall relationship of activation levels with each system was quite similar to that observed in the reporter assays (Fig. 1-16b). Importantly, the average score of MCP–8× GCN4-containing TREE system was about sixfold

greater than that of SAM system, although statistical significance could not be confirmed because of the score variability.

Similar to the *CDH1* locus, I constructed the sgRNA #1 or #2-expressing dCas9–VP64, SAM, and dCas9–VPR, and validated their functionality by reporter assay and qPCR analysis. Significant reporter activation was observed in all the TREE vectors constructed compared to the conventional systems (Fig. 1-17a). It is particularly noteworthy that my TREE vector containing sgRNA #1 retained about a half activity of sgRNA #2-containing TREE vector, while none of the previous systems resulted in high-level activation. My TREE vectors also highly activated the endogenous *RANKL* with sgRNA #2 (Fig. 1-17b), and their average activation levels were higher than those with previous systems. Statistically significant difference of TREE versus previous systems was also observed, except TREE versus dCas9–VPR.

Discussion

In summary, I established a novel hybrid system of the previously characterized SAM and SunTag activators, enabling sequential recruitment of the tag arrays and effector molecules, and showed high-powered transcriptional activation efficacy at two gene loci in different cell lines. My achievements are somewhat contrary to a previous paper by Chavez et al. reporting that various combinations of second-generation systems (e.g., dCas9–10× GCN4 [conventional SunTag] + sgRNA2.0 + MCP–p65–HSF1 + scFv–VP64) did not show any additive or synergistic activation effects (Chavez et al., 2016). One possible explanation for this contradiction is that a simple “addition” of various systems was not effective for further accumulation of the effector molecules, but my tree-shaped hierarchical configuration (i.e., “integration” of SAM and SunTag systems) was effective to highly accumulate *trans*-activators. Chavez et al. just collectively used the previous systems, while I built the high-order system by

using newly created MCP–22sTag proteins as the adapter molecules. In addition, there is still room for improvement in optimizing the expression levels of the TREE components, which might result in greater efficiency of transcriptional activation. In fact, I set the mass ratio of three TREE vectors as 1:1:1 throughout this study, but this ratio might be suboptimal because every "root" (MS2 stem loop) requires two "branches" (MCP–22sTag proteins) and every "branch" requires either four or eight "leaves" (scFv–activators). Thus, further optimization in terms of the stoichiometry of these components will be required to achieve maximum level of activation.

Moreover, based on their architecture, my TREE system might overcome the obstacles that were difficult to solve using the existing methods by exploiting its particular attributes. For example, the conventional MCP–effector system has the ability to distribute several types of effector to independent target loci by using multiple types of RNA-protein interaction (Ma et al., 2016; Zalatan et al., 2015). However, this system has a limit on the number of recruitable effector molecules. In contrast, the dCas9–SunTag system can accumulate more effectors, but it cannot discriminate each locus to assign various effectors. In contrast, using my TREE system, highly accumulated recruitment of different effectors would be achieved in a locus-specific manner, by simultaneously using the 22sTag and other tag proteins fused with multiple RNA-binding proteins (e.g., MCP–22sTag and PCP–another tag).

Another anticipated application of the TREE system is targeted gene activation *in vivo*. Recently, gene activation in mice with adeno-associated virus (AAV)-mediated delivery of a SAM-like system was reported (Liao et al., 2017). AAV has a strict size limit. Thus, direct fusion of large tag arrays or effectors such as SunTag or VPR is not applicable in AAV-mediated delivery. However, such tags and effectors can be supplied independently in my TREE system, in which the lengths of coding sequences of MCP–22sTag and scFv–effectorare capable of loading in AAV: MCP–4× GCN4 (22sTag), 1,218 bp; MCP–8× GCN4 (22sTag), 1,710 bp;

scFv–p65–HSF1, 2,823 bp; scFv–VPR, 3,441 bp. Therefore, my system would also be suitable for *in vivo* application, although the actual applicability has not yet experimentally confirmed.

My investigations also revealed some challenges and open questions. First, the DNA and amino acid sequences of MCP–22sTag might be reconsidered to achieve further enhancement in accordance with the Western blot analysis of the corresponding proteins. In previous papers reporting original and modified SunTag systems (Morita et al., 2016; Tanenbaum et al., 2014), their protein expression was not examined. Therefore, it should be clarified whether this phenomenon is 22sTag-specific. Second, the cytotoxicity analyses of the TREE system using two cell lines resulted in different outcomes. No significant toxicity was observed in MIA-PaCa2 cells, whereas the system components showed slight toxicity in HEK293T cells. This cell type-specific toxicity should be further characterized. Third, both the sufficient level of transcriptional activation to induce protein upregulation and a ceiling to the level of transcriptional activation regardless of the basal expression of the target transcript, observed in the previous paper (Konermann et al., 2015), were not examined in the context of this study. Fourth, the comparative analyses between my TREE system and the previous systems should be more thoroughly examined at various target loci in various cell lines to confirm the robust superiority of my system further. Especially with only one sgRNA, there was high variability in endogenous gene expression with the TREE activation. Such variability should be caused by the lack of tight robustness of the current TREE system with one sgRNA. Therefore, the robust activation not with the multiple TREE but with the single TREE will be the future avenue of this technology. Fifth, the specificity of this system should be assessed by comprehensive RNA-seq analysis. These points should be clarified in the future study.

Potential application of the TREE system is not limited to simple transcriptional activation. The MS2- or SunTag-mediated accumulation of various molecules has already been reported in various applications, including targeted DNA demethylation (Morita et al., 2016; Xu et al., 2016), targeted histone modification (Liu et al., 2018), visualization of specific

chromosomal regions (Fu et al., 2016), and directed evolution with saturation mutagenesis (Hess et al., 2016).

Conclusion

My TREE system not only has the potential to be a promising system of artificial transcriptional activator but also would contribute to a broad range of biological analyses assisted by the CRISPR system with various effectors, boosting and adding depth to life science studies.

Materials and Methods

Construction of sgRNA2.0 and dCas9–VP64 expression vectors

The previously established all-in-one CRISPR–Cas9 vector system (Sakuma et al., 2014) was modified to express sgRNA2.0 and dCas9–VP64. To construct the gene-specific vectors, the oligonucleotides for the sgRNA templates, listed in the Sequence 1-1, were annealed and inserted in accordance with a previously reported protocol (Ran et al., 2013). Subsequently, all required sgRNA and dCas9–VP64 expression cassettes were integrated into a single vector using Golden Gate assembly, following a previously described protocol (Sakuma et al., 2017).

Construction of MCP–effector and scFv–effector expression vectors

The coding sequences of MCP–p65–HSF1, VPR (VP64–p65–Rta), and scFv–sfGFP–VP64–GB1 were obtained from Addgene (plasmid #61423, #63798, and #60904). Then, cloning and substitutions were carried out by polymerase chain reaction (PCR) amplification and In-Fusion cloning (Takara) as follows: coding sequence of MCP–p65–HSF1 was cloned into CMV-expression vector. For MCP–VPR, the sequence of p65–HSF1 was substituted by VPR. For scFv–p65–HSF1 and scFv–VPR, the coding sequence of scFv–sfGFP–VP64–GB1 was cloned

into CMV-expression vector. Then, the sequence of VP64 was substituted into p65–HSF1 or VPR.

Construction of MCP–22sTag and sgRNA/dCas9–22sTag expression vectors

The coding sequence of 4× GCN4 (22sTag) was synthesized by gBlocks (IDT). The synthesized sequence is described in the Sequence 1-1. For MCP–4× and 8× GCN4, one or two 4× GCN4 sequences were inserted downstream of MCP. For MCP–16× and 24× GCN4, the sequence of 8× GCN4 was amplified from MCP–8× GCN4 vector. Then, one or two of these sequences were inserted downstream of MCP–8× GCN4. For dCas9–4× and 8× GCN4, the sequences of 4× and 8× GCN4 were amplified from MCP–4× and 8× GCN4 vectors, respectively, and inserted downstream of dCas9. Insertions of sequences were performed using an In-Fusion HD Cloning Kit with the primers listed in the Sequence 1-1.

Construction of reporter vectors

The promoter and 5′ UTR regions of *CDH1* and *RANKL* were amplified from the genomic DNA collected from HEK293T cells. Then, these sequences were inserted upstream of Luc2 coding sequence (Promega) using an In-Fusion HD Cloning Kit.

Cell culture

MIA-PaCa2 cells were maintained in Dulbecco's modified Eagle's medium (DMEM; high glucose) with L-glutamine and Phenol Red (FUJIFILM-Wako), supplemented with 10% fetal bovine serum (FBS; Thermo Fisher Scientific), 2.5% horse serum (Thermo Fisher Scientific), and 1% penicillin–streptomycin (FUJIFILM-Wako). HEK293T and HCT116 cells were maintained in DMEM (high glucose) with L-glutamine and Phenol Red (FUJIFILM-Wako), supplemented with 10% FBS (Thermo Fisher Scientific), 1% minimum essential medium non-essential amino acids (Thermo Fisher Scientific), and 1% penicillin–streptomycin (FUJIFILM-

Wako). All cell lines were tested negatively for mycoplasma contamination using an e-Myco Mycoplasma PCR Detection Kit (iNtRON Biotechnology) and authenticated by short tandem repeat analysis (Takara).

Transfection for the detection of MCP–22sTag proteins

A total of 60,000 cells were transfected with 200 ng of MCP–4×, 8×, 16×, or 24× GCN4 expression vector, or control pcDNA plasmid.

Transfection for the reporter assays

A total of 60,000 cells were transfected with the vectors mixed as follows, using a Lipofectamine LTX reagent (Thermo Fisher Scientific) in a 96-well plate: a 1:1:1 mass ratio of the following three vectors: (1) sgRNAs/dCas9–effector all-in-one vector, sgRNAs/dCas9–*n*× GCN4 all-in-one vector, or pcDNA; (2) MCP–*n*× GCN4 expression vector, MCP–effector expression vector, or pcDNA; and (3) scFv–effector expression vector or pcDNA (50 ng in total for Figs. 1-7a, b, 1-12a, b; 100 ng in total for Figs. 1-5, 1-6, 1-7a, b, 1-12a, b), and 100 and 20 ng of reporter vector and *R*Luc expression vector for reference, respectively.

Transfection for the quantitative PCR and endogenous protein detection analyses

A total of 30,000 cells (for quantitative PCR) or 60,000 cells (for endogenous protein detection) were transfected with the vectors mixed as follows, using a Lipofectamine LTX reagent in a 96-well plate: a 1:1:1 mass ratio of the following three vectors: (1) sgRNAs/dCas9–effector all-in-one vector or pcDNA; (2) MCP–*n*× GCN4 expression vector, MCP–effector expression vector, or pcDNA; and (3) scFv–effector expression vector or pcDNA (200 ng in total). For the endogenous protein detection analyses, untransfected HCT116 cells were also used as positive controls, which were previously characterized as the *CDH1*-positive cells.

Transfection for the cytotoxicity analysis

A total of 30,000 cells were transfected with the vectors mixed as follows, using a Lipofectamine LTX reagent in a 96-well plate: a 1:1:1 mass ratio of the following three vectors: (1) non-targeting dCas9–effector expression vector or pcDNA; (2) MCP–*n*× GCN4 expression vector, MCP–effector expression vector, or pcDNA; and (3) scFv–effector expression vector or pcDNA (200 ng in total).

Luciferase assay

At 24 h post transfection, dual luciferase activity was measured using a Dual-Glo Luciferase Assay System (Promega) on a TriStar LB 941 Multimode Microplate Reader (Berthold Technologies).

Analysis of endogenous CDH1 and RANKL mRNA expression

At 48 h post transfection, cell lysis and reverse transcription were performed using a SuperPrep Cell Lysis & RT Kit for qPCR (Toyobo), in accordance with the manufacturer's instructions. Relative mRNA expression levels were quantified by quantitative real-time PCR (qRT-PCR) using a KOD SYBR qPCR Mix (Toyobo) on a StepOnePlus Real-Time PCR System (Thermo Fisher Scientific). Expression levels of *CDH1* and *RANKL* were normalized by that of *RPL8*. Relative expression changes were calculated using the relative standard curve method. The primers used are listed in the Sequence 1-1.

Immunoblotting

At 24 h post transfection, the cells were collected and seeded onto a six-well plate. After 48 h, the cells were lysed and sonicated. Then, protein concentrations of lysates were measured using a Protein Assay Kit (Bio-Rad). The lysates were re-suspended in Laemmli buffer, denatured for 5 min at 98°C, and separated by Tris-glycine denaturing sodium dodecyl sulfate polyacrylamide

gel electrophoresis. Proteins were blotted onto polyvinylidene fluoride membranes, blocked in 5% milk, and incubated overnight with the following primary antibodies: for endogenous protein detection, anti-E-cadherin (ab40772; Abcam, Cambridge, United Kingdom) or anti-α-tubulin (ab11304; Abcam); for MCP–22sTag protein detection, anti-HA (ab49969; Abcam) at a 1:1,000 (ab40772 and ab11304) or 1:2,000 (ab49969) dilution ratio in Can Get Signal Solution 1 (Toyobo) at 4°C. Subsequently, the proteins were incubated with the corresponding horseradish peroxidase-conjugated secondary antibodies (Thermo Fisher Scientific) at a 1:250 dilution ratio in Can Get Signal Solution 2 (Toyobo) for 1 h at room temperature. Chemiluminescent signals were generated using a SuperSignal West Pico Plus Chemiluminescent Substrate (Thermo Fisher Scientific) and captured on X-ray films (Fujifilm, Tokyo, Japan). The films were scanned, and signal intensities were quantified using ImageJ software (https://imagej.nih.gov/ij/).

Fluorescence immunocytochemistry

At 24 h post transfection, the cells were collected and seeded onto a 24-well plate. After 48 h, the cells were fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) for 15 min at room temperature. After washing with PBS, the cells were permeabilized with 1% Triton X-100 in PBS for 20 min at room temperature, and subsequently rinsed with PBS. The cells were covered with 1% bovine serum albumin in PBS for 1 h at room temperature, and were then incubated overnight with rabbit anti-E-cadherin (ab40772; Abcam) at a 1:100 dilution ratio at 4°C. After washing with PBS, the cells were stained with an Alexa 647-conjugated anti-rabbit secondary antibody (Thermo Fisher Scientific) at a 1:1,000 dilution ratio for 1 h at room temperature. After washing with PBS, nuclei were counterstained with DAPI. After washing with PBS, the images of DIC, DAPI, sfGFP, and E-cadherin (Alexa 647, pseudocolored red) were obtained with an FV-1000D confocal laser scanning microscope (Olympus).

Cytotoxicity analysis

At 48 h post transfection, the cell viability was measured with ATP activity using a CellTiter-Glo Luminescent Cell Viability Assay (Promega) on a TriStar LB 941 Multimode Microplate Reader (Berthold Technologies) according to the manufacturers' instructions.

Statistical analysis

All statistical analyses were performed with a Student's *t*-test.

References

Beerli, R.R., Segal, D.J., Dreier, B., and Barbas, C.F. (1998). 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. Proc. Natl. Acad. Sci. U. S. A. *95*, 14628–14633.

Chakraborty, S., Ji, H., Kabadi, A.M., Gersbach, C.A., Christoforou, N., and Leong, K.W. (2014). A CRISPR/Cas9-based system for reprogramming cell lineage specification. Stem Cell Reports *3*, 940–947.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B., Tuttle, M., Iyer, E., Kiani, S., Guzman, C., Wiegand, D., Ter-Ovanesyan, D., et al. (2015). Highly-efficient Cas9-mediated transcriptional programming. Nat. Methods *12*, 326–328.

Chavez, A., Tuttle, M., Pruitt, B.W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S.J., Cecchi, R.J., Kowal, E.J.K., Buchthal, J., et al. (2016). Comparison of Cas9 activators in multiple species. Nat. Methods *13*, 563–567.

Cho, H.S., Kang, J.G., Lee, J.H., Lee, J.J., Jeon, S.K., Ko, J.H., Kim, D.S., Park, K.H., Kim, Y.S., and Kim, N.S. (2015). Direct regulation of E-cadherin by targeted histone methylation of TALE-SET fusion protein in cancer cells. Oncotarget *6*, 23837–23844.

Fu, Y., Rocha, P.P., Luo, V.M., Raviram, R., Deng, Y., Mazzoni, E.O., and Skok, J.A. (2016). CRISPR-dCas9 and sgRNA scaffolds enable dual-colour live imaging of satellite sequences and repeat-enriched individual loci. Nat. Commun. *7*, 11707.

Gao, X., Yang, J., Tsang, J.C.H., Ooi, J., Wu, D., and Liu, P. (2013). Reprogramming to pluripotency using designer TALE transcription factors targeting enhancers. Stem Cell Reports *1*, 183–197.

Garcia-Bloj, B., Moses, C., Sgro, A., Plani-Lam, J., Arooj, M., Duffy, C., Thiruvengadam, S., Sorolla, A., Rashwan, R., Mancera, R.L., et al. (2016). Waking up dormant tumor suppressor genes with zinc fingers, TALEs and the CRISPR/dCas9 system. Oncotarget *7*, 60535–60554.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442–451.

Hess, G.T., Frésard, L., Han, K., Lee, C.H., Li, A., Cimprich, K.A., Montgomery, S.B., and Bassik, M.C. (2016). Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nat. Methods *13*, 1036–1042.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

Li, K., Pang, J., Cheng, H., Liu, W.P., Di, J.M., Xiao, H.J., Luo, Y., Zhang, H., Huang, W.T., Chen, M.K., et al. (2015). Manipulation of prostate cancer metastasis by locus-specific modification of the CRMP4 promoter region using chimeric TALE DNA methyltransferase and demethylase. Oncotarget *6*, 10030–10044.

Liao, H.K., Hatanaka, F., Araoka, T., Reddy, P., Wu, M.Z., Sui, Y., Yamauchi, T., Sakurai, M., O'Keefe, D.D., Núñez-Delicado, E., et al. (2017). In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans-epigenetic Modulation. Cell *171*, 1495–1507.

Liu, P., Chen, M., Liu, Y., Qi, L.S., and Ding, S. (2018). CRISPR-Based Chromatin Remodeling of the Endogenous Oct4 or Sox2 Locus Enables Reprogramming to Pluripotency. Cell Stem Cell *22*, 252–261.

Ma, H., Tu, L.C., Naseri, A., Huisman, M., Zhang, S., Grunwald, D., and Pederson, T. (2016). Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. Nat. Biotechnol. *34*, 528–530.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., et al. (2011). A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. *29*, 143–150.

Morita, S., Noguchi, H., Horii, T., Nakabayashi, K., Kimura, M., Okamura, K., Sakai, A., Nakashima, H., Hata, K., Nakashima, K., et al. (2016). Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. Nat. Biotechnol. *34*, 1060–1065.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., et al. (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat. Methods *10*, 973–976.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., and Yamamoto, T. (2014). Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. Sci. Rep. *4*, 5400.

Sakuma, T., Sakamoto, T., and Yamamoto, T. (2017). All-in-one CRISPR-Cas9/Foki-dCas9 vector-mediated multiplex genome engineering in cultured cells. Methods Mol. Biol. *1498*, 41–56.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell *159*, 635–646.

Tang, H.M., Kuay, K.T., Koh, P.F., Asad, M., Tan, T.Z., Chung, V.Y., Lee, S.C., Thiery, J.P., and Huang, R.J. (2016). An epithelial marker promoter induction screen identifies histone deacetylase inhibitors to restore epithelial differentiation and abolishes anchorage independence growth in cancers. Cell Death Discov. *2*, 16041.

Xu, X., Tao, Y., Gao, X., Zhang, L., Li, X., Zou, W., Ruan, K., Wang, F., Xu, G.L., and Hu, R. (2016). A CRISPR-based approach for targeted DNA demethylation. Cell Discov. *2*, 16009.

Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi, L.S., et al. (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell *160*, 339–350.

Figures

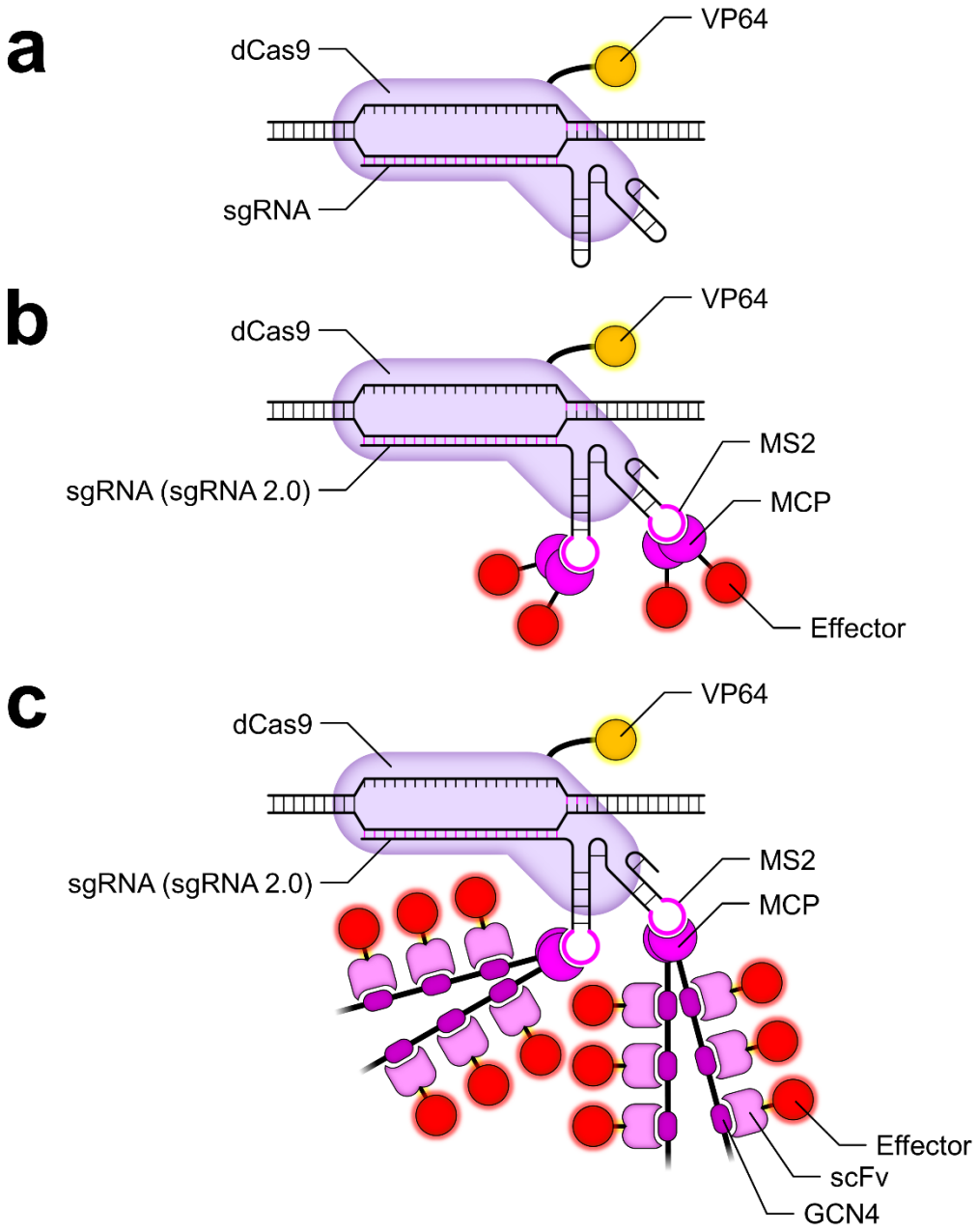

Figure 1-1. Schematics of artificial transcriptional activation systems. (a) The first-generation system consisting of dCas9 fused with VP64 (dCas9–VP64) and sgRNA. **(b)** The second-generation MCP–effector system. MCPs directly fused to the effector molecules bind to MS2 stem loops of sgRNA 2.0. **(c)** The TREE system established in this study. MCP–22sTags are recruited as with the MCP–effector system. Then, scFv antibodies carrying effectors bind to GCN4 epitopes of 22sTags, theoretically resulting in the high accumulation of effector molecules around the target site.

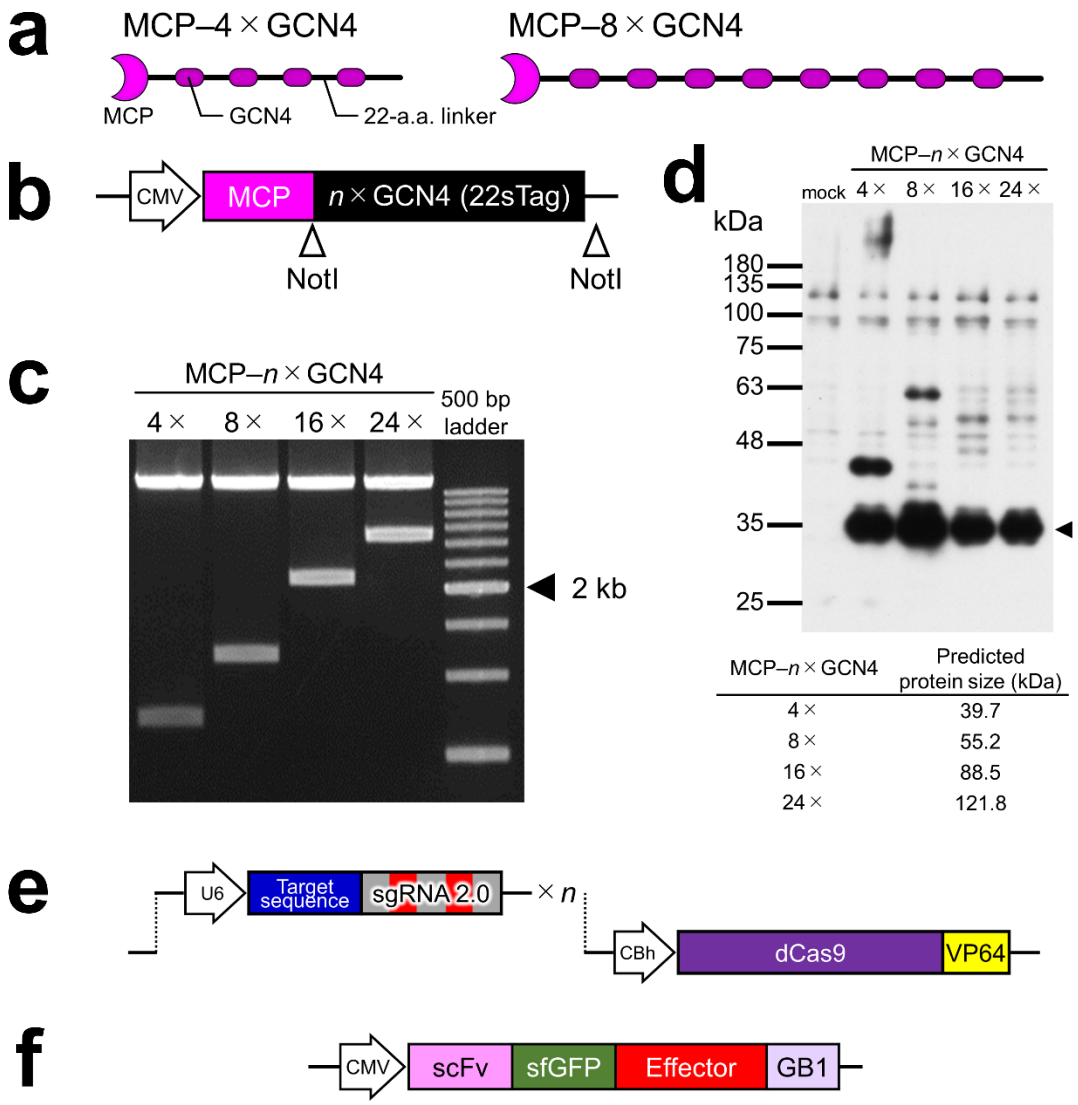

Figure 1-2. Construction of the TREE system. (a) Schematics of MCP–22sTag proteins. 4× or 8× repeats of GCN4 epitopes, each spaced of longer amino acids, are fused to MCPs. **(b)** Schematic of the MCP–22sTag expression vectors. **(c)** Validation of MCP–22sTag vectors by restriction digestion. A gel image of MCP–22sTag vector digested with NotI is shown. **(d)** Western blot analysis of MCP–22sTag proteins expressed in HEK293T cells. 10 μg each of cell lysate was loaded and the corresponding protein was detected with anti-HA antibody. Predicted protein sized are shown at the bottom table. A filled triangle indicates the bands of truncated proteins, possibly containing MCP and partial GCN4 tag arrays. **(e)** Schematic of modified all-in-one CRISPR vector expressing multiple sgRNA2.0s and dCas9–VP64. **(f)** Schematic of the scFv–effector vectors. p65–HSF1 or VPR was used as the effector.

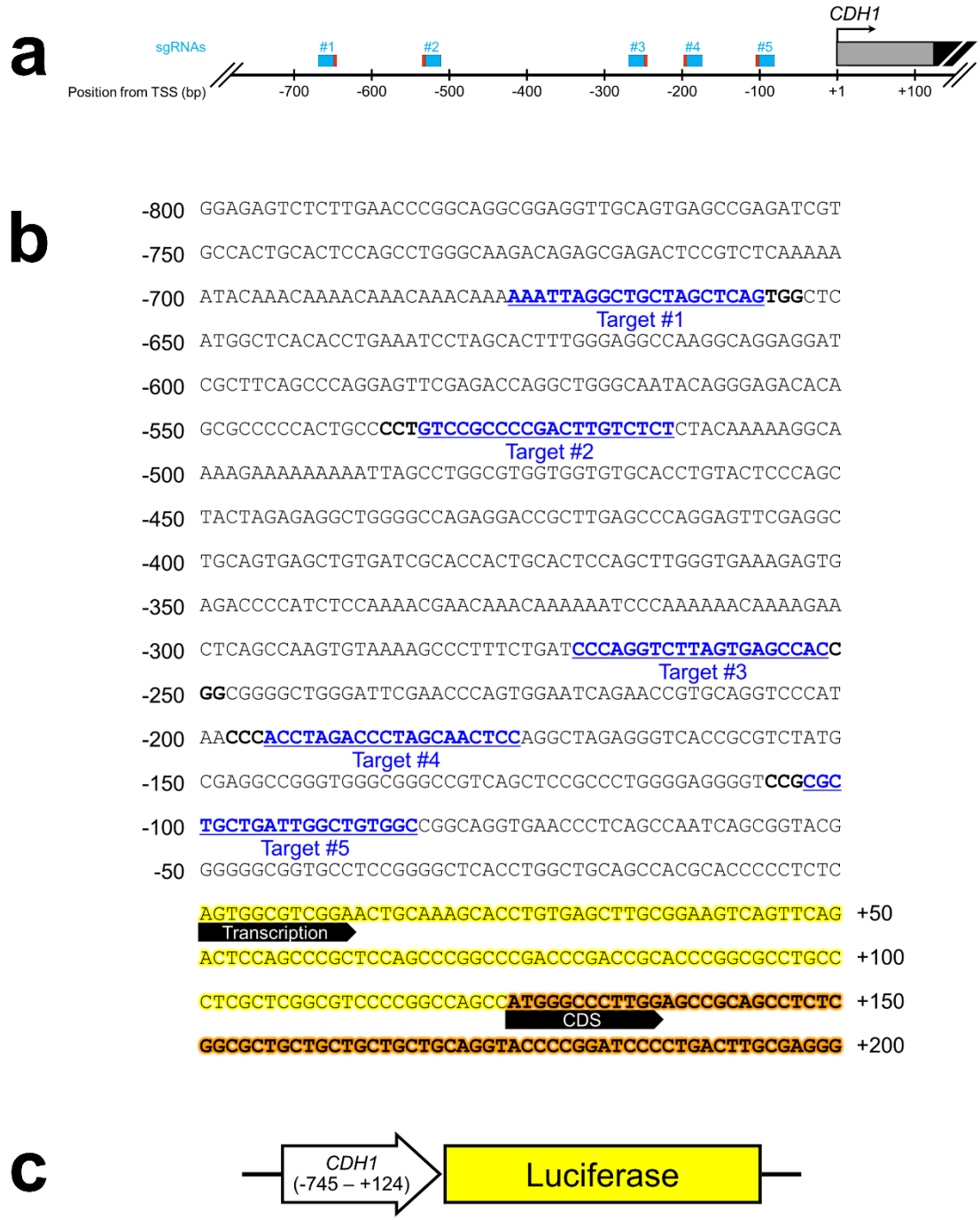

Figure 1-3. Design of the sgRNAs and the reporter for the activation of *CDH1*. (a) Schematic illustration of the positions of sgRNAs used or the activation of *CDH1*. Blue, red, gray, and black boxes show the protospacer, PAM, 5′ UTR, and CDS, respectively. **(b)** Detailed design of sgRNAs used for the activation of *CDH1*. **(c)** Schematic illustration of luciferase reporter vector containing *CDH1* promoter and 5′ UTR.

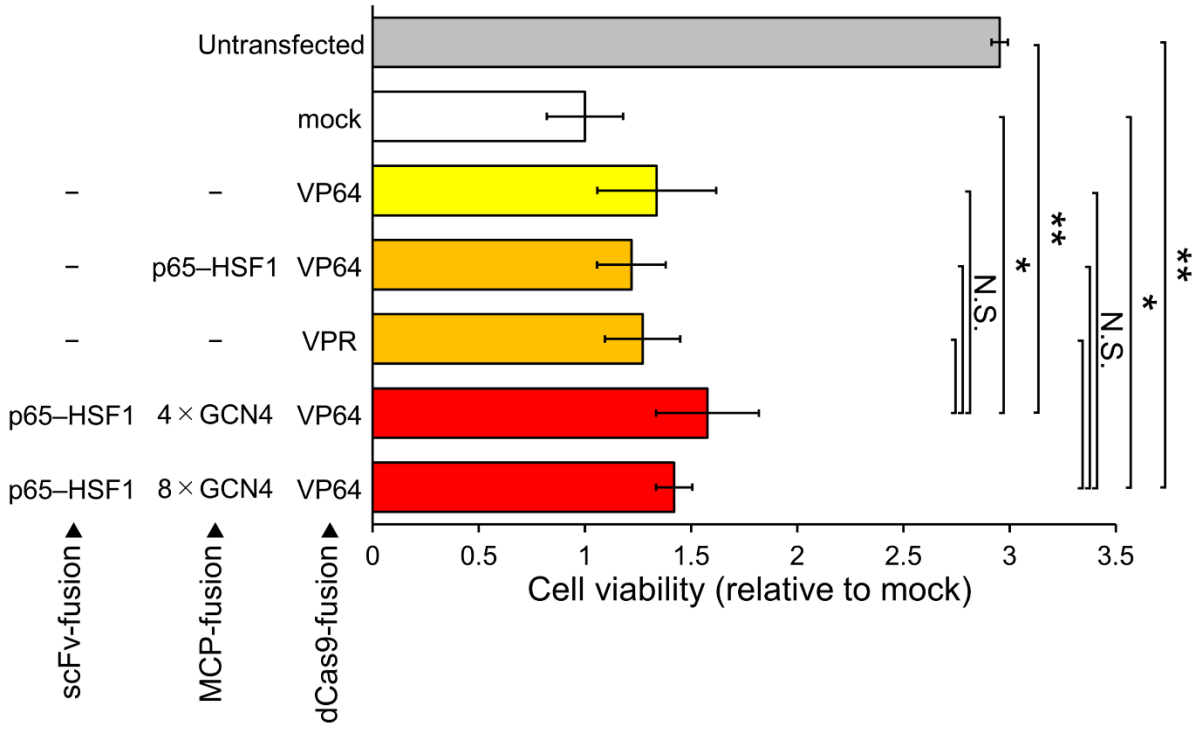

Figure 1-4. Investigation of the cell toxicity of the TREE systems in MIA-PaCa2 cells. Viability of the cells transfected with the vectors shown in the left. Data are shown as the mean ± standard deviation (S.D.; $n = 4$). $^{**}P < 0.01$; $^{*}P < 0.05$. N.S., not significant.

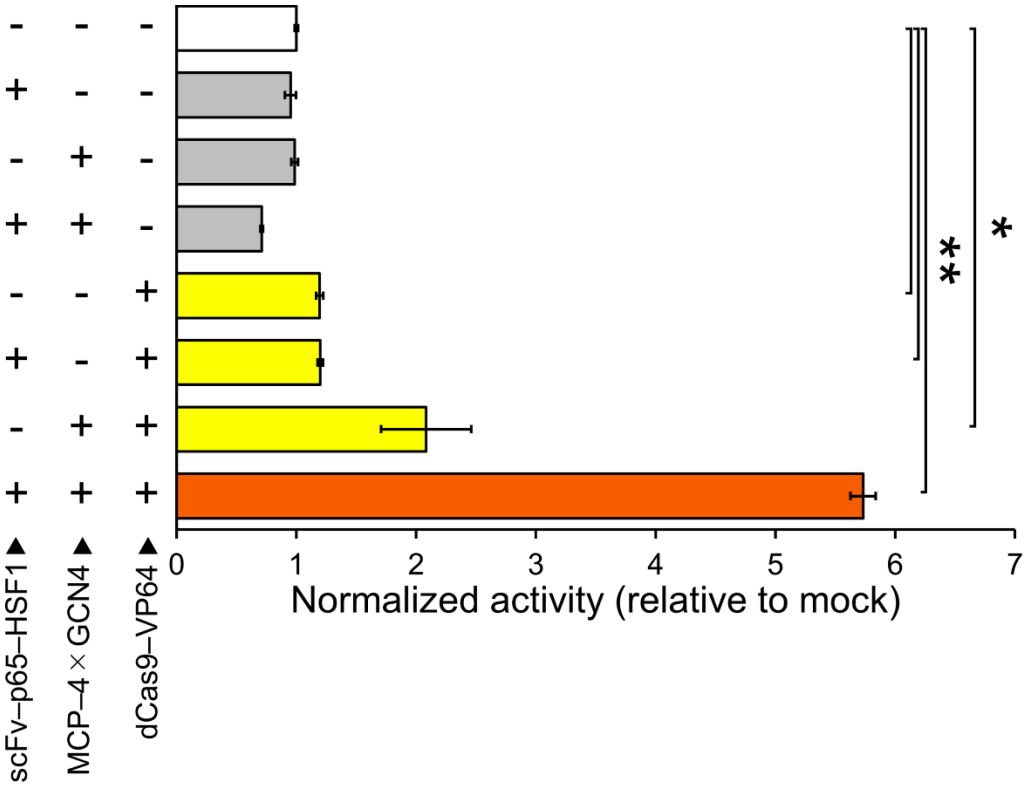

Figure 1-5. Initial validation of the TREE system by reporter assay. All possible patterns of non-, single, double, and triple administration of the TREE components were tested in MIA-PaCa2 cells. Data are shown as the mean ± S.D. ($n = 4$). **$P < 0.01$; *$P < 0.05$.

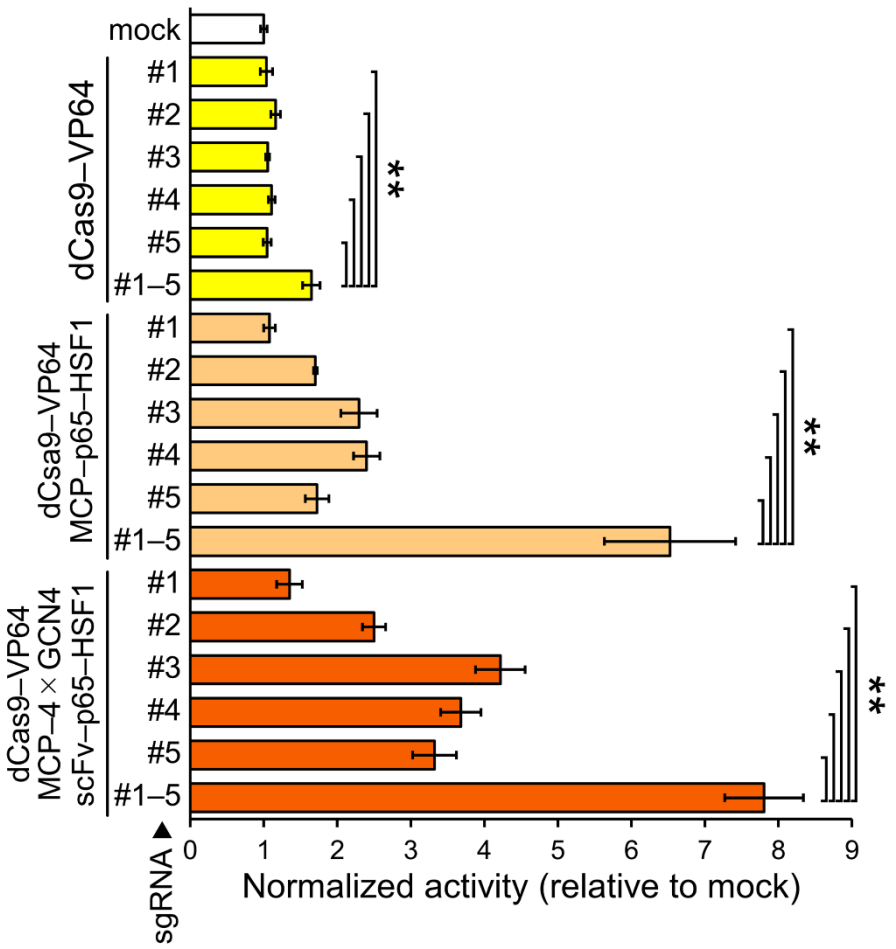

Figure 1-6. Investigation of the synergistic effect of using multiple sgRNAs. Activity comparison among single and five sgRNAs in MIA-PaCa2 cells, using the dCas9–VP64, SAM, and TREE systems by the *CDH1* reporter assay. Data are shown as the mean ± S.D. (n = 4). $^{**}P < 0.01$.

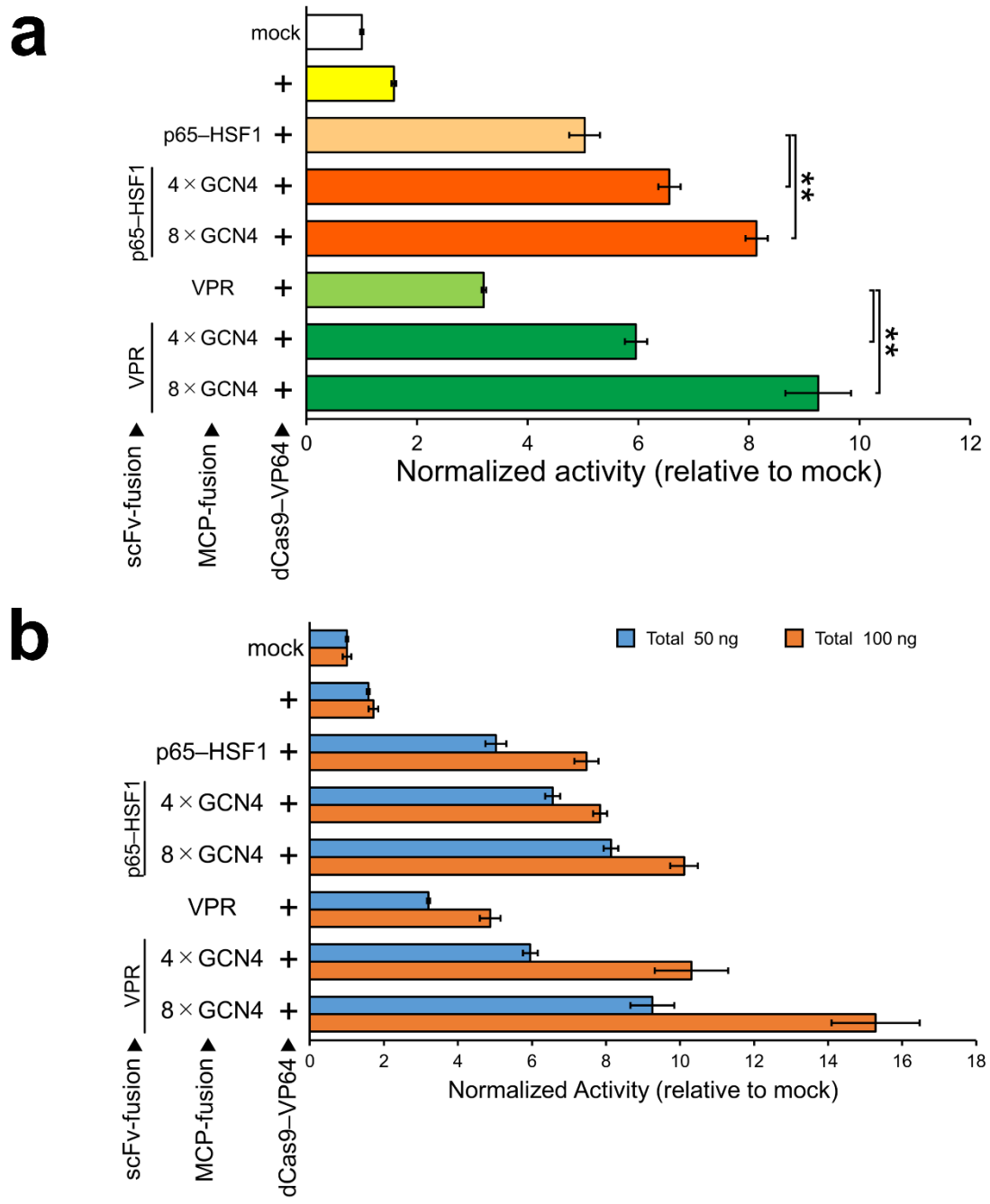

Figure 1-7. Activation of *CDH1* reporter in MIA-PaCa2 cells. (a) Activity comparison among the effectors and the repeat numbers of GCN4 epitopes, as well as the MCP–effector system, by *CDH1* reporter assay. Data are shown as the mean ± S.D. ($n = 4$). $^{**}P < 0.01$; $^{*}P < 0.05$. **(b)** Dose response of the vectors transfected (50 ng vs. 100 ng in total), related to **(a)**. Data are shown as the mean ± S.D. ($n = 4$).

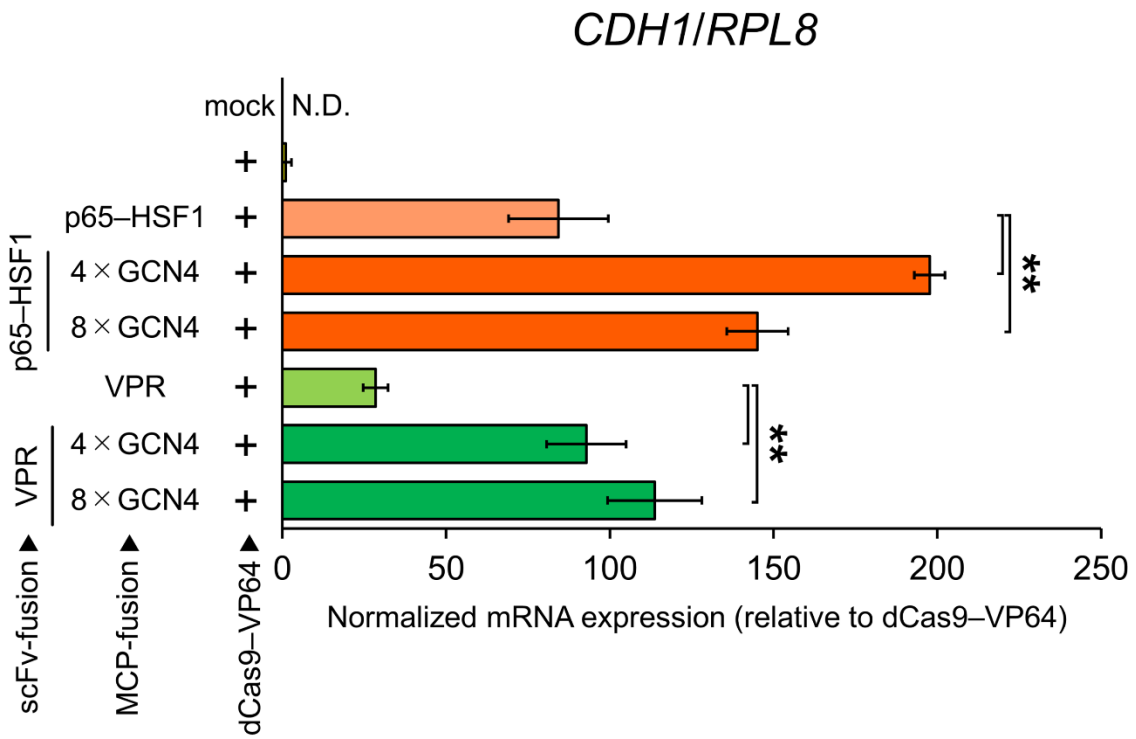

Figure 1-8. Endogenous *CDH1* gene activation in MIA-PaCa2 cells. Endogenous *CDH1* expression quantified by polymerase chain reaction (qPCR). Data are shown as the mean ± S.D. ($n = 4$). $^{**}P < 0.01$; $^{*}P < 0.05$. N.D., not detected.

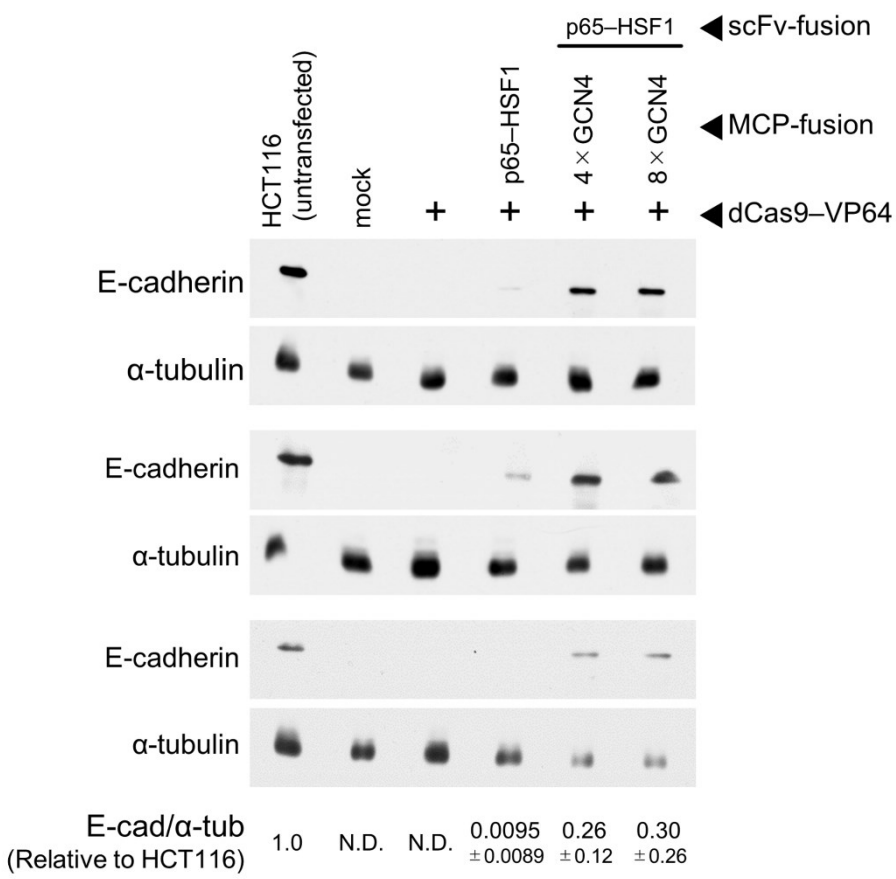

Figure 1-9. Immunoblotting of E-cadherin protein. Detection of E-cadherin and α-tubulin proteins by immunoblotting. Loaded protein is as follows: for the detection of E-cadherin in HCT116 and α-tubulin in all samples, 3 μg; for the detection of E-cadherin in all samples other than HCT116, 10 μg. Data are shown as the mean ± S.D. (n = 3). N.D., not detected.

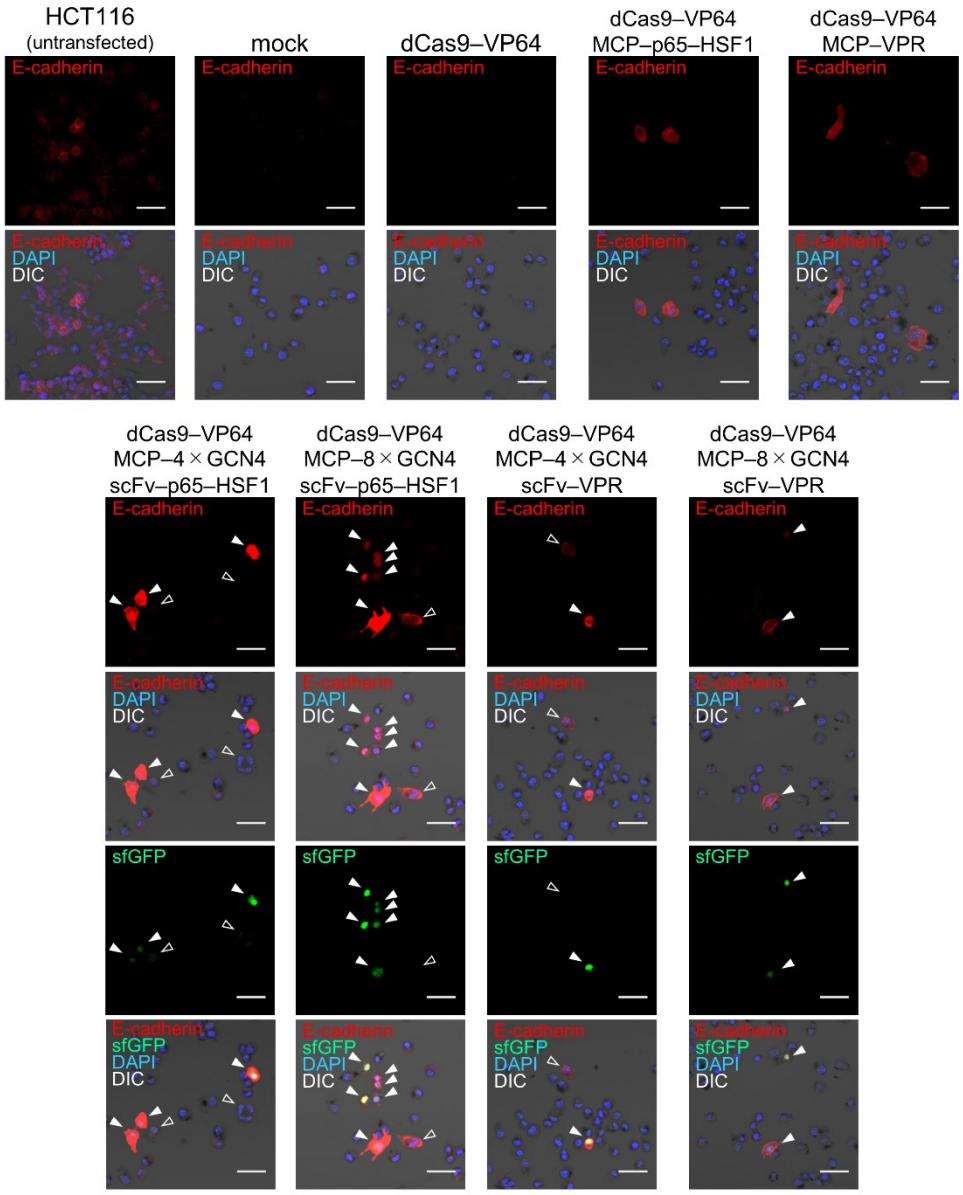

Figure 1-10. Detection of E-cadherin protein by immunostaining. Filled and open triangles indicate E-cadherin/sfGFP- and sfGFP-positive cells, respectively. Scale bars, 50 μm.

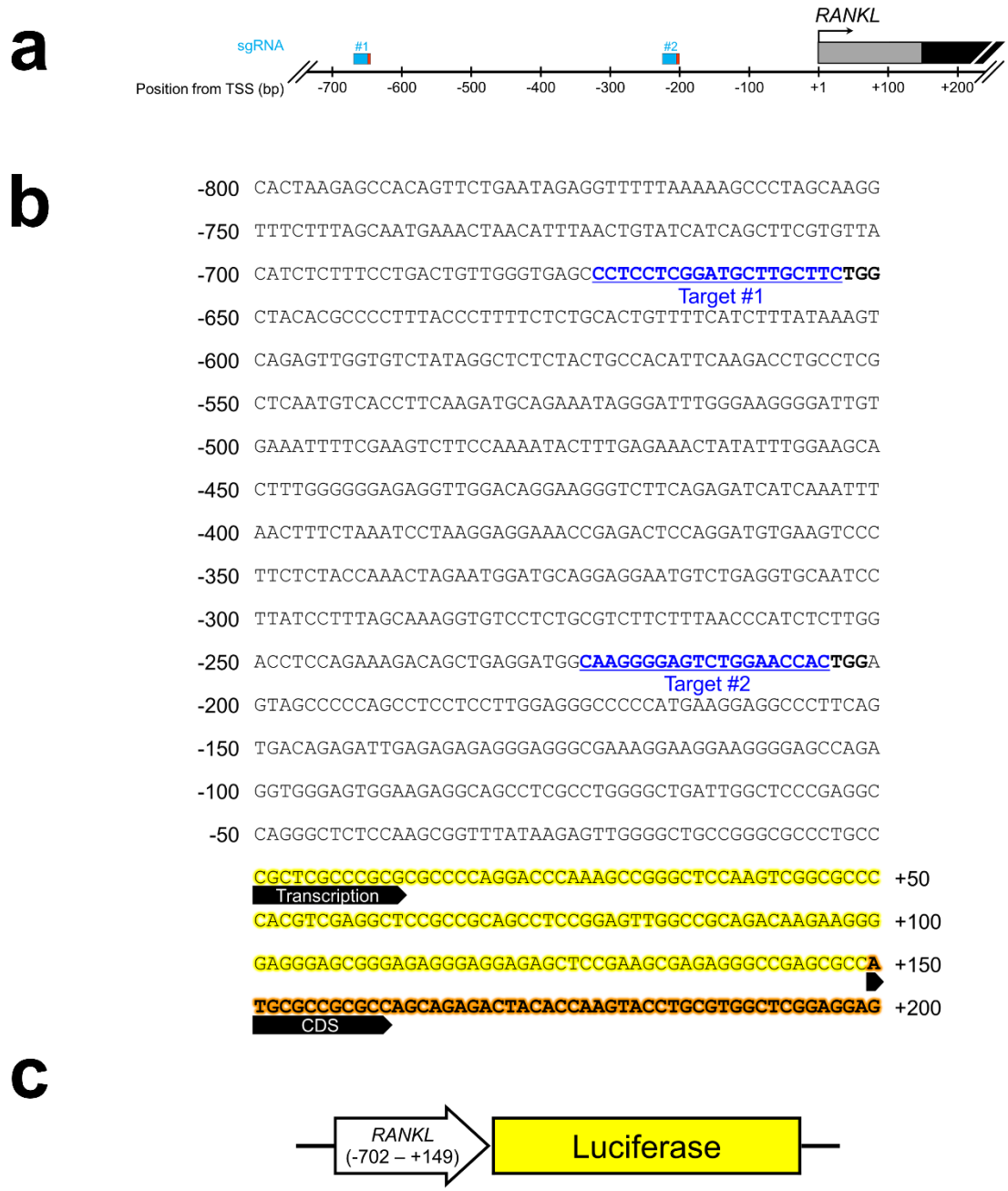

Figure 1-11. Design of the sgRNAs and the reporter for the activation of *RANKL*. (a) Schematic illustration of the positions of sgRNAs used or the activation of *RANKL*. Blue, red, gray, and black boxes show the protospacer, PAM, 5′ UTR, and CDS, respectively. **(b)** Detailed design of sgRNAs used for the activation of *RANKL*. **(c)** Schematic illustration of luciferase reporter vector containing *RANKL* promoter and 5′ UTR.

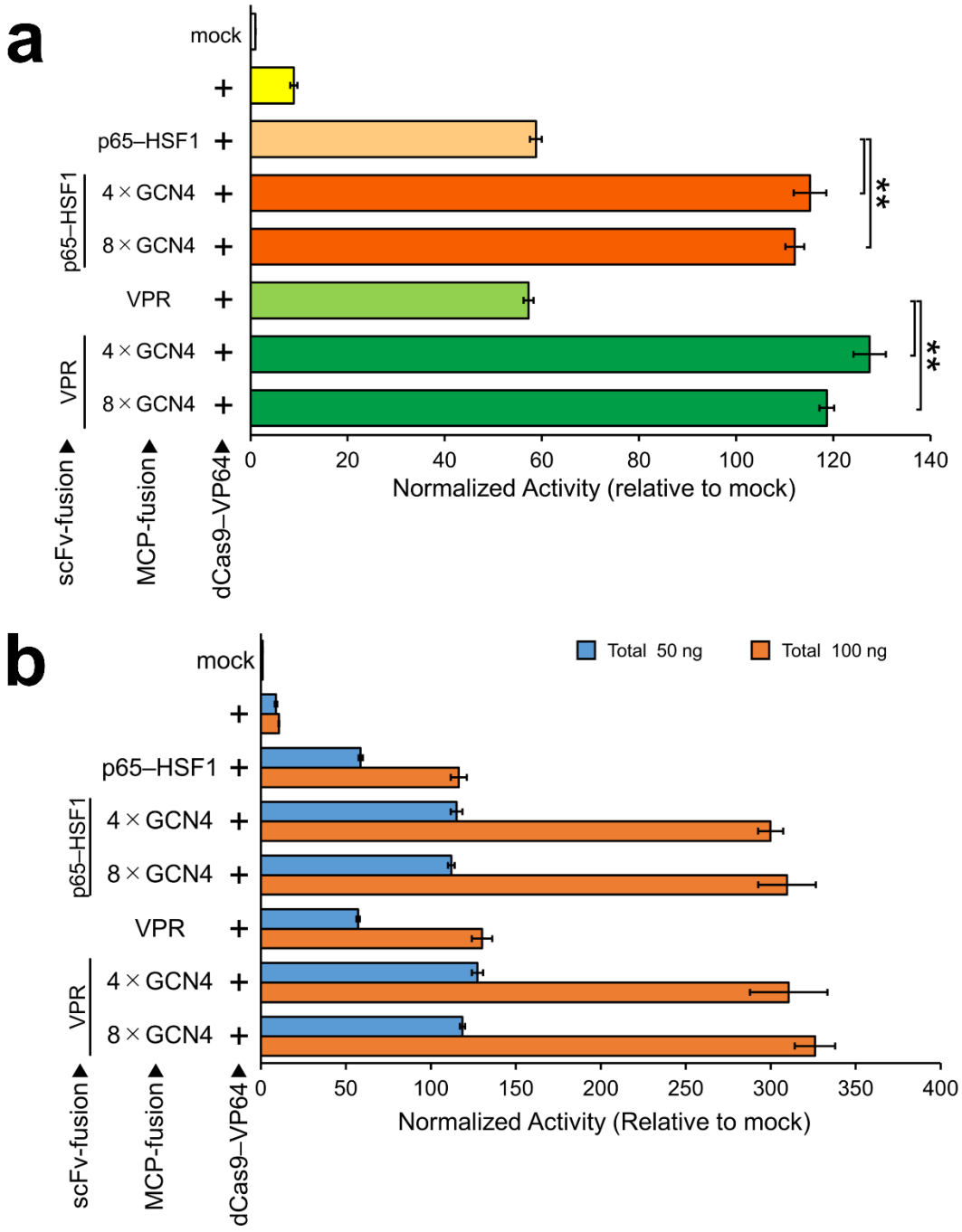

Figure 1-12. Activation of *RANKL* reporter in HEK293T cells. (a) Activity comparison among the effectors and the repeat numbers of GCN4 epitopes, as well as the MCP–effector system, by *RANKL* reporter assay. Data are shown as the mean ± S.D. ($n = 4$). $^{**}P < 0.01$; $^{*}P < 0.05$. **(b)** Dose response of the vectors transfected (50 ng vs. 100 ng in total), related to **(a)**. Data are shown as the mean ± S.D. ($n = 4$).

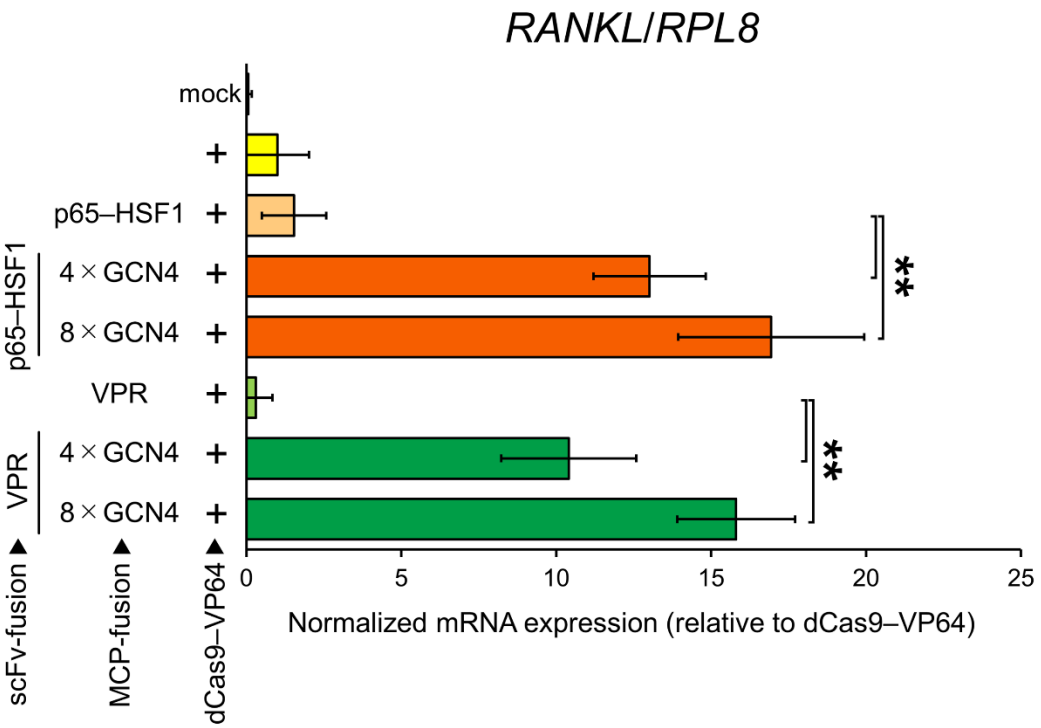

Figure 1-13. Endogenous *RANKL* gene activation in HEK293T cells. Endogenous *RANKL* expression quantified by polymerase chain reaction (qPCR). Data are shown as the mean ± S.D. ($n = 4$). $^{**}P < 0.01$; $^{*}P < 0.05$.

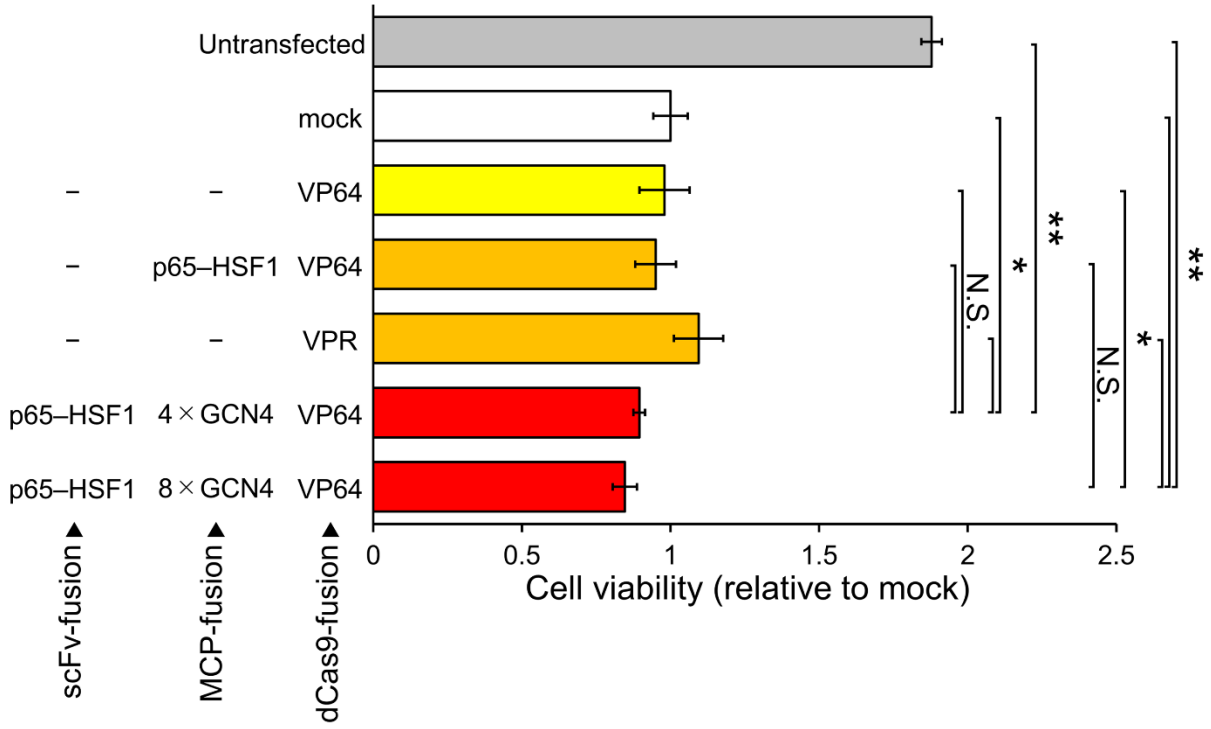

Figure 1-14. Investigation of the cell toxicity of the TREE systems in HEK293T cells. Viability of the cells transfected with the vectors shown in the left. Data are shown as the mean ± S.D ($n = 4$). $^{**}P < 0.01$; $^{*}P < 0.05$. N.S., not significant.

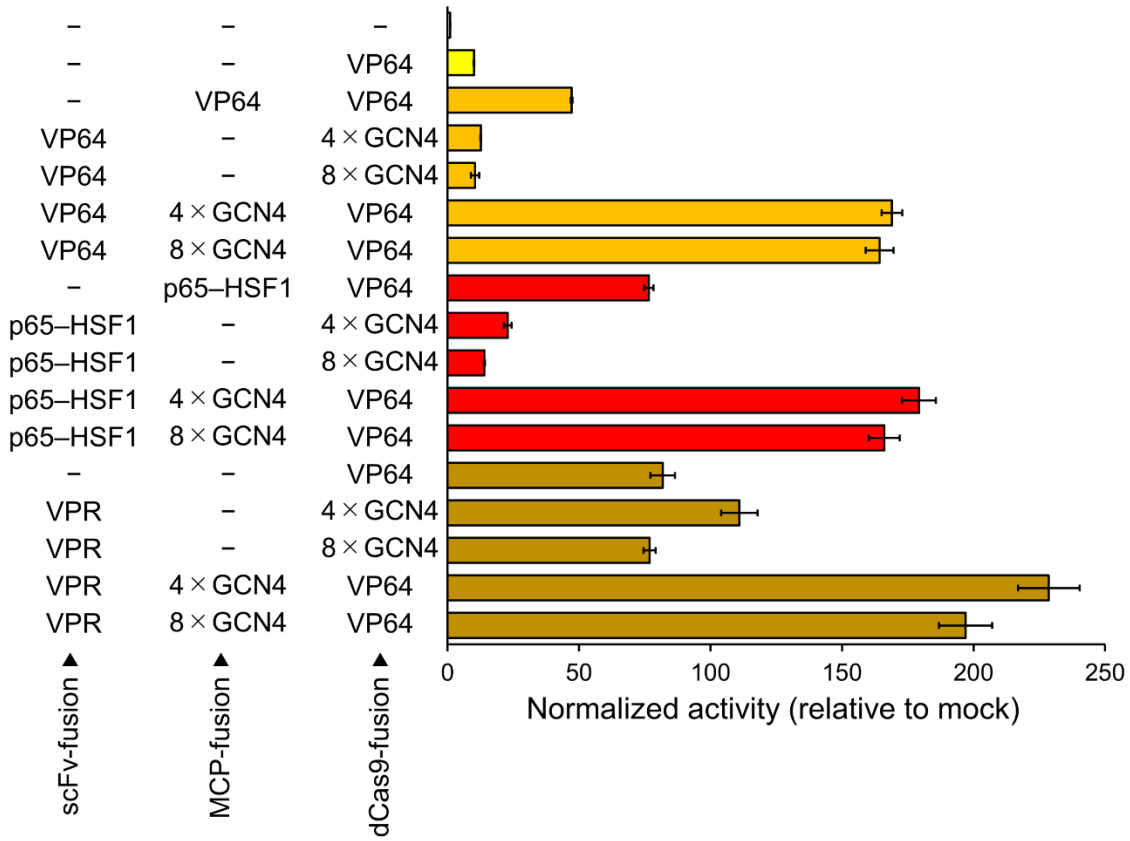

Figure 1-15. Comparison of activity between the MCP-22sTag (TREE) and dCas9-22sTag systems by *RANKL* reporter assay in HEK293T cells. For dCas9–4× and 8×GCN4 (22sTag), sgRNAs without MS2 stem loops were used instead of sgRNA2.0. Data are shown as the mean ± S.D ($n = 4$).

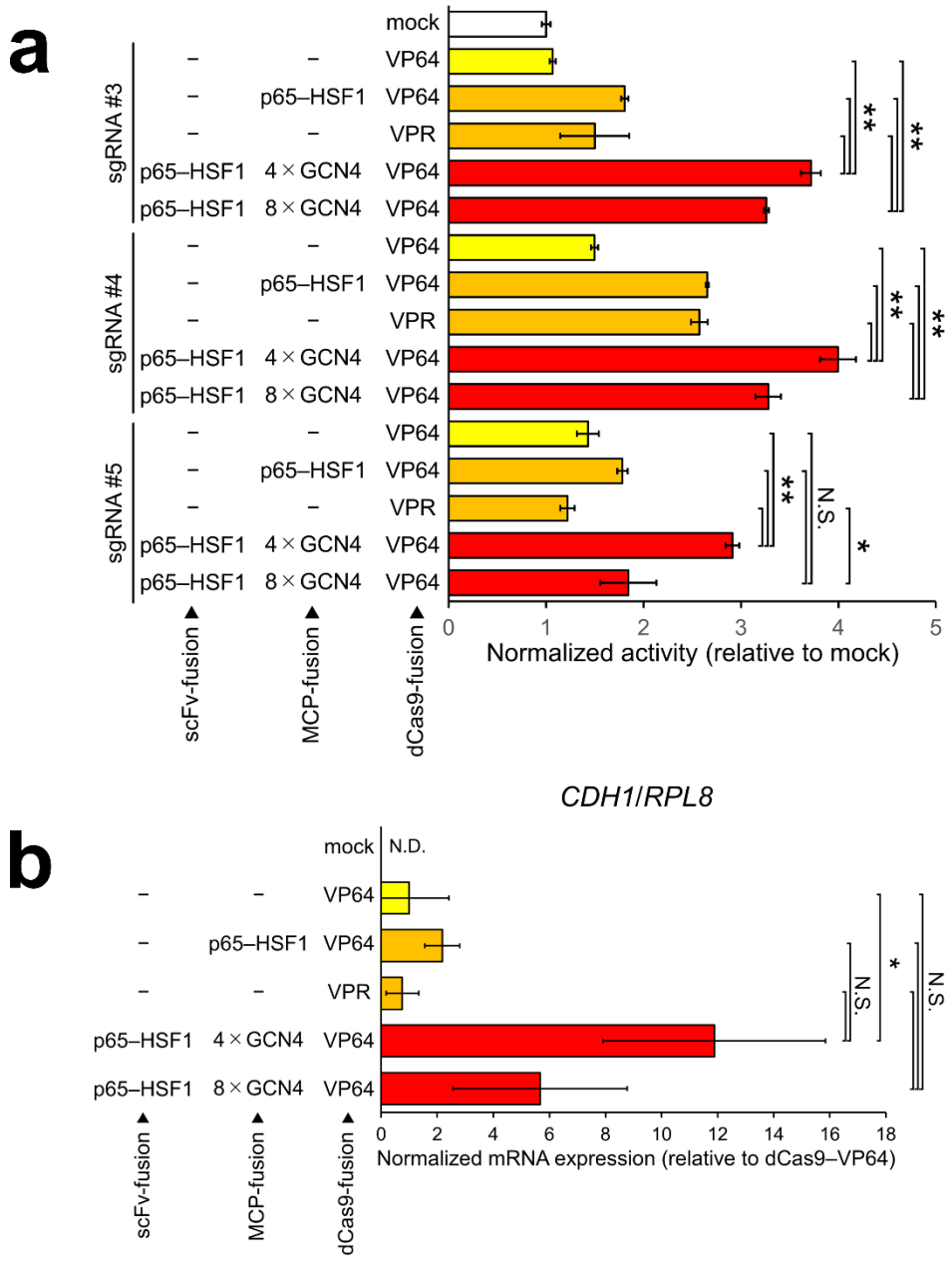

Figure 1-16. Activation of *CDH1* in MIA-PaCa2 cells with one sgRNA. (a) Assessment of the activation capacity of single TREE systems containing sgRNA #3, #4, or #5 by *CDH1* reporter assay with the comparison with the previous systems. Data are shown as the mean ± S.D (n = 3). **P < 0.01; *P < 0.05. N.S., not significant. **(b)** Assessment of the activation capacity of single TREE systems containing sgRNA #5 by qPCR analysis of endogenous *CDH1* gene with the comparison with the previous systems. Data are shown as the mean ± S.D (n = 3). *P < 0.05. N.S., not significant; N.D., not detected.

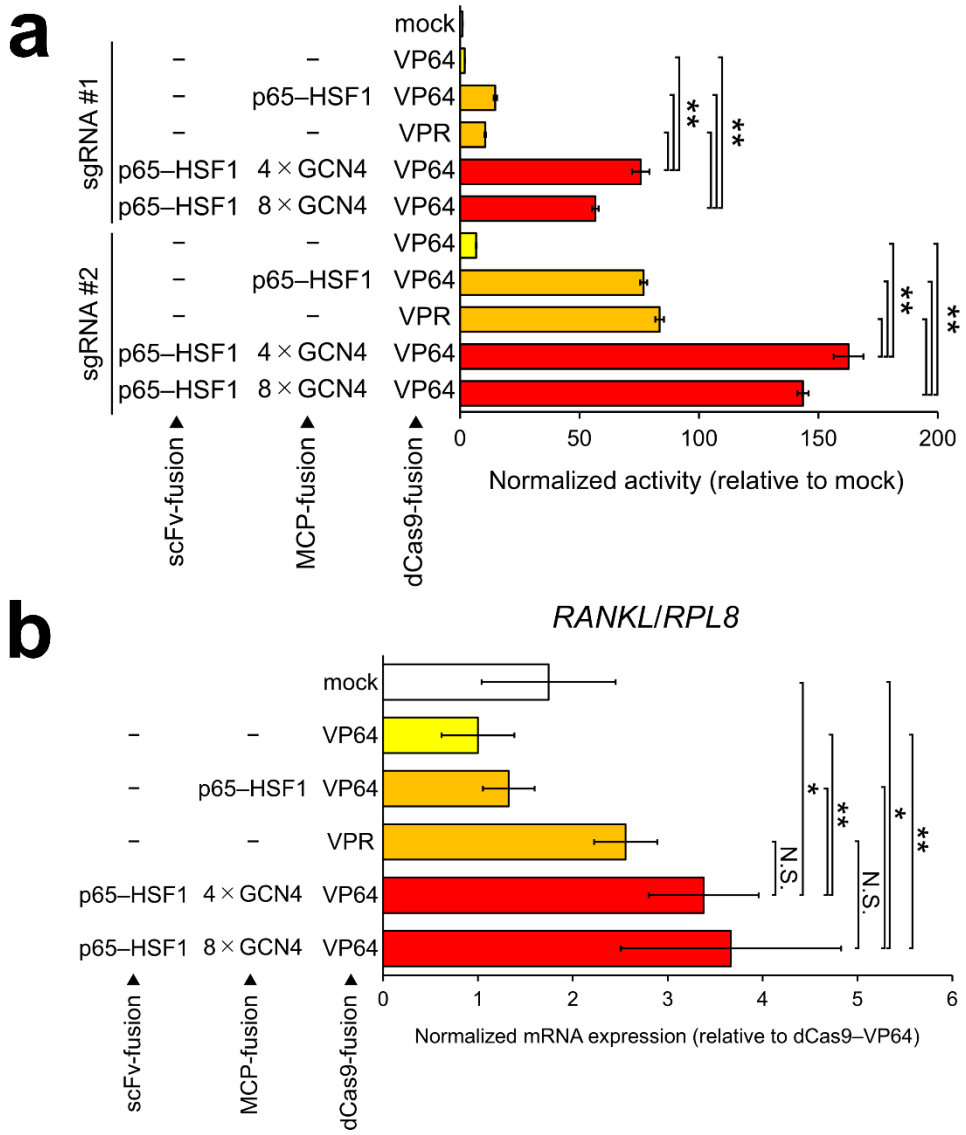

Figure 1-17. Activation of *RANKL* in HEK293T cells with one sgRNA. (a) Assessment of the activation capacity of single TREE systems containing sgRNA #1 or #2 by *RANKL* reporter assay with the comparison with the previous systems. Data are shown as the mean ± S.D (n = 3). $^{**}P < 0.01$. **(b)** Assessment of the activation capacity of single TREE systems containing sgRNA #2 by qPCR analysis of endogenous *RANKL* gene with the comparison with the previous systems. Data are shown as the mean ± S.D (n = 3). $^{**}P < 0.01$; $^{*}P < 0.05$. N.S., not significant.

Sequence 1-1

Oligonucleotides used for the sgRNA templates

Oligo name	Sequence (5'–3')
sgRNA_CDH1_1_s	caccGAAATTAGGCTGCTAGCTCAG
sgRNA_CDH1_1_as	aaacCTGAGCTAGCAGCCTAATTTC
sgRNA_CDH1_2_s	caccGAGAGACAAGTCGGGGCGGAC
sgRNA_CDH1_2_as	aaacGTCCGCCCCGACTTGTCTCTC
sgRNA_CDH1_3_s	caccGCCCAGGTCTTAGTGAGCCAC
sgRNA_CDH1_3_as	aaacGTGGCTCACTAAGACCTGGGC
sgRNA_CDH1_4_s	caccGGAGTTGCTAGGGTCTAGGT
sgRNA_CDH1_4_as	aaacACCTAGACCCTAGCAACTCC
sgRNA_CDH1_5_s	caccGCCACAGCCAATCAGCAGCG
sgRNA_CDH1_5_as	aaacCGCTGCTGATTGGCTGTGGC
sgRNA_RANKL_1_s	caccGCCTCCTCGGATGCTTGCTTC
sgRNA_RANKL_1_as	aaacGAAGCAAGCATCCGAGGAGGC
sgRNA_RANKL_2_s	caccGCAAGGGGAGTCTGGAACCAC
sgRNA_RANKL_2_as	aaacGTGGTTCCAGACTCCCCTTGC

Primers used for the construction of MCP–n× GCN4 (22sTag) expression vectors

Oligo name	Sequence (5'–3')
22sTag_vec_4n-V_F	ACCGCGGTAAACATAGGTGGTGG
22sTag_vec_1-V_R	TTCTTCCGCGGCGTCAGTC
22sTag_ins_V-1_F	gacgccgcggaaGAAGAGCTCCTTAGTAAGAACTATCATCTGG
22sTag_ins_4-5_F	ggctcagggtctgggGAAGAGCTCCTTAGTAAGAACTATCATCTGG
22sTag_ins_V-4n_R*	tatgtttaccgcggtGCCTGAGCCTGAGCCCTTC

22sTag_ins_5-4_R	CCCAGACCCTGAGCCACC
22sTag_Vec_8n-V_F	GATCTACAGCGGCCGCAAGG
22sTag_Vec_16-8_R	CATCGGACCGGTTCCACCAC
22sTag_Ins_8-16_F	ggaaccggtccgatggatctacagcggccgAGCAGCGGATCCA ACGGTCC
22sTag_Ins_24-16_R	gttggatccgctgctCGGCCGCTGTAGATCCATCG
22sTag_Ins_16-24_F	AGCAGCGGATCCAACGGTCC
22sTag_Ins_V-8n_R	CGGCCGCTGTAGATCCATCG

*The design of this primer was inappropriate, but no frame-shift mutation was observed in the constructed plasmid.

Primers used for qRT-PCR

Oligo name	**Sequence (5'–3')**
CDH1_Forward	GACTCGTAACGACGTTGCAC
CDH1_Reverse	CCGCTTCCTTCATAGTCAAA
RANKL_Forward	ATCAGAGCAGAGAAAGCGATG
RANKL_Reverse	ACTCACTTTATGGGAACCAGATG
RPL8_Forward	AGGGCATCGTCAAGGACATC
RPL8_Reverse	CGTCCGCTTCTTAAACCGATAC

Synthesized 4× GCN4 (22sTag) sequence

GAAGAGCTCCTTAGTAAGAACTATCATCTGGAAAATGAGGTAGCGCGCTTAAAGAAAGGGTC
GGGAAGTGGCGGCAGCGGAAGTGGGAGTGGAGGGAGCGGTTCTGGCGGTTCCGGCAGTGGAG
AGGAGTTGCTGTCTAAGAACTACCACTTAGAAAACGAAGTCGCACGGCTAAAAAAAGGTTCT
GGCTCCGGAGGCAGTGGTTCTGGAAGCGGTGGCAGCGGGTCAGGTGGAAGCGGATCAGGTGA
GGAATTGCTTTCCAAAAACTACCACCTTGAGAATGAGGTGGCCAGGTTAAAGAAGGGCAGCG
GCTCGGGGGGTAGTGGATCGGGGAGTGGCGGGTCAGGAAGCGGTGGTAGCGGAAGCGGGGAG
GAGCTGCTCTCGAAGAATTACCATTTGGAGAACGAAGTGGCGAGACTAAAGAAGGGAAGCGG
CAGTGGTGGATCTGGGTCTGGTTCAGGTGGGAGTGGGAGCGGTGGCTCAGGGTCTGGG

Chapter 2

Systematic optimization of the artificial transcription activators using variously patterned RNA aptamers and protein tags

Abstract

Programmable gene activation tools are useful for various fields of biological and biomedical research such as gene function analysis, direct cell reprogramming, and therapeutics. In the past half-decade, CRISPR–Cas9-based transcriptional activation systems have been improved by tandem fusion or *trans*-accumulation of multiple effector domains. Despite the various strategies for efficient gene activation, it remains unclear which approach is the most effective. Here, I incorporate and refine multiple types of RNA–protein binding and protein tagging systems, enabling systematic comparison of variously patterned accumulation platforms. I show that at least one system of my lineup significantly outperforms the conventional toolkits by comparing their abilities to induce strong gene expression in multiple contexts. A series of my accumulation platforms, named the "EARTH" collection, provides a strategy to construct "tailor-made" gene activators and can potentially contribute to improving other CRISPR-based applications.

Introduction

The CRISPR–Cas9 system has been widely used as a highly efficient genome editing tool, in which a complex of Cas9 nuclease and sgRNA induces a DSB at a target genomic site (Cong et al., 2013; Mali et al., 2013). Extending beyond DSB-dependent genome engineering, the CRISPR–Cas9 system has been diverted to recruit and accumulate a variety of effector domains, such as transcription activators and repressors (Gilbert et al., 2013), epigenomic factors (Choudhury et al., 2016; Hilton et al., 2015), cytidine and adenosine deaminases (Gaudelli et al., 2017; Komor et al., 2016), and fluorescent proteins (Chen et al., 2013).

Artificial transcription activators, including CRISPR activation (CRISPRa) systems, are promising tools for functional genomics (Gilbert et al., 2014), direct cell reprogramming (Chakraborty et al., 2014), and regulating or modeling diseases (Garcia-Bloj et al., 2016; Li et al., 2015). As the first report of CRISPRa, VP64 was directly fused to dCas9 (Gilbert et al., 2013), although its efficiency was often insufficient. Subsequently, several types of improved "second-generation" system were developed, including dCas9–VPR (Chavez et al., 2015), SAM (Konermann et al., 2015), and dCas9–SunTag (Tanenbaum et al., 2014). However, it was reported that the most effective tool depends on the tested organism, cell line, and target gene (Chavez et al., 2016), suggesting that there is room for further refinement of the transcriptional activation platform. In Chapter 1, I described the construction of a hybrid system of SAM and SunTag, named TREE, in which MCP-fused modified GCN4 epitope repeats (22sTags) were recruited to sgRNA harboring MS2 aptamers; then, anti-GCN4 scFv antibody-fused effectors accumulated at each GCN4, resulting in more efficient gene activation.

There are several platforms of RNA–protein binding and protein tagging, besides the MS2 and SunTag systems, some of which have been adopted to CRISPR-based applications (Boersma et al., 2019; Kamiyama et al., 2016; Kunii et al., 2020; Ma et al., 2016; Zalatan et al., 2015). However, it remains unclear which of them is optimal for efficient transcriptional activation. This issue prompted me to optimize the gene activation tools, based on SAM and

TREE, expanding the lineup of RNA aptamers and proteins tags. In this study, I incorporate four types of aptamer and three types of tag, enabling systematic comparison of variously patterned SAM-like RNA-binding domain (RBD)–effector fusion systems and TREE-like RBD–tag fusion systems. Using this lineup, named the "Effectors Accumulated by RNAs and Tags for High activity" (EARTH) collection, I perform powerful gene activation in multiple contexts and provide a strategy for creating "tailor-made" activation machinery.

Results and Discussion

Concept of the EARTH collection

In RBD–effector systems represented by SAM (Konermann et al., 2015), which is one of the most effective tools among the CRISPR-based transcription activators (Chavez et al., 2016), effector domains are fused to RBDs and accumulate at sgRNAs via aptamer–RBD binding (Fig. 2-1a). In contrast, in RBD–tag systems like TREE, RBD-fused protein tags are recruited to sgRNAs; then, tag-binding domain (TBD)-fused effectors accumulate at high levels (Fig. 2-1b). Here, I use four types of RNA-binding protein system, namely, MS2 (aptamer: MS2; RBD: MCP), PP7 (aptamer: PP7; RBD: PCP), boxB (aptamer: boxB; RBD: λN22), and com (aptamer: com; RBD: Com), and three types of protein tagging system, namely, 22sTag (Modified SunTag; tag: GCN4; TBD: anti-GCN4 scFv antibody) (Morita et al., 2016; Tanenbaum et al., 2014), sfGFP11 tag (tag: sfGFP11; TBD: sfGFP1–10) (Kamiyama et al., 2016), and MoonTag [tag: gp41; TBD: anti-gp41 Nanobody (Nb)] (Boersma et al., 2019). Incorporating and combining these systems, I construct variously patterned platforms to accumulate effector domains, in the formats of RBD–effectors and RBD–tags (Fig. 2-1c).

Basic validation of the EARTH collection

To express the systems of the EARTH collection in cultured human cells, I constructed three categories of plasmid: CRISPR vectors (Fig. 2-2a), RBD–effector/tag expression vectors (Fig. 2-2b), and TBD–effector expression vectors (Fig. 2-2c). I used the *RANKL* (also known as *TNFSF11*) reporter vector constructed as described in Chapter 1 and sgRNA_B designed as described in a previous report (Xu et al., 2016) to assess the activity of the systems in HEK293T cells (Fig. 2-3). First, I optimized and validated the four types of RNA–protein binding system: MS2, PP7, boxB, and com. For the systems other than com, I used the same sequences of sgRNA scaffolds as CRISPRainbow (Ma et al., 2016), a technology to visualize multiple regions of chromosomes with variously colored fluorescent proteins. Their sgRNA scaffolds started from 5′-GUUUG-3′, to avoid transcriptional termination due to T_4 sequences (5′-GUUUU-3′) under U6 promoters (Fig. 2-4a, b). This substitution slightly improved the recruitment of dCas9–VPR and did not affect the accumulation of MCP–VPR (Fig. 2-4c). In the com system used in the previous study, one sequence of the com aptamer was added at the 3′ end of the sgRNA scaffold, which started from 5′-GUUUA-3′ and had an elongated repeat at the root of the tetraloop (com #1) (Fig. 2-5a) (Zalatan et al., 2015). To refine the system, first, I changed this scaffold so that it started with 5′-GUUUG-3′, followed by the normal repeat (com #2) (Fig. 2-5b). Next, I added two aptamers as well as the other three systems, using the minimal and extended com sequences derived from the original com system of bacteriophage MU (com #3, #4) (Hattman, 1999), followed by inducing base-pairing mutations in an attempt to improve the stability (com #5, #6) (Fig. 2-5c). Although all six types of sgRNA–com showed functionality as sgRNAs by recruiting dCas9–VPR, the ability of sgRNA–com #2 and #5 to accumulate Com–VPR was quite low (Fig. 2-5d). Conversely, sgRNA–com #3, adopted as sgRNA–com in later investigations, significantly outperformed the previous sgRNA–com #1 by approximately twofold (Fig. 2-5e), suggesting that the use of two aptamer sequences conferred an improvement. I confirmed that all four of the RBD–effector systems correctly

activate the luciferase reporter without “cross-accumulation,” such as via interaction between MS2 aptamer and PCP (Fig. 2-6).

Next, I optimized the protein tagging systems. First, I refined the coding sequence of 22sTag, an optimized form of GCN4 tag (SunTag), to accumulate large effectors (Morita et al., 2016). As described in Chapter 1, immunoblotting of the MCP–GCN4 products resulted in the appearance of strong signals of unexpectedly truncated forms. RT-PCR followed by sequencing analysis suggested that the truncated forms were produced due to undesired RNA splicing, possibly in a “GU-AG” manner (Fig. 2-7a). I introduced two silent mutations to avoid “GT-AG” sequences at putative splicing sites. After that, the signals of the truncated products mostly disappeared, as confirmed by RT-PCR (Fig. 2-7b) and immunoblotting analysis targeting HA and ACV5 tags, which were located separately from each other (Fig. 2-7a, c). The refined MCP–GCN4 (22sTag) retained the ability to activate exogenous *RANKL* reporter at approximately the same level as before (Fig. 2-7d). Additionally, the activities of MCP–4× and 8×GCN4 were comparable. As other tag systems, I validated and optimized the sfGFP11 tag and MoonTag in the MCP–tag format. For the MCP–sfGFP11 tag, I used 15-a.a. peptide linkers, as described previously (Kamiyama et al., 2016). Although the most effective number of sfGFP11s depended on the sfGFP1–10-fused effector (Fig. 2-8), I adopted the 14×sfGFP11 tag in later experiments, considering the usage of VPR. For the MCP–gp41 tag (MoonTag), when the Nb-fused VP64 and VPR accumulated via MCP–4×gp41, there were no significant differences between 5-a.a. and 22-a.a. linkers (Fig. 2-9a). I chose the 22-a.a. linker for RBD–gp41 (MoonTag), the same as for the RBD–GCN4 tag (22sTag) (Morita et al., 2016). Upon comparing with MCP–4×gp41 with regard to VPR accumulation, MCP–8×gp41 showed a comparable level but MCP–12×gp41 had a significantly lower one (Fig. 2-9b).

Combining the platforms optimized as described above, I performed basic validation of the 12 patterns of RBD–tag systems (Fig. 2-10a). Strong luciferase activity was obtained only

when the full set of three plasmid vectors with the correct sgRNA was introduced, suggesting that the constructed RBD–tag systems work as expected.

Comparison of variously patterned RBD–effector and RBD–tag systems in HEK293T cells

To compare the activity of the systems of the EARTH collection, I performed *RANKL* reporter assay in HEK293T cells, using dCas9 or dCas9–VP64, with *trans*-p65–HSF1 or VPR (Fig. 2-11). I found that the activity of the RBD–4×GCN4 (22sTag) systems was higher than that of the RBD–effectors (MS2: #09 vs. #10; PP7: #14 vs. #15; boxB: #19 vs. 20; com: #24 vs. 25), except the boxB systems with *trans*-VPR. Additionally, including the other tags, the MCP–tag systems (#10–14) introduced activity that was comparable to or higher than that with the MCP–effectors (#09).

Following the results of the reporter assays, I next tested the activation of endogenous genes. I selected two additional target genes, *MMP9* and *CTCFL* (also known as *BORIS*), both of whose expression levels were on the order of ~10^{-3}, compared with the reference genes, *GUSB* and *TBP*, in HEK293T cells (Fig. 2-12a). Then, I designed two additional sgRNAs on the *RANKL* promoter (sgRNA_A and C) and three (sgRNA_A–C) on that of the other two genes (Fig. 2-12b). First, I tested p65–HSF1 accumulation in all nine cases (Fig. 2-13a–c). Surprisingly, focused on the results of activation with *RANKL* sgRNA_B, the order among the tested systems differed quite markedly from that of the reporter assay (Figs. 2-11b, 13a middle). In particular, the activity of PCP–4×GCN4 was slightly higher than that of PCP–p65-HSF1 with statistical significance in the reporter assay (#14 vs. #15) (Fig. 2-11b), while the latter was approximately tenfold higher in terms of endogenous *RANKL* activation (#07 vs. #08) (Fig. 2-13a middle). I found that the preference among the systems depended on the target gene rather than the target sequence. Interestingly, however, the activation magnitude was strongly affected by the sgRNA site. For example, λN22–p65–HSF1 (#12) with sgRNA_B and C upregulated *MMP9* expression ~500-fold and ~7,000-fold, respectively (Fig. 2-13b), although

these two sgRNAs were only 74 bp apart (Fig. 2-12b). This was consistent with previous research comparing the previous second-generation systems (Chavez et al., 2016). Although the most effective platform depended on the target gene and sgRNA site, MCP–4×GCN4 (#03), PCP–p65–HSF1 (#07), λN22–p65–HSF1 (#12), and Com–4×GCN4 (#18) consistently showed relatively high performance.

To investigate the dependence of these observations on the effector, I tested the VPR accumulation using a selected sgRNA that showed the highest magnitude with *trans*-p65–HSF1 on each target gene (*RANKL*: B; *MMP9*: C; *CTCFL*: A) (Figs. 2-13, 2-14). The activity of λN22–VPR (#12) remarkably exceeded that of the other systems, except Com–4×GCN4 (#18) in the *CTCFL* gene. Interestingly, the efficacy of the PCP–effector (VPR; #07) was comparable to that of PCP–4×GCN4 (#08), unlike in the case of p65–HSF1 accumulation, suggesting that the format preference depends on the accumulated effector. Throughout these experiments, I found that there was a tendency that the use of RBD–4×GCN4 (22sTag) was preferable for the MS2- and com-based systems, while the PP7- and boxB-based ones showed the best performance when the effector domain was directly fused to RBDs.

In most of the results, I could not observe an additive or synergistic effect in accordance with the number of effector domains to be accumulated (Fig. 2-13a–c). For example, PCP–effector, PCP–4×GCN4/gp41, PCP–8×gp41, and PCP–14×sfGFP11 systems could as expected accumulate 4, 16, 32, and 56 effector domains at most, respectively, while PCP–effectors were comparable to or more effective than PCP–tags. Although I speculate that there is still room to improve the functionality of the RBD–tag systems by refining the stability of the protein expression, induction dose, and mass ratio of the plasmids, among others, these points cannot explain the differences depending on the target gene, suggesting a need to investigate the biological background.

"Tailor-made" activators outperform conventional second-generation platforms in multiple cell lines and target genes

I next tested whether part of the EARTH collection outperformed the conventional activation tools that had been used as standard. I compared the selected five types of the EARTH collection, in accordance with the results of the full-set comparison performed as described above (Figs. 2-13, 2-14), with conventional dCas9–VP64 (#02), dCas9–VPR (#03), SAM (#04), and dCas9–SunTag (#05), in HEK293T and another cell line, MCF-7, which showed low or undetectable expression of *RANKL*, *MMP9*, and *CTCFL* genes (Figs. 2-15–2-17). Note that I display the same data on log-scaled graphs in Figs. 2-15b, 2-16b, and 2-17b.

In all six cases, at least one of the systems of the EARTH collection significantly outperformed the conventional second-generation tools (Figs. 2-15–2-17). Interestingly, the order regarding the performance level among the systems depended on not only the target gene but also the tested cell line. These results suggest that it is important to optimize the activation platform for each target gene and that the EARTH collection has potential to optimize the context-specific strategy of transcriptional activation as a "tailor-made" toolkit.

In a similar investigation, Martella and colleagues evaluated CRISPRa and CRISPR inhibition (CRISPRi) systems using MCP– and PCP–effector-based accumulation in human cells (Martella et al., 2019). In that study, PP7 system was more efficient for activation, which matches most results in my study (Figs. 2-13, 2-14), while MS2 system showed better performance for inhibition. Focusing on other applications, Chao and colleagues performed the accumulation of cytidine and adenosine deaminases via MS2, PP7, boxB, and com systems, named SWISS, in rice protoplasts (Li et al., 2020). It is interesting to divert the EARTH collection into applications other than transcriptional activation, such as transcriptional repression, epigenome editing, chromosome visualization, and deaminase-mediated saturation mutagenesis.

Materials and Methods

Construction of CRISPR vectors

The all-in-one CRISPR–Cas9 vector system (Sakuma et al., 2014) was modified to express dCas9–effector/tag and sgRNA–aptamer. coding sequences of dCas9 and dCas9–effector/tag were inserted downstream of the CBh promoter, replacing the original Cas9 sequence using PCR and an In-Fusion HD Cloning Kit (Takara). Modified sgRNA was inserted downstream of the U6 promoter, replacing the original sgRNA with a PCR-based method using an In-Fusion Cloning Kit. To express the sequence-specific sgRNAs, the annealed oligonucleotides, listed in Sequence 2-1, were annealed and inserted into the cloning sites, as described previously (Ran et al., 2013).

Construction of the RBD–effector/tag expression vectors

MCP–p65–HSF1, MCP–VPR, MCP–4×, and 8×GCN4 (22sTag) expression vectors, constructed as described in Chapter 1, were modified to express the RBD–effector/tag fusion proteins. For PCP and Com, their coding sequences were synthesized by gBlocks (IDT) and replaced the MCP sequence. For λN22, the coding sequence was added by PCR, using the primers adapted with a λN22 sequence at the 5′ end. For RBD–GCN4 (22sTag), the ACV5 tag for immunoblotting was inserted at the C terminus and silent mutations at the putative splicing sites were induced by PCR. For RBD–7×, 14×, and 21×sfGFP11, and RBD–4×, 8×, and 12×gp41 (MoonTag), the coding sequences of 7×sfGFP11 and 4×gp41 were synthesized by gBlocks and one to three of them were inserted downstream of RBDs, respectively. Insertion of the sequence was performed using an In-Fusion HD Cloning Kit.

Construction of the TBD–effector expression vectors

For scFv–effectors, previously constructed scFv–VP64, scFv–p65–HSF1, and scFv–VPR were used. For Nb effectors, coding sequence of gp41 was synthesized by gBlocks and replaced the

scFv sequence of scFv effectors. For sfGFP1–10–effectors, coding sequence of sfGFP1–10 was synthesized by gBlocks and replaced the MCP sequence of MCP–effectors. Insertion of the sequence was performed using an In-Fusion HD Cloning Kit.

Cell culture

HEK293T and MCF-7 cells, obtained from ATCC, were maintained in D-MEM (high glucose) with L-glutamine and Phenol Red (FUJIFILM Wako), supplemented with 10% FBS (Thermo Fisher Scientific), 1% minimum essential medium non-essential amino acids (Thermo Fisher Scientific), and 1% penicillin–streptomycin (FUJIFILM Wako). Both cell lines were tested to confirm the absence of mycoplasma contamination using e-Myco Mycoplasma PCR Detection Kit (iNtRON Biotechnology) and outsourced for authentication as follows: for HEK293T: short tandem repeat analysis (Takara); and for MCF-7: human cell line authentication test (STR-PCR) (Promega).

Transfection for the reporter assays

A total of 60,000 cells were transfected with the vectors mixed as follows, using a Lipofectamine LTX reagent (Thermo Fisher Scientific) in a 96-well plate. For experiments including RBD–tag systems (Figs. 2-7a, 2-8, 2-9a, b, 2-10, 2-11a–d), a 1:1:1 mass ratio of the following three vectors was used: (1) sgRNA/dCas9–effector all-in-one vector, or pcDNA; (2) RBD–effector/tag expression vector or pcDNA; (3) TBD–effector expression vector or pcDNA (25 ng in total for Fig. 2-11a–d; 100 ng in total for Figs. 2-7d, 2-8, 2-9a, b, 2-10), and 100 ng of the luciferase reporter vector and 20 ng of the *R*Luc expression vector for normalization. For validation of RBD–effector systems (Figs. 2-4c, 2-5d, e, 2-6), a 1:1 mass ratio of the following two vectors was used: (1) sgRNA/dCas9–effector all-in-one vector, or pcDNA; (2) RBD–effector/tag expression vector or pcDNA (100 ng in total), and 100 ng of the luciferase reporter vector and 20 ng of the *R*Luc expression vector.

Transfection for analysis of endogenous gene activation

A total of 30,000 cells were transfected with the vectors mixed as follows, using Lipofectamine LTX reagents in a 96-well plate. A 1:1:1 mass ratio of the following three vectors was used: (1) sgRNA/dCas9–effector all-in-one vector or pcDNA; (2) RBD–effector/tag expression vector or pcDNA; (3) TBD–effector expression vector or pcDNA (200 ng in total).

Transfection for expression analysis of MCP–GCN4 (22sTag)

For RT-PCR followed by sequencing analysis, a total of 30,000 cells were transfected with 200 ng of MCP–GCN4 (22sTag) expression vector or pcDNA using Lipofectamine LTX reagent in a 96-well plate. For immunoblotting analysis, a total of 60,000 cells were transfected with 200 ng of MCP–22sTag expression vector or pcDNA.

Luciferase assay

At 24 h post-transfection, the cells were lysed and dual luciferase activity was measured using a Dual-Glo Luciferase Assay System (Promega) on a TriStar LB Multimode Microplate Reader (Berthold Technologies).

Analysis of endogenous mRNA expression with RT-ddPCR

At 48 h post-transfection, cell lysis and extraction of total RNA were performed using a NucleoSpin RNA Plus (Takara). Reverse transcription was performed using a High-Capacity cDNA Reverse-Transcription Kit (Thermo Fisher Scientific). Relative mRNA expression levels were quantified by droplet digital PCR (ddPCR) using a ddPCR Supermix for Probes (no dUTP) (Bio-Rad) and TaqMan Gene Expression Assays (Thermo Fisher Scientific) as follows: *RANKL* (Assay ID Hs00243522_m1; FAM); *MMP9* (Assay ID Hs00959562_m1; FAM); *CTCFL* (Assay ID Hs00540740_m1; FAM); *GUSB* (Assay ID Hs99999908_m1; VIC); and *TBP* (Assay ID Hs99999910_m1; VIC) on a QX200 Droplet Digital PCR System (Bio-

Rad). The expression levels of *RANKL*, *MMP9*, and *CTCFL* were normalized by those of *GUSB* and *TBP*.

RT-PCR and sequencing analysis of MCP–GCN4 (22sTag)

At 48 h post-transfection, total RNA was extracted using a TRIzol Plus RNA Purification Kit (Thermo Fisher Scientific). cDNA was synthesized using a High-Capacity cDNA Reverse-Transcription Kit (Thermo Fisher Scientific) and amplified using a KOD One PCR Master Mix (Toyobo). For sequencing analysis, the amplicon was cloned into pTA2 plasmid using a TArget Clone -Plus- (Toyobo). The insert sequences were analyzed by a DNA sequence analysis service (FASMAC) using M13 forward/reverse primers. The signals were visualized on SnapGene Viewer software version 5.2.2.

Immunoblotting of MCP–22sTag proteins

At 24 h post-transfection, the cells were transferred onto a six-well plate. After 48 h, they were lysed, sonicated, and centrifuged for 1 min at 12,000 × *g* and 4°C. Then, the protein concentration of the supernatant was measured using a Protein Assay Kit (Bio-Rad). A total of 10 μg of each protein sample was denatured in Laemmli buffer for 5 min at 98°C, followed by electrophoresis on 10% polyacrylamide gel. The proteins were blotted onto a polyvinylidene fluoride membrane and blocked with 5% dry milk in 1× PBST for 1 h. The membranes were incubated overnight with anti-HA tag (ab49969; Abcam) or anti-ACV5 tag (A2980; Sigma-Aldrich) antibody at a 1:2,000 dilution ratio in Can Get Signal Solution 1 (Toyobo). Then, the proteins were incubated with HRP-conjugated anti-mouse IgG secondary antibody (Thermo Fisher Scientific) at a 1:2,000 dilution ratio in Can Get Signal Solution 2 (Toyobo) for 1 h at room temperature. Chemiluminescent signals were generated using a SuperSignal West Pico Plus Chemiluminescent Substrate (Thermo Fisher Scientific) and captured on a ChemiDoc Touch MP (Bio-Rad).

Statistical analysis

Statistical analyses were performed by one-tailed or two-tailed Welch's *t*-test on R software version 4.0.3.

References

Boersma, S., Khuperkar, D., Verhagen, B.M.P., Sonneveld, S., Grimm, J.B., Lavis, L.D., and Tanenbaum, M.E. (2019). Multi-Color Single-Molecule Imaging Uncovers Extensive Heterogeneity in mRNA Decoding. Cell *178*, 458–472.

Chakraborty, S., Ji, H., Kabadi, A.M., Gersbach, C.A., Christoforou, N., and Leong, K.W. (2014). A CRISPR/Cas9-based system for reprogramming cell lineage specification. Stem Cell Reports *3*, 940–947.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B., Tuttle, M., Iyer, E., Kiani, S., Guzman, C., Wiegand, D., Ter-Ovanesyan, D., et al. (2015). Highly-efficient Cas9-mediated transcriptional programming. Nat. Methods *12*, 326–328.

Chavez, A., Tuttle, M., Pruitt, B.W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S.J., Cecchi, R.J., Kowal, E.J.K., Buchthal, J., et al. (2016). Comparison of Cas9 activators in multiple species. Nat. Methods *13*, 563–567.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479–1491.

Choudhury, S.R., Cui, Y., Lubecka, K., Stefanska, B., and Irudayaraj, J. (2016). CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. Oncotarget *7*, 46545–46556.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Garcia-Bloj, B., Moses, C., Sgro, A., Plani-Lam, J., Arooj, M., Duffy, C., Thiruvengadam, S., Sorolla, A., Rashwan, R., Mancera, R.L., et al. (2016). Waking up dormant tumor suppressor genes with zinc fingers, TALEs and the CRISPR/dCas9 system. Oncotarget *7*, 60535–60554.

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature *551*, 464–471.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442–451.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Hattman, S. (1999). Unusual transcriptional and translational regulation of the bacteriophage Mu mom operon. Pharmacol. Ther. *84*, 367–388.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Kamiyama, D., Sekine, S., Barsi-Rhyne, B., Hu, J., Chen, B., Gilbert, L.A., Ishikawa, H., Leonetti, M.D., Marshall, W.F., Weissman, J.S., et al. (2016). Versatile protein tagging in cells with split fluorescent protein. Nat. Commun. *7*, 11046.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420–424.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588..

Kunii, A., Yamamoto, T., and Sakuma, T. (2020). Various strategies of effector accumulation to improve the efficiency of genome editing and derivative methodologies. Vitr. Cell. Dev. Biol. - Anim. *56*, 359–366.

Li, C., Li, C., Zong, Y., Jin, S., Jin, S., Zhu, H., Zhu, H., Lin, D., Lin, D., Li, S., et al. (2020). SWISS: multiplexed orthogonal genome editing in plants with a Cas9 nickase and engineered CRISPR RNA scaffolds. Genome Biol. *21*, 141.

Li, K., Pang, J., Cheng, H., Liu, W.P., Di, J.M., Xiao, H.J., Luo, Y., Zhang, H., Huang, W.T., Chen, M.K., et al. (2015). Manipulation of prostate cancer metastasis by locus-specific modification of the CRMP4 promoter region using chimeric TALE DNA methyltransferase and demethylase. Oncotarget *6*, 10030–10044.

Ma, H., Tu, L.C., Naseri, A., Huisman, M., Zhang, S., Grunwald, D., and Pederson, T. (2016). Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. Nat. Biotechnol. *34*, 528–530.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Martella, A., Firth, M., Taylor, B.J.M., Göppert, A., Cuomo, E.M., Roth, R.G., Dickson, A.J., and Fisher, D.I. (2019). Systematic Evaluation of CRISPRa and CRISPRi Modalities Enables Development of a Multiplexed, Orthogonal Gene Activation and Repression System. ACS Synth. Biol. *8*, 1998–2006.

Morita, S., Noguchi, H., Horii, T., Nakabayashi, K., Kimura, M., Okamura, K., Sakai, A., Nakashima, H., Hata, K., Nakashima, K., et al. (2016). Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. Nat. Biotechnol. *34*, 1060–1065.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., and Yamamoto, T. (2014). Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. Sci. Rep. *4*, 5400.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell *159*, 635–646.

Xu, X., Tao, Y., Gao, X., Zhang, L., Li, X., Zou, W., Ruan, K., Wang, F., Xu, G.L., and Hu, R. (2016). A CRISPR-based approach for targeted DNA demethylation. Cell Discov. *2*, 16009.

Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi, L.S., et al. (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell *160*, 339–350.

Figures

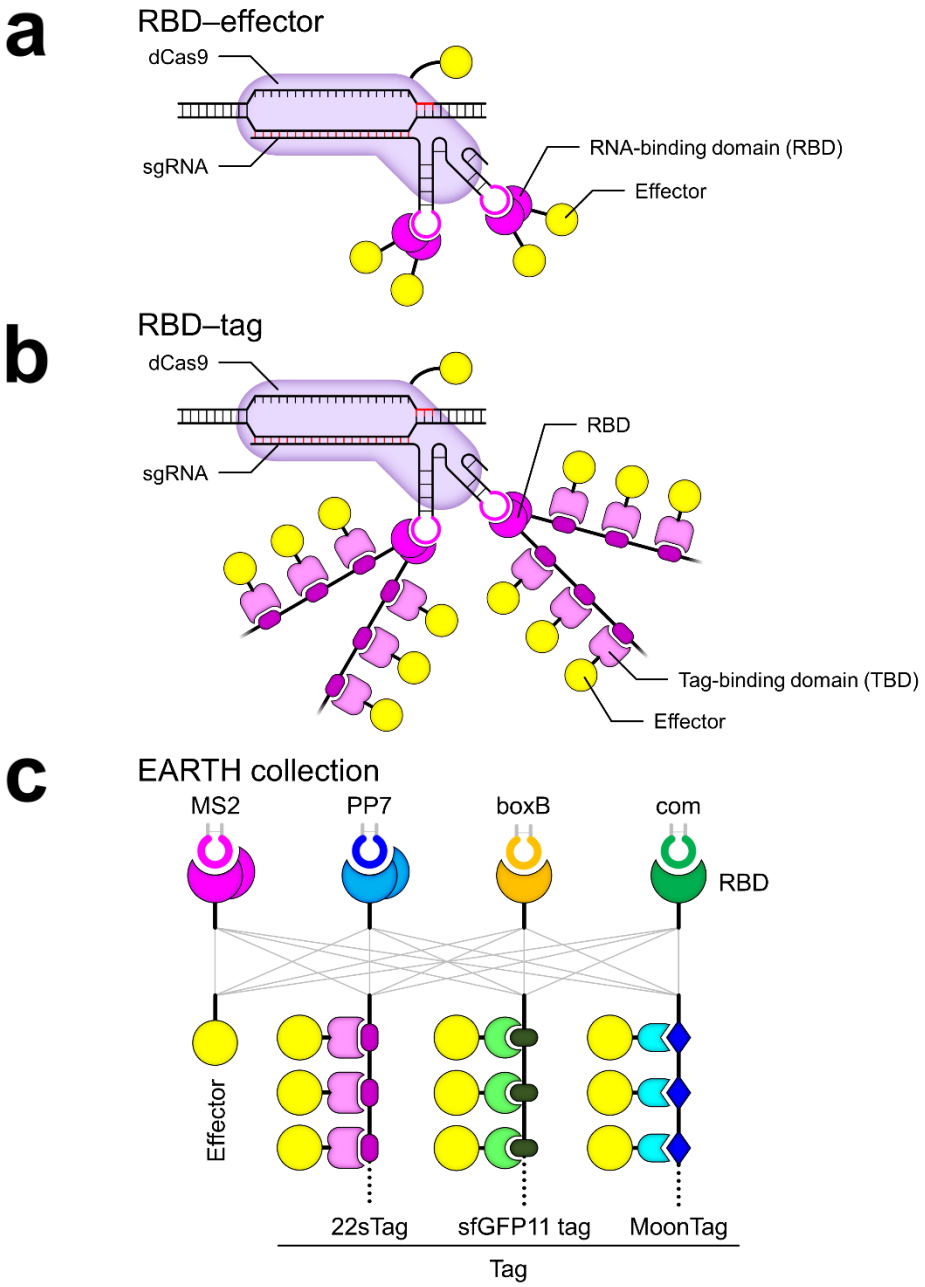

Figure 2-1. Concept of the EARTH collection. (a) Schematic illustration of the generalized form of the RBD–effector system. **(b)** Schematic illustration of the generalized form of the RBD–tag system. **(c)** Schematic illustrations of the EARTH collection, enabling the use of variously patterned platforms to accumulate effector domains.

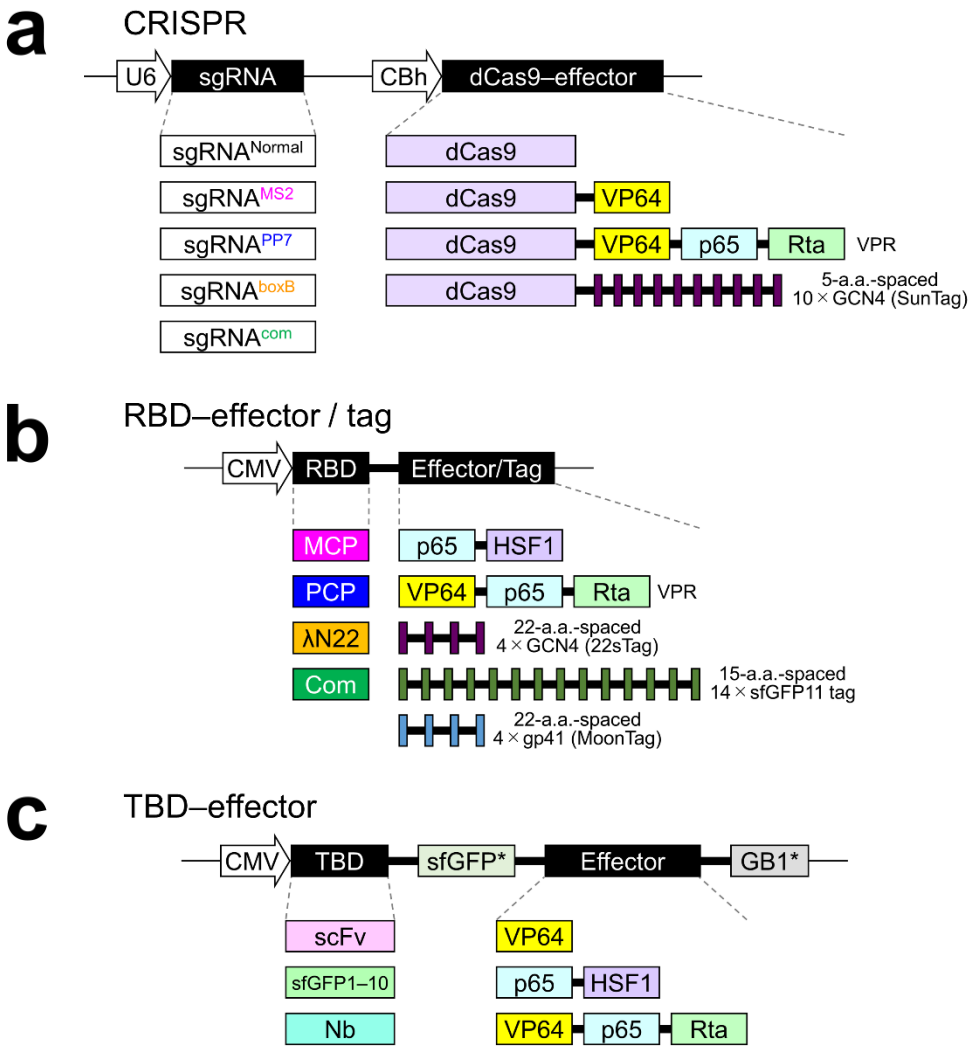

Figure 2-2. Schematic illustration of the expression vectors used in Chapter 2. (a) Modified all-in-one CRISPR–Cas9 vectors expressing sgRNA with/without aptamers and dCas9/dCas9–VP64/dCas9–VPR/dCas9–10×GCN4. sgRNA^{Normal}, dCas9–VPR, and dCas9–10×GCN4 (SunTag) were used for the conventional systems. sgRNA^{Normal} indicates sgRNA without aptamers. dCas9–VPR was also used to assess the activity of the modified sgRNAs. **(b)** Constructs to express the fusion proteins of MCP/PP7/λN22/Com and p65–HSF1/VPR/GCN4 (22sTag)/sfGFP11 tag/gp41 (MoonTag). **(c)** Constructs to express the fusion proteins of scFv (for 22sTag)/sfGFP1–10 (for sfGFP11 tag)/Nb (for MoonTag) and VP64/p65–HSF1/VPR. *sfGFP and GB1 were not included in the sfGFP1–10–effector fusion proteins.

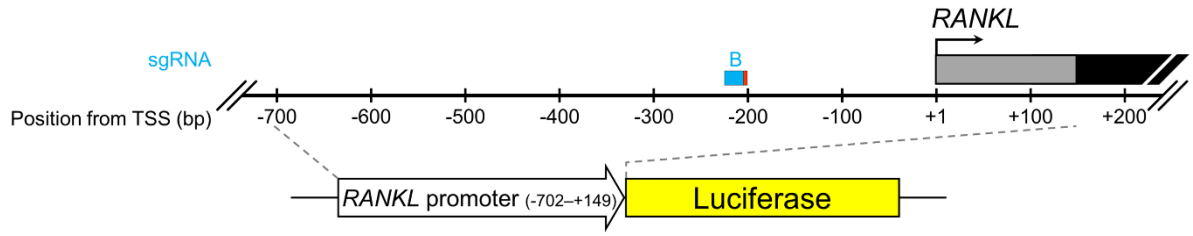

Figure 2-3. Schematic illustration of the *RANKL* reporter. Partial promoter and 5′ UTR sequence of *RANKL* was inserted upstream of the luciferase CDS. sgRNA_B was used for the reporter assay. Blue, red, gray, and black boxes show the protospacer, PAM, 5′ UTR, and CDS, respectively.

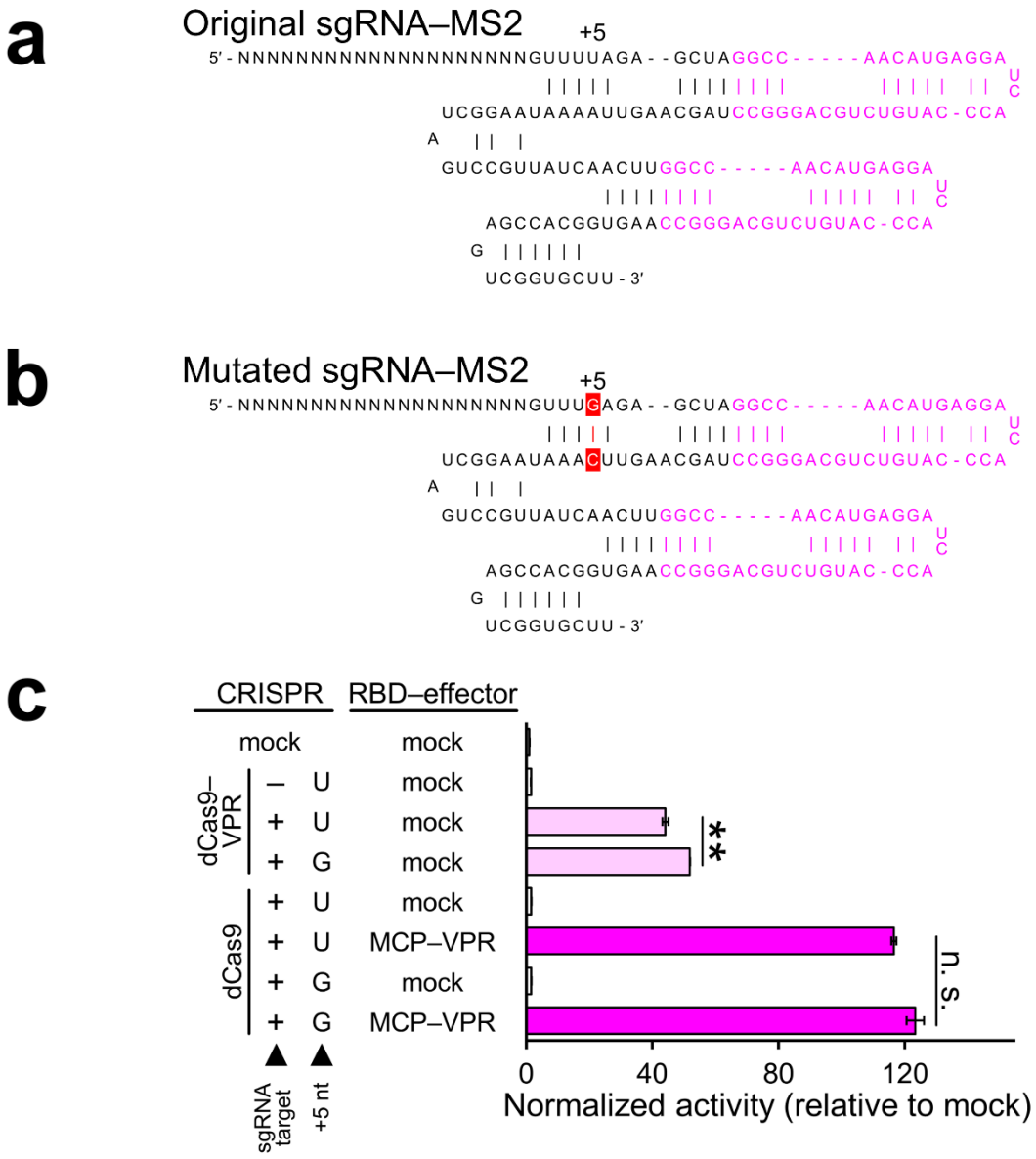

Figure 2-4. Optimization of sgRNA–MS2. (a), (b) Schematic illustration of sgRNA–MS2. The +5 nucleotide was altered from U (**a**, original) to G (**b**), accompanied by a change in its complementary nucleotide (from A to C). **(c)** Activity comparison between the original and mutated sgRNA–MS2. dCas9–VPR was used to assess the activity the sgRNAs themselves, and dCas9 + MCP–VPR were used to assess the efficiency of effector accumulation. Data are shown as the mean ± S.D. ($n = 3$). The differences between two samples with different sgRNA scaffolds were analyzed by two-tailed Welch's *t*-test. $^{**}P < 0.01$. n.s., not significant.

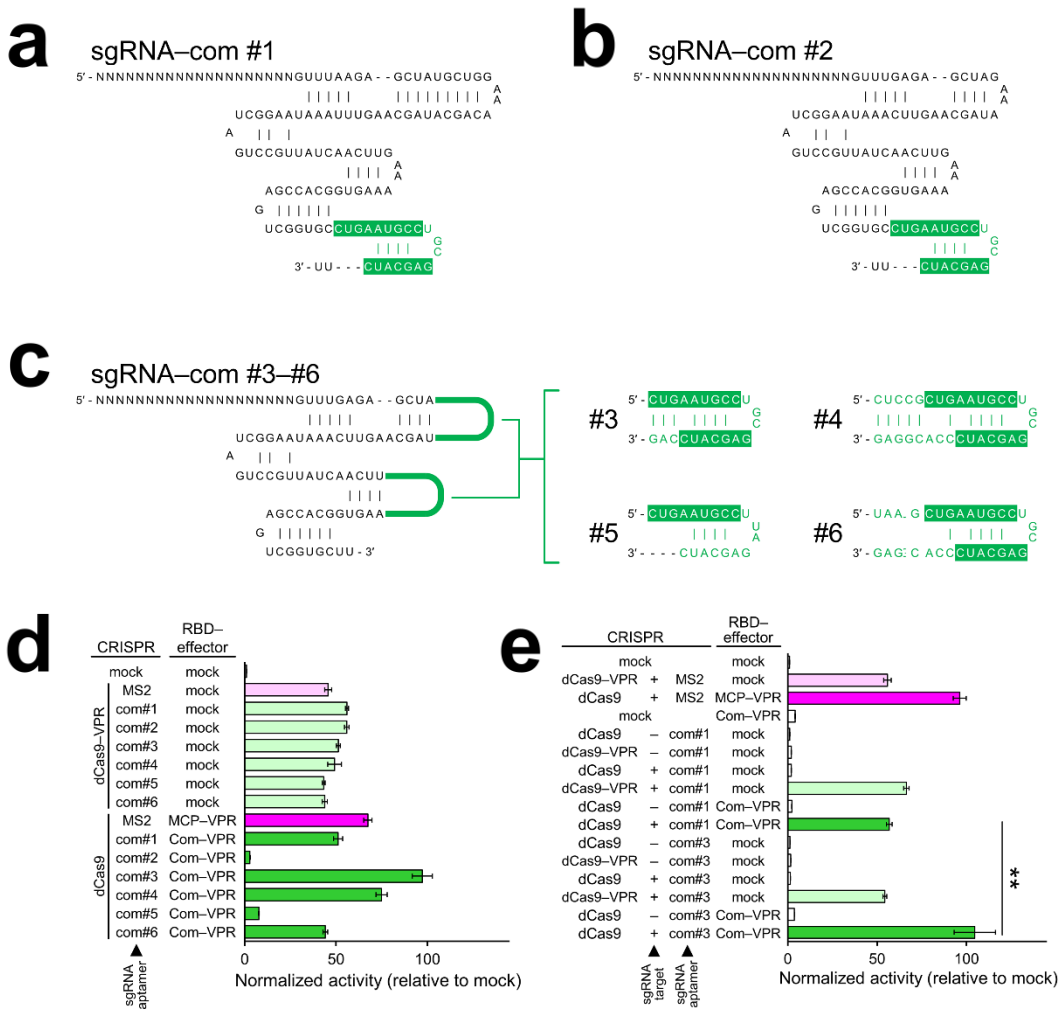

Figure 2-5. Optimization of sgRNA–com. (a)–(c) Schematic illustration of the lineup of the sgRNAs with com aptamers tested as described in Chapter 2. (**a**; com #1): The same sgRNA–com as in a previous report (Zalatan et al., 2015). (**b**; com #2): The scaffold sequence was partially altered from com #1. (**c**; com #3–#6) Two of the corresponding aptamer sequences shown on the right were added on each loop shown by green curves. The nucleotides shown in white letters with a green background have been reported to be required for binding of the Com (Hattman, 1999). **(d)** Activity comparison among six types of sgRNA–com and sgRNA–MS2 as a control. Data are shown as the mean ± S.D. (n = 3). **(e)** Validation of the conventional (com #1) and optimized (com #3) sgRNA–com-based RBD–effector systems. Data are shown as the mean ± S.D. (n = 3). The difference between two samples with VPR accumulation was determined by one-tailed Welch's t-test. $^{**}P < 0.01$.

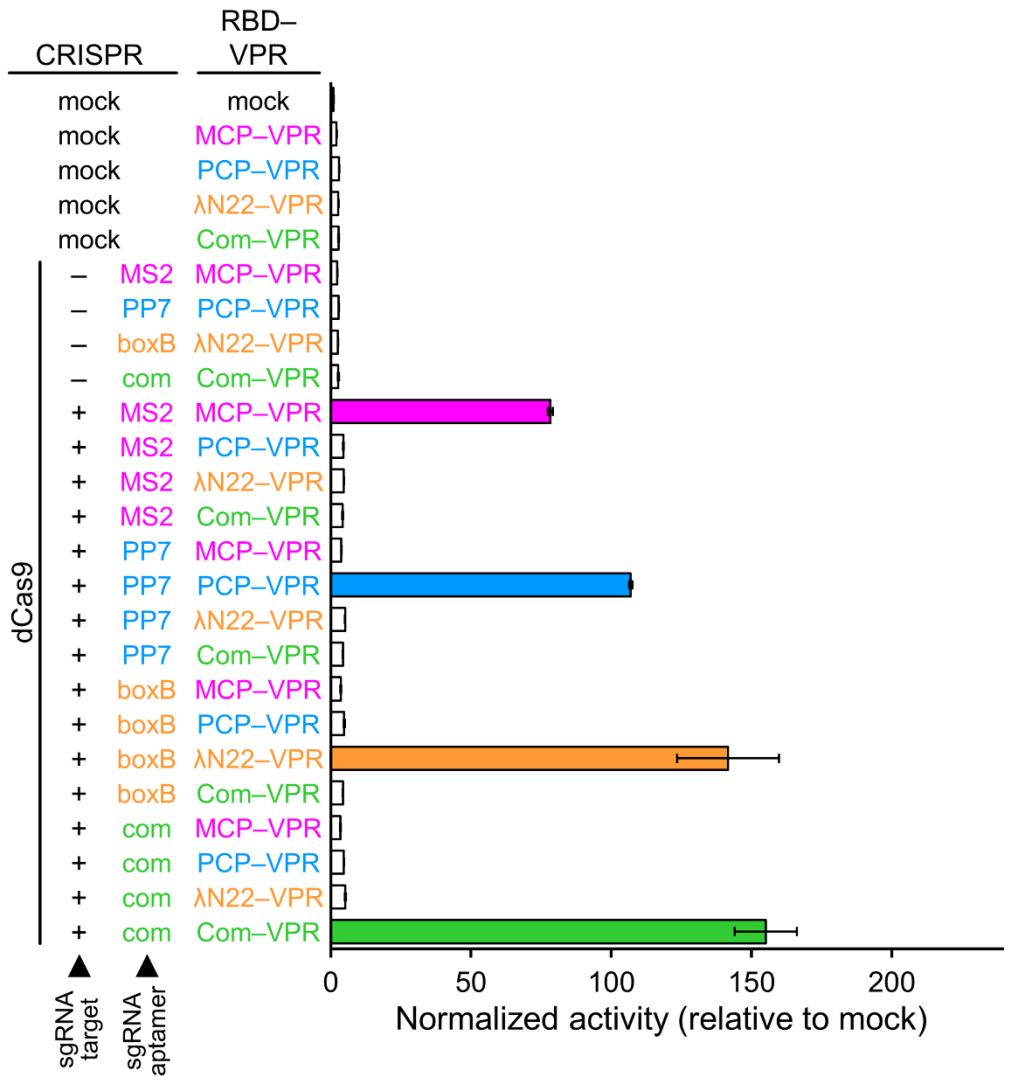

Figure 2-6. Validation of the four RBD–effector systems. The activity of the four types of RBD–effector system and "cross-accumulation" among them were investigated. Data are shown as the mean ± S.D. (n = 3).

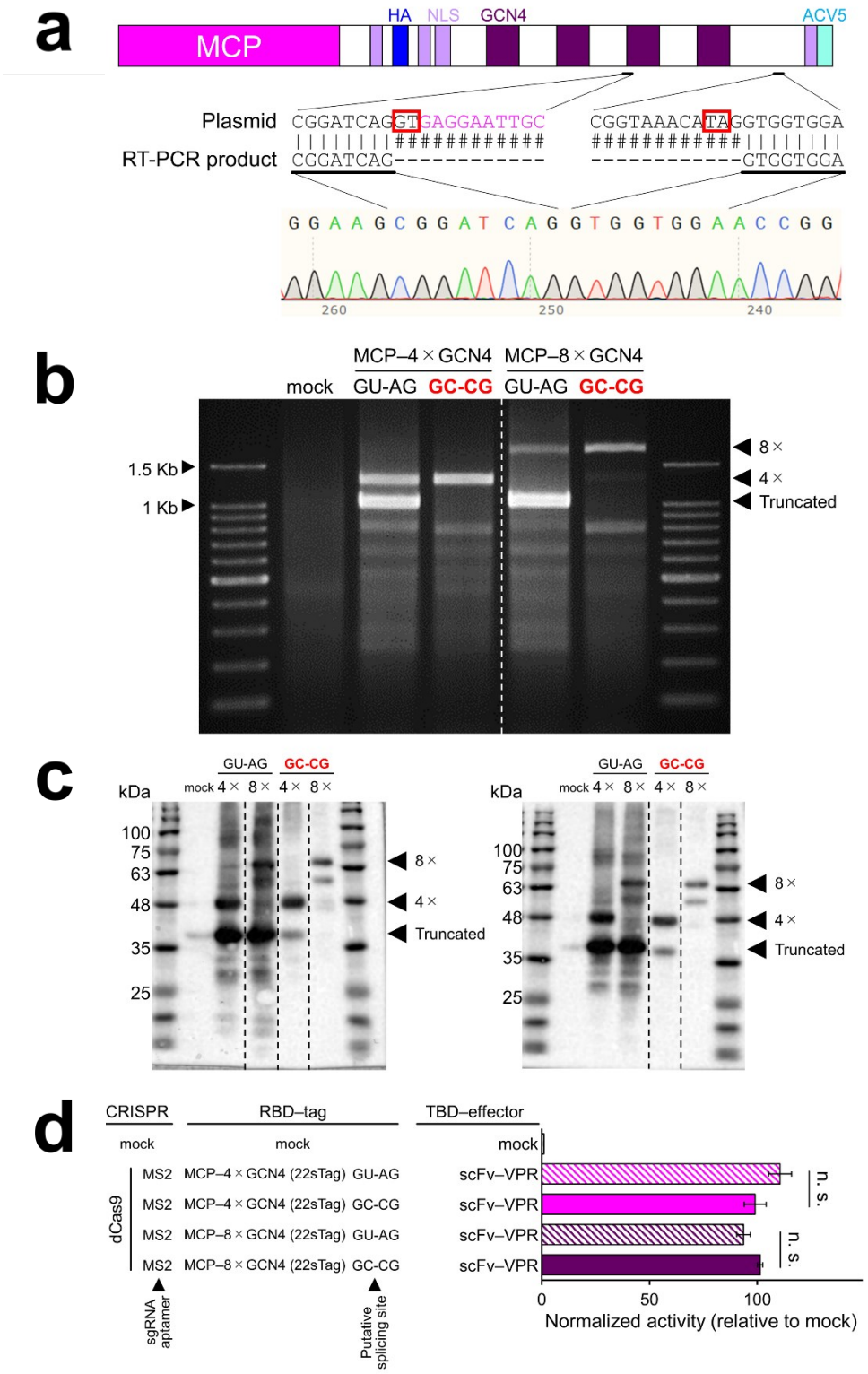

Figure 2-7 (*legend on next page*)

Figure 2-7. Optimization of the coding sequence of 22sTag. (a) Schematic illustration of the original MCP–4×GCN4 protein and the results of sequencing analysis of the cDNA from MCP–4×GCN4-introduced HEK293T cells. The red frames show the putative splicing sites in the "GU-AG" manner. The signals were visualized on SnapGene Viewer software version 5.2.2. **(b)** RT-PCR analysis of the original and splicing-mutated MCP–GCN4. The full lengths of the cDNA sequences are as follows: MCP–4×GCN4: 1,290 bp; MCP–8×GCN4: 1,782 bp. **(c)** Immunoblotting analysis of the original and splicing-mutated MCP–GCN4, using the anti-HA (left) and anti-ACV5 (right) antibodies. The putative size of the proteins was as follows: MCP–4×GCN4: 42.2 kDa, MCP–8×GCN4: 57.5 kDa. **(d)** Activity comparison between the original and splicing-mutated MCP–GCN4. Data are shown as the mean ± S.D. (n = 3). The differences were analyzed by two-tailed Welch's t-test. n.s., not significant.

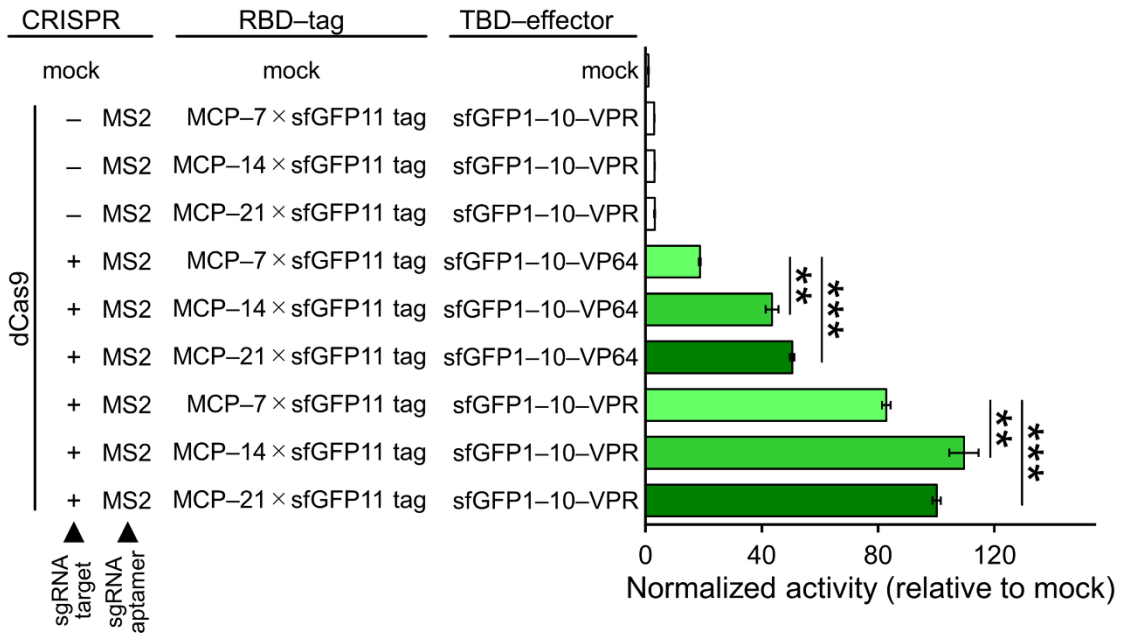

Figure 2-8. Optimization of sfGFP11 tag. Activity comparison among different tag numbers of MCP–sfGFP11. Data are shown as the mean ± S.D. (n = 3). The differences between MCP–7×sfGFP11 and MCP–14× or –21×sfGFP11 were analyzed by two-tailed Welch's t-test. **P < 0.01; ***P < 0.001.

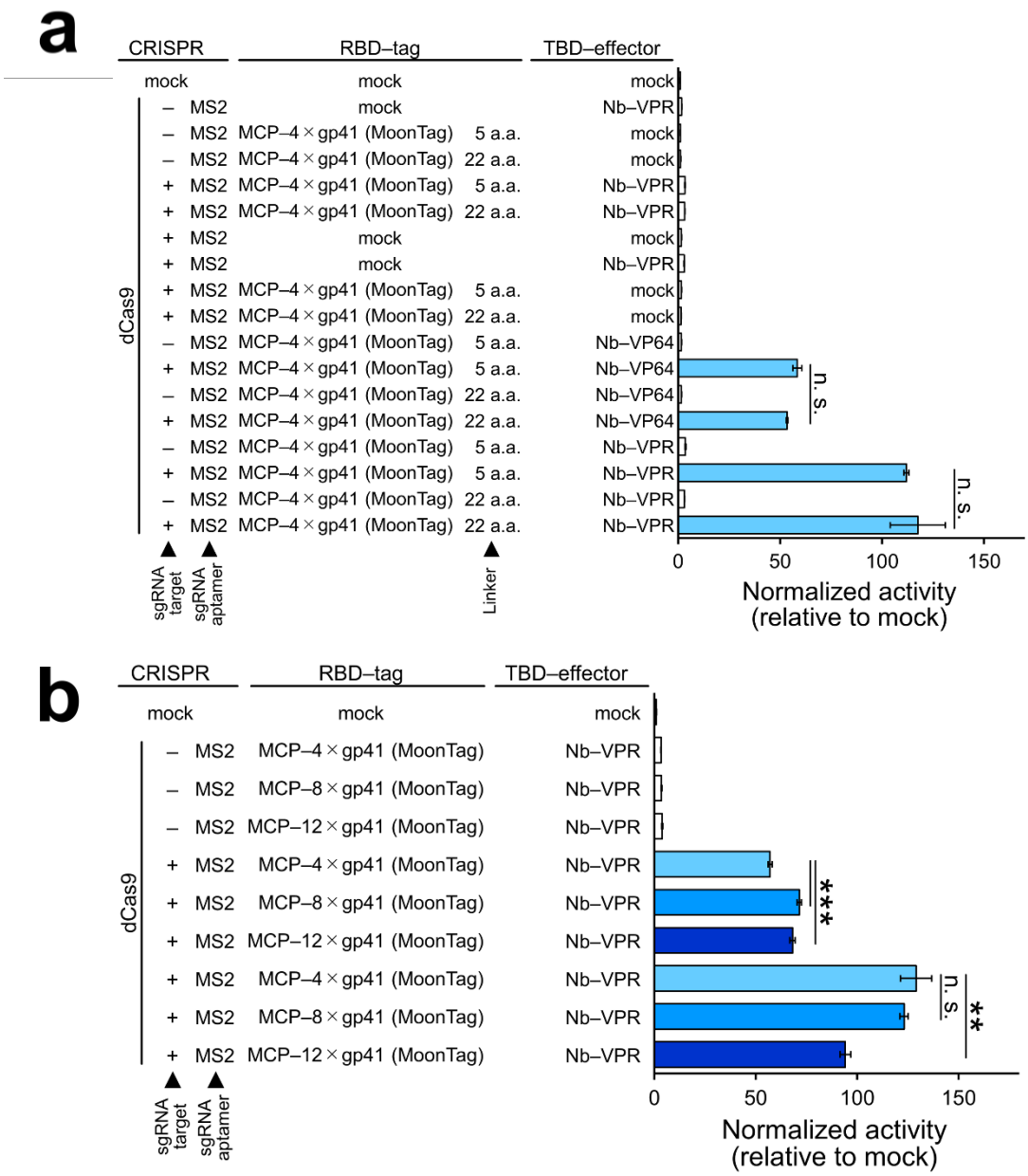

Figure 2-9. Optimization of MoonTag. (a) Activity comparison between 5-a.a.- and 22-a.a.-spaced MCP–4×gp41. Data are shown as the mean ± S.D. (n = 3). The differences were analyzed by two-tailed Welch's t-test. n.s., not significant. **(b)** Activity comparison among different tag numbers of MCP–gp41. Data are shown as the mean ± S.D. (n = 3). The differences between MCP–4×gp41 and MCP–8× or –12×gp41 were analyzed by two-tailed Welch's t-test. $^{**}P < 0.01$; $^{***}P < 0.001$. n.s., not significant.

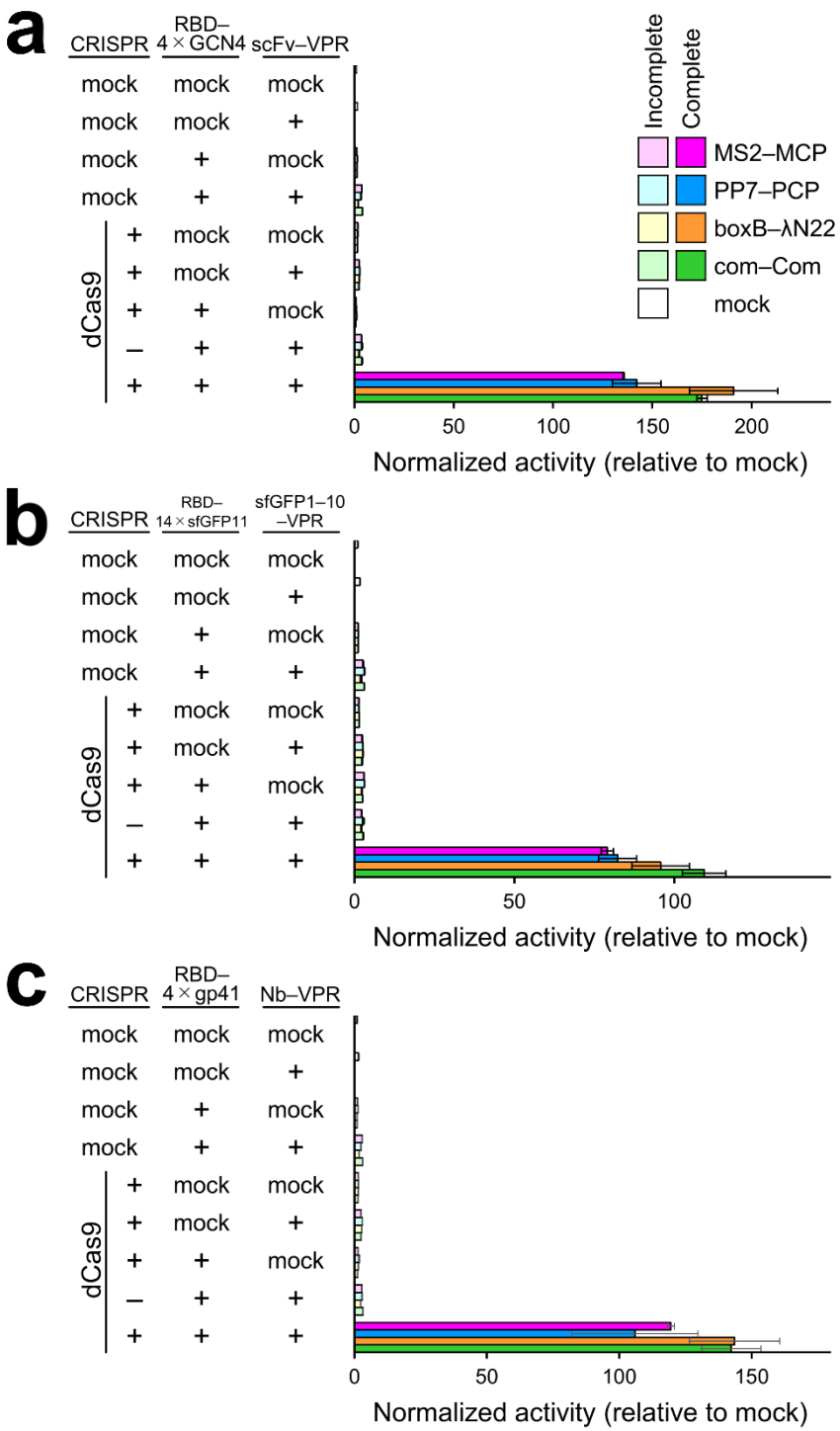

Figure 2-10. Basic validation of RBD–tag systems. Validation of RBD–GCN4 (22sTag; **a**), RBD–sfGFP11 tag (**b**), and RBD–gp41 (MoonTag; **c**) was performed by *RANKL* reporter assay in HEK293T cells.

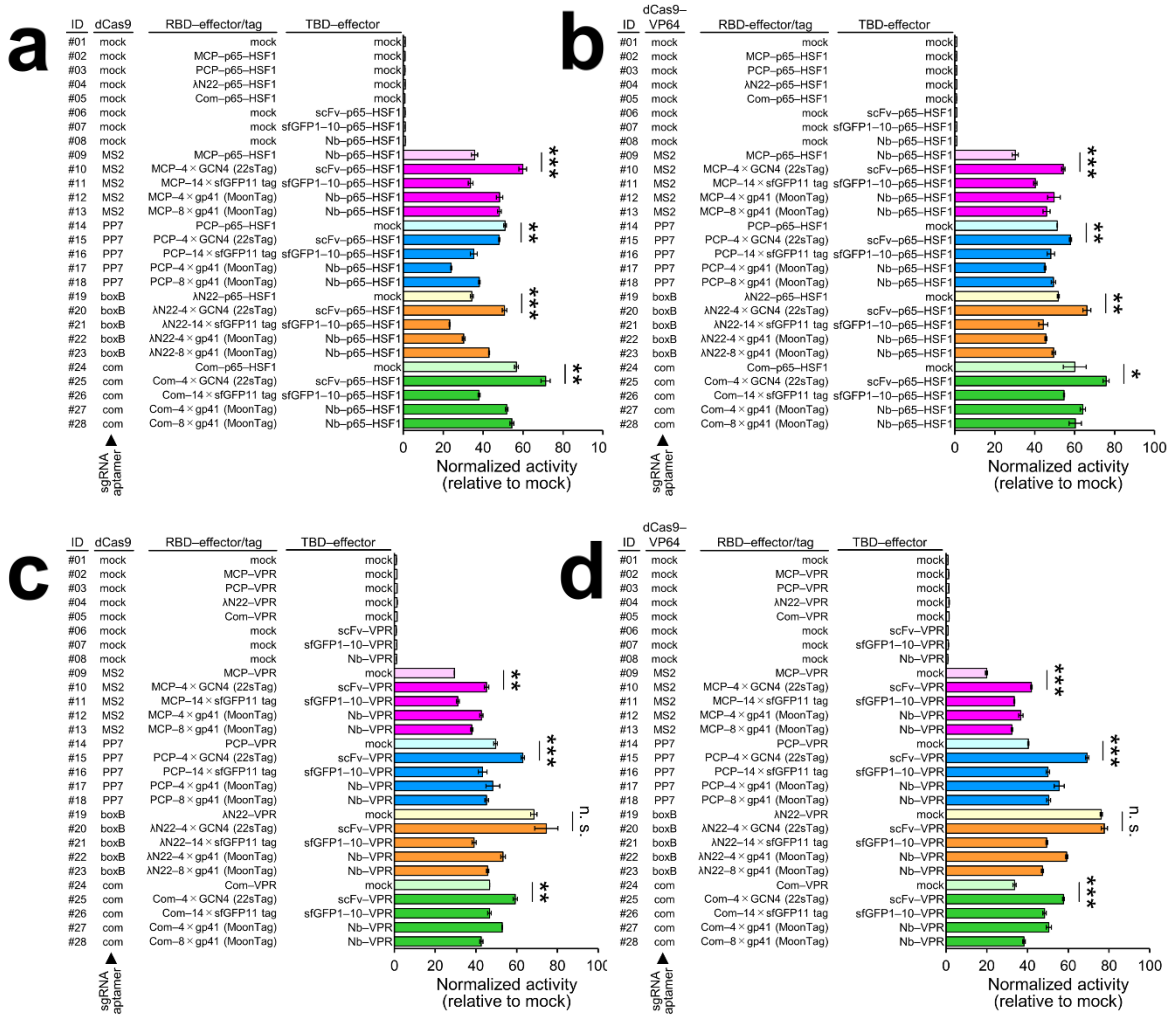

Figure 2-11. Activity comparisons among the EARTH collection by reporter assay. Activity comparisons among the platforms of the EARTH collection were performed using dCas9 + *trans*-p65–HSF1 (**a**), dCas9–VP64 + *trans*-p65–HSF1 (**b**), dCas9 + *trans*-VPR (**c**), and dCas9–VP64 + *trans*-VPR (**d**). Data are shown as the mean ± SD (n = 3). The differences between the RBD–effector and RBD–4×GCN4 systems were determined by two-tailed Welch's *t*-test. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. n.s., not significant.

a

Cell line	*RANKL*/*GUSB*	*RANKL*/*TBP*	Average
HEK293T	6.24×10^{-5}	3.34×10^{-5}	4.79×10^{-5}
MCF-7	n. d.	n. d.	n/a

Cell line	*MMP9*/*GUSB*	*MMP9*/*TBP*	Average
HEK293T	9.20×10^{-4}	1.42×10^{-3}	1.17×10^{-3}
MCF-7	5.20×10^{-4}	7.50×10^{-4}	6.34×10^{-4}

Cell line	*CTCFL*/*GUSB*	*CTCFL*/*TBP*	Average
HEK293T	2.05×10^{-3}	3.54×10^{-3}	2.80×10^{-3}
MCF-7	3.44×10^{-5}	n. d.	n/a

b

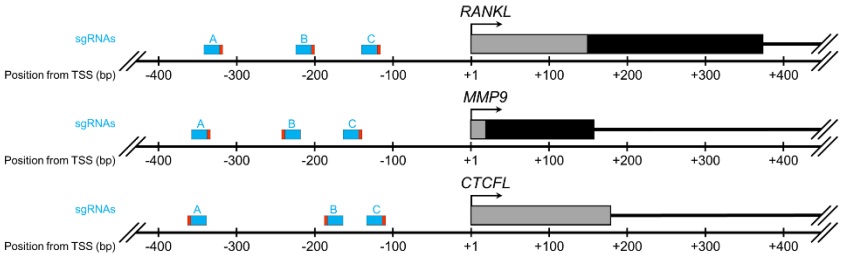

Figure 2-12. Designed sgRNAs and background expression levels of the three target genes. **(a)** Background expression levels of the three target genes in HEK293T and MCF-7 cells. Data are shown as the ratios of each target gene (*RANKL*, *MMP9*, and *CTCFL*) and reference gene (*GUSB* and *TBP*) ($n = 1$). n.d., not detected. n/a, not available. **(b)** Three sgRNAs were designed on each of *RANKL* (top), *MMP9* (middle), and *CTCFL* (bottom). Blue, red, gray, and black boxes show the protospacer, PAM, 5′ UTR, and CDS, respectively.

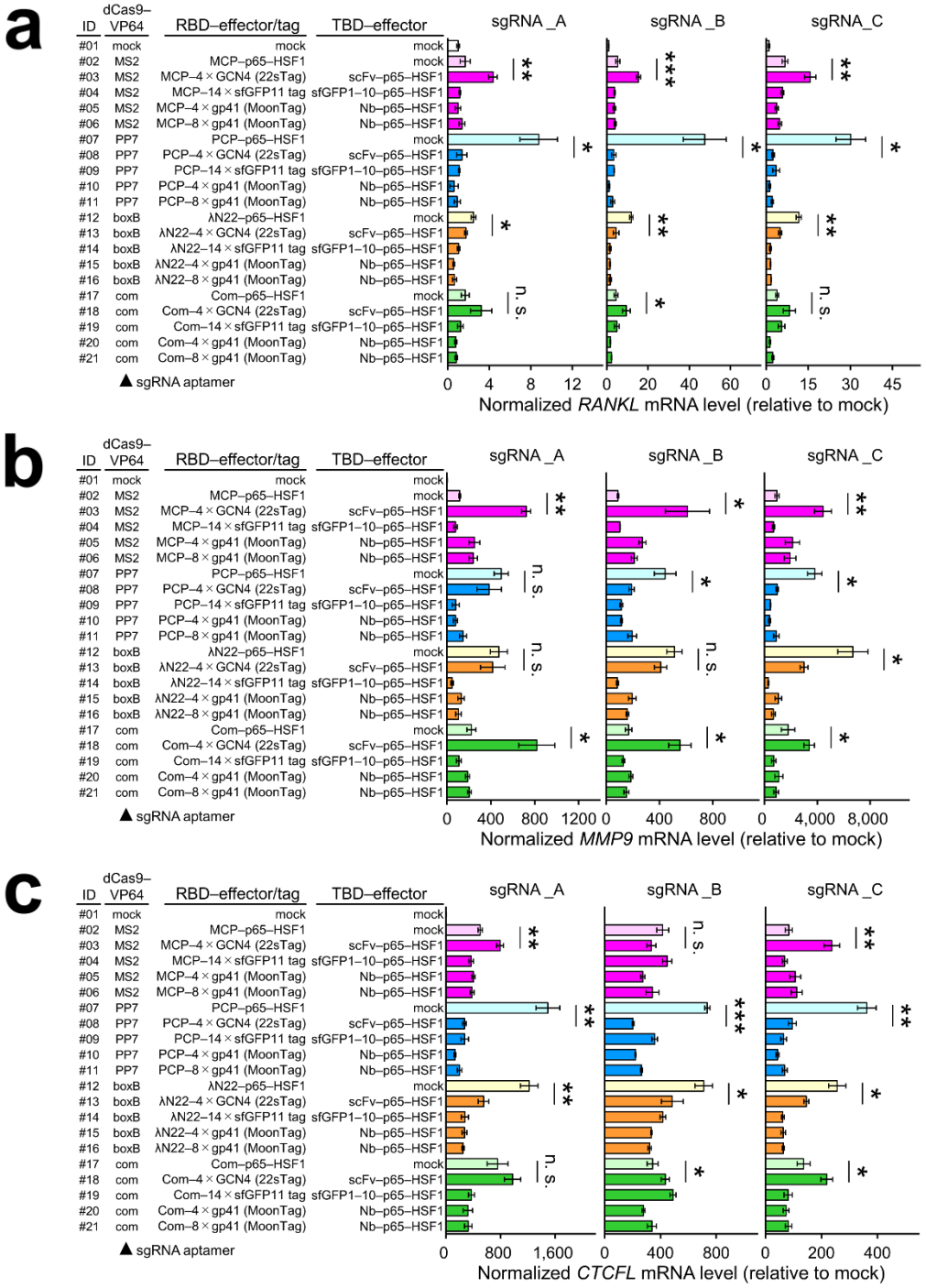

Figure 2-13. Activation of endogenous genes by p65–HSF1 accumulation via various platforms of the EARTH collection. Activation of endogenous *RANKL* (**a**), *MMP9* (**b**), and *CTCFL* (**c**) genes, using dCas9–VP64 + *trans*-p65–HSF1 in HEK293T cells. Data are shown as the mean ± S.D. ($n = 3$). The differences between the RBD–p65–HSF1 and RBD–4×GCN4 systems were analyzed by two-tailed Welch's *t*-test. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. n.s., not significant.

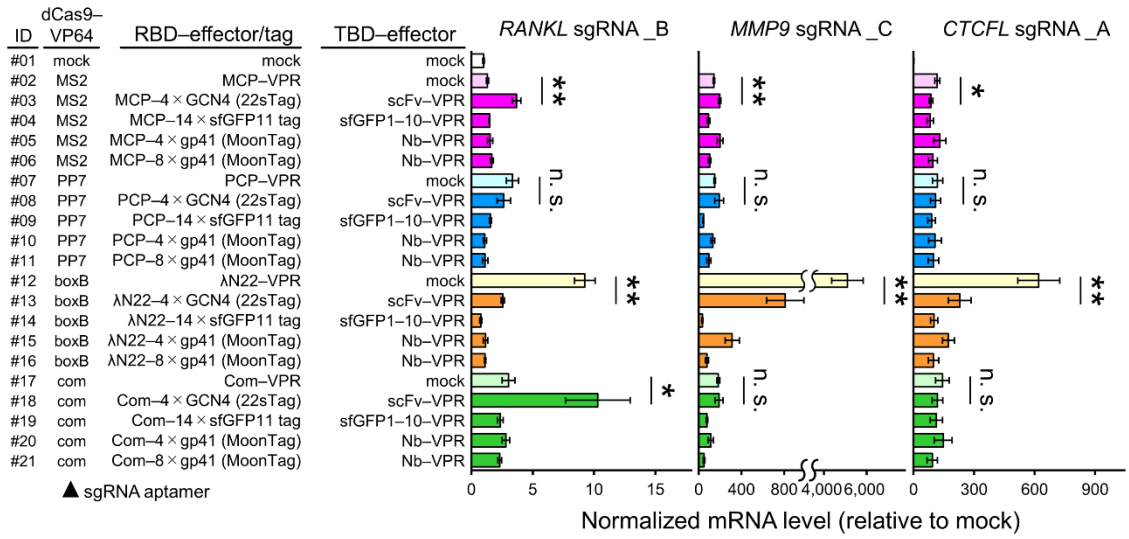

Figure 2-14. Activation of endogenous genes by VPR accumulation via various platforms of the EARTH collection. Activation of endogenous *RANKL* (left), *MMP9* (middle), and *CTCFL* (right) genes, using dCas9–VP64 + *trans*-VPR in HEK293T cells. Data are shown as the mean ± S.D. (n = 3). The differences between the RBD–VPR and RBD–4×GCN4 systems were analyzed by two-tailed Welch's t-test. *P < 0.05; **P < 0.01. n.s., not significant.

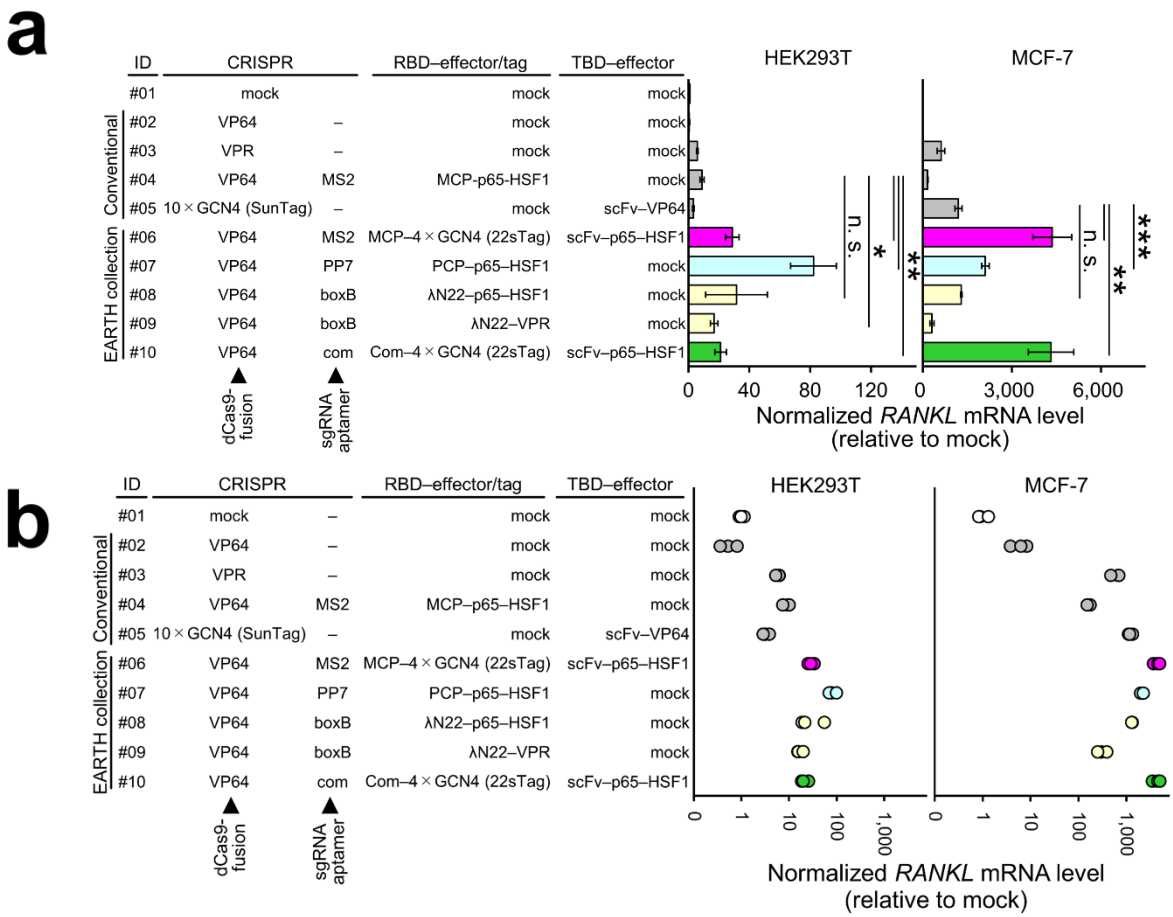

Figure 2-15. Activity comparison among the conventional second-generation systems and selected platforms from the EARTH collection targeting endogenous *RANKL* gene. (a) Activation of endogenous *RANKL* gene using conventional CRISPRa platforms (#02: dCas9–VP64; #03: dCas9–VPR; #04: SAM; #05: dCas9–SunTag) and selected ones from the EARTH collection (#06–#10) with sgRNA_B in HEK293T (left) and MCF-7 (right) cells. Data are shown as the mean ± S.D. (n = 3). The differences between two samples, one being the system with the highest average value among the conventional CRISPRa toolkits (#02–05) and the other being the system of the EARTH collection with a higher average value than the former, were analyzed by one-tailed Welch's t-test. *P < 0.05; **P < 0.01; ***P < 0.001. n.s., not significant. **(b)** Log-scaled description of the same data as in (**a**).

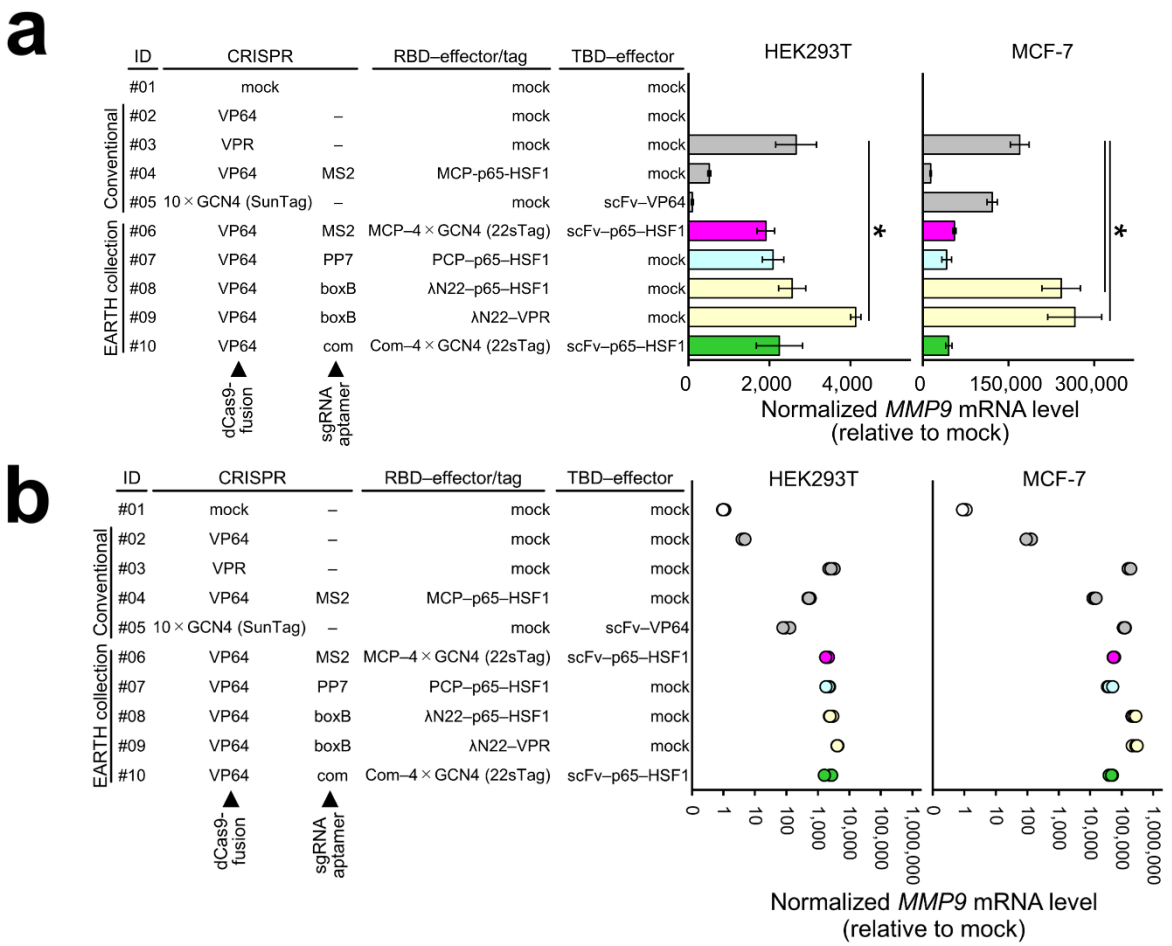

Figure 2-16. Activity comparison among the conventional second-generation systems and selected platforms from the EARTH collection targeting endogenous *MMP9* gene. (a) Activation of endogenous *MMP9* gene using conventional CRISPRa platforms (#02: dCas9–VP64; #03: dCas9–VPR; #04: SAM; #05: dCas9–SunTag) and selected ones from the EARTH collection (#06–#10) with sgRNA_C in HEK293T (left) and MCF-7 (right) cells. Data are shown as the mean ± S.D. (n = 3). The differences between the two samples, one being the system with the highest average value among the conventional CRISPRa toolkits (#02–05) and the other being the system of the EARTH collection with a higher average value than the former, were analyzed by one-tailed Welch's t-test. *P < 0.05; **P < 0.01; ***P < 0.001. n.s., not significant. **(b)** Log-scaled description of the same data as in (**a**).

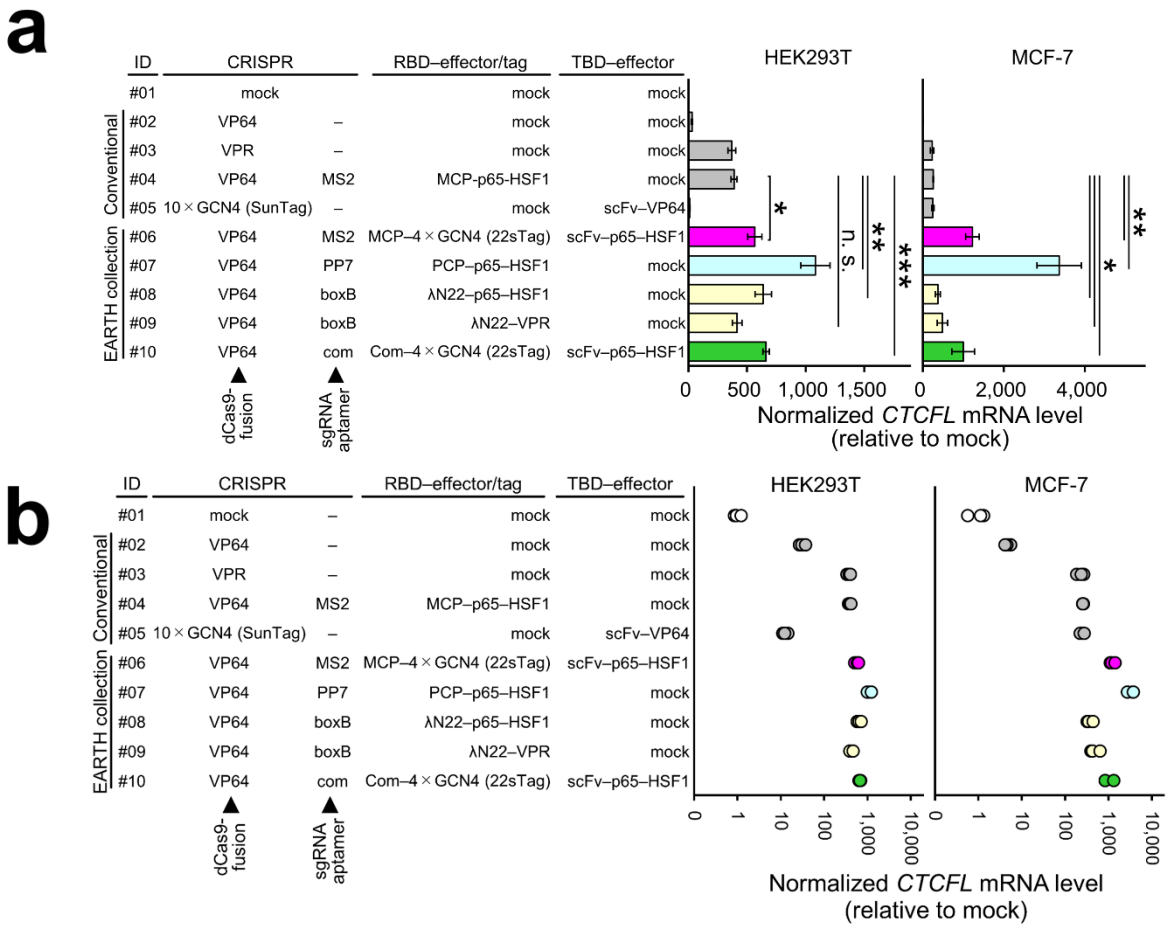

Figure 2-17. Activity comparison among the conventional second-generation systems and selected platforms from the EARTH collection targeting endogenous *CTCFL* gene. (a) Activation of endogenous *CTCFL* gene using conventional CRISPRa platforms (#02: dCas9–VP64; #03: dCas9–VPR; #04: SAM; #05: dCas9–SunTag) and selected ones from the EARTH collection (#06–#10) with sgRNA_A in HEK293T (left) and MCF-7 (right) cells. Data are shown as the mean ± S.D. (n = 3). The differences between the two samples, one being the system with the highest average value among the conventional CRISPRa toolkits (#02–05) and the other being the system of the EARTH collection with a higher average value than the former, were analyzed by one-tailed Welch's t-test. *P < 0.05; **P < 0.01; ***P < 0.001. n.s., not significant. **(b)** Log-scaled description of the same data as in (**a**).

Sequence 2-1

Modified sgRNAs

sgRNA–MS2 (+5th nt: U)

NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGGCCAACATGAGGATCACCCATGTCTGCAG
GGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAACATGAGGATCACCCAT
GTCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTTTTT

sgRNA–MS2 (+5th nt: G)

NNNNNNNNNNNNNNNNNNNNGTTTGAGAGCTAGGCCAACATGAGGATCACCCATGTCTGCAG
GGCCTAGCAAGTTCAAATAAGGCTAGTCCGTTATCAACTTGGCCAACATGAGGATCACCCAT
GTCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTTTTT

sgRNA–PP7

NNNNNNNNNNNNNNNNNNNNGTTTGAGAGCTACCGGAGCAGACGATATGGCGTCGCTCCGGT
AGCAAGTTCAAATAAGGCTAGTCCGTTATCAACTTGGAGCAGACGATATGGCGTCGCTCCAA
GTGGCACCGAGTCGGTGCTTTTTTT

sgRNA–boxB

NNNNNNNNNNNNNNNNNNNNGTTTGAGAGCTAGGGCCCTGAAGAAGGGCCCTAGCAAGTTCA
AATAAGGCTAGTCCGTTATCAACTTGGGCCCTGAAGAAGGGCCCAAGTGGCACCGAGTCGGT
GCTTTTTTT

sgRNA–com#1

NNNNNNNNNNNNNNNNNNNNGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAA
GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCCTGAATGCCTGCGAGCAT
CTTTTTTT

sgRNA–com#2

NNNNNNNNNNNNNNNNNNNNGTTTGAGAGCTAGAAATAGCAAGTTCAAATAAGGCTAGTCCG
TTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCCTGAATGCCTGCGAGCATCTTTTTTT

sgRNA–com#3

NNNNNNNNNNNNNNNNNNNNGTTTGAGAGCTACTGAATGCCTGCGAGCATCTAGCAAGTTCA
AATAAGGCTAGTCCGTTATCAACTTCTGAATGCCTGCGAGCATCAAGTGGCACCGAGTCGGT
GCTTTTTTT

sgRNA–com#4

NNNNNNNNNNNNNNNNNNNNGTTTGAGAGCTATAATGCTGAATGCCTGCGAGCATCCCACGG
AGTAGCAAGTTCAAATAAGGCTAGTCCGTTATCAACTTTAATGCTGAATGCCTGCGAGCATC
CCACGGAGAAGTGGCACCGAGTCGGTGCTTTTTTT

sgRNA–com#5

NNNNNNNNNNNNNNNNNNNNGTTTGAGAGCTACTGAATGCCTGCGAGCATCCAGTAGCAAGT
TCAAATAAGGCTAGTCCGTTATCAACTTCTGAATGCCTGCGAGCATCCAGAAGTGGCACCGA
GTCGGTGCTTTTTTT

sgRNA–com#6

NNNNNNNNNNNNNNNNNNNNGTTTGAGAGCTACTCCGCTGAATGCCTGCGAGCATCCCACGG
AGTAGCAAGTTCAAATAAGGCTAGTCCGTTATCAACTTCTCCGCTGAATGCCTGCGAGCATC
CCACGGAGAAGTGGCACCGAGTCGGTGCTTTTTTT

Oligonucleotides used for the sgRNA templates

sgRNA	Strand*	Sequence (5′–3′)	Position from TSS (bp)
*RANKL*_A	s	caccgCAAACTAGAATGGATGCAGG	-342 – -323
	as	aaacCCTGCATCCATTCTAGTTTGc	
*RANKL*_B	s	caccgCAAGGGGAGTCTGGAACCAC	-224 – -205
	as	aaacGTGGTTCCAGACTCCCCTTGc	
*RANKL*_C	s	caccgTGAGAGAGAGGGAGGGCGAA	-140 – -121
	as	aaacTTCGCCCTCCCTCTCTCTCAc	
*MMP9*_A	s	caccgACTGGAGGCTTTCAGACCAA	-357 – -338
	as	aaacTTGGTCTGAAAGCCTCCAGTc	
*MMP9*_B	s	caccgCCAGACCCGCAGGAAACCGC	-238 – -219
	as	aaacGCGGTTTCCTGCGGGTCTGGc	
*MMP9*_C	s	caccGTGTAAGCCCTTTCTCATGC	-164 – -145
	as	aaacGCATGAGAAAGGGCTTACAC	
*CTCFL*_A	s	caccGGGAGCCCTGCGGGGGGCGA	-359 – -340
	as	aaacTCGCCCCCCGCAGGGCTCCC	
*CTCFL*_B	s	caccGGGAGCCCTGCGGGGGGCGA	-184 – -165
	as	aaacTCGCCCCCCGCAGGGCTCCC	
*CTCFL*_C	s	caccgTCCCCTCCCCTAGGTGTCAC	-133 – -114
	as	aaacGTGACACCTAGGGGAGGGGAc	

*s, sense strand; as, antisense strand

Synthesized dsDNA sequences for RBDs and tags

PCP

ATGGGTTCCAAAACCATCGTTCTTTCGGTCGGCGAGGCTACTCGCACTCTGACTGAGATCCA
GTCCACCGCAGACCGTCAGATCTTCGAAGAGAAGGTCGGGCCTCTGGTGGGTCGGCTGCGCC
TCACGGCTTCGCTCCGTCAAAACGGAGCCAAGACCGCGTATCGCGTCAACCTAAAACTGGAT
CAGGCGGACGTCGTTGATTCCGGACTTCCGAAAGTGCGCTACACTCAGGTATGGTCGCACGA
CGTGACAATCGTTGCGAATAGCACCGAGGCCTCGCGCAAATCGTTGTACGATTTGACCAAGT
CCCTCGTCGCGACCTCGCAGGTCGAAGATCTTGTCGTCAACCTTGTGCCGCTGGGCCGT

Com

ATGAAATCAATTCGCTGTAAAAACTGCAACAAACTGTTATTTAAGGCGGATAGTTTTGATCA
CATTGAAATCAGGTGTCCGCGTTGCAAACGTCACATCATAATGCTGAATGCCTGCGAGCATC
CCACGGAGAAACATTGTGGGAAAAGAGAAAAAATCACGCATTCTGACGAAACCGTGCGTTAT

7×sfGFP11 tag

ATGCGTGACCACATGGTCCTTCATGAGTATGTAAATGCTGCTGGGATTACAGGTGGCTCTGG
AAGTTCAGGTGGAGGCTCGGGTGGCGGCAGTTCGAGAGATCATATGGTTCTCCACGAATACG
TTAACGCCGCAGGCATCACTGGCAGTGGTGGATCTGGCAGCGGGAGCGGCTCTGGAGGTAGC
AGTCGCGACCATATGGTACTACATGAATATGTCAATGCAGCCGGAATAACCGGATCCGGAAG
TGGCTCAAGCGGAGGAGGAAGTAGTGGAAGTTCTCGGGATCACATGGTGCTGCATGAGTATG
TGAACGCGGCGGGTATAACTGGTTCGGGAGGCTCAGGTAGCGGCAGTTCAGGAGGAAGCGGG
TCCCGAGACCATATGGTGCTTCACGAATACGTAAACGCAGCTGGCATTACTGGGTCAGGAGG
TTCAGGAGGGTCTGGTTCTGGATCAGGAGGTAGCAGGGATCACATGGTACTCCATGAGTACG
TGAACGCTGCTGGAATCACAGGCGGTAGCAGTGGTGGAAGTAGCGGCAGCGGCGGCAGTAGC
TCACGGGACCATATGGTCCTGCACGAATATGTCAATGCTGCCGGTATCACCGGGAGTGGTGG
GTCCGGCGGGAAATTCATG

4×gp41 (MoonTag; 5-a.a.-spaced)

AAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGGGTTCTGGCAGTGGAAA
AAACGAACAGGAACTGTTGGAGTTGGACAAATGGGCTAGCTTGGGGAGCGGAAGTGGGAAGA
ATGAGCAAGAGTTGCTGGAACTGGATAAATGGGCCTCCCTGGGGTCCGGATCGGGTAAAAAT
GAACAAGAATTGTTGGAATTGGATAAGTGGGCTTCCTTG

4×gp41 (MoonTag; 22-a.a.-spaced)

AAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGGGTTCTGGCTCCGGAGG
CAGTGGTTCTGGAAGCGGTGGCAGCGGGTCAGGTGGAAGCGGATCAGGTAAAAACGAACAGG
AACTGTTGGAGTTGGACAAATGGGCTAGCTTGGGGTCGGGAAGTGGCGGCAGCGGAAGTGGG
AGTGGAGGGAGCGGTTCTGGCGGTTCCGGCAGTGGAAAGAATGAGCAAGAGTTGCTGGAACT
GGATAAATGGGCCTCCCTGGGCAGCGGCTCGGGGGGTAGTGGATCGGGGAGTGGCGGGTCAG
GAAGCGGTGGTAGCGGAAGCGGGAAAAATGAACAAGAATTGTTGGAATTGGATAAGTGGGCT
TCCTTG

sfGFP1–10

ATGTCCAAAGGAGAAGAACTGTTTACCGGTGTTGTGCCAATTTTGGTTGAACTCGATGGTGA
TGTCAACGGACATAAGTTCTCAGTGAGAGGCGAAGGAGAAGGTGACGCCACCATTGGAAAAT
TGACTCTTAAATTCATCTGTACTACTGGTAAACTTCCTGTACCATGGCCGACTCTCGTAACA
ACGCTTACGTACGGAGTTCAGTGCTTTTCGAGATACCCAGACCATATGAAAAGACATGACTT
TTTTAAGTCGGCTATGCCTGAAGGTTACGTGCAAGAAAGAACAATTTCGTTCAAAGATGATG
GAAAATATAAAACTAGAGCAGTTGTTAAATTTGAAGGAGATACTTTGGTTAACCGCATTGAA
CTGAAAGGAACAGATTTTAAAGAAGATGGTAATATTCTTGGACACAAACTCGAATACAATTT
TAATAGTCATAACGTATACATCACTGCTGATAAGCAAAAGAACGGAATTAAAGCGAATTTCA
CAGTACGCCATAATGTAGAAGATGGCAGTGTTCAACTTGCCGACCATTACCAACAAAACACC
CCTATTGGAGACGGTCCGGTACTTCTTCCTGATAATCACTACCTCTCAACACAAACAGTCCT
GAGCAAAGATCCAAATGAAAAA

anti-gp41 Nanobody (Nb)

ATGGAGGTGCAGCTGGTGGAATCTGGGGGCGGACTGGTGCAGCCCGGGGGATCTCTGCGGCT
GTCCTGCGCCGCCTCTGGCTCCATCTCTAGCGTGGATGTGATGTCCTGGTACAGGCAGGCCC
CTGGCAAGCAGAGGGAACTGGTGGCCTTCATTACTGATAGGGGAAGAACCAATTACAAAGTG
AGCGTGAAAGGCCGCTTCACTATCAGCCGGGATAATTCTAAGAATATGGTGTATCTGCAGAT
GAACAGCCTGAAGCCAGAAGACACTGCCGACTATCTGTGCAGGGCTGAGTCCCGGACTTCCT
GGTCCAGCCCATCTCCTCTGGATGTGTGGGGGCGCGGCACTCAGGTGACTGTGTCCTCCTTG
GATCCAGGTGGAGGTGGAAGCGGT

Conclusions

In my series of studies, I first developed a novel format of CRISPR-based artificial transcription activators, named "TREE," which enabled high-powered induction of gene expression by accumulating a number of effector domains via hierarchical architecture of RNA aptamers and protein tags. Second, I constructed a series of variously patterned RNA-binding protein (RBD)–effector and RBD–tag systems, named the "EARTH" collection, and performed systematic comparisons of the activity of its lineup.

Following the publication of the first report on the dCas9–VP64 system in 2013, several types of highly efficient gene activator, generally called "second-generation" systems, have been developed (Chavez et al., 2015; Gilbert et al., 2013; Konermann et al., 2015; Tanenbaum et al., 2014). However, their efficacy has been reported to be affected by factors such as the tested organism, cell line, and target gene (Chavez et al., 2016).

My TREE system showed higher activity than SAM, which was reported to often be the best among the second-generation systems (Chavez et al., 2016), and therefore it would potentially be helpful for the activation of strongly silenced genes, and the identification of novel genes in genome-wide activation screening (Gilbert et al., 2014). Additionally, my EARTH collection, some of which outperformed the three most widely used second-generation systems, would provide a strategy to optimize the activation platform on any target gene.

My systems described above can potentially be diverted to CRISPR-based technologies other than transcriptional activation, such as transcriptional inhibition (Gilbert et al., 2013), epigenome editing (Hilton et al., 2015), base editing (Komor et al., 2016), and chromosome visualization (Chen et al., 2013). It would also be possible to perform several dCas9-based applications simultaneously. I hope that the systems that I constructed will be used by researchers globally and contribute to various fields of research in the future.

References

Chavez, A., Scheiman, J., Vora, S., Pruitt, B., Tuttle, M., Iyer, E., Kiani, S., Guzman, C., Wiegand, D., Ter-Ovanesyan, D., et al. (2015). Highly-efficient Cas9-mediated transcriptional programming. Nat. Methods *12*, 326–328.

Chavez, A., Tuttle, M., Pruitt, B.W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S.J., Cecchi, R.J., Kowal, E.J.K., Buchthal, J., et al. (2016). Comparison of Cas9 activators in multiple species. Nat. Methods *13*, 563–567.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479–1491.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442–451.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420–424.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell *159*, 635–646.

Acknowledgements

I would like to express my sincere appreciation to Prof. Takashi Yamamoto and Dr. Tetsushi Sakuma for giving me a chance to perform this study, and for their guidance, discussion, invaluable suggestions, and constant support. I also thank to Dr. Naoaki Sakamoto and Dr. Shota Nakade for guide on my experiments.

I wish to thank Mr. Yoshihiro Hara and Ms. Shiho Nakamura for helpful discussions and scientific advices. I also thank Mr. Mitsumasa Takenaga and Ms. Tomoko Kurisu for supporting me to construct expression vectors. I thank Dr. Kazuki Nakamae for giving me advice on statistical analysis.

I am grateful to Prof. Tomoji Mashimo, Dr. Kazuto Yoshimi, Dr. Toshikazu Ushijima, Dr. Naoko Hattori, and Dr. Takuya Fukazawa for providing valuable advices.

I thank Edanz (https://en-author-services.edanz.com/ac) for editing the English text of a draft of this thesis.