博 士 論 文

Photoactivatable CRISPR-Cas9 systems

for optogenetic genome engineering

（ゲノムの光操作を実現する

光活性化型 CRISPR-Cas9 システムの開発）

平成 28 年 12 月　博士（学術）申請

東京大学大学院総合文化研究科

広域科学専攻

二本垣　裕太

Acknowledgements

The present thesis study has been carried out under the supervision of Prof. Moritoshi Sato. I would like to express my sincere gratitude for his continuous support of my graduate study and research, his patience, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to thank Sato laboratory's members for the meaningful discussions. In particular, I am grateful to Dr. Hideyuki Suzuki for insightful comments and passionate guidance.

I would like to acknowledge the financial support by the Japan Society for the Promotion of Science (JSPS).

Finally, I would like to sincerely thank my family and friends for supporting me throughout this study and my life.

Table of Contents

Chapter 1.

General introduction

1-1. Genome editing

Genome editing refers to the technologies for the targeted specific modification of the genetic information of living organisms[1]. This is now a very active research area because of the wide range of possible applications. In the area of biological research, genome editing enables to investigate the function of a gene by gene knockout. In the area of agricultural biotechnology, it offers the development of new generations of genetically modified foods and plants. Furthermore, in human health, this technology has the potential to correct a gene causing genetic disease.

Genome editing is facilitated in many cases by targeted nucleases (**Figure 1-1**). The targeted nuclease creates specific double-strand break at desired locations in the genome, and this double-strand break are repaired by cellular endogenous mechanisms, nonhomologous end-joining pathway and homology-directed repair pathway. In most cases, double-strand break are repaired by nonhomologous end-joining pathway. This repair pathway is error-prone, so that small deletions or insertions are introduced in the repaired targeted gene. Because genes have information for protein synthesis in the form of triplet DNA code, the indel mutation induces DNA frame-shift and disturb the correct translation, resulting in translation of non-functional protein. Therefore, this NHEJ process induced by targeted double-strand break offers gene knockout. In contrast to NHEJ-mediated error-prone repair pathway, HDR pathway offers precise error-free gene modification. In HDR process, double-strand break are repaired with reference to template sequences. By using the homologous exogenous DNA template

containing point mutations or exogenous gene, the targeted gene in the genome can be mutated precisely and can be tagged with exogenous proteins.

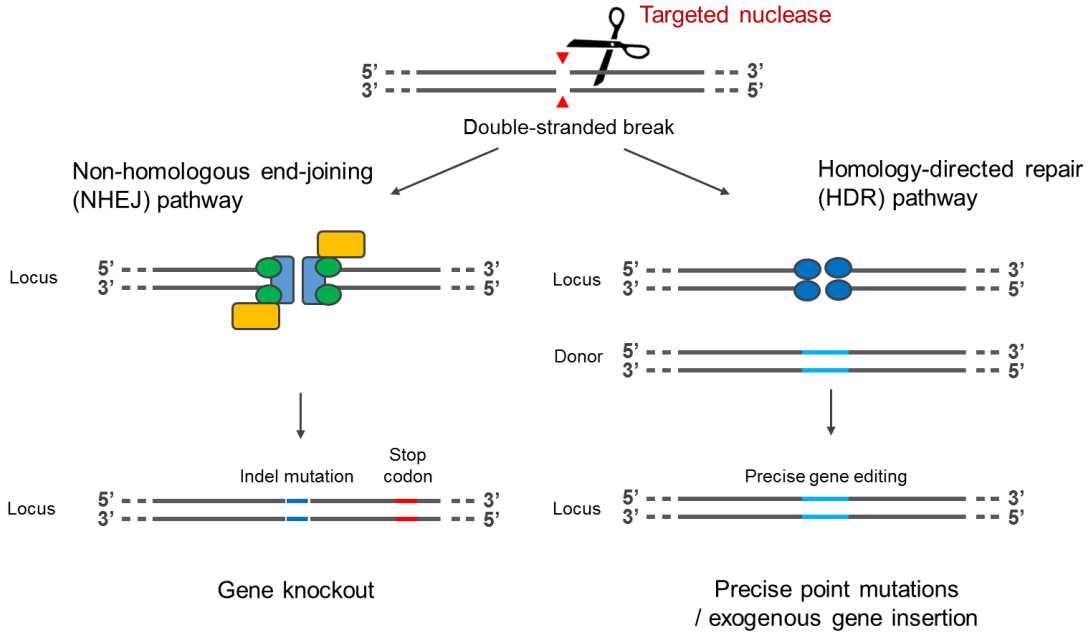

Figure 1-1. Schematics of genome editing.

1-2. CRISPR-Cas9

There are three kinds of targeted nucleases, zinc finger nucleases (ZFNs)[2], transcription activator-like effector nucleases (TALENs)[3], and CRISPR-Cas9[4–6] (**Figure 1-2**). Compared to previous genome engineering techniques, such as ZFNs and TALENs, CRISPR-Cas9 offers more simple approach to edit the genome. The major difference between previous techniques and CRISPR-Cas9 is how targeted nucleases recognize DNA. ZFNs and TALEN recognizes DNA by protein repeats. ZFNs consist of six repeated zinc finger DNA-binding units and a DNA-cleavage domain. One zinc finger units recognize three base pair of DNA, so that six repeated zinc finger units recognize eighteen base pair of DNA. DNA-cleavage domain can work only when this domain

dimerizes. Therefore, using a pair of ZFNs, targeted double stranded break in the genome is achievable, facilitating genome modification. TALENs also consist of repeated DNA-binding units and a DNA-cleavage domain. Unlike ZFNs, one TALE units recognize one base pair of DNA. ZFNs and TALENs used repeated protein for targeting DNA sequence. So, scientists who want to edit the gene have to reassembly DNA constructs encoding repeated protein units that can bind new target sequence. This plasmid construction is time-consuming and requires specialized expertise.

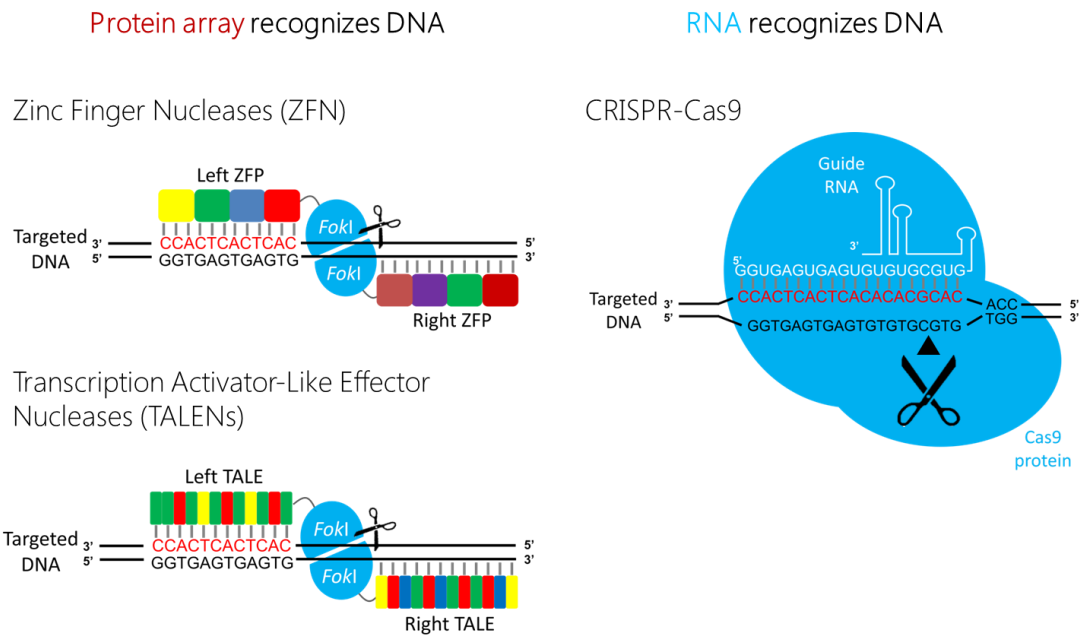

Figure 1-2. Schematics of three targeted nucleases, zinc finger nucleases, transcription activator-like effector nucleases and CRISPR-Cas9.

On the other hands, CRISPR-Cas9 recognizes DNA by RNA. CRISPR-Cas9 has completely different DNA recognition system. Cas9 is targeted to specific DNA using a guide RNA, and this feature enables to edit the genome with ease. Genome engineering with Cas9 requires transfection of two components, Cas9 protein and guide RNA. Cas9

protein is complexed with guide RNA. The Cas9:guide RNA complex can bind to and cleave a DNA sequence that is complementary to the first 20 nucleotides of a guide RNA. So, just redesigning the first 20 nucleotides of guide RNA, Cas9 can be targeted to new DNA sequence. Note that Cas9 can target only DNA sequence adjacent to a protospacer-adjacent motif (PAM) of the form of NGG. After targeted double stranded break, this break is repaired by NHEJ pathway or HDR pathway, resulting in genome modification, such as gene knockout, precise point mutation and insertion of exogenous gene.

1-3. Other applications of CRISPR-Cas9

CRISPR-Cas9 offers not only genome engineering but also transcription control[7,8]. Cas9 binds to DNA targeted by guide RNA and cleaves it. Cas9 has two catalytic active centers, D10 and H840. Cas9 mutant, having mutations of D10A and H840A, shows no nuclease activity. However, this mutant still has DNA binding activity. This mutant is called catalytically dead Cas9 (dCas9). Targeting dCas9 close to transcription start site or to the coding region of gene, dCas9 can block RNA polymerase and transcript elongation, reducing the transcript level. This system is called CRISPR interference (CRISPRi). In addition, fusing transcriptional activation domain with dCas9, endogenous gene activation can be achievable[9,10]. Fusing epigenetic modifying enzymes with dCas9, targeted epigenetic modification, such as histone-methylation and histone-acetylation, is also available[11,12]. Furthermore, fusing GFP with dCas9, researchers now can visualize targeted genome loci in living cells[13].

1-4. Purpose of the present study

CRISPR-Cas9 can control various molecular processes on the genome, and this is an absolutely potent tool. In the present study, I set out to develop spatiotemporally controllable CRISPR-Cas9 systems. Spatiotemporally controllable Cas9 means that Cas9 can be switched on in defined region at defined time, and Cas9 can be switched off at defined time. It will significantly expand the application possibility of CRISPR-Cas9 technology in many fields of basic biological research and medical applications.

For spatiotemporally controlling Cas9-mediated genome manipulation, I focused on methods to photoactivation of proteins, called optogenetics[14,15]. Optogenetics offer optical control of biomolecular activity using light-controllable protein derived from plants, algaes and fungies. Because light has high spatiotemporal resolution, molecular activity can be controlled with high spatiotemporal resolution. Taken together, I sought to combine optogenetics with CRISPR-Cas9 for spatiotemporal control of Cas9.

In Chapter 2, I engineered a photoactivatable Cas9, which enables optical control of CRISPR-Cas9 genome editing in human cells. In Chapter 3, I described CRISPR-Cas9-based photoactivatable transcription system for photoactivatable endogenous gene activation. In Chapter 4, I engineered an improved optogenetic CRISPR-Cas9-based transcription system, named paCas9-A. The present photoactivatable CRISPR-Cas9 systems offer simple and versatile methods for precise and robust genome manipulation in basic biological researches, biotechnology applications, and future medical applications.

1-5. References

1. Gaj, T., Gersbach, C. A & Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31,** 397–405 (2013).
2. Porteus, M. H. & Carroll, D. Gene targeting using zinc finger nucleases. *Nat. Biotechnol.* **23,** 967–973 (2005).
3. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29,** 143–148 (2011).
4. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–826 (2013).
5. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
6. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2,** e00471 (2013).
7. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152,** 1173–1183 (2013).
8. Gilbert, L. a *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442–451 (2013).
9. Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10,** 977–979 (2013).
10. Perez-pinera, P. *et al.* RNA-guided gene activation by CRISPR- Cas9 – based transcription factors. *Nat. Methods* **10,** 973–976 (2013).

11. Kearns, N. a *et al.* Functional annotation of native enhancers with a Cas9 – histone demethylase fusion. *Nat. Methods* **12,** 401–403 (2015).

12. Morita, S. *et al.* Targeted DNA demethylation in vivo using dCas9–peptide repeat and scFv–TET1 catalytic domain fusions. *Nat. Biotechnol.* **34,** 1060–1065 (2016).

13. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155,** 1479–91 (2013).

14. Brieke, C., Rohrbach, F., Gottschalk, A., Mayer, G. & Heckel, A. Light-controlled tools. *Angew. Chem. Int. Ed. Engl.* **51,** 8446–76 (2012).

15. Lee, H., Larson, D. & Lawrence, D. Illuminating the chemistry of life: design, synthesis, and applications of 'caged' and related photoresponsive compounds. *ACS Chem. Biol.* **4,** 409–427 (2009).

Chapter 2.

Photoactivatable CRISPR-Cas9 for optogenetic genome editing

2-1. Introduction

The methods that enable targeted genome modification are crucial for dissecting gene functions and regulatory mechanisms in various biological phenomena. For this purpose, recently emerged Cas9 from microbial type II CRISPR (clustered regularly interspaced palindromic repeats) system, has offered potent tools[1–3]. The *Streptococcus pyogenes* Cas9 nuclease (hereafter referred to as Cas9) can be bound on and cleave a target DNA sequence that is complementary to the first 20 nucleotide of the sgRNA and is adjacent to a protospacer-adjacent motif (PAM) of the form NGG. The Cas9-induced DNA double-strand breaks are repaired by non-homologous end joining (NHEJ) or homology-directed repair (HDR) in mammalian cells, enabling targeted genome editing.

For further development of CRISPR-Cas9 technology, several groups developed the methods to chemically control nuclease activity of Cas9, such as doxycycline-regulated Cas9 expression[4,5], rapamycin-inducible split-Cas9[6] and transient delivery of purified Cas9:sgRNA complex[7–9]. These chemical methods have been utilized for beneficial applications, such as conditional gene knockout and reduction of off-target genome modification. However, these methods using chemical reagents have an inherent limitation in spatiotemporal resolution because these chemicals are freely diffuse and hard to remove rapidly. In addition, some chemicals have adverse effects. For example, rapamycin can induce undesirable biological effects by perturbing endogenous mammalian target of rapamycin (mTOR) pathway[10].

An ideal regulation system of Cas9 nuclease activity should allow spatial, temporal and reversible control and should be non-invasive. Considering these requirements, a method to control Cas9 nuclease activity by light will be ideal because light has high spatiotemporal resolution and is non-invasive[11,12]. It will significantly expand the application possibility of CRISPR-Cas9 technology in many fields of basic biological research and medical applications; however, the methods for optically controlling nuclease activity of Cas9 itself remain elusive.

For providing a simple and versatile method for optogenetic control of Cas9, here I developed photoactivatable Cas9 (paCas9). paCas9 consists of two split Cas9 fragments combined with photoactivatable dimerization domains (**Figure 2-1**). Upon blue light irradiation, photoactivatable dimerization domains are heterodimerized, and consequently the two split Cas9 fragments are brought together to be reconstituted as functional Cas9, breaking targeted DNA strands for genome editing. I showd that paCas9 can induce targeted genome modification via NHEJ and HDR pathways by light. Furthermore, I showed that paCas9 can be spatially activated and precisely switched on and off. My results show that paCas9 can offer simple and versatile tool for targeted genome editing with high spatial and temporal resolution.

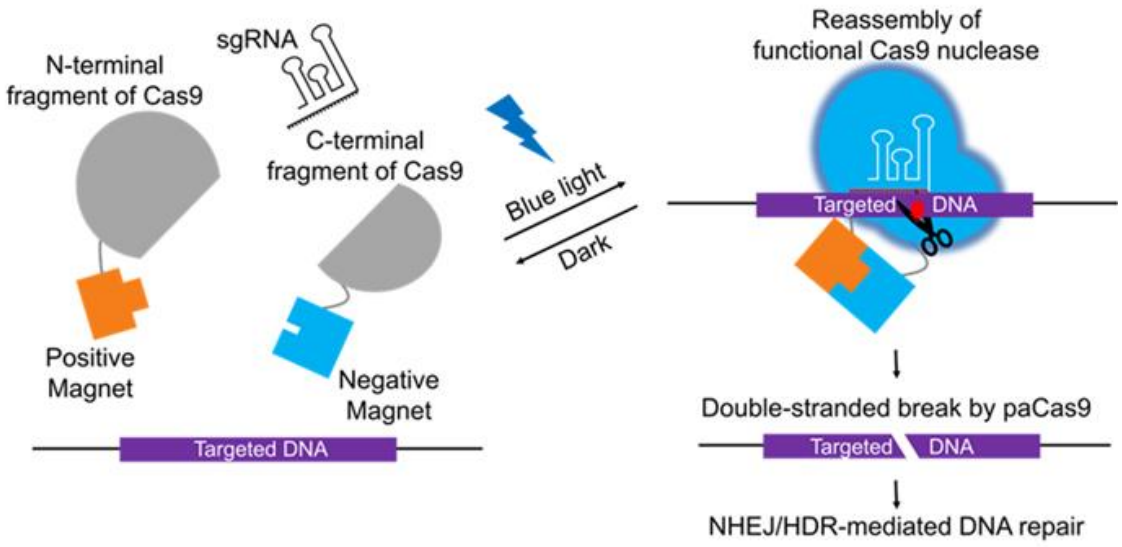

Figure 2-1. Schematic of the photoactivatable Cas9 (paCas9). Cas9 is split into two fragments without nuclease activity, and the Cas9 fragments are combined with photoactivatable dimerization domains, (here positive Magnet and negative Magnet). Blue light irradiation induces heterodimerization between positive Magnet and negative Magnet, and therefore two split Cas9 fragments are reassociated, reconstituting RNA-guided nuclease activity.

2-2. Materials and Methods

2-2-1. Inducible Cas9 constructions

cDNAs encoding the N- and C- fragments of codon-optimized *Streptococcus pyogenes* Cas9 combined with NLS from SV40 were amplified from Addgene plasmid 42230. cDNAs encoding FKBP and FRB were amplified from a human cDNA library. cDNA encoding CRY2PHR was amplified from Addgene plasmid 26871. The plasmid containing CIB1 was obtained from RIKEN BioResource Center (Resource Number: pda10875). cDNA encoding pMag, nMagHigh1 and nMag were prepared as previously described[18]. These inducible dimerization domains were amplified by standard PCR using primers that add glycine-serine linker sequences at the 5' and 3' ends. The inducible Cas9 constructs based on the N-terminal and C-terminal Cas9 fragments combined with dimerization domains were cloned into *Hin*dIII/*Eco*RI and *Hin*dIII/*Xho*I sites of pcDNA3.1 V5/His-A (Invitrogen), respectively. To construct paCas9 nickase and padCas9, D10A mutation in N-fragment of Cas9 and H840A in C-fragment of Cas9 were introduced by Multi Site-Directed Mutagenesis Kit (MBL) according to manufacturer's directions.

2-2-2. sgRNA constructions

The sgRNAs targeting StopFluc reporters, *CCR5*, *EMX1*, *VEGFA*, *AAVS1* and destabilized luciferase reporter were generated by annealed oligo cloning using *Bbs*I site of Addgene plasmid 47108.

2-2-3. Reporter constructions

StopFluc reporters for plasmid HDR assay were constructed by inserting firefly luciferase sequence amplified from pGL4.31 vector (Promega) into *Hin*dIII and *Xho*I sites in pcDNA 3.1/V5-HisA, and introducing stop codons and/or mutant PAM by Multi Site-Directed Mutagenesis Kit. Luciferase donor vector was constructed by inserting inverted firefly luciferase sequence into *Xho*I and *Hin*dIII in bacteria-expression pColdI vector (Clontech). Destabilized luciferase reporter was constructed by inserting firefly luciferase with PEST sequence amplified from pGL4.31 vector into *Kpn*I and *Xba*I sites in pcDNA 3.1/V5-HisA, and introducing 5 copies of mRNA-destabilizing nonamer sequence[27] (5'-TTATTTATT-3') into *Xba*I and *Apa*I sites by annealed oligo cloning. Surrogate EGFP reporter was constructed by inserting mCherry and out-of frame EGFP into *Hin*dIII and *Xho*I sites in pcDNA 3.1/V5-HisA, and introducing *EMX1* target site between mCherry and EGFP using *Eco*RI and *Bam*HI sites by annealed oligo cloning.

2-2-4. Cell culture

HEK293T and HeLa cells (ATCC) were cultured at 37 °C under 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM, Sigma Aldrich) supplemented with 10% FBS (HyClone), 100 unit/ml penicillin and 100 μg/ml of streptomycin (GIBCO).

2-2-5. Luciferase plasmid HDR assay

HEK293T cells were plated at approximately 2.0×10^4 cells/well in 96-well black-walled plate (Thermo Fisher Scientific), and cultured for 24 h at 37 °C in 5% CO_2.

The cells were then transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocols. Plasmid encoding N-fragments of Cas9 combined with dimerization domain, C-fragments of Cas9 with dimerization domain, sgRNAs, StopFluc reporter and luciferase donor were transfected at a 2.5:2.5:5:1:4 ratio. The total amount of DNA was 0.2 μg/well. Twenty hours after the transfection, samples were incubated at 37 °C in 5% CO_2 under continuous blue light irradiation or in the dark. Blue light irradiation was performed using a 470 nm ± 20 nm LED light source (CCS Inc.). Intensity of blue light was 1.2 W/m^2. For chemically-inducible reassembly of split-Cas9, the culture medium was replaced with 100 μl of DMEM containing 10 nM of rapamycin instead of light irradiation. After 48 h incubation, the culture medium was replaced with 100 μl of phenol red-free DMEM (Sigma Aldrich) containing 500 μM of D-luciferin (Wako Pure Chemical Industries) as a substrate. After 30 min incubation, bioluminescence measurements were performed using Centro XS^3 LB 960 plate-reading luminometer (Berthold Technologies). For comparison between DNA recognition ability of paCas9 and full-length Cas9, plasmids encoding full-length Cas9, sgRNAs, StopFluc reporter and luciferase donor were transfected at a 5:5:1:4 ratio. The total amount of DNA was 0.2 μg/well. After 48 h incubation, the culture medium was replaced with D-luciferin-containing phenol red-free DMEM and bioluminescence measurements were performed as described above.

2-2-6. Optogenetic genome editing experiments

For NHEJ-mediated indel mutation experiments, HEK293T cells were plated at approximately 1.0 × 10^5 cells/well in 24-well plate (Thermo Fisher Scientific), and

cultured for 24 h at 37 °C in 5% CO_2. The cells were then transfected with Lipofectamine 2000 according to the manufacturer's protocols. Plasmids encoding N713-pMag, nMagHigh1-C714 and sgRNAs were transfected at a 1:1:1 ratio. As a positive control, plasmids encoding full-length Cas9 and sgRNAs were transfected at a 2:1 ratio. The total amount of DNA was 0.9 μg/well. Twenty hours after the transfection, samples were incubated at 37 °C in 5% CO_2 under continuous blue light irradiation or in the dark as described above. After 24 h incubation, genomic DNA was isolated using Blood Cultured Cell Genomic DNA Extraction Mini Kit (Favorgen) according to the manufacturer's instructions.

For HDR-mediated genome editing experiments, 6.0×10^5 HEK293T cells were nucleofected with 125 ng of N713-pMag, 125 ng of nMagHigh1-C714, 250 ng of sgRNA targeting *EMX1* and 10 μM of single-stranded oligonucleotide donor using the SF Cell line 4D-Nucleofector X Kit S (Lonza) and the CA-189 program. Transfected cells were plated at 2.0×10^5 cells/well in 24-well plate. Twenty hours after the nucleofection, samples were incubated at 37 °C in 5% CO_2 under continuous blue light irradiation or in the dark. After 48 h incubation, genomic DNA was isolated as described above.

In **Figure 2-13** experiment, cells were plated and cultured as described above NHEJ-mediated indel mutation experiments. The cells were then transfected with Lipofectamine 3000 (Invitrogen) according to the manufacturer's protocols. Plasmids encoding N713-pMag, nMagHigh1-C714 and sgRNAs were transfected at a 1:1:1 ratio. The total amount of DNA was 0.5 μg/well. Immediately after the transfection, samples were incubated at 37 °C in 5% CO_2 under continuous blue light irradiation. After 6 h, I

split and incubated the cells in dark and light state, and performed second transfection of sgRNA targeting *EMX1* with Lipofectamine 3000. The DNA amount was 0.5 μg/well. The samples in dark and light state just before second transfection were incubated again in dark and light state, respectively. After incubation for 30 h, the genomic DNA was isolated as described above.

2-2-7. Mismatch-sensitive T7E1 assay for quantifying indel mutation of endogenous genes

The genomic region containing paCas9 target site was PCR-amplified using Pyrobest DNA polymerase (TaKaRa) using nested PCR for *CCR5* and *AAVS1* (First PCR: 98 °C, 3 min; (98 °C, 10 sec; 55 °C, 30 sec; 72 °C, 1 min) × 20 cycles; 72 °C, 3 min. Second PCR: 98 °C, 3 min; (98 °C, 10 sec; 55 °C, 30 sec; 72 °C, 1 min.) × 35 cycles; 72 °C, 3 min), two-step PCR with 5% DMSO for *EMX1* (98 °C, 3 min; (98 °C, 10 sec; 72 °C, 30 sec) × 35 cycles; 72 °C, 5 min) or touchdown PCR for *VEGFA* (98 °C, 3 min; (98 °C, 10 sec; 72-62 °C, -1 °C/cycle, 30 sec; 72 °C, 30 sec) × 10 cycles; (98 °C, 10 sec; 62 °C, 30 sec; 72 °C, 30 sec) × 25 cycles; 72 °C, 3 min). The PCR amplicons were purified using FastGene Gel/PCR Extraction Kits (Nippon Genetics) following the manufacturer's protocol. Purified PCR products were mixed with 2 μl of 10× M buffer for restriction enzyme (TaKaRa) and ultrapure water to a final volume of 20 μl, and re-annealed to form heteroduplex DNA (95 °C, 10 min; 90-15°C, -1 °C/ 1 min). After re-annealing, heteroduplexed DNA were treated with 5 units of T7 endonuclease I (New England Biolabs) for 30 min at 37 °C and then analyzed by agarose gel electrophoresis.

Gels were stained with GRR-500 (BIO CRAFT) and imaged with E-shot II gel imaging system (ATTO). Quantification was based on relative band intensities. Following equation is used to calculate the percentage of indel mutation by paCas9: $100 \times (1 - (1 - (b + c)/(a + b + c))^{1/2})$, where a is the intensity of the undigested PCR product, and b and c are the intensities of each T7E1-digested PCR product.

2-2-8. Sequence analysis

Purified PCR products used for the T7E1 assay were inserted into *EcoR*V sites in pcDNA3.1/V5-HisA vector. Plasmid DNAs were isolated by standard alkaline lysis miniprep, and sequenced using a T7 forward primer by the Sanger method.

2-2-9. RFLP assay for detecting HDR-mediated modification in endogenous human gene

The genomic PCR and purification were performed as described above. Purified PCR products were mixed with 30 units of *Hin*dIII (TaKaRa), 2 µl of 10× M buffer for restriction enzyme and ultrapure water to a final volume of 20 µl, and incubated for 30 min at 37°C. The digested products were analyzed by agarose gel electrophoresis. Gel staining and imaging were performed as described above. Quantification was based on relative band intensities. Following equation is used to calculate the percentage of HDR by paCas9: $100 \times (b + c)/(a + b + c)$, where a is the intensity of the undigested PCR product, and b and c are the intensities of each *Hin*dIII-digested product.

2-2-10. Spatial surrogate reporter activation

HEK293T cells were plated at 8.0×10^5 cells/dish on 35 mm dish (Iwaki Glass) coated with fibronectin (BD Biosciences), and cultured for 24 h at 37 °C in 5% CO_2. The cells were then transfected with Lipofectamine 2000 according to the manufacturer's protocols. Plasmids encoding N713-pMag, nMag-Cas9, sgRNAs targeting *EMX1* and NHEJ-mediated surrogate EGFP reporter containing *EMX1* targeting site were transfected at a 1:1:2:6 ratio. The total amount of DNA was 4.0 μg/dish. Twenty hours after the transfection, samples were illuminated by slit-patterned blue light using a photomask for 24 h at 37 °C in 5% CO_2. The width of slit is 2 mm. Cells were fixed with 4% paraformaldehyde in PBS for 15 min. Images were acquired using Axio Zoom.V16 stereo zoom microscope (Zeiss), and analyzed using Metamorph (Molecular Devices).

2-2-11. Photoactivatable CRISPR interference

HEK293T cells were plated at 2.0×10^4 cells/well in 96-well black-walled plate, and cultured for 24 h at 37 °C in 5% CO_2. The cells were then transfected with Lipofectamine 3000 according to the manufacturer's protocols. Plasmids encoding N713 (D10A)-pMag, nMag-Cas9 (H840A), mRNA-destabilized luciferase-PEST reporter and indicated sgRNAs targeting luciferase reporter were transfected at a 2.5:2.5:1:4 ratio. For the sample transfected with triple sgRNAs, the ratio of three sgRNAs was 1:1:1. The total amount of DNA was 0.1 μg/well. In **Figure 2-14b** experiment, twenty hours after the transfection, samples were incubated at 37 °C in 5%

CO_2 under continuous blue light irradiation or in the dark as described above. After 30 h, the culture medium was replaced with 100 μl of phenol red-free DMEM containing 500 μM of D-luciferin. After 1 h incubation, bioluminescence measurements were performed. In **Figure 2-14d** experiment, samples were incubated at 37 °C in 5% CO_2 under continuous blue light immediately after transfection. After 30 h, the culture medium was replaced with phenol red-free DMEM containing D-luciferin as described above. After 1 h incubation, light-illuminated samples were incubated again upon continuous blue light or in the dark, and bioluminescence measurements were performed at indicated time points.

2-2-12. Reproducibility and statistics

No sample size estimates were performed, and my sample sizes are consistent with that normally used in the genome editing and gene regulation experiments. No sample exclusion was carried out. No randomization was used. No blinding was used. In **Figure 2-11b and Figure 2-14b**, variances estimated by *F*-tests were equal, and two-sided Welch's *t*-tests were performed. I also confirmed the same statistical results by nonparametric two-sided Mann-Whitney tests because the normality of data was not tested. *P* values by Mann-Whitney tests are following; In **Figure 2-11b**, paCas9-1 and paCas9-2 in light showed no significant difference: $P = 0.773$. In **Figure 2-14b**, light and dark samples transfected with padCas9 targeting luciferase showed significant differences: $P = 0.00395$. Light and dark samples transfected with padCas9 and negative control sgRNA showed no significant difference: $P = 0.749$.

2-3. Results

2-3-1. Development of paCas9

In initial experiments, to determine the split site of Cas9 for inducible control with high efficiency, I generated the various Cas9 fragments combined with rapamycin-inducible dimerization system, FKBP-FRB[13] (**Figure 2-2**). I selected 18 candidate split sites based on the analysis of the crystal structure of Cas9 in complex with single-guide RNA[14,15] (sgRNA). All candidate positions were loop regions exposed to solvent. I assessed the rapamycin-induced nuclease activity of each split-Cas9 pair using a luciferase-reporter plasmid HDR assay. In this assay, the CMV promoter-driven luciferase reporter with an in-frame stop codon (StopFluc-1) is cleaved by split-Cas9, and then recovers full-length luciferase expression through homologous recombination with promoter-less luciferase donor vector. Eight combinations of N- and C-terminal Cas9 fragments showed significant rapamycin-induced reporter upregulation in HEK293T cells (**Figure 2-3**). In the following experiments, I used the N-terminal fragment of Cas9 (residues 2-713) and C-terminal fragment of Cas9 (residues 714-1368), termed N713 and C714, respectively.

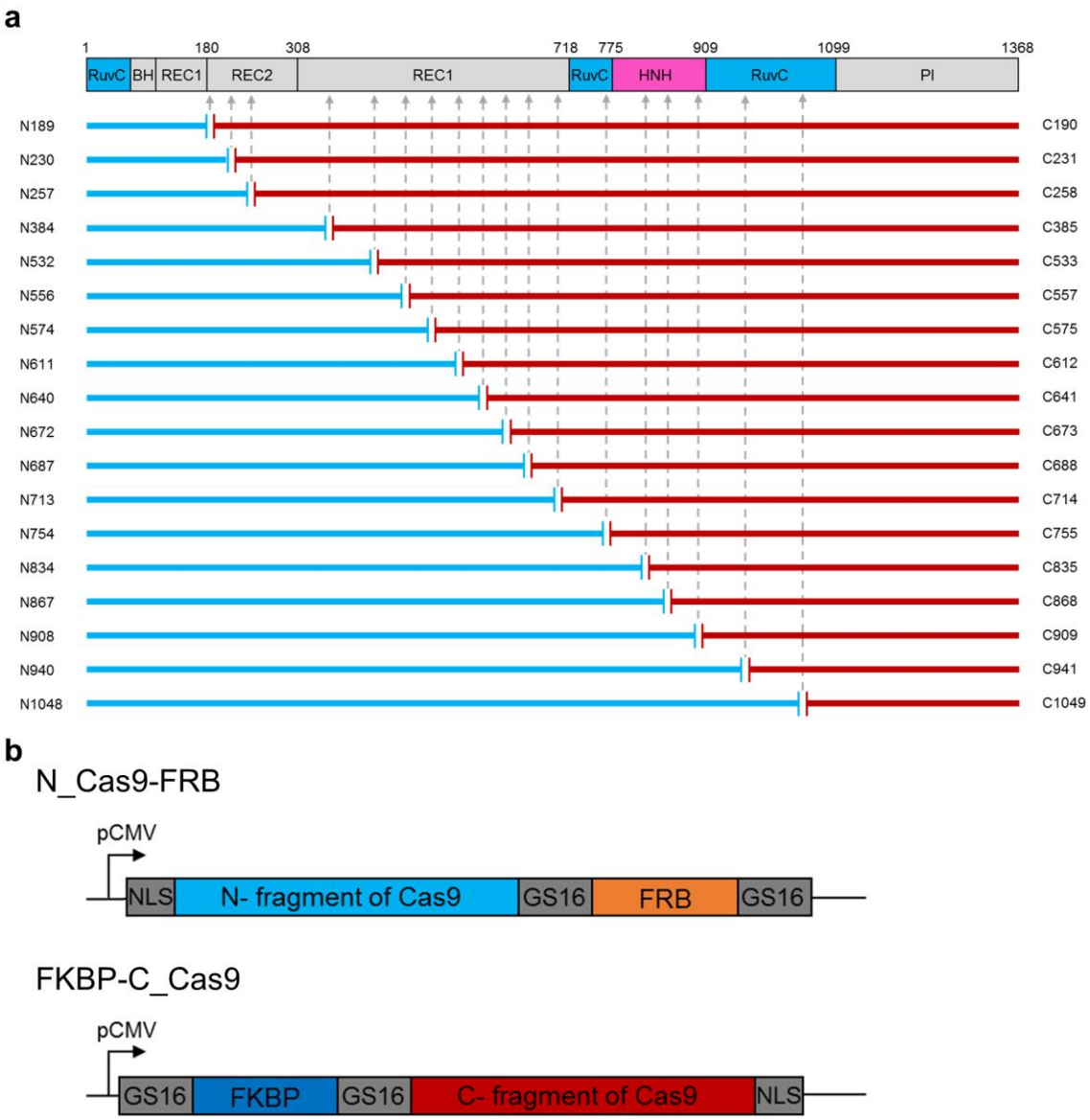

Figure 2-2. Construction of rapamycin-inducible split-Cas9. (a) Candidate 18 split sites of *Streptococcus pyogenes* Cas9 protein. (b) Constructs of rapamycin-inducible Cas9. N- and C- fragments of Cas9 were combined with FRB and FKBP, respectively.

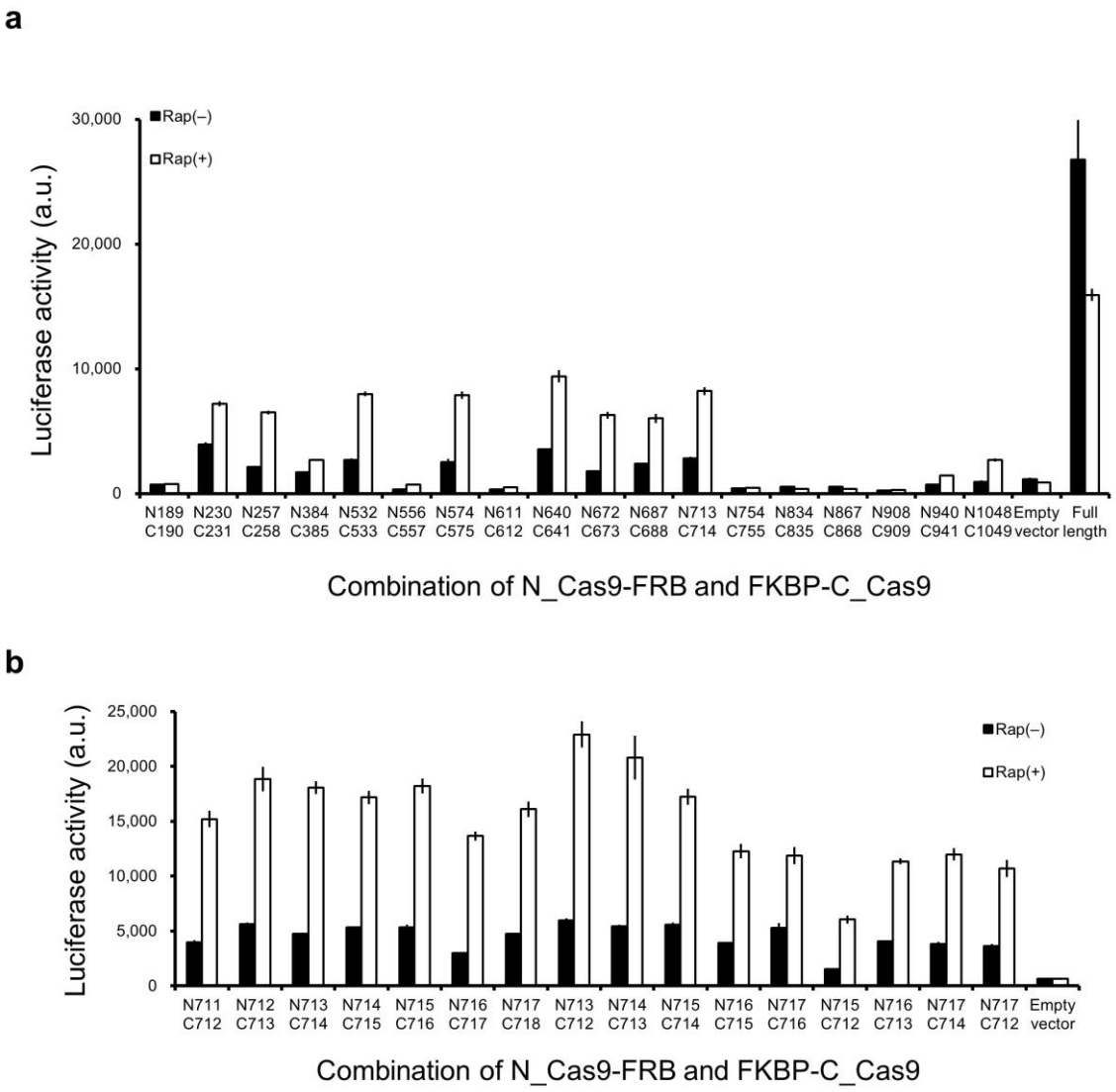

Figure 2-3. Screening of split-Cas9 fragments (**a**) Ligand-induced Cas9 activity measured by luciferase reporter plasmid HDR assay. HEK293T cells were transfected with N- fragment of Cas9 combined with FRB, C-fragment of Cas9 combined with FKBP, stop codon-inserted luciferase reporter and promoter-less luciferase donor vector. (**b**) Second screening of Cas9 split sites in the vicinity of residues 713 and 714. Data are shown as the mean ± s.e.m. (n=6 from two individual experiments with biological triplicates).

Next, I combined photoactivatable dimerization domains with N713 and C714 (**Figure 2-4**). Firstly, I tested CRY2-CIB1 photoactivatable dimerization system. The CRY2-CIB1 system is based on blue light-dependent protein interaction between *Arabidopsis thaliana* cryptochrome 2 (CRY2) and its binding partner CIB1[16]. This system has been widely used for optogenetic control of protein-protein interaction in mammalian cells. I generated N713 and C714 combined with the photolyase homology region of CRY2 (CRY2PHR) and CIB1 and tested its induction potency using luciferase plasmid HDR assay. However, N713 and C714 combined with CRY2PHR and CIB1 system did not show light-induced Cas9 activity. There are several possible explanations for this failure. One is a steric hindrance caused by the large molecular size of CRY2PHR (498 amino acids) and CIB1 (335 amino acids), impeding reassembly of split-Cas9. Another possibility is that the oligomerization property of CRY2PHR might disturb one-to-one interaction of N-terminal and C-terminal fragments of Cas9[17]. For these assumptions, I focused on recently developed photoactivatable dimerization system, termed Magnets[18]. Magnet system consists of paired photoswitchable proteins, named positive Magnet (pMag) and negative Magnet (nMag). Upon blue light irradiation, pMag and nMag are heterodimerized. In contrast to CRY2-CIB1, pMag and nMag are much smaller, only 150 amino acids, comparable to FKBP (107 amino acids) and FRB (93 amino acids). The dynamic range and dissociation kinetics of Magnet system can be tuned by introducing mutation into pMag and/or nMag. In this study, I employed the combination of pMag and nMagHigh1, nMag harboring M135I and M165I mutation[18]. To test whether Magnets could provide effective light-triggered reassembly of split-Cas9, I generated N713 and C714 combined with Magnets. I found that combinations of N713 and C714 combined with pMag and nMagHigh1 showed

substantial light-induced Cas9 activity. In particular, the combination of N713-pMag and nMagHigh1-C714 yielded higher fold-induction (16.4-fold) and lowest background activity. Therefore, I used N713-pMag and nMagHigh1-C714 in the following studies, and called this paCas9-1.

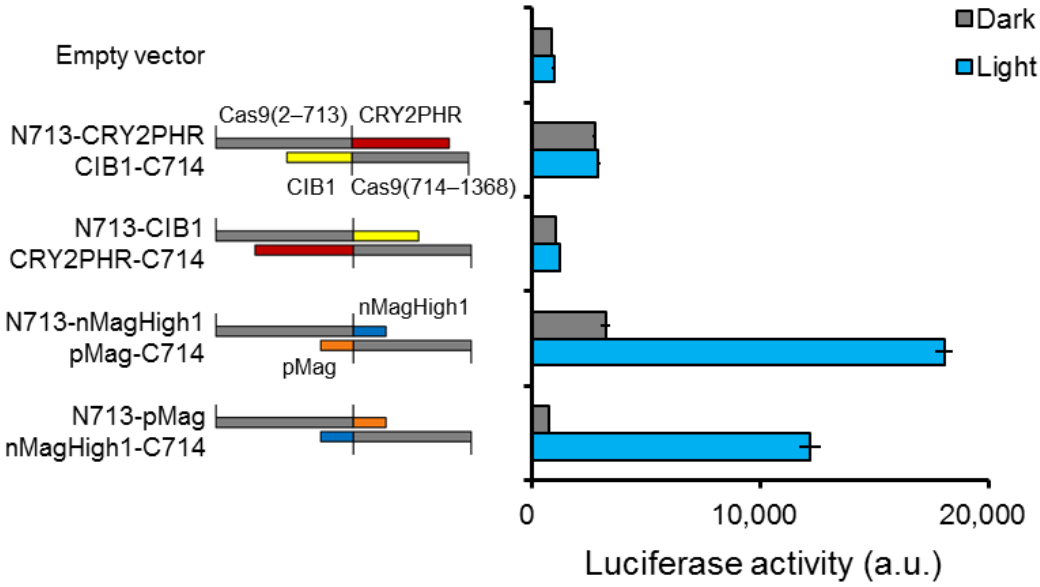

Figure 2-4. Light-induced reporter activity in HEK293T cells by N713 and C714 fragments of Cas9 combined with indicated photoactivatable dimerization domains. Data are shown as the mean ± s.e.m. (n=6 from two individual experiments with biological triplicates).

2-3-2. PAM requirement and DNA targeting specificity of paCas9

To investigate whether paCas9-1 recognizes PAM in the same manner with full-length Cas9, I generated a set of stop codon-inserted luciferase reporters harboring point mutation in NGG PAM (**Figure 2-5a, b**). I tested two different luciferase reporters containing an internal stop codon in different sites (StopFluc-1 and StopFluc-2), and confirmed that the Cas9-induced activities of the luciferase reporters that have non-canonical PAM were lower than that of luciferase reporters having the canonical PAM of the form NGG. I found that there is no significant difference in the normalized luciferase activities between paCas9-1 and full-length Cas9. I also evaluated the DNA targeting specificity of paCas9-1 using luciferase plasmid HDR assay (**Figure 2-5c**). To do this, I generated a series of sgRNAs for StopFluc-1 harboring single-nucleotide Watson-Crick transversion mutations. I found that there is no significant difference in the DNA targeting specificity between paCas9-1 and full-length Cas9. To further investigate DNA targeting specificity of paCas9-1, I performed this specificity assay in additional two different reporters containing an internal stop codon in different sites (StopFluc-2 and StopFluc-3), and confirmed that the DNA targeting specificity of paCas9-1 is comparable to that of full-length Cas9 (**Figure 2-5d, e**). Consistent with previous studies, the sensitivity patterns to single sgRNA-DNA mismatch differed among target sequences[19–21]. From these experiments, I concluded that the PAM requirement and target specificity of paCas9-1 were indistinguishable from those of full-length Cas9.

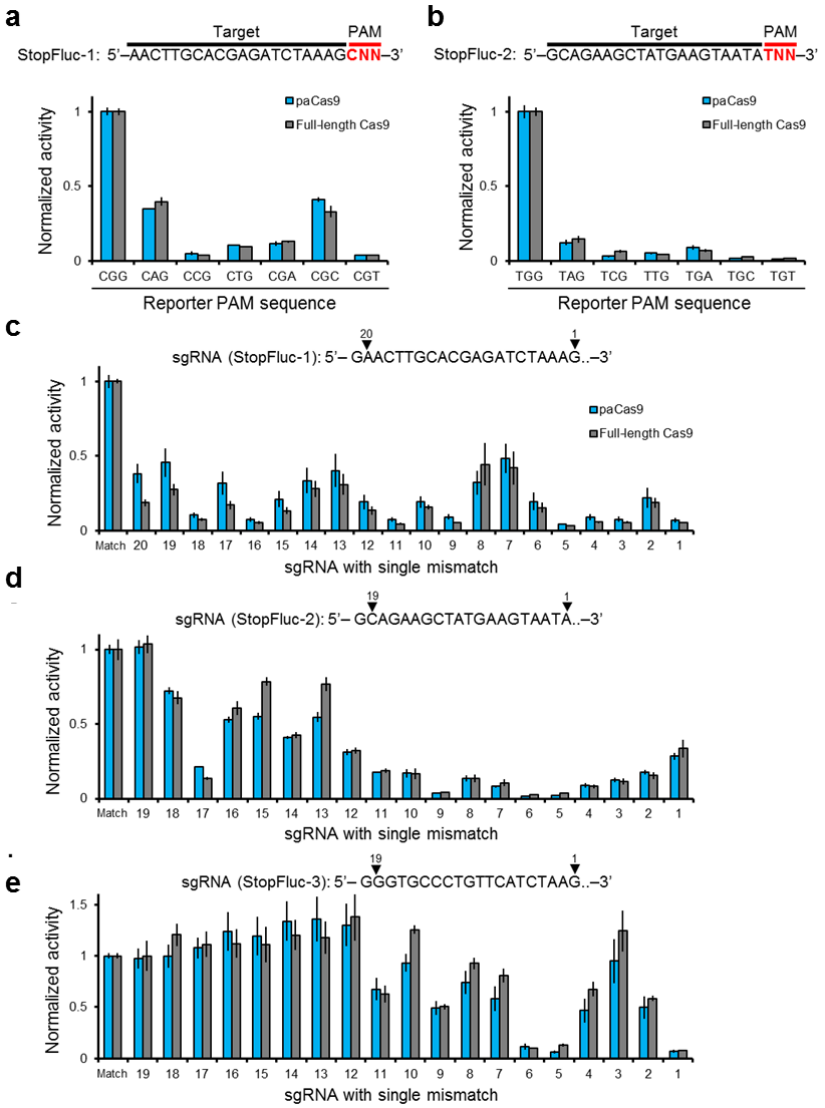

Figure 2-5. Activities of paCas9-1 and full-length Cas9 targeting StopFluc-1 (**a**) and StopFluc-2 (**b**) harboring indicated mutations within the PAM. Values are normalized to positive control with luciferase reporter having canonical PAM (NGG). (**c–e**) Activities of Cas9 and paCas9 targeting StopFluc-1 (**c**), StopFluc-2 (**d**) and StopFluc-3 (**e**) with a set of sgRNAs harboring single-nucleotide Watson-Crick transversion mutations. sgRNAs mutated in 5'-end G to C were not tested because 5'-end G is necessary for efficient expression from the U6 promoter. The positions of point mutations in each sgRNA are indicated at the bottom of each panel. Values are normalized to positive control with perfectly matched sgRNAs. Data are shown as the mean ± s.e.m. (n=6 from two individual experiments with biological triplicates).

2-3-3. Optogenetic genome editing

To show paCas9-1 could cleave a targeted endogenous genomic locus in mammalian cells and induce indel mutation via non-homologous end joining (NHEJ) in a light-dependent fashion, I transfected HEK293T cells with paCas9-1 and sgRNA targeting human *CCR5* locus (**Figure 2-6a**). I quantitatively evaluated the ability of paCas9-1 to induce indel mutations by light using the mismatch-sensitive T7 endonuclease I (T7E1), which cleaves heteroduplexed DNA strand formed by hybridization between mutant and wild-type DNA. In the dark, cells transfected with paCas9-1 targeting *CCR5* showed only 1.1% indel rates. However, upon blue light irradiation, cells transfected with paCas9-1 targeting *CCR5* exhibited significantly higher indel rates (20.5%) in human *CCR5* locus. The frequency of indel mutation induced by paCas9-1 is approximately sixty percent compared with full-length Cas9 (34.4%) (**Figure 2-6a**). By Sanger sequencing, I also verified that paCas9-1-mediated indel mutations occurred in targeted region of human *CCR5* locus (**Figure 2-6b**). To explore the generalizability of optogenetic RNA-guided genome editing with paCas9-1, I constructed sgRNAs for additional four sites in three human genes (*EMX1*, *VEGFA* and *AAVS1*). In all the cases using each sgRNA, light-induced indel mutations were observed (**Figure 2-6c**).

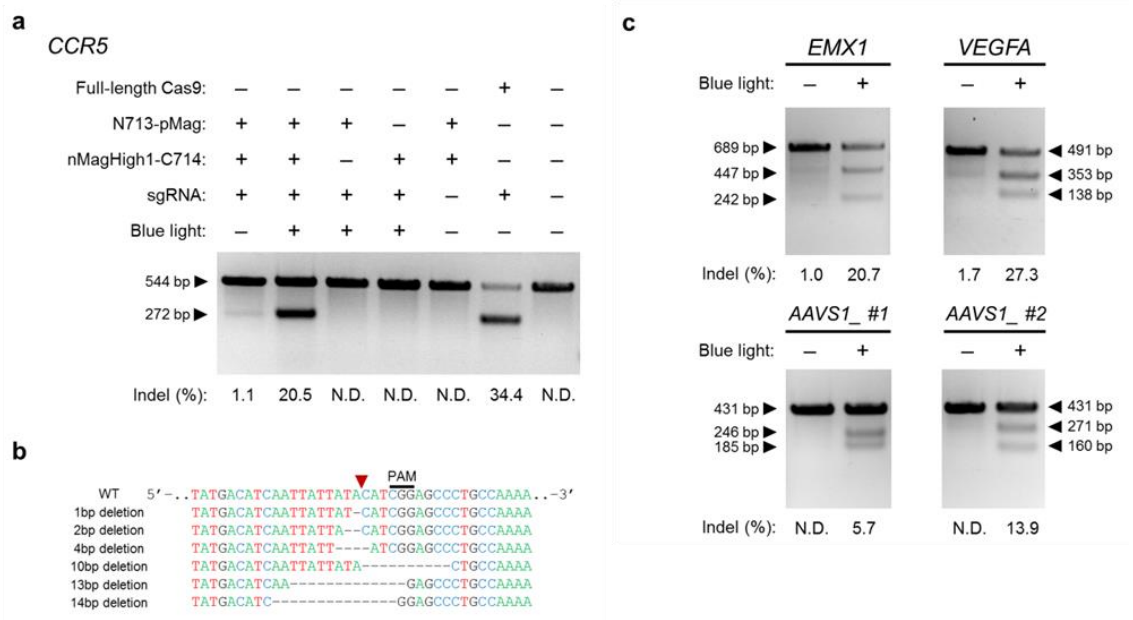

Figure 2-5. Optogenetic genome editing of mammalian endogenous genes by the photoactivatable Cas9. (**a**) Light-mediated indel mutation of human *CCR5* locus by paCas9-1. Indel mutation frequencies were evaluated by mismatch-sensitive T7E1 assay. (**b**) Example sequences of the human *CCR5* locus targeted by paCas9. By Sanger sequencing of 27 individual cDNA clones, six clones showed deletion mutations. (**c**) paCas9 could target a variety of endogenous genes. Cells were transfected with paCas9-1 and sgRNA targeting the *EMX1* site, the *VEGFA* site and the two sites in *AAVS1*.

I also performed the time-course analysis of indel mutations in *EMX1* locus induced by paCas9-1 (**Figure 2-6**). I observed that the frequency of indel mutation increased as blue light irradiation time increased. To test whether paCas9-1 could induce indel mutation in other cell lines, I transfected paCas9-1 and sgRNA targeting human *EMX1* into HeLa cells (**Figure 2-7**). As expected, light-induced indel mutation in *EMX1* locus was observed in HeLa cells. I also tested whether paCas9-1 could induce the indel mutations in multiple target sites (**Figure 2-8**). Using two sgRNAs targeting *EMX1* and *VEGFA* simultaneously, paCas9-1 induced indel mutations in both human *EMX1* and

VEGFA loci by light. These results showd that paCas9-1 can be widely used for optogenetic control of NHEJ-mediated indel mutation in mammalian genome sequence with multiplexed fashion.

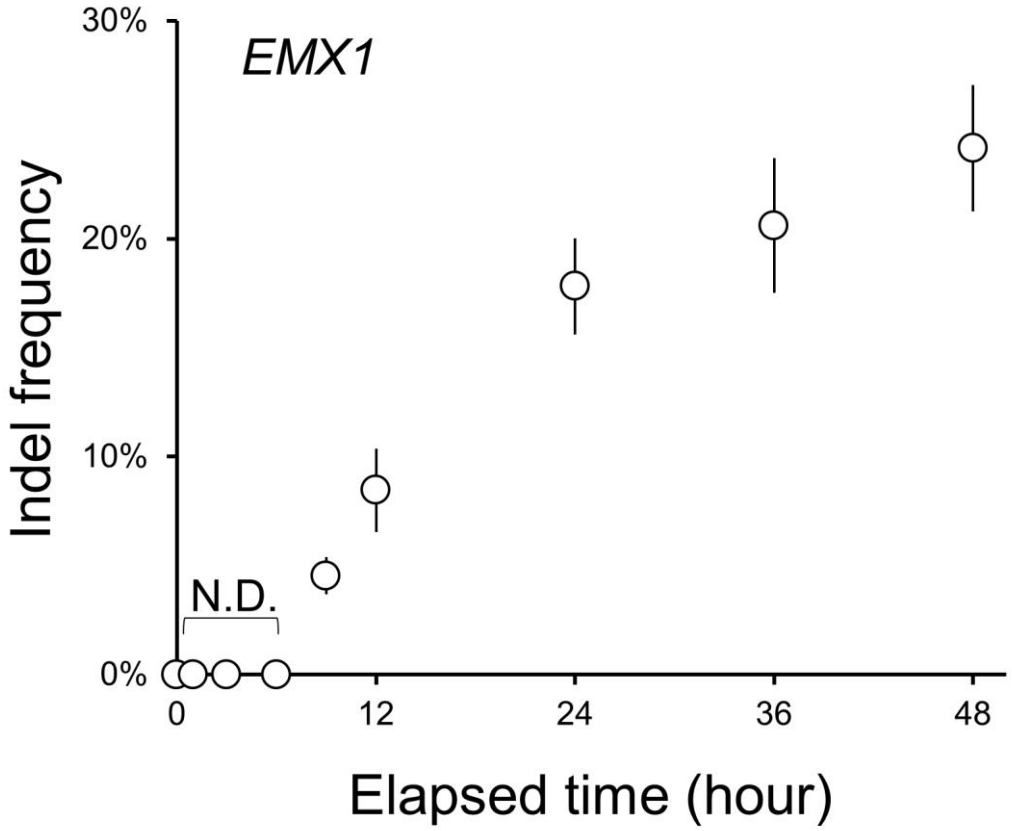

Figure 2-6. Time-course analysis of paCas9-mediated indel mutation in EMX1 locus. HEK293T cells were transfected with paCas9-1 targeting *EMX1*. After 20 h incubation, the cells were illuminated by 1.2 W/m^2 blue light and genomic DNA was extracted at indicated time points. Indel mutation frequencies were evaluated by mismatch-sensitive T7E1 assay. Data are shown as the mean ± s.e.m. (n=4 from two individual experiments with biological duplicates). Indel frequencies at 0, 1, 3 and 6 h were below the detection limit (1%).

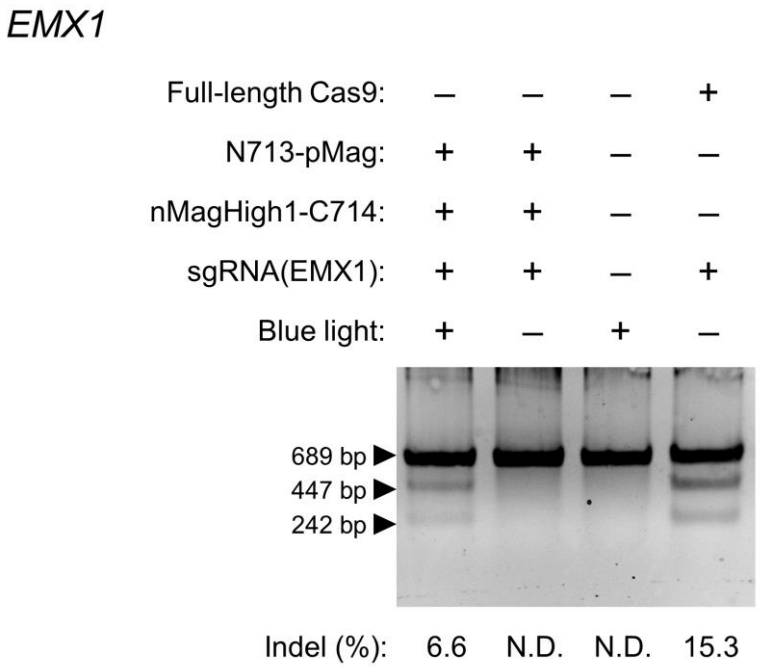

Figure 2-7. paCas9-mediated indel mutation of human *EMX1* locus in HeLa cells. Indel mutation frequencies were evaluated by mismatch-sensitive T7E1 assay.

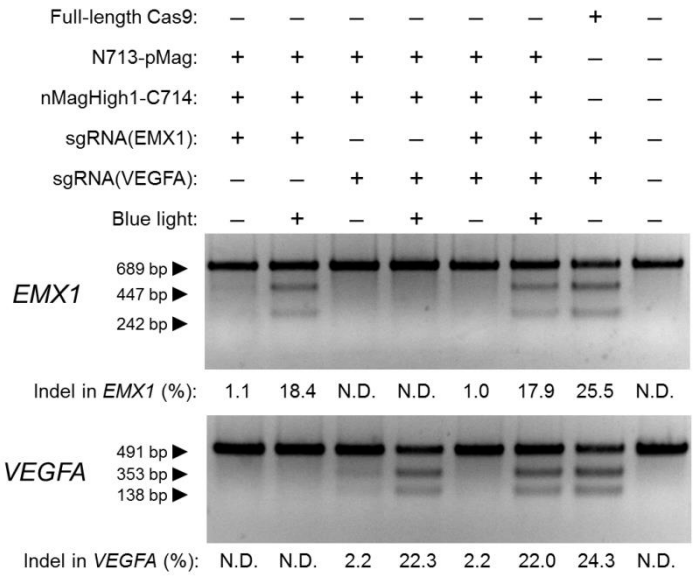

Figure 2-8. Light-induced multiplex genome editing by paCas9. HEK293T cells were transfected with paCas9-1 and the indicated sgRNAs.

Next, I investigated whether paCas9-1 can be used for genome editing via HDR (**Figure 2-9a**, **b**). In this study, I used single-stranded oligodeoxynucleotides (ssODN) as a donor template. I transfected HEK293T cells with paCas9-1 targeting *EMX1* and ssODN containing *Hin*dIII site, and analyzed the frequency of HDR in EMX1 locus by restriction fragment length polymorphism (RFLP) assay. I found that paCas9-1 can induce *Hin*dIII site integration in human *EMX1* locus at a frequency of 7.2%. This result showed that paCas9-1 can induce not only random indel mutation but also designed genome sequence modification via HDR by light.

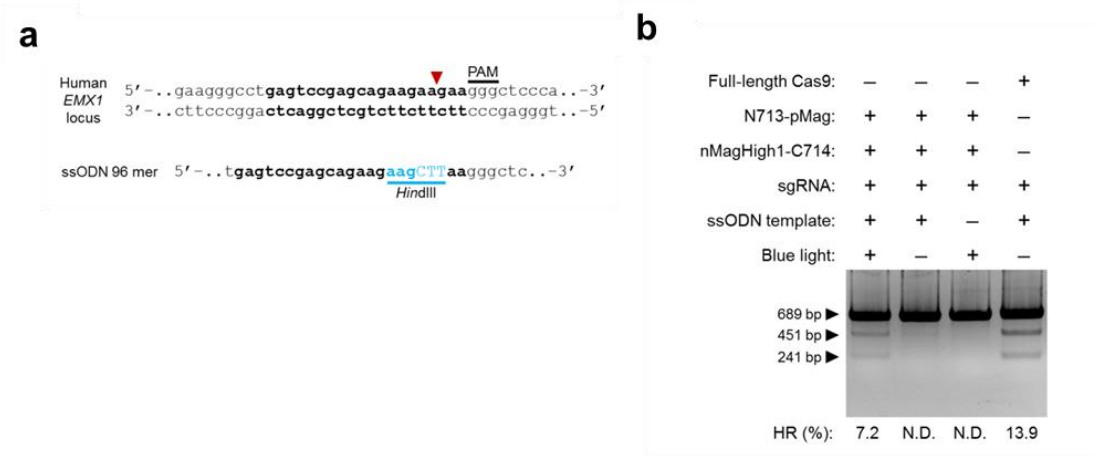

Figure 2-9. (**a**) Schematic of paCas9-mediated precise genome editing experiments. The red arrowhead indicates the putative paCas9 cleavage site. A 96 mer single-stranded oligodeoxynucleotide (ssODN) donor template is designed to insert *Hin*dIII site (blue) into *EMX1* locus. Target sequence is shown in bold lowercase. 3 base pair insertion sequence is shown in uppercase. (**b**) paCas9-mediated precise genome editing by HDR via a single-stranded oligodeoxynucleotide template. Successful HDR frequency was determined by restriction-fragment length polymorphism (RFLP) analysis, and calculated as the ratio of *Hin*dIII digested product to substrate.

For reducing off-target activity of Cas9, paired nicking strategy based on Cas9 D10A variant, which nicks targeted DNA instead of cutting double strands, has been showd[22]. To explore whether the paCas9-1 could also be converted into photoactivatable nickase, I generated paCas9-1 containing D10A mutation (paCas9 nickase) (**Figure 2-10**). paCas9 nickase with single sgRNA targeting *EMX1* did not induce indel mutation in human EMX1 locus. However, using a pair of sgRNAs targeting opposite strands of *EMX1* site, paCas9 nickase showed light-induced indel mutation. This result indicates that paCas9 nickase can also employ Cas9-mediated double nicking strategy for reducing off-target genome modifications.

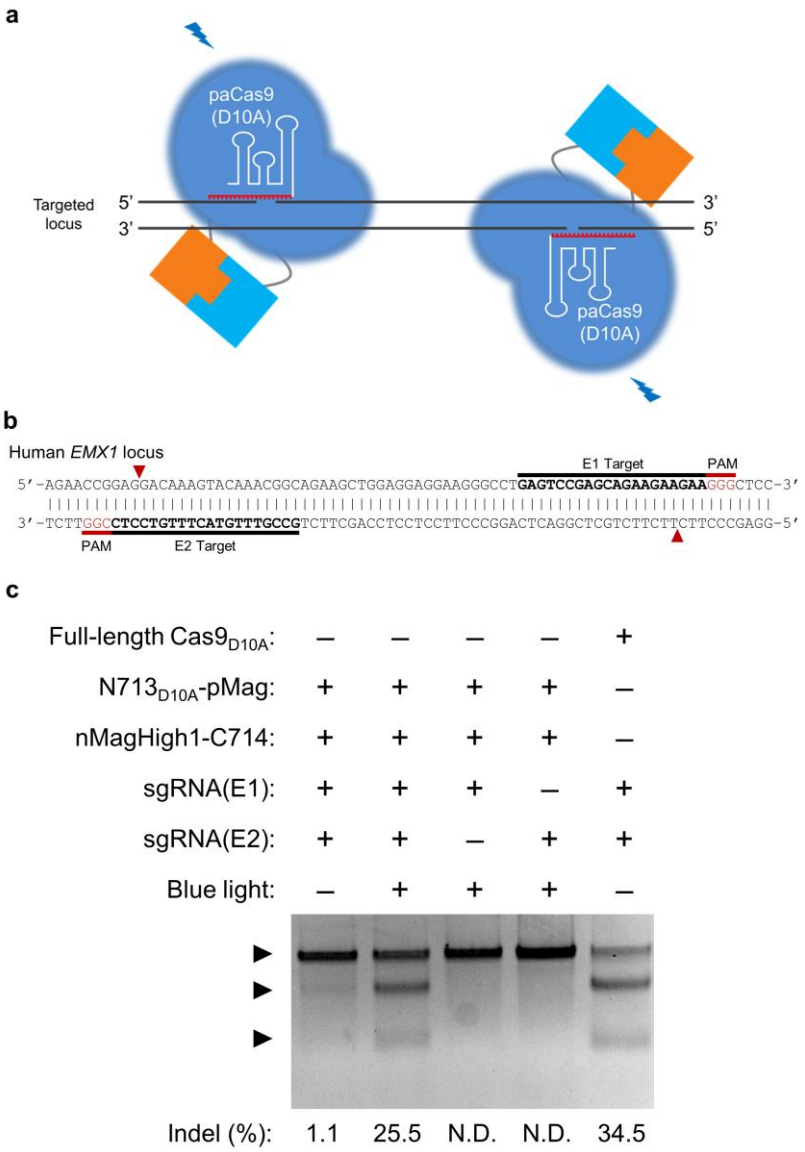

Figure 2-10. paCas9 nickase can facilitate efficient genome editing using a pair of sgRNAs. (**a**) Schematic of light-induced DNA double-strand breaks using a pair of sgRNAs with paCas9 D10A nickase. By introducing D10A mutation in N-terminal fragment of Cas9, paCas9 nuclease can be converted into paCas9 nickase. Using a pair of sgRNAs targeting opposite strands of targeted gene, paCas9 nickase can offer optical control of DNA double-stranded break in targeted locus. (**b**) Schematic of the human *EMX1* locus presenting the target region of a paired sgRNAs indicated by blue lines. PAMs are also shown in red. Red allows indicate putative cleavage site. (**c**) Representative T7E1 assay result for calculating indel mutation rates induced by paCas9 nickase (mean, n=2 from independent experiments).

2-3-4. Reducing background activity of paCas9

Although paCas9-1 can induce efficient NHEJ-mediated indel mutation, paCas9-1 shows slight background activity in dark state (**Figure 2-5a, c**). To reduce the background activity of paCas9-1, I focused on the dynamic-range tunability of Magnets. I replaced nMagHigh1 with nMag because the combination of pMag and nMag shows lower background activity than that of pMag and nMagHigh1[18]. To test this, I transfected HEK293T cells with the pair of N713-pMag and nMag-C714 (named paCas9-2) and sgRNA targeting *VEGFA* locus, and measured induced indel mutations in light and dark state (**Figure 2-11a**). I found that indel mutation in dark state by paCas9-2 was reduced to an undetectable level by the T7E1 assay. Note that the light-induced indel frequency with paCas9-2 is comparable to that with paCas9-1 (**Figure 2-11b**). Therefore, I used paCas9-2 in the following studies.

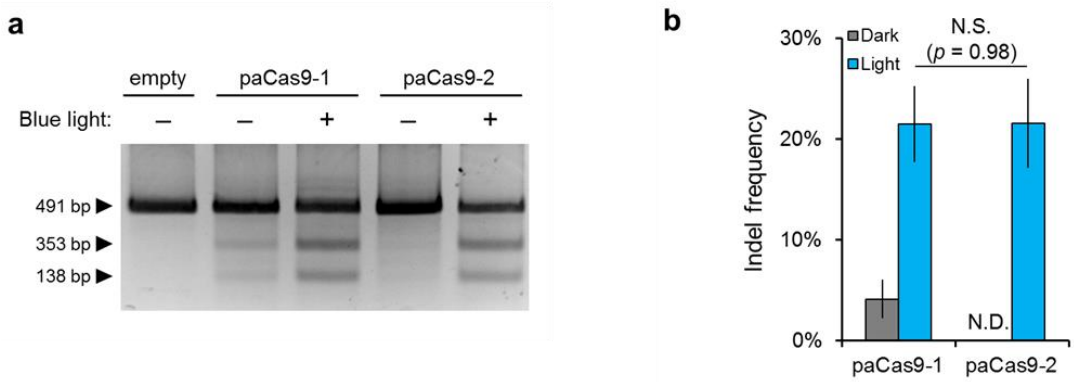

Figure 2-11. (**a**) paCas9-2 dramatically reduces background indel mutation in human *VEGFA* locus with maintaining light-induction potency. (**b**) Quantification of **a**. Data are represented as mean ± s.d. (n=3 from individual experiments). Two-sided Welch's *t*-test was performed. N.S., not significant.

2-3-5. Spatiotemporal control of paCas9

Next, I tested whether paCas9 can provide spatial activation of Cas9 nuclease (**Figure 2-12a, b**). To visualize Cas9-induced NHEJ-mediated indel mutations in living cells, I used a surrogate EGFP reporter system that expresses EGFP fluorescence when a double-strand break is introduced into the target sequence by Cas9[23,24]. HEK293T cells transfected with paCas9-2, surrogate EGFP reporter and sgRNA targeting reporter were irradiated with slit-patterned blue light. After 24 h, the slit pattern of EGFP-expression was observed, demonstrating that paCas9-2 can spatially control gene editing by light.

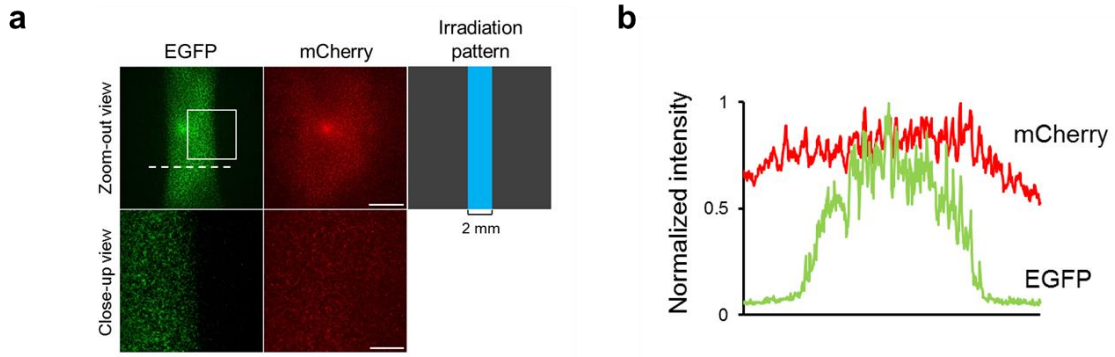

Figure 2-12. (**a**) Spatial activation of paCas9. HEK293T cells were transfected with paCas9-2, NHEJ-dependent EGFP-expressing surrogate reporter and sgRNA targeting the surrogate reporter. 20 h post transfection, samples were irradiated by slit-patterned blue light using a photomask for 24 h. The width of slit is 2 mm. Close-up view within white square region is also shown. Scale bar = 3 mm (zoom-out view) and 1 mm (close-up view), respectively. (**b**) Line scan intensity profile of EGFP (green) and mCherry (red) in **a**.

I also investigated whether paCas9 activation is reversible (**Figure 2-13a, b**). To test this, I firstly transfected HEK293T cells with paCas9-2 and sgRNA targeting *VEGFA*, and incubated these samples upon blue light irradiation for paCas9-2 activation. After 6 h, I split and incubated the cells in dark and light state, and performed second transfection of sgRNA targeting *EMX1*. The samples in dark and light state just before second transfection were incubated again in dark and light state, respectively. After incubation, the genomic DNA were isolated and analyzed by T7E1 assay. Cells irradiated with blue light after first transfection of paCas9-2 and sgRNA targeting *VEGFA* showed indel mutation in *VEGFA* locus, indicating paCas9-2 was activated by blue light. After second transfection of sgRNA targeting *EMX1*, cells irradiated continuously with blue light showed indel mutation in *EMX1* locus; however, cells shifted to dark state did not show indel mutation in *EMX1* locus. This result indicates that paCas9-2 can be switched off by turning off blue light and offer reversible regulation of Cas9 nuclease activity.

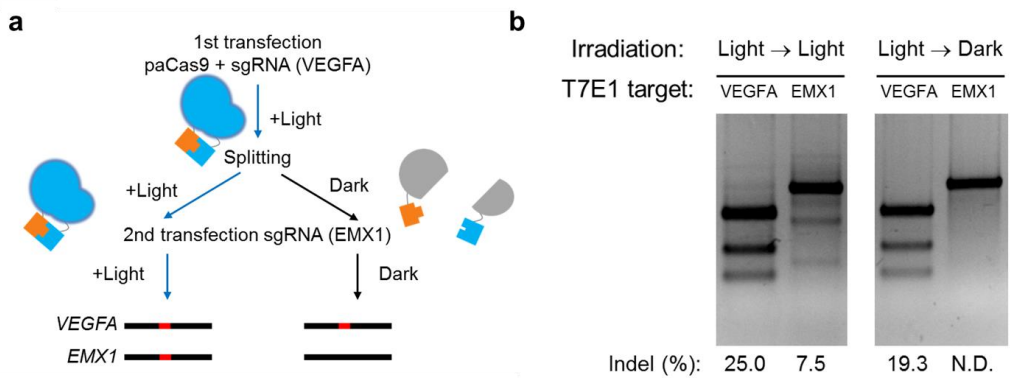

Figure 2-13. (**a**) Experimental scheme to test whether paCas9 activation is reversible. First, HEK293T cells were transfected with paCas9-2 and sgRNA targeting *VEGFA*. After 20 h, cells were illuminated by blue light for 6 h, then split and incubated in light and dark state. After 6 h incubation, the cells in light and dark state were transfected with only sgRNA targeting *EMX1*, and incubated again in light and dark state, respectively. 30 h later, genomic DNA was extracted. If paCas9 activation is reversible, the cells shifted to the dark state before the second transfection of sgRNA targeting *EMX1* show the indel mutation in only *VEGFA* locus. (**b**) A representative gel of the T7E1 assay in **e**. The indel frequencies of *EMX1* and *VEGFA* locus are shown at the bottom of the gel. (n=4 from two independent experiments with biological dupilicates)

2-3-6. Reversible regulation of RNA-guided transcription interference

Several reports have shown that RNA-guided targeting of catalytically inactive Cas9 (dCas9) to a specific gene can sterically block RNA polymerase and transcript elongation, enabling RNA-guided gene silencing, termed CRISPR interference (CRISPRi)[25,26]. To further show the reversibility of paCas9, I also tried photoactivatable and reversible control of CRISPRi. This is named photoactivatable CRISPRi (paCRISPRi) after conventional CRISPRi. To do this, I generated paCas9-2 containing D10A and H840A mutations (padCas9) (**Figure 2-14a**). I also designed three sgRNAs targeting different regions of the CMV promoter-driven luciferase reporter containing PEST and mRNA destabilizing sequences[27]. padCas9 with each sgRNA targeting the luciferase reporter showed light-induced repression of luciferase reporter activity (**Figure 2-14b, c**), demonstrating that my paCas9 platform also provides optogenetic control of RNA-guided transcription. Next, I tested whether padCas9-mediated gene repression can be switched off by halting light irradiation (**Figure 2-14d**). I found that the reporter activity recovered gradually after stopping light irradiation, showing padCas9 has the reversibility. Collectively, these results show that paCas9 can offer spatiotemporal control of RNA-guided genome editing and transcription regulation in a reversible manner.

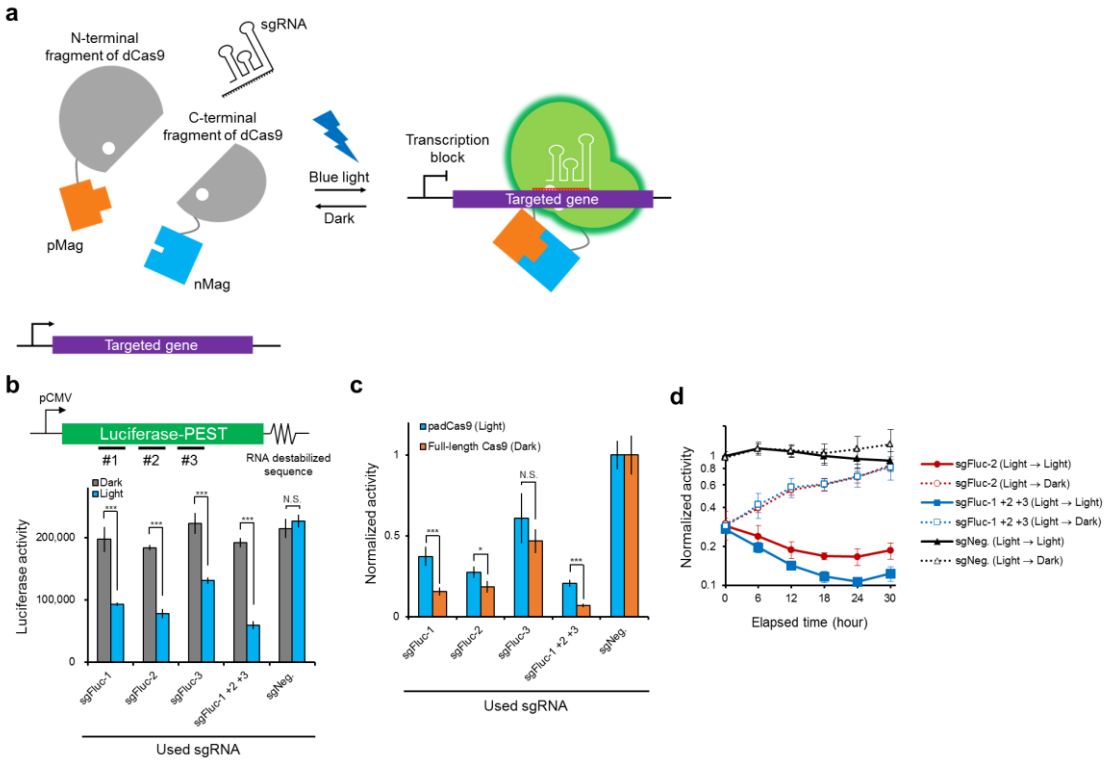

Figure 2-14. Optogenetic control of RNA-guided transcription interference with padCas9. (**a**) Schematic of photoactivatable CRISPR interference (paCRISPRi) with paCas9-2 harboring D10A H840A mutations (padCas9). In the dark, N713 (D10A)-pMag and nMag-C714 (H840A) were inactive. Upon blue light illumination, pMag and nMag are heterodimerized, and consequently N713 (D10A) and C714 (H840A) are reconstituted as functional dCas9, enabling sgRNA-guided transcription interference. (**b**) padCas9 can repress gene expression in light-dependent manner. HEK293T cells were transfected with N713 (D10A)-pMag, nMag-C714 (H840A), luciferase reporter and either indicated sgRNAs targeting luciferase (sgFluc-1, -2 and -3) or an unmatched control sgRNA (sgNeg.). 20 h post transfection, samples were illuminated by 1.2 W/m^2 blue light or kept in the dark for 30 h before measuring luciferase bioluminescence. In this experiment, luciferase reporter containing PEST and mRNA destabilized sequence is used. Data are represented as the mean ± s.e.m. (n=6 from two individual experiments with biological triplicates). Two-sided Welch's *t*-test was performed. N.S., not significant; ***p < 0.005 versus the sample in the dark. (**c**) Comparison of the repression efficiency of padCas9 and full-length dCas9.

Values are normalized to negative control with an unmatched control sgRNA (sgNeg.). Data are represented as the mean ± s.e.m. (n=6 from two individual experiments with biological triplicates). Two-sided Welch's *t*-test was performed. N.S., not significant; *p < 0.05; ***p < 0.005 versus the full-length dCas9 in the dark. (**d**) Time-course of restored luciferase reporter activity post blue light irradiation. HEK293T cells were transfected with N713 (D10A)-pMag, nMag-C714 (H840A), luciferase reporter and the indicated sgRNAs. Immediately post transfection, samples were illuminated by 1.2 W/m^2 blue light for 30 h before measuring bioluminescence at time 0. After time 0 measurement, samples were illuminated by 1.2 W/m^2 blue light (solid lines) or kept in the dark (dotted lines), and bioluminescence was measured every six hours. Data are shown as the mean ± s.e.m. (n=6 from two individual experiments with biological triplicates) normalized to negative control cells (under continuous light irradiation) at time 0.

2-4. Discussion

In conclusion, I have successfully developed photoactivatable Cas9. In initial experiments, I showed that rapamycin-inducible Cas9 activation can be achieved using many split-Cas9 pairs (**Supplementary Figs. 1 and 2**). In preparing this manuscript, a rapamycin-inducible Cas9 similar to mine has been reported, based on the split-Cas9 architecture[6]. However, these rapamycin-inducible Cas9 could not offer spatial and reversible activation because rapamycin is freely diffuse and hard to remove[28,29]. Furthermore, rapamycin can perturb endogenous mTOR signaling pathways[10]. In this study, to achieve spatiotemporal, reversible and non-invasive Cas9 regulation, I further converted rapamycin-inducible Cas9 into photoactivatable Cas9, paCas9. To do this, I firstly applied CRY2-CIB1 system, the most commonly used photoactivatable dimerization system, into split-Cas9 system. However, the CRY2-CIB1 system could not provide optogenetic control of Cas9 nuclease activity. I discovered that Magnet system can provide optogenetic control of Cas9. In addition, applying the tunability of Magnets, I also succeeded in the development of paCas9-2, which provides efficient optogenetic control of Cas9 activity with diminished background activity. Furthermore, I showd that paCas9 can be spatially activated and precisely switched on and off.

I provide the first example of optogenetic tool for spatiotemporally controlling nuclease activity of Cas9. I also report the first evidence that PAM requirement and target specificity of split-Cas9 are indistinguishable from those of full-length Cas9. I also showd that paCas9 is applicable to multiplex indel mutations, HDR-mediated genome modification, DNA double-nicking and transcription regulation. The spatiotemporal and reversible property of paCas9 is well suited for dissection of causal

gene function in various biological processes and medical application, such as *in vivo* and *ex vivo* gene therapies. Also, paCas9 has the potential to reduce off-target indel frequencies in Cas9-based genome editing. There are several studies showing that transient introduction of Cas9:sgRNA complex prepared *in vitro* can improve the specificity of genome editing[7,9,8]. Because paCas9 can be switched off by stopping light irradiation, optically controlling the duration of paCas9 activation would contribute to reducing off-target gene modification.

This paCas9 platform has the potential to further facilitate CRISPR-Cas9 applications. For example, Cas9 applications *in vivo* have been limited due to the packaging size limitation of viral vectors. Because the cDNA length of paCas9 components is smaller than full-length Cas9, it will be possible to package separately each paCas9 fragment into viral vectors having size limitation, expanding the opportunities of *in vivo* genome editing. Another potential application of paCas9 is an intersectional control of Cas9 activity, enabling conditional genome editing with high precision. Recently, conditional gene knockout strategy by combining Cas9 and tissue-specific promoter has been developed[30]. By expressing each paCas9 component from two different tissue-specific promoters, Cas9 activity could be controlled by two promoter activities and light, enabling gene modification with ultrahigh precision.

I and other group have recently developed dCas9-based photoactivatable transcription systems[31,32]. These dCas9-based optogenetic systems can activate targeted endogenous gene expression. Unlike these systems, the present paCas9 can offer a pioneering approach to photoactivatable targeted genome editing via NHEJ and HDR pathways. In addition, I showed that padCas9 can offer paCRISPRi, an optogenetic control of targeted gene silencing. The present paCas9 will contribute to expand the

possibilities of optogenetic regulation, genome engineering and biomedical applications.

2-5. References

1. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
2. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
3. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471 (2013).
4. Dow, L. E. *et al.* Inducible in vivo genome editing with CRISPR-Cas9. *Nat. Biotechnol.* **33**, 390–394 (2015).
5. González, F. *et al.* An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* **15**, 215–226 (2014).
6. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33**, 139–142 (2015).
7. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012–1019 (2014).
8. Zuris, J. a *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* **33**, 73-80 (2015).
9. Ramakrishna, S. *et al.* Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome Res.* **24**, 1020–1027 (2014).

10. Laplante, M. & Sabatini, D. M. mTOR signaling in growth control and disease. *Cell* **149**, 274–293 (2012).

11. Brieke, C., Rohrbach, F., Gottschalk, A., Mayer, G. & Heckel, A. Light-controlled tools. *Angew. Chem. Int. Ed. Engl.* **51**, 8446–8476 (2012).

12. Lee, H., Larson, D. & Lawrence, D. Illuminating the chemistry of life: design, synthesis, and applications of "caged" and related photoresponsive compounds. *ACS Chem. Biol.* **4**, 409–427 (2009).

13. DeRose, R., Miyamoto, T. & Inoue, T. Manipulating signaling at will: chemically-inducible dimerization (CID) techniques resolve problems in cell biology. *Pflugers Arch.* **465**, 409–417 (2013).

14. Nishimasu, H. *et al.* Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156**, 935-949 (2014).

15. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573 (2014).

16. Kennedy, M. J. *et al.* Rapid blue-light–mediated induction of protein interactions in living cells. *Nat. methods* **7**, 973–975 (2010).

17. Bugaj, L. J., Choksi, A. T., Mesuda, C. K., Kane, R. S. & Schaffer, D. V. Optogenetic protein clustering and signaling activation in mammalian cells. *Nat. Methods* **10**, 249–252 (2013).

18. Kawano, F., Suzuki, H., Furuya, A. & Sato, M. Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nat. Commun.* **6**, 6256 (2015).

19. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).

20. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).

21. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).

22. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).

23. Kim, H. *et al.* Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat. Methods* **8**, 941–943 (2011).

24. Ramakrishna, S. *et al.* Surrogate reporter-based enrichment of cells containing RNA-guided Cas9 nuclease-induced mutations. *Nat. Commun.* **5**, 3378 (2014).

25. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).

26. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).

27. Voon, D. C. *et al.* Use of mRNA- and protein-destabilizing elements to develop a highly responsive reporter system. *Nucleic Acids Res.* **33**, e27 (2005).

28. Lin, Y. *et al.* Rapidly reversible manipulation of molecular activity with dual chemical dimerizers. *Angew. Chem. Int. Ed. Engl.* **52**, 6450-6454 (2013).

29. Hosoi, H. *et al.* Rapamycin causes poorly reversible inhibition of mTOR and induces p53- independent apoptosis in human rhabdomyosarcoma cells. *Cancer Res.* **59**, 886–894 (1999).

30. Shen, Z. *et al.* Conditional knockouts generated by engineered CRISPR-Cas9 endonuclease reveal the roles of coronin in C. elegans neural development. *Dev. Cell* **30**, 625–636 (2014).

31. Chapter 3 in this thesis

32. Polstein, L. R. & Gersbach, C. A. A photoactivatable CRISPR-Cas9 system for control of endogenous gene activation. *Nat. Chem. Biol.* **11**, 198–200 (2015).

Chapter 3.

CRISPR-Cas9-based Photoactivatable Transcription System

3-1. Introduction

Precise spatial and temporal pattern of multiple gene expression is crucial for various biological phenomena, such as cellular programming, metabolism, homeostasis, memory formation and circadian rhythm. To understand gene functions in these phenomena, the methods that enable to regulate endogenous gene expression are required. To date, synthetic transcription factors based on programmable DNA-binding domains of zinc-finger proteins and transcription activator-like effectors (TALE) have offered potent tools for targeted endogenous gene expression[1–3]. In particular, LITE system based on TALE is an outstanding technology, enabling spatiotemporal control of endogenous gene expression[4]. LITE system is an optogenetic two-hybrid system, consisting of the DNA-binding domain based on TALE, and the light-sensitive cryptochrome 2 (CRY2) and its binding partner CIB1 from *Arabidopsis thaliana*[5]. This system can be customized to target a wide range of endogenous genomic loci by redesigning TALE. However, due to their repetitive DNA sequences, constructing TALE for a given target site requires complicated and time-consuming multistage DNA assembly. This causes not only the entry barrier for researchers but also the difficulty of multiplexed user-defined genes regulation.

To address this, we focused on the new class of programmable genome targeting technology, CRISPR (clustered regularly interspaced palindromic repeats)-Cas (CRISPR-associated) system. CRISPR-Cas systems from bacteria have enabled the development of easy-to-use targeted genome editing technology[6–8]. By single-guide

RNA (sgRNA), the *Streptococcus pyogenes* Cas9 protein can be bound on and cleave a target endogenous genome sequence that is complementary to the first 20 nucleotide (nt) of the sgRNA and is adjacent to a protospacer-adjacent motif (PAM) of the form NGG. This Cas9-induced double-strand break in chromosome is repaired by the homologous recombination and non-homologous end joining, enabling targeted genome editing. In CRISPR-Cas system, targeting Cas9 into given genome sequences requires just designing the first complementary 20 nt of the sgRNA done by simple plasmid construction, and therefore Cas9 technology offers a simple method for simultaneously targeting multiple genomic loci. In addition, several studies have shown that catalytically inactive Cas9 (dCas9) provides the programmable RNA-guided DNA binding domain[9,10]. Recently, dCas9 combined with transcriptional activator domain enables synergistic and multiplexed mammalian gene activation[11–13]. Here, we showd targeted genome photoactivation system based on dCas9 and thereby developed the simple and versatile tool for spatiotemporally activating multiple user-defined endogenous genes.

3-2. Materials and Methods

3-2-1. Plasmid construction

cDNAs encoding the codon-optimized *Streptococcus pyogenes* Cas9, single guide RNA, CRY2PHR, CIB1(*NLS), VP64 and p65 activator domain (residues 286-550) were amplified from Addgene plasmids 42230, 44248, 26871, 28240, 43884, 21966, respectively, using standard PCR approach. The plasmid containing CIB1 were obtained from RIKEN BioResource Center (Resource Number: pda10875). dCas9 (Cas9-D10A, H840A) was generated by Multi Site-Directed Mutagenesis Kit (MBL) according to manufacturer's directions. LITE2.0 anchor (TALE(*Neurog2*)-trCIB1(NLS*)) was amplified from Addgene plasmids 47458 and inserted into *Bam*HI and *Eco*RI sites of pcDNA3.1 V5/His-A (Invitrogen). LITE2.0 activator (NLS-CRY2PHR-NLS-VP64) was amplified from Addgene plasmids 47457 and inserted into *Hin*d III and *Xba*I sites of pcDNA3.1 V5/His-A. Other chimeric fusion constructs encoding the genomic anchor probes and activator probes were inserted into *Hin*d III and *Eco*RI sites of pcDNA3.1 V5/His-A. The sgRNAs targeting GAL4UAS-Luciferase reporter, *ASCL1*, *MYOD1*, *NANOG* and *IL1RN* were generated by standard PCR, and these fragments were inserted into *Bst*XI and *Not* I sites of Addgene plasmids 44248. The sgRNA targeting tet-inducible reporter and *Neurog2* was generated by annealed oligo cloning using BbsI site of Addgene plasmid 47108. GAL4UAS-Luciferase reporter plasmid was constructed by replacing firefly luciferase sequence including protein degradation sequence to firefly luciferase sequence in pGL4.31 vector (Promega) using *Hin*d III and *Xba* I sites. pNeurog2-Luciferase plasmid, luciferase reporter containing the part of Neurog2 promoter sequence, was constructed by replacing the part of GAL4UAS to the

part of Neurog2 promoter sequence with additional PAM sequence (5'-GTGAATGATGATAATACGATGCGG-3') in GAL4UAS-Luciferase reporter using *Xho* I *and Hin*d III sites. Tet-inducible mCherry and luciferase reporter plasmids were generated by standard PCR, and the fragment were inserted into *Eco*RI and *Bgl* II sites of pKM006. cDNAs encoding tetracycline-controlled transactivator 3 was amplified from Addgene plasmid 26429, and was inserted into *Eco*RI and *Xho* I sites in pcDNA 3.1/V5-HisA.

3-2-2. Cell culture

HEK293, HeLa, COS-7 and HEK293T cells were cultured at 37 °C under 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM, Sigma Aldrich) supplemented with 10% FBS (HyClone), 100 unit/ml penicillin and 100 μg/ml of streptomycin (GIBCO).

3-2-3. Bioluminescence assay

For reporter gene expression assay, HEK293, HeLa, COS-7 and HEK293T cells were plated at approximately 2.0×10^4 cells/well in 96-well black-walled plate (Thermo Fisher Scientific), and cultured for 24 h at 37 °C in 5% CO_2. The cells were then transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocols. cDNAs encoding genomic anchor probes, activator probes, sgRNAs and reporter were transfected at a 1:1:1:1 ratio. In this experiment, the ratio of three sgRNAs was 1:1:1. The total amount of DNA was 0.2 μg/well. Twenty hours after the transfection, the culture medium was replaced with 100 μl of phenol red-free DMEM (Sigma Aldrich) containing 500 μM of D-luciferin (Wako Pure Chemical Industries) as a

substrate. After incubation for 24 h at 37 °C in 5% CO_2 under continuous blue light irradiation or in the dark, bioluminescence measurements were performed using Centro XS^3 LB 960 plate-reading luminometer (Berthold Technologies). In time-course experiment, bioluminescence measurements were performed at indicated time points. For Tet-inducible expression, 1.0 μg/ml doxycycline (Dox) was used. Blue light irradiation was performed using a 470 nm ± 20 nm LED light source (CCS Inc.). Intensity of blue light was 1.5 W/m^2.

3-2-4. Spatial gene activation experiment

HEK293T cells were plated at 8.0×10^5 cells/dish on 35 mm dish (Iwaki Glass) coated with fibronectin (BD Biosciences), and cultured for 24 h at 37 °C in 5% CO_2. The cells were then transfected with Lipofectamine2000 according to the manufacturer's protocols. cDNAs encoding genomic anchor probes, activator probes, sgRNAs and mCherry reporter were transfected at a 1:1:1:1 ratio. The total amount of DNA was 4.0 μg/dish. Twenty hours after the transfection, samples were illuminated by slit-patterned blue light using a black masking tape for 24 h at 37 °C in 5% CO_2. The widths of slits are 2.5 and 1.5 mm, respectively. Images were acquired using Leica M205 FA stereomicroscope (Leica), and analyzed using Metamorph (Molecular Devices).

3-2-5. Quantitative real-time PCR analysis

For measuring relative mRNA amount by quantitative real-time PCR, HEK293T cells were plated at approximately 2.0×10^4 cells/well in 96-well black-walled plate, and cultured for 24 h at 37 °C in 5% CO_2. The cells were then transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocols. cDNAs encoding genomic anchor probes, activator probes, sgRNAs were transfected at a 1:1:1 ratio. For synergistic activation by multiple sgRNAs, the ratio of four sgRNAs was 1:1:1:1. In multiplexed photoactivation, the ratio of set of four sgRNAs targeting *ASCL1* and *MYOD1* is 1:1. The total amount of DNA was 0.2 μg/well. Twenty hours after the transfection, samples were incubated at 37 °C in 5% CO_2 under continuous blue light irradiation or in the dark. After 24 h incubation, total RNA isolation and reverse transcription PCR were performed using cells-to-ct kit (Life technologies) according to the manufacturer's protocols. Real-time PCR was performed by the StepOnePlus system (Life technologies) using TaqMan Gene Expression Master Mix (Life technologies). TaqMan probes for detecting each target gene and *GAPDH* as endogenous control were used (Life technologies, TaqMan Gene Expression Assay IDs are following; *ASCL1*: Hs04187546_g1, *MYOD1*: Hs02330075_g1, *NANOG*: Hs04260366_g1, *IL1RN*: Hs00893626_m1, *GAPDH*: Hs99999905_m1). Standard $\Delta\Delta C_t$ method was used to obtain each relative mRNA level to negative controls transduced with empty vector only in the dark.

3-2-6. Statistical analysis

All values are expressed as mean ± SEM. Statistical significance was determined using a Student's *t* test with two-tailed distribution.

3-3. Results

3-3-1. Design and optimization of CRISPR-Cas9-based photoactivatable transcription system

This CRISPR-Cas9-based photoactivatable transcription system consists of two fusion proteins and sgRNAs (**Figure 3-1a**). The first fusion protein is the genomic anchor probe, containing dCas9 and CIB1 (dCas9-CIB1). This anchor probe binds to targeted genome sequence by sgRNAs. The second fusion protein is the activator probe, which includes the photolyase homology region of CRY2 and transcriptional activator domain p65 (CRY2PHR-p65). In the absence of blue light, dCas9-CIB1 binds the promoter region of the targeted gene by sgRNAs while CRY2PHR-p65 is freely diffuse within the nucleus. Upon blue light irradiation, CRY2PHR and CIB1 are heterodimerized, and consequently the transcriptional activator domain is recruited to the target locus to activate gene expression.

I generated the two fusion proteins and assessed the induction potency of each their combination by measuring reporter gene activity in both light and dark state conditions. In this assessment, I constructed firefly luciferase reporter under the control of upstream activator sequence of Gal4 (UAS). There have been several studies showing that the level of gene activation by synthetic transcription factor based on dCas9 and activation domain can be synergistically enhanced by using multiple sgRNAs targeting different sites in the promoter region of the same locus[5–7]. To obtain robust reporter gene activation in this experiment, I simultaneously transfected triple sgRNAs targeting different regions of UAS into HEK293 cells with genomic anchor probe, activator probe and luciferase reporter.

To build the effective photoactivatable transcription system, I generated the several genomic anchors and activators having different additional motifs (**Figure 3-1b**). Nuclear localization signal (NLS) derived from SV40 Large T-antigen is used to localize these probes into the nucleus. For adequate nuclear localization, I also tested three tandem repeat of the NLS sequence (NLSx3). Furthermore, I made alternations in CIB1 domain. I tested C-terminal truncated (Δ308-334) CIB1 (trCIB1), reported as the mutant of which transcriptional factor-like domain has been partially removed[9]. I also examined a CIB1 variant that is mutated at internal NLS (CIB1(*NLS))[8].

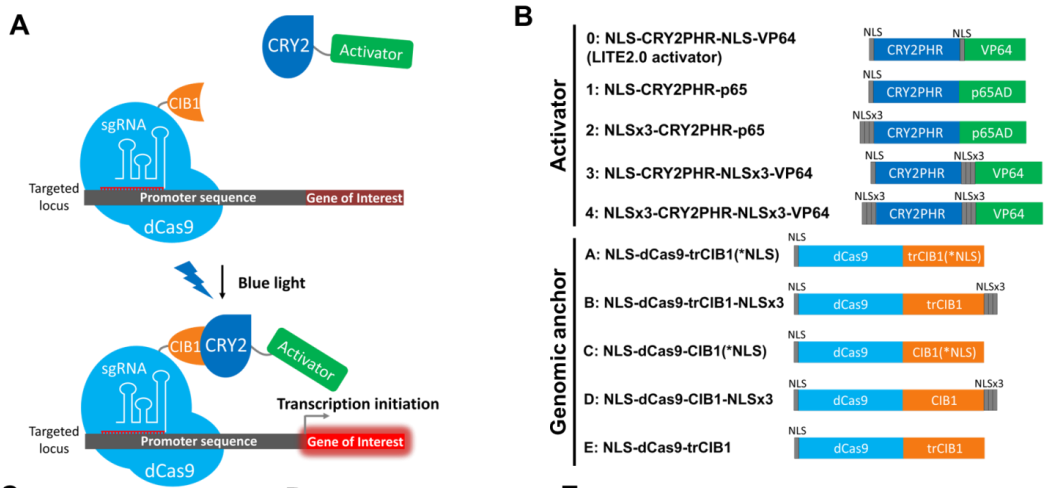

Figure 3-1. Design and optimization of CRISPR-Cas9-based photoactivatable transcription system. (**A**) Schematic of the CRISPR-Cas9-based photoactivatable transcription system. (**B**) Constructs of genomic anchor probes (A-E) and activator probes (0-4).

Firstly, I tested which activator probes could most efficiently induce reporter gene expression with the genomic anchor, NLS-dCas9-trCIB1(*NLS) (**Figure 3-2A**). Four of the five activator units showed light-induced reporter upregulation in HEK293 cells. In particular, NLSx3-CRY2PHR-p65 yielded highest reporter activation in light state as

well as fold induction (16.1-fold). Conversely, I found that the A0 combination of NLS-dCas9-trCIB1(*NLS) and NLS-CRY2PHR-NLS-VP64, which is similar to the design and combination of optimized LITE system, optical gene regulation system based on transcription activator-like effector (TALE), could not significantly induce reporter gene expression[9]. This result showed that just replacing TALE with dCas9 in LITE system is insufficient to build CRISPR-Cas9-based photoactivatable transcription system.

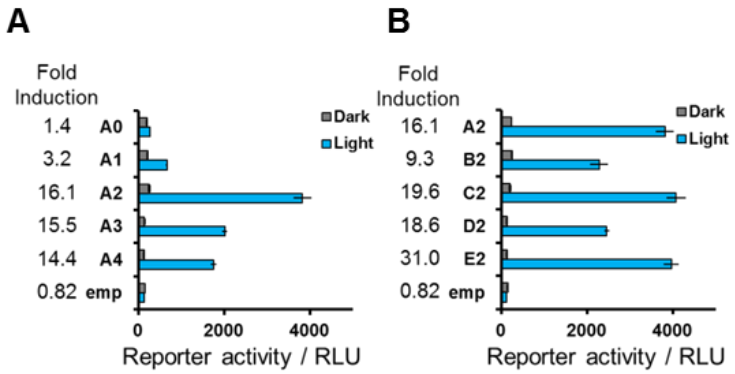

Figure 3-2. (**A**) Luciferase reporter activity induced by each combination of the genomic anchor probe (A: NLS-dCas9-trCIB1(*NLS)) and activator probes (0-4). (**B**) Luciferase reporter activity induced by each combination of the genomic anchor probe (A-E) and activator probes (2: NLSx3-CRY2PHR-p65).

Next, I optimized genomic anchor for reducing background activity (**Figure 3-2B**). I compared five genomic anchor units and found that the E2 combination of NLS-dCas9-trCIB1 and NLSx3-CRY2PHR-p65 yielded lowest background activity and highest fold induction (31.0-fold).

For further functional validation of this combination, I transfected these constructs into HeLa and COS-7 cells (**Figure 3-3A, B**). All samples showed light-induced reporter gene expression, demonstrating the wide applicability of this photoactivatable transcription system. In addition, I compared the gene expression kinetics between this

Cas9-based photoactivatable transcription system and tetracycline-inducible system, Tet-On system (**Figure 3-3C, D**). I confirmed that, in terms of the gene expression kinetics, developed Cas9-based photoactivatable transcription system can match Tet-On system, which is commonly used in biological sciences. In the following experiments, I used this E2 combination constructs.

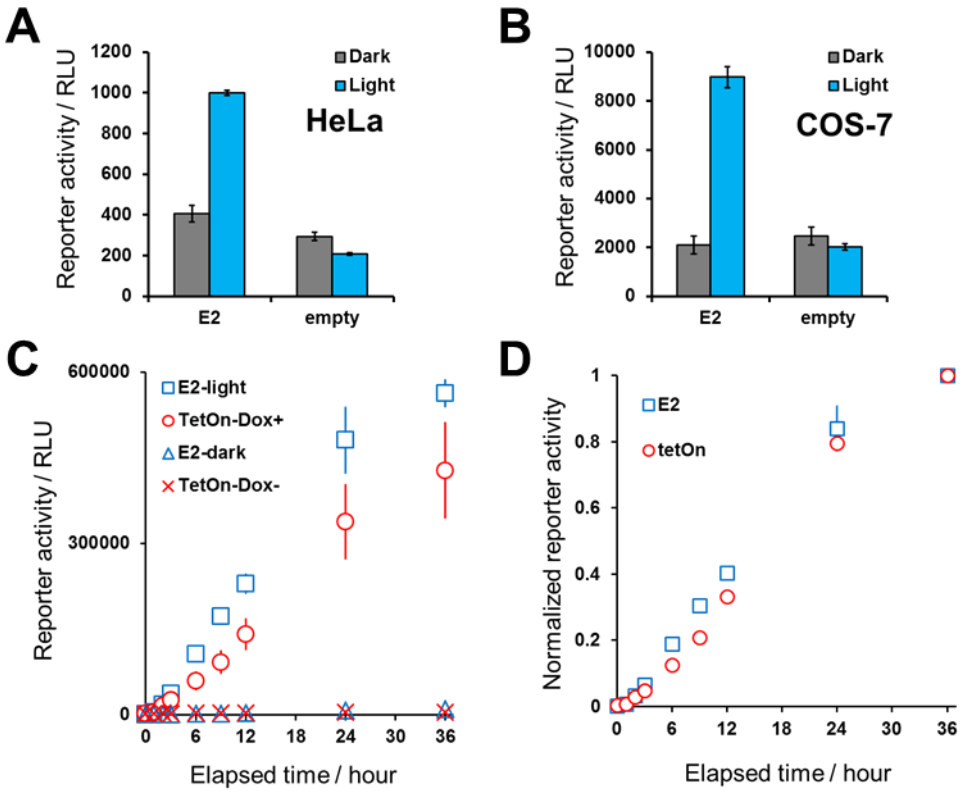

Figure 3-3. (**A, B**) Light-induced luciferase reporter activity in HeLa cells (**A**) and COS-7 cells (**B**). (**C**) Comparison of the gene expression kinetics between the CRISPR-Cas9-based photoactivatable transcription system and Tet-On system. (**D**) Normalized data from **C**. Error bars, SEM (n=6 from two individual experiments with biological triplicates).

3-3-2. Spatial gene activation by patterned illumination

I tested whether this system offer spatial gene activation by light (**Figure 3-4**). To perform this, I generated mCherry reporter for visualizing the expression pattern of reporter gene with a fluorescence stereomicroscope. HEK293T cells transfected with NLS-dCas9-trCIB1, NLSx3-CRY2PHR-p65, mCherry reporter, sgRNAs targeting mCherry reporter, and EGFP as transfection marker were irradiated with slit-patterned blue light. After 24 h, the slit pattern of mCherry-expressing cells according with irradiation pattern was observed (**Figure 3-4**), demonstrating that this system could spatially control gene expression.

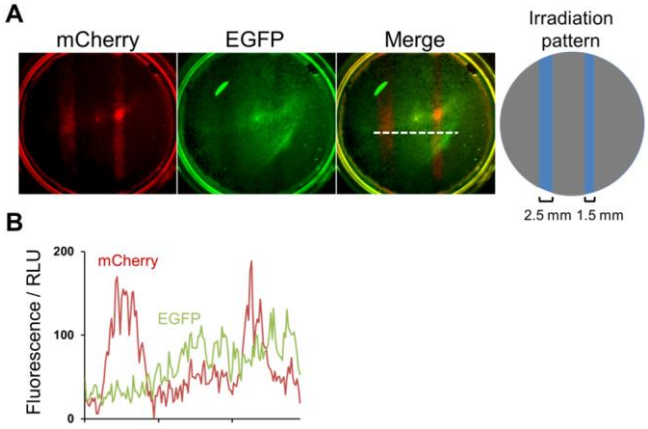

Figure 3-4. Spatial gene activation by CRISPR-Cas9-based photoactivatable transcription system. (**A**) Slit-patterned mCherry expression in HEK293T cells illuminated by blue light with a spatial pattern using a black masking tape. The widths of slits are 2.5 and 1.5 mm, respectively. (**B**) Linescan intensity profile of mCherry (red) and EGFP (green) in (A).

3-3-3. Optogenetic activation of endogenous gene by the CRISPR-Cas9-based photoactivatable transcription system

I next tested whether this system could also optically activate endogenous gene expression in HEK293T cells (**Figure 3-5A**). To do this, I generated four sgRNAs

targeting the different sequences in the promoter of human *ASCL1* gene, which encodes the transcription factor regulating neural differentiation. To determine which sgRNA can most efficiently induce *ASCL1* expression, I transfected NLS-dCas9-trCIB1 and NLSx3-CRY2PHR-p65 with individual sgRNAs into HEK293T cells. By comparative quantitative PCR, I confirmed that *ASCL1* has been successfully activated in all the cases using each sgRNA (**Figure 3-5A**). When I simultaneously transfected all the four sgRNAs targeting the promoter region of *ASCL1, ASCL1* expression in the light state was significantly enhanced, compared to using individual sgRNAs as expected (**Figure 3-5A**)[5–7]. It is notable that the transfection of multiple sgRNAs targeting *ASCL1* did not affect the *ASCL1* expression in the dark and showed substantially high induction ratio of *ASCL1* expression by light (~50-fold). These results show that photoactivatable transcription system can induce endogenous gene expression by light, and the level of gene expression could be synergistically enhanced by multiple sgRNAs.

Next, I investigated the time course of light-induced *ASCL1* transcription. I found that three hours of blue light irradiation was enough to induce significant *ASCL1* mRNA transcription (~10-fold) (**Figure 3-5B**). I also tested whether endogenous gene activation by this system is reversible and repeatable. Eighteen hours incubation in the dark after first blue light irradiation reduced *ASCL1* mRNA expression level to baseline level, and second light irradiation can induce *ASCL1* mRNA expression again (**Figure 3-5C**). These results show that this system can offer rapid, reversible and repeatable endogenous gene activation.

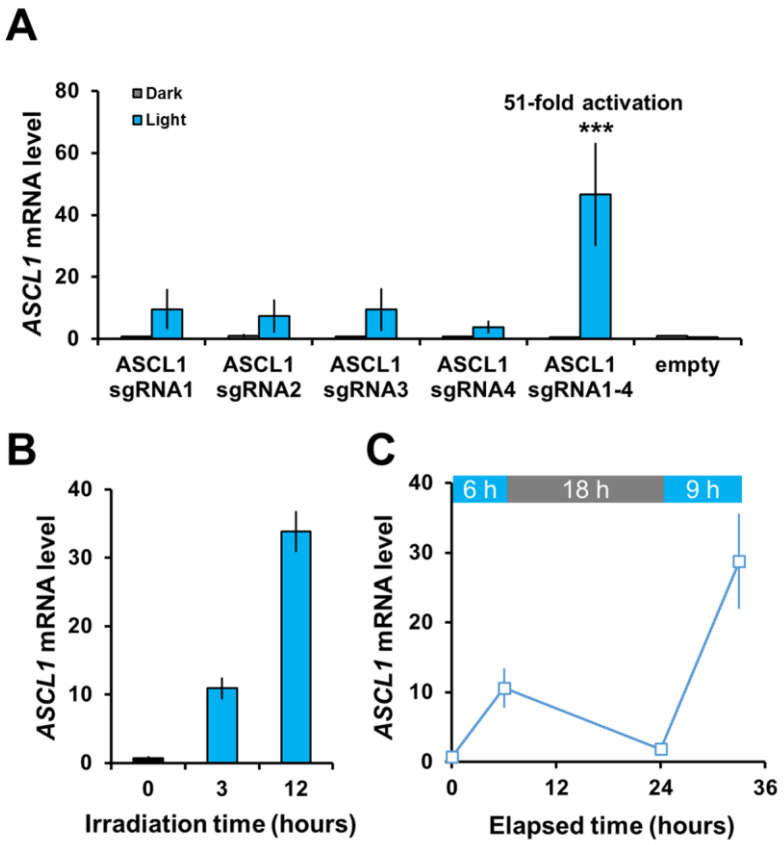

Figure 3-5. Optogenetic activation of endogenous *ASCL1* gene in HEK293T cells by the CRISPR-Cas9-based photoactivatable transcription system. (**A**) Light-induced *ASCL1* expression in HEK293T cells measured by qRT-PCR. The four sgRNAs targeting the promoter region of *ASCL1* were transfected individually or in combination, as indicated. (**B**) The time course of light-induced *ASCL1* transcription. (**C**) Reversible and repeatable activation of *ASCL1* transcription. In these experiments, the data are expressed as relative mRNA amount to the negative control transfected with empty vector in the dark. Error bars, SEM (In A, n≥6 from at least two individual experiments with biological triplicates. In **B** and **C**, n=3 from the same experiment.). Student's two-tailed *t* test was performed. *p < 0.05, **p < 0.01, ***p < 0.001 versus the sample in the dark.

3-3-4. Multiplexed photoactivation of user-defined endogenous genes

To show that this system allows the photoactivation of various endogenous genes, I generated four each sgRNAs targeting the promoter regions of human *MYOD1*, *NANOG*, and *IL1RN* gene and tested their light-induced transcription in HEK293T cells (**Figure 3-6A,B,C**). The light-dependent transcription of each gene was observed when HEK293T cells were cotransfected with the four sgRNAs. Next, I tested multiple photoactivation of *ASCL1* and *MYOD1* genes in HEK293T cells (**Figure 3-6D**). In the sample transfected with NLS-dCas9-trCIB1, NLSx3-CRY2PHR-p65 and multiple sgRNAs targeting the promoters of *ASCL1* and *MYOD1*, I observed the light-dependent transcription of both *ASCL1* and *MYOD1*. I found no significant difference in *ASCL1* and *MYOD1* expression levels between the multiple and single photoactivation experiments ($P > 0.20$), indicating that *ASCL1* and *MYOD1* activation are saturated in my transfection conditions. I also confirmed that cells transfected with sgRNAs targeting the promoter of *ASCL1* showed the light-induced transcription of *ASCL1* without affecting *MYOD1* expression, and vice versa. These results show that this system can be used for multiplexed photoactivation of user-defined endogenous genes.

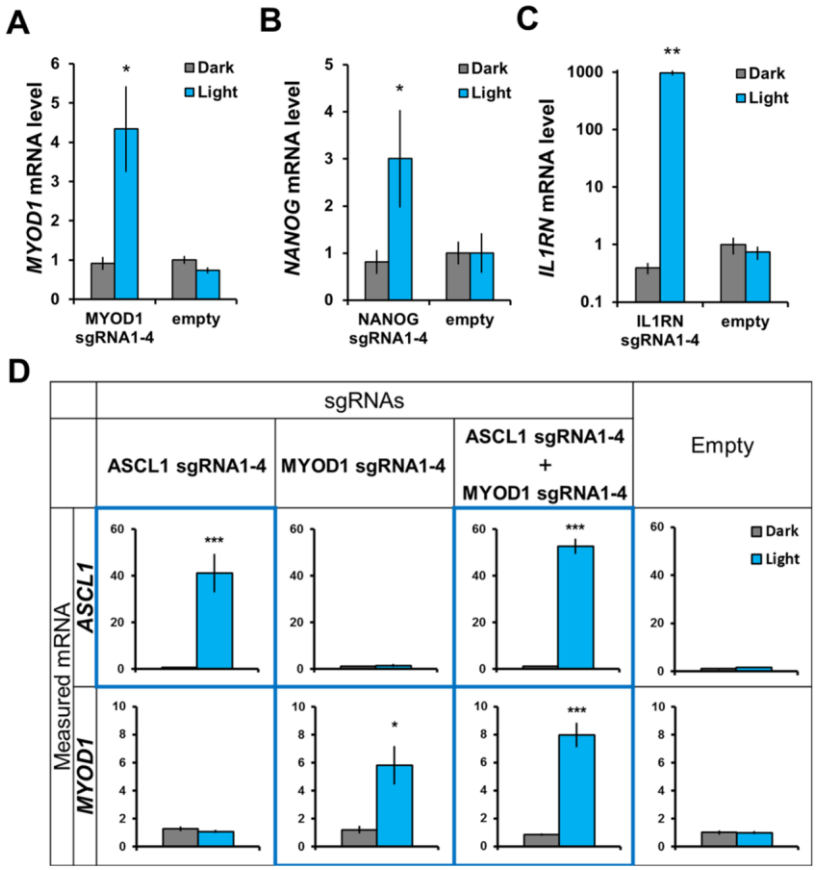

Figure 3-6. Optogenetic activation of various endogenous genes in HEK293T cells by the CRISPR-Cas9-based photoactivatable transcription system. (**A,B,C**) Light-induced *MYOD1* (A), *NANOG* (B), and *IL1RN* (C) expression in HEK293T cells measured by qRT-PCR. The four sgRNAs targeting the promoter region of each gene were transfected in combination. (**D**) Multiplexed endogenous genes photoactivation. HEK293T cells were transfected with cDNAs encoding NLS-dCas9-trCIB1, NLSx3-CRY2PHR-p65 and the indicated sgRNAs. In these experiments, the data are expressed as relative mRNA amount to the negative control transfected with empty vector in the dark. Error bars, SEM (n≥6 from at least two individual experiments with biological triplicates.). Student's two-tailed *t* test was performed. *p < 0.05, **p < 0.01, ***p < 0.001 versus the sample in the dark.

3-4. Discussion

In conclusion, I have developed the CRISPR-Cas9-based photoactivatable transcription system. Previously, it has been reported that optical endogenous gene activation can also be achieved by LITE system based on TALE[9]. Unlike LITE system, which requires complex and time-consuming DNA assembly for targeting a given sequence, my CRISPR-Cas9-based transcription system provides easy-to-use user-defined endogenous gene activation. The capacity of CRISPR-Cas9-based photoactivatable transcription system was exemplified by synergistic photoactivation of endogenous genes. Also, I firstly showd multiplexed photoactivation of user-defined endogenous genes. These features of the present photoactivation system are due to easy programmability and highly parallel applicability of the CRISPR-Cas technology. I exemplified multiplexed gene expression by using 8 different sgRNAs. The number of sgRNA is readily increased to a large extent, according to the recent report by Chen *et al*, in which ~70 sgRNAs were used for imaging specific genomic loci based on dCas9 tagged with EGFP[10]. Thus, the present optogenetic system based on the CRISPR-Cas technology with easy programmability and highly parallel applicability could achieve highly synergistic and massively multiplexed spatiotemporal activation of endogenous genes with light.

Konermann and colleagues mentioned that TALE DNA-binding domain of LITE system might be replaced with dCas9[9]. However, I found that replacing TALE with dCas9 in LITE system could not provide dCas9-based optical gene activation. One possible explanation for this is that the larger and more complex structure of dCas9 than that of TALE may deteriorate the localization and dimerization of the two fusion

proteins. To overcome this, I generated the originally designed molecular probe and finally established the robust dCas9-based optical gene activation system. To my knowledge, the present CRISPR-Cas9-based photoactivatable transcription system is the first optogenetic tool employing Cas9 technology. Furthermore, by replacing transcriptional activator p65 with other domains, such as epigenetic-modifying enzymes[11,12], nucleases[13] and recombinases[14], this system could also offer various types of Cas9-guided genome regulation with high spatiotemporal resolution. The present CRISPR-Cas9-based photoactivatable transcription system will contribute to expanding possibilities of optogenetic regulation, mammalian genome engineering and biotechnology applications[15,16].

3-5. References

1. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343,** 1247997 (2014).
2. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156,** 935–949 (2014).
3. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152,** 1173–1183 (2013).
4. Gilbert, L. a *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442–451 (2013).
5. Perez-pinera, P. *et al.* RNA-guided gene activation by CRISPR- Cas9 – based transcription factors. *Nat. Methods* **10,** 973–976 (2013).
6. Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10,** 977–979 (2013).
7. Cheng, A. W. *et al.* Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23,** 1163–1171 (2013).
8. Kennedy, M. J. *et al.* Rapid blue-light–mediated induction of protein interactions in living cells. *Nat. methods* **7**, 973–975 (2010).
9. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500,** 472–476 (2013).
10. Chen, B. *et al.* Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **155,** 1479–1491 (2013).

11. de Groote, M. L., Verschure, P. J. & Rots, M. G. Epigenetic Editing: targeted rewriting of epigenetic marks to modulate expression of selected target genes. *Nucleic Acids Res.* **40,** 10596–613 (2012).

12. Maeder, M. L. *et al.* Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat. Biotechnol.* **31,** 1137–1142 (2013).

13. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29,** 143–8 (2011).

14. Mercer, A. C., Gaj, T., Fuller, R. P. & Barbas, C. F. Chimeric TALE recombinases with programmable DNA sequence specificity. *Nucleic Acids Res.* **40,** 11163–72 (2012).

15. Lienert, F., Lohmueller, J. J., Garg, A. & Silver, P. A. Synthetic biology in mammalian cells : next generation research tools and therapeutics. *Nat. Rev. Mol. Cell. Biol.* **15**, 95-107 (2014)

16. Brieke, C., Rohrbach, F., Gottschalk, A., Mayer, G. & Heckel, A. Light-controlled tools. *Angew. Chem. Int. Ed. Engl.* **51,** 8446–76 (2012).

Chapter 4.

Highly efficient CRISPR-Cas9-based Photoactivatable Transcription System

4-1. Introduction

Complex and dynamic gene expression is crucial for various biological phenomena, such as cell division, cellular differentiation, memory formation and circadian rhythm. To dissect gene functions in these phenomena, the methods that enable to precisely and efficiently manipulate expression of targeted endogenous genes are required. For manipulating targeted endogenous gene expression, the new class of programmable genome targeting technology, CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) system, has offered potent tools[1]. By single-guide RNA (sgRNA), the catalytically inactive Cas9 protein (dCas9) derived from *Streptococcus pyogenes* can be bound on a target endogenous genome sequence that is complementary to the first 20 nucleotide (nt) of the sgRNA and is adjacent to a protospacer-adjacent motif (PAM) of the form NGG[2,3]. Several initial studies have shown that dCas9 combined with transcriptional activator domain, such as dCas9-VP64 and dCas9-p65AD, enables the activation of user-defined endogenous gene[2,4,5].

These dCas9 activators can be further combined with inducible dimerization domains to spatiotemporally control gene expression, which is useful for dissecting dynamic and complex gene networks. I and other group previously created photoactivatable dCas9-based transcriptional activators, named CRISPR-Cas9-based photoactivatable transcription system[6] (CPTS) and light-activated CRISPR-Cas9 effector[7] (LACE), respectively. In these systems, dCas9 and a transcriptional activator

VP64 or p65AD are combined with blue-light-sensitive heterodimerization proteins CIB1 and cryptochrome 2 (CRY2), respectively[8]. Blue light stimulation induces dimerization of CRY2-CIB1, recruiting VP64 or p65AD to dCas9-targeted locus for activating endogenous gene. These systems can activate targeted endogenous gene in a spatial, temporal and reversible manner using blue light. However, the magnitude of transcriptional upregulation by these photoactivatable dCas9 activators is severely low to elicit a biological response and limits the application possibility. To overcome this, here I present highly efficient CRISPR-Cas9-based photoactivatable transcription systems via sgRNA-mediated recruitment of transcriptional activators into targeted genomic loci in a light-dependent manner. First, I engineer a highly efficient Cas9-based transcriptional photoactivation system based on photoactivatable Cas9[9], named photoactivatable Cas9-based activator (paCas9-A). I show that paCas9-A can activate more efficiently the expression of targeted endogenous genes in various human cells than CPTS and LACE. I then investigate whether CPTS and paCas9-A could induce neuronal differentiation of human iPSCs via optical upregulation of endogenous *NEUROD1* gene. CPTS could activate slightly *NEUROD1* but could not induce neuronal differentiation of human iPSCs. In contrast, paCas9-A can upregulate efficiently endogenous *NEUROD1* gene, which represents 860-fold improvement in activation efficiency of *NEUROD1* in human iPSCs compared with CPTS, and therefore induce neuronal differentiation of human iPSCs. Next I find that paCas9-A offers spatial and sustained targeted gene activation. For robust and rapidly reversible gene activation, I further developed another CRISPR-Cas9-based photoactivatable transcription system, named CPTS 2.0. CPTS 2.0 can also activate more efficiently endogenous gene in

human cells than existing systems. In addition, CPTS 2.0 can be rapidly switched off by halting light irradiation and activated repeatedly by resuming light irradiation.

4-2. Materials and Methods

4-2-1. Photoactivatable dCas9 constructions

cDNAs encoding the N- and C- fragments of codon-optimized *Streptococcus pyogenes* Cas9 were amplified from Addgene plasmid 42230. To eliminate nuclease activity of Cas9, D10A mutation in N-fragment of Cas9 and H840A in C-fragment of Cas9 were introduced by Multi Site-Directed Mutagenesis Kit (MBL) according to manufacturer's directions. cDNA encoding pMag, nMagHigh1 and nMag were prepared as previously described[10]. cDNAs encoding VP64 activator domain were amplified from Addgene plasmid 61422. These photoactivatable dimerization domains were amplified by standard PCR using primers that add glycine-serine linker sequences at the 5' and/or 3' ends. The photoactivatable dCas9 constructs based on the N-terminal and C-terminal dCas9 fragments combined with photoactivatable dimerization domains and/or VP64 activator domain were cloned into pcDNA3.1 V5/His-A (Invitrogen) for expression via CMV promoter and pCAGGS vector (RIKEN Bio Resource Center, RDB08938) for expression via CAG promoter.

4-2-2. MS2-p65-HSF1 activator constructions

cDNAs encoding MS2 N55K mutant and p65-HSF1 were amplified from Addgene plasmid 61423. cDNAs encoding MS2 delta FG mutant was amplified from Addgene plasmid 27122. These were amplified by standard PCR using primers that add glycine-serine linker sequences and nuclear localization signals at the 5' and 3' ends.

The MS2-p65-HSF1 activator constructs were cloned into pcDNA3.1 V5/His-A and pCAGGS vector.

4-2-3. sgRNA constructions

I used pSPgRNA vector (Addgene 47108) for expressing sgRNA in mammalian cells. For expressing sgRNA 2.0, sgRNA having MS2-binding sequences, cDNAs encoding sgRNA 2.0 was amplified from sgRNA(MS2) cloning backbone (Addgene 61424) and was cloned into pSPgRNA vector. The sgRNAs targeting *ASCL1*, *IL1R2, NEUROD1, IL1RN, MYOD1, HBG1,* GAL4UAS-mCherry reporter and luciferase reporter were generated by annealed oligo cloning using *Bbs*I site of these vectors.

4-2-4. Reporter constructions

GAL4UAS-mCherry reporter plasmid was constructed by replacing firefly luciferase sequence including protein degradation sequence to mCherry in pGL4.31 vector (Promega) using *Hin*d III and *Xba* I sites. Luciferase reporter was constructed by inserting luciferase sequence and the part of Neurog2 promoter sequence with additional PAM sequence (5'-GTGAATGATGATAATACGATGCGG-3') in pGL4.31 vector using Xho I and XbaI sites.

4-2-5. Cell culture

HEK293T cells (ATCC) were cultured at 37 °C under 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM, Sigma Aldrich) supplemented with 10% FBS

(HyClone), 100 unit/ml penicillin and 100 μg/ml of streptomycin (GIBCO). HeLa cells (ATCC) were cultured at 37 °C under 5% CO_2 in Minimum Crucial Media (MEM, Sigma Aldrich) supplemented with 10% FBS, 100 unit/ml penicillin and 100 μg/ml of streptomycin. Human fetal fibroblasts (RIKEN Bio Resource Center, RCB0698) were cultured at 37 °C under 5% CO_2 in Nutrient Mixture F-12 Ham (Sigma Aldrich) supplemented with 15% FBS, 100 unit/ml penicillin and 100 μg/ml of streptomycin.

4-2-6. Optogenetic gene activation experiments

HEK293T cells were plated at 2.0×10^4 cells/well in 96-well plate (Thermo Scientific), and cultured for 24 h at 37 °C in 5% CO_2. HEK293T cells were transfected with Lipofectamine 3000 (Thermo Scientific) according to the manufacturer's protocols. HeLa cells were plated at 1.0×10^4 cells/well in 96-well plate, and cultured for 24 h at 37 °C in 5% CO_2. HeLa cells were transfected with X-tremeGENE 9 DNA Transfection Reagent (Roche) according to the manufacturer's protocol. Plasmids encoding pCMV-NES-N713d-pMag-NES, pCMV-nMagHigh1-C714d-NLS-VP64, pCMV-NLS-MS2-NLS-p65-HSF1 and sgRNA were transfected at a 1:1:1:1 ratio. Plasmids encoding pCMV-dCas9-CIB1 (or pCMV-CIB1-dCas9-CIB1), pCMV-CRY2-p65 (or pCMV-CRY2FL-VP64) and sgRNA were transfected at a 2:1:1 ratio. Plasmids encoding pCMV-dCas9-VP64 and sgRNA were transfected at a 3:1 ratio. Plasmids encoding pCMV-dCas9-VP64, pCMV-MS2-NLS-p65-HSF1 and sgRNA were transfected at a 2:1:1 ratio. For synergistic activation by multiple sgRNAs, the ratio of four sgRNAs was 1:1:1:1. In multiplexed photoactivation, the ratio of sgRNAs targeting

ASCL1, *MYOD1, IL1R2* and *HBG1* is 1:1:1:1. The total amount of DNA was 0.1 μg/well. For human fetal fibroblasts, 3.0×10^5 cells were nucleofected using the P2 Primary Cell 4D-Nucleofector X Kit S (Lonza) and the DT-130 program. The total amount of DNA was 2.0 μg/sample. Transfected cells were plated at 2.0×10^4 cells/well in 96-well plate, and cultured for 24 h at 37 °C in 5% CO_2. Twenty hours after the transfection, samples were incubated at 37 °C in 5% CO_2 under continuous blue light irradiation or in the dark. Blue light irradiation was performed using a 470 nm ± 20 nm LED light source (CCS Inc.). Intensity of blue light was 1 W/m^2. After 24 h incubation, total RNA were extracted for quantitative PCR analysis.

4-2-7. Quantitative real-time PCR analysis

Total RNA isolation and reverse transcription PCR were performed using Cells-to-Ct kit (Thermo Fisher Scientific) according to the manufacturer's protocols. Real-time PCR was performed by the StepOnePlus system (Thermo Fisher Scientific) using TaqMan Gene Expression Master Mix (Thermo Fisher Scientific). TaqMan probes for detecting each target gene and *GAPDH* as endogenous control were used (Life technologies, TaqMan Gene Expression Assay IDs are following; *ASCL1*: Hs04187546_g1, *MYOD1*: Hs02330075_g1, *IL1RN*: Hs00893626_m1, *IL1R2*: Hs01030384_m1, *NEUROD1*: Hs01922995_s1, *HBG1*: Hs00361131_g1, *GAPDH*: Hs99999905_m1). Standard $\Delta\Delta C_t$ method was used to obtain each relative mRNA level to negative controls transduced with empty vector only in the dark.

4-2-8. Bioluminescence assay

For reporter gene expression assay, HEK293T cells were plated at approximately 2.0×10^4 cells/well in 96-well black-walled plate, and cultured for 24 h at 37 °C in 5% CO_2. The cells were then transfected with Lipofectamine 3000 according to the manufacturer's protocols. Plasmids encoding pCMV-NES-N713d-pMag-NES, pCMV-nMagHigh1-C714d-NLS-VP64, pCMV-NLS-MS2-NLS-p65-HSF1, sgRNA and luciferase reporter were transfected at a 1:1:1:1:1 ratio. Plasmids encoding pCMV-dCas9-VP64, sgRNA and luciferase reporter were transfected at a 3:1:1 ratio. The total amount of DNA was 0.1 μg/well. Twenty hours after the transfection, samples were incubated at 37 °C in 5% CO_2 under continuous blue light irradiation or in the dark. After incubation for 24 h, the culture medium was replaced with 100 μl of phenol red-free DMEM (Sigma Aldrich) containing 500 μM of D-luciferin (Wako Pure Chemical Industries) as a substrate. Bioluminescence measurements were performed using Centro XS^3 LB 960 plate-reading luminometer (Berthold Technologies).

4-2-9. Spatial gene activation experiment

HEK293T cells were plated at 4.0×10^5 cells/dish on 35 mm dish (Iwaki Glass) coated with fibronectin (BD Biosciences), and cultured for 24 h at 37 °C in 5% CO_2. The cells were then transfected with Lipofectamine 3000 according to the manufacturer's protocols. Plasmids encoding NES-N713d-pMag-NES, nMagHigh1-C714d-NLS-VP64, NLS-MS2dFG-NLS-p65-HSF1, mCherry reporter, sgRNAs targeting mCherry reporter and EGFP transfection marker were transfected at a

1:1:1:8:1:8 ratio. The total amount of DNA was 5.0 μg/dish. Twenty hours after the transfection, samples were illuminated by slit-patterned blue light using a photomask for 24 h at 37 °C in 5% CO_2. The width of slit is 2.6 mm. Cells were fixed with 4% paraformaldehyde in PBS for 15 min. Images were acquired using Leica M205 FA stereomicroscope (Leica), and analyzed using Metamorph (Molecular Devices).

4-2-10. hiPSC culture, transfection and optogenetic neural induction

454E2 and 201B7 human iPSC cells were obtained from RIKEN BioResource Center (HPS0077, HPS0063) and maintained on Matrigel (Corning, #354230) coated 6-well culture plates (Thermo Fisher Scientific) in mTeSR1 medium (Stemcell Technologies). 1.0×10^6 cells were nucleofected with pCAG-NES-N713d-pMag-NES, pCAG-nMagHigh1-C714d-NLS-VP64, pCAG-NLS-MS2dFG-NLS-p65-HSF1 and an sgRNA targeting *NEUROD1* using the P3 Primary Cell 4D-Nucleofector X Kit S (Lonza) and the CA-137 program. Transfected cells were plated at 2.5×10^5 cells/well in Matrigel-coated 8-well chamber slide (Thermo Scientific) for immunostaining and 5.0×10^4 cells/well in Matrigel-coated 96-well plate for quantitative real-time PCR with mTeSR1 and 10 μM ROCK inhibitor (WAKO). Six hours after the nucleofection, samples were incubated at 37 °C in 5% CO_2 under continuous blue light irradiation or in the dark. Fresh mTeSR1 medium with 10 μM ROCK inhibitor was added every day. After 4 d incubation, samples were analyzed by immunofluorescence and quantitative real-time PCR.

4-2-11. Immunostaining of light-induced neurons with paCas9-A

The following steps are performed at room temperature. Samples were washed twice with PBS(+), fixed with 4% paraformaldehyde (WAKO) for 10 min and then permeabilized with 0.2% Triton X-100/PBS for 10 min. Samples were washed twice with PBS(+) and then blocked with 3% BSA and 10% FBS for 1 h and stained for 3 h with anti-beta III tubulin eFluor 660 conjugate (eBioscience, catalog no. 5045-10, clone 2G10-TB3) diluted into the blocking reagent at a 1:500 dilution. Samples were washed twice with PBS(+) and then stained with DAPI (Thermo Scientific) for 10 min. Samples were imaged using 20× objective on a Zeiss LSM710 confocal microscope (Carl Zeiss).

4-2-12. Reproducibility and statistics

Pairwise comparisons between samples incubated under blue light and in the dark were made using Welch's two-tailed *t*-test. No sample size estimates were performed, and the sample sizes are consistent with that normally used in the gene regulation experiments. No sample exclusion was carried out. No randomization was used. No blinding was used.

4-3. Results

First, to design a dCas9 optogenetic system for activating targeted endogenous gene by light with high efficiency, I focused on my recently developed photoactivatable Cas9[9] (paCas9), which enable light-controlled Cas9 activity in living cells. In paCas9, N-terminal and C-terminal fragment of Cas9 are combined with photoactivatable dimerization domains named positive Magnet (pMag) and negative Magnet (nMag), respectively[10]. Blue light irradiation induces heterodimerization between pMag and nMag, which enables split Cas9 fragments to reassociate, thereby reconstituting RNA-guided Cas9 nuclease activity. I also showd that catalytically dead paCas9 (padCas9) can bind to targeted DNA, enabling to repress targeted gene expression. However, application of padCas9 for endogenous gene activation remains elusive. Recently, several groups have developed highly efficient dCas9 transcriptional activators, named synergistic activation mediator[11] (SAM), SunTag[12] and VPR[13] respectively. Among these systems, it was shown that SAM could induce high levels of gene expression most consistently[14]. SAM utilizes dCas9-VP64 with a sgRNA bearing MS2 RNA aptamers (known as sgRNA 2.0) to recruit the MS2 coat protein combined with two activators, p65 and heat shock factor1 (HSF1). sgRNA-mediated p65 and HSF1 recruitment dramatically improves the induction potency of targeted endogenous genes by dCas9-VP64. Therefore, I hypothesized that optical control of sgRNA-mediated transcriptional activator recruitment with padCas9 could provide robust Cas9-targeted gene photoactivation.

Toward this end I generated an optogenetic transcriptional activation system consisting of padCas9 combined with VP64, MS2 coat protein combined with p65 and HSF1 activatior domain and sgRNA 2.0 (padCas9-SAM) (**Figure 4-1a**). To investigate

whether this padCas9-SAM architecture is the most effective, I also generated optogenetic transcriptional activation systems combining padCas9 with SunTag (padCas9-SunTag) and VPR (padCas9-VPR) (**Figure 4-1b, c**). I transfected human embryonic kidney (HEK) 293T cells with plasmids encoding these optogenetic systems and an sgRNA targeting the promoter region of human *ASCL1* locus. After 24 h blue light irradiation, I found that padCas9-VPR, padCas9-SunTag and padCas9-SAM showed significant light-induced *ASCL1* activation (**Figure 4-1d**). In particular, padCas9-SAM exhibited the highest light-induced *ASCL1* activation. Therefore I focused on further improvement of this padCas9-SAM.

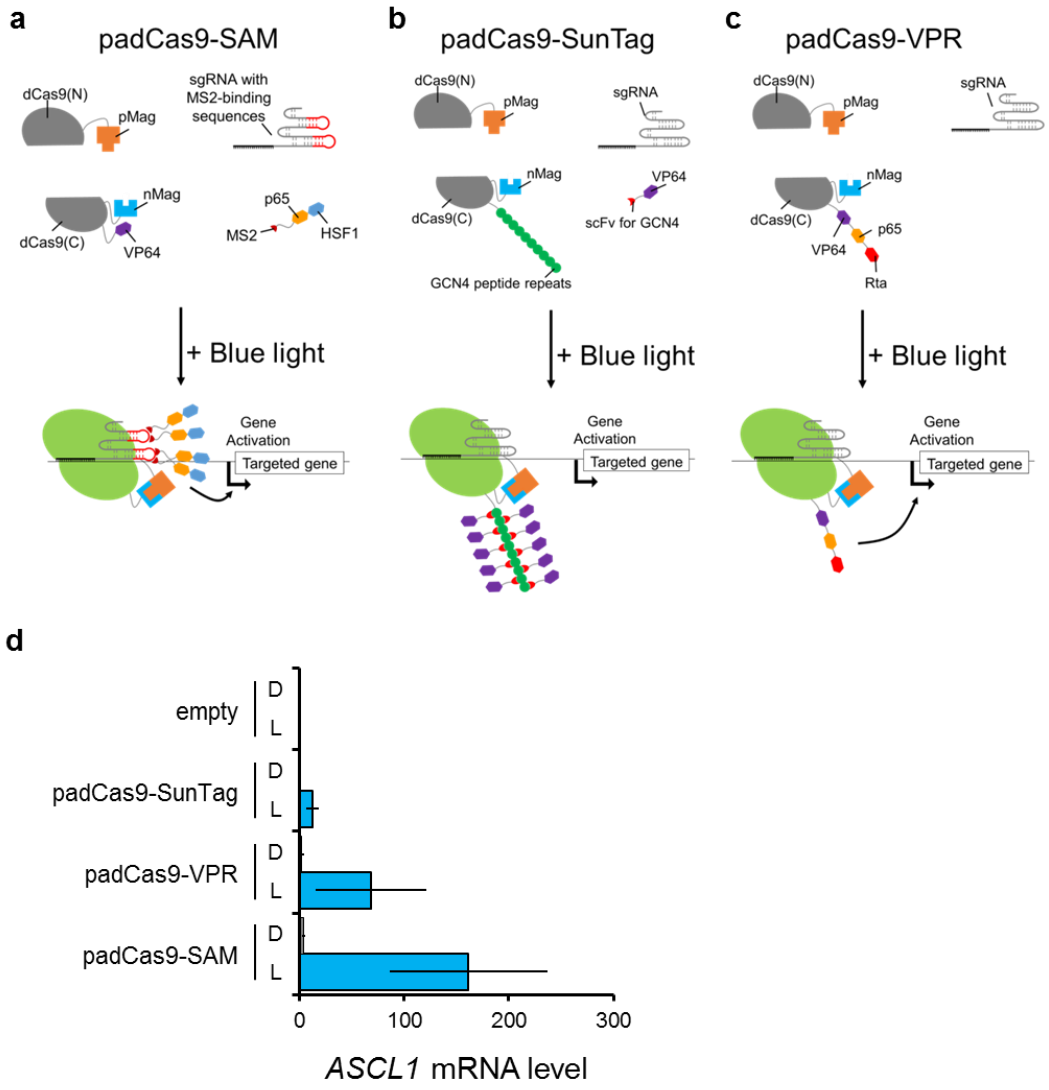

Figure 4-1. Design and characterization of padCas9-SAM, padCas9-SunTag and padCas9-VPR. (**a**) Schematics of padCas9-SAM. padCas9-SAM consists of three chimeric proteins, N713d-pMag, C-terminal fragment of dCas9 combined with VP64 and nMag (nMag-C714d-VP64) and MS2-coat protein combined with p65 and HSF1 activator domains (MS2-p65-HSF1), and an sgRNA having MS2 aptamers in stem loops, known as sgRNA 2.0. MS2 aptamers in sgRNA recruits MS2-p65-HSF1, so this system can simultaneously recruit VP64, p65 and HSF1. (**b**) Schematics of padCas9-Suntag-VP64. padCas9-Suntag-VP64 consists of three chimeric proteins, N-terminal fragment of dCas9 combined with pMag (N713d-pMag), C-terminal fragment of dCas9 combined with nMag and 10 copies of a GCN4 peptide (nMag-C714d-10xGCN4), and single-chain variable fragment nanobody (scFv) for GCN4

combined with VP64 activator domains (scFv-VP64), and an sgRNA. Because scFv-VP64 strongly binds to GCN4 peptides, this system can recruit multiple VP64 into targeted locus by light. (**c**) Schematics of padCas9-VPR. VPR is a potent tripartite activator consisting of VP64, p65AD and the Epstein-Barr virus R transactivator Rta. padCas9-VPR consists of two chimeric proteins, N713d-pMag and C-terminal fragment of dCas9 combined with nMag and VPR (nMag-C714d-VPR), and an sgRNA. (**d**) Light-induced endogenous ASCL1 activation in HEK293T cells with padCas9-Suntag-VP64, padCas9-VPR and padCas9-SAM with a single sgRNA targeting the promoter region of *ASCL1*. Data are expressed as relative mRNA amount to the negative control transfected with empty vector in the dark and represented as mean ± s.e.m. (n=3 from three individual experiments).

padCas9-SAM consists of three chimeric proteins, N-terminal fragment of dCas9 combined with pMag and a nuclear localization signal (NLS) (NLS-N713d-pMag), C-terminal fragment of dCas9 combined with VP64, nMag and NLS (nMag-C714d-NLS-VP64) and MS2-coat protein combined with p65 and HSF1 activator domains and NLS (MS2-NLS-p65-HSF1), and an sgRNA 2.0. For further improving light-induction efficiency of padCas9-SAM, I replaced nMag photoactivatable dimerization domain in nMag-C714d-NLS-VP64 with nMagHigh1 (nMagHigh1-C714d-NLS-VP64) for enhancing dimerization efficiency[10]. In parallel, I modified the subcellular localization of NLS-N713d-pMag for suppressing background activity, replacing the nuclear localization signal in NLS-N713d-pMag with the two nuclear export signals (NES) (NES-N713d-pMag-NES)[15]. The resultant combination of NES-N713d-pMag-NES and nMagHigh1-C714d-NLS-VP64 with MS2-NLS-p65-HSF1 showed higher gene activation efficiency than the original padCas9-SAM (**Figure 4-2a**). Next, I tested several MS2-effector units that have different designs. Surprisingly, I found that MS2-p65-HSF1 activator containing non-aggregating delta FG mutant of MS2[16] and two copies of NLS (NLS-MS2dFG-NLS-p65-HSF1) can provide significantly high *ASCL1* activation (**Figure 4-2b**). Therefore, I used NES-N713d-pMag-NES, nMagHigh1-C714d-NLS-VP64 and NLS-MS2dFG-NLS-p65-HSF1 in following studies and referred to this as paCas9-A.

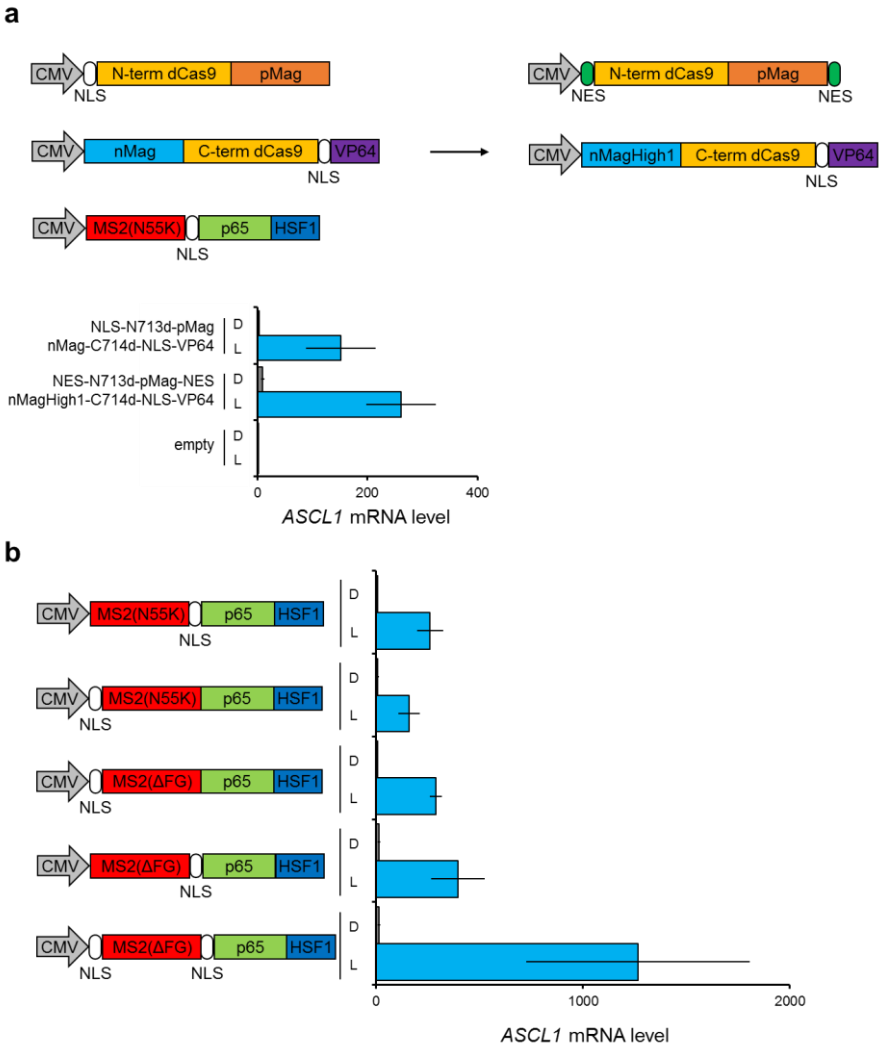

Figure 4-2. Optimization of padCas9-SAM. (**a**) Optimizing subcellular localization of padCas9 fragments and replacing nMag photoactivatable dimerization domain. The nuclear localization signal (NLS) of N713d-pMag was removed and two copies of nuclear export signal (NES) were combined with N713d-pMag. I replaced nMag of nMag-C714d-VP64 into nMagHigh1 for enhancing dimerization efficiency. (**b**) Screening of MS2-p65-HSF1 units that have different designs. About MS2-coat protein, I tested MS2(N55K) mutant, which is originally used in SAM, and non-aggregating MS2 deltaFG mutant. I also tested whether the number and position of NLS effects induction potency. Data are expressed as relative mRNA amount to the negative control transfected with empty vector in the dark and represented as mean ± s.e.m. (n=4 from two individual experiments with biological duplicates).

For initial characterization of paCas9-A, I tested whether paCas9-A could activate endogenous human *ASCL1* gene more potently than two previously reported CRISPR-Cas9-based optogenetic transcription systems[6,7], named CPTS and LACE, respectively (**Figure 4-3a**). HEK293T cells transfected with CPTS and LACE targeting *ASCL1* with a single sgRNA showed 22-fold and 9-fold light-dependent upregulation of *ASCL1* mRNA compared with cells transfected with empty vector, respectively. In contrast, paCas9-A with an sgRNA targeting *ASCL1* showed 1200-fold upregulation of *ASCL1* mRNA. Note that the induction efficiency of paCas9-A is comparable to that of SAM. I also tested additional sgRNAs targeting different sites in the promoter region of *ASCL1* (**Figure 4-3b, c**). I found that all six tested sgRNAs with paCas9-A showed higher *ASCL1* activation than those with CPTS and observed the maximum *ASCL1* activation by paCas9-A when sgRNAs targeting the sequences at -250 to -50 bp upstream from the transcription start site were used.

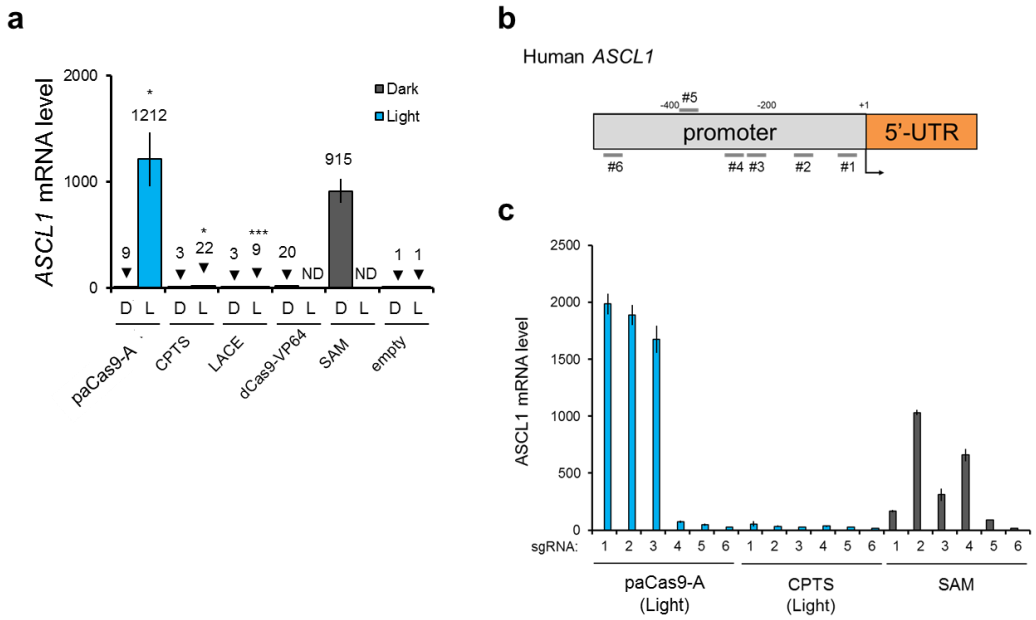

Figure 4-3. (**a**) Light-induced *ASCL1*, upregulation in HEK293T cells using paCas9-A, CPTS, LACE, dCas9-VP64 and SAM with a single sgRNA targeting each gene. (**b**) Maps showing the sgRNA targeting sites in the human *ASCL1* loci. (**c**) *ASCL1* activation by paCas9-A, CPTS and SAM. Used *ASCL1*-targeting sgRNA are indicated. Data are shown as the mean ± s.e.m. (n = 2 from biological duplicates).

To explore the generalizability of robust optogenetic endogenous gene activation with paCas9-A, I also tested whether paCas9-A can efficiently upregulate human *IL1R2*, *NEUROD1*, *IL1RN* and *MYOD1* genes in light-dependent manner (**Figure 4-4**). In all these loci, although the level of transcriptional activation by CPTS and LACE is severely low or ineffective, paCas9-A can activate these endogenous genes potently with a single sgRNA. I also confirmed that paCas9-A with an sgRNA can activate robustly endogenous *ASCL1* gene in HeLa cells and primary human fibroblasts, which previously reported Cas9-based photoactivatable transcription systems could not (**Figure 4-5**). These results show that paCas9-A can activate more potently targeted

endogenous gene with a single sgRNA than previously reported CRISPR-Cas9-based optogenetic transcription systems.

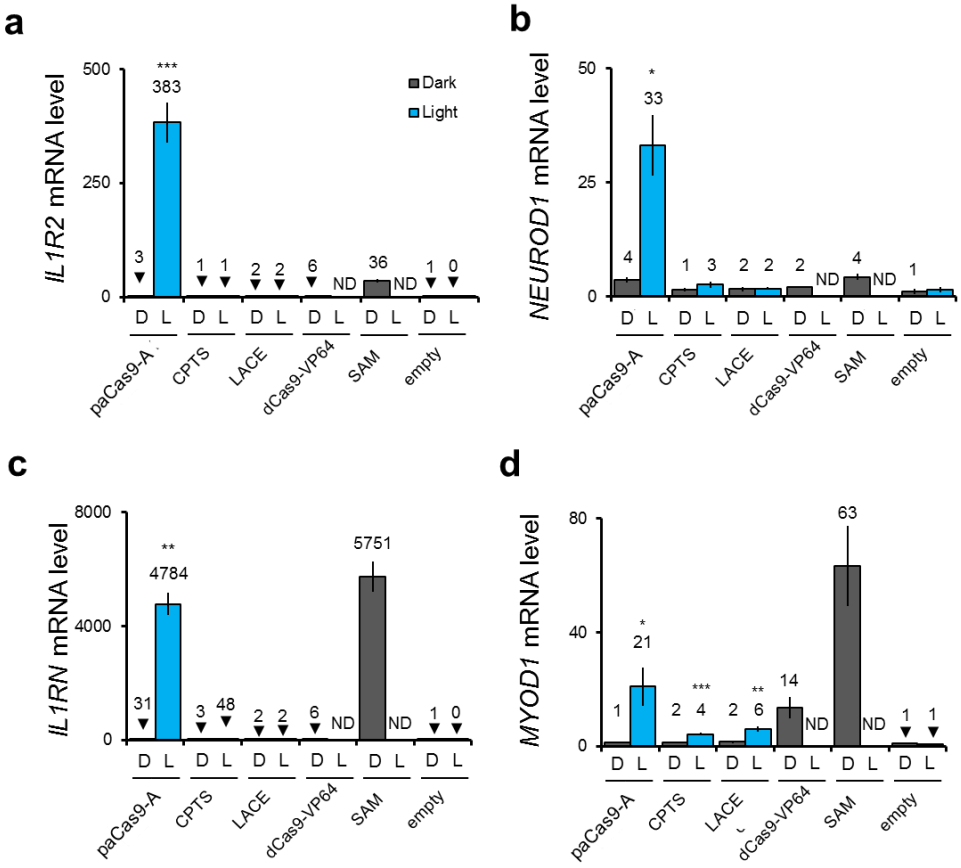

Figure 4-4. Light-induced *IL1R2* (**a**), *NEUROD1* (**b**), *IL1RN* (**c**) and *MYOD1* (**d**) upregulation in HEK293T cells using paCas9-A, CPTS, LACE, dCas9-VP64 and SAM with a single sgRNA targeting each gene. Data are expressed as relative mRNA amount to the negative control transfected with empty vector in the dark and represented as mean ± s.e.m. (n=4 from two individual experiments with biological duplicates) Welch's two-tailed *t* test was performed. *p < 0.05, **p < 0.01, ***p < 0.005 versus the sample in the dark.

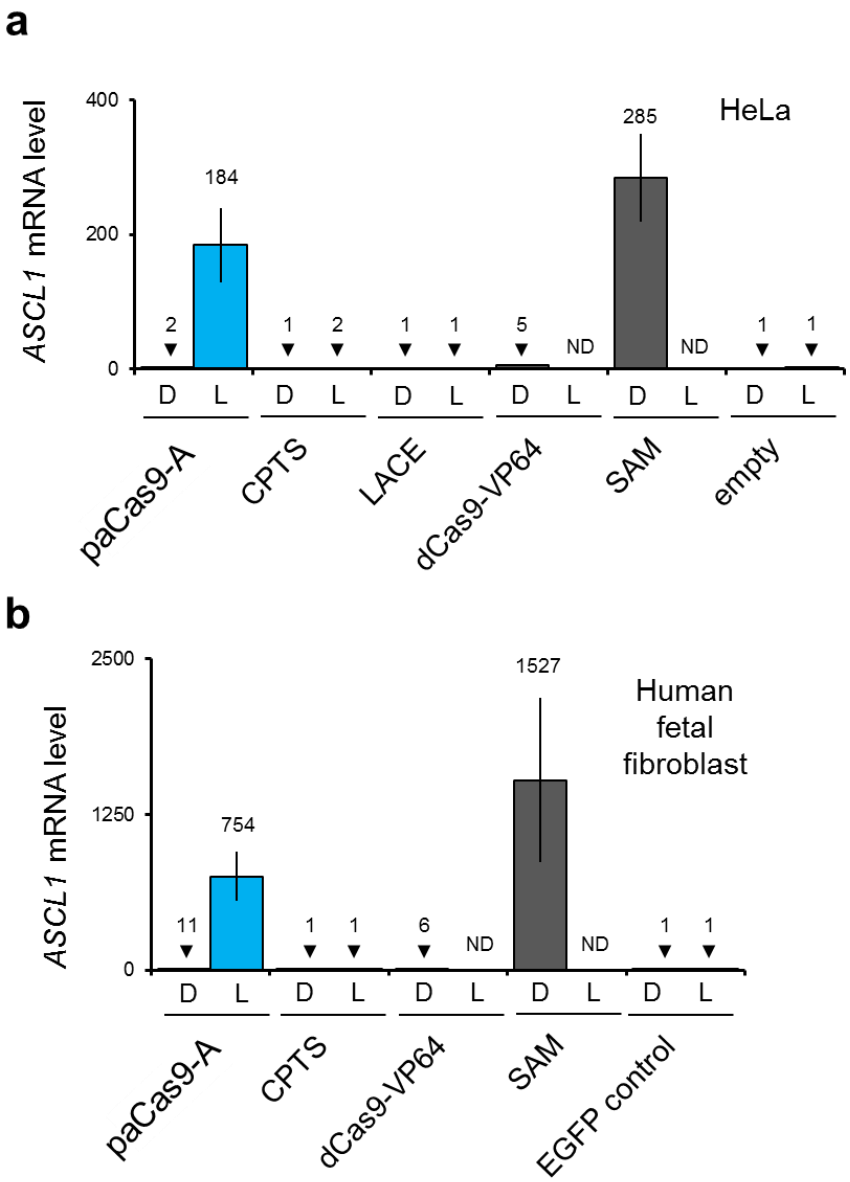

Figure 4-5. Light-induced *ASCL1* upregulation in HeLa cells (**a**) and human fetal fibroblasts (**b**) using paCas9-A, CPTS, LACE, dCas9-VP64 and SAM with a single sgRNA targeting each gene. Data are expressed as relative mRNA amount to the negative control transfected with empty vector in the dark and represented as mean ± s.e.m. (n=2 from individual experiments with biological duplicates)

The level of gene activation by previous Cas9-based photoactivatable transcription systems can be synergistically enhanced by using multiple sgRNAs targeting different sites in the promoter region of the same locus[4–7]. I then compared the induction efficiency of paCas9-A and CPTS with a single sgRNA or four sgRNAs targeting human *MYOD1* locus (**Figure 4-6a**). As previously shown, the activation efficiency of CPTS with four sgRNAs is higher than that of CPTS with a single sgRNA. Nevertheless, I found that paCas9-A with a single sgRNA targeting *MYOD1* can activate *MYOD1* more efficiently than CPTS with four sgRNAs. In addition, simultaneous four *MYOD1* sgRNAs transfection can further enhance the induction efficiency of paCas9-A. A similar result was observed in endogenous *IL1RN* gene activation (**Figure 4-6b**). These data indicate that the activation efficiency of paCas9-A with a single sgRNA is higher than that of CPTS with four sgRNAs and can be further enhanced using multiple sgRNAs. I next tested whether paCas9-A could activate efficiently multiple target genes (**Figure 4-6c**). Using four different sgRNAs targeting *ASCL1*, *MYOD1*, *IL1R2* and *HBG1* simultaneously, paCas9-A activated more potently these genes than CPTS in response to light. This result shows that paCas9-A can be deployed for robust and multiplexed activation of targeted endogenous genes in light-dependent manner.

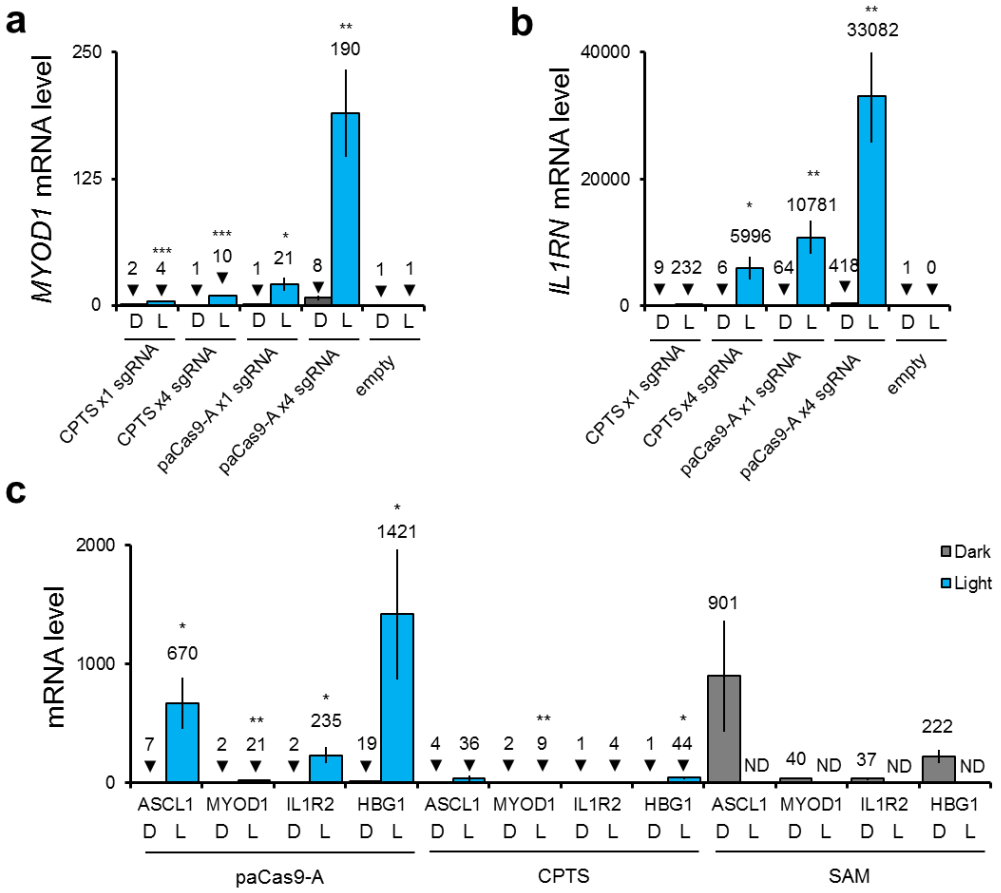

Figure 4-6. (**a,b**) Light-induced *MYOD1* (**a**) and *IL1RN* (**b**) upregulation in HEK293T cells using paCas9-A and CPTS with quadruple sgRNAs targeting *MYOD1* and *IL1RN*, respectively. (**c**) Simultaneous photoactivation of endogenous genes in HEK293T cells using paCas9-A, CPTS and SAM with a mixture of four sgRNAs, each targeting *ASCL1*, *MYOD1*, *IL1R2* and *HBG1*. Data are expressed as relative mRNA amount to the negative control transfected with empty vector in the dark and represented as mean ± s.e.m. (n=6 from three individual experiments with biological duplicates (**a–b**), n=8 from four individual experiments with biological duplicates (**c**)).

To investigate whether the DNA targeting specificity of paCas9-A is comparable to that of conventional dCas9-VP64 activator, I generated a series of sgRNAs for luciferase reporter harboring single-nucleotide and adjacent double-nucleotides Watson-Crick transversion mutations (**Figure 4-7**). By luciferase reporter activation assay with a series of mutated sgRNAs, I confirmed that induced luciferase activities by paCas9-A and dCas9-VP64 with mutated sgRNAs were lower than that with the perfectly matched sgRNA. I found no significant difference in the normalized reporter activities induced by paCas9-A and dCas9-VP64, suggesting that the target specificities of paCas9-A are indistinguishable from those of dCas9-VP64.

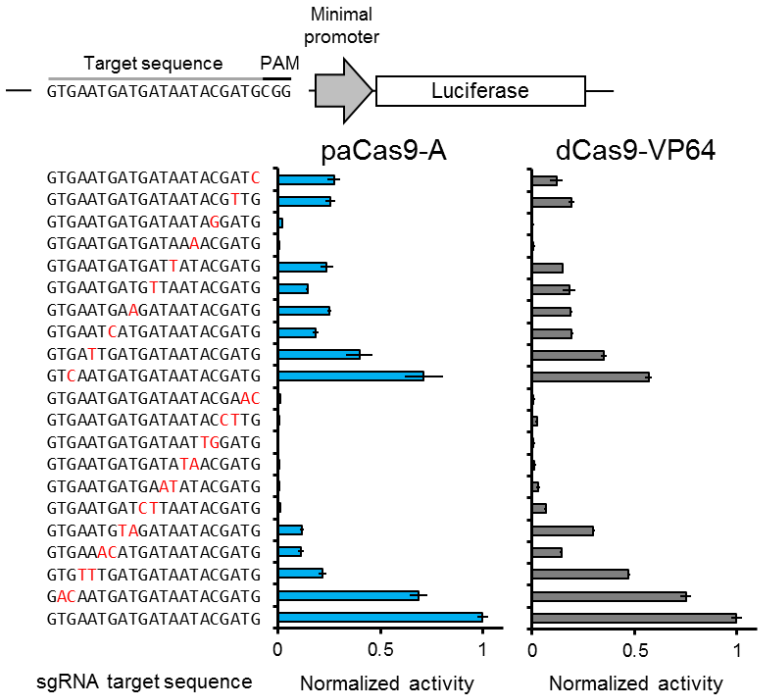

Figure 4-7. DNA targeting specificity analysis of padCas9-SAM and dCas9-VP64 using luciferase reporter plasmid activation assay. Luciferase reporter is constructed of Cas9-binding sites upstream and luciferase gene downstream of minimal CMV promoter. The positions of mutations in each sgRNA are indicated. Values are normalized to positive control with perfectly matched sgRNAs. Data are shown as the mean ± s.e.m. (n=6 from two individual experiments with separately seeded biological triplicates).

Although several optogenetic systems for targeted endogenous gene activation have been developed[6,7,17], optical regulation of cellular phenotype by these endogenous gene photoactivation systems has not been showd. I then investigated whether CPTS and paCas9-A can change cellular phenotype by robust endogenous gene activation with light. It was previously shown that *NEUROD1* upregulation induced neuronal differentiation of human iPSCs[13,18]. In this experiment, I tested whether paCas9-A with only a single sgRNA targeting *NEUROD1* could induce optogenetically neuronal differentiation of human iPSCs. iPSCs were transfected with paCas9-A or CPTS and sgRNAs targeting *NEUROD1*. After 1 d blue light irradiation, CPTS with four sgRNAs showed 29-fold upregulation of *NEUROD1* mRNA level in human iPSCs (**Figure 4-8a**). In contrast, paCas9-A with a single sgRNA targeting *NEUROD1* showed 25000-fold upregulation of *NEUROD1* mRNA, representing 860-fold improvement in *NEUROD1* activation efficiency in human iPSCs compared with CPTS. After 4 d blue light irradiation, while cells expressing CPTS did not show *NEUROD1* upregulation, cells transfected with paCas9-A and a *NEUROD1* sgRNA showed significant 3.3-fold *NEUROD1* upregulation (**Figure 4-8b**).

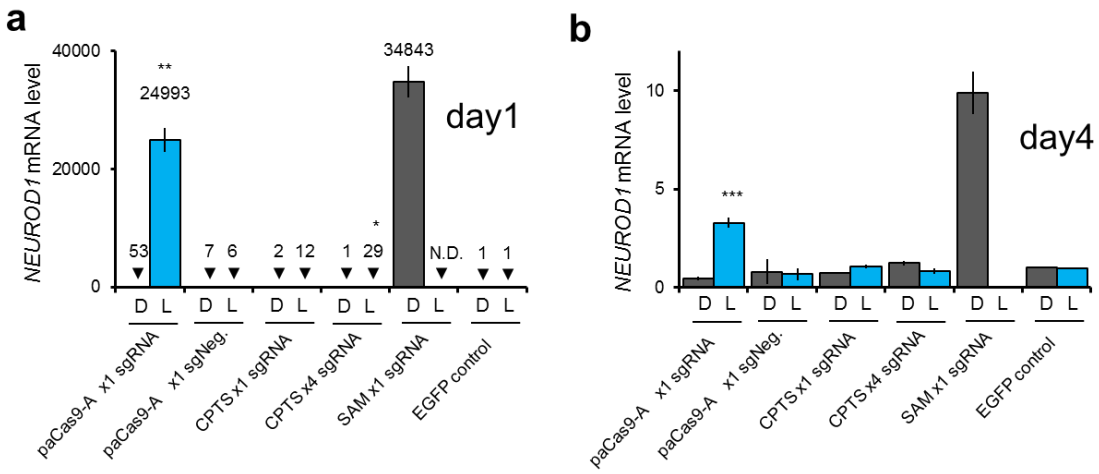

Figure 4-8. Light-induced *NEUROD1* upregulation in human iPSCs using indicated Cas9-activators with indicated sgRNAs at day 1 (**a**) and day 4 (**b**). Data are expressed as relative mRNA amount to the negative control transfected with EGFP vector in the dark and represented as mean ± s.e.m. (n=3 from one of two representative experiments). Welch's two-tailed *t* test was performed. *p < 0.05, ** < 0.01, ***p < 0.005 versus the sample in the dark.

I then performed immunostaining analysis at day 4 and found that light-stimulated iPSCs expressing paCas9-A and an sgRNA targeting *NEUROD1* stained positively for the neuronal markers beta III tubulin (Tuj1) (**Figure 4-9**). The dark-incubated iPSCs transfected with paCas9-A and an sgRNA targeting *NEUROD1* did not express beta III tubulin. In contrast to paCas9-A, CPTS with *NEUROD1*-targeting sgRNAs could not induce neural differentiation of human iPSCs. These results show that CPTS could not induce sufficient *NEUROD1* upregulation for neuronal conversion of human iPSCs, but paCas9-A can efficiently activate *NEUROD1* and successfully induce neural differentiation of iPSCs in light-dependent manner.

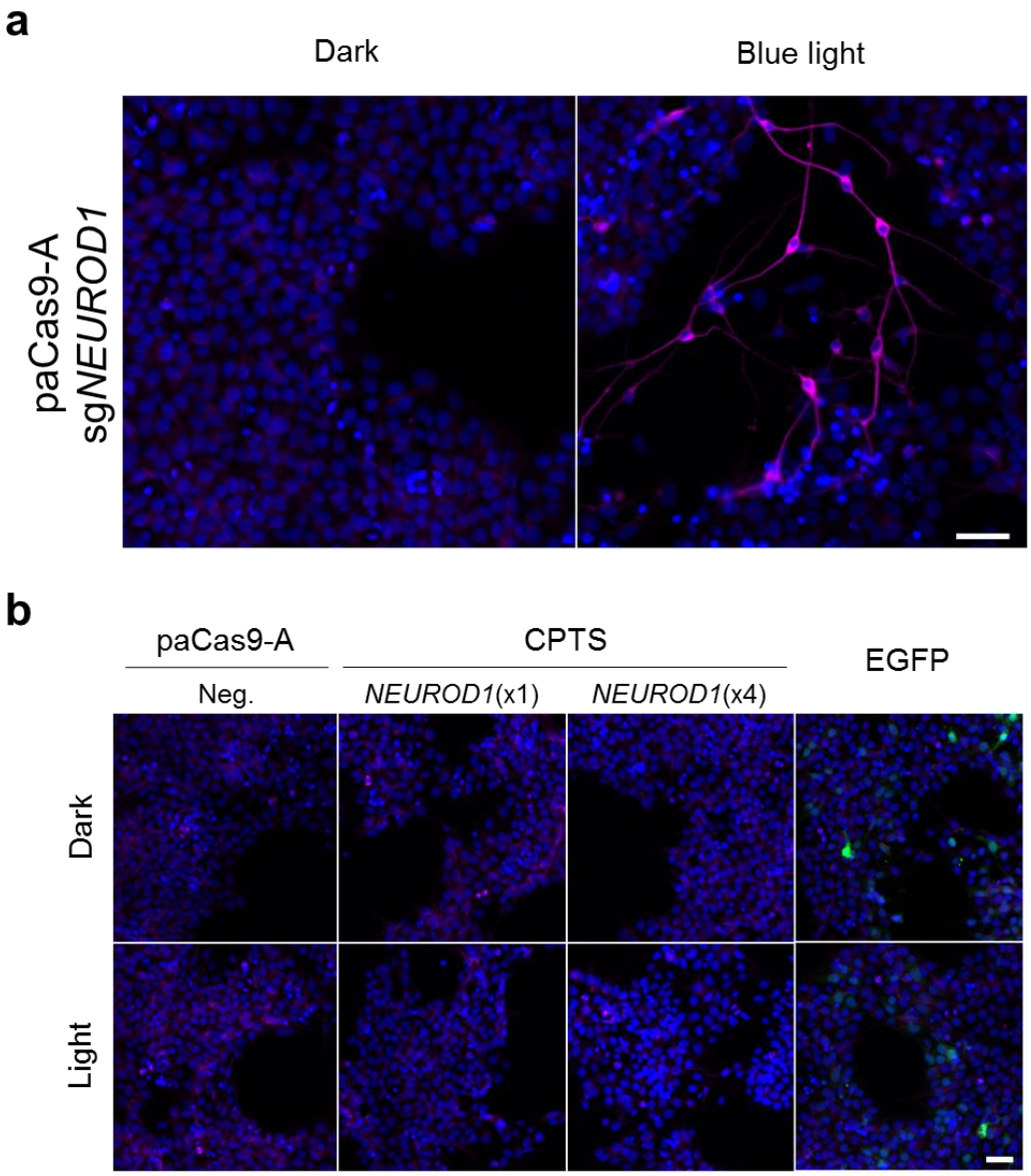

Figure 4-9. Optogenetic control of iPSC neuronal differentiation with paCas9-A. (**a**,**b**) Immunofluorescence images of human iPSCs for beta III tubulin (magenta) and DAPI (blue) after 96 h blue light irradiation or incubation in the dark. Transfected plasmids are indicated. Images are representative from three individual experiments. Scale bar = 50 μm.

I tested whether paCas9-A can enable light-induced spatial gene activation (**Figure 4-10**). To visualize paCas9-A-mediated gene activation in living cells, I used mCherry reporter. HEK293T cells transfected with paCas9-A targeting mCherry reporter were irradiated with slit-patterned blue light. After 24 h irradiation, the slit pattern of mCherry-expressing cells according to the irradiation pattern was observed, demonstrating that paCas9-A can spatially control gene expression.

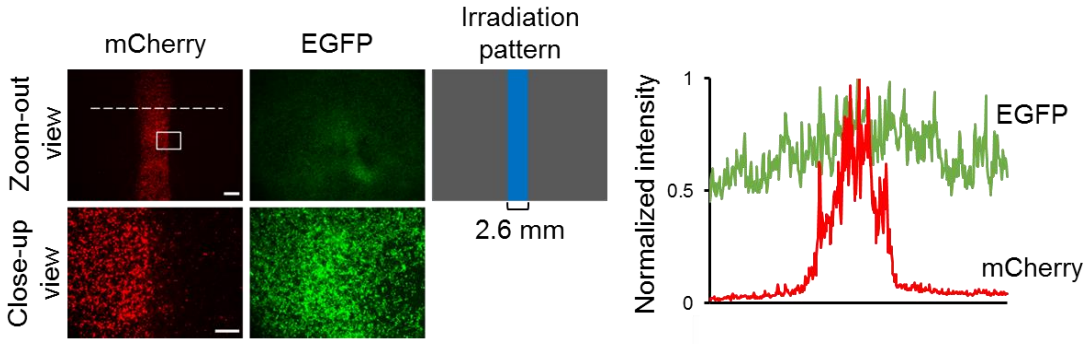

Figure 4-10. Spatial gene activation by paCas9-A. Slit-patterned mCherry expression in HEK293T cells illuminated by blue light with a spatial pattern using a photomask. The width of slit is 2.6 mm. Scale bar, 2 mm (zoom-out view) and 0.5 mm (close-up view), respectively.

Next I investigated the kinetics of light-induced gene activation by paCas9-A. HEK293T cells transfected with paCas9-A and a single sgRNA targeting *ASCL1* showed a significant 45-fold upregulation of *ASCL1* mRNA levels by only 3 h of blue light irradiation. The level of *ASCL1* mRNA increased as the blue light stimulation time increased (**Figure 4-11a**). I also tested whether paCas9-A-mediated gene activation could be switched off by halting light irradiation (**Figure 4-11b**). After I turned off the blue light post 12 h light irradiation, cells showed sustained 400-fold *ASCL1* expression for about 12 h. After 24 h incubation in the dark, *ASCL1* mRNA decreased to 160-fold, but further 12 h dark incubation did not decrease the level of *ASCL1* mRNA. This result

shows that paCas9-A provide sustained endogenous gene activation for several hours after extinguishing blue light.

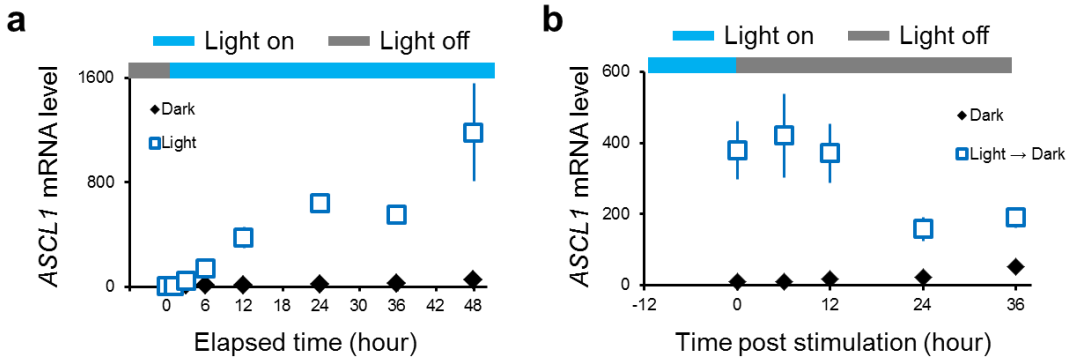

Figure 4-11. (**a**) The time course of light-induced *ASCL1* transcription by paCas9-A. (**b**) Decay kinetics of light-induced *ASCL1* transcription by paCas9-A after removal of blue light irradiation. Data are expressed as relative mRNA amount to the negative control transfected with empty vector in the dark and represented as mean ± s.e.m. (n=4 from two individual experiments with biological duplicates). Welch's two-tailed *t* test was performed. *p < 0.05, ** < 0.01, ***p < 0.005 versus the sample in the dark.

In addition to sustained gene activation, rapidly reversible gene regulation with high induction efficiency is also important, considering the fact that some biological processes are based on temporal gene expression pattern. For robust and reversible Cas9-targeted gene regulation, I developed another optogenetic Cas9-based transcription system, named CPTS 2.0 (**Figure 4-12a**). CPTS 2.0 consists of three chimeric proteins, dCas9 combined with two copies of NLS, MS2-coat protein combined with CIB1 and three tandem NLS, CRY2 combined with p65 and HSF1 activator domains, and an sgRNA 2.0. First, I tested whether CPTS 2.0 could activate efficiently endogenous *ASCL1* gene (**Figure 4-12b**). Cells transfected with CPTS 2.0 and a single sgRNA targeting *ASCL1* showed 200-fold upregulation of *ASCL1* mRNA.

Using four sgRNAs targeting *ASCL1*, the induced *ASCL1* mRNA level by CPTS 2.0 was enhanced to 700-fold. From this result, the magnitude of transcriptional upregulation by CPTS 2.0 is significantly higher than that by CPTS. A similar result was observed in endogenous *MYOD1* gene activation (**Figure 4-12c**).

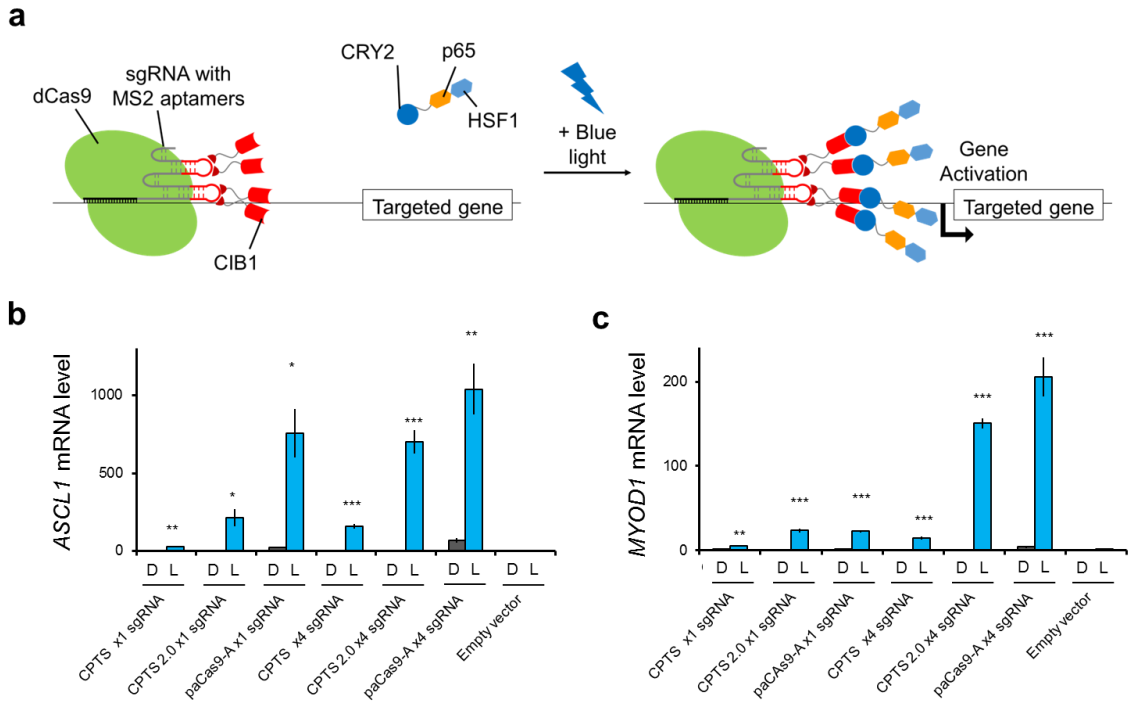

Figure 4-12. (**a**) Schematic of the CPTS 2.0. (**b**,**c**) Light-induced *ASCL1* (**b**) and *MYOD1* (**c**) upregulation in HEK293T cells indicated Cas9-activators with indicated sgRNAs. Data are expressed as relative mRNA amount to the negative control transfected with empty vector in the dark and represented as mean ± s.e.m. (n=4 from two individual experiments with biological duplicates). Welch's two-tailed *t* test was performed. *p < 0.05, ** < 0.01, ***p < 0.005 versus the sample in the dark.

I then investigated the kinetics of light-induced *ASCL1* upregulation by CPTS 2.0. HEK293T cells transfected with CPTS 2.0 and four sgRNAs targeting *ASCL1* showed a significant 79-fold upregulation of *ASCL1* mRNA levels by 3 h of blue light irradiation

(**Figure 4-13a**). Subsequently, the level of *ASCL1* mRNA increases as the blue light stimulation time increased. Next I tested whether CPTS 2.0 can rapidly switch off *ASCL1* gene activation (**Figure 4-13b**). After I turned off the blue light post 12 h light irradiation, the level of *ASCL1* mRNA rapidly decreases from 164-fold to 12-fold for 6 h. At 24 h post light irradiation, *ASCL1* mRNA returns to baseline level. I further tested whether CPTS 2.0 can activate repeatedly endogenous *ASCL1* gene (**Figure 4-13c**). Second light irradiation increased *ASCL1* mRNA levels to 176-fold over control within 12 h. These results showed that CPTS 2.0 can offer rapid, reversible and repeatable endogenous gene activation with high efficiency.

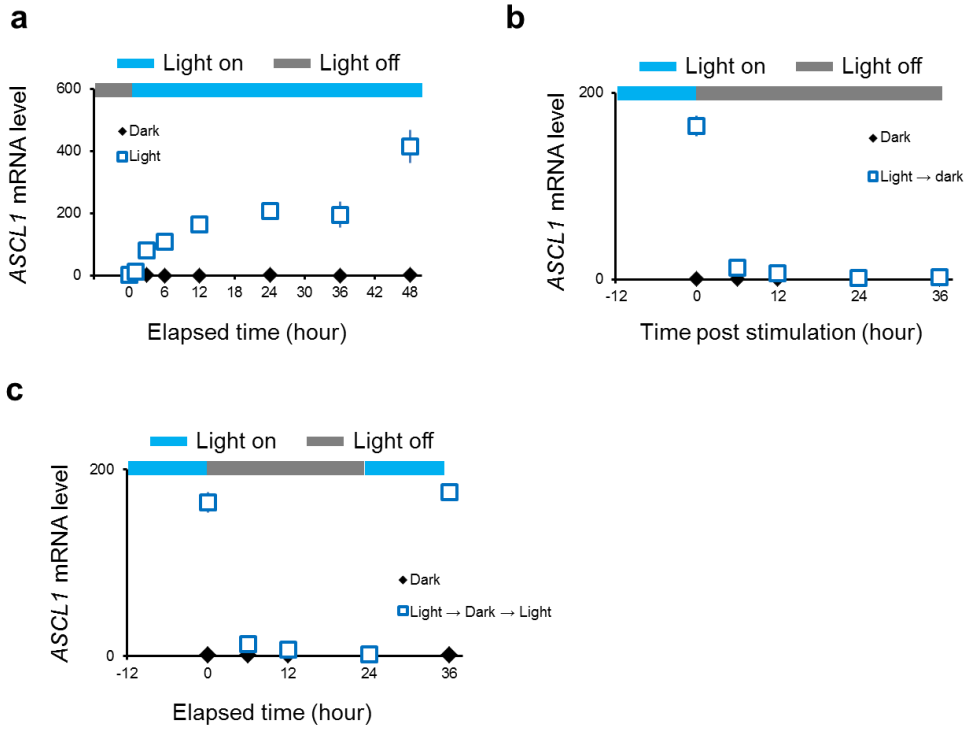

Figure 4-13. (**a**) The time course of light-induced *ASCL1* transcription by CPTS 2.0. (**b**) Decay kinetics of light-induced *ASCL1* transcription by CPTS 2.0 after removal of blue light irradiation. (**c**) Reversible and repeatable activation of endogenous *ASCL1* with CPTS 2.0. Data are expressed as relative mRNA amount to the negative control transfected with empty vector in the dark and represented as mean ± s.e.m. (n=4 from two individual experiments with biological duplicates). Welch's two-tailed *t* test was performed. *p < 0.05, ** < 0.01, ***p < 0.005 versus the sample in the dark.

4-4. Discussion

In conclusion, I have developed highly efficient CRISPR-Cas9-based photoactivatable transcription system, named paCas9-A. Unlike previously developed dCas9-based photoactivatable transcription systems, named CPTS and LACE, the present paCas9-A can offer potent endogenous gene activation with spatiotemporal resolution. As I showed in this study, CPTS with four sgRNAs showed only weak activation of *NEUROD1* in human iPSCs and could not induce neuronal differentiation. In contrast, compared with CPTS with four *NEUROD1* sgRNAs, paCas9-A with a single sgRNA targeting *NEUROD1* showed 860-fold improvement in activation of *NEUROD1* in human iPSCs and could induce efficiently neuronal differentiation (**Fig. 2a**). This result indicates that the magnitude of gene expression induced by an engineered Cas9-based transcription system is crucial to successfully control cellular function. In addition, I have also developed another efficient CRISPR-Cas9-based photoactivatable transcription system, named CPTS 2.0. CPTS 2.0 can provide efficient and rapidly reversible endogenous gene regulation, while paCas9-A can offer more efficient and sustained activation. My study shows that paCas9-A and CPTS 2.0 represent potent systems to manipulate various biological phenomena via highly efficient endogenous gene activation with high spatiotemporal resolution, and will be useful for investigations of gene function in various biological phenomena, biotechnological applications and future biomedical applications.

The modular design of paCas9-A and CPTS 2.0 would provide system extensibility. By replacing transcriptional activator domains into epigenetic modifiers[19–21], optogenetic control of epigenetic states in targeted genomic regions will be available. Furthermore, using chemically-inducible dimerization domains instead of

photoactivatable dimerization domains[22,23], these systems would be regulated with chemicals. The present platforms will offer simple and versatile methods for precise and robust endogenous genome manipulation in basic biological research and biotechnology application.

4-5. References

1. Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* **17,** 5–15 (2015).
2. Gilbert, L. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442–51 (2013).
3. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152,** 1173–1183 (2013).
4. Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10,** 977–979 (2013).
5. Perez-pinera, P. *et al.* RNA-guided gene activation by CRISPR- Cas9 – based transcription factors. *Nat. Methods* **10,** 973–976 (2013).
6. Chapter 3 in this thesis
7. Polstein, L. R. & Gersbach, C. A. A photoactivatable CRISPR-Cas9 system for control of endogenous gene activation. *Nat. Chem. Biol.* **11**, 198–200 (2015).
8. Kennedy, M. J. *et al.* Rapid blue-light–mediated induction of protein interactions in living cells. *Nat. methods* **7**, 973–975 (2010).
9. Chapter 2 in this thesis
10. Kawano, F., Suzuki, H., Furuya, A. & Sato, M. Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nat. Commun.* **6**, 6256 (2015).

11. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2014).

12. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell* **159**, 635–646 (2014).

13. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).

14. Chavez, A. *et al.* Comparison of Cas9 activators in multiple species. *Nat. Methods* **13**, 563–567 (2016).

15. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33,** 139–142 (2015).

16. Keryer-Bibens, C., Barreau, C. & Osborne, H. B. Tethering of proteins to RNAs by bacteriophage proteins. *Biol. Cell* **100,** 125–38 (2008).

17. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500,** 472–476 (2013).

18. Zhang, Y. *et al.* Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78,** 785–798 (2013).

19. Vojta, A. *et al.* Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res.* **44**, 5615–5628 (2016).

20. Liu, X. S. *et al.* Editing DNA Methylation in the Mammalian Genome. *Cell* **167,** 233–247.e17 (2016).

21. Amabile, A. *et al.* Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. *Cell* **167,** 219–232.e14 (2016).

22. Miyamoto, T. *et al.* Rapid and orthogonal logic gating with a gibberellin-induced dimerization system. *Nat. Chem. Biol.* **8,** 465–70 (2012).

23. DeRose, R., Miyamoto, T. & Inoue, T. Manipulating signaling at will: chemically-inducible dimerization (CID) techniques resolve problems in cell biology. *Pflugers Arch.* **465,** 409–17 (2013).

Chapter 5.

General conclusion

In the present study, I developed photoactivatable CRISPR-Cas9 systems and showd spatiotemporal control of Cas9-based genome manipulation in living cells.

In Chapter 2, I engineered photoactivatable Cas9, paCas9, which enables optogenetic control of Cas9 nuclease activity. paCas9 can optically control targeted genome editing via NHEJ and HDR pathways. I showd that paCas9 can be spatially activated by patterned light irradiation, and switched off by stopping illumination. The present paCas9 offers an optogenetic method for targeted genome editing with high spatial and temporal resolution, which will facilitate understanding complex gene networks and has potential in future biomedical applications.

In Chapter 3, I have developed CRISPR-Cas9-based photoactivatable transcription system, which can activate targeted endogenous gene by light. I showd that this system offers spatiotemporal control of endogenous gene expression in mammalian cells.

In Chapter 4, I described an improved optogenetic CRISPR-Cas9-based transcription system, named paCas9-A. paCas9-A with a single sgRNA could optically activate far more efficiently targeted endogenous genes in human cells than previously reported Cas9-based photoactivatable transcription systems. I show that previously reported Cas9-based photoactivatable transcription system could activate slightly *NEUROD1* gene but could not induce neuronal differentiation of human iPSCs. In contrast, paCas9-A represents an 860-fold improvement in activation of *NEUROD1* gene and induces neuronal differentiation of human iPSCs.

The present systems offer simple and versatile methods for optogenetic manipulation of genome, and will be useful for investigations of gene function in various biological phenomena, biotechnological applications and future biomedical applications.