Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology) 2022 23(5):382-391
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Research Article

https://doi.org/10.1631/jzus.B2100710

Can SpRY recognize any PAM in human cells?

Jinbin YE[1*], Haitao XI[1*], Yilu CHEN[1], Qishu CHEN[1], Xiaosheng LU[1], Jineng LV[2], Yamin CHEN[2], Feng GU[2✉], Junzhao ZHAO[1✉]

[1]*Reproduction Center, Department of Obstetrics and Gynecology, the Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou 325000, China*
[2]*School of Ophthalmology and Optometry, Eye Hospital, Wenzhou Medical University, State Key Laboratory and Key Laboratory of Vision Science, Ministry of Health and Zhejiang Provincial Key Laboratory of Ophthalmology and Optometry, Wenzhou 325000, China*

Abstract: The application of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated proteins (Cas) can be limited due to a lack of compatible protospacer adjacent motif (PAM) sequences in the DNA regions of interest. Recently, SpRY, a variant of *Streptococcus pyogenes* Cas9 (SpCas9), was reported, which nearly completely fulfils the PAM requirement. Meanwhile, PAMs for SpRY have not been well addressed. In our previous study, we developed the PAM Definition by Observable Sequence Excision (PAM-DOSE) and green fluorescent protein (GFP)-reporter systems to study PAMs in human cells. Herein, we endeavored to identify the PAMs of SpRY with these two methods. The results indicated that 5'-NRN-3', 5'-NTA-3', and 5'-NCK-3' could be considered as canonical PAMs. 5'-NCA-3' and 5'-NTK-3' may serve as non-priority PAMs. At the same time, PAM of 5'-NYC-3' is not recommended for human cells. These findings provide further insights into the application of SpRY for human genome editing.

Key words: CRISPR/Cas; SpRY; Protospacer adjacent motif (PAM); Recognize

1 Introduction

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein 9 (Cas9) nucleases trigger targeted gene editing in a wide variety of organisms and living cells, and provide a potential avenue for improved therapy in genetic diseases (Hsu et al., 2014; Jiang and Doudna, 2017; Komor et al., 2017). For target site recognition, Cas9 is programmed by a chimeric single-guide RNA (sgRNA) that encodes a sequence complementary to a target protospacer (Jinek et al., 2012), and the recognition of a short neighboring protospacer adjacent motif (PAM) is also required (Mojica et al., 2009; Jinek et al., 2012; Shah et al., 2013; Sternberg et al., 2014). The application of CRISPR/Cas9 can be limited due to a lack of compatible PAM sequences in the DNA region of interest. The most commonly used CRISPR nuclease *Streptococcus pyogenes* Cas9 (SpCas9) strictly recognizes a 5'-NGG-3' PAM (Hsu et al., 2013), thereby restricting the targetable genomic loci, although 5'-NAG-3' and 5'-NGA-3' PAMs are considered to be non-canonical PAM sequences (Hsu et al., 2013; Zhang et al., 2014).

In order to expand the genome-targeting scope of CRISPR/SpCas9, several SpCas9 variants recognizing additional PAM have been reported, including xCas9 and SpCas9-NG (Hu et al., 2018; Nishimasu et al., 2018). As to xCas9, it was initially reported to have flexible PAM selectivity, while a further study revealed the low activity at the sites containing 5'-NGT-3', 5'-NGC-3', 5'-NGA-3', 5'-GAN-3', and 5'-NAA-3' PAMs, indicating its restriction on PAM (He et al., 2019). SpCas9-NG can recognize 5'-NGN-3' PAM, which greatly relaxes the PAM requirement at the third base. At the same time, a large number of genomic loci still cannot be targeted

✉ Junzhao ZHAO, z.joyce08@163.com
Feng GU, gufenguw@gmail.com
* The two authors contributed equally to this work
Junzhao ZHAO, https://orcid.org/0000-0002-2385-205X
Feng GU, https://orcid.org/0000-0002-9600-8407

Received Aug. 12, 2021; Revision accepted Jan. 27, 2022; Crosschecked May 4, 2022

with it (Nishimasu et al., 2018). Recently, SpRY, a variant of *S. pyogenes* Cas9 (SpCas9), has been reported as a near-PAMless CRISPR/Cas9 variant (5'-NRN-3'>5'-NYN-3', R: A/G, Y: C/T), which nearly completely relaxes the PAM requirement of SpCas9 (Walton et al., 2020). With SpRY, it is possible to generate previously inaccessible disease-relevant genetic variants, increasing the utility of high-resolution targeting across genome editing applications (Walton et al., 2020). Thus far, SpRY has been harnessed for the manipulation of the genomes of different species, including *Oryza sativa* (rice), *Candida albicans*, and *Dictyostelium discoideum* (Asano et al., 2021; Evans and Bernstein, 2021; Xu et al., 2021), demonstrating the flexible PAM requirements. Recently, based on SpRY, new variants for enhanced fidelity have been identified, which would partially solve the editing scope and fidelity issue (Zhang et al., 2021).

The characterization of PAM preferences could be a key step for the development or engineering of novel CRISPR enzymes. To date, different methods have been proposed for this task, which can be broadly divided into in vitro, bacterial cell-based, and mammalian cell-based approaches (Walton et al., 2021). The high-throughput PAM determination assay (HT-PAMDA) method enables the characterization of the targeting ranges of Cas nucleases on the basis of the in vitro cleavage of plasmid libraries harboring randomized PAMs, which has been adapted for the characterization of SpRY enzyme (Walton et al., 2020). As this nuclease is involved in the expression of Cas proteins in human cells and the in vitro cleavage of target sequence using a PAM library, this reaction condition cannot be identical to that of target sequences in human cells, which may have bias for PAM identification.

In our previous studies, we developed a positive screening system termed PAM Definition by Observable Sequence Excision (PAM-DOSE) to delineate the functional PAMs in human cells (Tang et al., 2019). Also, we generated a green fluorescent protein (GFP)-reporter system to quantitatively test the efficiency of CRISPR/Cas9-mediated DNA cleavage in human cells with different PAM sequences (Zhang et al., 2014). In the present study, we sought to identify the PAMs of SpRY using the PAM-DOSE and GFP-reporter systems, to augment the understanding and benchmarking of the performance of SpRY in genome editing applications.

2 Materials and methods

2.1 Plasmid construction

The plasmids pX330 (harboring the wild-type SpCas9 coding sequence), pX330-NG (harboring the SpCas9-NG coding sequence), pX601, and SpRY-ABEmax (7.10) were obtained from Addgene (plasmids #42230, #117919, #61591, and #140003; Addgene, Watertown, MA, USA). To make the constructs comparable, we generated all constructs using the same backbone. The coding sequences for SpCas9 from SpCas9-NG and SpRY-ABEmax (7.10) were amplified and inserted into the backbone of pX330 without the SpCas9-coding sequence. The coding sequence of SpCas9 from SpRY-ABEmax (7.10) was initially inactive and was mutated to the active form using site-directed mutagenesis for this study. The primers used to construct the plasmid pX330-SpRY are listed in Table S1. SgRNA oligos were annealed and inserted into the backbone of pX330 using a standard protocol. The vector plasmid pmTmG, which contains *tdTomato*, enhanced green fluorescent protein (*EGFP*), and target sites, has been described in our previous study (Yang et al., 2017). The generation of PAM libraries was performed as detailed in our previous work (Tang et al., 2019). The target sequences and the oligonucleotide sequences for sgRNA construction of Zeo, Neo, and sloxP were listed in Tables S2 and S3, respectively. Plasmid DNA was isolated using established protocols. The DNA sequencing confirmed the presence of the specific sequences required in the construct.

2.2 Cells and their culture

HEK-293 cells obtained from the American Type Culture Collection (ATCC; CAT#CRL-1573) were grown at 37 °C under 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM; Life Technologies, Carlsbad, CA, USA), supplemented with 10% (volume fraction) heat-inactivated fetal bovine serum (FBS) and penicillin/streptomycin. HEK-293 cells expressing EGFP, named 293-SC1, were described previously (Zhang et al., 2014). To maintain EGFP expression, the medium for 293-SC1 culture included puromycin.

2.3 Transfection, images, and flow cytometry analysis

In the PAM-DOSE experiment, 1.8×10^5 HEK-293 cells were seeded per well in a 12-well plate on Day 1. On Day 2, HEK-293 cells were co-transfected with 400 ng of the pmTmG PAM library and 0.75 times the molar number of plasmids expressing Cas and sgRNA, using TurboFect (Thermo Fisher Scientific, Waltham, MA, USA). On Day 3, fresh medium was added to the wells of transfected HEK-293 cells. On Day 4, images were obtained under an inverted fluorescence microscope (Nikon TS2-FL, Nikon, Tokyo, Japan).

For the GFP-reporter system-based PAM study, 293-SC1 cells were plated at a density of 0.9×10^5 cells/well in a 24-well plate on Day 1, and transfected with 250 ng CRISPR/*Staphylococcus aureus* Cas9 (SaCas9) and 125 ng sgRNA plasmids with TurboFect on Day 2. Fresh medium was added to the wells of the transfected 293-SC1 cells on Day 3. The transfected cells were analyzed using a FACS-Calibur flow cytometer (BD Biosciences, NY, USA) on Day 4. The *EGFP* target sequences and the oligonucleotide sequences for sgRNA construction targeting them are listed in Tables S4 and S5.

2.4 Sequence analysis

The purification of plasmids from the cells was performed using standard protocols. Polymerase chain reaction (PCR) was used to amplify the sequences flanking the CRISPR-targeted locus, and the products were inserted into the vector pJET1.2 (CloneJET PCR Cloning Kit, Thermo Fisher Scientific) for Sanger sequencing. The fragment flanking the target sequence was amplified for next generation sequencing (NGS) via a two-round PCR protocol including barcodes for the amplicons. The detailed sequences of the primers used are listed in Table S6. Amplicons were subsequently subjected to paired-end read sequencing using the Hiseq-PE150 strategy of Novogene (Nanjing, China). The deep sequencing data are available at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject PRJNA727919 (SRA: SRR14460859; sample accession numbers, SAMN19065633). The open source software CRISPResso Version 1.0.10 (Pinello et al., 2016) was employed to analyze the status of base editing. The PAM regions without indels within three bases of the PAM were extracted. Then, PAMs were counted and used to generate sequence logos (Crooks et al., 2004).

3 Results

3.1 Identification of PAM preferences with PAM-DOSE

In our previous study, we developed a system termed PAM-DOSE to delineate the functional PAMs in human cells (Tang et al., 2019). Specifically, there are two target sequences, as illustrated in Fig. 1a, including sloxP for the guide RNA spacer for SaCas9 and spacer N_4 harboring a spacer plus four randomized nucleotides for the PAM identification. After cleavage, the *tdTomato* cassette is excised, and the CAG promoter drives the expression of *EGFP* gene directly. The corresponding CRISPR/Cas-mediated cleaved fragments with functional PAMs in human cells could be harvested for DNA sequencing (Fig. 1b). Also, it allows for live visualization under fluorescence microscope or detection via flow cytometry analysis (FCA) (Tang et al., 2019). Using this system, in our previous study, we investigated the PAM preferences of SpCas9, as well as three Cas12a family orthologs—FnCas12a, AsCas12a, and LbCas12a (Tang et al., 2019).

In order to identify the PAM preference of SpRY in human cells, we selected two target sequences (N_4-Zeo and N_4-Neo) (Fig. 1a). SpRY, SpCas9-NG, and SpCas9-WT (wild-type SpCas9) were tested with the PAM libraries of Zeo and Neo using PAM-DOSE, as we previously reported (Tang et al., 2019). Plasmids of SaCas9-sloxP, Zeo (or Neo) PAM libraries, and SpRY (or SpCas9-NG or SpCas9-WT) were co-transfected into HEK-293 cells. The fluorescence microscope images were obtained, which demonstrated the presence of more green cells in both the Zeo or Neo PAM library with the SpRY group, as compared with SpCas9-NG or SpCas9-WT (Figs. 1c and S1). Collectively, according to the data of fluorescence microscope, it was noted that SpRY is more flexible to recognize PAMs than SpCas9-NG or SpCas9-WT.

The above results showed that SpRY may be more flexible for the requirements of PAMs, while the PAMs of SpRY cannot be directly delineated. We inserted the corresponding CRISPR/Cas-mediated cleaved fragments harboring the PAM sequence into

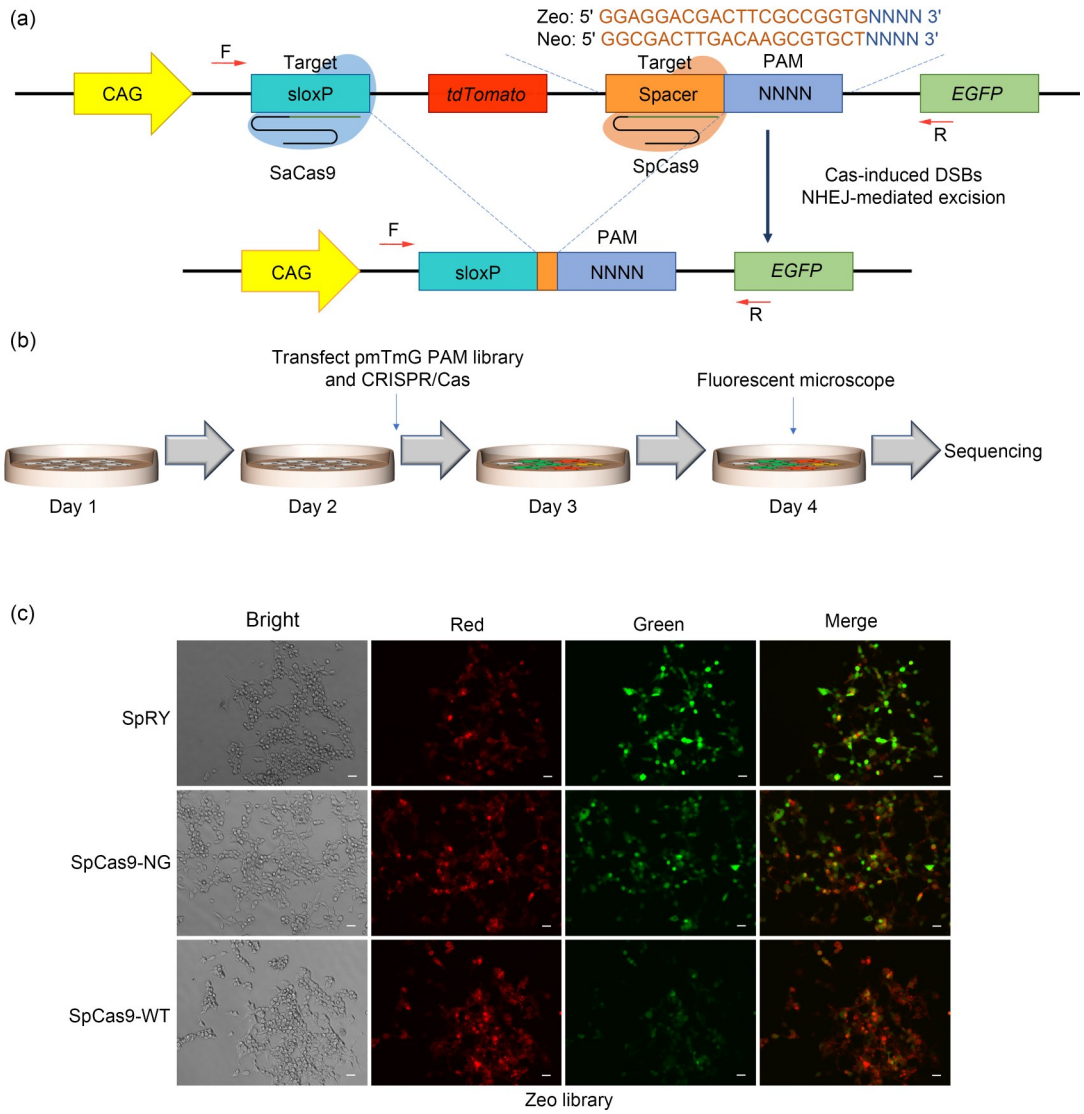

Fig. 1 PAM-DOSE and its results via fluorescent microscopy. (a) Illustration of the PAM-DOSE system. A library of PAM sequences was introduced into the constructs. In the presence of a functional PAM, cleavage-mediated *tdTomato* cassette excision was performed, leading to the expression of EGFP, which could be tracked via fluorescence microscopy. (b) Illustration of each step of PAM-DOSE. The red cells represent cells expressing *tdTomato*, the green cells represent cells expressing GFP, the yellow cells represent cells expressing both two colors, and the gray cells represent cells without fluorescence. (c) HEK-293 cells were co-transfected with a pmTmG-N_4-Zeo PAM library and the plasmids expressing Cas and sgRNA for the cleavage, and images were obtained 48 h after transfection. Scale bars are 100 μm. Cas9: clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein 9; DSB: double-strand break; EGFP: enhanced green fluorescent protein; F: forward; GFP: green fluorescent protein; NHEJ: non-homologous end joining; PAM: protospacer adjacent motif; PAM-DOSE: PAM Definition by Observable Sequence Excision; R: reverse; SaCas9: *Staphylococcus aureus* Cas9; sgRNA: single-guide RNA; SpCas: *Streptococcus pyogenes* Cas9; SpRY: a variant of SpCas9; SpCas9-NG: a variant of SpCas9 (PAM is 5'-NGN-3'); WT: wild type.

the pJET vectors and transformed them into competent *Escherichia coli*. We randomly picked 15 individual colonies in each of the two sites. The results of Sanger sequencing showed that the ratio of 5'-NGN-3' PAM to the total number was 0.53 to 0.60, and the ratio of 5'-NWN-3' (W: A/T) PAM to the total was 0.33 to 0.47, with a small proportion 5'-NCN-3' to the total of 0 to 0.07 (Figs. 2a and 2b). These results roughly illustrated that 5'-NDN-3' (D: A/G/T) PAM may trigger efficient cleavage.

Because of the low throughput of Sanger sequencing, we performed NGS for PAM identification. Prior

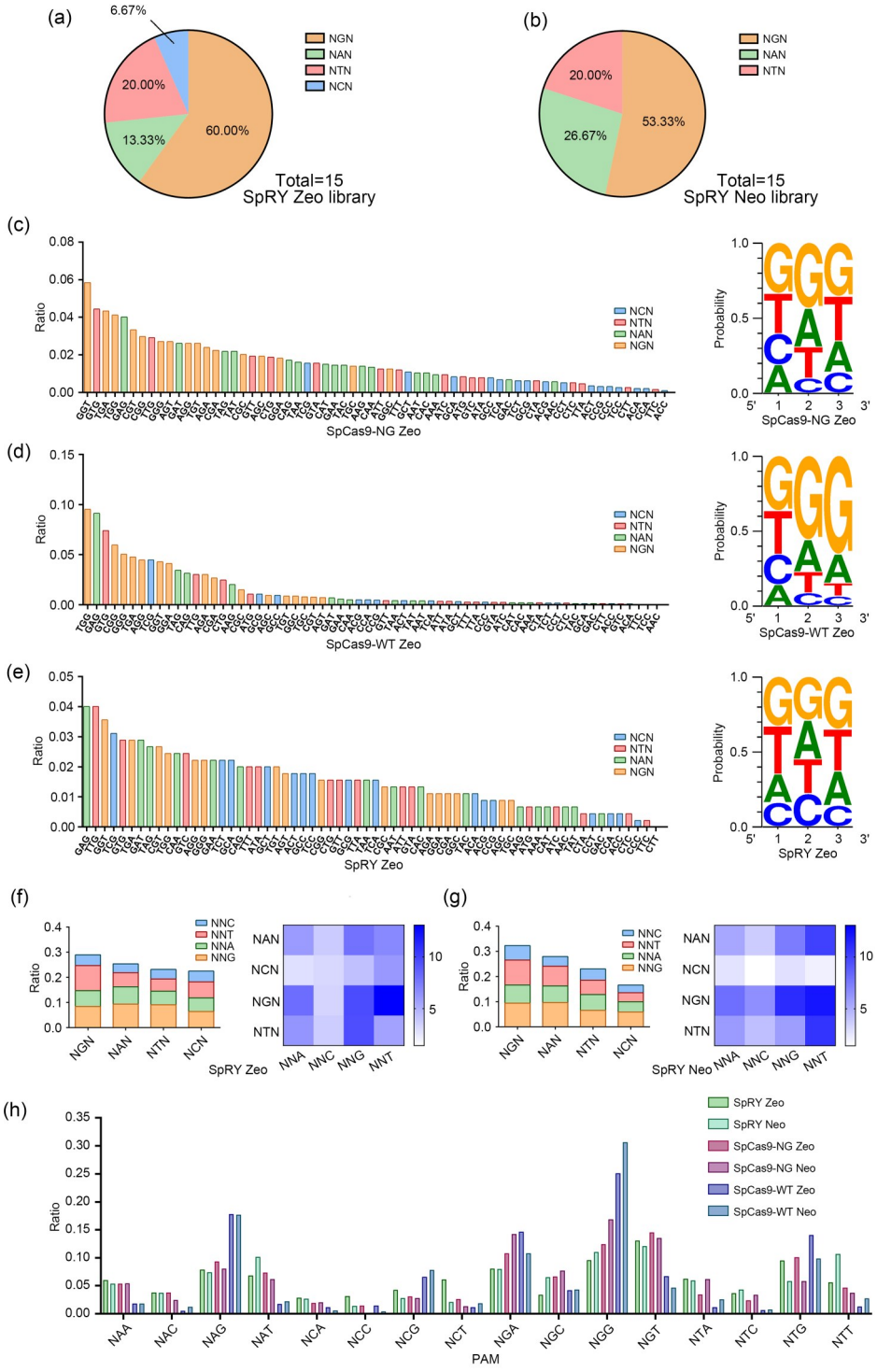

Fig. 2 Sequencing data of SpRY, SpCas9-NG, and SpCas9-WT using PAM-DOSE. (a, b) Pie chart of the PAM results for SpRY at the Zeo and Neo sites via Sanger sequencing. (c, d, e) The ratio of recognized PAMs calculated and the sequence logo based on NGS results at the Zeo sites of SpCas9-NG, SpCas9-WT, and SpRY. (f, g) The stacked bars and the heatmaps show the PAM preferences of SpRY at the Zeo and Neo sites, based on the NGS results. (h) The ratios of SpRY, SpCas9-NG, and SpCas9-WT at the Zeo and Neo sites, based on NGS results, with 16 5'-NNN-3' PAMs. NGS: next generation sequencing; PAM: protospacer adjacent motif; PAM-DOSE: PAM Definition by Observable Sequence Excision; SaCas9: *Staphylococcus aureus* Cas9; SpRY: a variant of SpCas9; SpCas9-NG: a variant of SpCas9 (PAM is 5'-NGN-3'); WT: wild type.

to data analysis, we tested whether the PAMs for SpCas9-WT and SpCas9-NG would be consistent as that in the literature (Hsu et al., 2013; Hirano et al., 2016; Nishimasu et al., 2018). Our results confirmed that SpCas9-WT is specific to 5'-NGG-3' PAM, and SpCas9-NG preferentially recognizes 5'-NG-3' PAM. In addition, SpCas9-WT and SpCas9-NG recognize 5'-NAG-3' and 5'-NAN-3' as non-canonical PAMs, respectively (Figs. 2c, 2d, S2a, and S2b). These results showed that the system enables highly reproducible characterizations that are consistent with the literature (Hsu et al., 2013; Hirano et al., 2016; Nishimasu et al., 2018).

Compared with SpCas9-NG and SpCas9-WT, SpRY is more flexible for the recognition of PAMs (Figs. 2e, S2c, and Table S7). Similar to that of SpCas9-NG, 5'-NGN-3' PAM accounts for the largest proportion of PAMs in SpRY. With different spacer sequences, SpRY has different preferences for 5'-NGC-3' and 5'-NCT-3' PAMs (Figs. 2f and 2g, Table S7). Specifically, if we set more than 5% of the total reads as PAMs in the Zeo library, the functional PAMs of SpRY were 5'-NDD-3' and 5'-NCT-3'. Meanwhile, in the Neo library, the functional PAMs of SpRY are 5'-NDD-3' or 5'-NGC-3' (Fig. 2h). These results indicated that SpRY has the potential to recognize 5'-NDD-3' PAM, and the second and third base C may affect the preference of SpRY (Figs. 2f and 2g). Notably, there is no evidence to support that SpRY can target sequences with a 5'-NCV-3' (V: A/C/G) PAM at these two tested sites, which is consistent with the result of HT-PAMDA (Walton et al., 2020). Collectively, these results demonstrated that 5'-NDD-3' may be served as PAMs for SpRY-mediated genome editing in human cells.

3.2 Identification of PAMs of SpRY, SpCas9-NG, and SpCas9-WT with the GFP-reporter system

As PAM-DOSE can only show the proportion of recognized PAMs rather than provide direct editing efficiency, we sought to test these results in 293-SC1 cells, which harbor *EGFP* gene in their genome. The generation and characteristics of the 293-SC1 cell line have been described in our previous study (Zhang et al., 2014). First, sgRNAs designed for a variety of PAMs targeting the *EGFP* gene were inserted into vectors. When a PAM is recognized, Cas9 can induce DNA double-strand breaks in the *EGFP* gene, which can create frameshift indel mutations, resulting in a loss of function. The green fluorescence of the 293-SC1 cells would therefore be quenched, and the loss of activity is easily detectable using flow cytometry (Fig. 3a). The activity of genome editing was measured by identifying the proportion of GFP negative cells.

A total of 46 sites containing 5'-NNN-3' PAMs were evaluated with three SpCas9 nucleases (Fig. 3b). We set the PAMs with an editing efficiency of more than 15% as functional PAMs. We found that SpRY induced the highest editing efficiency with 5'-NRN-3' PAM, which is consistent with the results for PAM-DOSE and that reported by Walton et al. (2020). The top seven sites were all harboring the 5'-NGN-3' PAM. We noticed that four of these were 5'-NGC-3', compared with a low proportion with the Zeo site of PAM-DOSE or low editing efficiency in *O. sativa* (Li et al., 2021; Xu et al., 2021). Similar results have been found with 5'-NCT-3' PAM. Surprisingly, the efficiencies of most 5'-NCD-3' PAMs were higher than those of 5'-NTN-3' PAMs, which is not consistent with the results of PAM-DOSE (Fig. 3c). Especially, the efficiencies of target sequences with 5'-NCD-3' PAMs were all above 20%, while only target sequences with 5'-NTA-3' PAM had an efficiency higher than 20% among the 5'-NTN-3' PAMs. Collectively, with the GFP-reporter system, 5'-NRN-3', 5'-NCD-3', and 5'-NTA-3' may be considered as functional PAMs.

Our above results illustrated that SpCas9 variants possess different PAM preferences. We then wanted to evaluate the performance of the variants and its parental one with the identical target sequence. We compared the results of SpRY with those of SpCas9-WT and SpCas9-NG. The activity of SpRY and SpCas9-WT with 5'-NGG-3' PAMs was very similar. The activity of SpRY was much higher across most of the NGN sites than that of SpCas9-NG in the GFP-reporter system. Considering the non-NG PAMs, the efficiencies of SpRY with 5'-NAN-3' and 5'-NCN-3' PAMs were around three-fold higher than that of SpCas9-NG (3.08-fold for 5'-NAN-3' and 3.58-fold for 5'-NCN-3') and SpCas9-WT (3.97-fold for 5'-NAN-3' and 4.07-fold for 5'-NCN-3'). The efficiency of SpRY with 5'-NTN-3' PAM was approximately two-fold higher than those of SpCas9-NG (1.57-fold) and SpCas9-WT (2.16-fold) (Fig. 3d). Therefore, SpRY improved the efficiency with almost all of the non-NG PAMs, especially 5'-NAN-3', 5'-NCD-3', and 5'-NTA-3' PAMs. Notably, none of the three

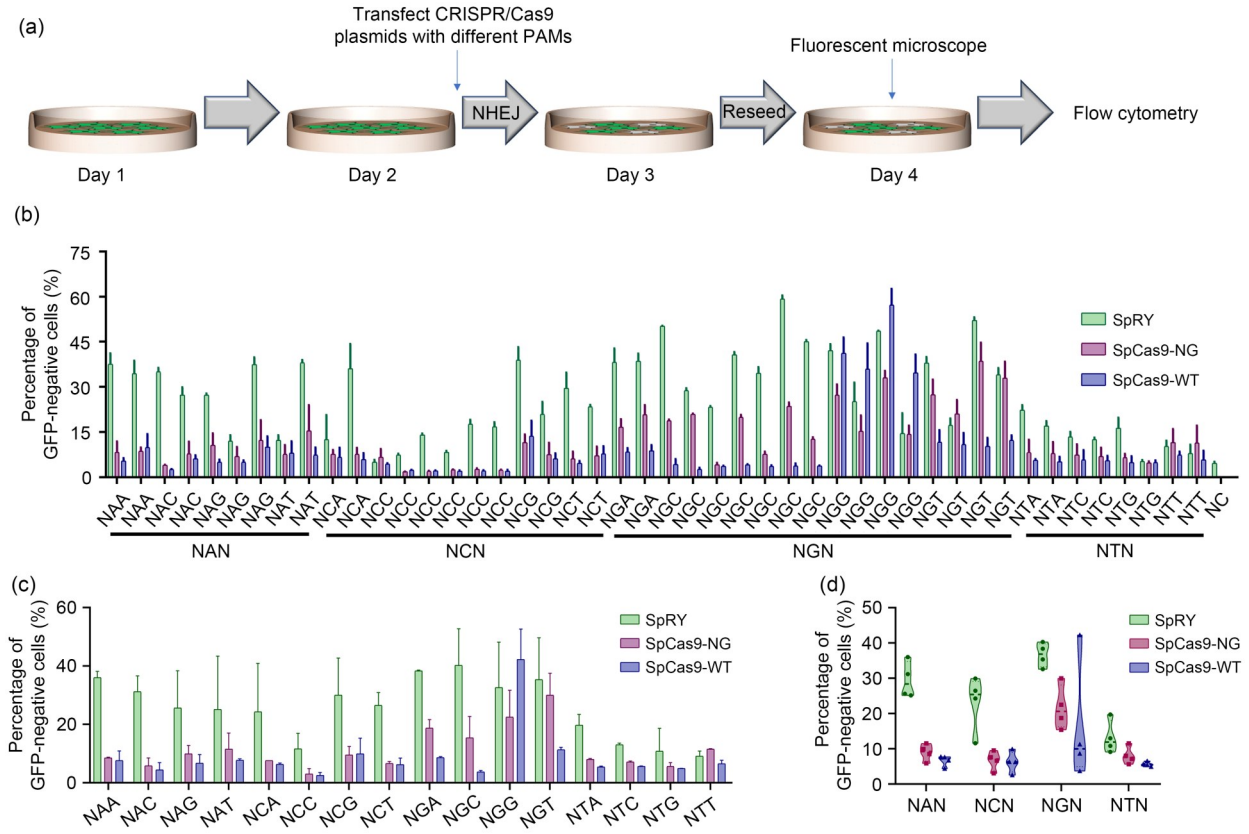

Fig. 3 Identification of the PAMs of SpRY, SpCas9-NG, and SpCas9-WT using the GFP-reporter system. (a) Illustration of the application of the GFP-reporter system. Parental cells are in green. After the editing, the loss of green fluorescence will be observed in partial cells. (b) The results of the editing efficiencies of SpRY, SpCas9-NG, and SpCas9-WT in *EGFP* gene via FCA. Data are expressed as mean±SD, *n*=3. (c) The average efficiency of FCA as 16 5'-NNN-3' PAMs. Data are expressed as mean±SD. (d) The median results of FCA as four 5'-NNN-3' PAMs, where the pots represent the average efficiency of 5'-NNN-3' sites tested via GFP-reporter. CRISPR: clustered regularly interspaced short palindromic repeats; Cas9: CRISPR-associated protein 9; *EGFP*: enhanced green fluorescent protein; FCA: flow cytometry analysis; GFP: green fluorescent protein; NHEJ: non-homologous end joining; PAM: protospacer adjacent motif; SD: standard deviation; SpCas: *Streptococcus pyogenes* Cas9; SpRY: a variant of SpCas9; SpCas9-NG: a variant of SpCas9 (PAM is 5'-NGN-3'); WT: wild type.

variants showed robust activity with 5'-NCC-3' PAM (Fig. 3c).

4 Discussion

It has been reported that SpRY nuclease has a more flexible PAM preference than SpCas9-WT and SpCas9-NG (Walton et al., 2020). Specifically, it relaxes or almost entirely removes the dependence of SpCas9 on specific PAMs, extending its applicability to sites with 5'-NGN-3' and 5'-NAN-3' PAM and to many sites with 5'-NCN-3' or 5'-NTN-3' PAM, albeit at a reduced relative efficiency. Therefore, SpRY greatly expands the target choices. Further studies have been performed for the genome manipulation of rice (Li et al., 2021; Xu et al., 2021). For human therapeutic purposes, there is no doubt that there is a demand for research on PAM identification using human target cells. To address this need, in the present study, we investigated the PAM preference in human cells with two systems (PAM-DOSE and GFP-reporter) developed in our previous studies. PAM-DOSE provides the PAM preference of SpRY at a specific target sequence. However, since SpRY could be utilized for the manipulation of the human genome for molecular therapeutic purpose, the results for the GFP-reporter may provide more insight into the selection of target sequence. Therefore, based on the results of these two methods, 5'-NRD-3' and 5'-NTA-3' may be treated as

canonical PAMs. 5'-NRC-3' and 5'-NCK-3' (K: G/T) triggered relative high activity in all sites with them in GFP-reporter but not in PAM-DOSE, which also could be served as PAMs. With PAM-DOSE but not with GFP-reporter, 5'-NCA-3' and 5'-NTK-3' have certain activity, and hence these may only be selected as non-priority PAMs.

Nevertheless, we acknowledge that these two assays have limitations. With respect to the PAM-DOSE assay, only the cleavage events that do not alter the PAM sequence are calculated, resulting in partial PAM preference loss. At the same time, the GFP-reporter assay only shows frameshift indels or function loss from key residues missing, which may underestimate cleavage efficiency. Therefore, we compared our results with those from the literature, and summarized them in Table 1. To allow for a more robust comparison, we re-analyzed the data from the literature and calculated the relative activity with the formula O/O_{NGG}, where O is the value of each PAM in the methods, and O_{NGG} is the value of 5'-NGG-3' PAM in each study (Table 1). When we set the PAMs with relative activity of more than 0.50 as canonical PAMs, all of the results revealed that 5'-NRN-3' could be considered as canonical PAMs. The 5'-NCG-3', 5'-NCT-3', and 5'-NTA-3' PAMs performed well in our study, and also were considered as canonical PAMs, which is not consistent with that of HT-PAMDA. It was revealed that the target sequence in plasmids (PAM-DOSE and HT-PAMDA) as episome or in the genome (GFP-reporter) may modulate the results of PAM identification as we reported previously (Tang et al., 2019). We speculated that the inconsistency between our study and HT-PAMDA may also be due to the different reaction conditions and/or different target sequences. Moreover, the relative activity of 5'-NGC-3' PAM was high in most studies in human cells but was low in *O. sativa* and *D. discoideum* genomes, which revealed that the PAM preference may not be the same among species. Furthermore, with the 5'-NYN-3' PAM, SpRY possessed different activity in the literature (Walton et al., 2020; Asano et al., 2021; Li et al., 2021; Xu et al., 2021), especially with 5'-NCA-3' and 5'-NTK-3' PAMs, indicating that their efficiencies of SpRY may be greatly affected by the target sequence or species. Notably, according to results shown in the table, we

Table 1 Comparison of the relative activity of SpRY with 5'-NNN-3' PAMs between different studies

PAM	Human				*Oryza sativa* (rice)		*Dictyostelium discoideum*
	PAM-DOSE	GFP-reporter system	HT-PAMDA	NGS of human genome	Sanger sequencing of rice genome	NGS of rice genome	*tdTomato* gene
NAA	**0.55**	**1.10**	**1.21**	**1.34**	**0.52**	0.46	0
NAC	0.36	**0.96**	**1.32**	**1.46**	**0.93**	**2.12**	**1.08**
NAG	**0.74**	**0.79**	**1.78**	**0.81**	0.39	**1.10**	**0.78**
NAT	**0.82**	**0.77**	**0.64**	**0.73**	**0.88**	0.43	**0.82**
NCA	0.27	**0.75**	0.15	**1.10**	0	**0.52**	0.03
NCC	0.22	0.35	0.07	0.38	0.12	0.18	0.38
NCG	0.34	**0.92**	0.36	0.20	0.05	**1.09**	0
NCT	0.40	**0.81**	0.13	**0.80**	0.04	0.12	0.05
NGA	**0.78**	**1.18**	**1.13**	**1.16**	**2.07**	0.39	**0.55**
NGC	0.48	**1.24**	**0.78**	**1.30**	0.02	0.26	0.32
NGG	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
NGT	**1.22**	**1.08**	**0.64**	**1.36**	**0.69**	**1.03**	**0.61**
NTA	**0.59**	**0.61**	0.22	**0.72**	0	0.18	0.47
NTC	0.39	0.40	0.03	0.46	**0.85**	0.18	**1.08**
NTG	**0.74**	0.33	0.38	**0.72**	0	0.32	0.01
NTT	**0.79**	0.28	0.04	0.18	0	0.11	0.11
Ref.	This study		Walton et al., 2020		Xu et al., 2021	Li et al., 2021	Asano et al., 2021

The relative activity was determined by the formula O/O_{NGG}, where O is the value of each PAM in the methods, and O_{NGG} is the value of 5'-NGG-3' PAM in each method. The values mean proportion in PAM-DOSE, rate constants (k) in HT-PAMDA, and efficiency in the others. Values higher than 0.50 are highlighted in bold type. GFP: green fluorescent protein; HT-PAMDA: high-throughput PAM determination assay; NGS: next generation sequencing; PAM: protospacer adjacent motif; PAM-DOSE: PAM Definition by Observable Sequence Excision; Ref.: reference; SpRY: a variant of *Streptococcus pyogenes* Cas9.

found that 5'-NYC-3' PAM cannot be efficiently recognized by SpRY in human cells, highlighting the fact that, as for the design of sgRNA, the fewest target sequences may be selected with these PAMs.

We acknowledge that PAM identification may be resolved with a PAM library integrated into the genome via site-specific recombination (Fareh et al., 2021), which may provide a better interpretation of DNA cleavage and repair in chromosomes. Further studies would be required to generate cells harboring PAM libraries for the identification of PAM preferences. The current study focused on the cleavage of SpRY instead of base editing with inactive SpRY; therefore it should be further investigated whether the situation is the same with the recognition of the target sequence for wild-type SpRY and inactive SpRY.

5 Conclusions

In this work, we identified the PAM sequences recognized by SpRY nucleases in human cells. The PAM sequences for the nuclease can be summarized as follows: 5'-NRN-3', 5'-NTA-3', and 5'-NCK-3' could be considered as canonical PAMs, while 5'-NCA-3' and 5'-NTK-3' may be served as non-priority PAMs; and 5'-NYC-3' PAM is not recommended.

Acknowledgments

This work was supported by Lin HE's Academician Workstation of New Medicine and Clinical Translation (No. 18331105) and the Program for Basic Science and Technology Cooperation Projects of Wenzhou City (No. H22010011), China.

Author contributions

Feng GU and Junzhao ZHAO conceived the idea and performed data analyses. Jinbin YE, Haitao XI, Yilu CHEN, Qishu CHEN, Xiaosheng LU, Jineng LV, Yamin CHEN, and Feng GU performed the experiments. Jinbin YE and Feng GU wrote the manuscript. All authors have read and approved the final manuscript, and therefore, have full access to all the data in the study and take responsibility for the integrity and security of the data.

Compliance with ethics guidelines

Jinbin YE, Haitao XI, Yilu CHEN, Qishu CHEN, Xiaosheng LU, Jineng LV, Yamin CHEN, Feng GU, and Junzhao ZHAO declare that they have no conflict of interest.

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Asano Y, Yamashita K, Hasegawa A, et al., 2021. Knock-in and precise nucleotide substitution using near-PAMless engineered Cas9 variants in *Dictyostelium discoideum*. *Sci Rep*, 11:11163. https://doi.org/10.1038/s41598-021-89546-0

Crooks GE, Hon G, Chandonia JM, et al., 2004. WebLogo: a sequence logo generator. *Genome Res*, 14(6):1188-1190. https://doi.org/10.1101/gr.849004

Evans BA, Bernstein DA, 2021. SpRY Cas9 can utilize a variety of protospacer adjacent motif site sequences to edit the *Candida albicans* genome. *mSphere*, 6(3):e00303-21. https://doi.org/10.1128/mSphere.00303-21

Fareh M, Zhao W, Hu WX, et al., 2021. Reprogrammed CRISPR-Cas13b suppresses SARS-CoV-2 replication and circumvents its mutational escape through mismatch tolerance. *Nat Commun*, 12:4270. https://doi.org/10.1038/s41467-021-24577-9

He XB, Wang YF, Yang FY, et al., 2019. Boosting activity of high-fidelity CRISPR/Cas9 variants using a $tRNA^{Gln}$-processing system in human cells. *J Biol Chem*, 294(23): 9308-9315. https://doi.org/10.1074/jbc.RA119.007791

Hirano S, Nishimasu H, Ishitani R, et al., 2016. Structural basis for the altered PAM specificities of engineered CRISPR-Cas9. *Mol Cell*, 61(6):886-894. https://doi.org/10.1016/j.molcel.2016.02.018

Hsu PD, Scott DA, Weinstein JA, et al., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*, 31(9):827-832. https://doi.org/10.1038/nbt.2647

Hsu PD, Lander ES, Zhang F, 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157(6):1262-1278. https://doi.org/10.1016/j.cell.2014.05.010

Hu JH, Miller SM, Geurts MH, et al., 2018. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, 556(7699):57-63. https://doi.org/10.1038/nature26155

Jiang FG, Doudna JA, 2017. CRISPR-Cas9 structures and mechanisms. *Annu Rev Biophys*, 46:505-529. https://doi.org/10.1146/annurev-biophys-062215-010822

Jinek M, Chylinski K, Fonfara I, et al., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096):816-821. https://doi.org/10.1126/science.1225829

Komor AC, Badran AH, Liu DR, 2017. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell*, 168(1-2):20-36. https://doi.org/10.1016/j.cell.2016.10.044

Li J, Xu RF, Qin RY, et al., 2021. Genome editing mediated by SpCas9 variants with broad non-canonical PAM compatibility in plants. *Mol Plant*, 14(2):352-360. https://doi.org/10.1016/j.molp.2020.12.017

Mojica FJM, Díez-Villaseñor C, García-Martínez J, et al., 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*

(Reading), 155(Pt 3):733-740. https://doi.org/10.1099/mic.0.023960-0

Nishimasu H, Shi X, Ishiguro S, et al., 2018. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science*, 361(6408):1259-1262. https://doi.org/10.1126/science.aas9129

Pinello L, Canver MC, Hoban MD, et al., 2016. Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol*, 34(7):695-697. https://doi.org/10.1038/nbt.3583

Shah SA, Erdmann S, Mojica FJM, et al., 2013. Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol*, 10(5):891-899. https://doi.org/10.4161/rna.23764

Sternberg SH, Redding S, Jinek M, et al., 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490):62-67. https://doi.org/10.1038/nature13011

Tang LC, Yang FY, He XX, et al., 2019. Efficient cleavage resolves PAM preferences of CRISPR-Cas in human cells. *Cell Regen*, 8(2):44-50. https://doi.org/10.1016/j.cr.2019.08.002

Walton RT, Christie KA, Whittaker MN, et al., 2020. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science*, 368(6488):290-296. https://doi.org/10.1126/science.aba8853

Walton RT, Hsu JY, Joung JK, et al., 2021. Scalable characterization of the PAM requirements of CRISPR-Cas enzymes using HT-PAMDA. *Nat Protoc*, 16(3):1511-1547. https://doi.org/10.1038/s41596-020-00465-2

Xu ZY, Kuang YJ, Ren B, et al., 2021. SpRY greatly expands the genome editing scope in rice with highly flexible PAM recognition. *Genome Biol*, 22:6. https://doi.org/10.1186/s13059-020-02231-9

Yang FY, Liu CB, Chen D, et al., 2017. CRISPR/Cas9-*loxP*-mediated gene editing as a novel site-specific genetic manipulation tool. *Mol Ther Nucleic Acids*, 7:378-386. https://doi.org/10.1016/j.omtn.2017.04.018

Zhang WW, Yin JH, Zhang-Ding ZR, et al., 2021. In-depth assessment of the PAM compatibility and editing activities of Cas9 variants. *Nucleic Acids Res*, 49(15):8785-8795. https://doi.org/10.1093/nar/gkab507

Zhang YL, Ge XL, Yang FY, et al., 2014. Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci Rep*, 4:5405. https://doi.org/10.1038/srep05405

Supplementary information

Figs. S1 and S2; Tables S1–S7