


RESEARCH PAPER



## Establishment of expanded and streamlined pipeline of PITCh knock-in – a web-based design tool for MMEJ-mediated gene knock-in, PITCh designer, and the variations of PITCh, PITCh-TG and PITCh-KIKO

Kazuki Nakamae, Yuki Nishimura, Mitsumasa Takenaga, Shota Nakade , Naoaki Sakamoto, Hiroshi Ide, Tetsushi Sakuma, and Takashi Yamamoto

Department of Mathematical and Life Sciences, Graduate School of Science, Hiroshima University, Hiroshima, Japan

### ABSTRACT

The emerging genome editing technology has enabled the creation of gene knock-in cells easily, efficiently, and rapidly, which has dramatically accelerated research in the field of mammalian functional genomics, including in humans. We recently developed a microhomology-mediated end-joining-based gene knock-in method, termed the PITCh system, and presented various examples of its application. Since the PITCh system only requires very short microhomologies (up to 40 bp) and single-guide RNA target sites on the donor vector, the targeting construct can be rapidly prepared compared with the conventional targeting vector for homologous recombination-based knock-in. Here, we established a streamlined pipeline to design and perform PITCh knock-in to further expand the availability of this method by creating web-based design software, PITCh designer (<http://www.mls.sci.hiroshima-u.ac.jp/smg/PITChdesigner/index.html>), as well as presenting an experimental example of versatile gene cassette knock-in. PITCh designer can automatically design not only the appropriate microhomologies but also the primers to construct locus-specific donor vectors for PITCh knock-in. By using our newly established pipeline, a reporter cell line for monitoring endogenous gene expression, and transgenesis (TG) or knock-in/knockout (KIKO) cell line can be produced systematically. Using these new variations of PITCh, an exogenous promoter-driven gene cassette expressing fluorescent protein gene and drug resistance gene can be integrated into a safe harbor or a specific gene locus to create transgenic reporter cells (PITCh-TG) or knockout cells with reporter knock-in (PITCh-KIKO), respectively.

### ARTICLE HISTORY

Received 24 January 2017  
Revised 22 March 2017  
Accepted 24 March 2017



### KEYWORDS

CRISPR–Cas9; gene knock-in; genome engineering; microhomology-mediated end-joining (MMEJ); web tool


## Introduction

Genome editing technology has rapidly become a standardized reverse genetics approach, leading to many important biologic and therapeutic discoveries. Genome editing allows site-specific gene engineering (e.g. gene knockout, gene knock-in, and various chromosomal rearrangements) via programmable nuclease-induced DNA double-strand break (DSB), followed by DSB repair.<sup>1</sup> In gene knock-in approaches, there are 3 accessible repair pathways: homologous recombination (HR), nonhomologous end-joining (NHEJ), and microhomology-mediated end-joining (MMEJ).<sup>2</sup> HR-mediated gene editing

enables precise knock-in<sup>3</sup>; however, it is time-consuming to make the targeting vector, and HR is only active during the late S/G<sub>2</sub> phases of the cell cycle, resulting in a laborious process and low efficiency. Conversely, homology-independent gene knock-in assisted by NHEJ can be performed in a cell cycle-independent manner in mammalian cells, resulting in highly efficient gene knock-in.<sup>4</sup> However, it is relatively error-prone due to the lack of homologous arms in the donor vector, and precise adjustment of junction sequences, such as intentional deletion/replacement/addition of several bases, is difficult because NHEJ simply re-joins the cut sites.

**CONTACT** Tetsushi Sakuma  [tetsushi-sakuma@hiroshima-u.ac.jp](mailto:tetsushi-sakuma@hiroshima-u.ac.jp)  Department of Mathematical and Life Sciences, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan.

Color versions of one or more of the figures in this article can be found online at [www.tandfonline.com/kbie](http://www.tandfonline.com/kbie).

 Supplemental data for this article can be accessed on the [publisher's website](#).

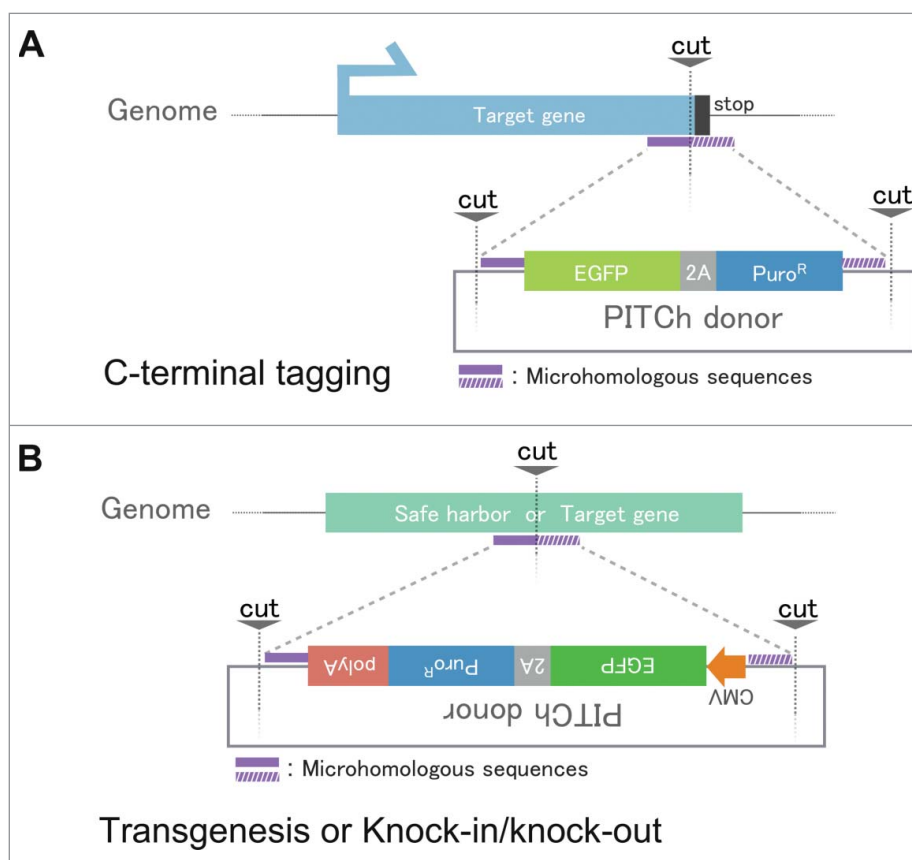
© 2017 Kazuki Nakamae, Yuki Nishimura, Mitsumasa Takenaga, Shota Nakade, Naoaki Sakamoto, Hiroshi Ide, Tetsushi Sakuma, and Takashi Yamamoto. Published with license by Taylor and Francis.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

We previously reported a gene knock-in approach with good efficiency, convenience, and flexibility, based on the MMEJ repair pathway, referred to as the PITCh (Precise Integration into Target Chromosome) system.<sup>2,5-8</sup> In the PITCh system, the targeting vector and the genomic target site are simultaneously cut by transcriptional activator-like effector nuclease (TALEN) or clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (Cas9), termed TAL-PITCh and CRIS-PITCh, respectively; then, the linearized DNA fragment is integrated into the genome via short microhomologies in the range of 8–40 bp, depending on MMEJ repair. The advantages of this approach are increased by the updated version of CRIS-PITCh, CRIS-PITCh (v2), using a generic single-guide RNA (sgRNA) to cleave the PITCh donor vector, designed outside of each microhomology, which confers high versatility.<sup>2</sup> The detailed protocol for designing, constructing, and transfecting the reagents

for CRIS-PITCh (v2) is available,<sup>2</sup> and the plasmids can be obtained from Addgene.<sup>2</sup>

The PITCh systems have 3 major characteristics. First, the active stages within the cell cycle for PITCh knock-in are different from those for HR knock-in. Since MMEJ can occur in most of the cell cycle, PITCh knock-in often results in higher efficiency than HR knock-in. In fact, we observed higher positive colony-forming efficiency using PITCh knock-in than using HR-mediated knock-in in HeLa cells.<sup>5</sup> In addition, a recent study showed that the overexpression of MMEJ-related genes such as exonuclease 1 can further enhance the knock-in efficiency.<sup>8</sup> Second, the lengths of the microhomologies on the PITCh donor vector are extremely short; thus, they can easily be added by a single PCR without any amplification from genomic DNA. Third, in contrast to NHEJ knock-in, such as HITI,<sup>4</sup> the CRIS-PITCh (v2) system can perform seamless knock-in, without carrying



**Figure 1.** Schematic illustrations of the PITCh knock-in for C-terminal tagging (A) and transgenesis or knock-in/knockout (B). In the C-terminal tagging, a promoterless EGFP-2A-Puro<sup>R</sup> cDNA is integrated into the genome to monitor endogenous gene expression and protein localization. In the transgenesis or knock-in/knockout, a constitutive promoter-driven CMV-EGFP-2A-Puro<sup>R</sup>-polyA cassette is integrated to establish a cell line with the stable expression of exogenous gene (PITCh-TG) or to disrupt gene function by knocking-in the exogenous gene cassette (PITCh-KIKO), respectively. The CMV-EGFP-2A-Puro<sup>R</sup>-polyA cassette should be integrated with the opposite orientation in the endogenous gene to avoid promoter interference. Puro<sup>R</sup>, puromycin resistance gene.

the partial target sequence of sgRNA, because the extra bases outside of microhomologies can be trimmed during the MMEJ pathway. With these advantageous characteristics, the PITCh system should expand the knock-in capacity in various cells and organisms.

Although CRIS-PITCh (v2) can be performed easily with the reported protocols and open-source reagents, as we mentioned earlier, the design of gene-specific sgRNAs and microhomologies has been left to the users themselves. In addition, reported examples and materials for CRIS-PITCh (v2) have been limited to the EGFP-labeling of an endogenous gene. In this study, we thus develop an online design tool for CRIS-PITCh (v2) and examine a knock-in of a constitutively expressed gene cassette to expand the utility of the PITCh system.

## Results and discussion

### Design principles of 2 types of PITCh knock-in

The current PITCh system is designed and applied for a promoterless EGFP cDNA fused with a 2A sequence, followed by a puromycin resistance gene (EGFP-2A-Puro<sup>R</sup>) to tag an endogenous gene without affecting the corresponding gene function (Fig. 1A). This approach is clearly useful and versatile for monitoring endogenous gene expression; however, there should be another type of versatile knock-in, namely, integration of a constitutive or controllable promoter-driven EGFP cassette into a safe harbor or a specific gene locus to create transgenic reporter cells (PITCh-TG) or knock-in/knockout (KIKO) cells (PITCh-KIKO), respectively (Fig. 1B). PITCh-TG is useful for establishing the cell line stably expressing an exogenous gene, where the knock-in cassette should be integrated in *AAVS1* locus (for human cells) or in *Rosa26* locus (for mouse cells). PITCh-KIKO can be used for gene disruption by biallelic knock-in of exogenous gene cassettes into the corresponding gene loci. Thus, the creation of a PITCh donor plasmid and CRISPR-Cas9 vector for such gene cassette knock-in should broaden the range of applications of PITCh knock-in.

### An experimental example of transgenesis using the PITCh system

To present an example of PITCh-TG, we designed and constructed a CRISPR-Cas9 vector and a PITCh donor vector targeting the *AAVS1* locus

(Fig. 2). The CRISPR-Cas9 vector was constructed to include a Cas9 nuclease expression cassette and 2 sgRNA cassettes targeting the PITCh donor vector and the genomic target site. On the PITCh donor vector, a cytomegalovirus (CMV) promoter-driven EGFP-2A-Puro<sup>R</sup> cassette with a polyA signal sequence was added in the opposite direction against the endogenous *AAVS1* gene to avoid promoter interference. The intended genomic sequence of the correct knock-in allele is shown in Fig. 3A. The CRISPR-Cas9 vector and the PITCh vector were cotransfected into HCT116 cells and puromycin-resistant clones were selected and isolated. Of 12 clones isolated, 8 were determined to have knock-in alleles by PCR amplification of 5' and 3' knock-in junctions (Fig. 3B). Further genotyping by out-out PCR revealed that 3 clones were monoallelic knock-in and 5 clones were biallelic knock-in (Fig. 3B). The genomic context of 4 selected clones was further confirmed by Southern blot analyses (Fig. 3C). These results proved that our materials for PITCh-TG functioned efficiently.

### Development of PITCh designer: A web-based design tool for PITCh knock-in

Toward an accessible design of PITCh knock-in, we created a free web-based design tool, named PITCh designer (Fig. 4A). This application only requires the sequence around the target region. After providing the sequence with some selectable options, the user selects the targeting base (Fig. 4B). Then, the application automatically designs the sgRNA, left and right microhomologies, and primers for the construction of the PITCh donor vector (Fig. 4C).

### Simple procedure of PITCh designer

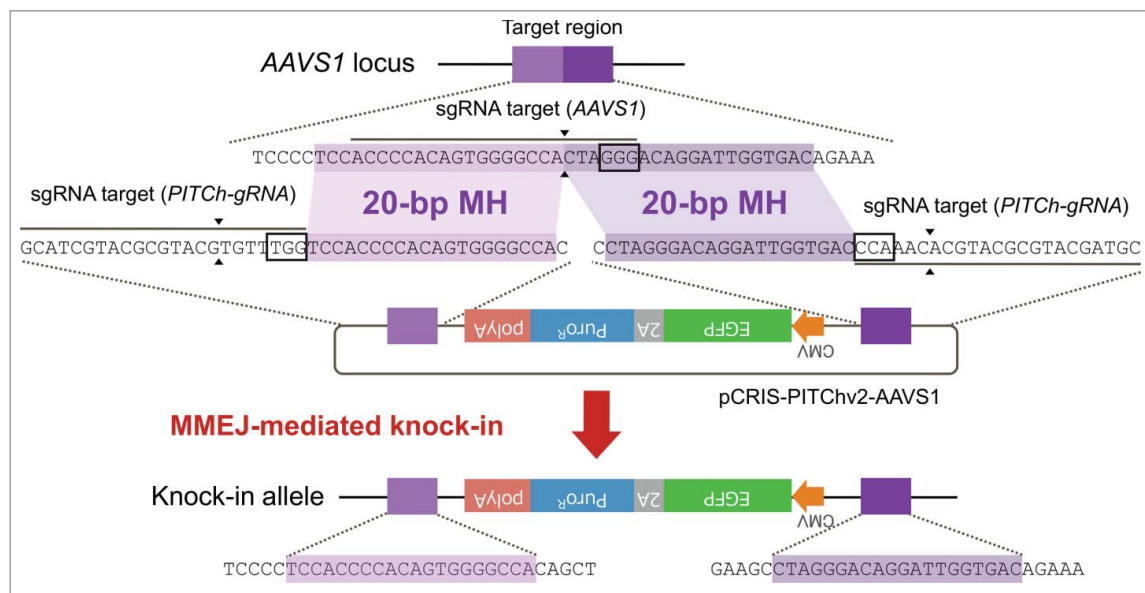
The procedure for the PITCh designer consists of the following 3 steps:

#### 1) Sequence submission

PITCh designer accepts an arbitrary nucleotide sequence in the FASTA format from 100 bp up to 1.5 kbp (Fig. 4A).

#### 2) Selecting the targeted nucleotide

PITCh designer analyzes the input sequence upon clicking of the "Submit" button, and provides a list of potential knock-in sites within the provided sequence. These potential knock-in sites are displayed as color-coded bases, depending on the presence of a PAM



**Figure 2.** Detailed design of PITCh-TG in human AAVS1 locus. Black lines indicate the target sequences of sgRNAs. Black boxes indicate PAM sequences. Black triangles indicate DSB sites. MH, microhomology. Puro<sup>R</sup>, puromycin resistance gene.

sequence within a window of 12 bp (6 bp upstream and 6 bp downstream) (Fig. 4B).

### 3) Displaying search results

After the user selects the desired knock-in position, the gene-specific sgRNA, microhomologies, and knock-in sequence are visually displayed. Furthermore, upon pushing the “Details” button, the textual information of gene-specific sgRNA, left and right microhomologies, primers for the construction of a PITCh donor vector according to the previous protocol,<sup>2</sup> and the 5′ and 3′ sequences of the PITCh donor is shown, along with images of the primer-binding sites on the donor vector (Fig. 4C). The sequence data and images can be downloaded as CSV and JPG files, respectively.

## Design options in PITCh designer

PITCh designer includes a variety of optional settings:

### 1) Reading frame

There are 4 options regarding reading frames: frame 1, frame 2, frame 3, and no frame, representing the open reading frames starting at the 1st, 2nd, and 3rd bases, and a sequence without frame specification, respectively.

### 2) Adjustment of reading frame

For endogenous gene tagging, the knock-in cassette has to be designed to produce a functional protein without containing a frame shift. PITCh designer provides the following 2 methods to adjust the reading frame.

The first method is “C-insertion,” which fills in the frame with additional cytosine(s). The addition of cytosine(s) never results in the production of a stop codon. The second method is “codon deletion,” which removes one or 2 bases to make the in-frame knock-in.

### 3) Knock-in cassette

As illustrated in Fig. 1A and B, 2 versatile knock-in systems are currently available in PITCh knock-in: the integration of promoterless EGFP-2A-Puro<sup>R</sup> or constitutive promoter-driven CMV-EGFP-2A-Puro<sup>R</sup>-polyA. PITCh designer supports the automated design of primers for each knock-in method.

### 4) Microhomology lengths

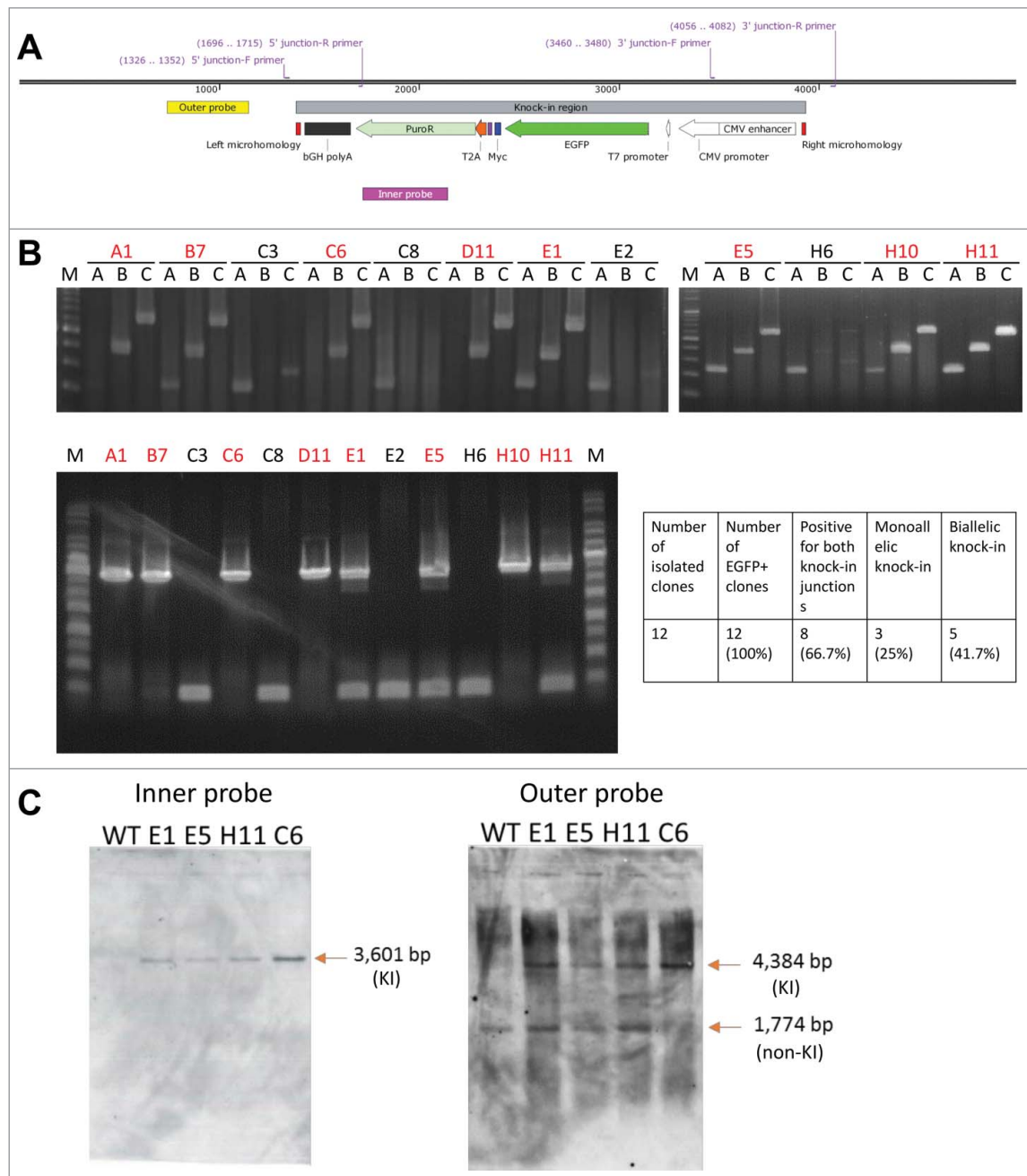
The lengths of left and right microhomologies can be independently set within the range of 10–40 bp.

### 5) PAM specificity

Although the typical PAM sequence of *Streptococcus pyogenes* Cas9 is 5′-NGG-3′, various PAM motifs have been identified and may be discovered in the future.<sup>9</sup> PITCh designer can accept any user-defined PAM sequences described with IUB codes.<sup>10</sup>

## PITCh designer algorithm can be implemented any website

While building PITCh designer, we created a JavaScript library, referred to as KnockinJS (<https://github.com/Kazuki-Nakamae/public/tree/master/KnockIn.js>). This library contains all of the JavaScript codes for

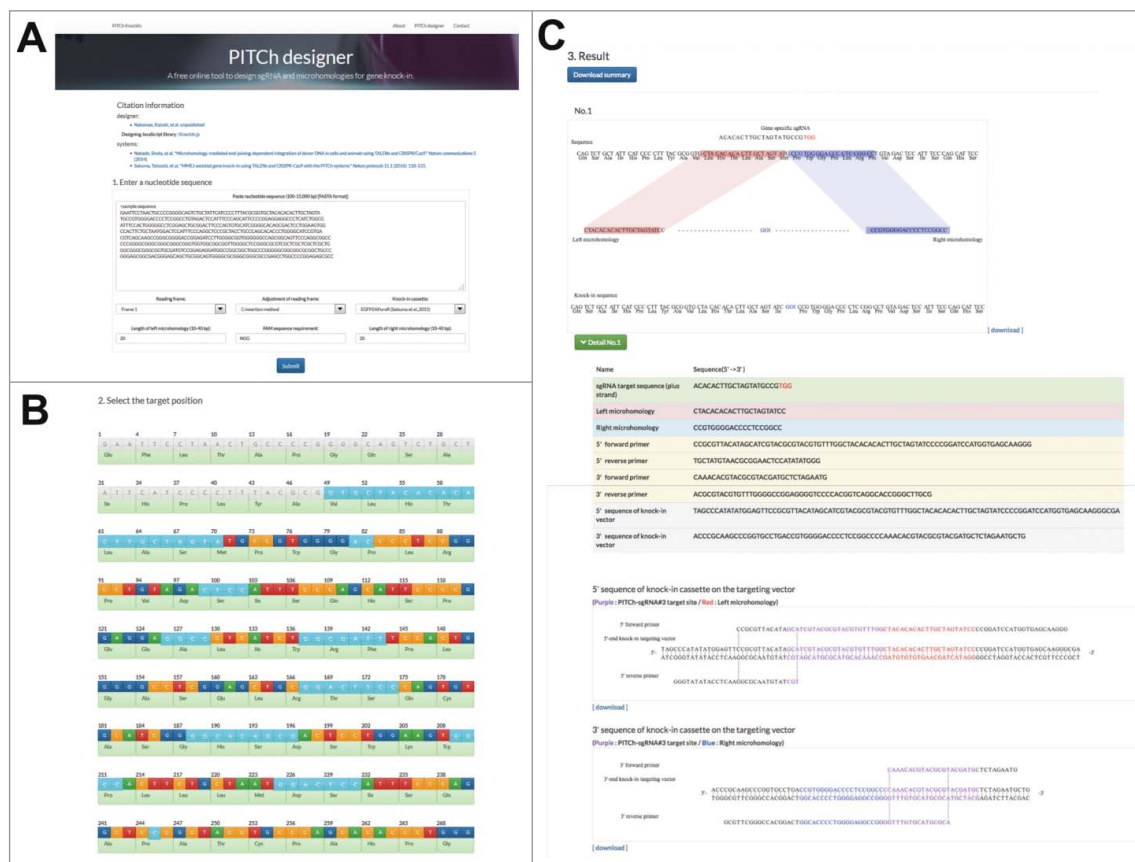


**Figure 3.** Experimental example of PITCh-TG. (A) Schematic illustration of knock-in allele. The genomic context was visualized using SnapGene Viewer software (Chicago, IL, USA) (<http://www.snapgene.com/>) with various annotations including primers used for genotyping and probes for Southern blotting. Red boxes indicate left and right microhomology regions. The cut region indicates a DNA fragment generated by restriction enzymes used in Southern blotting. (B) Gel images and summary of genotyping. The clone IDs are shown at the top of each panel. Red letters indicate the clones which the knock-in was successful. A, B, and C in the upper panels indicate the PCR products of non-knock-in allele, 5' junction, and 3' junction, respectively. The lower panel represents the results of the out-out PCR, simultaneously amplifying both the knock-in and the non-knock-in alleles. M, ladder marker. (C) Southern blotting. The results of inner and outer probes are shown. WT, wild type; KI, knock-in.

designing the PITCh knock-in. Anyone can implement the knock-in design algorithms used in PITCh designer on their own websites by adopting KnockinJS library in the HTML file. Furthermore, this portable JavaScript library has been written using pseudo-class-

based object-oriented programming (OOP) ([https://developer.mozilla.org/en/docs/Web/JavaScript/Introduction\\_to\\_Object-Oriented\\_JavaScript](https://developer.mozilla.org/en/docs/Web/JavaScript/Introduction_to_Object-Oriented_JavaScript)), enabling implementation in a wide range of class-based OOP languages, such as Java, C#, and Python.





**Figure 4.** Screenshots of PITCh designer, representing the sequence submission (A), selection of target bases (B), and the example of the design results (C).

In the field of biotechnology, computed and automated programs are currently used for various research purposes.<sup>11-13</sup> To make the most of these programs, it is important for them to be implemented in various forms, based on portable software systems. Our PITCh designer with the KnockinJS library is a good example of software portability in the field of genome editing.

## Materials and methods

### Vector construction

pX330A-AAVS1/PITCh, an all-in-one CRISPR-Cas9 vector for cutting the genomic *AAVS1* locus and the donor vector, was constructed using the Multiplex CRISPR/Cas9 Assembly System Kit (Addgene, Kit #1000000055) and pX330S-2-PITCh (Addgene, Plasmid #63670), in accordance with a described previously protocol.<sup>2</sup> The oligonucleotides for the template of sgRNA targeting *AAVS1* are listed in Supplemental Table 1. pCRIS-PITChv2-AAVS1, the donor vector for PITCh-knock-in, was

constructed using a standard PCR and In-Fusion cloning method (Takara).

### Cell culture, transfection, and puromycin selection

HCT116 cells were maintained in DMEM supplemented with 10% fetal bovine serum. Transfection, puromycin selection, and single cell cloning were performed as described previously<sup>2</sup> with slight modifications. Briefly,  $2.5 \times 10^5$  HCT116 cells were seeded in a 100-mm dish a day before transfection. The transfection was performed with 1.2  $\mu$ g each of pX330A-AAVS1/PITCh and pCRIS-PITChv2-AAVS1 using Lipofectamine LTX (Thermo Fisher Scientific). Puromycin selection was started at 72 h post-transfection, by supplementation with 1  $\mu$ g/mL puromycin. The culture medium was replaced with fresh medium containing puromycin on a daily basis. After a week of puromycin selection, single cell cloning was performed using a limiting dilution method with a 96-well plate, in accordance with the following protocol. The cells were trypsinized, collected, and adjusted to

7.5 cells/mL. Subsequently, 200  $\mu$ L of the suspended cells were moved to each well of a 96-well plate (1.5 cells/well).

### Genotyping by PCR

Genomic DNA was extracted using a DNeasy Blood & Tissue Kit (Qiagen). For genotyping, PCR amplification was performed using KOD FX Neo (Toyobo) or PrimeSTAR GXL (Takara) using the primers listed in Supplemental Table 1.

### Southern blotting

Southern blotting was performed as described previously.<sup>7</sup> The primers used to generate the inner and outer probes are listed in Supplemental Table 1.

### Disclosure of potential conflicts of interest

The authors report no conflict of interest.

### Acknowledgments

The authors wish to express their thanks to Dr. Yuki Naito at DBCLS for his useful suggestions on the development of the web tool.

### Funding

This study was supported in part by a Grant-in-Aid for Young Scientists (B) (16K18478) to T.S. and a Grant-in-Aid for JSPS Research Fellows (16J03164) to S.N. from the Japan Society for the Promotion of Science.

### ORCID

Shota Nakade  <http://orcid.org/0000-0003-4721-9408>

### References

- [1] Sakuma T, Woltjen K. Nuclease-mediated genome editing: At the front-line of functional genomics technology. *Dev Growth Differ* 2014; 56(1):2-13; PMID:24387662; <https://doi.org/10.1111/dgd.12111>
- [2] Sakuma T, Nakade S, Sakane Y, Suzuki KT, Yamamoto T. MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nat Protoc* 2016; 11(1):118-33; PMID:26678082; <https://doi.org/10.1038/nprot.2015.140>
- [3] Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, DeKaveler RC, Katibah GE, Amora R, Boydston EA, Zeitler B, et al. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol* 2009; 27(9):851-7; PMID:19680244; <https://doi.org/10.1038/nbt.1562>
- [4] Suzuki K, Tsunekawa Y, Hernandez-Benitez R, Wu J, Zhu J, Kim EJ, Hatanaka F, Yamamoto M, Araoka T, Li Z, et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* 2016; 540(7631):144-9; PMID:27851729; <https://doi.org/10.1038/nature20565>
- [5] Nakade S, Tsubota T, Sakane Y, Kume S, Sakamoto N, Obara M, Daimon T, Sezutsu H, Yamamoto T, Sakuma T, et al. Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat Commun* 2014; 5:5560; PMID:25410609; <https://doi.org/10.1038/ncomms6560>
- [6] Hisano Y, Sakuma T, Nakade S, Ohga R, Ota S, Okamoto H, Yamamoto T, Kawahara A. Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. *Sci Rep* 2015; 5:8841; PMID:25740433; <https://doi.org/10.1038/srep08841>
- [7] Sakuma T, Takenaga M, Kawabe Y, Nakamura T, Kamihira M, Yamamoto T. Homologous recombination-independent large gene cassette knock-in in CHO cells using TALEN and MMEJ-directed donor plasmids. *Int J Mol Sci* 2015; 16(10):23849-66; PMID:26473830; <https://doi.org/10.3390/ijms161023849>
- [8] Aida T, Nakade S, Sakuma T, Izu Y, Oishi A, Mochida K, Ishikubo H, Usami T, Aizawa H, Yamamoto T, et al. Gene cassette knock-in in mammalian cells and zygotes by enhanced MMEJ. *BMC Genomics* 2016; 17(1):979; PMID:27894274; <https://doi.org/10.1186/s12864-016-3331-9>
- [9] Nakade S, Yamamoto T, Sakuma T. Cas9, Cpf1 and C2c1/2/3—What's next? *Bioengineered* 2017; PMID:28140746; <https://doi.org/10.1080/21655979.2017.1282018>
- [10] Cornish-Bowden A. Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic Acids Res* 1985; 13(9):3021-30; PMID:2582368; <https://doi.org/10.1093/nar/13.9.3021>
- [11] Naito Y, Hino K, Bono H, Ui-Tei K. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* 2015; 31(7):1120-3; PMID:25414360; <https://doi.org/10.1093/bioinformatics/btu743>
- [12] Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* 2016; 34(2):184-91; PMID:26780180; <https://doi.org/10.1038/nbt.3437>
- [13] Pinello L, Canver MC, Hoban MD, Orkin SH, Kohn DB, Bauer DE, Yuan GC. Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol* 2016; 34(7):695-7; PMID:27404874; <https://doi.org/10.1038/nbt.3583>