

Article

# A Simple and Efficient CRISPR Technique for Protein Tagging

Fanning Zeng \*, Valerie Beck, Sven Schuierer, Isabelle Garnier, Carole Manneville, Claudia Agarinis, Lapo Morelli, Lisa Quinn, Judith Knehr, Guglielmo Roma, Frederic Bassilana and Mark Nash

Novartis Institutes for Biomedical Research, 4002 Basel, Switzerland; valerie.beck@novartis.com (V.B.); sven.schuierer@novartis.com (S.S.); isabelle.garnier@novartis.com (I.G.); carole.manneville@novartis.com (C.M.); claudia.agarinis@novartis.com (C.A.); lapo.morelli@novartis.com (L.M.); lisa.quinn@novartis.com (L.Q.); judith.knehr@novartis.com (J.K.); guglielmo.roma@novartis.com (G.R.); frederic.bassilana@novartis.com (F.B.); mark.nash@novartis.com (M.N.)

\* Correspondence: fanning.zeng@novartis.com; Tel.: +41-79-874-6670

Received: 21 October 2020; Accepted: 3 December 2020; Published: 5 December 2020



**Abstract:** Genetic knock-in using homology-directed repair is an inefficient process, requiring the selection of few modified cells and hindering its application to primary cells. Here, we describe Homology independent gene Tagging (HiTag), a method to tag a protein of interest by CRISPR in up to 66% of transfected cells with one single electroporation. The technique has proven effective in various cell types and can be used to knock in a fluorescent protein for live cell imaging, to modify the cellular location of a target protein and to monitor the levels of a protein of interest by a luciferase assay in primary cells.

**Keywords:** CRISPR knock in; non-homologous end-joining (NHEJ); protein tagging

## 1. Introduction

Genome editing has been revolutionized by the discovery of the CRISPR-Cas9 system. Cas9 nuclease creates a double-strand DNA break (DSB) at a target site specified by the guide RNA (gRNA) sequence [1,2]. In the presence of a donor DNA template, gene tagging can be achieved through homology-directed repair (HDR), the efficiency of which is low and not applicable for all cell types. Compared with HDR, non-homologous end-joining (NHEJ) is more efficient and is active in both proliferating and post-mitotic cells. Indeed, DSBs are preferentially repaired via the NHEJ pathway [3]. Although commonly used for gene knock-out by causing insertion or deletion (indel) mutations, NHEJ repair has been demonstrated to be intrinsically accurate and can also be used to create gene knock-ins [4–7].

The application of CRISPR knock-in in cell biology is hampered by its low efficiency and laborious procedure. The aim of this study was to develop a CRISPR knock-in protocol that is: (1) highly efficient; (2) easy to use; and (3) applicable to cell lines and primary cells. We chose electroporation as the delivery method for the gRNA/Cas9 nuclease/donor DNA complex, as it offers better genome editing efficiency [8–10]. We opted to use dsDNA without any homology arms as the donor and hypothesized that if dsDNA donors are available in abundance when and where DSB happens, they could be incorporated into the genome via NHEJ with high degrees of accuracy. With this strategy, we can efficiently tag and test the target proteins within a short timeframe. We name it Homology independent gene Tagging (HiTag).

## 2. Materials and Methods

**gRNAs.** We used the Alt-R CRISPR-Cas9 System from Integrated DNA Technologies (IDT) (Leuven, Belgium). The gRNAs were designed using the CRISPOR program [11] (<http://crispor.tefor.net/>) and are listed in Table S2. crRNA and tracrRNA were dissolved with Duplex buffer (IDT) to 169  $\mu\text{M}$  each. They were mixed at the equal volume, and the final RNA concentration was adjusted to 2  $\mu\text{g}/\mu\text{L}$  by adding Duplex buffer. The gRNA complex was formed by incubation at 95 °C for 5 min, followed by slowly cooling down to room temperature (RT), and then stored at  $-80$  °C.

**Cas9 nuclease.** Cas9 proteins were mostly produced in-house. Briefly, NLS-SpyCAS9-NLS-His6 was expressed in *E. coli* strain Rosetta 2 (DE3) in pET28a+, using 2xYT media plus 50  $\mu\text{g}/\text{mL}$  kanamycin overnight at 18 °C. Cas9 was then purified using Ni-NTA agarose (Qiagen, Germantown, MD, USA) capture/elution, followed by size exclusion chromatography on an S200 26/600 Superdex column (GE Lifesciences, Marlborough, MA, USA) using modified buffers [12] from Jinek et al. Then, Cas9 stocks (5.9  $\text{mg}/\text{mL}$ ) were aliquoted and stored at  $-80$  °C. Commercially available Cas9 nuclease (IDT) can also be used, with no difference observed in our hands (data not shown). When the HiTag protocol was tested in HCT116 cells stably expressing Cas9, the knock-in efficiency was equivalent to that by Cas9 proteins (data not shown).

**dsDNA donors.** Oligonucleotides were synthesized by Microsynth with PAGE purification. The forward and reverse strands were dissolved in Duplex buffer to 20  $\mu\text{M}$  each. To anneal the oligos, they were mixed at equal volume and incubated at 95 °C for 5 min and then allowed to cool down slowly to RT. The final concentration of the dsDNA stock is 10  $\mu\text{M}$ . Unless stated otherwise, final concentrations of dsDNA in the transfection mixtures are 750 nM.

**Transfection mixture.** For each transfection, we prepared 20  $\mu\text{L}$  of a cell/ribonucleoprotein/dsDNA mixture. Cells were trypsinized and washed once with phosphate-buffered saline (PBS) and then resuspended at the concentration of  $5.4 \times 10^6$  cells per 100  $\mu\text{L}$  buffer R (IDT). Ribonucleoprotein (RNP) was formed by incubating 4  $\mu\text{g}$  Cas9 (18.4 pmol or 0.68  $\mu\text{L}$ ) and 2  $\mu\text{g}$  gRNA (50 pmol or 1  $\mu\text{L}$ ) at RT for 10 min in a final volume of 5  $\mu\text{L}$  Duplex buffer. Then, 400,000 cells (in 7.5  $\mu\text{L}$  buffer R), 15 pmol dsDNA (1.5  $\mu\text{L}$ ) and 6  $\mu\text{L}$  buffer R were added to make up the transfection mixtures (Table S3).

For the knock-in of mScarlet, dsDNA donors were ordered as gBLOCK gene fragments (IDT). An amount of 500 ng DNA was dissolved in 5  $\mu\text{L}$  buffer R (IDT). The constitutions of the transfection mixtures are listed in Table S3.

**Electroporation.** The RNP and dsDNA templates were delivered into cells by electroporation with the Neon Transfection System (MPK5000) (ThermoFisher, Waltham, MA, USA). We used the 10  $\mu\text{L}$  kit (MPK1096) (ThermoFisher, Waltham, MA, USA), and transfection protocols for individual cell types are listed in Table S4. After transfection, cells were cultured in antibiotic-free medium for at least 24 h before replacing with normal growth medium.

**Cell culture.** All cells were maintained under a humidified atmosphere of 5%  $\text{CO}_2$  at 37 °C. A549, HCT116, HEK293, HeLa, SH-SY5Y and U-2 OS cell lines were purchased from ATCC and cultured according to the supplier's recommendation. The primary human skeletal muscle cells (CC-2561, Lonza, Basel, Switzerland) were maintained in Skeletal Muscle Growth Medium (Lonza, CC-3246), supplemented with 20% fetal calf serum (FCS).

**Schwann cells culture.** Animal experiments were performed in accordance with the Swiss ordinance on animal experimentation, after approval by Kantonales Veterinäramt Basel-Stadt. Primary Schwann cells were isolated as described previously by Kaewkhaw et al. [13] Briefly, sciatic nerves were isolated from adult rats and the epineurium was stripped. The tissues were teased with sharp tweezers and digested with 0.05% collagenase digestion at 37 °C for 2 h. The homogenous cell suspension was obtained by dissociating the tissue with a glass Pasteur pipette and filtered through a 40  $\mu\text{m}$  cell strainer. The cells were pelleted down and plated in poly-D-lysine-coated flasks. The cells were maintained in a custom-made D-valine DMEM medium (based on 10313039, L-Valine free, ThermoFisher, Waltham, MA, USA), supplemented with 10% FCS, B27, 10 mM Hepes, 2 mM glutamate and 5  $\mu\text{M}$  Forskolin. Once the cultures reach confluency, Schwann cells were dispatched by Dispase (Corning) digestion,

and they can be sub-cultured for up to 10 passages. The purities of Schwann cells were at least 95% at the time of transfection, judged by immunostaining with Schwann cell markers.

**Generation and maintenance of iPS cells.** Neonatal human dermal fibroblasts (ThermoFisher) were used for reprogramming using Sendai virus with the help of the CytoTune-iPS reprogramming kit according to the standard protocol. Colonies with the hallmark of a pluripotent morphology were readily visible between days 17 and 20 after transduction. These were picked and sub-cloned multiple times on plates coated with Matrigel (BD Biosciences, San Jose, CA, USA) in mTeSR1 medium (STEMCELL Technologies) until Sendai virus RNA could no longer be detected and the morphology looked stable. Pluripotency was controlled by FACS analysis. Potential for differentiation into the three germ layers was approved by using the TaqMan hPSC Scorecard Panel (ThermoFisher) according to the supplier's guideline. Karyotype analysis was performed by full-genome SNP analyses conducted with Life & Brain and showed no larger chromosomal aberrations. iPSCs were maintained on Matrigel-coated plates and grown in mTesR1.

**Generation and differentiation of iNgn2 iPSCs.** Human Ngn2 cDNA was synthesized using sequence information from the Ensembl database (Ensembl Gene ID ENSG00000178403) and cloned under the control of the TRE tight (Tetracycline Response Element) promoter in a PiggyBac/Tet-ON all-in-one vector [14]. This vector contains a CAG rtTA16 cassette allowing constitutive expression of the Tet-ON system and an Hsv-tkNeo cassette for generation of stable IPS clones. After trypsinization into single cells with Tryple express reagent (ThermoFisher), approximately  $1 \times 10^6$  iPS cells were nucleofected by an Amaxa nucleofector device using Human Stem Cell Nucleofector Kit 1 (Lonza, Basel, Switzerland) and Program B-016 with 4  $\mu\text{g}$  of Ngn2 plasmid and 1  $\mu\text{g}$  of the dual helper plasmid. Subsequently, cells were replated on Matrigel plates with mTeSR medium containing 10  $\mu\text{M}$  of Rock inhibitor. Antibiotic selection (G418 0.1 mg/mL) was applied 48 h later. Stable clones appeared within 1 week. To differentiate iNgn2 neurons,  $1 \times 10^6$  of iPS cells were plated on a 6 cm Matrigel plate in proliferation medium (DMEM/F12 with Glutamax supplemented with 2% B27 and 1% N2, 10 ng/mL hEGF, 10 ng/mL hFGF, 1% Pen/Strep (all from ThermoFisher) containing Rock inhibitor (10  $\mu\text{M}$ ) for 1 d and doxycycline (1  $\mu\text{g}/\text{mL}$ ) for 3 d. Three days later, cell-induced neurons were frozen down. Cells were thawed and recovered in differentiation medium for 24 h before transfection (Neurobasal supplemented with 2% B27, 1% N<sub>2</sub>, Pen/Strep, 1 mM Sodium Pyruvate, plus 10 ng/mL BDNF, GDNF and hNT3). To ensure the cells were post-mitotic at the time of electroporation, the cultures were treated with 10  $\mu\text{M}$  cytosine arabinoside 24 h before and after.

**Amplicon sequencing.** Total genomic DNA was isolated from transfected cells using a DNA extraction kit (Qiagen, DNeasy Blood and Tissue kit, Germantown, MD, USA). The region of theoretical insertion was amplified by PCR using primers hnRNPA2B1 fw (5'-tcccgtagcgagggtgctctcgcag) and hnRNPA2B1 re (5'-agctccgcagcctcgtcagagg). A 500 bp fragment could be obtained after 40 amplification cycles with denaturation at 95 °C for 30 s, annealing at 60 °C for 30 s and extension at 68 °C for 45 s using a proofreading Accuprime Pfx Mix (ThermoFisher). After gel purification, the PCR fragments were used to prepare sequencing libraries with the Ovation Low complexity Sequencing System (NuGEN, cat.9092-256). Libraries were sequenced in paired-end mode, at a read length of  $2 \times 300$  bp, using the MiSeq platform (Illumina, San Diego, CA, USA). We obtained 10,253,514 and 11,140,907 read pairs for the control and the CRISPR samples, respectively. Read quality was assessed by running FastQC (version 0.10) (Babraham Institute, Cambridge, UK) on the FASTQ files. Since the sequencing quality decreased continuously after base 200, we trimmed the reads to a length of 240 base pairs before alignment. The FASTQ files were aligned against an extended human reference genome (based on GRCh38) using BWA version 0.7.15 with default parameter settings [15]. The extension of the genome consisted of the addition of two artificial chromosomes which contained the insert in forward and reverse orientations concatenated with 220 bp of chromosome 7 before base 26,200,579 and 220 bp after base 26,200,580 on chromosome 7. The alignments were then processed by a custom script to compute the different alignment classes of fragments.

Western blot analysis. At 72 h after transfection, cells were washed with PBS and lysed in RIPA buffer (ThermoFisher) containing a protease inhibitor cocktail (Roche). Protein concentrations in the samples were determined using a colorimetric assay based on the Lowry method (DC Protein Assay, Biorad) following the manufacturer's instructions. Equal amounts (ranging from 5 to 10  $\mu$ g) of total protein extracts were loaded after a 10 min denaturation step at 70 °C on Nu-PAGE 4–12% Bis-Tris Gels (ThermoFisher) and separated by electrophoresis using MOPS buffer under reducing conditions. Proteins were transferred on a 0.2  $\mu$ m PVDF membrane by electroblotting (TransBlot Turbo system, BioRad). After one hour of room-temperature blocking of the membrane in Odyssey PBS-Blocking buffer (LI-COR), the commercial primary antibodies anti-V5 (ThermoFisher), anti-Vimentin (Abcam) and anti hnRNPA2B1 (Abcam) were diluted 1:2000 in fresh blocking buffer, added on the membrane and incubated overnight at 4 °C. Membranes were washed twice in TBS buffer containing 0.1% Tween 20 (TBST) and incubated with fluorescence-labeled secondary antibodies (Goat anti-Mouse IR Dye 680 RD, or Goat anti-Rabbit IR Dye 800CW, LI-COR) diluted 1:10,000 for 1 h at room temperature. After 3 wash steps in TBST, the labeled proteins were visualized using an Odyssey Infrared Imaging system (LI-COR, Lincoln, CA, USA).

Immunostaining, imaging and FACS sorting. Immunostaining was performed according to standard protocols. Briefly, 72 h after transfection, cells grown on 96-well plates were fixed and permeabilized for 20 min at 4 °C with Fixation/Permeabilization solution (BD Biosciences). Cells were washed 3 times in PBS and then blocked in PBS containing 5% donkey serum, 1% bovine serum albumin (BSA) and 10% Perm/Wash buffer (BD Biosciences) at RT for 1 h. For V5 tag staining, cells were incubated at RT for 4 h with Alexa647-conjugated anti-V5 antibody, followed by rinsing 3 times in PBS before imaging. For other antigens, cells were incubated overnight at 4 °C with primary antibodies as indicated (Table S5), washed three times with PBS and then incubated with 1:1000 Alexa Fluor-conjugated secondary antibodies (ThermoFisher) for 1 h at RT. The nuclei were visualized by counterstaining with NucBlue (ThermoFisher). Finally, the cells were rinsed a further 3 times in PBS before reading the plate on the Operetta Imaging System (PerkinElmer). The image analyses were conducted with the built-in Harmony software (PerkinElmer, Waltham, MA, USA).

For live cell imaging of mScarlet-Actin/Vimentin, HeLa cells were cultured on a 96-well plate for 7 days after transfection. The nuclei were stained with NucBlue Live ReadyProbe (ThermoFisher) and the culture media were replaced with FluoroBrite DMEM (ThermoFisher) supplemented with 3% FCS. The images were taken on Operetta and analyzed with Harmony software.

To enrich mScarlet-positive cells, HeLa cells were subjected to FACS sorting by a standard protocol. Briefly, cells were trypsinized and resuspended in PBS without  $\text{Ca}^{2+}/\text{Mg}^{2+}$ , with 2% FCS and 2 mM EDTA to obtain the single cell solutions ( $1 \times 10^6$  cells/mL). Cells were sorted on BD FACSAria Fusion, with a 561 nm laser for excitation and a 610/20 nm filter for emission. Positive cells were collected in a 6-well plate within the normal HeLa cell growth medium. Once the cells were confluent, they were split in a 96-well plate in FluoroBrite DMEM plus 3% FCS. Images were collected on Cell Voyager 7000 (Yokogawa, Tokyo, Japan) over 16 h, with one image every 10 min. The movie was generated with ImageJ.

Reverse transcription quantitative PCR. Total RNA was extracted from rat primary Schwann cells at the indicated time points with RNeasy (Qiagen) according to the manufacturer's instructions. RNA concentration was determined by using a Nanodrop and reverse transcription was performed with 300 ng of total RNA by the Reverse Transcription kit (Thermo Fisher Scientific, Waltham, MA, USA). Quantitative PCR (qPCR) reactions were carried out on a 7900 HT AB instrument with TaqMan universal buffer (Thermo Fisher Scientific) and primers master mix designed in-house and produced by Microsynth (Table S6). mRNA levels were normalized to levels of PpiB and ActB mRNA (Taqman probe from Thermo Fisher Scientific) in each sample.

HiBiT Nano-Glo assays. A palindrome-like HiBiT sequence (Table S1) was knocked into Oct-6 in rat SCs, and cells can be expanded for a further 5–6 passages before use. The cells were split in a 384-well plate with 10,000 cells per well. Cells were cultured for 3 days in FSK-free medium before FSK

was added to the wells at various concentrations. The expression level of Oct-6-HiBiT was determined with the NanoGlo HiBiT lytic assay kit (Promega, N3040). Briefly, Nano-Glo substrate and LgBiT protein were mixed with the lytic buffer (1:50 and 1:100, respectively). The volume of culture media was adjusted to 30  $\mu$ L each well and then an equal volume of reagents was added. After 10 min, luminescence signals were measured on an EnVision Plate Reader (PerkinElmer, Waltham, MA, USA).

A palindrome-like HiBiT sequence (Table S1) was knocked into IFIT1 in SkMCs, and cells were expanded for 10 days in the growth medium. The cells were split in a 384-well plate with 5000 cells per well. After cells attached, Interferon- $\beta$  (IF014, Merck, Kenilworth, NJ, USA) was added to the wells at various concentrations. The expression level of IFIT1-HiBiT was determined with the NanoGlo HiBiT lytic assay kit.

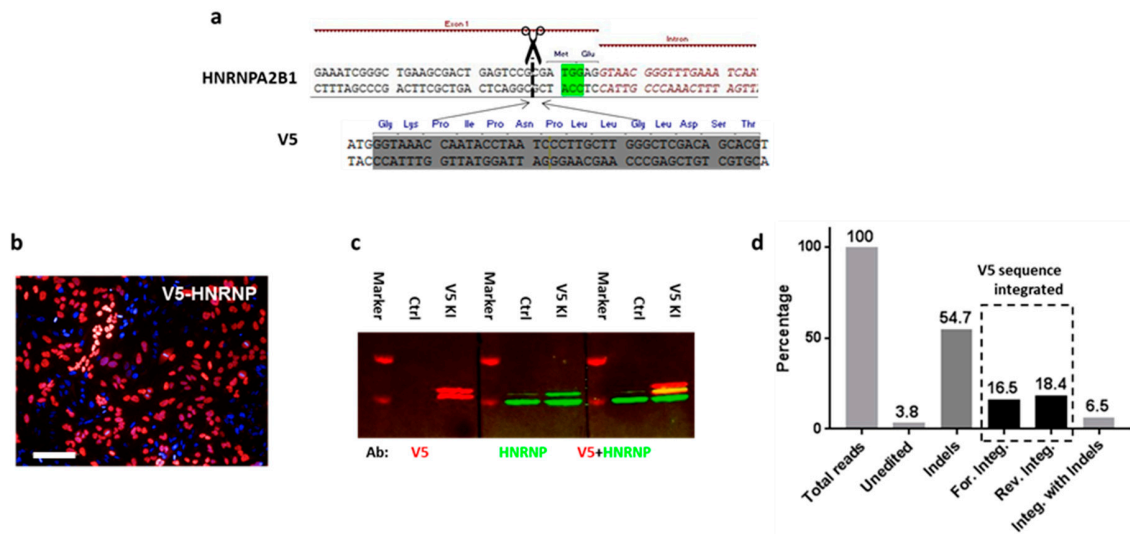
Statistical analysis. All statistical analyses were conducted using the GraphPad Prism. Data were expressed as the mean  $\pm$  standard deviation. Unpaired *t*-test is used in Figure S8.

### 3. Results

To increase the chance of target integration, we limited the size of donor DNA to 100 bp and thus focused initially on peptide tags. First, we aimed to label the heterogeneous nuclear ribonucleoprotein A2/B1 (HnRNP) in HeLa cells with a V5 tag. A gRNA was designed to target the 5' UTR (DSB at 2 bp before ATG) of HNRNPA2B1, and the dsDNA donor was synthesized as complementary oligonucleotides and annealed. To ensure that the V5 tag is translated in the frame, an ATG was added to its 5' end, as well as 1 bp to the 3' end (Figure 1a). Three days after transfection, knock-in efficiency was assessed by anti-V5 immunostaining:  $40 \pm 1.5\%$  of the cells displayed strong V5 signals that exclusively localized in the nucleus (Figure 1b). To further confirm the correct insertion of the V5 tag into the HnRNP, we performed Western blotting to check the size of the tagged proteins: two bands were visualized by the V5 antibody, with their sizes corresponding to the predicted chimeras (Figure 1c). Although this result already suggested that the V5 sequence was integrated into HNRNPA2B1 as intended, we went on further to confirm its correct location at nucleotide precision by amplicon sequencing.

To overcome the genomic instability issue of HeLa cells, we used HCT116 cells for NGS-based analysis of targeting events (Figures S1 and S2). We classified the reads from the amplicon sequencing into different alignment categories: 3.8% of the fragments aligned to the wildtype genome and contained no indels, whereas 54.7% of the fragments aligned to the wildtype genome and contained at least one indel. A total of 16.5% of the fragments aligned to the genome sequence containing the forward orientation of the insert without an indel and 18.4% of the fragments aligned to the genome sequence containing the 46 bp reverse insert without an indel. The forward or reverse direction of the alignment was determined with respect to the orientation of the gene HNRNPA2B1. A total of 6.5% of the fragments aligned to the genome sequence containing the 46 bp insert and at least one indel (Figure 1). The ratio of forward and reverse integrations of the V5 sequence is about 50:50, and only forward integration (16.5%) can result in a functional expression of the V5 tag.

To calculate the functional V5 tagging at the cellular level, we use the following equation:  $C = \{1 - (1 - g/100)^2\} \times 100$ , where *g* stands for the percentage of functional genome integration, which is 16.5 in this case; and *C* represents the percentage of functional V5 expression at the cellular level, which is 30.3. The number matches well with the immunostaining result (Figure S1), which suggests  $31.3 \pm 3.9\%$  of HCT116 cells express V5-HnRNP.

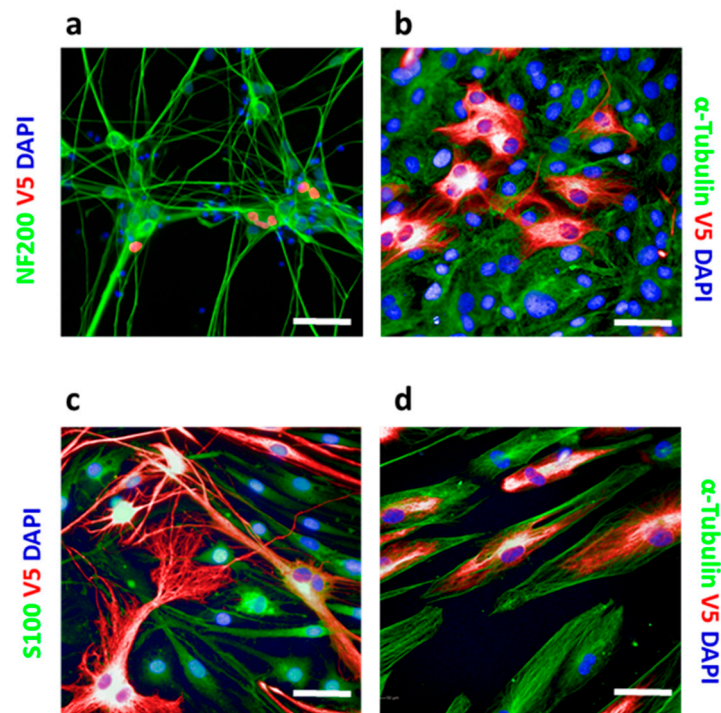


**Figure 1.** HiTag enables highly efficient CRISPR knock-in for imaging. (a) Sequences for HNRNPA2B1 (5' end) with the PAM site (green) and cleavage site (dotted line); V5 sequence with ATG and an extra base at 3'. (b) Immunostaining with V5 tagged HnRNP in HeLa cells ( $n = 6$ ). Scale bar = 100  $\mu$ m. (c) Lysates from untransfected or V5-HnRNP knock-in HeLa cells were subjected to Western blotting. Both V5 (left) and HNRNP (middle) antibodies visualized two bands in Western blots, corresponding to A2/B1 isoforms. V5-labeled fragments are slightly larger than those recognized by the HnRNP antibody (right). (d) The ratio of forward and reverse integrations of the V5 sequence into the HnRNP genomic sequence is about 50:50, and only forward integration (16.5%) can result in a functional expression of the V5 tag.

To test HiTag's efficiency with another target, we tagged Vimentin in HeLa cells. Not only was high knock-in efficiency (over 60%) achieved, but a clear overlap of V5 and Vimentin antibody staining was also observed (Figure S3a). Furthermore, V5-tagged Vimentin could be recognized by both antibodies in the Western blot (Figure S3b). Similarly,  $\beta$ -actin and a cytosolic/nucleoli protein (RLP10A) could also be tagged with the V5 epitope in HeLa cells (Figure S4). The method achieved similarly high efficiencies in other common cell lines (31.3~50.7%) (Figure S5). We also achieved positive knock-in in human induced pluripotent stem (iPS) with the HiTag method (Figure S6).

CRISPR knock-in is generally not applicable to post-mitotic cells since their HDR mechanism is inactive. However, by using iPS-derived iNgn2 neurons, we demonstrated that the HiTag approach also worked in these cells (Figure 2a). Primary cells can be problematic for CRISPR knock-in too, as they are not suitable for clonal selection. We extended the application of HiTag into primary cells: as shown in Figure 2b,c, Vimentin labeling could be achieved in  $23.9 \pm 0.6\%$  of primary rat tenocytes, as well as  $36.3 \pm 2.5\%$  of rat Schwann cells (SCs), while V5-tagged Sox10 was visualized in  $9.9 \pm 0.2\%$  of cells (Figure S7). Surprisingly, the HiTag protocol was even more efficient in primary human skeleton muscle cells (SkMC): V5-Vimentin signals could be observed in  $66.3 \pm 2.5\%$  of cells (Figure 2d).

Gene tagging has applications in cell biology far beyond just labeling a protein for immunostaining. For example, an endogenous protein can be tracked in live cells by tagging with fluorescent proteins [5]. We introduced a dsDNA fragment encoding mScarlet [16] into Vimentin or  $\beta$ -actin in HeLa cells via HiTag. Filament structures were observed in 0.6~1.5% of live cells after 7 days (Figure 3a), and mScarlet-positive cells were enriched by fluorescence-activated cell sorting. The subcellular location of mScarlet-Vimentin could then be tracked using time-lapse fluorescence imaging (Supplementary Video).

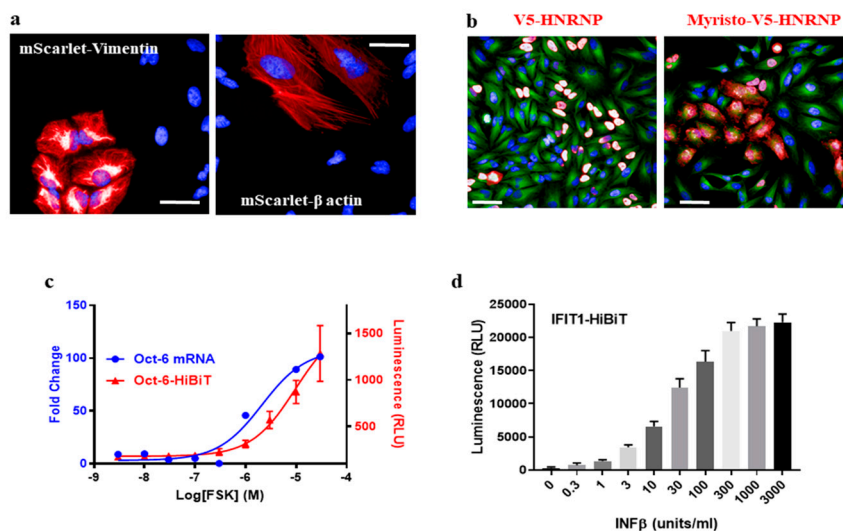


**Figure 2.** HiTag enables highly efficient CRISPR knock-in for imaging. (a) A total of  $5.9 \pm 1.6\%$  of iNgn2 cells display V5-HNRNP (red) signals. Cells were co-stained with neuronal marker NF200 (green) ( $n = 4$ ). (b) V5-Vimentin (red) in primary rat tenocytes, co-stained with  $\alpha$ -Tubulin (green) ( $n = 10$ ). (c) V5-Vimentin (red) in primary rat Schwann cells (SCs), co-stained with SC marker S100 (green) ( $n = 8$ ). (d) V5 knock-in Vimentin (red) in primary human skeleton muscle cells, co-stained with  $\alpha$ -Tubulin (green) ( $n = 4$ ). Scale bar = 50  $\mu\text{m}$ .

Other possible applications include the ability to modify subcellular protein localization, expression level or stability by knock-in signal peptides, promoters/Kozak sequence or degrons [17]. As a proof of concept, a myristoylation sequence was added to V5-HnRNP (Table S1) to disrupt its nucleus location: while  $96.9 \pm 0.9\%$  of control cells display V5 signals in nuclei,  $71.1 \pm 8.5\%$  of signals were cytosolic or on cell membranes when the myristoylation peptide was integrated (Figure 3b).

Given the high efficiency of the HiTag protocol, we wondered whether it was possible to develop functional assays in primary cells via CRISPR knock-in. Oct-6 is a key transcription factor in the myelination pathway in Schwann cells. Its mRNA expression is up-regulated upon stimulation of the cAMP pathway [18] (Figure 3c). To measure this up-regulation at the protein level, a HiBiT sequence (an 11-amino acid subunit of Nanoluc luciferase [19]) was attached to the C-terminus of the Oct-6 protein in rat primary Schwann cells by HiTag. After the cells were incubated with Forskolin (FSK), which increases the intracellular cAMP level, up-regulated Oct-6 protein expressions were evident in the Nanoluc assay (Figure 3c).

To further validate this approach, the HiBiT sequence was attached to the C-terminus of the IFIT1 protein in human SkMC. After incubation with IFN $\beta$ , which induces IFIT1 expression [5], a dose-dependent response was observed in the luciferase assay (Figure 3d). Therefore, we were able to develop a high-throughput format assay to measure endogenous protein expression in primary cells via CRISPR knock-in.



**Figure 3.** Multiple applications of HiTag in cell biology. (a) mScarlet was inserted into the N-terminus of Vimentin or the 5' UTR of  $\beta$ -actin in HeLa cells. Live cell fluorescence images were taken 7 days after transfection ( $n = 10$  each). Scale bar = 50  $\mu$ m. (b) V5 HNRNP (left) and Myristoylated V5-HNRNP (right) were visualized by V5 antibody staining (red) in HELA cells. Addition of myristoylation sequence targets the endogenous protein to the cell membrane ( $n = 5$ ). Cells were co-stained with  $\alpha$ -Tubulin (green). Scale bar = 50  $\mu$ m. (c) Primary rat SCs were incubated with FSK for 24 h, and Oct-6 mRNA expression was determined by qPCR (blue trace) ( $n = 3$ ). To measure Oct-6 at the protein level, the HiBiT sequence was inserted at its C-terminus. The cells were treated with FSK at various concentrations and subjected to NanoLuc assay after 48 h (red trace) ( $n = 4$ ). EC<sub>50</sub> for FSK was 2.1 and 9.6  $\mu$ M in qPCR and HiBiT assay, respectively. (d) IFIT1 was tagged with the HiBiT sequence at its C-terminus in primary human skeleton muscle cells (SkMC). The cells were treated with IFN $\beta$  at various concentrations and subjected to NanoLuc assay after 16 h ( $n = 10$ ). The calculated EC<sub>50</sub> for IFN $\beta$  was 27.5 unit/mL. The images shown are representative of the indicated number of separate experiments shown in brackets.

#### 4. Discussion

Here, we report HiTag as a simple and efficient protocol for the CRISPR knock-in method that can be applied to various targets in a wide range of cell types. Although its exact mechanism remains elusive, HiTag knock-in seems to involve DNA-dependent protein kinase (DNA-PK), as the DNA-PK inhibitor [20] abolished the knock-in effect. In contrast, SCR7, a commonly used NHEJ pathway inhibitor [21], did not block HiTag knock-in at all (Figure S8).

After cutting of the genomic DNA by Cas9, knock-in of dsDNA donors needs to compete with the self-ligation of free ends. It is conceivable that the more dsDNA is used in transfection, the higher the knock-in efficiency would be. Indeed, this is shown in Figure S9a. In some cell types, we observed cell toxicity issues when 1  $\mu$ M or higher concentrations of donor DNAs were used, especially when they are more than 100 bp in length (data not shown). An inverse correlation between knock-in efficiency and the length of dsDNA donors is also evident (Figure S9b). The efficiency was reduced even further (around 1%) when we tried to knock-in fluorescent proteins such as mScarlet (Figure 3a).

The knock-in efficiency leading to the expression of the desired tagged protein is probably also linked to the intrinsic properties of the cells: genome stability and accessibility, repair machinery accountable for appropriate integration and regulation of translation. We have observed more than 50% efficiency with V5 knock-in in HnRNP in HEK293 cells (Figure S5a), but only 31% in HCT116 cells (Figure S1a). Highlighted by our deep sequencing results, it is important to keep in mind that the efficiency of genome editing is not directly linked to accurate integration of dsDNA: the 96% modified HCT cells contained only 31% of the integration events, and amongst these, only half were in the needed orientation to get a tagged protein. These numbers will probably differ from one cell type



to another and can result in a variety of experimental outcomes. Interestingly, once dsDNA donors were incorporated, the chance of precise integration (~35%) was far higher than that with an indel (6.5%). This suggests the intrinsic accuracy of the NHEJ mechanism [4–7], and the observed high indel frequency without donor integration is probably the result of forced errors with the Cas9 presence.

Like other CRISPR knock-in methods, the successful application of HiTag depends on the targeting and cutting efficiency of the gRNA/Cas9 complex, and the detection of the tagged proteins heavily relies on their endogenous expression level. Issues such as off-target editing, monoallelic editing and indels will also confound the biological readout. Ideally, deep sequencing should be performed for each knock-in experiment. For cell lines, some of the problems could be addressed by the labor-intensive clone selection, whereas for primary cells, the potential impact of the “unwanted” editing needs to be considered when results are interpreted.

The HiTag approach also has its own intrinsic limitations: (1) The orientations of the integrated donor DNAs are non-controllable and might result in non-functional knock-in. If the tags are targeted to the C-terminus of a protein, this problem could be partially overcome by using palindromic sequences, as shown in Figure S7. However, use of longer palindromic donors would also reduce the knock-in efficiency. (2) Compared to HDR-based protocols, the HiTag method is more likely to result in “off-target” knock-in. Therefore, when designing their experiments, future users of the HiTag protocol should evaluate potential off-target sites of their sgRNA sequences with an off-target prediction program such as Cas-OFFinder (<http://www.rgenome.net/cas-offinder/>). (3) Compared to HDR-based knock-in protocols, HiTag is even more restricted by the availability of suitable gRNA binding sites for the insertion. For example, knock-in can be efficient within  $\pm 10$  bp around DSB sites for HDR-based protocols [22]; however, it can only happen at DSB sites for HiTag. For tagging the N-terminus of a protein, the insertion site could be either before or after ATG. If it is within the 5' UTR, one must consider that an additional sequence (between the DSB site and the original ATG) will be attached to the 3' end of dsDNA donors and might have unwanted consequences. For tagging at the C-terminus of a protein, the DSB site must be the 5' of the stop codon. Knock-in might result in the loss of a few residues at the C-terminus (encoded by the sequence between the DSB site and the stop codon), although this could be compensated by designing a dsDNA donor containing the deleted genomic sequence (with a mutated gRNA binding site). (4) The delivery of RNP into cells via electroporation is essential for HiTag, thus its throughput is rather low in the current format. However, electroporation devices are available in 96-well format as the alternative.

Despite being potentially less specific compared with some HDR knock-in methods [23,24], one apparent advantage of the HiTag method lies in its efficiency. While many published data have claimed high knock-in efficiency, the numbers quoted were obtained after selection or enrichment [5,25,26]. With the HiTag protocol, we have been able to achieve 30–60% efficiency (at the cellular level) without selection. This unselected cell population might actually be advantageous for certain experiments, as it minimizes some of the biases inherent to clone selection. However, if a uniform cell population is preferred, the high efficiency of HiTag might help facilitate the selection process.

The second benefit of our protocol lies in its ease and simplicity: while other NHEJ-based methods rely on the vector expression of gRNA and the Cas9 protein [5,6,25,26], which requires intensive and expensive laboratory work, all reagents used in HiTag methods are readily available from commercial suppliers. In particular, we realize that synthetic dsDNAs are far more cost-effective donors compared to PCR amplicons or plasmid fragments. Combined with the fact that clonal selection is not necessary with HiTag, experiments on target proteins could be finished within two weeks, from design to a functional test.

Lastly, the importance of our method also lies in its diverse application. Primary cells are generally not suitable for CRISPR knock-in. However, given the high efficiency of HiTag methods, we have established several high-throughput luciferase assays to monitor the expression of endogenous proteins in primary cells.

In summary, HiTag offers a highly efficient CRISPR knock-in strategy that can be quickly applied to many target genes and cell types. We expect the simplicity of this approach will encourage more studies to be performed on endogenous proteins in their native environments.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4409/9/12/2618/s1>, Figure S1: HiTag enables high-efficiency CRISPR knock-in of HCT116 cells; Figure S2: Genotyping result in HCT116 cells; Figure S3: V5-tagged Vimentin displays expected cellular localization and size; Figure S4: V5 tagging of  $\beta$ -actin and RLP10a in HeLa cells; Figure S5: HiTag efficiency in other common cell lines; Figure S6: HiTag efficiency in iPS cells; Figure S7: HiTag efficiency in primary Schwann cells; Figure S8: Effect of NU-7441 and SCR7 on HiTag knock-in efficiency; Figure S9: Effect of various parameters on HiTag knock-in efficiency; Video S1: mScarlet-Vimentin in HeLa cells. Table S1: Sequences and reagents; Table S2: gRNA sequences; Table S3: transfection mixture; Table S4: Electroporation protocol; Table S5: Antibodies.

**Author Contributions:** F.Z. designed the experiments and performed CRISPR knock-in and imaging; V.B. performed the Western blots and PCRs; S.S., J.K. and G.R. were involved in amplicon sequencing; I.G. performed HiBiT assay on Schwann cells; C.M. produced iPS and iNgn2 cells; C.A. performed FACS sorting; L.M. performed knock-in on T cells; L.Q. produced Cas9 protein; F.Z., G.R., F.B. and M.N. were involved in the writing and editing of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Cong, L.; Ran, F.A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P.D.; Wu, X.; Jiang, W.; Marraffini, L.A.; et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **2013**, *339*, 819–823. [[CrossRef](#)]
- Hsu, P.D.; Lander, E.S.; Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **2014**, *157*, 1262–1278. [[CrossRef](#)]
- Mao, Z.; Bozzella, M.; Seluanov, A.; Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair* **2008**, *7*, 1765–1771. [[CrossRef](#)]
- Suzuki, K.; Tsunekawa, Y.; Hernandez-Benitez, R.; Wu, J.; Zhu, J.; Kim, E.J.; Hatanaka, F.; Yamamoto, M.; Araoka, T.; Li, Z.; et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nat. Cell Biol.* **2016**, *540*, 144–149. [[CrossRef](#)]
- Lackner, D.H.; Carré, A.; Guzzardo, P.M.; Banning, C.; Mangena, R.; Henley, T.; Oberndorfer, S.; Gapp, B.V.; Nijman, S.M.; Brummelkamp, T.R.; et al. A generic strategy for CRISPR-Cas9-mediated gene tagging. *Nat. Commun.* **2015**, *6*, 10237. [[CrossRef](#)] [[PubMed](#)]
- He, X.; Tan, C.; Wang, F.; Wang, Y.; Zhou, R.; Cui, D.; You, W.; Zhao, H.; Ren, J.; Feng, B. Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res.* **2016**, *44*, e85. [[CrossRef](#)] [[PubMed](#)]
- Fei, J.-F.; Schuez, M.; Knapp, D.; Taniguchi, Y.; Drechsel, D.N.; Tanaka, E. Efficient gene knockin in axolotl and its use to test the role of satellite cells in limb regeneration. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 12501–12506. [[CrossRef](#)] [[PubMed](#)]
- Leonetti, M.D.; Sekine, S.; Kamiyama, D.; Weissman, J.S.; Huang, B. A scalable strategy for high-throughput GFP tagging of endogenous human proteins. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E3501–E3508. [[CrossRef](#)] [[PubMed](#)]
- Liang, X.; Potter, J.; Kumar, S.; Ravinder, N.; Chesnut, J.D. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *J. Biotechnol.* **2017**, *241*, 136–146. [[CrossRef](#)]
- Burger, A.; Lindsay, H.; Felker, A.; Hess, C.; Anders, C.; Chiavacci, E.; Zaugg, J.; Weber, L.M.; Catena, R.; Jinek, M.; et al. Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Development* **2016**, *143*, 2025–2037. [[CrossRef](#)]
- Haeussler, M.; Schönig, K.; Eckert, H.; Eschstruth, A.; Mianné, J.; Renaud, J.-B.; Schneider-Maunoury, S.; Shkumatava, A.; Teboul, L.; Kent, J.; et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* **2016**, *17*, 1–12. [[CrossRef](#)] [[PubMed](#)]
- Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J.A.; Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337*, 816–821. [[CrossRef](#)] [[PubMed](#)]

13. Kaewkhaw, R.; Scutt, A.M.; Haycock, J.W. Integrated culture and purification of rat Schwann cells from freshly isolated adult tissue. *Nat. Protoc.* **2012**, *7*, 1996–2004. [[CrossRef](#)] [[PubMed](#)]
14. Lacoste, A.; Berenshteyn, F.; Brivanlou, A.H. An Efficient and Reversible Transposable System for Gene Delivery and Lineage-Specific Differentiation in Human Embryonic Stem Cells. *Cell Stem Cell* **2009**, *5*, 332–342. [[CrossRef](#)]
15. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]
16. Bindels, D.S.; Haarbosch, L.; van Weeren, L.; Postma, M.; Wiese, K.E.; Mastop, M.; Aumonier, S.; Gotthard, G.; Royant, A.; Hink, M.A.; et al. mScarlet: A bright monomeric red fluorescent protein for cellular imaging. *Nat. Methods* **2017**, *14*, 53–56. [[CrossRef](#)]
17. Varshavsky, A. N-degron and C-degron pathways of protein degradation. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 358–366. [[CrossRef](#)]
18. Kawasaki, T.; Oka, N.; Tachibana, H.; Akiguchi, I.; Shibasaki, H. Oct6, a transcription factor controlling myelination, is a marker for active nerve regeneration in peripheral neuropathies. *Acta Neuropathol.* **2003**, *105*, 203–208. [[CrossRef](#)]
19. Dixon, A.S.; Schwinn, M.K.; Hall, M.P.; Zimmerman, K.; Otto, P.; Lubben, T.H.; Butler, B.L.; Binkowski, B.F.; Machleidt, T.; Kirkland, T.A.; et al. NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chem. Biol.* **2016**, *11*, 400–408. [[CrossRef](#)]
20. Robert, F.; Barbeau, M.; Ethier, S.D.; Dostie, J.; Pelletier, J. Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome Med.* **2015**, *7*, 1–11. [[CrossRef](#)]
21. Maruyama, T.; Dougan, S.K.; Truttmann, M.C.; Bilate, A.M.; Ingram, J.R.; Ploegh, H.L. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **2015**, *33*, 538–542. [[CrossRef](#)] [[PubMed](#)]
22. Paquet, D.; Kwart, D.; Chen, A.; Sproul, A.; Jacob, S.; Teo, S.; Olsen, K.M.; Gregg, A.; Noggle, A.S.S.J.S.; Tessier-Lavigne, M. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nat. Cell Biol.* **2016**, *533*, 125–129. [[CrossRef](#)] [[PubMed](#)]
23. Nakajima, K.; Zhou, Y.; Tomita, A.; Hirade, Y.; Gurumurthy, C.B.; Nakada, S. Precise and efficient nucleotide substitution near genomic nick via noncanonical homology-directed repair. *Genome Res.* **2017**, *28*, 223–230. [[CrossRef](#)]
24. Chen, X.; Janssen, J.M.; Liu, J.; Maggio, I.; Jong, A.E.J.T.; Mikkers, H.M.; Gonçalves, M.A.F.V. In trans paired nicking triggers seamless genome editing without double-stranded DNA cutting. *Nat. Commun.* **2017**, *8*, 1–15. [[CrossRef](#)]
25. Schmid-Burgk, J.L.; Höning, K.; Ebert, T.S.; Hornung, V. CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. *Nat. Commun.* **2016**, *7*, 12338. [[CrossRef](#)]
26. Sawatsubashi, S.; Joko, Y.; Fukumoto, S.; Matsumoto, T.; Sugano, S.S. Development of versatile non-homologous end joining-based knock-in module for genome editing. *Sci. Rep.* **2018**, *8*, 1–10. [[CrossRef](#)]

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).