Please cite this article in press as: Li et al., Precise Correction of the Dystrophin Gene in Duchenne Muscular Dystrophy Patient Induced Pluripotent Stem Cells by TALEN and CRISPR-Cas9, Stem Cell Reports (2015), http://dx.doi.org/10.1016/j.stemcr.2014.10.013

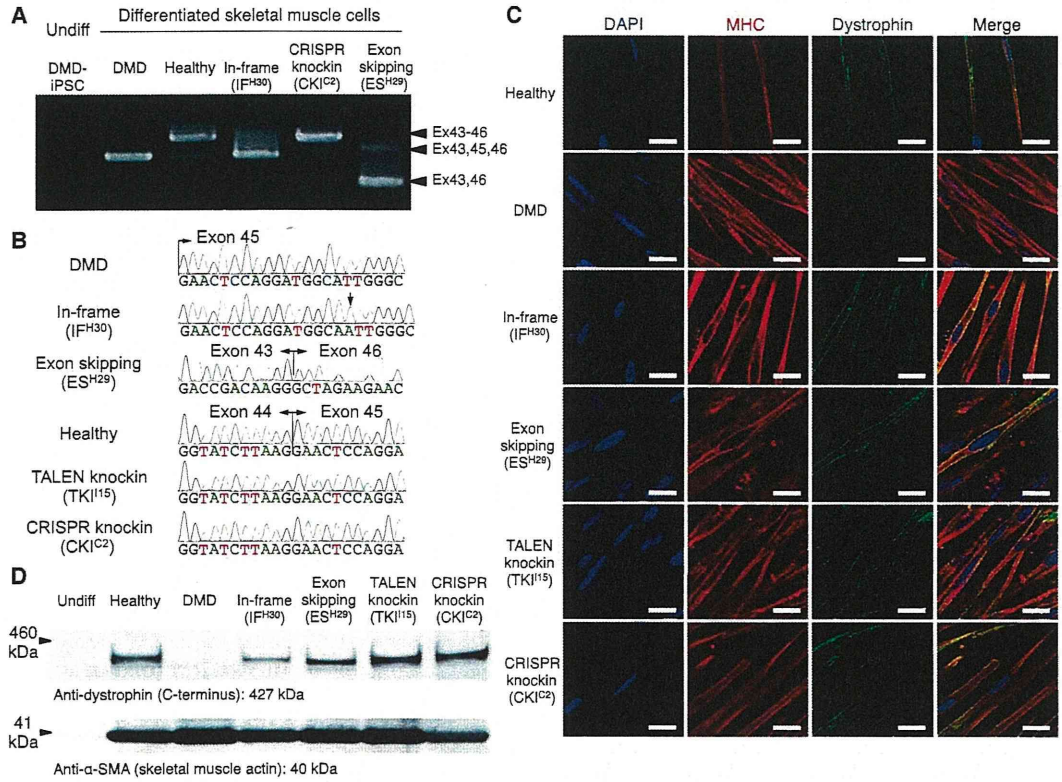

Figure 6. Restoration of the Dystrophin Protein in Differentiated Myogenic Cells

(A) RT-PCR analysis for dystrophin cDNA from iPSCs and cells differentiated from the corrected clones toward skeletal muscle lineage. The original DMD patient and an IF clone (IF^{H30}) with 1 bp insertion corresponded to the 452 and 453 bp PCR bands, respectively. The healthy control and knockin clones (TALEN-mediated TKI^{I15} and CRISPR-mediated CKI^{C2}) corresponded to the 600 bp bands; and the ES clone (ES^{H29}) corresponded to the 276 bp band.

(B) Sanger sequence analysis of dystrophin cDNAs from differentiated skeletal muscle cells. The IF clone (IF^{H30}) exhibited a 1 bp insertion (A, black arrow), the ES clone (ES^{H29}) exhibited the conjugation of exons 43 and 46 due to the skipping of exon 45, and knockin clones (TKI^{I15} and CKI^{C2}) exhibited the complete restoration of exon 44 in front of exon 45 as the healthy control.

(C) Immunofluorescence staining of skeletal muscle cells differentiated from the corrected clones. A z axis section of the confocal microscopy image shows submembrane localization of the dystrophin protein in the healthy control and all corrected clones, but not in the uncorrected original DMD iPSCs. The cells were stained by DAPI, a marker of skeletal differentiation (myosin heavy chain [MHC]), in red and an antibody that detects the rod domain of dystrophin (DYS1) in green. Scale bar, 50 μm.

(D) Western blot analysis to estimate the molecular weight of the dystrophin protein in the corrected clones. Expected molecular weight: 420 kDa for the reading-frame-corrected clone, 414 kDa for the exon-skipping clone, and 427 kDa for the exon 44 knockin clones and healthy control. An anti-dystrophin C terminus (amino acids 3661–3677) antibody was used to detect dystrophin protein, and an anti-α-SMA antibody was used as the sample loading control.

we did not detect a band in the original DMD clone. Together, our data indicate that genetically corrected iPSCs can express the dystrophin protein once they differentiate into myogenic cells.

DISCUSSION

Here, we have demonstrated that three distinct methods can correct the dystrophin gene: exon skipping, frameshifting, and exon knockin. All three approaches restored dystrophin protein expression in differentiated skeletal muscle cells. However, only the exon knockin approach restored the full-length dystrophin protein. We took advantage of the ability to expand iPSCs limitlessly and achieved a high percentage of knockin events by incorporating a drug selection system (up to 84% in the present study). Based on its precision and efficacy, we conclude that the knockin approach is preferable for correcting the dystrophin gene in iPSCs.

Please cite this article in press as: Li et al., Precise Correction of the Dystrophin Gene in Duchenne Muscular Dystrophy Patient Induced Pluripotent Stem Cells by TALEN and CRISPR-Cas9, Stem Cell Reports (2015), http://dx.doi.org/10.1016/j.stemcr.2014.10.013

Regarding the nuclease specificity, both TALENs and CRISPR-sgRNA can bind to DNA despite a few base mismatches (Hsu et al., 2013). Therefore, it is critical to target a unique region in the genome with a minimal number of off-targets, as otherwise multiple targets may be cleaved. Several web-based programs can be used to search for off-target sites with a given target sequence region (e.g., CRISPR Design Tool [Hsu et al., 2013], Cas-OFFinder [Bae et al., 2014b], and E-CRISP [Heigwer et al., 2014]). However, these programs provide the predicted number of off-target sites only within a small region (typically ~500 bp) at any given time. Our unique *k*-mer approach allows the visualization of targetable regions in the entire genome, so users can select the targetable region(s) before checking the number of off-targets with other programs.

The risk of off-target mutagenesis is one of the most important obstacles to the therapeutic use of programmable nucleases. We performed a T7EI assay and amplicon deep sequencing to detect rare mutations at the target site, but did not detect an increased mutation rate from our results. To further assess the risk of target-sequence-independent off-target mutations, we employed combinations of rigorous genome-wide mutation analyses, such as the G-band for karyotyping, SNP array for detecting CNVs, and exome sequencing for searching SNVs and small indels. Since none of these methods alone sufficiently covers the large spectrum of mutations (from the single-nucleotide level to the chromosome level), it is important to combine several methods before applying gene therapy.

To achieve a therapeutic effect with genetically corrected iPSCs for an autologous ex vivo gene therapy approach, we must still overcome several hurdles, such as the successful transplantation of iPSC-derived myogenic cells. Since MYOD1-induced muscle cells from iPSCs have the ability to fuse (Goudenege et al., 2012; Tanaka et al., 2013), a corrected copy of the dystrophin gene may be able to contribute to an entire myofiber. Moreover, for long-term repopulation, the differentiation of iPSCs toward muscle progenitor cells (i.e., satellite cells) could be ideal for restoring damaged muscle in DMD patients (Darabi et al., 2012). In addition, an immunogenic response to the newly corrected gene product is possible (Mendell et al., 2010), although the response may be hindered by transient immunosuppression.

In summary, we have demonstrated the restoration of the dystrophin protein in patient-derived iPSCs by three different approaches. TALEN and CRISPR were equally effective and had minimal effects on off-target mutagenesis when they were targeted to a unique sequence region. Our efficient and precise correction method using TALEN and CRISPR technologies should provide a framework for future ex vivo gene therapy using patient-specific human iPSCs.

EXPERIMENTAL PROCEDURES

Generation of Integration-Free DMD iPSCs

DMD fibroblasts were derived from a DMD patient lacking exon 44 of the dystrophin gene after the subject provided written informed consent. The use of patient-derived samples and the genomic analysis were approved by the Ethics Committee of Kyoto University (no. 824 and no. G259, respectively). DMD fibroblasts were cultured in Dulbecco's modified Eagle's medium supplemented with 5% fetal bovine serum. To generate integration-free DMD iPSCs, we transfected 6×10^5 DMD fibroblasts with three episomal vectors (pCXLE-hOCT3/4-shp53-F, pCXLE-hSK, and pCXLE-hUL) by Neon electroporation (1650 V, 10 ms, 3 pulses) as described previously (Okita et al., 2011). The iPSC colonies that emerged were picked up and plated onto 24-well plates with feeders on day 31 and then expanded.

To screen for iPSC clones that were negative for residual episomal vectors, iPSC pellets were lysed with 500 μl of lysis solution (200 μg ml^{-1} proteinase K) at 55°C for 3–16 hr. Genomic DNA was purified by phenol-chloroform extraction and ethanol precipitation, and then used for quantitative PCR analyses using the primers listed in Table S6. An episomal plasmid was used to determine the standard curve, and DMD-iPSC clones with fewer than 0.01 copies were deemed integration-free iPSC clones. The original DMD-iPSC #1 (clone ID: CiRA00111) and the corrected clones, including the IF clone IF^{D28} (clone ID: CiRA00111-IF-D28), ES clone ES^{H19} (clone ID: CiRA00111-ES-H19), TALEN-mediated knockin clone TKI^{I15} (clone ID: CiRA00111-TKI-I15), and CRISPR-mediated knockin clone CKI^{C2} (clone ID: CiRA00111-CKI-C2), will be available from the RIKEN BRC Cell Bank (cell no. HPS0383-HPS0387).

Unique *k*-mer Sequence Database

To identify unique sequence regions and avoid repeated sequences in the human genome, we generated all possible combinations of small *k*-mer sequences (≤16 bp) by a custom Perl script. We then mapped the *k*-mer sequences onto the human genome (hg19) using Bowtie (Langmead et al., 2009), with no mismatch allowed. Only uniquely mapped *k*-mer sequences were pooled as the data set. To visually show the stack of unique *k*-mer sequences, the mapping data were converted to the BEDGRAPH format by genomeCoverageBed and then converted to TDF format by igvtools or to the bigWig format by bedGraphToBigWig. The unique k-mer sequence (Unik) database will be available on our website (https://apps.cira.kyoto-u.ac.jp/igeats).

Transfection of TALEN and CRISPR into Human iPSCs

Target iPSCs were pretreated with a ROCK inhibitor (Y-27632; Sigma) at 10 μM for at least 1 hr before electroporation. The cells were washed with PBS and treated with CTK solution for 1–3 min at 37°C to remove feeders and then were washed with PBS twice. Next, the iPSCs were further dissociated into single cells by a 0.25% Trypsin solution for 5–8 min at 37°C and were neutralized with culture medium containing 10% fetal bovine serum. We electroporated 10 μg of nuclease-expressing plasmids (TALENs: 5 μg left and 5 μg right; CRISPR: 5 μg Cas9 and 5 μg sgRNA) and 5 μg donor plasmid (if applicable) into 1×10^6 cells using a NEPA 21 electroporator (poring pulse: pulse voltage, 125 V; pulse

width, 5 ms; pulse number, 2; Negagene). Cells were plated onto one well of a six-well plate with feeders in the presence of 10 μM Y-27632 for 1–2 days.

Analysis of Indel Patterns by Deep Sequencing

The dystrophin gene target region was PCR amplified with barcoding primers (DMD-MiSeq-Rd1-fwd1 and DMD-MiSeq-Rd2-rev1) and then adaptor primers (Multiplex P5 fwd and Multiplex P7 rev) using a high-fidelity PCR enzyme. The resultant PCR products were gel purified and quantified by a Qubit 2.0 Fluorometer (Life Technologies) and the KAPA Library Quantification Kit for Illumina (KAPA Biosystems). Each DNA sample was adjusted to 2 nM and denatured by 0.2 N NaOH solution for 5 min. The samples were further diluted to 12 pM, mixed with 4 pM of PhiX spike-in DNA, and run on MiSeq using the MiSeq Reagent Kit v2 for 2 × 150 bp sequencing. The generated FASTQ sequence files were filtered by the fastq_quality_filter program from the FASTX-Toolkit to remove low-quality sequencing reads. After removal of the PhiX sequences, the remaining sequencing reads were split based on the barcode indices by the fastx_barcode_splitter program. The resultant reads were mapped to the target sequences by BWA, and the mutation patterns were extracted from the CIGAR code and MD tag.

Frameshift Screening without a Template Donor

Genomic DNAs from the transfected iPSCs were analyzed by the T7EI assay and restriction enzyme (XcmI) sensitivity assay to monitor the efficiency of the nuclease-mediated mutagenesis. Then the cells were dissociated into single cells and diluted to 200–500 cells per 10 cm dish with feeders. The subclonal colonies that emerged were picked on days 11–13 after reseeding. From the genomic PCR sequencing, the indels at dystrophin exon 45 with the $(3n + 2)$ bp deletion or $(3n + 1)$ bp insertion (where n is a nonnegative integer) were further expanded for later experiments.

TALEN- or CRISPR-Mediated Exon 44 Knockin

For the knockin experiment, 5 μg of donor vector was cotransfected with TALEN expression vectors (5 μg for left TALEN and 5 μg for right TALEN) or Cas9 and sgRNA expression vectors (5 μg for Cas9 and 5 μg for sgRNA) using NEPA 21 as described above. Hygromycin B (25 μg ml^{-1}; Invitrogen) selection was applied after iPSC colonies were recovered (4–5 days after transfection). The resulting hygromycin-resistant colonies were dissociated into single cells and plated at 200–500 cells per 10 cm dish with feeders. Each subclone was screened by genomic PCR (with P1-P2 primer pairs, amplifying a fragment from upstream of the 3′ arm to the EF1α-promoter, and P3-P4 primer pairs, amplifying a fragment from exon 44 to downstream of the 5′ arm). Homologous recombinants were further verified by Southern blot analysis using EcoRI and probe intron 45. After establishing the single-copy knockin clones, we electroporated the cells with 10 μg of the *Cre* expression vector pCXW-Cre-puro using NEPA 21. Clone isolation was carried out as described above, and excision of the hygromycin-selection cassette was confirmed by PCR screening with primers P1 and P4 and Southern blot analysis with EcoRI digestion and probe intron 45.

Skeletal Muscle Differentiation by Dox-Inducible MYOD1

The induction of skeletal muscle differentiation from iPSCs was described previously (Tanaka et al., 2013). Briefly, a Dox-inducible MYOD1-expressing *piggyBac* vector, PB-TetO-MyoD, was coelectroporated with the *piggyBac* transposase vector PBaseII (Matsui et al., 2014) using NEPA 21 (125 V, 5 ms). G418 (Calbiochem) selection (100 μg ml^{-1}) was applied to select stable PB-TetO-MyoD clones. Among the several G418-resistant clones, we screened for clones with a high mCherry induction rate upon addition of 1 μg ml^{-1} doxycycline (Funakoshi). Successful differentiation was confirmed by a spindle-shape-like morphology and immunocytochemical staining with myosin heavy chain (MHC) and α-skeletal muscle actin (α-SMA) antibodies on day 9 postdifferentiation.

ACCESSION NUMBERS

The plasmid DNAs used in this study are available from Addgene (https://www.addgene.org/) under accession numbers 60599–60605. DMD-patient-derived iPSCs and genetically corrected subclones are available from the RIKEN BRC Cell Bank (http://www.brc.riken.jp/lab/cell/english/) under accession numbers HPS0383–HPS0387.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2014.10.013.

AUTHOR CONTRIBUTIONS

H.L.L., S.Y., and A.H. conceived and designed the project. H.L.L., N.F., N.S., S.S., T.O., N.A, A.W., and A.H. performed the experiments. T.S. and T.Y. provided the TALEN construction platforms. M.T. constructed the website. H.S. provided the skeletal muscle differentiation protocol. H.L.L. and A.H. interpreted the data and wrote the manuscript.

ACKNOWLEDGMENTS

We thank the anonymous DMD patient and his family for kindly providing the biopsy. We also thank Dr. Megumu K. Saito and Takayuki Tanaka for preparing the cell samples; Dr. Kazutoshi Takahashi, Dr. Yusuke Echigoya, and Emi Shoji for their technical advice; Dr. Peter Karagiannis for a critical reading of the manuscript; Dr. Knut Woltjen and Dr. Keisuke Okita for providing vectors; Yumie Tokunaga for the SNP array analysis; and Osamu Ohta for computational advice. This research was supported in part by JSPS KAKENHI, JST PRESTO, JST Yamanaka iPS Cell Special Project, and the JST Research Center Network for Realization of Regenerative Medicine. H.L.L. is a recipient of a JSPS DC1 fellowship. S.Y. is a scientific advisor of iPS Academia Japan without salary.

Received: September 4, 2014
Revised: October 24, 2014
Accepted: October 24, 2014
Published: November 26, 2014

Please cite this article in press as: Li et al., Precise Correction of the Dystrophin Gene in Duchenne Muscular Dystrophy Patient Induced Pluripotent Stem Cells by TALEN and CRISPR-Cas9, Stem Cell Reports (2015), http://dx.doi.org/10.1016/j.stemcr.2014.10.013

REFERENCES

Aartsma-Rus, A., Fokkema, I., Verschuuren, J., Ginjaar, I., van Deutekom, J., van Ommen, G.J., and den Dunnen, J.T. (2009). Theoretic applicability of antisense-mediated exon skipping for Duchenne muscular dystrophy mutations. Hum. Mutat. *30*, 293–299.

Amps, K., Andrews, P.W., Anyfantis, G., Armstrong, L., Avery, S., Baharvand, H., Baker, J., Baker, D., Munoz, M.B., Beil, S., et al.; International Stem Cell Initiative (2011). Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. Nat. Biotechnol. *29*, 1132–1144.

Bae, S., Kweon, J., Kim, H.S., and Kim, J.S. (2014a). Microhomology-based choice of Cas9 nuclease target sites. Nat. Methods *11*, 705–706.

Bae, S., Park, J., and Kim, J.S. (2014b). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics *30*, 1473–1475.

Canver, M., Bauer, D., Dass, A., Yien, Y., Chung, J., Masuda, T., Maeda, T., Paw, B., and Orkin, S. (2014). Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. J. Biol. Chem. *289*, 21312–21324.

Choi, S.M., Kim, Y., Shim, J.S., Park, J.T., Wang, R.H., Leach, S.D., Liu, J.O., Deng, C., Ye, Z., and Jang, Y.Y. (2013). Efficient drug screening and gene correction for treating liver disease using patient-specific stem cells. Hepatology *57*, 2458–2468.

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

Darabi, R., Arpke, R.W., Irion, S., Dimos, J.T., Grskovic, M., Kyba, M., and Perlingeiro, R.C. (2012). Human ES- and iPS-derived myogenic progenitors restore DYSTROPHIN and improve contractility upon transplantation in dystrophic mice. Cell Stem Cell *10*, 610–619.

Ding, Q., Lee, Y.K., Schaefer, E.A., Peters, D.T., Veres, A., Kim, K., Kuperwasser, N., Motola, D.L., Meissner, T.B., Hendriks, W.T., et al. (2013a). A TALEN genome-editing system for generating human stem cell-based disease models. Cell Stem Cell *12*, 238–251.

Ding, Q., Regan, S.N., Xia, Y., Oostrom, L.A., Cowan, C.A., and Musunuru, K. (2013b). Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. Cell Stem Cell *12*, 393–394.

Filareto, A., Parker, S., Darabi, R., Borges, L., Iacovino, M., Schaaf, T., Mayerhofer, T., Chamberlain, J.S., Ervasti, J.M., McIvor, R.S., et al. (2013). An ex vivo gene therapy approach to treat muscular dystrophy using inducible pluripotent stem cells. Nat. Commun. *4*, 1549.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. *31*, 822–826.

Goudenege, S., Lebel, C., Huot, N.B., Dufour, C., Fujii, I., Gekas, J., Rousseau, J., and Tremblay, J.P. (2012). Myoblasts derived from normal hESCs and dystrophic hiPSCs efficiently fuse with existing muscle fibers following transplantation. Mol. Ther. *20*, 2153–2167.

Heigwer, F., Kerr, G., and Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. Nat. Methods *11*, 122–123.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C., et al. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nat. Biotechnol. *29*, 731–734.

Hotta, A., Cheung, A.Y., Farra, N., Vijayaragavan, K., Séguin, C.A., Draper, J.S., Pasceri, P., Maksakova, I.A., Mager, D.L., Rossant, J., et al. (2009). Isolation of human iPS cells using EOS lentiviral vectors to select for pluripotency. Nat. Methods *6*, 370–376.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

Koehler, R., Issac, H., Cloonan, N., and Grimmond, S.M. (2011). The uniqueome: a mappability resource for short-tag sequencing. Bioinformatics *27*, 272–274.

Langmead, B., Schatz, M.C., Lin, J., Pop, M., and Salzberg, S.L. (2009). Searching for SNPs with cloud computing. Genome Biol. *10*, R134.

Laurent, L.C., Ulitsky, I., Slavin, I., Tran, H., Schork, A., Morey, R., Lynch, C., Harness, J.V., Lee, S., Barrero, M.J., et al. (2011). Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. Cell Stem Cell *8*, 106–118.

Li, H.L., Nakano, T., and Hotta, A. (2014). Genetic correction using engineered nucleases for gene therapy applications. Dev. Growth Differ. *56*, 63–77.

Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M., Vertino, P.M., Stewart, F.J., and Bao, G. (2014). CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucleic Acids Res. *42*, 7473–7485.

Ma, N., Liao, B., Zhang, H., Wang, L., Shan, Y., Xue, Y., Huang, K., Chen, S., Zhou, X., Chen, Y., et al. (2013). Transcription activator-like effector nuclease (TALEN)-mediated gene correction in integration-free β-thalassemia induced pluripotent stem cells. J. Biol. Chem. *288*, 34671–34679.

Maetzel, D., Sarkar, S., Wang, H., Abi-Mosleh, L., Xu, P., Cheng, A.W., Gao, Q., Mitalipova, M., and Jaenisch, R. (2014). Genetic and chemical correction of cholesterol accumulation and impaired autophagy in hepatic and neural cells derived from Niemann-Pick Type C patient-specific iPS cells. Stem Cell Rep. *2*, 866–880.

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

Matsui, H., Fujimoto, N., Sasakawa, N., Ohinata, Y., Shima, M., Yamanaka, S., Sugimoto, M., and Hotta, A. (2014). Delivery of full-length factor VIII using a piggyBac transposon vector to correct a mouse model of hemophilia A. PLoS ONE *9*, e104957.

Mendell, J.R., Campbell, K., Rodino-Klapac, L., Sahenk, Z., Shilling, C., Lewis, S., Bowles, D., Gray, S., Li, C., Galloway, G., et al. (2010).

Dystrophin immunity in Duchenne's muscular dystrophy. N. Engl. J. Med. *363*, 1429–1437.

Okada, T., and Takeda, S. (2013). Current challenges and future directions in recombinant AAV-mediated gene therapy of Duchenne muscular dystrophy. Pharmaceuticals (Basel) *6*, 813–836.

Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K., et al. (2011). A more efficient method to generate integration-free human iPS cells. Nat. Methods *8*, 409–412.

Osborn, M.J., Starker, C.G., McElroy, A.N., Webber, B.R., Riddle, M.J., Xia, L., DeFeo, A.P., Gabriel, R., Schmidt, M., von Kalle, C., et al. (2013). TALEN-based gene correction for epidermolysis bullosa. Mol. Ther. *21*, 1151–1159.

Ousterout, D.G., Perez-Pinera, P., Thakore, P.I., Kabadi, A.M., Brown, M.T., Qin, X., Fedrigo, O., Mouly, V., Tremblay, J.P., and Gersbach, C.A. (2013). Reading frame correction by targeted genome editing restores dystrophin expression in cells from Duchenne muscular dystrophy patients. Mol. Ther. *21*, 1718–1726.

Park, I.H., Arora, N., Huo, H., Maherali, N., Ahfeldt, T., Shimamura, A., Lensch, M.W., Cowan, C., Hochedlinger, K., and Daley, G.Q. (2008). Disease-specific induced pluripotent stem cells. Cell *134*, 877–886.

Pichavant, C., Aartsma-Rus, A., Clemens, P.R., Davies, K.E., Dickson, G., Takeda, S., Wilton, S.D., Wolff, J.A., Wooddell, C.I., Xiao, X., and Tremblay, J.P. (2011). Current status of pharmaceutical and genetic therapeutic approaches to treat DMD. Mol. Ther. *19*, 830–840.

Popplewell, L., Koo, T., Leclerc, X., Duclert, A., Mamchaoui, K., Gouble, A., Mouly, V., Voit, T., Pâques, F., Cédrone, F., et al. (2013). Gene correction of a duchenne muscular dystrophy mutation by meganuclease-enhanced exon knock-in. Hum. Gene Ther. *24*, 692–701.

Rousseau, J., Chapdelaine, P., Boisvert, S., Almeida, L.P., Corbeil, J., Montpetit, A., and Tremblay, J.P. (2011). Endonucleases: tools to correct the dystrophin gene. J. Gene Med. *13*, 522–537.

Sakuma, T., Hosoi, S., Woltjen, K., Suzuki, K., Kashiwagi, K., Wada, H., Ochiai, H., Miyamoto, T., Kawai, N., Sasakura, Y., et al. (2013a). Efficient TALEN construction and evaluation methods for human cell and animal applications. Genes Cells *18*, 315–326.

Sakuma, T., Ochiai, H., Kaneko, T., Mashimo, T., Tokumasu, D., Sakane, Y., Suzuki, K., Miyamoto, T., Sakamoto, N., Matsuura, S., and Yamamoto, T. (2013b). Repeating pattern of non-RVD variations in DNA-binding modules enhances TALEN activity. Sci. Rep. *3*, 3379.

Sebastiano, V., Maeder, M.L., Angstman, J.F., Haddad, B., Khayter, C., Yeo, D.T., Goodwin, M.J., Hawkins, J.S., Ramirez, C.L., Batista, L.F., et al. (2011). In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. Stem Cells *29*, 1717–1726.

Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., Wang, Y., Brodsky, R.A., Zhang, K., Cheng, L., and Ye, Z. (2014). Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. Cell Stem Cell *15*, 12–13.

Soldner, F., Laganière, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S., et al. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell *146*, 318–331.

Sugiura, M., Kasama, Y., Araki, R., Hoki, Y., Sunayama, M., Uda, M., Nakamura, M., Ando, S., and Abe, M. (2014). Induced pluripotent stem cell generation-associated point mutations arise during the initial stages of the conversion of these cells. Stem Cell Rep. *2*, 52–63.

Suzuki, K., Yu, C., Qu, J., Li, M., Yao, X., Yuan, T., Goebl, A., Tang, S., Ren, R., Aizawa, E., et al. (2014). Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specific induced pluripotent stem cell clones. Cell Stem Cell *15*, 31–36.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Tanaka, A., Woltjen, K., Miyake, K., Hotta, A., Ikeya, M., Yamamoto, T., Nishino, T., Shoji, E., Sehara-Fujisawa, A., Manabe, Y., et al. (2013). Efficient and reproducible myogenic differentiation from human iPS cells: prospects for modeling Miyoshi Myopathy in vitro. PLoS ONE *8*, e61540.

Tuffery-Giraud, S., Béroud, C., Leturcq, F., Yaou, R.B., Hamroun, D., Michel-Calemard, L., Moizard, M.P., Bernard, R., Cossée, M., Boisseau, P., et al. (2009). Genotype-phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD-DMD database: a model of nationwide knowledgebase. Hum. Mutat. *30*, 934–945.

Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Talkowski, M.E., and Musunuru, K. (2014). Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. Cell Stem Cell *15*, 27–30.

Ye, L., Wang, J., Beyer, A.I., Teque, F., Cradick, T.J., Qi, Z., Chang, J.C., Bao, G., Muench, M.O., Yu, J., et al. (2014). Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Δ32 mutation confers resistance to HIV infection. Proc. Natl. Acad. Sci. USA *111*, 9591–9596.

Zou, J., Mali, P., Huang, X., Dowey, S.N., and Cheng, L. (2011a). Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. Blood *118*, 4599–4608.

Zou, J., Sweeney, C.L., Chou, B.K., Choi, U., Pan, J., Wang, H., Dowey, S.N., Cheng, L., and Malech, H.L. (2011b). Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. Blood *117*, 5561–5572.

nature COMMUNICATIONS

ARTICLE

Received 26 Jun 2014 | Accepted 11 Oct 2014 | Published 20 Nov 2014

DOI: 10.1038/ncomms6560 OPEN

Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9

Shota Nakade[1,*], Takuya Tsubota[2,*], Yuto Sakane[1,*], Satoshi Kume[1], Naoaki Sakamoto[1], Masanobu Obara[3], Takaaki Daimon[4], Hideki Sezutsu[2], Takashi Yamamoto[1], Tetsushi Sakuma[1] & Ken-ichi T. Suzuki[1]

Genome engineering using programmable nucleases enables homologous recombination (HR)-mediated gene knock-in. However, the labour used to construct targeting vectors containing homology arms and difficulties in inducing HR in some cell type and organisms represent technical hurdles for the application of HR-mediated knock-in technology. Here, we introduce an alternative strategy for gene knock-in using transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated 9 (Cas9) mediated by microhomology-mediated end-joining, termed the PITCh (Precise Integration into Target Chromosome) system. TALEN-mediated PITCh, termed TAL-PITCh, enables efficient integration of exogenous donor DNA in human cells and animals, including silkworms and frogs. We further demonstrate that CRISPR/Cas9-mediated PITCh, termed CRIS-PITCh, can be applied in human cells without carrying the plasmid backbone sequence. Thus, our PITCh-ing strategies will be useful for a variety of applications, not only in cultured cells, but also in various organisms, including invertebrates and vertebrates.

[1] Department of Mathematical and Life Sciences, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan. [2] Transgenic Silkworm Research Unit, National Institute of Agrobiological Sciences, 1-2 Owashi, Tsukuba, Ibaraki 305-8634, Japan. [3] Department of Biological Science, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan. [4] Insect Growth Regulation Research Unit, National Institute of Agrobiological Sciences, 1-2 Owashi, Tsukuba, Ibaraki 305-8634, Japan. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to T.S. (email: tetsushi-sakuma@hiroshima-u.ac.jp) or to K.T.S. (email: suzuk107@hiroshima-u.ac.jp).

Programmable nucleases, such as transcription activator-like effector nucleases (TALENs) and RNA-guided endonucleases, that is, clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated 9 (Cas9), have been used widely for genetic engineering, including gene knock-out, knock-in and various chromosomal rearrangements[1,2]. Gene knock-in has generally been achieved by co-introduction of programmable nucleases and single-stranded oligonucleotides[3,4] or a targeting vector harbouring left and right homology arms[5,6], inducing homologous recombination (HR)-dependent gene addition. Although HR-mediated gene knock-in allows precise insertion of large DNA fragments, construction of targeting vectors is often laborious and targeting efficiency depends on the substantial variation in the frequency of HR induction among cell types and organism species. However, the addition of complementary overhangs to donor DNA fragments or simple linearization of donor DNA plasmids has been shown to facilitate targeted integration mediated by non-homologous end-joining both in cultured cells[7,8] and in zebrafish[9]. Obligate ligation-gated recombination has also reportedly been able to integrate plasmid DNA into a targeted genomic locus[10]. These methods use programmable nucleases to make a DNA double-strand break (DSB) that leaves 5′ overhangs (zinc finger nucleases (ZFNs) and TALENs) or blunt ends (CRISPR/Cas9), and then rely on the ligation of similar ends on the chromosomal target site and the insert. These targeted integrations can thus be considered to represent 'simple ligation'.

Conversely, microhomology-mediated end-joining (MMEJ)-dependent mutations have frequently been found in programmable nuclease-mediated gene disruption without exogenous donors[11]. MMEJ is a DSB repair mechanism that uses microhomologous sequences (5–25 bp) for error-prone end-joining[12]. In the cell cycle, MMEJ repair is active during G1/early S phases, whereas HR is active during late S/G2 phases[13]. Therefore, we devised a novel MMEJ-mediated gene knock-in strategy, referred to as the PITCh (Precise Integration into Target Chromosome) system, which enables efficient targeted integration of large DNA fragments in a wide range of cells and organisms, even those with low HR activity. We demonstrate the insertion of exogenous reporter genes into human cells and animals using the PITCh system with TALENs and CRISPR/Cas9. Our PITCh methods provide a new insight into the targeted insertion of exogenous donor DNA and an alternative way of making knocked-in cells and organisms.

Results

TAL-PITCh design and application in human cells. We first demonstrated the PITCh system in TALEN-mediated knock-in (TAL-PITCh). In TAL-PITCh, a single pair of TALENs and a TAL-PITCh vector containing a TALEN target site are constructed and co-introduced (Fig. 1a, left panel). To generate microhomologous sequences, the TALEN target site on the TAL-PITCh vector should contain a different spacer sequence compared with the original genomic sequence, in which the anterior half and posterior half are switched. The genomic sequence and the TAL-PITCh vector can be cut by the same TALEN pair, and the linearized TAL-PITCh vector contains microhomologous DNA ends corresponding to the genomic cleavage site. After MMEJ-dependent integration, the whole vector is precisely incorporated into the genome with two TALEN target sites (Fig. 1a, right panel). However, these TALEN target sites are

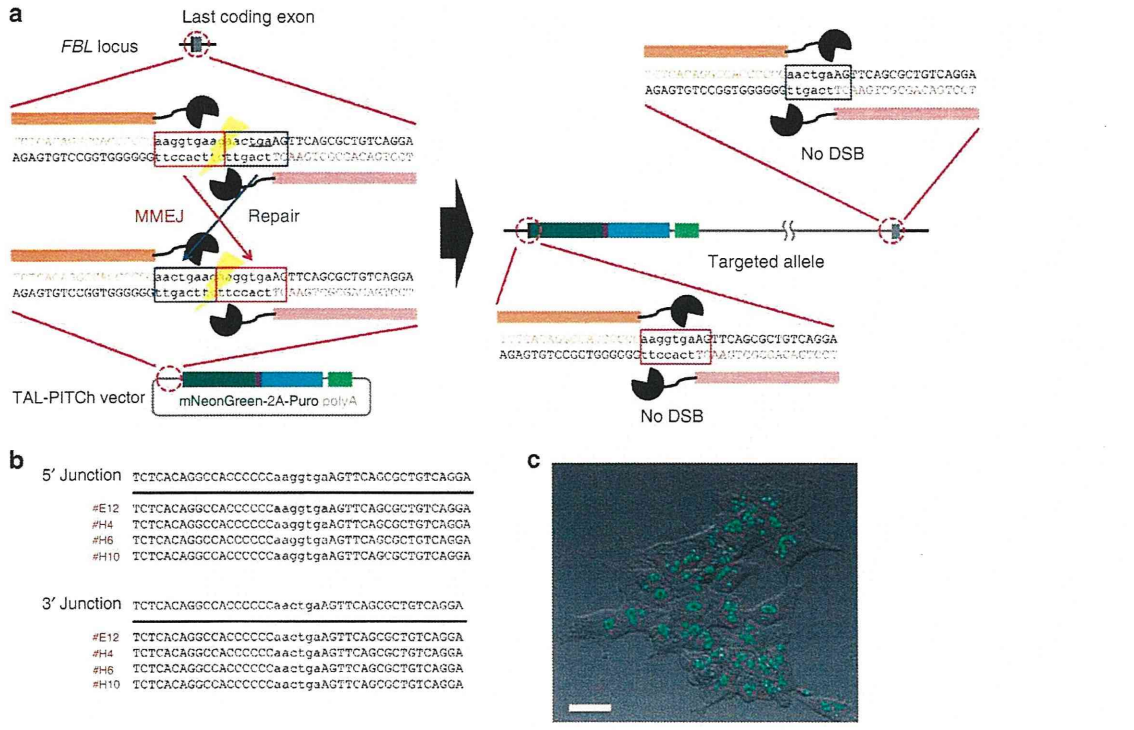

Figure 1 | TAL-PITCh in human cells. (**a**) Schematic illustration of TAL-PITCh at the human *FBL* locus. Orange and pink letters indicate the left and right TALEN target sites, respectively. Red and blue boxes indicate the microhomologous sequences. The stop codon is underlined. (**b**) Sequences of knocked-in clones. The intended knocked-in sequence is shown at the top. TALEN target sites are shown in capital letters. Red letters indicate correctly knocked-in clones. (**c**) Confocal laser scanning microscopy image of knocked-in cells showing nucleolar localization of mNeonGreen fluorescence (clone #H6). Scale bar, 30 μm.

hardly cut by TALENs, because they contain shortened spacer regions, which are out of the optimal range for DSB introduction by TALENs[14].

As a proof-of-principle experiment, we first demonstrated the TAL-PITCh system in cultured cells. We targeted the last coding exon of the human *fibrillarin* (*FBL*) gene using Platinum TALENs[15], and knocked-in the TAL-PITCh vector in HEK293T cells, resulting in a C-terminal fusion of *mNeonGreen*, reported recently as an ultra-bright fluorescent protein gene[16], followed by *2A-puromycin*. The TAL-PITCh vector contains no promoter for mammalian cell expression; therefore, the *FBL-mNeonGreen-2A-puromycin* gene expression should be driven by the endogenous *FBL* promoter. After puromycin selection, single cells were isolated by limiting dilution and cultured independently. Six potential knocked-in clones were analysed by DNA sequencing and laser-scanning fluorescence microscopy. Genomic regions around the 5′ and 3′ junctions could be amplified by PCR and sequenced from four of the six clones (Supplementary Fig. 1a,b; Supplementary Table 1). All the sequenced clones had correctly targeted alleles mediated by MMEJ (Fig. 1b), and showed nucleolar fluorescence, which is consistent with a previous report[16] (Fig. 1c).

To test the applicability of TAL-PITCh for another genomic locus and another cell line, we targeted the human *β-actin* (*ACTB*) gene in HeLa cells (Supplementary Fig. 2a). Six potentially knocked-in cell clones showing fluorescence were established and their junctions were analysed by PCR. Four of the six were selected as correctly PITCHed candidate clones (Supplementary Fig. 3a,b; Supplementary Table 1). In this case, one of the four clones contained a 3-bp insertion at the 5′ junction and three of the clones contained 5–27-bp insertions and deletions at the 3′ junction; one clone had correct junctions at both sides (Supplementary Fig. 2b). Fluorescence was observed at stress fibres in this clone (Supplementary Fig. 2c). We further confirmed correct integration by southern blot analysis, indicating that no random integration occurred (Supplementary Fig. 4). We also confirmed the higher colony-forming efficiency than TALEN-assisted gene knock-in mediated by HR, suggesting the superiority of MMEJ-mediated integration compared with the conventional method (Supplementary Fig. 5).

TAL-PITCh in animals. To check the applicability of the TAL-PITCh system *in vivo*, we next examined TAL-PITCh in silkworms (*Bombyx mori*) and frogs (*Xenopus laevis*). In silkworms, TALENs can induce highly efficient mutagenesis of the target genes, and the mutation rates in G_0 gametes can exceed 50%[17]. Nevertheless, a successful knock-in of a long gene cassette using

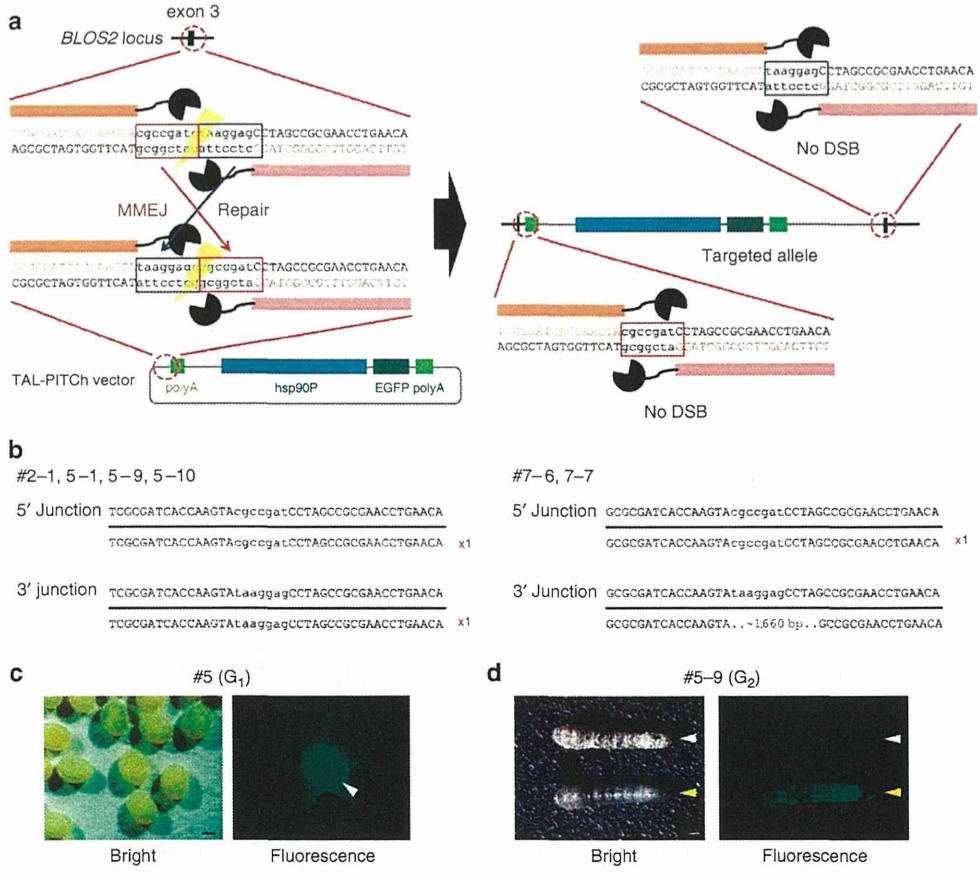

Figure 2 | TAL-PITCh in silkworms. (**a**) Schematic illustration of TAL-PITCh at the *B. mori BLOS2* locus. Orange and pink letters indicate the left and right TALEN target sites, respectively. Red and blue boxes indicate the microhomologous sequences. hsp90P, *hsp90* promoter. (**b**) Sequences of knocked-in alleles from the six G_1 worms (#2-1, #5-1, #5-9, #5-10, #7-6 and #7-7). The intended knocked-in sequence is shown at the top. TALEN target sites are shown in capital letters. Red letters indicate correctly knocked-in alleles. Blue letters indicate insertions. (**c**) Bright-field and fluorescence microscopy images of the G_1 embryos in the #5 batch. Strong EGFP expression could be observed in the putative knock-in embryo, according to the activity of the *hsp90* promoter (arrowhead). Scale bar, 0.5 mm. (**d**) Bright-field and fluorescence microscopy images of the G_2 larva derived from #5-9 batch. White and yellow arrowheads indicate wild-type and knock-in silkworms, respectively. Note that the knock-in larva shows an oily skin phenotype. Scale bar, 1 mm.

TALENs has not yet been achieved. This is presumably because HR activity is very low in germline cells of silkworms[17] and suggests that conventional knock-in methods mediated by HR are not promising. Therefore, we conceived the idea of applying the MMEJ-mediated TAL-PITCh system in silkworms.

We targeted the silkworm *BLOS2* gene, because its efficient knockout has previously been achieved using TALENs[18]. Messenger RNA (mRNA) of TALENs designed against exon 3 of the *BLOS2* gene was injected together with the TAL-PITCh vector harbouring the *hsp90* promoter-enhanced green fluorescent protein (*EGFP*) expression cassette[19] (Fig. 2a), and the EGFP expression in their progeny was examined. Remarkably, a number of G_1 embryos showed strong EGFP expression (Fig. 2c; Supplementary Fig. 6; Supplementary Table 2). This result suggested that the knock-in had occurred successfully in the G_0 gametes. We checked the genotype of each EGFP-positive G_1 individual and found that the TAL-PITCh vector was integrated into the *BLOS2* locus in six worms (Fig. 2b; Supplementary Fig. 7; Supplementary Table 2). Four of them showed precise integration, whereas two of them had ~1,660-bp extra sequence containing a partial *EGFP* sequence and the genomic sequence at ~2.6-kb downstream of the TALEN target site in the 3′ junction (Fig. 2b; Supplementary Table 2). The targeted integration into the *BLOS2* locus was further supported by the fact that these individuals exhibited an oily skin, a phenotype caused by the disruption of *BLOS2* gene[18] (Fig. 2d). Thus, we concluded that the TAL-PITCh system is quite effective in silkworms.

We subsequently tried *EGFP* knock-in at endogenous gene loci in *X. laevis* embryos as a model of vertebrates, because gene knock-in in frogs including *X. laevis* has not yet been achieved, although targeted mutagenesis can be performed efficiently using TALENs[20,21]. Thus, we first targeted the *no29* locus, one of the histone chaperone paralogues[22] in *X. laevis*, using TAL-PITCh (Fig. 3a). In this case, we designed TALENs around the start codon of the *no29* gene and knocked-in the *no29-EGFP* fusion

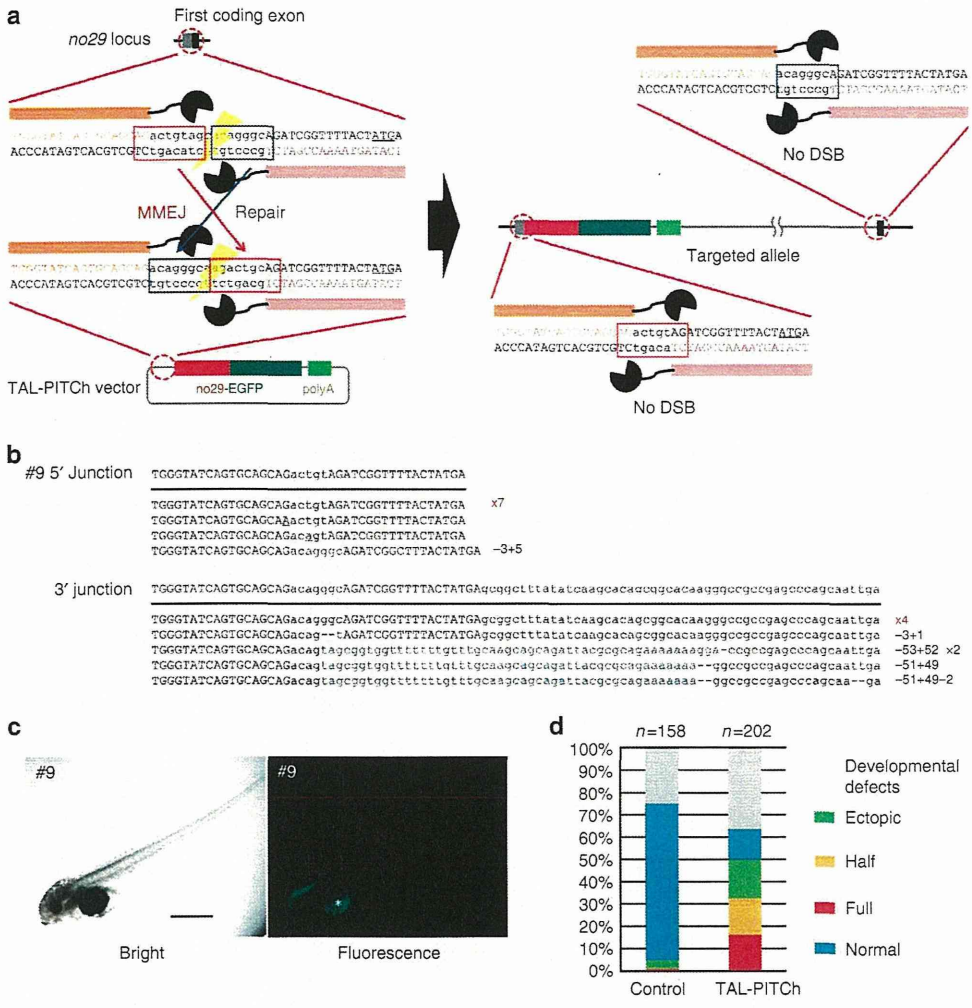

Figure 3 | TAL-PITCh at the *no29* locus in frog embryos. (**a**) Schematic illustration of TAL-PITCh at the *X. laevis no29* locus. Orange and pink letters indicate the left and right TALEN target sites, respectively. Red and blue boxes indicate the microhomologous sequences. The start codons are underlined. (**b**) Sequences of knocked-in alleles from embryo #9. The intended knocked-in sequence is shown at the top. TALEN target sites are shown in capital letters. Red letters indicate correctly knocked-in alleles. Blue letters indicate insertions. Dashes indicate deletions. Substitutions are underlined. (**c**) Bright-field and fluorescence microscopy images of embryo #9. An asterisk indicates yolk autofluorescence. Scale bar, 1 mm. (**d**) Percentage of phenotypes in the control embryos and the TAL-PITChed embryos. For the control, the vector from which the TALEN target site was removed was used instead of the TAL-PITCh vector. Except for abnormally developed embryos, phenotypes were divided into four groups (full, half, ectopic and normal) according to the expressed region of EGFP. Total numbers of individuals are shown at the top of each graph.

complementary DNA. Although the spatial expression pattern of *no29* during early development of *X. laevis* has never been elucidated, we observed an obvious expression tendency in the central nervous system (Fig. 3c; Supplementary Fig. 8b). Overall, ~15% of the embryos injected with the TALEN mRNAs and the TAL-PITCh vector showed full expression and another 15% of the embryos showed half expression, that is, the left half or the right half of the body, in the central nervous system (Fig. 3d). Three individuals showing the intended sizes of amplicons for both the 5′ and 3′ junctions were sequenced, and all three had precisely PITCHed alleles at least in the 5′ junction. In the 3′ junction, however, not all the individuals contained precisely PITCHed alleles (#1, 9 and 10; Fig. 3b; Supplementary Fig. 8a,b).

We next demonstrated TAL-PITCh-mediated *in vivo* gene knock-in at the *keratin* (*fgk*; fin and gill keratin) locus in *X. laevis*, in a manner similar to that in human cells (Fig. 4a). The *EGFP* gene was inserted just before the endogenous stop codon to express an *fgk–EGFP* fusion gene. Regarding *fgk*, transgenic *X. laevis* embryos have reportedly shown specific expression in the fin and the gill[23]. Consistent with this report, we obtained several embryos showing fluorescence specifically localized in the fin (Fig. 4b) and the gill (Fig. 4c) with precise 5′ and 3′ junctions, although one of them also contained subtle mutations at the both junctions (Fig. 4b,c). Furthermore, fusional expression enabled us to observe a cytoskeletal localization (Fig. 4b,c).

CRIS-PITCh design and application in human cells. Another important facet of the PITCh system is whether CRISPR/Cas9 could be used instead of TALENs. Thus, we targeted the *FBL* locus in HEK293T cells using CRISPR/Cas9-mediated PITCh (CRIS-PITCh) (Fig. 5a, left panel). The principles of inducing DSBs with TALENs and CRISPR/Cas9 are totally different; therefore, we modified the targeting strategy (Fig. 5a).

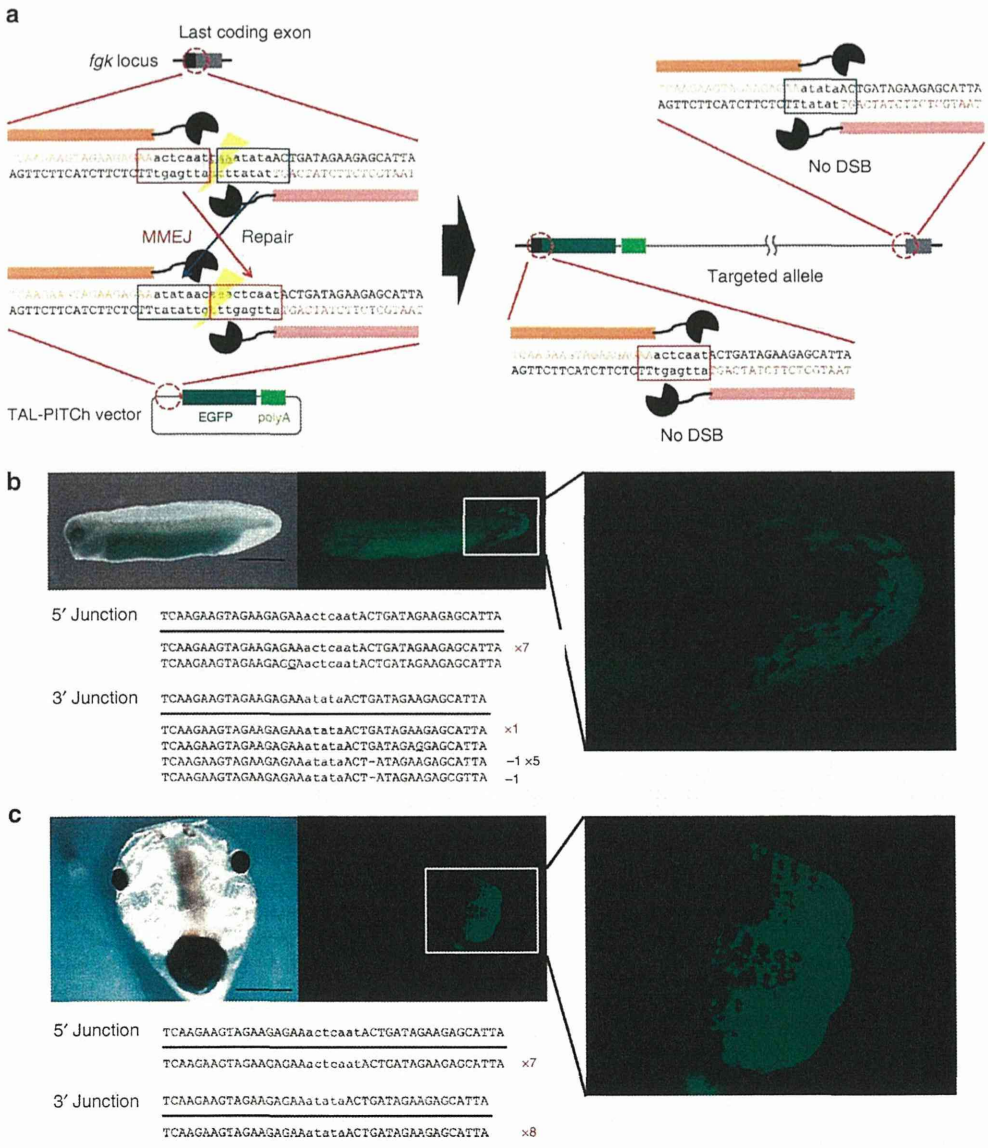

Figure 4 | TAL-PITCh at the *fgk* locus in frog embryos. (**a**) Schematic illustration of TAL-PITCh at the *X. laevis fgk* locus. Orange and pink letters indicate the left and right TALEN target sites, respectively. Red and blue boxes indicate the microhomologous sequences. The stop codon is underlined. (**b**,**c**) Microscopic images and sequences of the TAL-PITCHed embryos showing EGFP expression in the fin (**b**) and the gill (**c**). TALEN target sites are shown in capital letters. Red letters indicate correctly knocked-in alleles. Substitutions are underlined. Scale bars, 1 mm.