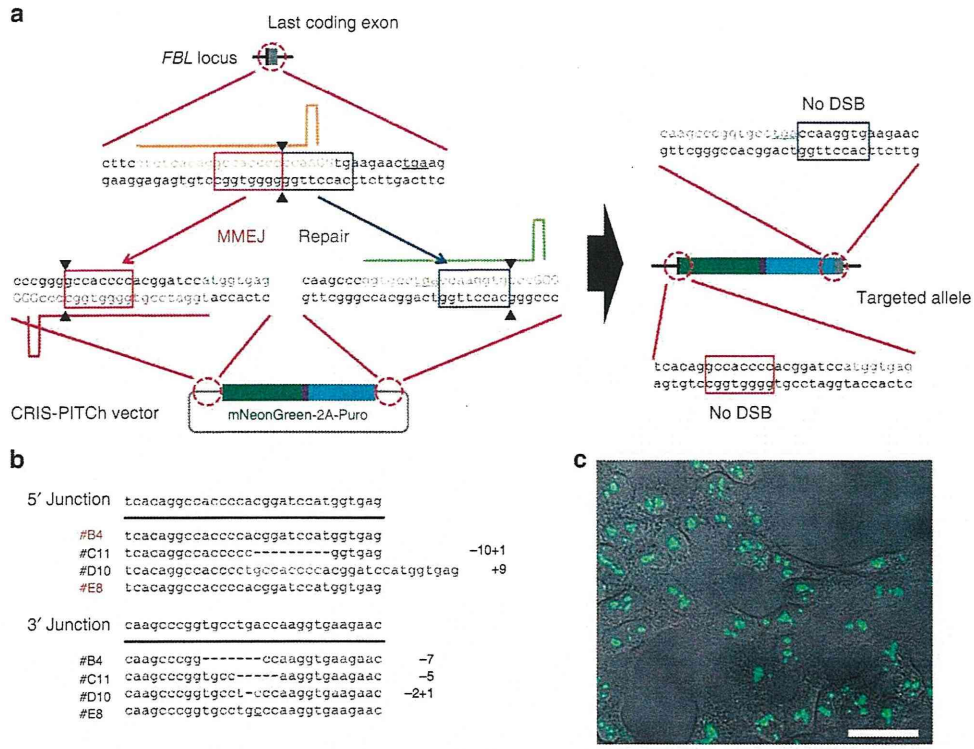


Figure 5 | CRIS-PITCh in human cells. (**a**) Schematic illustration of CRISPR/Cas9-mediated targeted integration using CRIS-PITCh. Orange, pink and green letters indicate the gRNA target sites. Red and blue boxes indicate the microhomologous sequences. The stop codons are underlined. (**b**) Sequences of knocked-in clones. The intended knocked-in sequence is shown at the top. Dashes indicate deletions. Blue letters indicate the insertion. The substitution is underlined. (**c**) Confocal laser scanning microscopy image of knocked-in cells showing nucleolar localization of mNeonGreen fluorescence (clone #B4). Scale bar, 30 μm.

In CRIS-PITCh, three guide RNAs (gRNAs) and Cas9 nuclease should be coexpressed, and two different gRNA target sites should be added to the CRIS-PITCh vector. Strategies used for TAL-PITCh can, of course, also be applicable to CRIS-PITCh; however, there are several reasons for improving the system: to remove unnecessary vector backbone and to abolish restriction on the gRNA target sequence (see Discussion for details). Using this improved CRIS-PITCh system enabling cassette integration, we could produce knocked-in cells without any additional sequence (Fig. 5a, right panel). After transfection of CRISPR/Cas9 and CRIS-PITCh vectors, followed by puromycin selection and single-cell cloning, genomic DNA was extracted and knocked-in alleles were amplified by PCR (Supplementary Fig. 9). DNA sequencing revealed that two of the four sequenced clones contained precisely joined 5′ junctions, while they had substitutions, insertions or deletions at the 3′ junction (Fig. 5b). Nucleolar localization of fluorescence was observed, similar to the TAL-PITCh experiment (Fig. 5c).

Finally, we investigated whether the TAL-PITChed and CRIS-PITChed cell clones contained off-target mutations, especially because the CRISPR/Cas9 system can reportedly induce substantial off-target mutations in human cell lines, such as U2OS, K562 and HEK293 cells[24–26]. The top six potential off-target sites were sequenced for the TAL-PITChed HEK293T cell clones (#H4 and #H6 in Fig. 1b and Supplementary Fig. 1), and the top three potential off-target sites of each gRNA were sequenced for the CRIS-PITChed HEK293T cell clones (#B4 and #E8 in Fig. 5b and Supplementary Fig. 9b). To our relief, none of the sequenced off-target candidates were mutated (Supplementary Tables 3 and 4).

Discussion

Along with the HR-independent knock-in strategies reported so far, our PITChing strategy enables flexible gene knock-in in cells and animals. In this study, we examined full-plasmid integration using TAL-PITCh and cassette integration using CRIS-PITCh. However, cassette integration using TAL-PITCh without the vector backbone and full-plasmid integration using CRIS-PITCh may also be possible. When performing cassette integration using TAL-PITCh, two TALEN target sites should be added at both ends of the cassette (Supplementary Fig. 10). The left half of the spacer sequence of the genomic TALEN target site should be placed at the right half of the spacer region of the left TALEN site on the TAL-PITCh vector, and the right half of the spacer sequence of the genomic TALEN target site should be placed at the left half of the spacer region of the right TALEN site on the vector. Alternatively, additional Cre-loxP- or Flp-FRT-mediated excision after establishing knocked-in cells or minicircle DNA generation[27] after constructing plasmid vectors could also be performed, as discussed in the previous study[10].

There are two ways to perform full-plasmid integration using CRIS-PITCh, and restriction of the target genomic sequence varies between them. When gRNA for the CRIS-PITCh vector is designed against the sense strand, sequence restriction would be 5′-GGNNNNNNGG-3′, if the microhomologous sequence is set to 8 bp (Supplementary Fig. 11a). However, when gRNA for the CRIS-PITCh vector is designed against the antisense strand, the sequence restriction would be 5′-CCGG-3′, if the microhomologous sequence is set to 8 bp (Supplementary Fig. 11b). Modifying the length of the microhomology sequence can change these restrictions of the target sequence.

Other tips for the TAL-PITCh and CRIS-PITCh are described below. Regarding TAL-PITCh, the left and right TALEN target sites on the TAL-PITCh can be shuffled. If they are shuffled, the 5′ junction of the targeted allele would contain the left TALEN site, a shortened spacer and an inverted left TALEN site. The 3′ junction of the targeted allele would contain the right TALEN site, a shortened spacer and the inverted right TALEN site. By performing shuffling, we can completely avoid recutting the targeted allele by TALENs, if we use TALENs containing heterodimeric FokI nuclease domains[28–30]. Regarding CRIS-PITCh-mediated cassette integration, the left and right gRNA target sites on the CRIS-PITCh vector should be designed to target the antisense and sense strands, respectively, as shown in Fig. 5a. Otherwise, some sequence restriction other than the original protospacer adjacent motif of the genomic CRISPR/Cas9 target site will arise. When targeting the antisense strand at the genomic CRISPR/Cas9 target site, the left and right gRNA target sites on the CRIS-PITCh vector should be designed to target the sense and antisense strands, respectively, to minimize the limitation of the genomic CRISPR/Cas9 target site.

In our TAL-PITCh and CRIS-PITCh experiments, the 5′ junctions had a high tendency to be joined precisely using the microhomologous sequence; however, the 3′ junctions were not necessarily joined by MMEJ. This tendency is likely to depend on the selection method that we adopted in this study. Puromycin selection used in this study can exclude out-of-frame clones; however, the 3′ junction is thought to have little influence on selection. In the case where both junctions need to be joined correctly for drug or fluorescence selection, it is likely that the bias observed in this study will disappear. In addition, enhancement of the MMEJ repair pathway and/or suppression of the NHEJ repair machineries might also increase the targeted integration with precise junctions and decrease the NHEJ-dependent erroneous integration, including false-positive clones, such as #F3 and #B10 in Supplementary Fig. 1b, #2 and #6 in Supplementary Fig. 3b and #G10 in Supplementary Fig. 9, which supposedly have unintended knocked-in alleles; for example, integration of concatemerized vectors, integration with large deletion or addition and random integration. Further studies are needed to clarify the mechanism of integration and improve its accuracy.

Overall, we proved that TAL-PITCh-mediated gene knock-in could be applied in human cells and other animals, suggesting broad applicability of the strategy. To the best of our knowledge, this is the first report to show that targeted insertions can occur via very short microhomologies, both in cultured cells and in animals. In addition, we demonstrated successful CRIS-PITCh-mediated gene knock-in in human cells without carrying over a vector backbone sequence. We anticipate that our PITCh systems will enhance the usefulness of genome engineering techniques in a variety of cells and organisms, especially in those in which gene knock-in is difficult because of low HR efficiency.

Methods

Construction of TALEN plasmids. For human cell and frog experiments, a two-step Golden Gate assembly method using the Platinum Gate TALEN Kit (Addgene; cat#1000000043)[15] was used to construct Platinum TALEN plasmids containing the homodimer-type FokI nuclease domain. Briefly, single DNA-binding repeats were assembled into the intermediate array vectors. The assembled repeat arrays were subsequently inserted into the final destination vectors, ptCMV-153/47-VR. For the silkworm experiments, the Golden Gate TALEN and TAL Effector Kit 2.0 (Addgene; cat#1000000024)[31] were used to construct TALENs (BLTS-5A and BLTS-4B[18]), and the repeat arrays were inserted into the scaffold plasmid, pBlue-TAL[18].

Construction of CRISPR/Cas9 plasmids. The multiplex CRISPR/Cas9 assembly system[32] was used to construct the all-in-one CRISPR/Cas9 plasmids. Briefly, pX330 vector (Addgene; Plasmid 42230) was modified to unify multiple gRNA-expressing cassettes into a single vector using the Golden Gate assembly method. Oligonucleotides for gRNA templates were synthesized, annealed and inserted into the corresponding vectors. A list of the oligonucleotides used is shown in Supplementary Table 5. Golden Gate assembly was used to assemble the constructed vectors into an all-in-one CRISPR/Cas9 vector for the *FBL* gene, termed pX330A-*FBL*-3gRNAs, harbouring three gRNA cassettes and a Cas9 cassette.

Construction of PITCh and HR vectors. TAL-PITCh, CRIS-PITCh and HR vectors were constructed using PCR and In-Fusion cloning (Clontech) or by standard molecular-cloning methods. The full-plasmid sequences are shown in Supplementary Fig. 12.

Cell culture and transfection. HEK293T and HeLa cells, obtained from ATCC, were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum. Lipofectamine LTX (Life Technologies) and Opti-MEM (Life Technologies) were used to transfect plasmids, according to the supplier's protocols. Plasmid concentrations, cell numbers and dishes used were as follows: 200 ng each for ptCMV left and right TALEN vectors and for TAL-PITCh vector into 1×10^5 cells using a six-well plate in the experiments in Fig. 1 and Supplementary Fig. 2; 1.6 μg each for ptCMV left and right TALEN vectors and for TAL-PITCh or HR vector into 5×10^5 cells using a 100-mm dish in the experiments of Supplementary Fig. 5; 400 ng for pX330A-*FBL*-3gRNAs CRISPR/Cas9 vector and 200 ng for CRIS-PITCh vector into 1×10^5 cells using a six-well plate in the experiments of Fig. 5. After transfection, cells were cultured in the growth medium described above for 3 days and then selected with 1 μg ml^{-1} puromycin for 6 days. For DNA sequencing, microscopy and southern blotting, the selected cells were cloned using the limiting dilution method in 96-well plates.

mRNA synthesis and microinjection for the silkworm experiments. mMessage mMachine T7 Ultra Kit (Life Technologies) was used to synthesize *B. mori BLOS2* TALEN mRNA (BLTS-5A and BLTS-4B[18]). mRNA was precipitated with LiCl, washed with 70% ethanol three times and air-dried and dissolved in 0.5 mM phosphate buffer (pH 7.0) containing 5 mM KCl. TALEN mRNA (250 ng μl^{-1} each) and 500 ng μl^{-1} donor vector were injected into embryos of silkworm *w1-pnd* strain that were collected between 1- and 2 h after egg laying at the syncytial preblastderm stage. After injection, the opening was sealed with glue and the embryos were incubated at 25 °C. The hatched silkworm larvae were reared on an artificial diet (Nihon Nosan Kogyo) at 25 °C under a 12-h light/dark photoperiod. Each injected individual was crossed with non-injected worms. EGFP expressions of G_1 individuals were observed during the embryonic stage, and the EGFP-positive embryos were selected for further rearing. The G_1 adults were crossed with the *w-c* diapausing strain.

mRNA synthesis and microinjection for the frog experiments. Fertilized *X. laevis* eggs were obtained from wild-type adults injected with human chorionic gonadotropin (Aska Pharmaceutical). Eggs were dejellied with 2% cysteine and washed in 0.1 × Marc's modified ringer (MMR). Washed eggs were transferred into 5% Ficoll (Sigma-Aldrich) in 0.3 × MMR and were co-injected with a pair of TALEN mRNAs (250 pg each), synthesized using a mMessage mMachine T7 Ultra Kit (Life Technologies), and TAL-PITCh vectors (100 pg) at the one-cell stage using Nanoject II (Drummond). Injected embryos were reared to the swimming stage in 0.1 × MMR at 20 °C. Animals were maintained and used in accordance with the Hiroshima University guidelines for the use and care of experimental animals.

Microscopy. For human cell experiments, cells were moved to collagen-coated glass-bottom 24-well plates and fixed with 4% paraformaldehyde in PBS. Fluorescence was observed and cell images were captured with a 488-nm laser using a confocal laser-scanning microscope (Olympus FV-1000D). For *B. mori* experiments, fluorescence was observed using a fluorescence stereomicroscope (Olympus SZX16). For *X. laevis* experiments, fluorescence was observed using a fluorescence stereomicroscope (Leica MZ10F).

Genomic PCR and DNA sequencing. A DNeasy Blood and Tissue kit (Qiagen) was used to extract genomic DNA from cell pellets, frog embryos and silkworm larvae or adults. Genomic PCR was performed using KOD FX (Toyobo), KOD FX Neo (Toyobo) or LA *Taq* (Takara) with the primers listed in Supplementary Table 6. For the human cell and silkworm experiments, the PCR products were subjected to direct DNA sequencing. For the *X. laevis* experiments, the PCR products were cloned and transformed into bacteria using a TOPO TA Cloning Kit with PCR2.1 TOPO (Life Technologies). Subsequently, colony PCR products were used as templates for DNA sequencing. DNA sequencing was performed using an ABI 3130xl Genetic analyzer (Life Technologies) with a BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies).

Off-target analyses. The PROGNOS tool (http://baolab.bme.gatech.edu/cgi-bin/prognos/prognos.cgi)[33] was used to identify potential off-target sites for the *FBL* TALENs. Maximum mismatches per half-site and spacer lengths were set to 6 and

12–24, respectively. The CRISPR design tool (http://crispr.mit.edu/)[25] was used to identify potential off-target sites for the three gRNAs against the genomic locus and CRIS-PITCh vector. Genomic regions around each candidate site were amplified by PCR using primers listed in Supplementary Table 7 and the sequence was confirmed by direct sequencing.

Southern blot analyses. Southern blotting was carried out according to the previously report[34] with some modifications, as described below. Five-μg aliquots of genomic DNA were digested with PstI, and 2 μg and 1 μg for the outer and the mNG probes, respectively, were resolved on 0.8% agarose gels. Digoxigenin-labelled DNA probes were made by PCR using KOD FX Neo (Toyobo) and DIG DNA labelling mix (Roche) with primers listed in Supplementary Table 6. Membrane transfer (Hybond-N+; GE Healthcare), ultraviolet cross-linking (120 mJ cm^{-2}), pre-hybridization and hybridization were performed according to the instructions for DIG Easy Hyb Granules (Roche). The CDP-Star Detection Reagent (Roche) was used to develop the membrane, following the manufacturer's instructions. The chemiluminescent signal was detected using Amersham Hyperfilm ECL (GE Healthcare).

References

1. Carroll, D. Genome engineering with targetable nucleases. *Annu. Rev. Biochem.* **83,** 409–439 (2014).
2. Kim, H. & Kim, J. S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15,** 321–334 (2014).
3. Chen, F. *et al.* High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat. Methods* **8,** 753–755 (2011).
4. Miyaoka, Y. *et al.* Isolation of single-base genome-edited human iPS cells without antibiotic selection. *Nat. Methods* **11,** 291–293 (2014).
5. Hockemeyer, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* **27,** 851–857 (2009).
6. Hockemeyer, D. *et al.* Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol.* **29,** 731–734 (2011).
7. Orlando, S. J. *et al.* Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Res.* **38,** e152 (2010).
8. Cristea, S. *et al.* In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnol. Bioeng.* **110,** 871–880 (2013).
9. Auer, T. O., Duroure, K., De Cian, A., Concordet, J. P. & Del Bene, F. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res.* **24,** 142–153 (2014).
10. Maresca, M., Lin, V. G., Guo, N. & Yang, Y. Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Res.* **23,** 539–546 (2013).
11. Bae, S., Kweon, J., Kim, H. S. & Kim, J. S. Microhomology-based choice of Cas9 nuclease target sites. *Nat. Methods* **11,** 705–706 (2014).
12. McVey, M. & Lee, S. E. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet.* **24,** 529–538 (2008).
13. Taleei, R. & Nikjoo, H. Biochemical DSB-repair model for mammalian cells in G1 and early S phases of the cell cycle. *Mutat. Res.* **756,** 206–212 (2013).
14. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29,** 143–148 (2011).
15. Sakuma, T. *et al.* Repeating pattern of non-RVD variations in DNA-binding modules enhances TALEN activity. *Sci. Rep.* **3,** 3379 (2013).
16. Shaner, N. C. *et al.* A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nat. Methods* **10,** 407–409 (2013).
17. Daimon, T., Kiuchi, T. & Takasu, Y. Recent progress in genome engineering techniques in the silkworm, *Bombyx mori*. *Dev. Growth Differ.* **56,** 14–25 (2014).
18. Takasu, Y. *et al.* Efficient TALEN construction for *Bombyx mori* gene targeting. *PLoS ONE* **8,** e73458 (2013).
19. Tsubota, T. *et al.* Identification of a novel strong and ubiquitous promoter/enhancer in the silkworm, *Bombyx mori*. *G3* **4,** 1347–1357 (2014).
20. Suzuki, K. T. *et al.* High efficiency TALENs enable F0 functional analysis by targeted gene disruption in *Xenopus laevis* embryos. *Biol. Open* **2,** 448–452 (2013).
21. Sakane, Y. *et al.* Targeted mutagenesis of multiple and paralogous genes in *Xenopus laevis* using two pairs of transcription activator-like effector nucleases. *Dev. Growth Differ.* **56,** 108–114 (2014).
22. Motoi, N., Hasebe, T., Suzuki, K. T. & Ishizuya-Oka, A. Spatiotemporal expression profile of no29/nucleophosmin3 in the intestine of *Xenopus laevis* during metamorphosis. *Cell Tissue Res.* **344,** 445–453 (2011).
23. Suzuki, K. T. *et al.* Characterization of a novel type I keratin gene and generation of transgenic lines with fluorescent reporter genes driven by its promoter/enhancer in *Xenopus laevis*. *Dev. Dyn* **239,** 3172–3181 (2010).
24. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31,** 822–826 (2013).
25. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827–832 (2013).
26. Cradick, T. J., Fine, E. J., Antico, C. J. & Bao, G. CRISPR/Cas9 systems targeting β-globin and *CCR5* genes have substantial off-target activity. *Nucleic Acids Res.* **41,** 9584–9592 (2013).
27. Kay, M. A., He, C. Y. & Chen, Z. Y. A robust system for production of minicircle DNA vectors. *Nat. Biotechnol.* **28,** 1287–1289 (2010).
28. Miller, J. C. *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25,** 778–785 (2007).
29. Szczepek, M. *et al.* Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.* **25,** 786–793 (2007).
30. Doyon, Y. *et al.* Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat. Methods* **8,** 74–79 (2011).
31. Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* **39,** e82 (2011).
32. Sakuma, T., Nishikawa, A., Kume, S., Chayama, K. & Yamamoto, T. Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci. Rep.* **4,** 5400 (2014).
33. Fine, E. J., Cradick, T. J., Zhao, C. L., Lin, Y. & Bao, G. An online bioinformatics tool predicts zinc finger and TALE nuclease off-target cleavage. *Nucleic Acids Res.* **42,** e42 (2014).
34. Ochiai, H. *et al.* TALEN-mediated single-base-pair editing identification of an intergenic mutation upstream of *BUB1B* as causative of PCS (MVA) syndrome. *Proc. Natl Acad. Sci. USA* **111,** 1461–1466 (2014).

Acknowledgements

We thank Dr Hiroshi Ochiai (Hiroshima University, Hiroshima, Japan) for sharing the synthesized mNeonGreen cDNA, Dr Keiro Uchino (National Institute of the Agrobiological Sciences, Ibaraki, Japan) for injecting the silkworm, Dr Yoko Takasu (National Institute of the Agrobiological Sciences, Ibaraki, Japan) for instructing the TALEN mRNA synthesis, Mr Kaoru Nakamura and Mr Toshihiko Misawa (National Institute of the Agrobiological Sciences, Ibaraki, Japan) for rearing the silkworms and Ms Satoko Kawamoto (National Institute of the Agrobiological Sciences, Ibaraki, Japan) for technical assistance. This work was supported by the Japan Society for the Promotion of Science (25890014 to T.S. and 25124708 to K.T.S.), the Ministry of Agriculture, Forestry and Fisheries, and the NIAS Strategic Research Fund.

Author contributions

T.T., T.S. and K.T.S. designed the work. S.N., T.T., Y.S., S.K., T.S. and K.T.S. performed the experiments. T.T., T.S. and K.T.S. wrote the manuscript with support from all the authors. N.S., M.O., T.D. and H.S. provided instructions. T.Y. supervised the work.

Additional information

Accession codes: Sequences of TAL-PITCh, CRIS-PITCh, and HR vectors have been deposited in the NCBI Genbank nucleotide database under accession codes LC008486, LC008487, LC008488, LC008489, LC008490, LC008491 and LC008492.

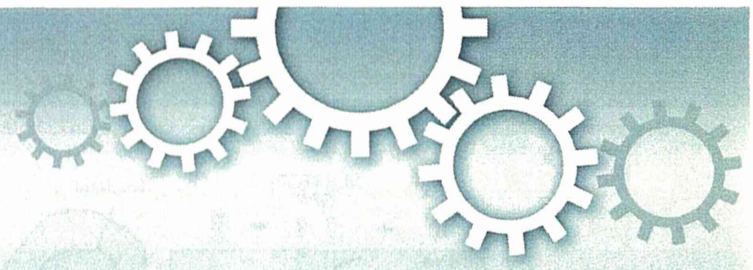
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Nakade, S. *et al.* Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat. Commun.* 5:5560 doi: 10.1038/ncomms6560 (2014).

SCIENTIFIC REPORTS



OPEN

SUBJECT AREAS:
TRANSCRIPTION
EMBRYONIC STEM CELLS
CELLULAR NOISE
SINGLE-CELL IMAGING

Received
7 August 2014

Accepted
30 October 2014

Published
20 November 2014

Correspondence and requests for materials should be addressed to H.O. (ochiai@hiroshima-u.ac.jp)

Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells

Hiroshi Ochiai[1], Takeshi Sugawara[1], Tetsushi Sakuma[2] & Takashi Yamamoto[1,2]

[1]Research Center for the Mathematics on Chromatin Live Dynamics (RcMcD), Hiroshima University, Higashi-Hiroshima 739-8530, Japan, [2]Department of Mathematical and Life Sciences, Graduate School of Science, Hiroshima University, Higashi-Hiroshima 739-8526, Japan.

Mouse embryonic stem cells (mESCs) are self-renewing and capable of differentiating into any of the three germ layers. An interesting feature of mESCs is the presence of cell-to-cell heterogeneity in gene expression that may be responsible for cell fate decisions. Nanog, a key transcription factor for pluripotency, displays heterogeneous expression in mESCs, via mechanisms that are not fully understood. To understand this variability, we quantitatively analyzed *Nanog* transcription and found that *Nanog* was both infrequently transcribed, and transcribed in a pulsatile and stochastic manner. It is possible that such stochastic transcriptional activation could contribute to the heterogeneity observed in *Nanog* expression as "intrinsic noise." To discriminate the effects of both intrinsic noise from other (extrinsic) noise on the expression variability of *Nanog* mRNA, we performed allele-specific single-molecule RNA fluorescent *in situ* hybridization in a reporter cell line and found that intrinsic noise contributed to approximately 45% of the total variability in *Nanog* expression. Furthermore, we found that *Nanog* mRNA and protein levels were well correlated in individual cells. These results suggest that stochastic promoter activation significantly affects the Nanog expression variability in mESCs.

mESCs are clonal cell lines derived from the pre-implantation epiblast. They are capable of self-renewing in media containing leukemia inhibitory factor (LIF), can differentiate into three germ layers, and display heterogeneity in gene expression. Uncovering the nature of this heterogeneity at the molecular level is important to the understanding of how stem cells modulate their stemness/differentiation balance. Nanog, a key transcription factor for pluripotency, exhibits a broad range of expression levels, and its heterogeneity seems to be related to the stemness/differentiation balance[1–4].

The autocrine fibroblast growth factor (FGF)/extracellular signal-regulated kinase (Erk) signaling that induces lineage commitment and inhibits Nanog expression is thought to be a source of Nanog heterogeneity[5–7], along with the positive[8] and negative feedback loops[9] in the gene regulatory networks responsible for pluripotency. Recently, it has been reported that *Nanog* is prone to be monoallelically transcribed (i.e., transcribed from only one copy of *Nanog*) in mESCs cultured in standard mESC medium containing serum and LIF (serum condition), even though mice have two copies of *Nanog*[9–11]. If this effect is a result of an intrinsically stochastic effect on promoter activation, it may also contribute to Nanog heterogeneity[12,13].

In this study, to gain insight into the contribution of *Nanog* transcriptional activity to Nanog heterogeneity, we quantitatively analyzed the transcription dynamics of endogenous *Nanog* and distribution of *Nanog* mRNA in a mESC population at single-cell resolution and observed infrequent and stochastic switching on and off of *Nanog* promoter states. Furthermore, we found that expression noise stemming from such promoter dynamics significantly affected heterogeneity in Nanog expression.

Results

Establishment of a Nanog-MS2 mESC line. To quantitatively analyze *Nanog* transcription in mESCs, we applied the MS2 system[14,15]. The transcribed MS2 sequence derived from MS2 bacteriophage forms a stem-loop structure, which is known to be bound by the MS2 coat protein (MCP) as a dimer (Figure 1a). Therefore, the integration of the 24 tandem MS2 sites into a specific gene of interest and expression of MCP fused with fluorescent protein enables the visualization of mRNA transcription as bright spots in the nucleus[16–18]. We applied transcription activator-like effector nuclease (TALEN)-mediated targeted integration of the MS2 repeat sequence into the

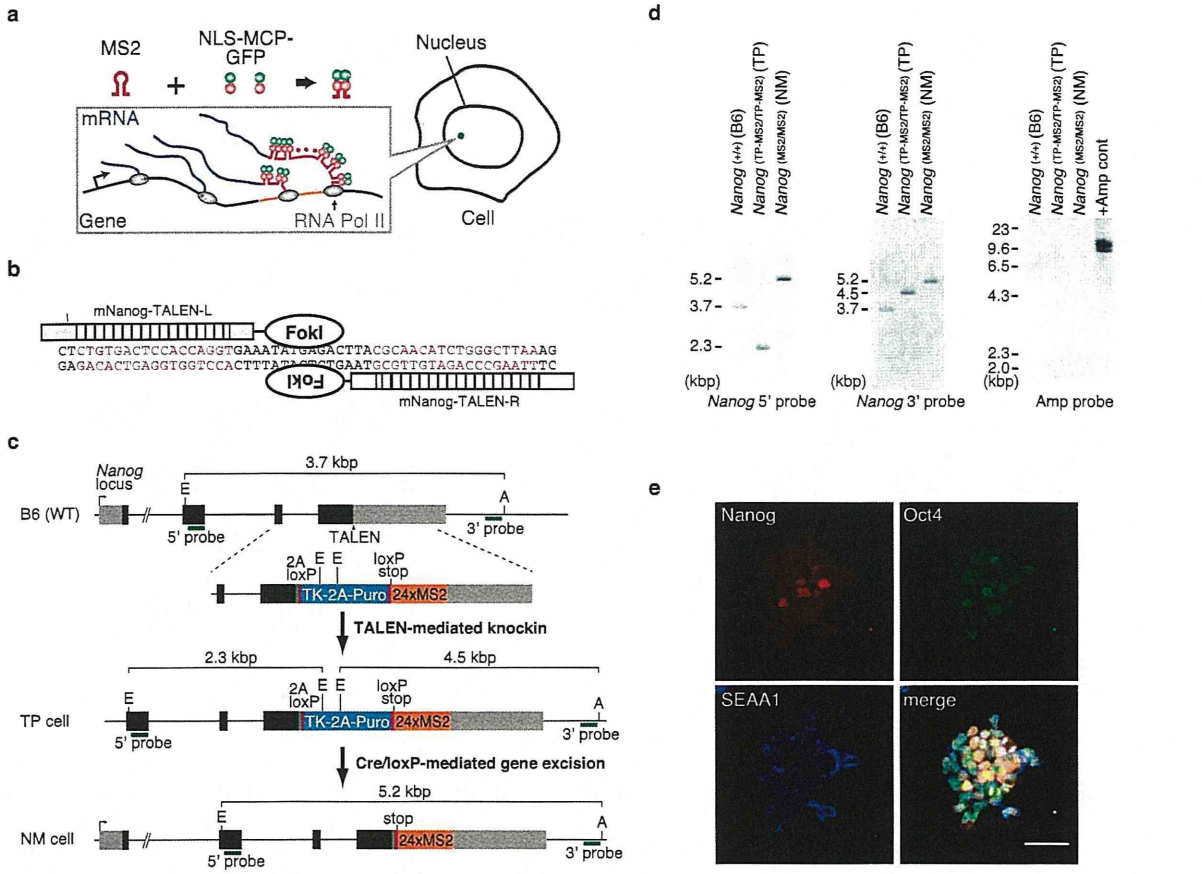


Figure 1 | Establishment of MS2-targeted ES cell line. (a) Schematic representation of the MS2 system. In the inset, black, orange, blue, and red lines represent a gene, its integrated MS2 repeat, the transcribed nascent mRNAs, and the transcribed MS2 repeats, respectively. Because of the accumulation of nuclear localization signal (NLS)-MS2 coat protein (MCP)-GFP fusion protein at the transcription site, the bright fluorescent spot could be observed at the site in the nucleus. (b) Schematic representations of transcription activator-like effector nucleases (TALENs) used in this study and their respective target nucleotide (highlighted in red letters). (c) The strategy for biallelic targeted gene integration into the *Nanog* locus. Grey, black, green, and magenta boxes indicate untranslated regions, Nanog coding sequences, and coding sequences for 2A peptide and loxP site, respectively. Green bars represent the positions of Southern blot probes. E, EcoNI; A, AflII. (d) Southern blot analyses showing TALEN-mediated biallelic insertion and Cre-mediated biallelic excision of the selection cassettes in modified mESC clones. (e) Immunofluorescence displays expression of pluripotent markers in the targeted cell line (NM cell). Each image is a maximum intensity projection of image stacks. Scale bar, 50 μm.

Nanog loci[19,20] (Figure 1b, c), and established a biallelically targeted mESC line (Figure 1d). Then, selection cassettes were removed by transient expression of Cre recombinase (Figure 1c, d). Southern blot analysis showed the expected band patterns and no random integrations (Figure 1d). We refer to the obtained cell line as Nanog-MS2 (NM) cells. The cell line expressed the undifferentiated embryonic stem-cell markers SSEA1 and Oct4 (also known as Pou5f1) (Figure 1e).

Next, to check the cell-to-cell variability of *Nanog* mRNA copy numbers in the NM cell line, we performed single-molecule fluorescent *in situ* hybridization (smFISH) using *Nanog* exonic probes (Figure 2a–c)[21]. In the serum conditions used (standard mESC culture medium containing LIF and serum), the mean count of *Nanog* mRNA copies in NM cells was slightly, but significantly, lower than that of the parental mESC line (133 for the parental mESC line and 77 for the NM cell line) (Figure 2c). However, the degrees of variability between the parental and derived lines were comparable (coefficient of variation [CV] for the parental and NM cell lines were 0.85 and 0.89, respectively). One of the possible causes of the difference in expression is the change in mRNA stability due to insertion of the MS2 repeats. To explore the possibility, we examined the half-lives of *Nanog* mRNA in NM and the parental mESCs (Supplementary Figure S1), and found that insertion of MS2 repeats slightly destabilizes *Nanog* mRNA. Therefore, the mean number of *Nanog* mRNA in NM cells appears lower than that in the parental mESCs.

It has been reported that the cytokine FGF4 is secreted by undifferentiated cells, acts in an autocrine manner[22,23], and represses *Nanog* expression through Erk signaling, suggesting that FGF/Erk signaling is a contributor to the transcription factor heterogeneity[6,24]. When cell lines were cultured in medium containing FGF/Erk and GSK3β inhibitors (2i conditions), the means of *Nanog* mRNA copies were increased, as expected (Figure 2c). Furthermore, the variability of these cells was lower than cells grown in the serum condition (CV for parental and NM cell lines in 2i conditions were 0.43 and 0.42, respectively). The CVs in parent and NM cells were comparable (Figure 2c). These findings suggest that the NM cell line displays similar heterogeneity to that of the parent cell line.

Next, to visualize transcriptional activity, we established a cell line that constitutively expresses nuclear localization signal (NLS)-MCP-green fluorescent protein (GFP) (mNeonGreen) (NM-G cell line). In this cell line, either none, one or two bright fluorescent spots were observed (Figure 2d). To confirm whether these spots corresponded with the *Nanog* transcription sites, we performed smFISH using a *Nanog* intronic probe set (Figure 2a, e). Although a majority of the smFISH and NLS-MCP-GFP spot signals coincided in cells cultured in 2i conditions (88%, n = 101), only one type of signal was detected

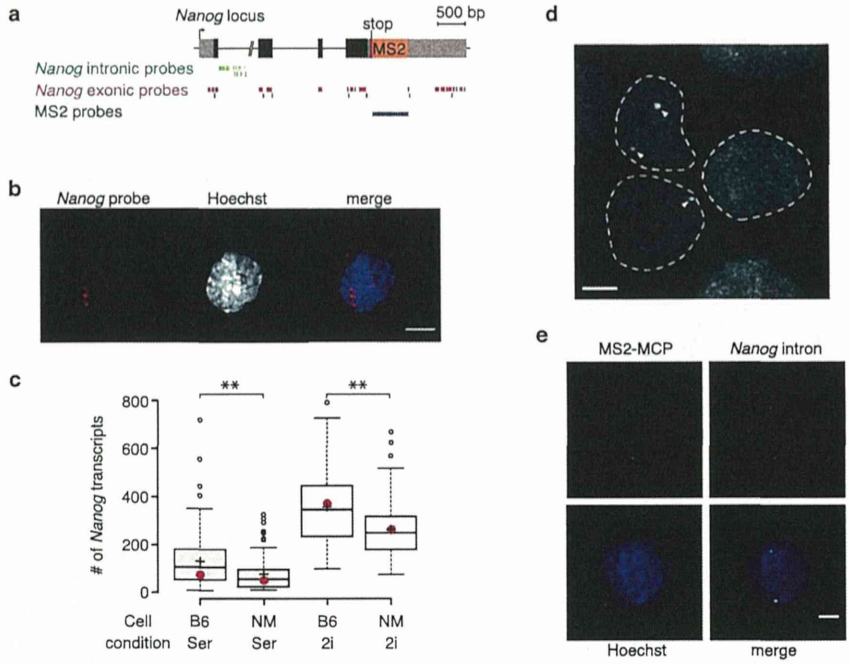


Figure 2 | Nanog-MS2 mESC line is useful for quantifying the *Nanog* transcription dynamics. (a) An illustration of the *Nanog* genomic locus and positions of smFISH probes used in this study. (b) Single-molecule fluorescent *in situ* hybridization (smFISH) analysis. *Nanog* mRNAs in a cell were visualized using *Nanog* exonic smFISH probes. Nuclei were counter-stained with Hoechst33342. Scale bar, 10 μm. (c) Distributions of *Nanog* mRNAs in mESC lines cultured in either serum (Ser) or 2i conditions (2i). The number of *Nanog* mRNAs in parental C57BL/6 (B6) or NM cell lines (NM) was counted by smFISH using *Nanog* exonic probes. In the box plot, center lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; outliers are represented by dots; crosses represent sample means. n = 129, 109, 102, and 119 sample points. Statistical significance of differences was assessed by a two-sample Kolmogorov-Smirnov test (**$P \ll 0.01$). Red dots represent the mean counts of *Nanog* transcripts predicted by the random telegraph model (see the main text). (d) Bright fluorescent spots as observed in NM-G cells. Representative NM-G cells in 2i conditions were subjected to live imaging. The image shown is a maximum intensity projection of image stacks. The dashed lines delineate individual nuclei of cells. The white arrowheads point out bright fluorescent spots assumed to be *Nanog* transcription sites. Scale bar, 5 μm. (e) Results of smFISH analysis in NM-G cells. A representative image shows the overlap of fluorescent spots of nuclear localization signal (NLS)-MS2 coat protein (MCP)-mNeonGreen (mNG) and *Nanog* intronic smFISH probes, indicating that the MS2-MCP spot signals coincide with *Nanog* transcription sites. Scale bar, 5 μm.

in some of the spots. This limitation could be attributed to the distance between the sites of *Nanog* intronic probes and the MS2 repeat (~4.5 kb, Fig. 2a). The RNA Pol II elongation rate in mESCs was reported to be 0.5–4 kb/min[25]; therefore, all signals of smFISH and MS2-MCP may not be simultaneously detected at the same position. These findings suggest that the MS2-MCP spot signals indicate the transcription of MS2-integrated *Nanog* alleles and that the NM-G cell line is useful for quantifying the *Nanog* transcription dynamics.

Quantification of *Nanog* transcription dynamics. To quantify *Nanog* transcription dynamics, we performed live imaging of NM-G cells cultured in serum or 2i conditions at 2-min intervals for 4 h (Figure 3a, b, Supplementary Video S1 and S2). In both conditions, *Nanog* was transcribed in a pulsatile manner, as reported in other model systems[26,27]. The transcription frequency over the 4-h time period was higher in 2i conditions than in serum conditions, consistent with previous reports (Figure 3c)[10,11].

Consistent with the smFISH data (Figure 2c), transcription frequencies were apparently variable among cells even in 2i conditions (Figure 3a–c). 2i conditions inhibit autocrine FGF/Erk signaling, one potential source of Nanog heterogeneity. This finding suggests that factors other than FGF/Erk signaling may significantly affect the expression variability of Nanog. Furthermore, transcription frequency during the 4-h imaging period seemed to show multimodal distributions in both conditions (Supplementary Figure S2). This suggests the existence of multiple states of *Nanog* expression that switch at intervals longer than 4 hours, as recently reported[28]. On the other hand, the distributions of transcriptional signal intensities in the two conditions were comparable (Figure 3d), suggesting that *Nanog* transcription was regulated not by modulation of the number of transcripts per transcriptional pulse, but rather by transcription frequency[27]. This kind of transcription dynamics might be explained by the random telegraph model[29–31] (Figure 3e). In this model, promoter states stochastically switch "ON", permitting transcription, and "OFF", deactivating transcription (Figure 3f). The ON and OFF times are predicted to be exponentially distributed. To determine those parameters from imaging data, transcriptional dynamics of each allele should be separately quantified. Although non-transcribing alleles were not actually traceable, their position is roughly estimated by the somewhat uneven nuclear background signals of NLS-MCP-mNeonGreen and changes in nuclear shape (Supplementary Video S1 and S2). By this qualitative estimation, distributions of ON and OFF time durations were obtained (Figure 3g). ON and OFF time distributions were well fitted by exponential distributions, consistent with the Poisson stochastic processes expected in the random telegraph model[30]. Furthermore, the mean duration of ON time in 2i conditions was longer than that in serum conditions. Conversely, the mean duration of OFF time in 2i conditions was shorter than that in serum conditions. These results are consistent with the increase in transcription frequency in 2i conditions (Figure 3c). Collectively, these findings suggest that *Nanog* transcription might be regulated by modulation of promoter states.

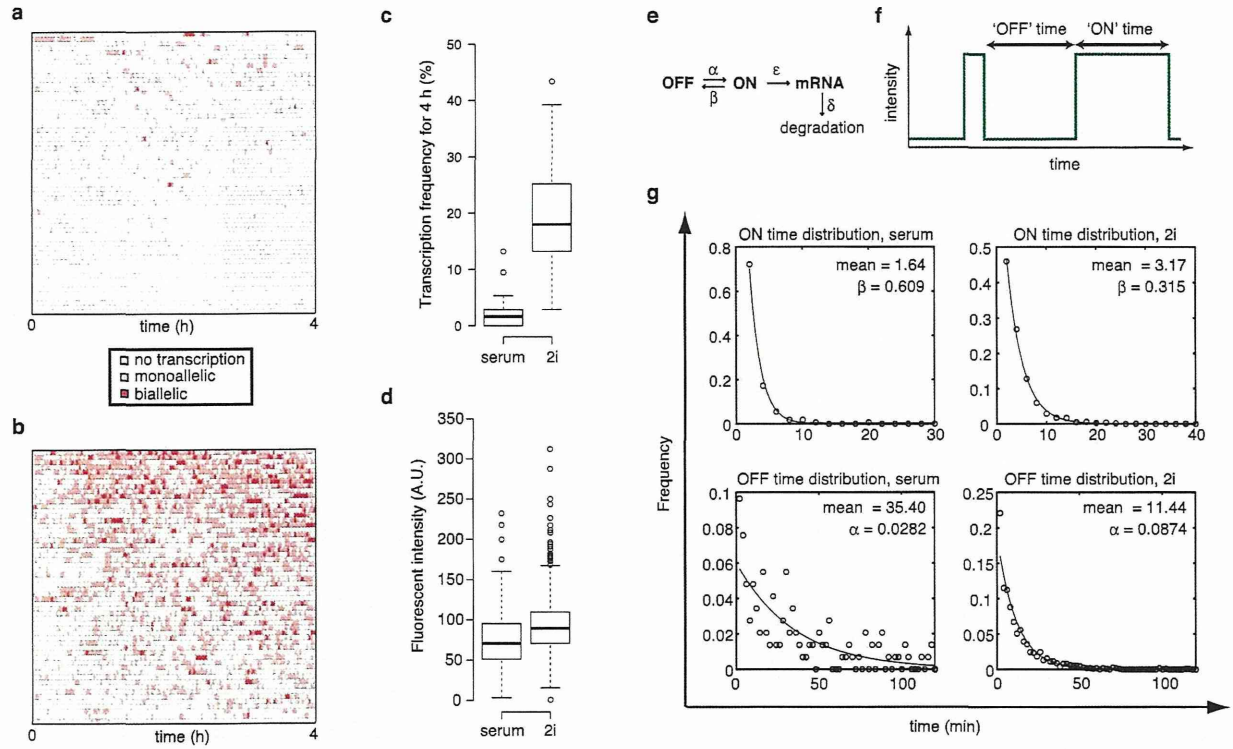


Figure 3 | Transcription dynamics of *Nanog* in mESCs. (a, b) NM-G cells cultured in serum (a) and 2i conditions (b) were imaged at 2-min intervals for 4 h. Each line of the color-coded graph represents the transcription dynamics of each cell. (c) Distribution of transcription frequencies over a 4-h period for NM-G cells cultured in serum or 2i conditions [(total counts of transcribed alleles in each cell at 121 time points)/242] (n = 62 and 56 for serum and 2i conditions, respectively). (d) Distribution of fluorescence intensity at the transcription site in cells cultured in serum or 2i conditions. Each value was subtracted by the mean fluorescent intensity of each nucleus (n = 171 and 565 sample points). (e) A schematic presentation of the random telegraph model. "OFF" and "ON" depicts the OFF and ON states of promoters, respectively. OFF to ON and ON to OFF transition rates were defined as α and β, respectively. mRNA is assumed to be transcribed from only ON state promoter at rate ε. mRNA is degraded at rate δ. (f) A schematic representation of fluorescence intensity dynamics of an allele. (g) Distributions of ON and OFF time of *Nanog* transcription in cells cultured in serum or 2i conditions. Each probability distribution was fitted with an exponential function. Mean values are displayed in each graph.

If the *Nanog* transcription kinetics could be explained by the random telegraph model, the mean quantity of mRNA at steady state can be predicted as $<mRNA> = \gamma(\alpha/(\alpha+\beta))(\varepsilon/\delta)$ (Figure 3e). Here, γ is the copy number of *Nanog* gene; α and β are the reciprocal of the means of OFF and ON time durations, respectively (Figure 3g). The degradation rate of MS2-integrated and wild-type (WT) *Nanog* mRNA, δ, was determined to be 0.00348 and 0.00245 min^{-1}, respectively (Supplementary Figure S1). A least square fit of the random telegraph model to the experimentally obtained means of *Nanog* transcripts in NM and parent cells cultured in serum and 2i conditions reveals that the transcription rate at ON state $\varepsilon = 2.11$ min^{-1} (Figure 2c). In both cell types cultured in serum conditions, the fitted means of *Nanog* mRNAs were considerably lower than those of the experimentally obtained values. Although we performed live imaging at 2-min intervals, the estimated mean duration of ON time in serum conditions was shorter than 2 min (1.64 min, Figure 3g). Therefore, it is possible that not all the transcriptional pulses had been detected and that we underestimated the transcriptional frequency in serum conditions. To confirm whether the value of ε is realistic or not, we performed further quantitative analysis. The value of ε could be predicted from the number of nascent mRNAs remaining at each transcription site. When we imaged the NM-G cells at higher magnification, not only transcriptional bright spots but also other relatively weak and fast-moving signals, which are assumed to be individual mRNAs, were observed (Supplementary Video S3 and Figure S3).

From the comparison of individual mRNA signals and transcriptional spot signals, we estimate that there are 2.79 ± 0.55 [mean ± standard deviation, n = 15] nascent NLS-MCP-mNeonGreen-tagged mRNAs per transcription site. The sum length of an MS2 repeat and the 3′ untranslated region of *Nanog* mRNA is 2.55 kb; therefore, each nascent mRNA is roughly estimated to be transcribed by RNA Pol II at 946 ± 184-bp intervals. Accordingly, the RNA Pol II elongation rate was estimated to be 1.99 ± 0.39 kb/min.Recently, the RNA Pol II elongation rate was reported to be 0.5–4 kb/min (mean and median are 1.793 and 1.824 kb/min, respectively) in mESCs[25] (Supplementary Figure S4), suggesting that the estimated ε seems to be a realistic value.

In the random telegraph model, mRNAs are transcribed in busts during promoter switches from an inactive to active state, and the average size of the transcriptional bursts should be described as ε/β[30]. Because of the increase in the mean ON time duration in 2i conditions, the burst size in 2i conditions is larger than that in serum conditions (3.46 and 6.69 for serum and 2i conditions, respectively). Collectively, these suggest that *Nanog* transcription can be explained by the random telegraph model.

Stochastic promoter activation significantly contributes to expression variability of Nanog. Heterogeneity in gene expression is induced by several factors[12]. One factor is the presence of stochasticity that is inherent to the biochemical process of gene expression, called the 'intrinsic noise'. Therefore, the stochastic promoter activation observed in *Nanog* transcription could be a result of intrinsic noise. On the other hand, other effects, including variability of cellular components (such as RNA Pol II or other regulatory molecules), asynchronous cell cycle, and heterogeneous

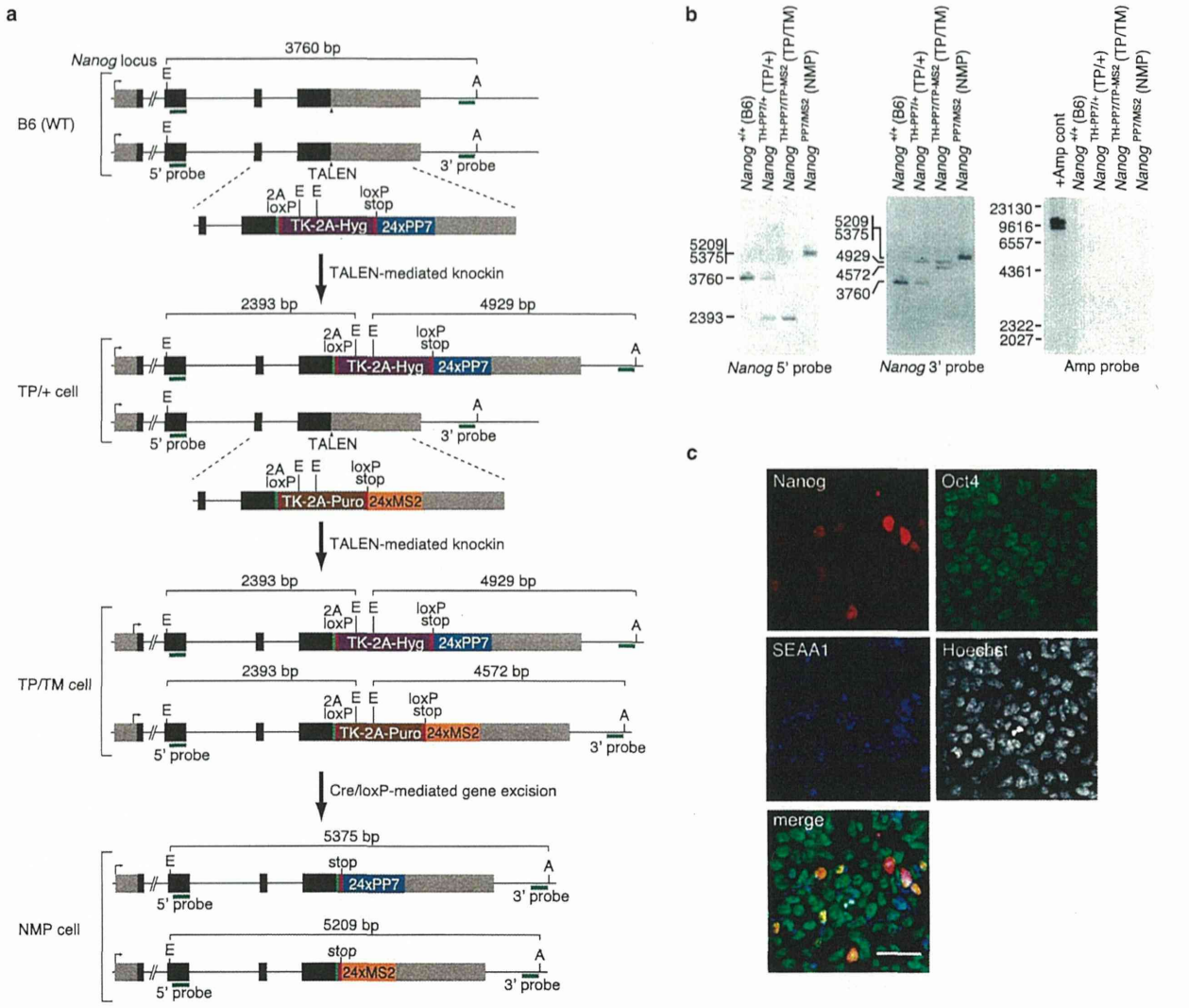


Figure 4 | Establishment of a reporter mESC line (NMP). (a) The strategy for targeted gene integration into the *Nanog* locus for establishment of MS2-PP7 targeted mouse embryonic stem cell (mESC) line (NMP). Grey, black, green, and magenta boxes indicate untranslated region, Nanog coding sequences, and coding sequences for 2A peptide and loxP site, respectively. Green bars represent the positions of Southern blot probes. E, EcoNI; A, AflII. (b) Southern blot analyses showing transcription activator-like effector nuclease (TALEN)-mediated insertion and Cre-mediated excision of the selection cassettes in modified mESC clones. (c) Immunofluorescence showed expression of pluripotent markers in the targeted cell line (NMP cells). Each image is a maximum intensity projection of image stacks. Scale bar, 50 μm.

inter-cellular signaling also affect gene expression variability and are called 'extrinsic noise'. These two types of noise can be discriminated by two-reporter assays[12,13]. If different reporters are integrated into each allele, distribution of each reporter enables the determination of how intrinsic noise contributes to the total expression variability in the population[12,13].

To determine the type of noise that is dominant in the expression of *Nanog*, we established another cell line (NMP) in which repeats of MS2 and PP7 were integrated immediately upstream of each allele of the endogenous *Nanog* stop codon (Figure 4a). PP7 is used to visualize mRNAs[32]. Southern blot analysis showed the expected band patterns and no random integrations (Figure 4b). The derived cells express the undifferentiated embryonic stem-cell markers, SSEA1 and Oct4 (Figure 4c). To quantify *Nanog* transcription dynamics on each allele, we introduced NLS-MCP-mNeonGreen and NLS-PP7 coat protein (PCP)-red fluorescent protein (RFP), which binds to the PP7 RNA stem loop. However, a tendency toward nucleoli localization of NLS-PCP-RFP prevents quantification of the transcription dynamics of *Nanog-PP7*. Using *Nanog* exonic and MS2 probes, we performed smFISH analysis of NMP cells cultured in serum and 2i conditions (Figure 2a and 5a). Among the *Nanog* probe-positive spots, those with an MS2 probe signal intensity above the threshold value were assumed to be mRNAs expressed from the *Nanog-MS2* allele; otherwise, mRNAs were considered to be expressed from the *Nanog-PP7* allele (Figure 5b, Supplementary Figure S5). To confirm the accuracy of this methodology, we also performed the smFISH using MS2 and PP7 probes in NMP cells cultured in 2i conditions (Supplementary Figure S5). The distribution of mRNA counts was similar to that obtained using *Nanog* exonic and MS2 probes (Figure 5b), suggesting that the method is reasonably accurate.

Intrinsic noise η^2_{int}, extrinsic noise η^2_{ext}, and total noise η^2_{tot} (sum of intrinsic and extrinsic noises) can be calculated using obtained allele-specific mRNA distribution data[12,13] (Figure 5b). Consistent with other reports, the total noise (heterogeneity in *Nanog* expression) in serum conditions was larger than that in 2i conditions (Figure 5b)[6]. Surprisingly, intrinsic noise contributed to approximately 45% of the total noise in both conditions, sug-

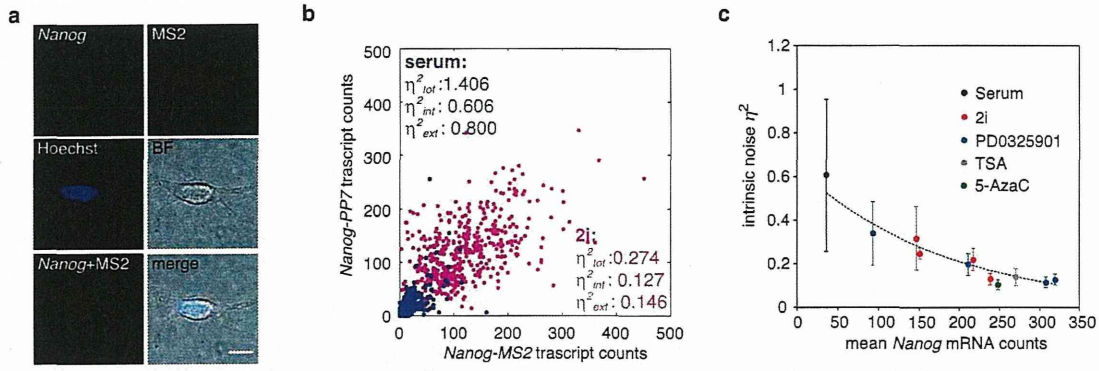


Figure 5 | Intrinsic noise significantly affects expression variability of *Nanog*. (a) Allele-specific single-molecule fluorescent *in situ* hybridization (smFISH) analysis. NMP cells cultured in serum condition were subjected to smFISH analysis using *Nanog* exonic smFISH probes and MS2 probes. Cellular nuclei were counter-stained with Hoechst33342. Each image is a maximum intensity projection of image stacks. The probe position is described in Figure 2a. Scale bar, 10 μm. (b) Scatter plot of *Nanog-MS2* and *-PP7* transcripts. *Nanog-MS2* and *-PP7* transcripts were counted in each cell by allele-specific smFISH. η^2_{int}, η^2_{ext}, and η^2_{tot} represent intrinsic, extrinsic and total noise, respectively. The Fano factors of *Nanog* mRNA are 39.9 and 50.3 in serum and 2i conditions, respectively. (c) Scatter plot of mean *Nanog* mRNA counts and intrinsic noise. Distributions of *Nanog-MS2* and *-PP7* mRNAs in cells cultured in several conditions were obtained by allele-specific smFISH analyses as in (b). Afterward, the mean *Nanog* mRNA counts and noise values were calculated and plotted. Dashed line represents a trend line obtained from data of cells cultured in media containing 2i and PD0325901 as calculated using Microsoft Excel. Error bars are 95% confidence intervals obtained by bootstrapping. Concentrations of inhibitors in each culture condition are listed in Supplementary Table S1.

gesting that intrinsic noise significantly affects *Nanog* expression variability.

Models of stochastic gene expression predict that intrinsic noise should increase as the amount of transcript decreases[33]. To change the mean expression level of *Nanog* mRNA, we cultured NMP cells in culture media containing several concentrations of 2i inhibitors and PD0325901, an FGF/Erk signal inhibitor (one of the 2i inhibitors) that increases *Nanog* expression[10] (Figure. 5c and Supplementary Table S1). As expected, intrinsic noise monotonously increased as the mean *Nanog* mRNA counts decreased (Figure 5c), suggesting that intrinsic noise significantly affects *Nanog* expression variability. We investigated the effects of histone acetylation and DNA methylation on intrinsic noise by treatments with Trichostatin A (TSA, a histone deacetylase inhibitor)[34] and 5-azacytidine (5-AzaC, a DNA-demethylating agent)[35] in 2i conditions (Fig. 5c), as these modifications influence gene activity. However, intrinsic noises in cells cultured with TSA and 5-AzaC were not strongly deviated from a trend line obtained from data of cells cultured in media containing 2i and PD0325901 (Fig. 5c), suggesting that histone acetylation and DNA methylation might not affect *Nanog* intrinsic noise at least in 2i conditions.

Given the above results, we evaluated whether the mRNA expression level in individual cells correlates with Nanog protein levels. To address this question, we performed smFISH using *Nanog* exonic probes followed by immunofluorescence using Nanog antibody (Figure 6), and found that *Nanog* mRNA and Nanog protein levels were well correlated in both serum and 2i conditions ($r = 0.85$ and $r = 0.72$ for serum and 2i conditions, respectively) (Figure 6). This finding suggests that Nanog protein heterogeneity originates from *Nanog* mRNA heterogeneity.

Discussion

We have demonstrated that the stochastic promoter activation significantly affects expression variability of Nanog in mESCs. Nanog expression variability is observed not only in mESCs cultured *in vitro* but also in pre-implantation embryonic inner cell mass (ICM) cells[36]. The biological meaning of heterogeneity is not fully understood, but some researches suggest that it may play a functional role in cell fate decisions[3,4,37]. It has been reported that the genome-wide epigenetic status of mESCs cultured in serum and 2i conditions resemble that of later ICM cells before differentiation and that of early ICM cells, respectively[38]. In addition, a recent report suggests that expression fluctuations observed in several genes including *Nanog* in both early- and late-stage ICM cells underlie lineage choices[39].

Our data indicated substantial expression variability of *Nanog* mRNA not only in serum but also in 2i conditions (Figure 2c and 5b). The Fano factor is the ratio between the variance and the mean of the mRNA copy number distribution, and is a key parameter used to quantify the deviation from Poisson statistics. For a complete Poisson distribution, the Fano factor equals 1. Distribution of *Nanog* mRNA transcripts in NMP cells were non-Poissonian (Fano factor $\gg 1$), suggesting that heterogeneity in *Nanog* expression is relatively high in both conditions. This observation is consistent with a previous report by Abranches et al. on the expression dynamics of the Nanog protein; using a bacterial artificial chromosome transgenic reporter and smFISH analysis, significant variability was observed in Nanog expression in mESCs cultured in 2i conditions[37]. Furthermore, we found that intrinsic noise derived from stochastic promoter activation significantly affects expression variability. Recently, Singer et al. reported that expression levels of genes, including *Nanog*, fluctuate between cells due not only to stochastic gene expression, but also to transitions between states with different gene activation potential[28]. This is consistent with our observation of multimodal distribution of transcription frequency over 4 h (Supplementary Figure S2). To compare the *Nanog* transcription dynamics among cells belonging to different states, we divided the cells into two groups according to their transcription frequency (lower or higher; Supplementary Figure S6). However, the distributions of ON and OFF time of *Nanog* transcription between the groups did not show statistically significant differences in any of the culture conditions (Supplementary Figure S6). Singer et al. suggested that metastable states are correlated with DNA methylation in mESCs[28]. However, inhibition of DNA methylation by DNA methyltransferase inhibitor 5-AzaC (our study) or inactivation of all three DNA methyltransferases[10,28] scarcely affects *Nanog* expression. Further analysis is necessary to understand what mechanisms determine the various *Nanog* expression states.

In our system, insertion of MS2 repeats into the *Nanog* locus slightly destabilizes *Nanog-MS2* mRNA (Supplementary Figure S1). In NMP cells, in which MS2 and PP7 repeats were integrated into the *Nanog* locus, *Nanog-MS2* and *Nanog-PP7* mRNA showed similar expression levels (Supplementary Figure S5), suggesting that

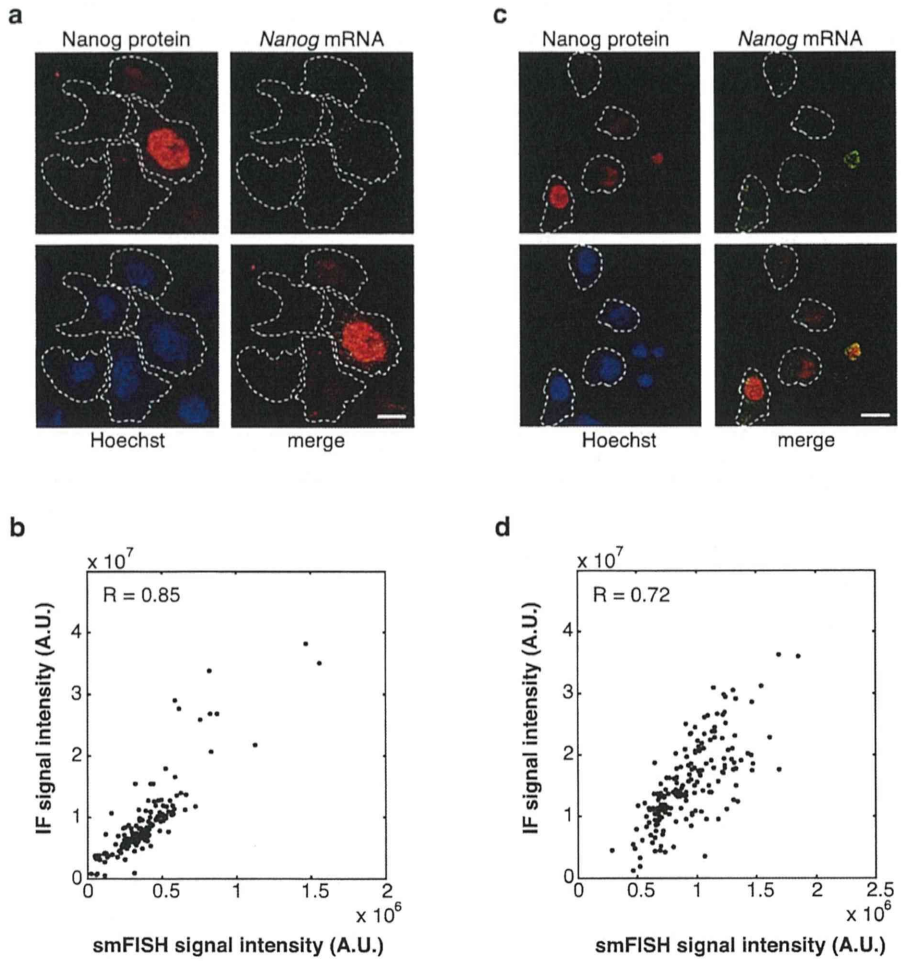


Figure 6 | Nanog protein heterogeneity originates from *Nanog* mRNA heterogeneity. C57BL/6 mESCs cultured in serum (a, b) and 2i conditions (c, d) were subjected to smFISH using *Nanog* exonic probes and followed by immunofluorescence (IF) using Nanog antibody. (a, c) Nuclei were counterstained with Hoechst33342 (Hoechst). Dashed lines represent edges of cell membranes of nondividing cells. Scale bar, 20 μm. (b, d) Scatter plot of fluorescence intensities of IF (Nanog protein) and smFISH (*Nanog* exonic probes).

integration of PP7 also affects *Nanog* mRNA degradation rate. Therefore, WT *Nanog* mRNA stays intact longer than *Nanog-MS2* (or *-PP7*) mRNA in cells. This might mask the cell-to-cell expression variability in *Nanog* mRNA in WT mESCs; in other words, it is possible that intrinsic noise of *Nanog* mRNA in WT mESCs is lower than that in NMP cells. However, it has been recently reported that considerable intrinsic noise in *Nanog* mRNA expression seemed to exist in hybrid mESCs without integration of reporter genes[11]. Therefore, it is possible that intrinsic noise has a considerable effect on Nanog expression variability in WT mESCs as well as in NMP cells.

Although pulsatile transcriptional events or "transcriptional bursting" has been reported in several model systems, the underlying molecular mechanisms are still elusive[15]. One of the models of transcriptional bursting is the chromatin-based model[13,15,40]. In this model, the efficiency of transcription depends on the absence of nucleosomes, which compete with the binding of transcription factors immediately upstream of the transcription start site (TSS). Therefore, the promoter activation timescale depends on relatively slower nucleosome turnover[41]. Recently, it has been reported that nucleosome occupancy immediately upstream of the *Nanog* TSS is inversely correlated with Nanog expression level[42], implying that the stochastic promoter activation of *Nanog* may originate from the relatively slow nucleosome dynamics.

Another potential source of transcriptional bursting includes DNA conformation changes involving efficient transcription. Some genes are regulated via long-range interaction between promoters and enhancers; because such long-range interactions seem to be variable among cells[43], the regulatory mechanism could be a source of expression variability[44]. It has been reported that the *Nanog* promoter region is associated with several regions genome-wide[45], suggesting that genome-wide stochastic association between *Nanog* promoters and enhancers may underlie transcriptional bursting. Further investigation is needed to understand the molecular mechanism of the *Nanog* promoter activation.

In summary, *Nanog* transcription dynamics were quantitated using the MS2 system in mESCs. We found that the promoter activation occurs in a pulsatile and stochastic manner. Furthermore, allele-specific smFISH analysis revealed that intrinsic noise considerably contributes to the Nanog heterogeneity. Therefore, we conclude that stochastic processes of promoter activation might be a key source of the intrinsic noise, and hence significantly affect Nanog expression variability. The combination of the MS2 system and smFISH analysis seems to be useful for evaluating stochastic promoter activation and expression variability at single-cell resolution. The techniques used in the present report will help further the understanding of the molecular basis of allelic expression in mESCs[46] and their heterogeneity.