Published online 15 June 2018

Nucleic Acids Research, 2018, Vol. 46, No. 17 ***e100***
doi: 10.1093/nar/gky501

CRISPR/Cas9-mediated knock-in of an optimized TetO repeat for live cell imaging of endogenous loci

Ipek Tasan[1], Gabriela Sustackova[2], Liguo Zhang[2], Jiah Kim[2], Mayandi Sivaguru[3], Mohammad HamediRad[3,4], Yuchuan Wang[5], Justin Genova[4], Jian Ma[5], Andrew S. Belmont[2,3,6] and Huimin Zhao[1,3,4,6,7,8,*]

[1]Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA, [2]Department of Cell and Developmental Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA, [3]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA, [4]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA, [5]Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA, [6]Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA, [7]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA and [8]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Received August 6, 2017; Revised May 17, 2018; Editorial Decision May 18, 2018; Accepted June 13, 2018

ABSTRACT

Nuclear organization has an important role in determining genome function; however, it is not clear how spatiotemporal organization of the genome relates to functionality. To elucidate this relationship, a method for tracking any locus of interest is desirable. Recently clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) or transcription activator-like effectors were adapted for imaging endogenous loci; however, they are mostly limited to visualization of repetitive regions. Here, we report an efficient and scalable method named SHACKTeR (Short Homology and CRISPR/Cas9-mediated Knock-in of a TetO Repeat) for live cell imaging of specific chromosomal regions without the need for a pre-existing repetitive sequence. SHACKTeR requires only two modifications to the genome: CRISPR/Cas9-mediated knock-in of an optimized TetO repeat and its visualization by TetR-EGFP expression. Our simplified knock-in protocol, utilizing short homology arms integrated by polymerase chain reaction, was successful at labeling 10 different loci in HCT116 cells. We also showed the feasibility of knock-in into lamina-associated, heterochromatin regions, demonstrating that these regions prefer non-homologous end joining for knock-in. Using SHACKTeR, we were able to observe DNA replication at a specific locus by long-term live cell imaging. We anticipate the general applicability and scalability of our method will enhance causative analyses between gene function and compartmentalization in a high-throughput manner.

INTRODUCTION

Spatiotemporal organization of the mammalian genome within the nucleus is highly regulated (1,2); however, the link between subnuclear localization and gene function remains elusive. To relate genome function to higher order nuclear organization, a direct, microscopy-based method for live cell tracking of the dynamics of any specific endogenous locus of interest is necessary.

DNA fluorescence *in situ* hybridization (DNA-FISH) is a commonly used method for imaging specific regions within the chromosome but is technically challenging. Because of the harsh DNA denaturation conditions required, structural preservation is poor, yet increasing fixation strength to counteract this structural perturbation results in decreased detection efficiency. DNA-FISH also frequently has high background with both false positive and negative rates. Finally, DNA-FISH is incompatible with tracking dynamics of DNA.

Live cell imaging of DNA was previously carried out by using a fluorescent repressor-operator system (3,4) to enrich fluorescent proteins (FPs) at a specific site on the DNA. In the two commonly used systems, repeating sequences of Lac operators (LacO) or Tet operators (TetO) are used as a DNA tag and FP-fused Lac repressor (LacI) or Tet repressor (TetR), respectively, is used for visualization of the tag. Most previous examples focused on plasmid or BAC (bacterial artificial chromosome) integrated transgenes, although their behavior may not fully recapitulate the behavior of the

*To whom correspondence should be addressed. Tel: +1 217 333 2631; Fax: +1 217 333 5052; Email: zhao5@illinois.edu

endogenous locus. In a more recent example, these operator arrays were targeted to endogenous loci using homologous recombination (HR) (5). However, a low targeting efficiency was observed, indicating the need for better targeting strategies with higher efficiency.

Discovery of novel modular proteins such as transcription activator-like effectors (TALEs) and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated protein 9 (Cas9), whose DNA recognition specificity can be easily tailored, lead to alternative approaches. CRISPR/Cas9 system includes a single guide RNA (sgRNA) that recognizes a specific 20 nt DNA sequence and recruits the Cas9 endonuclease to the target DNA (6). If the target sequence is followed by a protospacer adjacent motif (PAM), which is 5′-NGG-3′ for Cas9 from *Streptococcus pyogenes*, then Cas9 can induce a double strand break (DSB) ∼3 nt upstream of the PAM. Coupled with fluorescent tags, both TALEs and a catalytically inactive version of Cas9 (dCas9) were used for visualization of naturally occurring repetitive sequences within mammalian genomes (7–10). However, since repetitive sequences are not equally distributed throughout the genome, other strategies are necessary for visualizing non-repetitive regions. A set of many TALEs or sgRNAs would be required for visualizing non-repetitive regions, which could be technically challenging and inefficient.

In addition to imaging, increasing the efficiency of homology-directed genome editing by inducing DSBs is another application of CRISPR/Cas9. In mammalian cells, non-homologous end joining (NHEJ) is the predominant mechanism for repairing DSBs; however, in the presence of a DNA homologous to the sequences flanking the DSB site, such as an exogenous donor DNA with suitable homology arms (HAs), the DSB can be repaired by HR as well (11). Depending on the design of the donor, gene knock-in, deletion or replacement can be achieved. One unexplored application of the CRISPR-mediated knock-in is the tagging of the DNA itself to overcome the need for a pre-existing repeat for visualization of chromosome loci.

Here, we developed an easy, efficient and scalable method named SHACKTeR (short homology arm and CRISPR/Cas9 mediated knock-in of a TetO repeat) for tagging and live cell imaging of non-repetitive, endogenous chromosome regions. We created an optimized, irregular TetO repeat, which allowed polymerase chain reaction (PCR)-based creation of donors with short HAs and simplified the protocol. Irregular TetO repeats also enabled easy genotyping of positive TetO-labeled clones by PCR. SHACKTeR achieved targeted insertion at 10 different loci and we demonstrated the general applicability of SHACKTeR by showing the feasibility of knock-in into heterochromatin sites. We also showed that lamina-associated, heterochromatin sites prefer knock-in via NHEJ pathway and chromatin state can affect knock-in efficiency. SHACKTeR was highly specific given the lack of additional off-target spots. Moreover, integration of TetO repeats did not affect sub-nuclear localization, chromatin status and DNA replication dynamics of the integration site. The ease-of-use and the scalability of SHACKTeR will allow labeling large numbers of endogenous chromosome loci for both fixed and live-cell imaging, which could lead to an improved understanding of the relationship between gene compartmentalization and function.

MATERIALS AND METHODS

sgRNA and homology arm design

BACs containing human DNA inserts within selected regions were identified (Supplementary Table S1) and the DNA sequences corresponding to these inserts (region 1–10) were retrieved from UCSC Genome Browser (12). Transcriptional status of these regions were determined based on the ENCODE Caltech RNA-Seq data for HCT116 and K562 cells (13,14), which were displayed on the UCSC Genome Browser (15). RNA-Seq data for HCT116 (UCSC accession number: wgEncodeEH001425; GEO accession number: GSM958749) and K562 (UCSC accession number: wgEncodeEH000124; GEO accession number: GSM958729) were created by Wold lab as part of the ENCODE Project (13).

To design 20 nt guide sequences of sgRNAs, DNA sequences of the sub-regions without any exons were retrieved from the UCSC Genome Browser. UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly was used for all the analyses in this paper. Within the sub-regions, repetitive and low complexity DNA sequences were annotated by RepeatMasker (16) (http://www.repeatmasker.org) in the UCSC Genome Browser and sequences with masked repeats were retrieved. The unmasked regions were given to CRISPRdirect web tool (17) using a custom Application Program Interface developed in Python programming language to obtain a list of candidate guide sequences for each region. All guide sequences that could result in difficulty in DNA synthesis (more than six consecutive A's, C's and G's), sgRNA transcription (more than four consecutive T's) and DNA assembly (BpiI recognition site) were removed from the list. The remaining guide sequences were sorted based on GC content, specificity and efficiency using the scoring criteria described in Supplementary Table S2. The weight for the effect of each nucleotide on the efficiency is obtained from a previously published work (18).

Candidate sgRNAs with high scores (67 or more) were further evaluated in terms of their alignment with regulatory sites. UCSC Genome Browser (12) was used to search for possible transcription factor binding sites (TFBS). We used Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs track to find potential binding sites (12,15). We also retrieved enhancer-like sequences (http://zlab-annotations.umassmed.edu/enhancers/) and promoter-like sequences (http://zlab-annotations.umassmed.edu/promoters/) in HCT116 cells (13,15). sgRNA target sites that align with any of these three regulatory sites were excluded.

At last, sgRNAs with predicted DSB sites (3 nt upstream of PAM) closer than 50 nt to the masked repetitive regions were omitted to provide unique HAs. If an sgRNA with a good score could not be found within the regions in Supplementary Table S1, then the sequences that were ∼50 kb upstream or downstream of the region were also screened. A list of coordinates of selected sgRNA target sites can be found in Supplementary Table S1. A list of oligonucleotides

used for cloning the guide sequences can be found in Supplementary Table S3.

HAs were designed to exclude PAM sequence and also the first 3–4 nt upstream of the expected cut site (See Supplementary Table S4). Point mutations were introduced to the PAM if the HA included the PAM sequence (as for region 3). HAs were further optimized to minimize secondary structures.

PCR for creating linear donor DNA

Sequences of the primers used for creating linear donor DNAs can be found in Supplementary Table S5. Primers were ordered from Integrated DNA Technologies (IDT, Coralville, IA) and they contain phosphorothioate bonds between the last 3 nt at the 5′ end of the primers to reduce degradation of linear donor DNA (19). Either pSP2-48-merTetO-EFS-BlaR or pSP2-96-merTetO-EFS-BlaR plasmid was used as a PCR template, depending on the desired final donor DNA. PCRs were performed using Q5 High Fidelity DNA Polymerase (New England Biolabs (NEB), Ipswich, MA), using manufacturer's protocol with the following changes to the PCR cycle: 98°C for 3 min; followed by 30 cycles of 98°C for 30 s, 70°C for 15 s and 72°C for 3 min. The final extension was 72°C for 6 min. The donor DNA PCRs were gel-purified with Gel Extraction Kit (Qiagen, Valencia, CA).

Cell culture, transfection and clonal isolation

Human embryonic kidney (HEK) cell line HEK293T was maintained in Dulbecco's modified Eagle's Medium (Corning Life Sciences, Tewksbury, MA, USA) supplemented with 10% heat inactivated fetal bovine serum (FBS) (Life Technologies, Carlsbad, CA, USA). HCT116 cells were a gift from Mark S. Kuhlenschmidt and were cultured in McCoy's 5A medium without phenol red (UIUC Cell Media Facility), supplemented with 10% tetracycline-free FBS (Sigma-Aldrich, St. Louis, MO). Both cells were grown at 37°C in a humidified 5% CO_2 incubator. BioCoat Collagen-I coated plates (Corning Life Sciences) were used for better attachment of HCT116 cells.

To perform transfections for knock-in of the TetO donors, Fugene HD (Promega, Madison, WI, USA) was used, following manufacturer's protocol. For the initial transfections to test knock-in efficiencies of the 48-mer and 96-mer TetO donors at the *HSP70* locus, 1:1 CRISPR/Cas9 plasmid to donor DNA molar ratio was used and 1 μg CRISPR/Cas9 plasmid was transfected to the cells at 70% confluency in 6-well plates. Accordingly, 0.3 μg of 48-mer TetO EFS-BlaR donor and 0.5 μg of 96-mer TetO EFS-BlaR donor were used. Blasticidin (10 μg/ml) selection was started 1 day after transfection. Seven days after blasticidin selection, clonal isolation was started by limiting dilution in 96-well plate. For the later knock-ins, we used 2:1 CRISPR/Cas9 plasmid to donor DNA molar ratio. Before transfection, cells were grown in 24-well plates until 40–50% confluency. A total of 500 ng CRISPR/Cas9 plasmid was transfected and the necessary linear donor DNA amount was calculated accordingly (83 ng for 48-mer TetO donor DNA). One day after transfection, cells were plated onto 100 mm plates (together with 10 μg/ml blasticidin) at limited dilution for growth of isolated colonies. Clonal isolation was performed following a previously published protocol (20).

Magoh-mCherry plasmid (a gift from Kannanganattu V. Prasanth) was transfected using Lipofectamine 2000 (Thermo Fisher Scientific, Waltham, MA), following manufacturer's protocol. Briefly, cells were grown in T-25 flasks until 60–70% confluency and transfected with 3 μg plasmid. Selection with G418 was started 2 days after transfection and cells were analyzed once the selection was completed.

mCherry-laminB1 plasmid was transfected using Fugene HD (Promega), following manufacturer's instructions, as described above. Cells were analyzed 2 days after transfection. Same transfection protocol was followed for the transfection of the p3′SS TetR-EGFP plasmid.

Genotyping and sequencing

Once the clones grown in 24-well plates reached ∼80% confluency, half of the cells were collected and genomic DNA (gDNA) was extracted using QuickExtract DNA Extraction Solution (Epicentre, Madison, WI). Briefly, 50 μl QuickExtract solution was added per cell pellet, vortexed and the mixture was incubated at 65°C for 15 min, followed by 98°C for 10 min. A total of 4 μl from the gDNA was used for genotyping PCRs. All the genotyping PCRs were performed using Q5 High Fidelity DNA Polymerase (NEB), using manufacturer's protocol. We used the following PCR cycle: 98°C for 3 min; followed by 35 cycles of 98°C for 30 s, 68°C for 15 s and 72°C for 3 min (or 5 min for the 96-mer TetO knock-in clones). The final extension was at 72°C for 6 min.

Sequences of the primers used for genotyping PCRs can be found in Supplementary Table S6. For the clones with correct genotyping PCR results, knock-in band was gel-purified with Gel Extraction Kit (Qiagen) and was digest-tested with SpeI enzyme (NEB).

Sanger DNA sequencing was performed (Genewiz, South Plainfield, NJ) to verify sequences of the left and right junctions as well as the integrity of the TetO repeats. A list of sequencing primers can be found in Supplementary Table S7.

Clone 10–22 had both alleles targeted and since the integration was through NHEJ, the alleles had differences in the sequence. To sequence each allele separately, 10–22 knock-in band was amplified with a modified version of the forward and reverse genotyping primers to integrate 5′-BsmBI (giving AATT 5′ overhang) and 3′-BamHI sites, respectively. The PCR product was digested with BsmBI and BamHI enzymes, followed by cloning into the pSP2 backbone through EcoRI and BamHI sites. Plasmids from single colonies were purified and sent for sequencing.

Image acquisition and analysis

For analysis of the immuno-FISH and immunofluorescence results, 3D optical-section images were acquired with 0.2 μm z-steps using a DeltaVision OMX microscope system and DeltaVision SoftWorx software (GE Healthcare, Little Chalfont, UK) with a 100×, 1.4 NA objective lens and

an Evolve 512 Delta EMCCD Camera with conventional light path. Images were deconvolved using an enhanced ratio, iterative constrained algorithm (21) and registered based on Image Registration Target Slide (GE Healthcare) and TetraSpeck fluorescent beads (Molecular Probes, Eugene, OR, USA). Image deconvolution and registration were done with the SoftWorx 6.5.2 software (GE Healthcare). The distributions were graphed in the program OriginPro (OriginLab, Northampton, MA, USA).

Images from immunofluorescence staining of knock-in clones or immuno-FISH of wild-type (wt) HCT116 cells were analyzed to determine the distance of locus of interest from either nuclear speckles or lamina (Figure 5). The same analysis was also done for Supplementary Movies S3 and 4 (Figure 8). The shortest distances were measured from the center of each EGFP or FISH spot to the edge of its nearest nuclear speckle or to the center of its nearest lamina region using the Straight Line tool in ImageJ (National Institutes of Health). All microscopy images in Figures 3, 5, 7 and 8 were prepared using ImageJ software.

To determine percentage of cells with paired loci and for analysis of the distance between the paired loci (Figure 4), a Structured illumination Super resolution system (Elyra SIM, Carl Zeiss, Obercohen, Germany) was used to analyze fixed cells. The system configuration was the same as described before (22). Briefly, a 1.4 NA Plan Apochromat 63× objective with oil immersion was used together with a 488 nm excitation and the emission signals were picked up using an Andor Ixon EMCCD (EM gain setting of 50 and exposure time ∼100 ms) camera between 500–550 nm bandpass filter. For each focal plane, five rotations and five phases were used and a z-stack was made at 130 nm distance between slices and the final processed XY pixel size was 40 nm. The raw data was processed using the Zen software supplied by the manufacturer under Structured Illumination processing module with default (automatic processing) settings. The distance between paired spots was determined with the same program using the profile module and XY calipers between the two spots. Imaris program (Bitplane AG, Zurich, Switzerland) was used to generate the XY and YZ profiles to determine and display the single and double locus status across cell types. The distributions were graphed in the program OriginPro (OriginLab).

All other fixed samples were analyzed using Personal DeltaVision deconvolution microscope, which is equipped with Xenon lamp, 60× oil objective (NA 1.4) and CoolSNAP HQ slow-scan CCD camera (Roper Scientific, Vianen, Netherlands). After taking z-stack images, the image files were deconvoluted and, where indicated, were projected into a 2D image using Softworx program (GE Healthcare).

For live cell imaging, DeltaVision OMX microscope was used as described above. Cells were grown in collagen-coated 35 mm glass bottom dish (Mattek, Ashland, MA, USA), and maintained at 37°C and 5% CO_2 during imaging. Post-processing of the images was done as described above, except Supplementary Movies S1 and 2 were deconvolved using the enhanced additive algorithm. Maximum projection was performed for Supplementary Movies S2 and 4 using Softworx (GE Healthcare). Correction for the rigid-body translation and rotation of the nucleus (23) was performed for Supplementary Movie S4 using ImageJ (National Institutes of Health). For Supplementary Movies S1 and 2, cells with non-S phase proliferating cell nuclear antigen (PCNA) pattern were selected for imaging. We analyzed the cells where both G1/S and S/G2 transitions were captured.

RESULTS

Optimized TetO multimers enable PCR-based creation of a linear donor DNA with short HAs

A major bottleneck in the traditional HR method is the laborious construction of donor vectors with long HAs, which impedes high-throughput applications. Based on previous work (19), we aimed at devising a one-step, cloning-free, PCR-based method to create a linear donor DNA containing a TetO repeat flanked by short HAs. More specifically, our goal was to optimize a scalable system in which we could PCR-amplify the TetO repeat and selection marker using primers with 50 nt 5′ extension sequences homologous to the target site (Figure 1A). The purified PCR product could then be used directly for transfection as a donor construct.

In previous studies, direct repeats of TetO or LacO sequences were used as tagging sequences. Such repeats create significant plasmid instability when propagated in bacteria, making downstream applications difficult and time-consuming. Moreover, PCR-amplification of direct repeats is challenging (24), which would impede creation of linear donors as described above. It is also difficult to determine the presence of tandem repeats by PCR, requiring more tedious protocols such as Southern blot for genotyping. To overcome these difficulties, we created irregular 48-mer and 96-mer TetO repeats, containing the 19 nt TetO sequence separated by random 10 nt sequences. Since RecA pairing in bacteria requires at least 25 nt homology, the presence of 10 nt random sequences between TetO repeats reduces recombinational and replicational instability of these sequences (25).

We also designed these 48-mer and 96-mer TetO repeats to contain no CpG dinucleotides that could be templates for DNA methylation in mammalian cells. We rationalized that the absence of DNA methylation would reduce the tendency of these TetO arrays to form heterochromatin after integration into the mammalian chromosome. The cloning strategy (illustrated in Figure 1B–D) was designed to produce TetO arrays where 10 nt spacer sequences were unique. The irregular nature of these TetO repeats enabled PCR amplification of both 48-mer and 96-mer TetO donors (Figure 1E).

Comparison of the 48-mer and 96-mer TetO repeat knock-in

Although 96-mer TetO repeats were used previously for visualization (26), we predicted that insertion of the shorter 48-mer repeat into the genome might be less perturbing. To determine the feasibility of 48-mer and 96-mer TetO knock-ins, we targeted the heat shock protein 70 (*HSP70*) locus containing three *HSP70* homologous genes (*HSPA1A, HSPA1B, HSPA1L*). We used the genetically stable, near diploid human HCT116 colorectal carcinoma cell line (27)

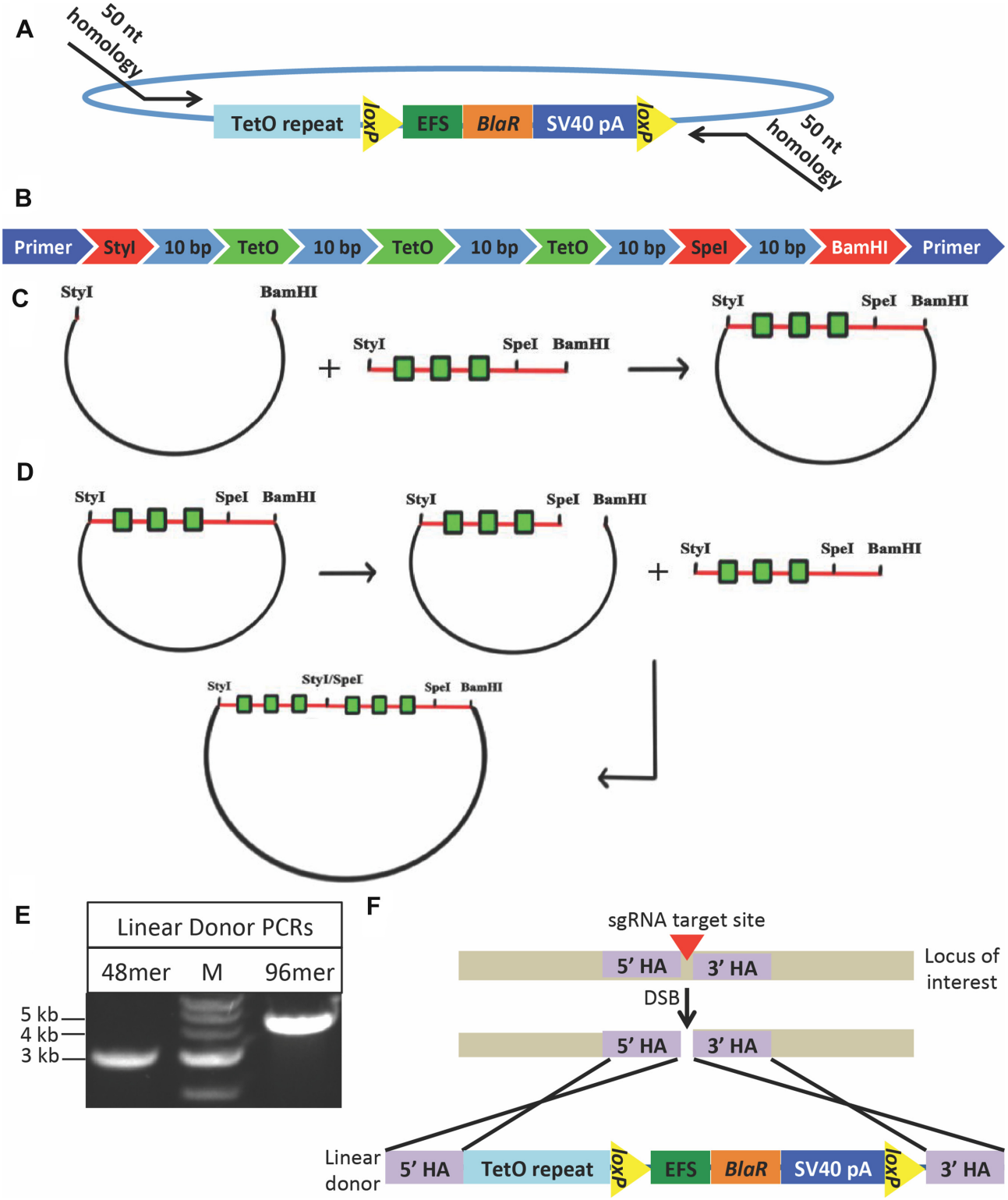

Figure 1. CRISPR/Cas9-mediated knock-in of an optimized TetO repeat using linear donor DNA with short homology arms (HAs). (**A**) Schematic of the strategy to create linear donor DNA with short HAs via PCR. The final plasmid used as a template is shown together with the primers used for amplification of the linear donor (black arrows). Blasticidin resistance (*BlaR*) gene under EFS promoter is for positive selection. Selection cassette is flanked with *loxP* sites. (**B**) Schematic of the complete initial cassette containing three TetO repeats (green), four random 10 bp (light blue) sequences, three restriction sites StyI, SpeI and BamHI (all of them are in red), and primer binding sites (dark blue) used for synthesis of the complementary strand and amplification of the product. (**C**) Schematic of the insertion of 3-mer TetO cassette into pSP2 vector through StyI and BamHI sites. Each TetO sequence is indicated with a green box. (**D**) Schematic of the assembly of the TetO repeat multimers. StyI/SpeI site, created by insertion of the TetO cassette, cannot be cut by either of these two enzymes. (**E**) Results of the PCRs to amplify 48-mer and 96-mer TetO repeat donors for the *HSP70* locus. The expected band sizes were around 3 kb for the 48-mer and around 4.6 kb for the 96-mer TetO donor. NEB 1 kb ladder (M) was used. (**F**) Schematic of the knock-in strategy. HAs are shown in purple. sgRNA target site is indicated with a red triangle. DSB: double strand break.

for all of our targeted insertions, which made genotyping results and the specificity of our targeting easier to interpret.

Since homology-directed repair of DSBs is inefficient in mammalian cells (28,29), we used CRISPR/Cas9 to increase the efficiency of the TetO repeat HR knock-in (Figure 1F). We designed an sgRNA targeting an intergenic region ∼16 kb downstream of the *HSPA1B* gene, avoiding regulatory regions in HCT116 cells. We also avoided repetitive or low complexity sequences within the sgRNA target site or HAs to minimize off-target integration.

PCR-amplified 48-mer and 96-mer TetO donors were transfected into HCT116 cells together with the CRISPR/Cas9 plasmid, which contained expression cassettes for both the sgRNA and the Cas9 protein. As a negative control for CRISPR/Cas9-mediated knock-in, we also transfected the cells with donor DNA only. After blasticidin selection, mixed pool cells were collected to determine the presence of cells with successful knock-in. Irregular TetO repeats made it possible to do genotyping by PCR. Primers for genotyping PCR were designed to specifically amplify the knock-in allele (Supplementary Figure S1A). After PCR analysis, we detected CRISPR/Cas9-dependent knock-in of both 48-mer and 96-mer TetO inserts (Supplementary Figure S1A).

We continued by isolating single cell clones, excluding all clones with noticeably slower rates of proliferation. The number of slow growing clones after CRISPR/Cas9-mediated knock-in of TetO repeats (1–3 clones) was not any different than the observations in wt cells when subcloned, which is likely due to the heterogeneity in the cell population. Genotyping of clones was first performed using primers which could amplify both knock-in and non-knock-in alleles (Supplementary Figure S1B), showing whether both, none or only one of the alleles were modified. We then performed PCRs to amplify left and right junctions of the knock-in cassette to determine the directionality of the insert (Supplementary Figure S1C) in the subset of clones that showed insertions at the targeted site. We digested the left junction PCR band with SpeI to further assay the correctness of the insert (Supplementary Figure S2A). At last, we sequenced the gDNA-insert junctions as well as the TetO repeats in the correctly edited clones. Sanger sequencing of TetO repeats was possible due to their irregular nature. In summary, 3/18 of the 48-mer clones (48-6, 48-15 and 48-17) and 2/17 of the 96-mer clones (clones 96–8 and 96-14) showed correctly targeted, heterozygous insertions of the intact TetO repeat. 48-6 was the only clone in which the integration did not happen through HR, but instead, the donor was directly ligated into the target site by NHEJ (Supplementary Text S1). Although integration occurred through NHEJ rather than HR, we counted this integration as correctly targeted because the 48-mer TetO repeat was confirmed to be intact by sequencing.

To test whether 48 TetO repeats were sufficient for visualization, clone 48-15 was transfected with a plasmid expressing TetR fused to EGFP. An analysis of mixed pool cells showed a single, clear EGFP-tagged spot in the nucleus of the cells expressing TetR-EGFP at appropriate levels (Supplementary Figure S2B), consistent with the heterozygous insertion demonstrated by genotyping. This microscopy observation of a single spot shows absence of additional off-target insertions of the donor.

Scalability of SHACKTeR: targeting nine additional chromosome loci with varying chromatin marks and sub-nuclear localizations

We used 48-mer TetO due to the possibility of a higher targeting efficiency and, more importantly, the expectation that the smaller repeat size would be less likely to perturb the normal chromatin structure.

Recent work suggested that the cutting efficiency of CRISPR/Cas9 may depend on the chromatin structure, with less efficient cutting observed at heterochromatic sites (30). It was also shown that nucleosome occupancy could result in decreased Cas9 cutting, which further supports that chromatin structure can affect CRISPR/Cas9-mediated editing (31,32). In addition to Cas9 cutting efficiency, previous work showed that HR efficiency can also be affected by the transcriptional and chromatin status of the region, with active gene regions with euchromatic histone marks showing higher HR efficiency (33). Moreover, lamina-associated domains were shown to prefer end joining for DSB repair (34). Thus, in addition to the *HSP70* locus that is known to be highly active in most of the cancer cells (35), we wanted to further test the efficiency of SHACKTeR targeting both euchromatic and heterochromatic loci localized in different nuclear compartments.

We used a new method, TSA-Seq (BioRxiv: https://doi.org/10.1101/307892), to guide selection of target chromosome loci showing different nuclear compartment localization. Using TSA-Seq, relative cytological distances of chromosomes genome-wide relative to the nuclear lamina and nuclear speckles was estimated in human K562 cells. Based on this TSA-mapping, we selected four chromosome regions (1, 2, 4 and 5) that were expected to be nuclear speckle-associated (data not shown). Nuclear speckles are enriched in pre-mRNA splicing factors and mostly interact with active gene regions (36–38). For instance, *HSP70* locus (region 3) and HSP70 transgenes were known to interact with nuclear speckles when active (39,40). In K562 cells, the *HSP70* locus also was predicted to be speckle-associated by TSA-Seq analysis. As expected, these five nuclear speckle-associated sites all corresponded to gene-rich and transcriptionally active regions in both K562 and HCT116 cells (Supplementary Figure S3).

Alternatively, we selected four chromosome regions (regions 6–9) that were predicted to associate with nuclear lamina in K562 cells based on both lamin TSA-Seq and lamin B1 DamID data (data not shown). Lamina-associated sites correspond to gene-poor, mostly inactive chromatin domains (41–43). These four regions all showed low levels of transcriptional activity in both K562 and HCT116 cells (Supplementary Figure S3). Additionally, we selected the β-globin locus (*HBB*) as our 10th targeting region. Although in K562 cells this locus is active and not associated with the lamina, in other tissues and cell lines in which this locus is inactive, the β-globin locus associates with the nuclear periphery, or centromeric heterochromatin in some cell types (44,45). Surprisingly, region 10 has high transcriptional activity in HCT116 cells (Supplementary Figure S3) but is still

associated with the lamina as determined by DNA-FISH. More information about all 10 regions can be found in Supplementary Table S1.

To knock-in an insert for the purpose of labeling a locus, high compaction of mammalian chromatin introduces a considerable flexibility for choosing an appropriate target sequence. Using FISH, hybridization of BAC probes of ∼200 kb still appears as a near diffraction-limited spot by wide-field light microscopy; therefore, a distance of up to ∼100 kb from the locus of interest will locate at the same location as resolved by conventional fluorescence microscope. Thus, we searched the entire region (Supplementary Table S1) to find the best sgRNA for knock-in into 9 additional loci. The criteria used to choose sgRNAs were similar to the ones described for the *HSP70* locus. Supplementary Table S1 lists the final target sites for all 10 regions.

Inspection of the ENCODE chromatin status in the 70 kb region surrounding each of these 10 target sites in HCT116 cells (Supplementary Table S8 and Figures S4 and 5) confirms that the lamina-associated regions have more heterochromatin characteristics compared to the speckle-associated regions, as expected (38,46,47). More specifically, all speckle-associated regions had enrichment of one or more of the chromatin marks (H3K4me3, H3K4me1, H3K27ac, H3K36me3) associated with active regions (Supplementary Figure S4). All lamina-associated sites lacked these euchromatin marks and some had enrichment for repressive chromatin marks (H3K9me3, H3K27me3) (Supplementary Figure S5).

To improve knock-in efficiency, we increased CRISPR plasmid to donor DNA molar ratio from 1:1 to 2:1. PCR-amplification of the donor DNA for all 10 regions was successful (Figure 2A). PCR analysis of mixed pool cells after selection showed successful knock-ins for 9 out of the 10 targeted regions. (Figure 2B and C). Region 8 was the only locus for which we could not detect a knock-in PCR product. Interestingly, the sgRNA for targeting region 8 is the only sgRNA that overlaps with a region enriched in the heterochromatin marker H3K9me3 (Supplementary Figure S5), which may have reduced knock-in efficiency at this location.

We performed Tracking of Indel DEcomposition analysis to determine cutting efficiency of region 8 sgRNA and compared it to region 3 sgRNA as a positive control (48). Accordingly, cutting efficiency at region 8 was non-zero but lower than in the euchromatic region 3 (Supplementary Figure S6). Our results are in line with the previous finding that CRISPR/Cas9 cutting efficiency is significantly reduced in heterochromatin regions enriched for the H3K9me3 mark (30).

We repeated PCR analysis of region 8 knock-in with modified PCR conditions in which we used four times more gDNA template and increased the cycle number from 30 to 40. With new PCR conditions, we were able to determine insertion into region 8 (Figure 2D). The reason why the knock-in band was visible only after increased cycle number and more gDNA template from mixed pool cells must be due to the low frequency of correctly edited cells in the mixed population. Our results demonstrate that knock-in into heterochromatic sites is possible; however, knock-in efficiency can be lower at sites enriched with heterochromatin marks.

Comparison of knock-in at two euchromatin and two heterochromatin regions: lamina-associated, heterochromatin sites prefer knock-in via NHEJ pathway

We isolated single clones after knock-in into two heterochromatin (regions 9 and 10) and two euchromatin target sites (regions 3 (*HSP70*) and 5). After clonal isolation, genotyping PCRs with Out-For and Out-Rev primers (Figure 2B) identified four clones from region 10 (10-13, 10-21, 10-22, 10-24), two clones from regions 3 (3-2, 3-9) and 5 (5-2, 5-6), and one clone from region 9 (9-44) as full length knock-ins (Figure 2E and F). We observed two double knock-in clones: 3-2 and 10–22 for regions 3 and 10, respectively, showing that labeling of both alleles of a gene is possible by SHACKTeR (Figure 2E and F). Further analysis of the full-length knock-in clones using SpeI digestion of the knock-in bands revealed all these clones were correct except for 10–13 (Supplementary Figure S7A). Analysis with both left and right junction PCRs further suggested most clones as correctly targeted and containing intact insert (Supplementary Figure S7B). However, unexpectedly we could not amplify right junction PCR band from the 9–44 clone, suggesting a local loss of donor or flanking DNA sequence.

We sequenced the gDNA-insert junctions for all eight remaining clones (3-2, 3–9, 5-2, 5–6, 9–44, 10–21, 10–22, 10–24) (Supplementary Text S1). Interestingly, although all targeted insertions into euchromatic sites occurred through HR, all four targeted integrations into the heterochromatic regions 9 and 10 occurred through NHEJ. For these heterochromatin regions, donor DNA was directly ligated after the expected cut site (Supplementary Text S1); moreover, various lengths of deletions of the linear donor DNA from their ends were observed in all clones (9-44, 10–21, 10–22, 10–24) which also explained why PCR-amplification of the right junction for clone 9–44 failed (Supplementary Text S1). We counted clones 9–44, 10–21, 10–22 and 10–24 as correctly targeted because the 48-mer TetO repeat was confirmed to be intact by sequencing. Our overall results show a pronounced bias toward NHEJ-based integrations into lamina-associated, heterochromatin sites (Supplementary Table S9).

48-mer TetO repeat allows visualization of both euchromatin and heterochromatin sites

Next, we tested whether all four of the tagged loci could be visualized by the expression of TetR-EGFP. For more efficient labeling, a lentiviral TetR-EGFP-IRES-PuroR vector was constructed. We used weak F9 promoter to reduce the background from unbound TetR-EGFP. Clones 3–9, 5-2, 9–44 and 10–21 were used for visualization. We also analyzed double knock-in clone 10–22 as an example to show that it is possible to label both alleles of a gene using SHACKTeR. Cells were transduced with the lentivirus at low multiplicity of infection (Figure 3A). After puromycin selection, nearly 100% of cells expressed TetR-EGFP. Signal-to-background intensity ratio is strongly dependent on the level of TetR-EGFP expression and when

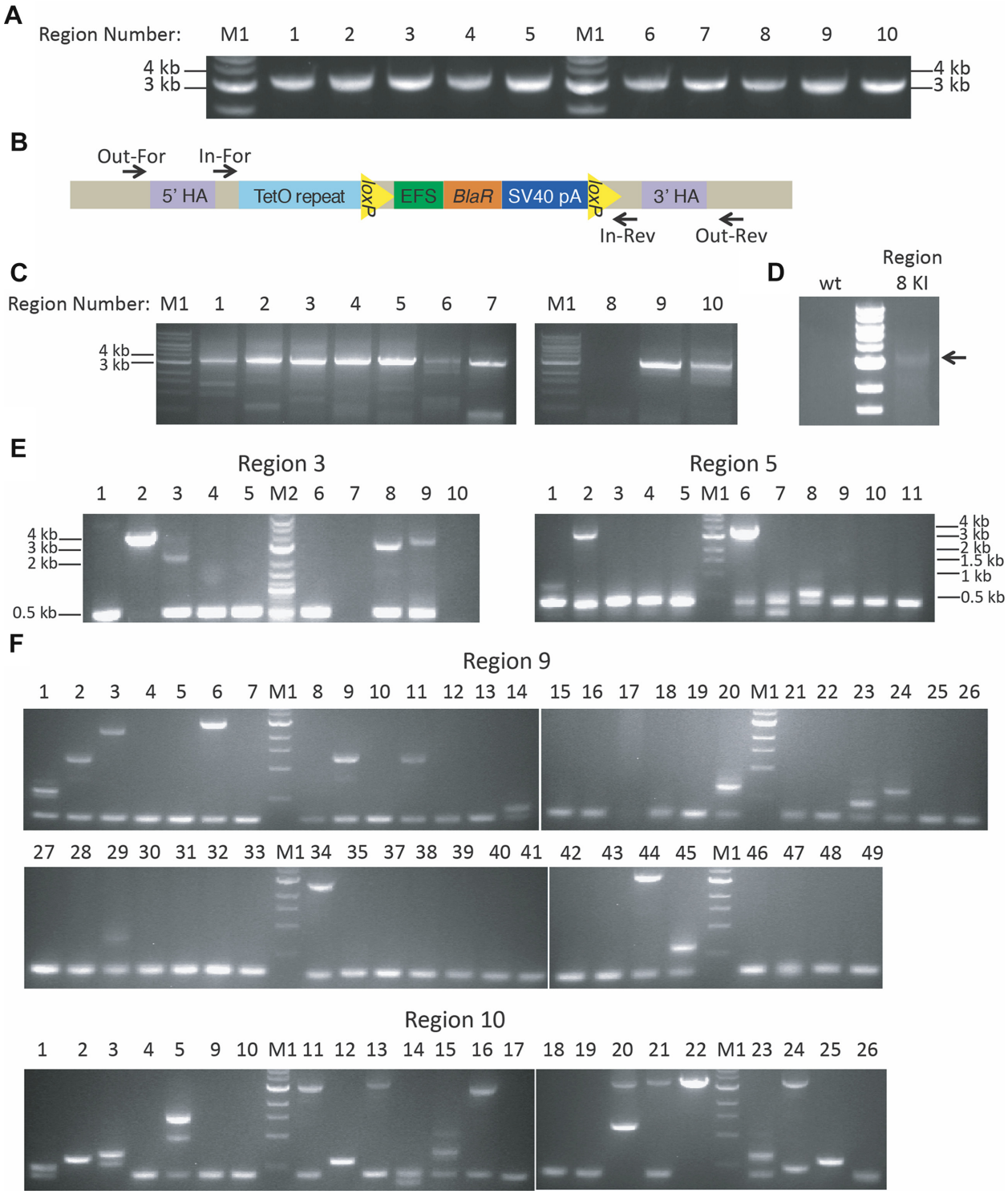

Figure 2. The 48-mer TetO knock-in into various loci. (**A**) PCR results for creating linear 48-mer TetO donor DNA for knock-in into 10 different regions. Donors are expected to be around 3.1 kb. (**B**) Schematic of the primers used for genotyping. (**C**) PCR analysis of mixed pool cells after transfection of the CRISPR/Cas9 plasmid and 48-mer TetO donors for knock-in into 10 different regions. Out-For and In-Rev primers, described in (B), were used for PCR-amplification of the knock-in band only. Knock-in band is expected to be around 3.1 kb for regions 3, 8 and 9; around 3.2 kb for regions 4, 5, 7 and 10; around 3.3 kb for region 2 and 6; and 3.6 kb for region 1. (**D**) Results of improved PCR conditions for analysis of mixed pool cells after transfection of the CRISPR/Cas9 plasmid and 48-mer TetO donor for knock-in into region 8. PCR using wt HCT116 gDNA was used as a negative control. Knock-in band is shown with an arrow. (**E**) Genotyping of the clones from 48-mer TetO knock-in into regions 3 and 5. Out-For and Out-Rev primers were used for the PCR. Expected band sizes from the knock-in alleles are around 3.5 kb for region 3 and 3.4 kb for region 5. Expected bands from the wt alleles are 528 bp for region 3 and 389 bp for region 5. (**F**) Genotyping of the clones from 48-mer TetO knock-in into regions 9 and 10. Out-For and Out-Rev primers were used for the genotyping PCR. Expected band sizes from the knock-in alleles are around 3.3 kb for region 9 and 3.4 kb for region 10. Expected bands from the wt alleles are 290 bp for region 9 and 407 bp for region 10. M1: NEB 1 kb ladder. M2: NEB 2-log ladder.

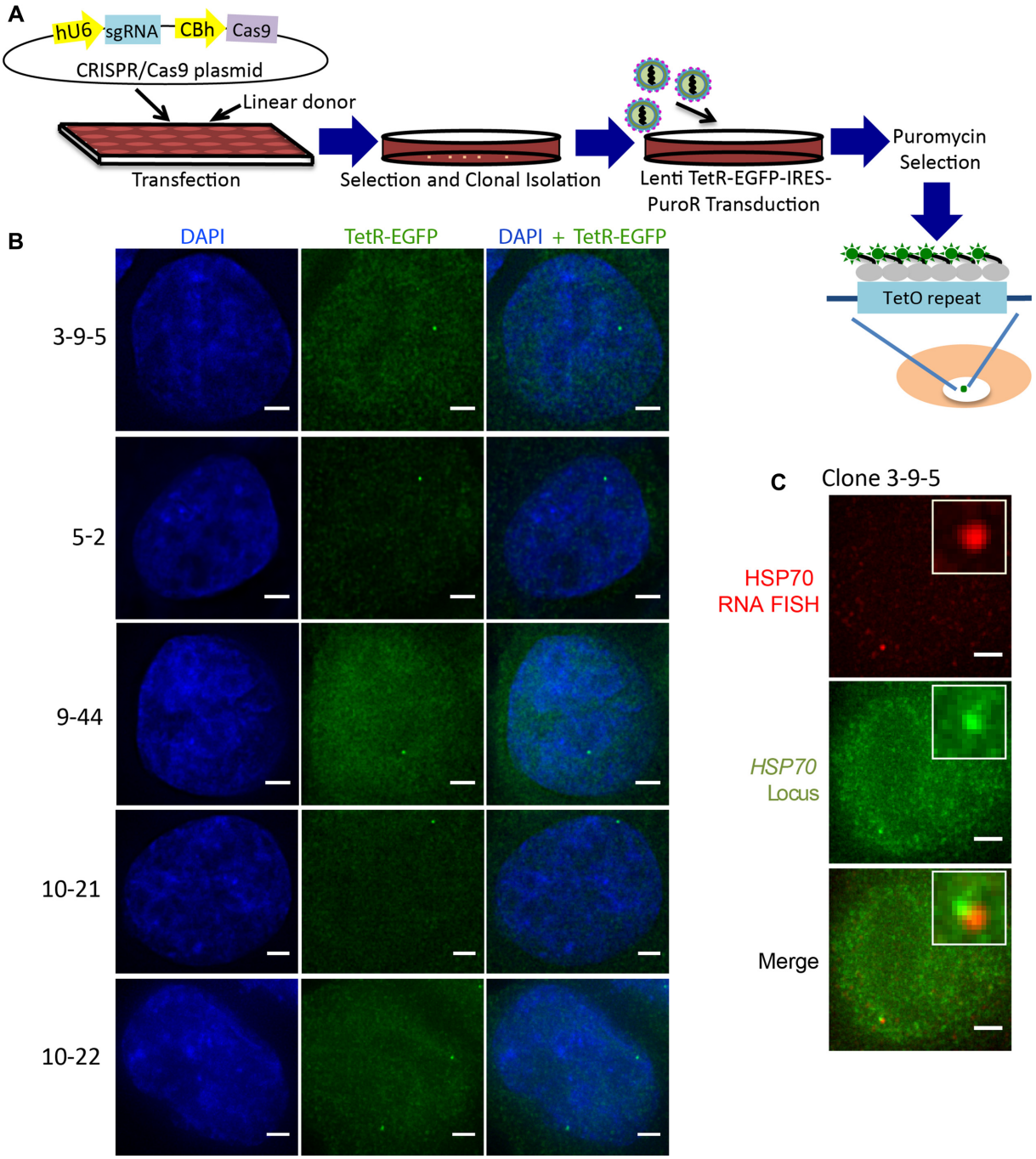

Figure 3. Visualization of 48-mer TetO-labeled endogenous loci by lentiviral expression of TetR-EGFP. (**A**) Overview of the workflow for the imaging of endogenous loci via knock-in of a 48-mer TetO repeat and lentiviral expression of TetR-EGFP. (**B**) Imaging of 48-mer TetO-labeled regions 3, 5, 9 and 10 in clones 3-9-5, 5-2, 9–44, 10–21 and 10–22, respectively, by lentiviral expression of TetR-EGFP (green). Cells were fixed and stained with DAPI (blue). Single z-sections are shown for 3-9-5, 5-2, 9–44 and 10–21. Z-projection is shown for 10–22. Scale bar: 2 μm. (**C**) Co-labeling of region 3 (*HSP70* locus) in clone 3-9-5 using TetO/TetR-EGFP system (green) and RNA FISH (red). All images are z-projections. Scale bar: 2 μm.

we analyzed non-clonal cells directly after puromycin selection, there was a variation in this ratio due to differences in TetR-EGFP expression. Still, over a range of expression levels of TetR-EGFP, we were able to observe a distinct spot for the TetO array (Figure 3B). Among the TetR-EGFP expressing cells, 40–70% of cells showed a distinct GFP spot. In previous work using the LacO/LacI system, detection sensitivity was a function of the LacR expression level, with low expression reducing the intensity over the LacO repeats but high expression creating a GFP background level which masked the LacO repeat signal (4). Subcloning TetR-EGFP stably expressing cells from the 3–9 original cell clone demonstrated that by selecting a subclone (3-9-5) with the appropriate level of TetR-EGFP ex-

pression (Figure 3B), distinct GFP spots could be visualized in nearly 100% of cells.

We used RNA FISH to confirm the specificity of the TetR-EGFP signal at the *HSP70* locus (region 3). *HSP70* genes at the *HSP70* locus are heat-inducible. After heat-shock, RNA FISH revealed an accumulation of HSP70 nascent transcripts immediately adjacent to the GFP spot tagging the *HSP70* locus (Figure 3C).

Using SR-SIM super-resolution light microscopy, we visualized GFP doublets in clones 3-9-5 and 10–21 as expected for sister chromatids following DNA replication (Figure 4A–C). The distribution of distances showed no cases in which the distance between the two spots exceeded 1 micron (Figure 4D), and no cells with more than two GFP spots were observed (Figure 4E). Clone 10–21 showed a lower percentage of cells with doublets versus a single spot as compared to the 3-9-5 clone, as predicted due to the expected late versus early DNA replication timing for heterochromatic region 10 and euchromatic region 3, respectively (49) (Figure 4E).

The successful knock-in of the TetO repeat and its visualization in multiple euchromatic and heterochromatic regions demonstrates the general applicability of our method. We observed no extra GFP spots in all 4 clones analyzed, which would have been derived from random integration of the donor DNA.

Knock-in of the 48-mer TetO repeat does not perturb the intranuclear localization of the target chromosome locus

To verify that the insertion of the TetO repeat and selection marker did not perturb the normal intranuclear targeting of the target loci, we compared intranuclear localization of the speckle-associated *HSP70* and lamina-associated *HBB* (regions 3 and 10, respectively) loci with or without the donor DNA insertion. We compared distance distribution of these chromosome loci from the nuclear lamina (*HBB*) and speckles (*HSP70*) in wt HCT116 cells versus for the knock-in alleles in clones 10–21 and 3-9-5, respectively (Figure 5A–D). The loci in wt HCT116 were labeled by DNA-FISH using the corresponding BAC probes. All the cells were co-labeled for nuclear speckles and lamina by immunofluorescence staining. No significant difference in distance distributions was observed between the loci with or without the donor DNA insertion (Figure 5C and D), indicating that insertion of the TetO repeat and selection marker did not perturb the normal intranuclear targeting of these loci. It is important to note that the signal from the nuclear speckle staining could bleed-through into FITC channel and increase the background signal after immunofluorescence staining.

Knock-in of 48-mer or 96-mer TetO repeats does not induce H3K9me3-mediated heterochromatinization flanking their integration sites

It is known that repetitive regions in the human genome can induce heterochromatin formation and these regions are mostly enriched for H3K9me3, a hallmark of constitutive heterochromatin (50,51). To test whether 48-mer and/or 96-mer TetO repeats induce H3K9me3 flanking their insertion sites, we performed a newly developed method called CUT&RUN (cleavage under targets and release using nuclease) (52,53). Similar to ChIP-Seq, CUT&RUN can identify genome-wide distribution of TFBS as well as specific histone modifications. However, CUT&RUN has significant advantages compared to ChIP-Seq including: lower background which allows for lower sequencing depth, higher signal-to noise ratio, higher resolution, fewer starting cells required and a faster and simpler protocol. The CUT&RUN protocol includes binding of the cells to concanavalin A-coated magnetic beads and permeabilization by digitonin followed by *in situ* binding of a primary antibody to the protein of interest. Following pA-MNase (protein A-micrococcal nuclease) binding to the antibody, targeted MNase digestion is started by adding Ca^{2+} at 0°C, releasing the target protein–DNA complexes into the supernatant. Chromatin fragments released by antibody targeted MNase cleavage can be purified and used for Next Generation sequencing. We performed CUT&RUN using an antibody specific for H3K9me3 to identify H3K9me3 enriched regions in the genome. H3K9me3 enrichment at region 3 was compared between clones with 96-mer or 48-mer TetO integration at region 3 (96-8 and 48-15, respectively) versus wt cells. We compared TetR-EGFP transduced 96–8 and/or 48-15 clones to both wt cells as well as TetR-EGFP transduced wt cells using a bin size of 1000 bp. We did not observe any significant difference in the H3K9me3 pattern at region 3 among the different cell lines analyzed (Figure 6A and B). Moreover, we observed that TetR-EGFP transduction of wt cells and cells carrying the TetO repeat also did not cause any significant change in the H3K9me3 pattern as compared to wt cells not expressing TetR-EGFP. Multi-sample correlation analysis showed that all the samples are highly correlated to one another (Supplementary Figure S8A and B).

Irregular 96-mer TetO repeats do not cause cell-cycle defects

A previous study proposed that TetO or LacO repeats bound by their corresponding repressors can cause replication fork barrier, accompanied by unusual, long-lived PCNA foci overlying the operator repeats, followed by mitotic defects (54). This previous study is inconsistent with other published studies, including live-cell imaging of LacO and PCNA dynamics during S-phase, which demonstrated that chromosome regions carrying large numbers of LacO arrays replicated at specific times during the cell cycle without any delay in DNA replication (55,56).

Given these conflicting studies, we directly examined segregation of mitotic chromosomes carrying the irregular 96-mer TetO repeat used in SHACKTeR. We also examined the DNA replication timing of the 96-mer TetO repeat and the spatial localization of this repeat relative to PCNA foci during DNA replication. No defects either in mitotic chromosome segregation or DNA replication timing were detected.

Specifically, we first analyzed 31 pairs of daughter cells at anaphase or telophase stages of mitosis (Figure 7A). We

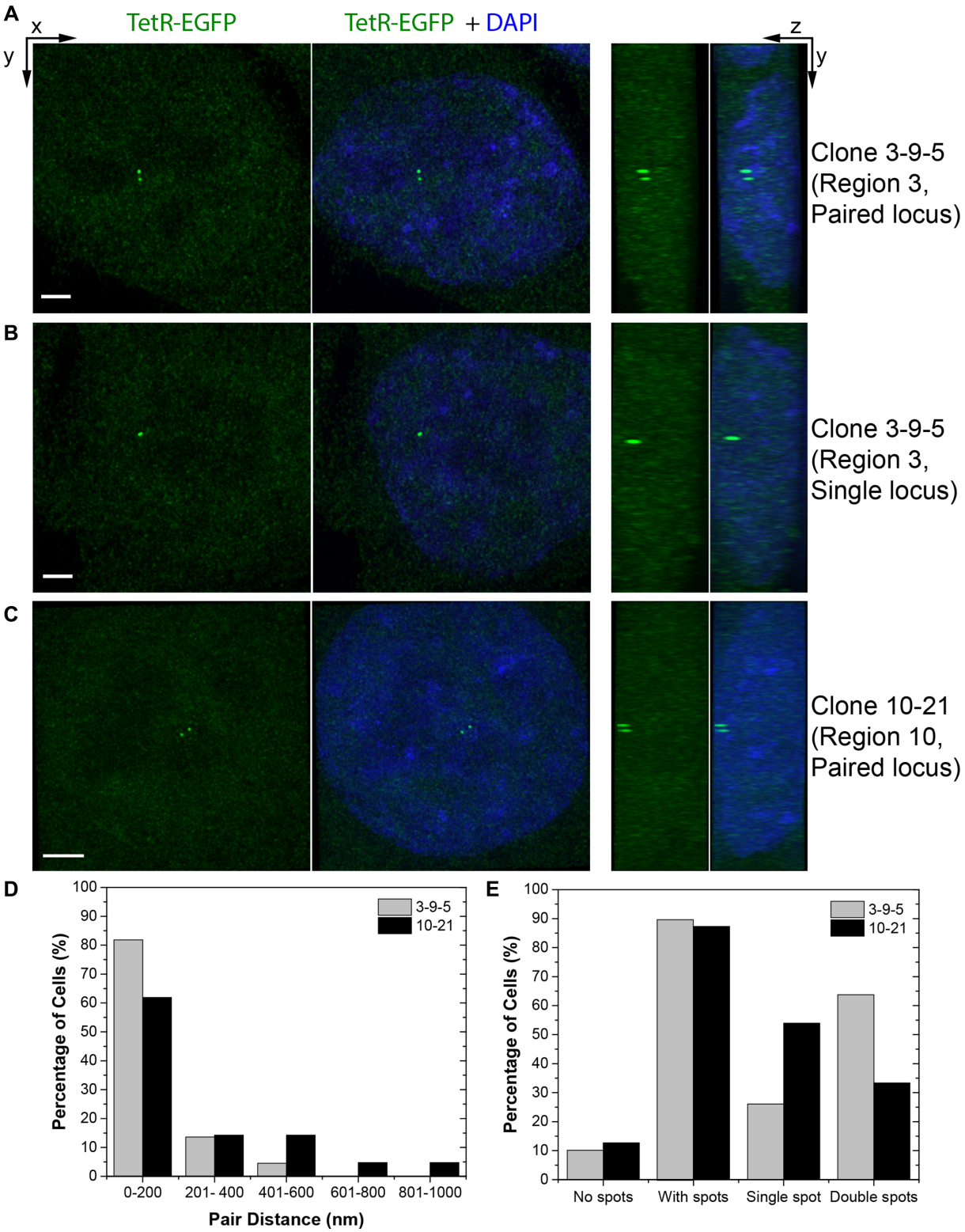

Figure 4. Analysis of the cells with paired loci. (**A**) 3D images of the TetR-EGFP transduced 3-9-5 clone in SR-SIM microscope. Images with EGFP only or EGFP+ DAPI are shown in XY (left 2 columns) and YZ (right 2 columns) projections. Here, images of a cell with paired spots (A) or a single spot (**B**) are shown. Scale bar: 0.7 μm. (**C**) 3D images of the TetR-EGFP transduced 10–21 clone in SR-SIM microscope. Images of a cell with paired spots are shown. In this picture, the two spots are at the top of the nucleus. Scale bar: 1 μm. (**D**) Distribution of the pair distance (in nm) between paired loci after DNA replication. Results for both region 3 (in clone 3-9-5, $N = 44$ cells) and region 10 (in clone 10–21, $N = 21$ cells) are shown. (**E**) Histograms of percentage of the cells with indicated number of spots observed per cell in 3-9-5 ($N = 69$ cells) and 10–21 ($N = 63$ cells) clones. Percentage of cells with spots demonstrates the percentage of the total counted cells that had either a single spot or double spots.

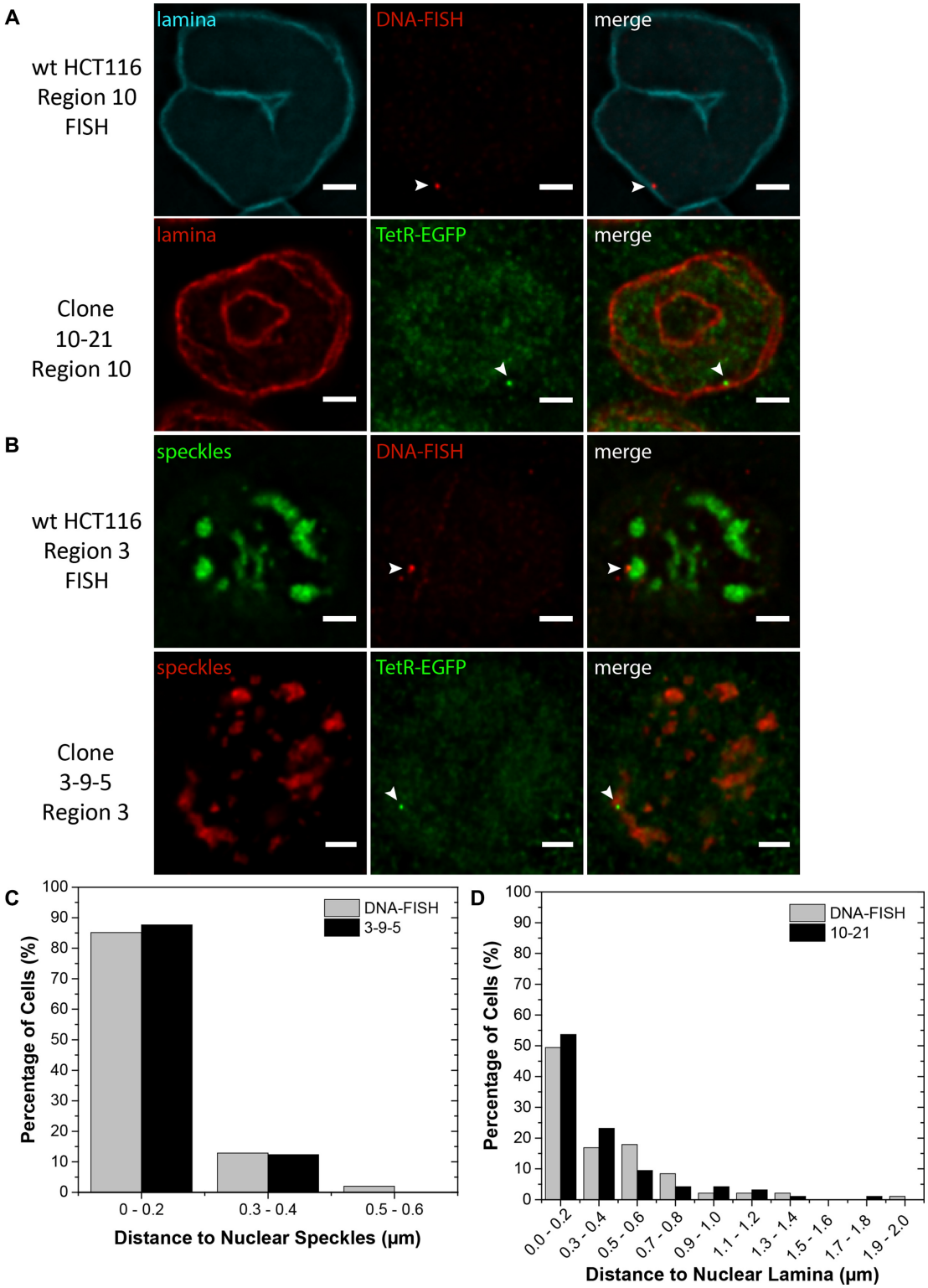

Figure 5. Controlling the effect of 48-mer TetO knock-in on the localization of endogenous loci. (**A**) Representative images of localization of the region 10 in wt HCT116 cells (top images), as determined by DNA-FISH (red), as well as in 10–21 48-mer TetO knock-in clone (bottom images), by expression of TetR-EGFP (green). Nuclear periphery was labeled with anti-laminB1/B2 antibody (shown in cyan for the top images and red for bottom images). The locus is pointed with an arrow. Single z-sections are shown. Scale bar: 2 μm. (**B**) Representative images of localization of the region 3 in wt HCT116 cells (top images), as determined by DNA-FISH (red), as well as in 3-9-5 48-mer TetO knock-in clone (bottom images), by expression of TetR-EGFP (green). Nuclear speckles were labeled with anti-SON antibody (shown in green for the top images and red for bottom images). The locus is pointed with an arrow. Single z-sections are shown. Scale bar: 2 μm. (**C**) Histogram of the distance between region 3 and the nuclear speckles in 3-9-5 knock-in clone ($N = 97$ cells) and wt HCT116 cells ($N = 101$ cells). (**D**) Histogram of the distance between region 10 and the nuclear lamina in 10–21 knock-in clone ($N = 95$ cells) and wt HCT116 cells ($N = 95$ cells).

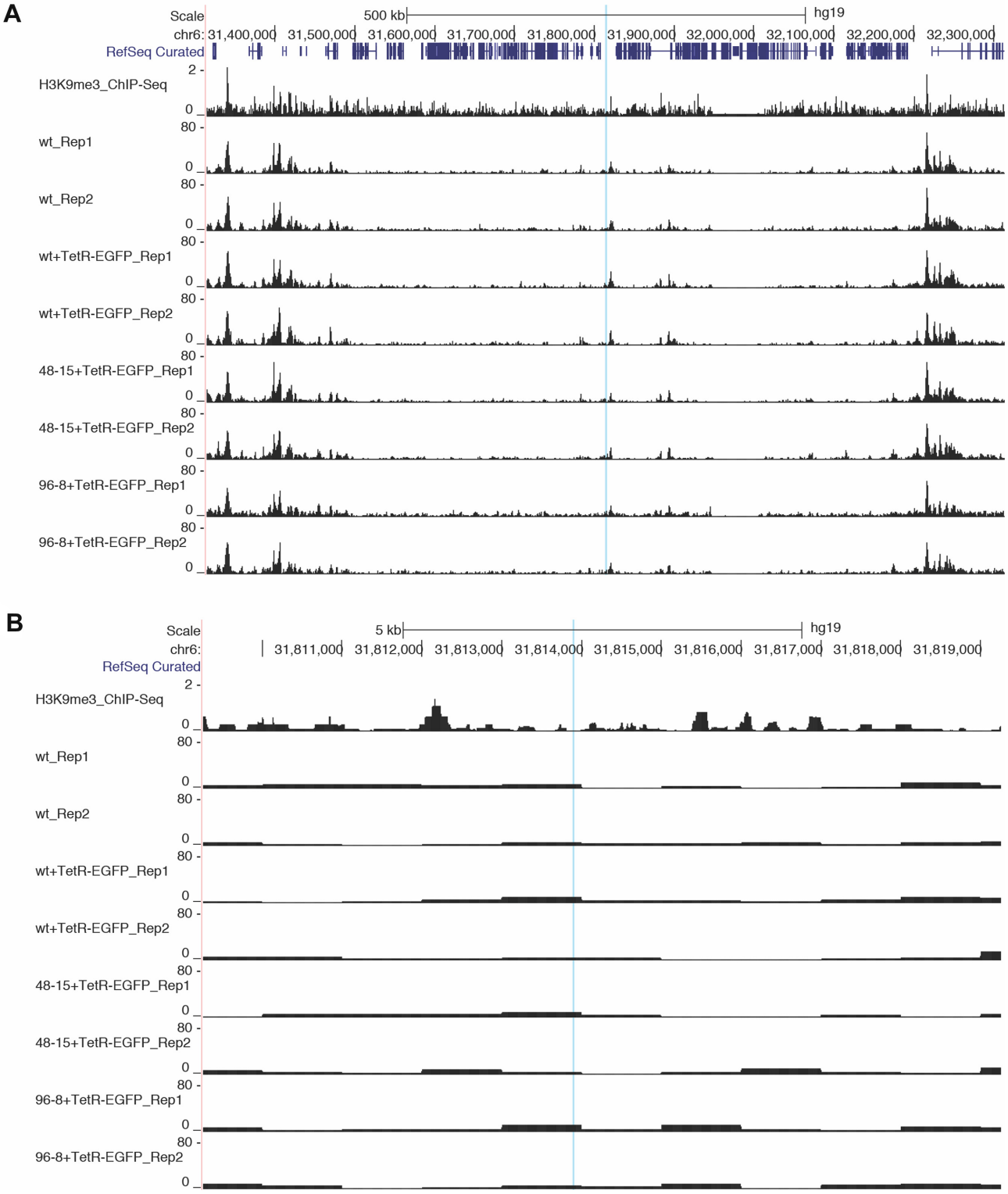

Figure 6. CUT&RUN results. UCSC Genome Browser images of the CUT&RUN signal tracks for H3K9me3 in non-transduced wt and TetR-EGFP transduced wt (wt+TetR-EGFP), 48-15 (48-15+TetR-EGFP) and 96–8 (96-8+TetR-EGFP) cells. Results are shown at (**A**) 1 Mb or (**B**) 10 kb resolution. Knock-in site is highlighted in blue. Ref-seq gene positions are demonstrated on the top. ChIP-Seq signal track for H3K9me3 in wt HCT116 cells is shown as a control. Both replicates (Rep1 and Rep2) are shown for each case.

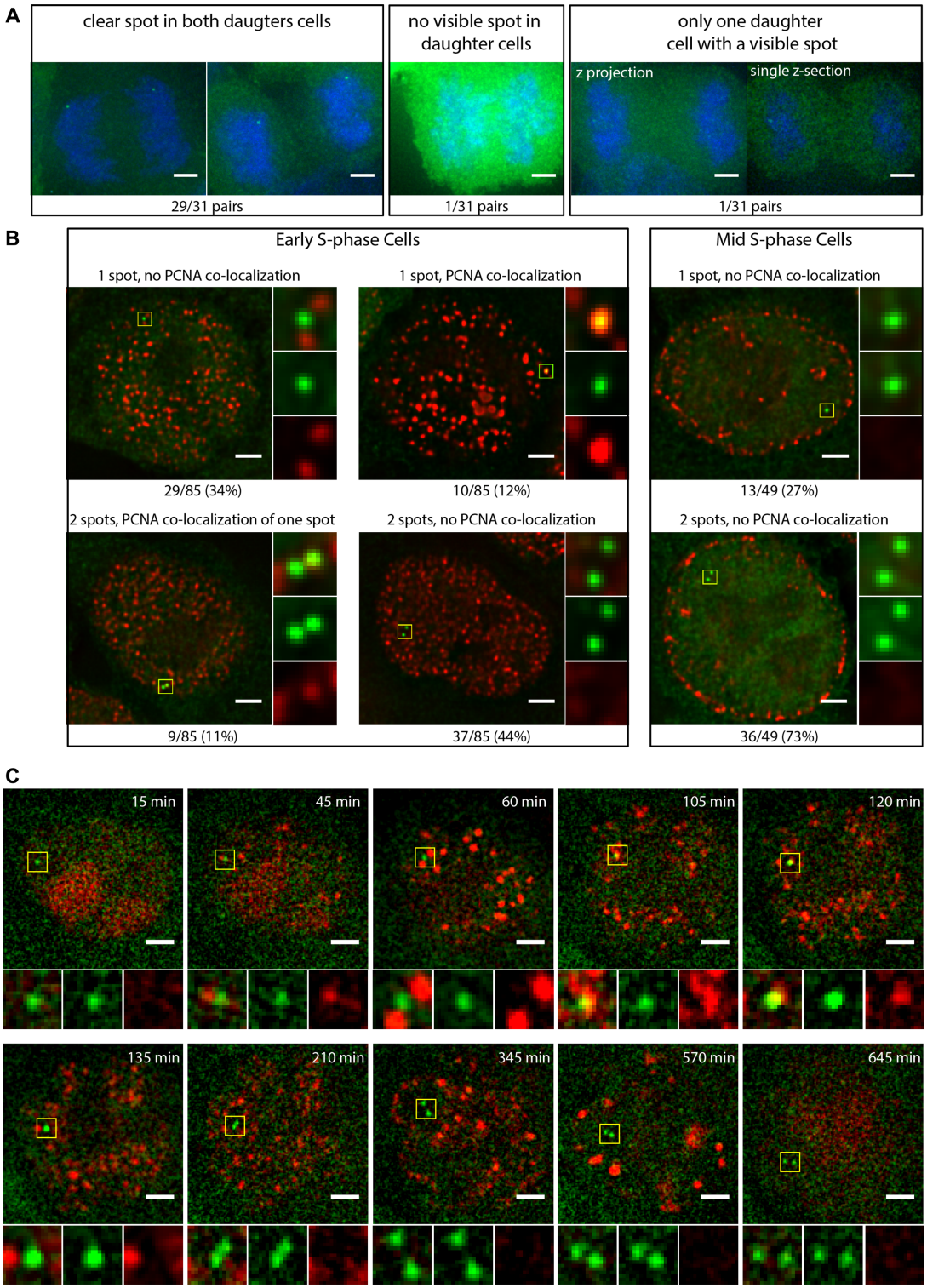

Figure 7. Analysis of cell cycle in 96-mer TetO labeled cells. (**A**) Pairs of daughter cells at anaphase and telophase were analyzed for equal and symmetric segregation of TetO repeats (green). Representative images for each observed case are shown, together with the number of analyzed pairs that showed each pattern. Thirty-one pairs of daughter cells were analyzed using TetR-EGFP transduced 96–8 clone. Cells were fixed and stained with DAPI (blue). All images are z-projections, unless indicated otherwise. Scale bar: 2 μm. (**B**) Analysis of PCNA (red)/96-mer TetO (green) co-localization at each subphase of S-phase. Cells were grouped based on PCNA co-localization at the locus and the number of spots observed. Representative images for each observed case are shown and number of cells observed with each pattern is reported below the images. A total of 85 early S-phase cells and 49 mid S-phase cells were analyzed. Single z-sections with merge signal are shown. Insets show 6× magnification of the boxed region and shows merge, green and red channels from top to bottom. Scale bar: 2 μm. (**C**) Snapshots from the live cell imaging of region 3 (green) and its interaction with PCNA (red). The movie was taken for 16 h with 15 min time-lapse. Single z-sections with merge signal are shown. Insets show 6× magnification of the boxed region and shows merge, green and red channels from left to right. Scale bar: 2 μm. Also see Supplementary Movies S1 and 2.

chose TetR-EGFP transduced 96–8 clone for analysis, since it was easier to detect 96-mer TetO in mitotic chromosomes compared to 48-mer TetO. Moreover, we reasoned that any tendency of the TetO repeat to perturb chromosome dynamics should be larger for the 96-mer versus 48-mer repeat; thus using the 96-mer repeat would provide a more sensitive assay for TetO induced changes in chromosome dynamics. We observed symmetric segregation of replicated TetO repeats between daughter cells in 29 of the pairs analyzed (Figure 7A). In the remaining two cases, we could not see a clear spot in daughter cells due to non-optimal expression level of TetR-EGFP (too high or too low).

We next analyzed 96–8 TetR-EGFP clone transduced with mTagRFP-PCNA (referred to as 96–$8^{\text{TetR-EGFP/mTagRFP-PCNA}}$) to test for delays in DNA replication timing induced by the TetO repeat. PCNA (proliferating cell nuclear antigen) is an essential protein in DNA replication and we used mTagRFP fused PCNA as a marker for DNA replication foci (57). In G1 and G2, cells have low levels of PCNA diffusely distributed throughout the nucleus, whereas in early S-phase PCNA forms many equally distributed bright foci in the nucleus, corresponding to regions of active DNA replication (57). PCNA foci are mostly seen along the nuclear periphery during mid S-phase; during late S-phase, a few large foci are seen in the middle of the nucleus (57).

The 2-fraction Repli-Seq data in HCT116 cells (58) (4D Nucleome accession number: 4DNESSYKYIK3, 4DNESLC8TDK4; data also available at www.replicationdomain.org) maps region 3 to an early replicating domain**.** We analyzed randomly selected early and mid S-phase cells in an unsynchronized population of 96–$8^{\text{TetR-EGFP/mTagRFP-PCNA}}$ cells and identified different groups of cells based on PCNA/TetO repeat co-localization and presence of single or double spots (Figure 7B). Thirty-four percent (29/85) of early S-phase cells showed one TetO spot without PCNA co-localization, which could either correspond to an unreplicated locus or pairing of two spots on sister chromatids, where the distance between the paired sister chromatids is below the x-y and/or z-resolution of the light microscope. Forty-four percent (37/85) of early S-phase cells showed two spots without PCNA co-localization, corresponding to separation of sister chromatids after replication. Twelve percent (10/85) of early S-phase cells showed one TetO spot with PCNA co-localization, consistent with this TetO-tagged locus undergoing active DNA replication. At last, 11% (9/85) of the early S-phase cells had two spots with one spot showing PCNA co-localization, which could correspond to loci immediately after replication (59). Zero percent (0/49) of mid S-phase cells showed PCNA co-localization, consistent with early replication of the TetO-labeled region 3 with no prolonged DNA replication block or delay. Seventy-three percent (36/45) of the mid S-phase cells showed two TetO spots with the remaining 27% (13/45) showing one spot (Figure 7B). The relatively high percentage of single TetO spots after the inferred period of DNA replication suggests frequent close proximity of replicated sister chromatids as measured by light microscopy. The similar fraction of single spots in cells with early S-phase versus mid S-phase PCNA foci patterns (34 versus 27%) further supports very early DNA replication of the TetO repeat.

Long-term imaging of DNA replication using SHACKTeR

We used live-cell imaging to directly observe the early DNA replication of the 96-mer TetO-tagged *HSP70* locus (region 3) and normal PCNA foci dynamics during this replication predicted by our fixed cell analysis. We performed live cell imaging of 96–$8^{\text{TetR-EGFP/mTagRFP-PCNA}}$ cells (Supplementary Movies S1 and 2; Figure 7C), selecting cells with a non-S phase PCNA pattern and then imaged them for 16 h, taking 3D z-stacks every 15 min. In the example shown in Supplementary Movie S1 (see Supplementary Movie S2 for z-projection), the TetO repeat-labeled locus was near PCNA foci but is not co-localized for the first ∼1 h of S-phase. Approximately 1 h after S-phase starts, the locus co-localizes with PCNA. After ∼30 min (two time points), PCNA foci disappeared at the locus indicating completion of DNA replication. After this brief period of PCNA co-localization, we saw frequent transitions between one brighter TetO spot and two dimmer TetO spots, consistent with sister chromatids tethered through sister chromatid cohesion. After this brief period of PCNA co-localization, the TetO spots showed no new period of PCNA co-localization through the remainder of S-phase.

These live-cell observations demonstrate DNA replication of the TetO-tagged *HSP70* locus between ∼1 and 2 h after the initiation of S-phase with normal dynamics between the TetO-array and PCNA foci ($N = 10$ cells), consistent with our fixed cell analysis. We therefore conclude that the TetO array with bound TetR-EGFP does not cause changes to DNA replication timing. Our results also demonstrate the feasibility of long-term imaging using SHACKTeR.

Live cell imaging of the TetO-tagged loci enables capturing the dynamics of loci with their associated nuclear compartments

At last, we performed live cell imaging of dynamics of the euchromatic and heterochromatic *HSP70* and β-globin loci (clones 3-9-5 and 10–21, respectively) with their associated nuclear compartments. First, we tested the stability and dynamics of the *HSP70* locus association with nuclear speckles. Measurement of distances to nuclear speckles in fixed cells (Figure 5C) had shown over 80% of loci within 0.2 microns of nuclear speckles, with nearly all remaining loci <0.5 microns from the speckles. Live cell imaging at 1 min intervals revealed that individual chromosome loci would occasionally move several tenths of a micron away from the nuclear speckle and back over short time intervals (Figure 8A and C; Supplementary Movie S3). Thus the distance distribution observed in fixed cells was derived partially from small fluctuations in speckle distance over time at individual loci in live cells; however, overall speckle association remained stable over the 28 min observation periods. Similarly, live-cell imaging revealed oscillations of several tenths of microns of the *HBB* locus toward and away from the nuclear lamina over several min time intervals (Figure 8B and D; Supplementary Movie S4), but with the locus remaining close to the nuclear lamina over the observation period.

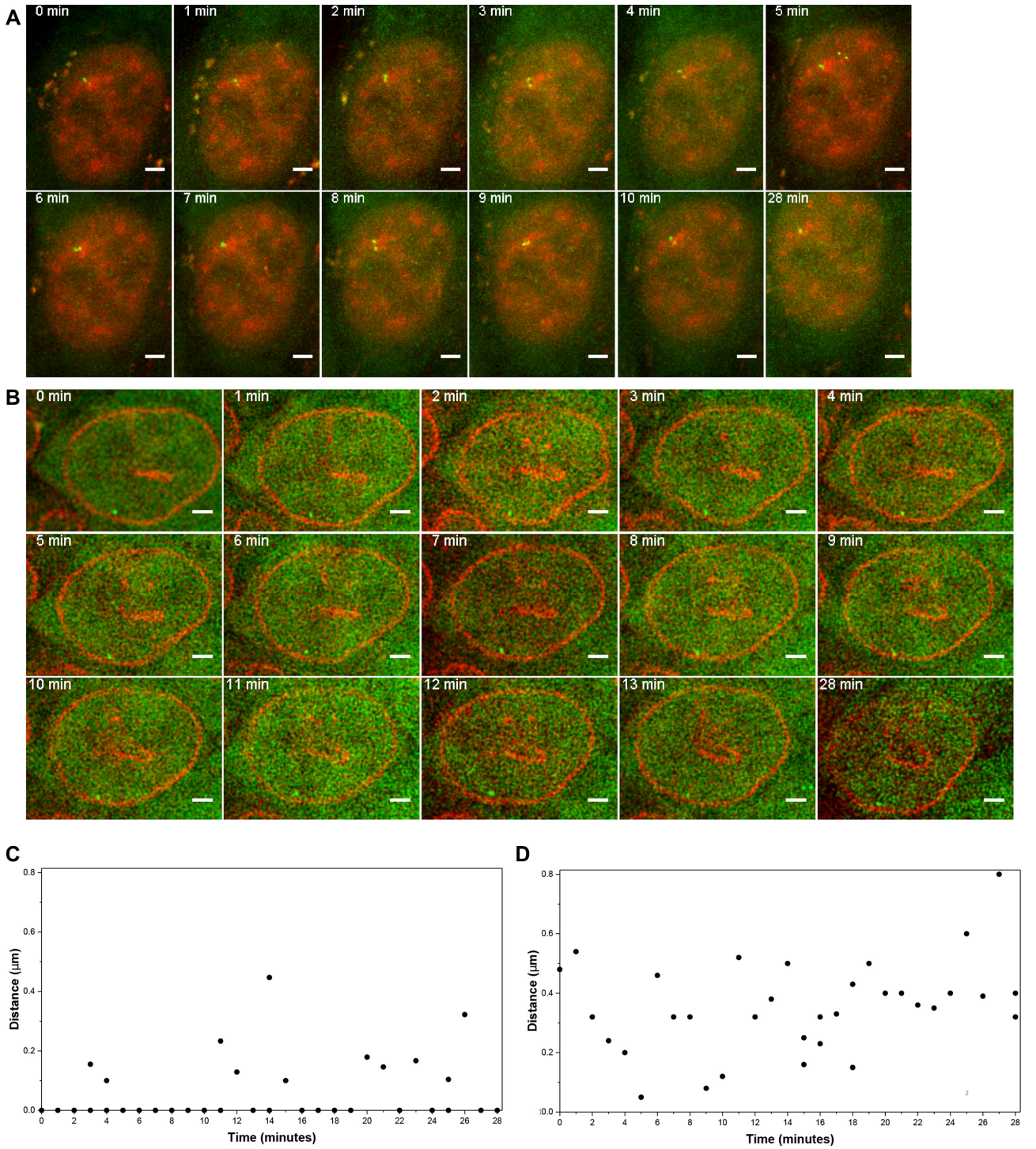

Figure 8. Dynamics of the nuclear speckle or lamina interactions. (**A**) Snapshots from the live cell imaging of region 3 (*HSP70* locus, shown in green) and its interaction with nuclear speckles (red). The movie was taken for 28 min with 1 min time-lapse. Images are from the 3-9-5 clone, transfected with Magoh-mCherry and transduced with TetR-EGFP. All images are z-projections. Scale bar: 2 μm. Also see Supplementary Movie S3. (**B**) Snapshots from the live cell imaging of region 10 (*HBB* locus, shown in green) and its interaction with nuclear lamina (red). The movie was taken for 28 min with 1 min time-lapse. Images are from the 10–21 clone, transfected with mCherry-LaminB1 and transduced with TetR-EGFP. Single z-sections are shown. Scale bar: 2 μm. Also see Supplementary Movie S4. (**C**) Histogram of the distance between region 3 and the nuclear speckles in 3-9-5 knock-in clone. (**D**) Histogram of the distance between region 10 and the nuclear lamina in 10–21 knock-in clone. For (C) and (D), when two spots were observed, both of them were plotted.

A preliminary mean square displacement analysis of clone 96–8 transduced with TetR-EGFP showed an estimated diffusion constant of 1.35×10^{-4} $\mu m^2/s$, which falls within the range of values observed previously in mammalian cells. Consistent with early results from Chubb *et al.* (60) who saw a reduced radius of constraint for chromosome loci tethered to a nuclear compartment (nucleolus), we saw a similarly low radius of constraint (~0.3 μm) consistent with the tethering of the *HSP70* locus to nuclear speckles.

DISCUSSION

We demonstrated high efficiency knock-in of an optimized 48-mer TetO repeat for labeling and live cell tracking of various endogenous loci, including heterochromatic sites. Our tagging method did not perturb the normal intranuclear localization of the *HSP70* and β-globin chromosome loci. Moreover, TetO repeats did not cause H3K9me3-mediated heterochromatinization flanking their integration sites. We showed the usefulness of SHACKTeR for cytological analysis in both fixed and live cells. 48-mer TetO repeat was sufficient to observe specific loci for fixed cell analysis or short-term movies. Long-term live cell imaging was feasible when 96-mer TetO was used for labeling. In the future, signal-to-noise ratio of 48-mer TetO could be improved by using split-GFP (61) or SunTag approaches (62) to allow longer-term imaging.

Random integration of exogenous DNA is a concern in knock-in strategies; however, our imaging of the targeted clones revealed only the number of GFP-tagged loci predicted by our genotyping and FISH analysis. The low off-target integration may be due to our exclusion of repetitive and low-complexity sequences from the short HAs and/or use of very low amount of donor DNA. Similar targeting conditions might be beneficial for other applications such as gene correction in various genetic diseases.

Our strategy to visualize non-repetitive regions requires minimal modification of the genome: integration of the 48-mer TetO repeat and the TetR-EGFP expression cassette. This is an advantage over previous CRISPR/Cas9-based imaging techniques that required integration of at least 26 sgRNAs, each from a separate lentivirus, per cell (7). For successful labeling, all the sgRNAs used need to have an optimum binding to their target sequence and binding efficiency can change from sgRNA to sgRNA. Thus, the number of required sgRNAs can be different for each case, depending on how many of the sgRNAs within the pool are functional. The same would be true for TALE labeling, although TALE imaging of non-repetitive loci has not been demonstrated yet. Thus, labeling of genes based on the known binding of TetR-TetO can be more robust for non-repetitive sites.

While our manuscript was in preparation, an improved CRISPR/Cas9-based method for visualizing non-repetitive regions was published (63). In this work, 3′ end of sgRNAs were tagged with multiple MS2 repeats and visualized by expression of MCP-FP fusion proteins. Although signal-to-noise ratio increased significantly, the system requires lentiviral delivery of a minimum of eight different sgRNAs for visualization using conventional fluorescence microscopy. The labeling efficiency of this technique was lower than what was observed with the dCas9-FP labeling, which could be due to the fact that adding MS2 repeats at 3′ end of sgRNAs can destabilize the sgRNA (64) and such long repeats can also be unstable in lentiviruses. Moreover, co-delivery of multiple plasmids can be challenging, which can decrease labeling efficiency. In terms of the protocol and number of cloning steps required, dCas9-based imaging and SHACKTeR are similar. For imaging using CRISPR/dCas9, dCas9-FP or the MCP-FP protein used by Qin *et al.* need to be stably expressed at an optimum level, requiring clonal isolation, which is comparable to stable TetR-EGFP expression in SHACKTeR. In order to get high labeling efficiency, ideally sgRNAs should be clonally expressed as well, requiring another clonal isolation step, which is comparable to the TetO repeat knock-in step in SHACKTeR. Indeed, in SHACKTeR strategy, when we analyzed mixed population cells with varying TetR-EGFP levels, we still could observe a clear spot in most of the cells analyzed, suggesting that clonal isolation of stable TetR-EGFP expressing cells is not necessary. We believe our method provides a complementary and orthogonal labeling approach that can be combined with dCas9 labeling for visualizing non-repetitive sequences. Although knock-in of repeats was not preferred previously due to the difficulty of the protocol, now our approach makes it a viable option. Thus, our method can also change the direction of future optimization efforts for imaging unique, non-repetitive endogenous loci.

We anticipate that using different repetitive sequences—for instance lacO repeats or repeats corresponding to binding targets of single TALE proteins or sgRNAs—will allow extension of our methodology to the tagging of multiple endogenous loci. What remain to be determined are the relative advantages and disadvantages of each of these different chromosome tagging systems. For instance, although CRISPR/dCas9 is being widely adopted for live-cell imaging, we still do not know whether the R-loops created by the CRISPR/dCas9 system will perturb cell physiology, as do naturally occurring R-loops (65). Moreover, a detailed analysis of TALE and CRISPR/dCas9 imaging in terms of cell cycle defects is necessary. The number of binding sites on the DNA for the proteins used for imaging as well as dynamics of these proteins at the target DNA site can affect the extent of deficiency caused at the target locus. Future comparisons of the same chromosome locus tagged with different systems (TALEs, CRISPR/dCas9 or TetO repeats) should prove useful in addressing these questions.

A disadvantage of our method was the low frequency of clones with biallelic knock-in of TetO repeats, which could be due to the low amount of donor DNA we used compared to what is typically used. In the future, ideal donor DNA amount can be optimized to achieve maximum biallelic labeling without compromising specificity. Novel strategies to increase CRISPR-mediated knock-in efficiency, such as use of Cas9/gRNA ribonucleoprotein complexes (66), novel Cas9 fusion proteins (67), inhibition of competing NHEJ pathway (68,69) or stimulating HR (70), could also be helpful. It is also possible to prepare two donor DNAs with two

different selectable markers, which previously was shown to increase biallelic knock-in efficiency (71).

Our results are also informative for understanding the relationship between chromatin state/nuclear position and the choice of DSB repair mechanism used at different chromosome loci. A previous publication suggested that heterochromatin and transcriptionally silent sites repair DSBs by NHEJ (33) and our results support this finding. It is likely that knock-in into heterochromatin, lamina-associated sites was possible due to our use of linear DNA that could directly ligate to DSB DNA ends by NHEJ.

At last, it is important to note that region 8, the only target region for which we saw a very faint knock-in band only after optimized PCR analysis, showed a noticeable enrichment of the H3K9me3 heterochromatin mark. Further analyses would clarify better which histone marks and subnuclear compartmentalization affect knock-in efficiency and/or DSB repair pathway choice, leading to improved rules for designing knock-in experiments.

DATA AVAILABILITY

CUT&RUN data from this study have been submitted to the Gene Expression Omnibus (GEO) under the accession number GSE113832.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Steven Henikoff, Jorja Henikoff and Hatice S. Kaya-Okur (Fred Hutchinson Cancer Research Center) for helpful discussions about CUT&RUN method. We thank Ayca Z. Ilter-Akulke (University of Illinois at Urbana-Champaign (UIUC)) for helping with optimization of the CUT&RUN experiments. We also thank Alvaro G. Hernandez and Chris L. Wright (High-Throughput Sequencing and Genotyping Unit, Roy J. Carver Biotechnology Center, UIUC) for assistance with next generation sequencing; Christopher J. Fields and Gloria Rendon (High Performance Biological Computing, Roy J. Carver Biotechnology Center, UIUC) for assistance with CUT&RUN data processing and analysis; and Barbara Pilas and Barbara Balhan (Flow Cytometry Facility, Roy J. Carver Biotechnology Center, UIUC) for FACS. Light microscopy was performed in the Department of Molecular and Cellular Biology Light Microscopy Facility (UIUC). SR-SIM microscopy was performed at the Core Facilities at the Carl R. Woese Institute for Genomic Biology. We thank David M. Gilbert (Florida State University) for helpful discussions about cell-cycle analyses. We also acknowledge Gilbert lab for creating HCT116 Repli-Seq data as part of the 4D Nucleome Project.

FUNDING

National Institutes of Health (NIH) [1U54DK107965 to H.Z., J.M. A.S.B., R01GM058460 to A.S.B]. Funding for open access charge: NIH [1U54DK107965].

Conflict of interest statement. None declared.

REFERENCES

1. Dixon,J.R., Selvaraj,S., Yue,F., Kim,A., Li,Y., Shen,Y., Hu,M., Liu,J.S. and Ren,B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
2. Misteli,T. (2007) Beyond the sequence: cellular organization of genome function. *Cell*, **128**, 787–800.
3. Belmont,A.S. (2001) Visualizing chromosome dynamics with GFP. *Trends Cell Biol.*, **11**, 250–257.
4. Robinett,C.C., Straight,A., Li,G., Willhelm,C., Sudlow,G., Murray,A. and Belmont,A.S. (1996) In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J. Cell Biol.*, **135**, 1685–1700.
5. Masui,O., Bonnet,I., Le Baccon,P., Brito,I., Pollex,T., Murphy,N., Hupé,P., Barillot,E., Belmont,A.S. and Heard,E. (2011) Live-cell chromosome dynamics and outcome of X chromosome pairing events during ES cell differentiation. *Cell*, **145**, 447–458.
6. Ran,F.A., Hsu,P.D., Wright,J., Agarwala,V., Scott,D.A. and Zhang,F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.
7. Chen,B., Gilbert,L.A., Cimini,B.A., Schnitzbauer,J., Zhang,W., Li,G.W., Park,J., Blackburn,E.H., Weissman,J.S., Qi,L.S. *et al.* (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, **155**, 1479–1491.
8. Ma,H., Reyes-Gutierrez,P. and Pederson,T. (2013) Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 21048–21053.
9. Miyanari,Y., Ziegler-Birling,C. and Torres-Padilla,M.E. (2013) Live visualization of chromatin dynamics with fluorescent TALEs. *Nat. Struct. Mol. Biol.*, **20**, 1321–1324.
10. Thanisch,K., Schneider,K., Morbitzer,R., Solovei,I., Lahaye,T., Bultmann,S. and Leonhardt,H. (2014) Targeting and tracing of specific DNA sequences with dTALEs in living cells. *Nucleic Acids Res.*, **42**, e38.
11. Johnson,R.D. and Jasin,M. (2001) Double-strand-break-induced homologous recombination in mammalian cells. *Biochem. Soc. Trans.*, **29**, 196–201.
12. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler,D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
13. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
14. Sloan,C.A., Chan,E.T., Davidson,J.M., Malladi,V.S., Strattan,J.S., Hitz,B.C., Gabdank,I., Narayanan,A.K., Ho,M., Lee,B.T. *et al.* (2016) ENCODE data at the ENCODE portal. *Nucleic Acids Res.*, **44**, D726–D732.
15. Rosenbloom,K.R., Sloan,C.A., Malladi,V.S., Dreszer,T.R., Learned,K., Kirkup,V.M., Wong,M.C., Maddren,M., Fang,R., Heitner,S.G. *et al.* (2013) ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.*, **41**, D56–D63.
16. Smit,A.F.A., Hubley,R. and Green,P. (1996–2010) RepeatMasker Open-3.0.
17. Naito,Y., Hino,K., Bono,H. and Ui-Tei,K. (2015) CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics*, **31**, 1120–1123.
18. Wang,T., Wei,J.J., Sabatini,D.M. and Lander,E.S. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, **343**, 80–84.
19. Orlando,S.J., Santiago,Y., DeKelver,R.C., Freyvert,Y., Boydston,E.A., Moehle,E.A., Choi,V.M., Gopalan,S.M., Lou,J.F., Li,J. *et al.* (2010) Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Res.*, **38**, e152.
20. Strukov,Y.G. and Belmont,A.S. (2008) Development of mammalian cell lines with lac operator-tagged chromosomes. *CSH Protoc.*,**2008**, doi:10.1101/pdb.prot4903.
21. Agard,D.A., Hiraoka,Y., Shaw,P. and Sedat,J.W. (1989) Fluorescence microscopy in three dimensions. *Methods Cell Biol.*, **30**, 353–377.

22. Sivaguru,M., Urban,M.A., Fried,G., Wesseln,C.J., Mander,L. and Punyasena,S.W. (2018) Comparative performance of airyscan and structured illumination superresolution microscopy in the study of the surface texture and 3D shape of pollen. *Microsc. Res. Tech.*, **81**, 101–114.
23. Thévenaz,P., Ruttimann,U.E. and Unser,M. (1998) A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process*, **7**, 27–41.
24. Hommelsheim,C.M., Frantzeskakis,L., Huang,M. and Ülker,B. (2014) PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. *Sci. Rep.*, **4**, 5052.
25. Lau,I.F., Filipe,S.R., Søballe,B., Økstad,O.A., Barre,F.X. and Sherratt,D.J. (2003) Spatial and temporal organization of replicating *Escherichia coli* chromosomes. *Mol. Microbiol.*, **49**, 731–743.
26. Normanno,D., Boudarène,L., Dugast-Darzacq,C., Chen,J., Richter,C., Proux,F., Bénichou,O., Voituriez,R., Darzacq,X. and Dahan,M. (2015) Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nat. Commun.*, **6**, 7357.
27. Haigis,K.M., Caya,J.G., Reichelderfer,M. and Dove,W.F. (2002) Intestinal adenomas can develop with a stable karyotype and stable microsatellites. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 8927–8931.
28. Mao,Z., Bozzella,M., Seluanov,A. and Gorbunova,V. (2008) Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst)*, **7**, 1765–1771.
29. Mao,Z., Bozzella,M., Seluanov,A. and Gorbunova,V. (2008) DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle*, **7**, 2902–2906.
30. Chen,X., Rinsma,M., Janssen,J.M., Liu,J., Maggio,I. and Gonçalves,M.A. (2016) Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Res.*, **44**, 6482–6492.
31. Horlbeck,M.A., Witkowsky,L.B., Guglielmi,B., Replogle,J.M., Gilbert,L.A., Villalta,J.E., Torigoe,S.E., Tjian,R. and Weissman,J.S. (2016) Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *Elife*, **5**, e12677.
32. Isaac,R.S., Jiang,F., Doudna,J.A., Lim,W.A., Narlikar,G.J. and Almeida,R. (2016) Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *Elife*, **5**, e13450.
33. Aymard,F., Bugler,B., Schmidt,C.K., Guillou,E., Caron,P., Briois,S., Iacovoni,J.S., Daburon,V., Miller,K.M., Jackson,S.P. *et al.* (2014) Transcriptionally active chromatin recruits homologous recombination at DNA double strand breaks. *Nat. Struct. Mol. Biol.*, **21**, 366–374.
34. Lemaître,C., Grabarz,A., Tsouroula,K., Andronov,L., Furst,A., Pankotai,T., Heyer,V., Rogier,M., Attwood,K.M., Kessler,P. *et al.* (2014) Nuclear position dictates DNA repair pathway choice. *Genes Dev.*, **28**, 2450–2463.
35. Murphy,M.E. (2013) The HSP70 family and cancer. *Carcinogenesis*, **34**, 1181–1188.
36. Hall,L.L., Smith,K.P., Byron,M. and Lawrence,J.B. (2006) Molecular anatomy of a speckle. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.*, **288**, 664–675.
37. Shopland,L.S., Johnson,C.V., Byron,M., McNeil,J. and Lawrence,J.B. (2003) Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains. *J. Cell Biol.*, **162**, 981–990.
38. Spector,D.L. and Lamond,A.I. (2011) Nuclear speckles. *Cold Spring Harb. Perspect. Biol.*, **3**, a000646.
39. Jolly,C., Vourc'h,C., Robert-Nicoud,M. and Morimoto,R.I. (1999) Intron-independent association of splicing factors with active genes. *J. Cell Biol.*, **145**, 1133–1143.
40. Khanna,N., Hu,Y. and Belmont,A.S. (2014) Hsp70 transgene directed motion to nuclear speckles facilitates heat shock activation. *Curr. Biol.*, **24**, 1138–1144.
41. Guelen,L., Pagie,L., Brasset,E., Meuleman,W., Faza,M.B., Talhout,W., Eussen,B.H., de Klein,A., Wessels,L., de Laat,W. *et al.* (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, **453**, 948–951.
42. Kind,J. and van Steensel,B. (2010) Genome–nuclear lamina interactions and gene regulation. *Curr. Opin. Cell Biol.*, **22**, 320–325.
43. van Steensel,B. and Belmont,A.S. (2017) Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell*, **169**, 780–791.
44. Hepperger,C., Mannes,A., Merz,J., Peters,J. and Dietzel,S. (2008) Three-dimensional positioning of genes in mouse cell nuclei. *Chromosoma*, **117**, 535–551.
45. Ragoczy,T., Bender,M.A., Telling,A., Byron,R. and Groudine,M. (2006) The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes Dev.*, **20**, 1447–1457.
46. Ernst,J., Kheradpour,P., Mikkelsen,T.S., Shoresh,N., Ward,L.D., Epstein,C.B., Zhang,X., Wang,L., Issner,R., Coyne,M. *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, **473**, 43–49.
47. Pombo,A. and Dillon,N. (2015) Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol.*, **16**, 245–257.
48. Brinkman,E.K., Chen,T., Amendola,M. and van Steensel,B. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.*, **42**, e168.
49. Gilbert,D.M. (2002) Replication timing and transcriptional control: beyond cause and effect. *Curr. Opin. Cell Biol.*, **14**, 377–383.
50. Saksouk,N., Simboeck,E. and Déjardin,J. (2015) Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin*, **8**, 3.
51. Nishibuchi,G. and Déjardin,J. (2017) The molecular basis of the organization of repetitive DNA-containing constitutive heterochromatin in mammals. *Chromosome Res.*, **25**, 77–87.
52. Skene,P.J. and Henikoff,S. (2017) An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife*, **6**, e21856.
53. Skene,P.J., Henikoff,J.G. and Henikoff,S. (2018) Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc*, **13**, 1006–1019.
54. Beuzer,P., Quivy,J.P. and Almouzni,G. (2014) Establishment of a replication fork barrier following induction of DNA binding in mammalian cells. *Cell Cycle*, **13**, 1607–1616.
55. Li,G., Sudlow,G. and Belmont,A.S. (1998) Interphase cell cycle dynamics of a late-replicating, heterochromatic homogeneously staining region: precise choreography of condensation/decondensation and nuclear positioning. *J. Cell Biol.*, **140**, 975–989.
56. Deng,X., Zhironkina,O.A., Cherepanynets,V.D., Strelkova,O.S., Kireev,I.I. and Belmont,A.S. (2016) Cytology of DNA replication reveals dynamic plasticity of large-scale chromatin fibers. *Curr. Biol.*, **26**, 2527–2534.
57. Leonhardt,H., Rahn,H.P., Weinzierl,P., Sporbert,A., Cremer,T., Zink,D. and Cardoso,M.C. (2000) Dynamics of DNA replication factories in living cells. *J. Cell Biol.*, **149**, 271–280.
58. Dekker,J., Belmont,A.S., Guttman,M., Leshyk,V.O., Lis,J.T., Lomvardas,S., Mirny,L.A., O'Shea,C.C., Park,P.J., Ren,B. *et al.* (2017) The 4D nucleome project. *Nature*, **549**, 219–226.
59. Shibahara,K. and Stillman,B. (1999) Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. *Cell*, **96**, 575–585.
60. Chubb,J.R., Boyle,S., Perry,P. and Bickmore,W.A. (2002) Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol.*, **12**, 439–445.
61. Cabantous,S., Terwilliger,T.C. and Waldo,G.S. (2005) Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol.*, **23**, 102–107.
62. Tanenbaum,M.E., Gilbert,L.A., Qi,L.S., Weissman,J.S. and Vale,R.D. (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell*, **159**, 635–646.
63. Qin,P., Parlak,M., Kuscu,C., Bandaria,J., Mir,M., Szlachta,K., Singh,R., Darzacq,X., Yildiz,A. and Adli,M. (2017) Live cell imaging of low- and non-repetitive chromosome loci using CRISPR-Cas9. *Nat. Commun.*, **8**, 14725.
64. Zalatan,J.G., Lee,M.E., Almeida,R., Gilbert,L.A., Whitehead,E.H., La Russa,M., Tsai,J.C., Weissman,J.S., Dueber,J.E., Qi,L.S. *et al.* (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*, **160**, 339–350.
65. Skourti-Stathaki,K. and Proudfoot,N.J. (2014) A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.*, **28**, 1384–1396.

66. Kim,S., Kim,D., Cho,S.W., Kim,J. and Kim,J.S. (2014) Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.*, **24**, 1012–1019.
67. Charpentier,M., Khedher,A.H.Y., Menoret,S., Brion,A., Lamribet,K., Dardillac,E., Boix,C., Perrouault,L., Tesson,L., Geny,S. *et al.* (2018) CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat. Commun.*, **9**, 1133.
68. Chu,V.T., Weber,T., Wefers,B., Wurst,W., Sander,S., Rajewsky,K. and Kühn,R. (2015) Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.*, **33**, 543–548.
69. Maruyama,T., Dougan,S.K., Truttmann,M.C., Bilate,A.M., Ingram,J.R. and Ploegh,H.L. (2015) Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.*, **33**, 538–542.
70. Song,J., Yang,D., Xu,J., Zhu,T., Chen,Y.E. and Zhang,J. (2016) RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nat. Commun.*, **7**, 10548.
71. Liu,Y., Han,X., Yuan,J., Geng,T., Chen,S., Hu,X., Cui,I.H. and Cui,H. (2017) Biallelic insertion of a transcriptional terminator via the CRISPR/Cas9 system efficiently silences expression of protein-coding and non-coding RNA genes. *J. Biol. Chem.*, **292**, 5624–5633.