SCIENCE CHINA
Life Sciences

•COVER ARTICLE•

May 2021 Vol.64 No.5: 667–677
https://doi.org/10.1007/s11427-020-1794-2

Illuminating single genomic loci in live cells by reducing nuclear background fluorescence

Song Lu[1,2], Dianbing Wang[2], Yu Hou[2,3], Dongge Guo[2], Yulin Deng[1†*] & Xian-En Zhang[2,3†*]

[1]*School of Life Sciences, Beijing Institute of Technology, Beijing 100081, China;*
[2]*National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China;*
[3]*University of Chinese Academy of Sciences, Beijing 100049, China*

Received June 30, 2020; accepted August 10, 2020; published online October 26, 2020

The tagging of genomic loci in living cells provides visual evidence for the study of genomic spatial organization and gene interaction. CRISPR/dCas9 (clustered regularly interspaced short palindromic repeats/deactivated Cas9) labeling system labels genes through binding of the dCas9/sgRNA/fluorescent protein complex to repeat sequences in the target genomic loci. However, the existence of numerous fluorescent proteins in the nucleus usually causes a high background fluorescent readout. This study aims to limit the number of fluorescent modules entering the nucleus by redesigning the current CRISPR/dCas9-SunTag labeling system consisting of dCas9-SunTag-NLS (target module) and scFv-sfGFP-NLS (signal module). We removed the nuclear location sequence (NLS) of the signal module and inserted two copies of EGFP into the signal module. The ratio of the fluorescent intensity of the nucleus to that of the cytoplasm (N/C ratio) was decreased by 71%, and the ratio of the signal to the background (S/B ratio) was increased by 1.6 times. The system can stably label randomly selected genomic loci with as few as 9 repeat sequences.

CRISPR/dCas9 labeling, background reduction, low-repetitive loci, live cell imaging

Citation: Lu, S., Wang, D., Hou, Y., Guo, D., Deng, Y., and Zhang, X.E. (2021). Illuminating single genomic loci in live cells by reducing nuclear background fluorescence. Sci China Life Sci 64, 667–677. https://doi.org/10.1007/s11427-020-1794-2

INTRODUCTION

Fluorescent labeling of genes provides visual information for cell biology research. Many methods have been developed for these purposes, such as fluorescence in situ hybridization (FISH) (Narayanswami and Hamkalo, 1990; Pinkel et al., 1986), transcription activator-like effector (TALE)-mediated genome visualization (Miyanari et al., 2013), and CRISPR/dCas9 (clustered regularly interspaced short palindromic repeats/deactivated Cas9) labeling (Chen et al., 2013). FISH can obtain the relative spatial position information for several gene loci (Boyle et al., 2001; Roix et al., 2003). However, it can only be used in fixed cells, mainly because the denaturation of DNA is required for the accessibility of the probe, and thus it cannot provide dynamic information. The TALE protein contains a central repeat domain, which mediates DNA recognition, and each repeat unit of 33–35 amino acids specifies one target base that is determined by two critical adjacent amino acids (repeat variable di-residue, RVD) (Miller et al., 2011). The TALE technique can flexibly target DNA sequences in genomic loci of interest by the editing of specific RVDs. This technique has been used to label highly repetitive sequences such as telomeres and centromeres.

†Contributed equally to this work
*Corresponding authors (Yulin Deng, email: deng@bit.edu.cn; Xian-En Zhang, email: zhangxe@ibp.ac.cn)

A CRISPR labeling system consists of a fluorescent protein-tagged dCas9 and a scaffold-optimized single-guide RNA (sgRNA) (Chen et al., 2013). The complex targets the locus of interest by recognition of the PAM sequence by the specific sgRNA sequence. This method can flexibly be used for genomic loci targeting by designing different specific sgRNA sequences (Nishimasu et al., 2014). If a dCas9-sgRNA complex carries only one copy of the fluorescent protein, the target loci are expected to contain repetitive sequences (referred to as repeats) with >30 copies, so that the binding can produce visible signals.

As most genomic loci have no repeats or have fewer than 20 repeats, developing methods with higher sensitivity is meaningful for genomic labeling (Ma et al., 2018). Multiple different sgRNAs have been co-expressed with the CRISPR/dCas9-EGFP system to bind multiple sequences of target genes to ensure sufficient EGFP allocation to the target, which requires extensive plasmid construction (Chen et al., 2013). Another strategy is to increase the copy number of the fluorescent proteins in a single sgRNA. One system is based on the modification of sgRNA scaffolds to introduce aptamer sequences into sgRNAs to recruit two or more fluorescent proteins. When tandem aptamers with more than 2 copies were integrated into the sgRNA scaffold, nonspecific foci may be visualized in the nucleus (Hong et al., 2018). Another system is based on the dCas9-SunTag system (Shao et al., 2018; Tanenbaum et al., 2014; Ye et al., 2017). It consists of two subassemblies; one is dCas9-SunTag-NLS, which is a fusion structure consisting of dCas9, a GCN4 peptide array with 10 or 24 copies and 4 copies of nuclear localization sequence (NLS), while the other is scFv-sfGFP-NLS, which is a fusion structure consisting of a GCN4-binding scFv, a superfolder GFP (sfGFP) and an NLS. The system is able to carry a number of fluorescent proteins equivalent to the number of GCN4 peptides to the nucleus to bind target loci. The dCas9-SunTag system has been successfully applied to label nonrepetitive genomic loci by using 20 sgRNAs and genomic loci that contain as few as 15 copies of a repeat (Ye et al., 2017). Nevertheless, among the tested genomic loci, only one locus with 15 repeats was detected, and loci with fewer than 15 repeats were not reported.

Therefore, it appears that visualization of low copy number repeat sequences is limited by merely increasing the number of fluorescent proteins. We then considered trying to decrease the nuclear background fluorescence rather than increasing the number of fluorescent proteins. Two attempts have already been made to do this. The bimolecular fluorescent complimentary system (BiFC) was first used to label genomic loci by fusing with TALE to decrease the nuclear background, which was then integrated into the dCas9-SunTag/sgRNA-aptamer system for labeling (Hong et al., 2018; Hu et al., 2017). The technique increased the signal-to-noise ratio of the detection of an *MUC4* locus that contained 90 repeats. One concern was the long maturation time (hours) and low recovery rate required BiFC to restore fluorescence; these problems need to be solved before this technique can be used to dynamically observe genome function (Hu et al., 2002; Shyu et al., 2006). Another attempt involved assembling dCas9-EGFP and Cy3-labeled gRNAs *in vitro* to form complexes and delivering them into cells to label the target sequence (Wang et al., 2019). More than 95% of the gRNA signals were degraded, while long-lasting labeling by Cy3-gRNA at the labeling site was observed. The signal-to-background ratio (S/B) of the gRNA was >4-fold higher than that of dCas9-EGFP. However, the authors did not explain why Cy3 did not contribute to the background signal after the degradation of the gRNA, and the system was used for target loci with more than 300 copies of repeats.

Here, based on a systematic experimental investigation, we propose a size-controlled background reduction method to improve the S/B ratio that uses the dCas9-SunTag (10×GCN4) system as a starting point (Figure 1A and B). First, we removed the NLS (SV40, Simian virus 40 large T antigen) from the signal modules of the system; i.e., the NLS was removed from scFv-sfGFP-NLS to generate scFv-sfGFP, which caused its migration into the nucleus to change from guided to passive diffusion. Second, we inserted two copies of EGFP into the C-terminus of scFv-sfGFP to increase its size to 117 kD to generate scFv-sfGFP-2EGFP, which was within the critical size range of the pore complex (110–135 kD) (Wang and Brattain, 2007). This design aimed to allow more scFv-sfGFP-2EGFP to remain in the cytoplasm. In addition, scFv-sfGFP-2EGFP would combine with dCas9-SunTag-NLS ($4\times NLS_{SV40}$) in the cytoplasm to form larger complexes that were too large to enter the nucleus, which would further increase the amount of scFv-sfGFP-2EGFP retained in the cytoplasm. As a result, the comprehensive effect of the design largely improved the S/B ratio and showed its capability to label genomic loci containing 9–19 copies of repeat sequences.

RESULTS

Reduction in the fluorescent nuclear-to-cytoplasm (N/C) ratio

We first tested the nuclear-cytoplasmic distribution of the signal module scFv-sfGFP without the NLS. It was constructed and expressed in HeLa cells. The ratio of the fluorescence mean intensity in the nucleus and the fluorescence mean intensity in the cytoplasm (fluorescent N/C ratio) was used to represent the nuclear-cytoplasmic distribution of the signal modules. The fluorescent N/C ratio ±S.D. was determined to be 1.31±0.13 (Table 1, Figure 2A). This indicates that scFv-sfGFP without the NLS was able to diffuse into the nucleus passively and that the nuclear entry

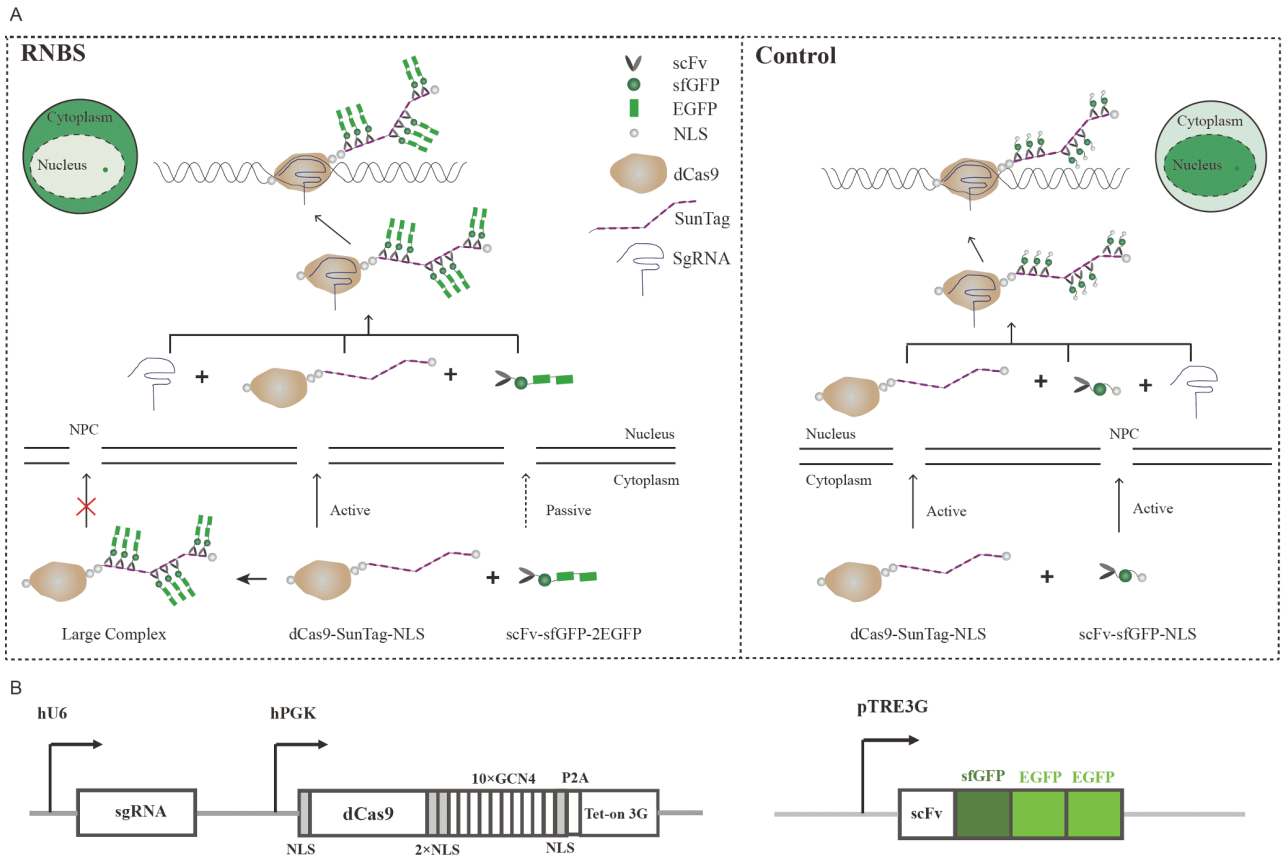

Figure 1 Design of the reduced nuclear background system (RNBS). A, Schematics of the reduced nuclear background labeling system and control system. For RNBS, dCas9-SunTag-NLS actively enters the nucleus, scFv-sfGFP-2EGFP passively enters the nucleus, and the large complex is blocked in the cytoplasm. B, Components of RNBS. dCas9 was fused with four nuclear localization signals (NLS) and $GCN4_{10\times}$ (SunTag), which were expressed using the human phosphoglycerate kinase (hPGK) promoter. A 2A self-cleaving peptide (P2A) linker was added following dCas9 to co-express the Teton3G protein. The sgRNA was constitutively transcribed under the control of the human U6 promoter and was added to the plasmid with dCas9. scFv-sfGFP was fused with two copies of EGFP and expressed under the control of a TRE3G promoter.

Table 1 Fluorescent N/C ratio of the signal modules and labeling module pairs[a)]

Signal module	Fluorescent N/C ratio	Labeling module pair	Fluorescent N/C ratio
scFv-sfGFP-NLS	Not shown	scFv-sfGFP-NLS & dCas9-SunTag-NLS	2.63±0.94 (*n*=35)
scFv-sfGFP	1.31±0.13 (*n*=32)	scFv-sfGFP & dCas9-SunTag-NLS	1.15±0.26 (*n*=38)
scFv-sfGFP-EGFP	1.08±0.10 (*n*=34)	scFv-sfGFP-EGFP & dCas9-SunTag-NLS	0.96±0.21 (*n*=34)
scFv-sfGFP-2EGFP	0.94±0.10 (*n*=36)	scFv-sfGFP-2EGFP & dCas9-SunTag-NLS	0.77±0.09 (*n*=36)
scFv-sfGFP-3EGFP	0.77±0.09 (*n*=36)	scFv-sfGFP-3EGFP & dCas9-SunTag-NLS	0.65±0.12 (*n*=36)

a) The data are presented as the mean±S.D. (cell number).

process was not limited by the nuclear pore complex. Then, a series of scFv-sfGFP-derived signal modules were constructed by inserting different numbers of copies of EGFP to enlarge its size (Figure 2; Table S1 in Supporting Information). As shown in Table 1 and Figure 2B, the fluorescent N/C ratio decreased with increasing signal module size. In other words, without the help of the NLS, the larger the molecular weight was, the greater the resistance to migration through the nuclear pore complex was, which resulted in more signal modules remaining in the cytoplasm.

We further tested the fluorescent N/C ratios of different signal module and target module (dCas9-SunTag-NLS plus sgRNA) pairs (labeling module pairs) (Table 1 and Figure 2C). The plasmid encoding a sgRNA scaffold and dCas9-SunTag-NLS was cotransferred with plasmids encoding different signal modules (Figure 1B). As predicted, the resulting fluorescent N/C ratios were ranked from high to low as follows: scFv-sfGFP-NLS & dCas9-SunTag-NLS, scFv-sfGFP & dCas9-SunTag-NLS, scFv-sfGFP-EGFP & dCas9-SunTag-NLS, scFv-sfGFP-2EGFP & dCas9-SunTag-NLS and scFv-sfGFP-3EGFP & dCas9-SunTag-NLS (Table 1 and Figure 2C). These results verified our design principle, in

that changing nuclear migration from guided to passive diffusion allowed a significant number of signal modules to be retained in the cytoplasm and that the pore size limitation mechanism successfully kept more signal modules (scFv-sfGFP-nEGFP) out of the nucleus (Figure S1 in Supporting Information). In addition, by comparing the Table 1 data, we found that the fluorescent N/C ratios produced by single signal modules were generally higher than those produced by co-expression of the labeling module pairs. This indicates that some of the signal modules and target modules had formed complexes in the cytoplasm, which prevented their entering the nucleus, as predicted.

Fortunately, the least fluorescent N/C ratio (0.65±0.12) indicates that a few of the signal modules were still able to enter the nucleus (Table 1). Next, we tested whether the reduction in the number of signal modules still allowed efficient labeling of genomic loci with a limited number of repeats.

Increase in the fluorescent S/B ratio

A total of 22 repeats containing the genomic locus (named locus #1) of chromosome 8 (Qin et al., 2017) and four labeling module pairs were used to evaluate the fluorescent S/B ratio (Table 2; Table S2 in Supporting Information). The pairs were co-expressed in HeLa cells. The original labeling system, the scFv-sfGFP-NLS/dCas9-SunTag-NLS pair, was used as a control for comparison. After transfection of the four labeling module pairs respectively, the fluorescent loci were observed (Figure 3A; Figure S2 in Supporting In-

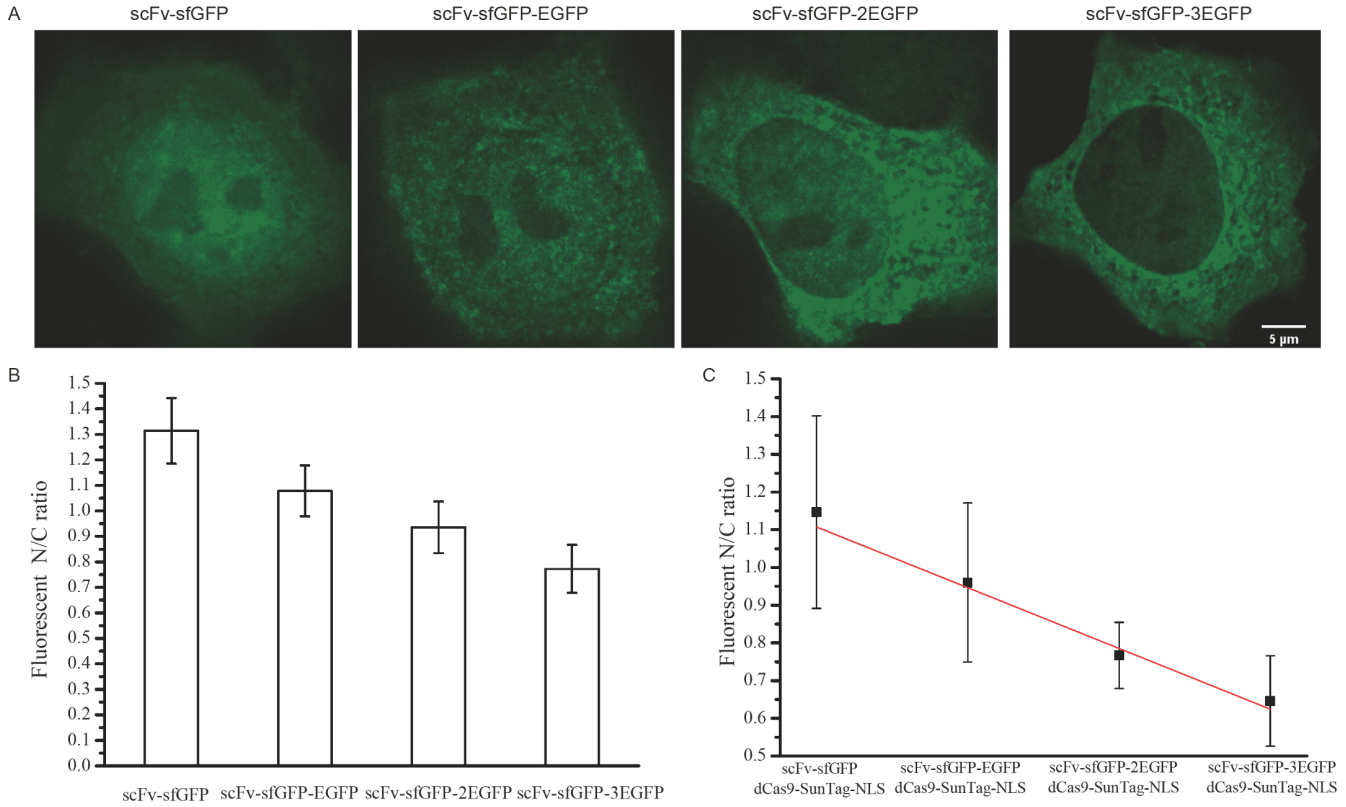

Figure 2 Development of RNBS. A, Representative images of scFv-sfGFP fused with different numbers of copies of EGFP. Scale bar, 5 μm. B, Fluorescent N/C ratio for scFv-sfGFP, scFv-sfGFP-EGFP, scFv-sfGFP-2EGFP, and scFv-sfGFP-3EGFP. All data are presented as the mean±S.D. The number of cells from left to right is 32, 34, 36, and 36. C, Fluorescent N/C ratio for the cotransfection of scFv-sfGFP, scFv-sfGFP-EGFP, scFv-sfGFP-2EGFP, and scFv-sfGFP-3EGFP with dCas9-SunTag-NLS. The number of cells from left to right is 38, 44, 40, and 21. All data are presented as the mean±S.D.

Table 2 Fluorescent S/B ratio of the four labeling module pairs[a)]

Experiment	Labeling module pair	Fluorescent S/B ratio
Control	scFv-sfGFP-NLS & dCas9-SunTag-NLS	4.49±2.87 (*n*=31)
1	scFv-sfGFP & dCas9-SunTag-NLS	5.63±3.56 (*n*=25)
2	scFv-sfGFP-EGFP & dCas9-SunTag-NLS	7.43±4.65 (*n*=28)
3	scFv-sfGFP-2EGFP & dCas9-SunTag-NLS	11.65±8.73 (*n*=31)

a) The data are presented as the mean±S.D. (cell number).

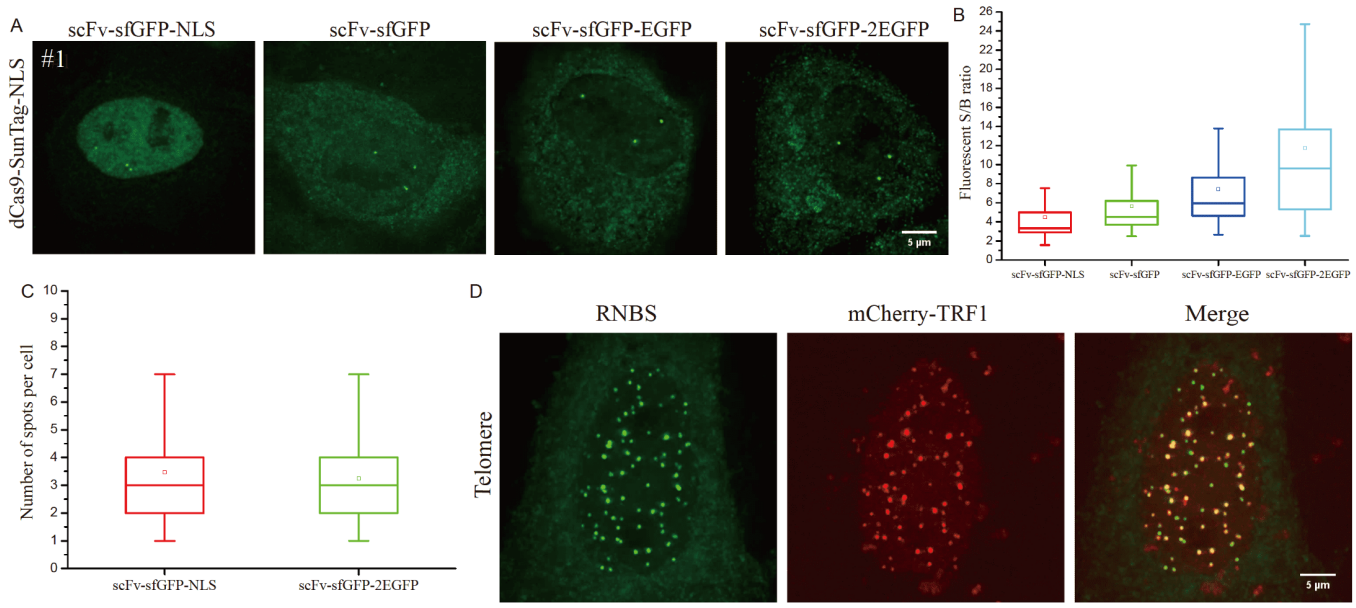

Figure 3 Labeling performance of RNBS and verification of RNBS reliability. A, Labeling of locus #1 using dCas9-SunTag-NLS with scFv-sfGFP, scFv-sfGFP-EGFP, scFv-sfGFP-2EGFP. Scale bar, 5 μm. B, Boxplot showing the fluorescent S/B ratio of locus #1 respectively labeled by four labeling module pairs. Number of cells from left to right are 31, 25, 28 and 31. C, Boxplot showing the number of spots for locus #1 per cell using scFv-sfGFP-NLS and scFv-sfGFP-2EGFP with dCas9-SunTag-NLS respectively. Number of cells from left to right are 50, 41. In the boxplot in B and C, the line and the dot within the boxplot respectively represent the median and the mean. The outer edges of the box are the 25th and 75th percentiles. The whiskers extend to the minimum and maximum values. D, Co-labeling of telomeres using RNBS (green) and mCherry-TRF1 (red). Scale bar, 5 μm.

formation). The S/B ratios are summarized in Table 2 and Figure 3B. It is clear that both removing NLS from the signal module and adding EGFP had an obvious influence on the fluorescent S/B ratio. The fluorescent S/B ratios of pair 1, 2, and 3 were 1.3-fold, 1.7-fold and 2.6-fold greater than those of the control pair, respectively. Therefore, we chose scFv-sfGFP-2EGFP & dCas9-SunTag-NLS as the newly built reduced nuclear background system (RNBS). The labeling efficiency of locus #1 by RNBS was 36.6% (41 of 112 cells). In addition, as the number of scFv-sfGFP-3EGFP molecules entering the nucleus decreased, the labeling efficiency decreased. Only 26 cells out of 166 cells exhibited successful labeling of the target loci.

The two fluorescent spots shown represent the pair of genomic loci on homologous chromosomes labeled by the system. Due to the karyotype heterogeneity of cancer cells and chromosome replication, the number of genomic loci per HeLa cell may be more than two. We compared the number of spots for locus #1 per cell labeled by the RNBS method with that of the control group. As shown in Figure 3C and Table 3, the spot number for the RNBS method was 3.46 ±1.67 (n=41), while the spot number in the control group was 3.24±1.24 (n=50). Both results were nearly the same, indicating the reliability of the RNBS. Then, we used the system to label the repetitive sequences of telomeres and used the telomere-binding protein fused with mCherry (mCherry-TRF1) to co-label telomeres (Chen et al., 2013). The colocalization ratio of the sfGFP-EGFP signal and mCherry signal was about 92% (n=12), indicating the specificity of the RNBS (Figure 3D; Figures S3 and S4 in Supporting Information).

Labeling of genomic loci with 9–24 copies of repeat sequences

As most genomic loci have fewer than 20 copies of repeat sequences (Figure 4A), we chose two genomic loci that contain repeats with <20 copies on chromosome 3 to test the RNBS method (Table 3; Table S2 in Supporting Information). One locus (locus #2) is located in gene *TFRC* and contains 18 repeats, and the other (locus #3) is located in the gene *TNK2* and contains 9 repeats. The two loci are both located on chromosome 3 and are located 316 kb and 96 kb away from a locus in the *MUC4* gene that contains 34 copies of repeats, which has been reported to be labeled successfully using SadCas9-mCherry (Figure 4B) (Chen et al., 2016). We thus chose the *MUC4* gene to verify the labeling specificity of the two loci. Both the RNBS and SadCas9-mCherry labeling systems were co-expressed in HeLa cells. As hoped, the two loci were labeled successfully and located close to *MUC4* spatially (Figure 4C). These results showed the specificity and reliability of the labeling of the two loci using RNBS. The labeling efficiencies for *TFRC* and *TNK2* were respectively 33.9% (41 of 121 cells) and 27.2% (34 of 125 cells). The fluorescent S/B ratios of *TFRC* and *TNK2* were 6.69±5.02 (n=23) and 5.20±3.54 (n=33), respectively (Figure 4D and Table 3). To further evaluate the labeling performance of RNBS, the control labeling pair scFv-sfGFP-

Table 3 Genomic loci labeled by RNBS[a)]

Genomic locus	Copies of repeat	Fluorescent spots per cell	Fluorescent S/B ratio
#1, chromosome 8	22	3.46±1.67 (*n*=41)	11.65±8.73 (*n*=31)
#2, chromosome 3	18	2.59±1.07 (*n*=41)	6.69±5.02 (*n*=23)
#3, chromosome 3	9	2.76±1.07 (*n*=34)	5.20±3.54 (*n*=33)
#4, chromosome 19	21	2.64±1.28 (*n*=28)	4.20±2.33 (*n*=13)
#5, chromosome 19	24	2.55±0.88 (*n*=31)	6.25±5.56 (*n*=13)
#6, chromosome 10	13	2.15±0.77 (*n*=27)	3.86±1.78 (*n*=13)
#7, chromosome 19	13	2.19±0.70 (*n*=31)	5.84±2.27 (*n*=14)
#8, chromosome 6	11	2.27±0.93 (*n*=22)	3.55±0.87 (*n*=13)
#9, chromosome 12	12	2.52±0.99 (*n*=31)	6.29±2.85 (*n*=13)
#10, chromosome 17	9	2.37±0.63 (*n*=27)	4.87±3.49 (*n*=13)
#11, chromosome 17	12	2.29±0.91 (*n*=24)	5.24±3.13 (*n*=13)

a) The data are presented as the mean±S.D. (cell number).

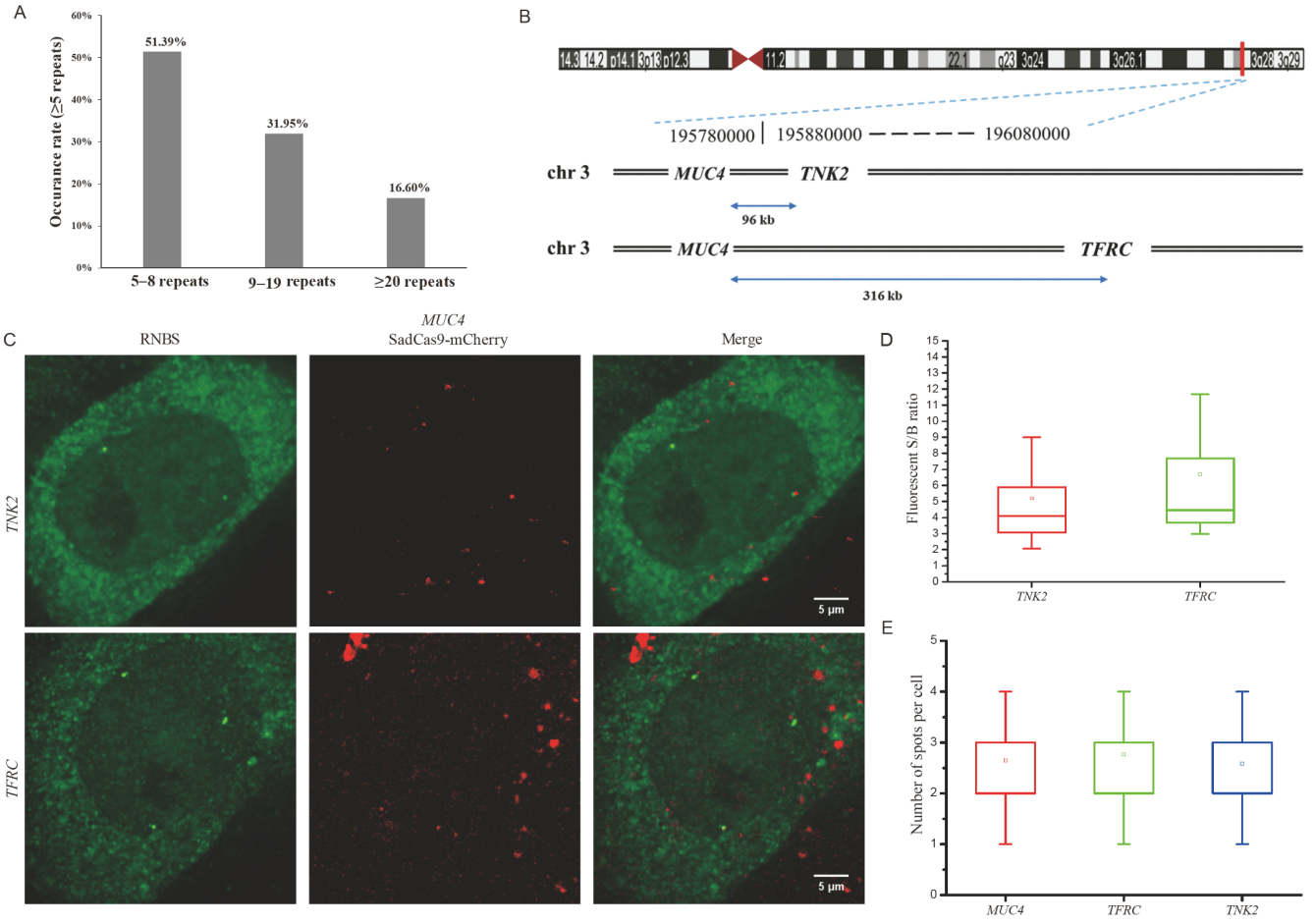

Figure 4 Labeling and verification of gene loci containing 9 copies and 18 copies of repeats. A, Histogram showing the occurrence rate of the genomic loci containing sequences≥5 repeats. B, Positions of *TNK2* and *TFRC*, with *MUC4* upstream, on chromosome 3. C, Colabeling of *TNK2* and *TFRC* with *MUC4* using the RNBS (green) and SadCas9-mCherry (red), respectively. Scale bar, 5 μm. D, Boxplot showing the signal-to-background ratios of loci in *TNK2* and *TFRC*. The number of cells from left to right is 23 and 33. E, Boxplot showing the number of spots per cell for *MUC4*, *TFRC* and *TNK2*. The number of cells from left to right is 28, 34, and 41. In the boxplots in D and E, the line and the dot within the boxplot represent the median and the mean, respectively. The outer edges of the box represent the 25th and 75th percentiles. The whiskers extend to the minimum and maximum values.

NLS & dCas9-SunTag-NLS (Chen et al., 2013) and scFv-sfGFP-EGFP & dCas9-SunTag-NLS were also used for calculation of the fluorescent S/B ratios for *TFRC*. As a result, RNBS showed the highest S/B ratio (Figure S5 in Supporting Information). Then, the previously reported *MUC4* locus with 32 copies of repeats located close to *TFRC*

and *TNK2* was labeled by the control labeling pair scFv-sfGFP-NLS & dCas9-SunTag-NLS. The number of fluorescent spots per cell for the three loci were 2.64±1.28 (n=28), 2.59±1.07 (n=41) and 2.76±1.07 (n=34) for *MUC4*, *TFRC* and *TNK2*, respectively (Figure 4E and Table 3; Figure S4 in Supporting Information). The three approximate results suggest that the RNBS was reliable when used to label genomic loci that each contained fewer than 20 copies of repeats (Figure 4E; Figure S6 in Supporting Information).

To further evaluate the performance of the RNBS, we chose 3 groups of target loci in the human genome (Table 3; Table S2 in Supporting Information). Group 1 contained two genomic loci that were not detectable in a previous report using the CRISPR-Sirius method with a sgRNA scaffold inserted by the 8×MS2 sequence (Ma et al., 2018), which is named locus #4 here and is located in the gene *DOT1L*; locus #5 is in the gene *FSD1*. Both are located on chromosome 19. The *DOT1L* locus has 21 repeats, and the *FSD1* locus has 24 repeats. RNBS was used to label the two loci (Figure 5A). The fluorescent S/B ratios of *DOT1L* and *FSD1* were 4.20 ±2.33 (n=13) and 6.25±5.56 (n=13) respectively (Table 3). The spot number for *DOT1L* per cell was 2.64±1.28 (n=28), and the spot number for *FSD1* per cell was 2.55±0.88 (n=31) (Figure 5B and Table 3), which was also in accordance with the karyotype of the cancer cell. Therefore, RNBS was able to detect genomic loci with low repeats which were not detected by CRISPR-Sirius.

Group 2 also had two loci, locus #6 in *DOCK1*, which is located in the lamina-associated domain (LAD), and locus #7 in *TDRD12*, which is located in the interior of the nucleus (non-LAD) (Qin et al., 2017). They both contain 13 repeats and have been successfully labeled by a previously reported method (Qin et al., 2017). The two loci were successfully labeled by RNBS (Figure 5B), and the fluorescent S/B ratios of *DOCK1* and *TDRD12* were 3.86±1.78 (n=13) and 5.84 ±2.27 (n=14) respectively (Table 3). The spot numbers per cell for the two loci were 2.15±0.77 (n=27) and 2.19±0.70 (n=31), respectively (Figure 5D and Table 3).

We then randomly selected more genomic loci from the human genome to test the applicability of the RNBS method. These loci (loci #8, #9, #10, and #11) are distributed on different chromosomes and contain a number of repeats ranging from 9 to 19. The results are shown in Table 3 and Figure 6. The loci were successfully labeled with an average spot number per cell more than 2, which indicated the robustness of the RNBS.

In addition, the dynamic movies of telomere and locus #3 in HeLa cells were also showed (Movies S1 and S2 in Supporting Information) by this method, presenting the ability for dynamic study in living cells using RNBS. We also labeled locus #1, locus #2 and locus #3 in HEK293T cells, which indicates the compatibility of RNBS (Figure S7

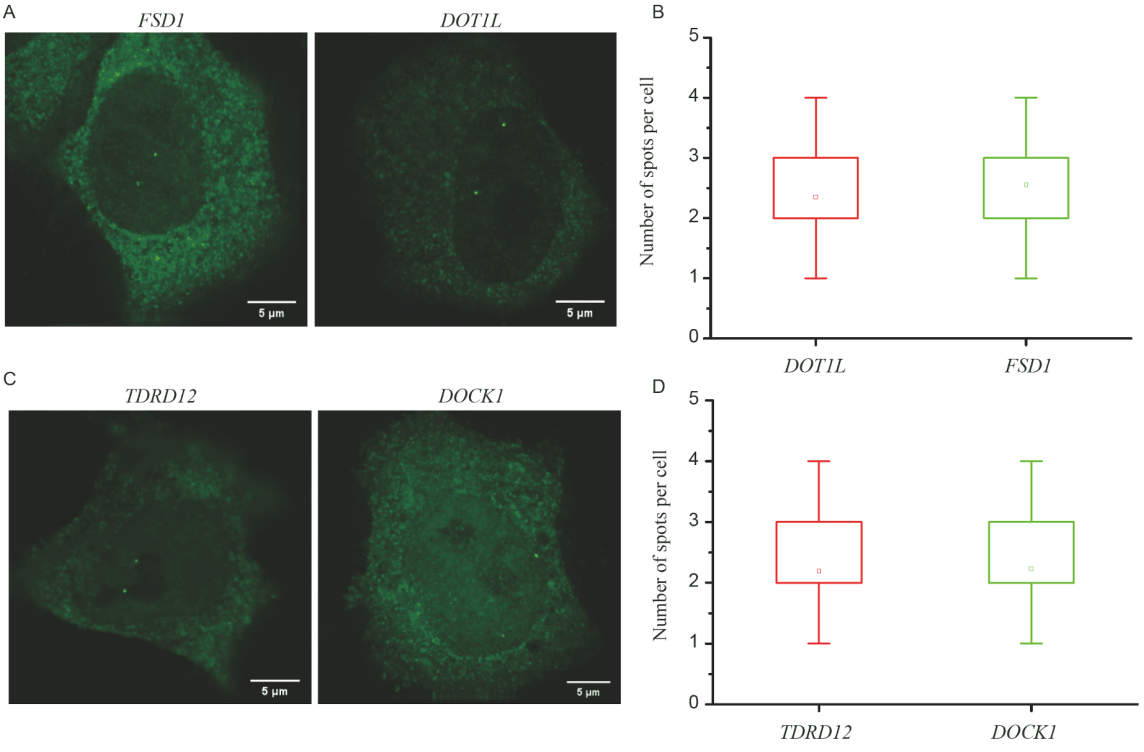

Figure 5 Labeling ability of the RNBS. A, Labeling of *FSD1* and *DOT1L* on chromosome 19. Scale bar, 5 μm. B, Boxplot showing the number of spots per cell for *DOT1L* and *FSD1*. The number of cells from left to right is 34 and 31. C, Labeling of *TDRD12* on chromosome 19 and *DOCK1* on chromosome 10. Scale bar, 5 μm. D, Boxplot showing the number of spots per cell for *TDRD12* and *DOCK1*. The number of cells from left to right is 31 and 27. In the boxplots in B and D, the line and the dot within the boxplot represent the median and the mean, respectively. The outer edges of the box represent the 25th and 75th percentiles. The whiskers extend to the minimum and maximum values.

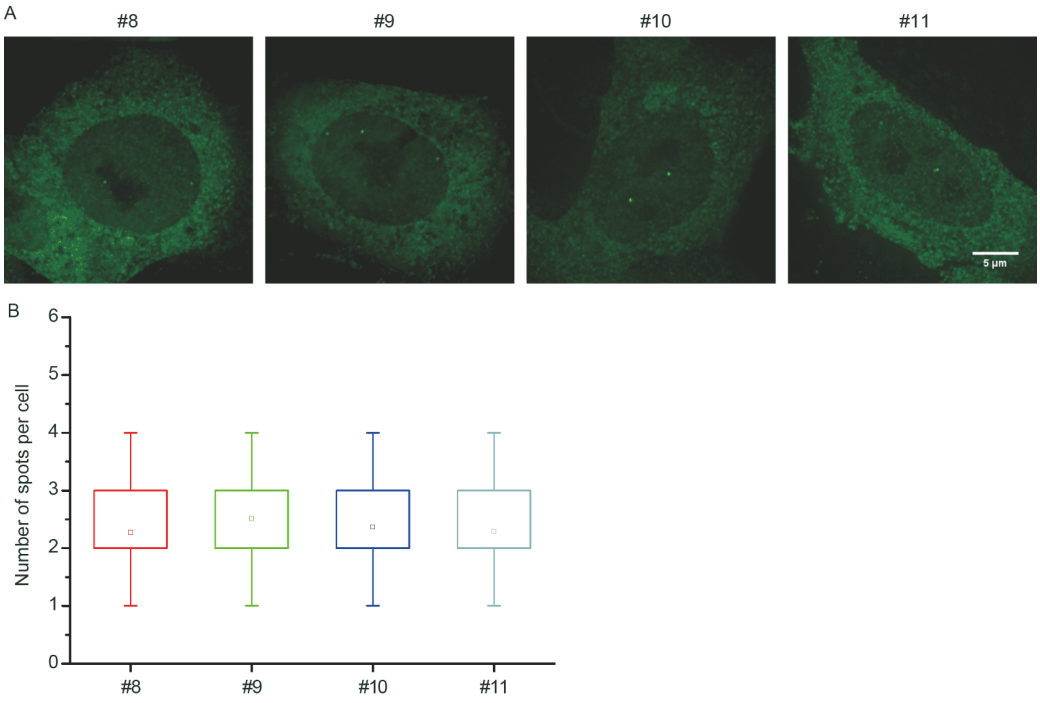

Figure 6 Robustness of RNBS. A, Labeling of locus #8 on chromosome 6, locus #9 on chromosome 12, and locus #10 and locus #11 on chromosome 17 using the RNBS. Scale bar, 5 μm. B, Boxplot showing the number of spots per cell for loci #8, #9, #10, and #11. The line and the dot within the boxplot represent the median and the mean, respectively. The outer edges of the box represent the 25th and 75th percentiles. The whiskers extend to the minimum and maximum values. The number of cells from left to right is 22, 31, 27, and 24.

in Supporting Information).

DISCUSSION

Great efforts have been made to label single genomic loci. It is still challenging to label low-copy repeat sequences containing genomic loci. In this study, we aimed to reduce the fluorescent N/C ratio to increase the fluorescent S/B ratio by size-controlled nuclear entry rather than by increasing the number of fluorescent proteins in the labeling module, which is the normal method. There are two ways for fluorescent proteins to enter the nucleus: passive diffusion and oriented transport guided by the NLS. It has been reported that in HeLa cells, tandem GFP proteins with sizes ranging from 90 to 110 kD were allowed to diffuse through the nuclear pore, and a 5-copy GFP fusion with a molecular weight of approximately 135 kD was mainly distributed in the cytoplasm (Wang and Brattain, 2007). Therefore, it can be inferred that 110–135 kD is the critical size range of the HeLa nuclear pore complex for the passive diffusion of molecules, which may vary for the different proteins studied. The molecular size of the scFv-sfGFP-nEGFP labeling modules without an NLS ranged from 62.7 kD to 145 kD, which was within the critical range of the nuclear pore size. The design that changed the entry method from nucleus-oriented movement to passive diffusion successfully reduced the fluorescent N/C ratio. It was also reported that an *S. thermophilus St1* dCas9 protein fused with three copies of GFP and three copies of NLS_{SV40} having a molecular weight of approximately 213 kD could barely enter into the nucleus (Ma et al., 2015). For any CRISPR/dCas9 labeling system, when the target module and the signal module are co-expressed in cells, some of these modules interact with each other to form a complex before they both enter the nucleus. Complexes with sizes exceeding the critical range of the nuclear pore size were restricted to the cytoplasm. In the RNBS system, the molecular weight of the complex of dCas9-SunTag-NLS and scFv-sfGFP-2EGFP reached 1,372 kD, which was too large to pass through the nuclear pore, and thus this decreased the overall intensity of fluorescence in the nucleus. This was also supported by our experimental evidence that the fluorescent N/C ratios of the co-expressed dCas9-SunTag-NLS and scFv-sfGFP-nEGFP (n=0–3) pairs were all lower than those of scFv-sfGFP alone. Removing the NLS and setting a size limitation for diffusion had obvious effects on the fluorescence N/C ratio. In the control experiment using the dCas9-SunTag-NLS/scFv-sfGFP-NLS pair, the fluorescent N/C ratio was 2.63, which implied that the majority of signal module molecules had entered the nucleus, while in the reduced nuclear background experiment using the dCas9-SunTag-NLS/scFv-sfGFP-2EGFP pair, the fluorescent N/C ratio was decreased to 0.77, which indicated that the ma-

jority of signal module molecules were unable to enter the nucleus. Meanwhile, the remaining target module molecules and signal module molecules could still enter the nucleus to form complexes and label the target genomic loci. These findings might be a clue for improving the gene editing efficiency of the CRISPR/Cas9 system, the efficiency of which is also affected by the abundances of the editing modules.

Using the proposed method, 11 genomic loci with repeat numbers ranging from 9 to 24 were successfully labeled. They are distributed on 7 chromosomes, two of which failed to be labeled by other methods in previous reports, demonstrating the reliability of this method. However, the system failed to label a genomic locus with 5 repeats (Figure S8 in Supporting Information), which indicates that there is room for improvement. In addition, another 2 loci with 9 repeats and 14 repeats were not noticeably labeled, which may be due to the poor accessibility of these genomic sites due to the complex spatial structure of the chromosomes. This may be related to the normal CRISPR/dCas9 gene editing technology with similar problem (Wang et al., 2014).

By using this method, we achieved labeling of low-repetitive loci containing 9–19 repeats. The method is robust, while the existing methods have enabled the robust labeling of loci with more than 20 repeats. This means our method broadens the taggable range of the genome, which has clear significance for the study of cell biology and genomics. Meanwhile, the reduction of background fluorescence in the nucleus provides an opportunity for visualizing the 3D spatial structure of the genome in live cells.

This method has two other advantages. It avoids the use of exogenous substances such as quantum dots and fluorescent dyes, which are often suspected to interfere with cellular processes. Endogenous labeling reagents (fluorescent protein modules) can avoid or minimize potential interference. Some labeling methods rely on multiple sgRNAs to bind different sequences in the target loci to ensure adequate labeling efficiency, but this requires to collectively transfer multiple kinds of plasmids, and the success rate is low. In the proposed RNBS method, only two plasmids are required, which makes the method simple and highly feasible.

As predicted, some labeling pairs formed aggregates due to accumulation of large complex in the cytoplasm (Figure 4C). In reality, when collecting target signals, the cells were sliced optically along the *z*-axis with a thickness of 250 nm. Because of the background reduction strategy in this study, the boundary between the nucleus and the cytoplasm is clear, so the optical slicing starts from the top of the nucleus, which can avoid the interference from the aggregation spot in the cytoplasm. After certain number of slicing, most of the labeled information can be obtained.

Still, it remains a challenge to label genomic loci containing fewer numbers of repeats and the one with non-repetitive sequences. The combined use of background fluorescence reduction and modules with brighter emission is a high priority.

MATERIALS AND METHODS

Construction of dCas9-SunTag, sgRNA and scFv-sfGFP-derived modules

The NLS_{SV40}-dCas9-NLS_{SV40} fragment was amplified from pSLQ1645-dCas9-GFP (Addgene Plasmid#51023). The pGK promoter was amplified from pLVX-T2A-mCherry (a gift from Chen Chang, Institute of Biophysics, CAS, Beijing). pGK-NLS_{SV40}-dCas9-2×NLS_{SV40}-10XGCN4_V4-NLS_{SV40} was generated by ligating the fragments into the pX330 vector using Gibson Assembly. Then, a P2A peptide was used to co-express NLS_{SV40}-dCas9-2×NLS_{SV40}-10XGCN4_V4-NLS_{SV40} and the Tet-on 3G system (Tet-on 3G Inducible Expression System, Clontech, USA). To simplify the system, the U6 promoter and sgRNA extensions containing two adjacent BbsI sites (Addgene Plasmid#42230) were also ligated into the dCas9 plasmid. SgRNA scaffolds contain A–U flip and hairpin extensions to increase the targeting efficiency. The NLS_{SV40}-dSaCas9-NLS_{SV40}-mCherry fragment (Addgene Plasmid# 85452) with the U6 promoter and sgRNA extensions containing two adjacent BbsI sites was constructed with similar strategies as those used for NLS_{SV40}-dCas9-NLS_{SV40}-mCherry-10XGCN4_V4. The scFv-GCN4-sfGFP fragment was amplified from pHR-scFv-GCN4-sfGFP-GB1-NLS_{SV40}-dWPRE (Addgene Plasmid#60906), and the TRE3G promoter was amplified from the pTRE3G vector (Tet-on 3G Inducible Expression System, Clontech). The TRE3G-scFv-GCN4-sfGFP sequence was generated by classical cloning into a pEGFP-N1(dam-) vector. Different copies of EGFP were inserted into the plasmid using Gibson Assembly.

Cell culture and plasmid transfections

HeLa cells and HEK293T cells were cultured in growth medium consisting of Dulbecco's modified Eagle medium (HyClone, USA) and 10% fetal bovine serum (Gibco, USA) at 37°C in 5% CO_2 in a humidified incubator. Cells were imaged 24 h following transfection. HeLa cells and HEK293T cells were then cultured to approximately 70% confluency in glass-bottomed dishes (Nunc) and co-transfected with the plasmid dCas9-SunTag-NLS/sgRNA (1 μg) without doxycycline and scFv-sfGFP-derived plasmids (1.5 μg) using Lipofectamine 3000 (Invitrogen, USA) under the guidance of the manufacturer's protocol. In brief, OPTI-MEM (GIBCO, USA) was used to dilute the DNA plasmids and transfection reagents. Cells were imaged 24–48 h after transfection.

Fluorescence microscopy and image processing

All 3D images of HeLa cells and HEK293T were acquired on the Delta-Vision OMX V3 imaging system (GE Healthcare, USA) with a 100× 1.4 NA oil-immersion objective (UPlanSApo, Olympus, Japan) and solid-state multimode lasers at 405 nm (for TagBFP/Hochst), 488 nm (for EGFP) and 561 nm (for mCherry). The images were recorded with a charge-coupled device (CCD) camera (Evolve 512×512, Photometrics, USA). Serial z-stack sectioning was performed at 250 nm intervals in conventional mode. Immersion oils with refractive indices of 1.512 were used for HeLa and HEK293T cells on glass coverslips to obtain optimal images. The microscope was routinely calibrated with 100 nm fluorescent spheres to calculate both the lateral and axial limits of the image resolution. The image stacks were reconstructed by SoftWoRx 6.1.1 (GE Healthcare) with the following settings: Wiener filter enhancement 0.900, winner filter smoothing 0.800, and further processed to obtain the maximum-intensity projections. Pixel registration was corrected to be less than 1 pixel for all channels using 100 nm Tetraspeck beads.

The reconstructed image datasets were then imported into Imaris software (version 8.1.3, Bitplane, USA). The mean fluorescent intensity of nuclear background EGFP fluorescence was measured and calculated by the Surface-function of Imaris. The Spot function in Imaris automatically located the labeled genomic loci based on size and intensity thresholds. The mean fluorescent intensity of the target spots was acquired by Imaris.

Data analysis

The z-stacks were converted into 2D projections in ImageJ. To define the nuclear region of interest (ROI), watershed segmentation was used for Hoechst images. For each nuclear ROI, a cytoplasmic "ring" ROI with a thickness of 150 nm was generated by dilating the nuclear ROI twice and performing exclusive-or (XOR) operation on these two dilated areas in ImageJ. The mean intensities of the nuclear and cytoplasmic ROIs were measured and exported into Microsoft Excel, where the nuclear to cytoplasmic EGFP signal (N/C) ratio was calculated. Data are represented as the mean ±S.D. The numbers of analyzed cells are mentioned in the text. For all statistical analyses, significance was determined using the unpaired *t*-test (***, $P<0.001$; **, $P<0.01$; *, $P<0.05$). Plots and statistics were generated in either GraphPad Prism version 6.00 or Origin 8.5.

The mean intensity values of the nucleus and the spots obtained from Imaris were then transferred to Microsoft Excel. The foci signal-to-nuclear background ratio was calculated using the following formula:

$$I_R = \frac{I_S - I_B}{I_N - I_B},$$

where I_R is the intensity ratio determined according to the mean intensity of the labeled loci (I_S) and the mean intensity of the nucleoplasm (I_N). The background fluorescence intensity (I_B) was subtracted from the same image based on the intensity in a dark region.

Compliance and ethics *The author(s) declare that they have no conflict of interest.*

Acknowledgements *This work was supported by the National Natural Science Foundation of China (21890743), Strategic Priority Research Program of the Chinese Academy of Sciences, China (XDB29050100) and National Key Research and Development Program of China (2017YFA0205503). We would like to thank the Center for Biological Imaging (CBI), IBP-CAS, particularly Shuoguo Li and Yun Feng for technical support with the Microscopy work.*

References

Boyle, S., Gilchrist, S., Bridger, J.M., Mahy, N.L., Ellis, J.A., and Bickmore, W.A. (2001). The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. Hum Mol Genet 10, 211–219.

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.

Chen, B., Hu, J., Almeida, R., Liu, H., Balakrishnan, S., Covill-Cooke, C., Lim, W.A., and Huang, B. (2016). Expanding the CRISPR imaging toolset with *Staphylococcus aureus* Cas9 for simultaneous imaging of multiple genomic loci. Nucleic Acids Res 44, e75.

Hong, Y., Lu, G., Duan, J., Liu, W., and Zhang, Y. (2018). Comparison and optimization of CRISPR/dCas9/gRNA genome-labeling systems for live cell imaging. Genome Biol 19, 10–1186.

Hu, C.D., Chinenov, Y., and Kerppola, T.K. (2002). Visualization of interactions among bZip and Rel family proteins in living cells using bimolecular fluorescence complementation. Mol Cell 9, 789–798.

Hu, H., Zhang, H., Wang, S., Ding, M., An, H., Hou, Y., Yang, X., Wei, W., Sun, Y., and Tang, C. (2017). Live visualization of genomic loci with BiFC-TALE. Sci Rep 7, 40192.

Ma, H., Naseri, A., Reyes-Gutierrez, P., Wolfe, S.A., Zhang, S., and Pederson, T. (2015). Multicolor CRISPR labeling of chromosomal loci in human cells. Proc Natl Acad Sci USA 112, 3002–3007.

Ma, H., Tu, L.C., Naseri, A., Chung, Y.C., Grunwald, D., Zhang, S., and Pederson, T. (2018). CRISPR-Sirius: RNA scaffolds for signal amplification in genome imaging. Nat Methods 15, 928–931.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., et al. (2011). A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29, 143–148.

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.
Narayanswami, S., and Hamkalo, B.A. (1990). High resolution mapping of *Xenopus laevis* 5S and ribosomal RNA genes by EM *in situ* hybridization. Cytometry 11, 144–152.
Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 156, 935–949.
Pinkel, D., Straume, T., and Gray, J.W. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. Proc Natl Acad Sci USA 83, 2934–2938.
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.
Roix, J.J., McQueen, P.G., Munson, P.J., Parada, L.A., and Misteli, T. (2003). Spatial proximity of translocation-prone gene loci in human lymphomas. Nat Genet 34, 287–291.
Shao, S., Chang, L., Sun, Y., Hou, Y., Fan, X., and Sun, Y. (2018). Multiplexed sgRNA expression allows versatile single nonrepetitive DNA labeling and endogenous gene regulation. ACS Synth Biol 7, 176–186.
Shyu, Y.J., Liu, H., Deng, X., and Hu, C.D. (2006). Identification of new fluorescent protein fragments for bimolecular fluorescence complementation analysis under physiological conditions. Biotechniques 40, 61–66.
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.
Wang, H., Nakamura, M., Abbott, T.R., Zhao, D., Luo, K., Yu, C., Nguyen, C.M., Lo, A., Daley, T.P., La Russa, M., et al. (2019). CRISPR-mediated live imaging of genome editing and transcription. Science 365, 1301–1305.
Wang, R., and Brattain, M.G. (2007). The maximal size of protein to diffuse through the nuclear pore is larger than 60 kDa. FEBS Lett 581, 3164–3170.
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.
Ye, H., Rong, Z., and Lin, Y. (2017). Live cell imaging of genomic loci using dCas9-SunTag system and a bright fluorescent protein. Protein Cell 8, 853–855.

SUPPORTING INFORMATION

The supporting information is available online at https://doi.org/10.1007/s11427-020-1794-2. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.