

Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow

Hanhui Ma¹, Li-Chun Tu², Ardalan Naseri³, Maximiliaan Huisman², Shaojie Zhang³, David Grunwald² & Thoru Pederson¹

A lack of techniques to image multiple genomic loci in living cells has limited our ability to investigate chromosome dynamics. Here we describe CRISPRainbow, a system for labeling DNA in living cells based on nuclease-dead (d) Cas9 combined with engineered single guide RNA (sgRNA) scaffolds that bind sets of fluorescent proteins. We demonstrate simultaneous imaging of up to six chromosomal loci in individual live cells and document large differences in the dynamic properties of different chromosomal loci.

Although currently CRISPR technology is mostly applied to gene editing and regulation¹, we and others have used it to label defined chromosomal loci to image the three-dimensional structure of the genome in live cells^{2–4}, complementing and extending information obtained from fluorescence *in situ* hybridization studies in fixed cells. We previously engineered three orthologous CRISPR systems for multicolor labeling of chromosomal loci in human cells⁴. In this report we extend the number of different colors that can be imaged simultaneously to six by engineering the sgRNAs instead of the Cas9 protein to bind combinations of different fluorescent proteins.

sgRNAs have previously been used for transcription regulation by fusing them to RNA aptamers that recruit transcription factors^{5,6} or to noncoding RNAs that target specific genomic loci⁷. We therefore reasoned that sgRNA scaffolds could be adapted to recruit fluorescent proteins for imaging. The sgRNA scaffold used for gene editing has previously been shown to be inefficient for DNA labeling and has been optimized by A-U flip and stem loop extension². Here we further optimized DNA labeling by replacing an A-U pair in the first stem loop of the sgRNA with G-C, which resulted in a substantial increase in the signal-to-noise ratio even without stem loop extension (Supplementary Fig. 1). All of the CRISPRainbow designs described below were built using this A-U to G-C mutation.

To establish a broad spectral range for multiloci labeling, we fused three hairpins, MS2, PP7, and boxB⁸, to stem loops or the 3'-end of the sgRNA in various combinations. The resulting sgRNAs each recruited a

pair of fluorescent proteins (FPs) that were fused to RNA hairpin binding domains recognizing the cognate RNA elements (Supplementary Fig. 2). These paired combinations generated three primary colors, blue, green, or red, plus three secondary colors, cyan, magenta, or yellow, with a triad of elements giving rise to white (Supplementary Table 1) by spectral overlapping from BFP, GFP, or RFP. The boxB/ λ N22 peptide pair previously used for RNA imaging⁸ showed inefficient DNA labeling in the CRISPRainbow system (Supplementary Fig. 3). However, an affinity-enhanced λ N22 peptide variant/boxB pair⁹ substantially increased the signal-to-noise ratio (Supplementary Fig. 3). The three primary colors were created by having two identical protein-binding RNA elements on the sgRNA (Fig. 1a). Live cell imaging in human osteosarcoma U2OS cells shows the expected labeling pattern for telomere-targeted sgRNAs (Supplementary Fig. 2). Dual-color labeling is achieved by combining different protein-binding elements that bind different combinations of fluorescent proteins on the same sgRNA (Fig. 1b). Live cell images again show the expected pattern of telomere labeling for each of the sgRNA variants. Telomere staining for the color white is shown in Figure 1c, in which the sgRNA carried all three primary colors.

To further evaluate the reliability of combinatorial colors in the CRISPRainbow system, we took z-stack images and captured signals for pairs of fluorescent proteins simultaneously for the secondary colors, and found that >97% of the telomere foci were labeled by dual color, and the brightness of distinct telomere foci were shown correspondingly in each color (Supplementary Fig. 4).

To confirm the specificity of CRISPRainbow in its primary color mode, we chose two or three distinct genomic loci (telomeres, the two aforementioned repeated sequences on chromosomes 3 and 9) with sgRNAs carrying either MS2, PP7, or boxB in pair-wise combinations, resulting in the expected dual-color or tri-color images (Supplementary Fig. 5).

We next explored the possibility of labeling four loci simultaneously using the three previous repeat regions and an additional repeat located in the subtelomeric region on the long arm of chromosome 13 (C13). Short sgRNAs with 11-mer sequences were chosen to target C3 and C13 here because they substantially increased the signal-to-noise ratio compared to 20-mer sequences for these two loci (Supplementary Fig. 6a–c). The telomeres and loci on C9 and C13 were readily labeled with three primary colors (blue, green, and red, respectively), whereas the locus on C3 was labeled by the fourth color, yellow, generated by the combination of green and red on that sgRNA (Supplementary Fig. 6d).

To demonstrate the color range of CRISPRainbow, we explored the possibility of simultaneously labeling six distinct targets. To identify the six chromosomal targets needed, a genome-wide survey of chromosome-specific repetitive DNAs was performed and thousands of sites were identified. We efficiently tested labeling at more than

¹Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts, USA. ²RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, Massachusetts, USA. ³Department of Computer Science, University of Central Florida, Orlando, Florida, USA. Correspondence should be addressed to H.M. (hanhui.ma@umassmed.edu).

Received 19 October 2015; accepted 4 March 2016; published online 18 April 2016; doi:10.1038/nbt.3526

100 chromosome-specific loci; finally, targets located on chromosomes 1, 3, 7, 13, 14, and X, respectively, were chosen for each CRISPRainbow color, i.e., the combination of three primary colors or three secondary colors (Supplementary Fig. 7). To coordinate the expression of sgRNAs in a single cell, we created a single plasmid to express all six sgRNAs (Fig. 2a). The location of each chromosomal site could be resolved by its expected CRISPRainbow color in living cells: blue for chromosome X (Cx), green for chromosome 14 (C14), red for chromosome 7 (C7), cyan for chromosome 1 (C1), yellow for chromosome 13 (C13), and magenta for chromosome 3 (C3) (Fig. 2b).

To interrogate the intranuclear dynamics of the four loci (C3, C9, C13, and telomeres) in living cells, we used time-lapse microscopy to simultaneously track their movement. The motion of different chromosomal loci varied both in range and direction (Fig. 2c, Supplementary Fig. 8 and Supplementary Movie 1). The range of movement was different even for different loci within a chromosome; for example, for C3-1 it was ~750 nm, and for C3-2 and C3-3, ~250 nm. Although previous studies have tracked the movements of single chromosomal sites, such as telomeres or integrated *lac* operator arrays¹⁰, CRISPRainbow offers an opportunity to do so for a number of endogenous chromosomal loci in the same cell. Moreover, although the tracking results shown here were deliberately obtained over a short interval (10 s), thus addressing dynamics on a short time scale, CRISPRainbow should, with sufficient control over photobleaching, enable longer-term observations of the three-dimensional intranuclear positioning¹¹ of various chromosomal sites during, for instance, progression through interphase, a program of cellular differentiation or chromosomal translocations.

Figure 1 A wide spectrum of colors generated by CRISPRainbow. (a) Primary colors for genomic DNA labeling. Two MS2 (top left), two PP7 (top middle) or two boxB (top right) elements were inserted into a human telomere-specific sgRNA to generate primary colors. Shown beneath each sgRNA is live cell labeling of telomeres in human U2OS cells following co-expression of dCas9, the indicated sgRNA, and the cognate RNA binding partners fused with different fluorescent proteins. (The overlay images are on the live cell phase-contrast micrographs in this and all other image figures in this paper.) Scale bar, 5 μ m. (b) Secondary colors. MS2 and PP7 (top left), PP7 and boxB (middle left) or boxB and MS2 (bottom left) were inserted into the sgRNA so as to generate cyan, yellow or magenta, respectively, by spectral overlapping of signals from a pair of cognate RNA binding partners fused with different fluorescent proteins. Images at the right of each secondary color design are the telomere labeling images obtained after co-expression of dCas9, the indicated sgRNA, and the cognate pair of fluorescent proteins. Scale bar, 5 μ m. (c) A tertiary 'color'. boxB, MS2, and PP7 were inserted into the sgRNA to generate white (left). Images at the right are telomere labeling following co-expression of dCas9, the triple element-bearing sgRNA, and the three cognate partners with distinct fluorescent proteins. Scale bar, 5 μ m. Data in all panels are representative of experiments performed at least three times.

Although earlier methods based on direct fusions of fluorescent proteins and Cas9 for live cell labeling of genomic sites supplanted transcription activator-like (TAL) effectors-base approaches¹², challenges remained for visualizing multiple genomic loci in live cells simultaneously, notwithstanding our introduction of a multicolor CRISPR system based on orthogonal Cas9s⁴. First, in that approach each Cas9 requires different PAM sequences¹³, which limits the range of target loci, plus the expression of the three Cas9s has to be balanced during multicolor labeling. Recently *Staphylococcus aureus* Cas9 was also engineered for CRISPR imaging¹⁴, which incrementally expands the multiple color range of the orthogonal Cas9-based system. By contrast, unlike orthogonal Cas9-based labeling, which requires the cognate sgRNAs for each Cas9, in CRISPRainbow a single dCas9 from *Streptococcus pyogenes* uses a number of different protein-binding sgRNAs. Thus, CRISPRainbow can be thought of as a 'spectral code', and the polychromatic range should be readily expandable, for example,

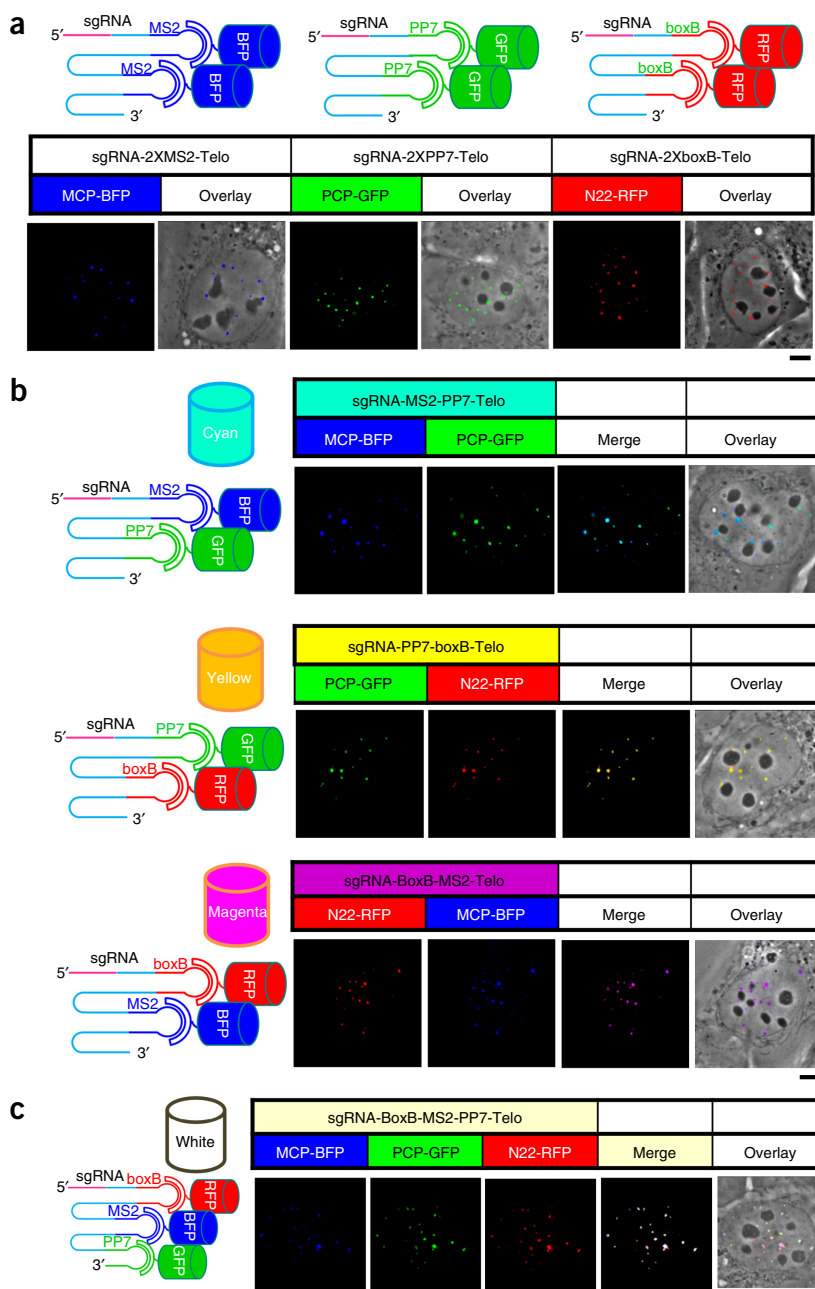
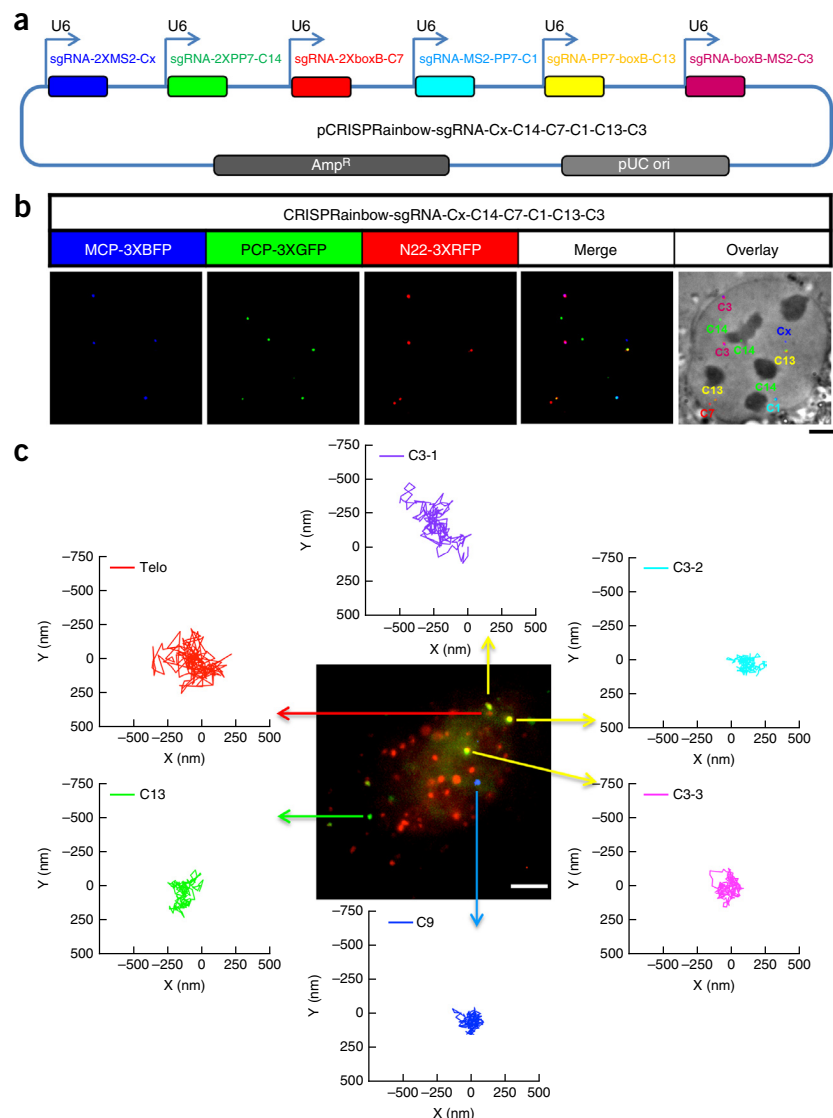


Figure 2 Localization and live tracking of multiple chromosome-specific loci simultaneously.

(a) Construct pCRISPRainbow-sgRNA-Cx-C14-C7-C1-C13-C3 for co-expression of six sgRNAs. sgRNA-2XMS2-Cx, sgRNA-2XPP7-C14, sgRNA-2XboxB-C7, sgRNA-MS2-PP7-C1, sgRNA-PP7-boxB-C3, and sgRNA-boxB-MS2-C3 were constructed individually and then subcloned into a single vector. (b) pCRISPRainbow-sgRNA-Cx-C14-C7-C1-C13-C3, dCas9, MCP-3XBFP, PCP-3XmNeonGreen, and N22-3XRFP were co-transfected into U2OS cells. Each CRISPRainbow color was dedicated to one chromosome locus: blue for chromosome X (Cx), green for chromosome 14 (C14), red for chromosome 7 (C7), cyan for chromosome 1 (C1), yellow for chromosome 13 (C13) and magenta for chromosome 3 (C3), respectively. Scale bar, 3 μ m. (c) Unique sites in chromosome 3, 9, and 13 as well as telomeres were labeled simultaneously using the CRISPRainbow colors blue, green, red, and yellow, respectively. The movements of these loci were recorded at 50 ms per frame for 10 s (200 total frames). All trajectories were shifted to start from the origin (0, 0) for easy comparison of the movement vectors. Scale bar, 5 μ m. Data in all panels are representative of experiments performed at least three times.



by use of a fourth RNA aptamer designed to be bound by, for example, a far-red fluorescent protein¹⁵. In principle, adding even one more color to CRISPRainbow would extend the simultaneous live-cell detection of genomic loci to 15.

A final point relates to our deliberate use of short sgRNA sequences, close to the length of the seed sequences¹⁶, for efficient labeling. This should make it possible to deploy a nuclease-active Cas9 for labeling owing to a lack of cleavage when using short sgRNAs^{17–19}. In such a format, one can envision a switchable CRISPR platform in which a live-cell genomic loci labeling mode with Cas9, instead of dCas9, is then redirected to gene editing by simply changing the expressed sgRNA to a form with a longer targeting sequence.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).

ACKNOWLEDGMENTS

We thank C. Smith in D. Grunwald laboratory for microscopy suggestions and S.A. Wolfe, M. Fatih Bolukbasi and A. Gupta of this institution for valuable discussion. This work was supported by the Vitold Arnett Fund to T.P. and NIH grants R01 GM102515 to S.Z. and U01 EB021238 to D.G.

AUTHOR CONTRIBUTIONS

H.M. and T.P. conceived the study and designed the experiments. H.M. and L.-C.T. performed experiments, added new innovations during the study and interpreted data. A.N. and S.Z. conducted bioinformatics mining of all the repeats in the human genome and identified the chromosome-specific repeats used as targets in this study. L.-C.T. and M.H. conducted image processing. H.M., D.G. and T.P. wrote the manuscript with the input of all the other authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Hsu, P.D., Lander, E.S. & Zhang, F. *Cell* **157**, 1262–1278 (2014).
- Chen, B. *et al. Cell* **155**, 1479–1491 (2013).
- Anton, T., Bultmann, S., Leonhardt, H. & Markaki, Y. *Nucleus* **5**, 163–172 (2014).
- Ma, H. *et al. Proc. Natl. Acad. Sci. USA* **112**, 3002–3007 (2015).
- Zalatan, J.G. *et al. Cell* **160**, 339–350 (2015).
- Konermann, S. *et al. Nature* **517**, 583–588 (2015).
- Shechner, D.M., Hacısuleyman, E., Younger, S.T. & Rinn, J.L. *Nat. Methods* **12**, 664–670 (2015).
- Daigle, N. & Ellenberg, J. *Nat. Methods* **4**, 633–636 (2007).
- Austin, R.J., Xia, T., Ren, J., Takahashi, T.T. & Roberts, R.W. *J. Am. Chem. Soc.* **124**, 10966–10967 (2002).
- Jegou, T. *et al. Mol. Biol. Cell* **20**, 2070–2082 (2009).
- Smith, C.S. *et al. J. Cell Biol.* **209**, 609–619 (2015).
- Ma, H., Reyes-Gutierrez, P. & Pederson, T. *Proc. Natl. Acad. Sci. USA* **110**, 21048–21053 (2013).
- Esvelt, K.M. *et al. Nat. Methods* **10**, 1116–1121 (2013).
- Chen, B. *et al. Nucleic Acids Res.* <http://dx.doi.org/10.1093/nar/gkv1533> (2016).
- Dean, K.M. & Palmer, A.E. *Nat. Chem. Biol.* **10**, 512–523 (2014).
- Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J.A. *Science* **348**, 1477–1481 (2015).
- Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. & Joung, J.K. *Nat. Biotechnol.* **32**, 279–284 (2014).
- Dahlman, J.E. *et al. Nat. Biotechnol.* **33**, 1159–1161 (2015).
- Kiani, S. *et al. Nat. Methods* **12**, 1051–1054 (2015).

ONLINE METHODS

Construction of CRISPRainbow plasmids. The expression vector for dCas9 (nuclease-dead), dCas9-mCherry and dCas9-GFP from *S. pyogenes* was described previously⁴. The sgRNA expression vector is based on the pLKO.1 lentiviral expression plasmid containing the *CcdB* gene between two BbsI sites for inserting guide sequences into the sgRNAs⁴. 2XMS2, 2XPP7, 2XboxB, MS2-PP7, PP7-boxB, boxB-MS2, and boxB-MS2-PP7 were inserted into the desired sgRNA (**Supplementary Fig. 2** and **Supplementary Table 1**) resulting, respectively, in pLH-sgRNA-2XMS2, pLH-sgRNA-2XPP7, pLH-sgRNA-2XboxB, pLH-sgRNA-MS2-PP7, pLH-sgRNA-PP7-boxB, pLH-sgRNA-boxB-MS2, and pLH-sgRNA-boxB-MS2-PP7. MCP-BFP or 3XBFP (mTagBFP2), PCP-GFP or 3XGFP (sfGFP), or PCP-mNeonGreen or PCP-3XmNeonGreen, N22-RFP or 3XRFP (mCherry) were subcloned into the pHAGE-TO-DEST or pHAGE-EFS-DEST vector⁴. The rapid guide RNA expression plasmid construction protocol was described previously⁴. For the assembly of six sgRNAs into a single plasmid, the Golden Gate cloning method was used, which had been used for assembly of repeating units in TALEs¹¹. All the plasmids reported here will be deposited at Addgene.

Cell culture and transfection. Human U2OS cells from ATCC were cultured on 35 mm glass bottom dishes (MatTek) at 37 °C in Dulbecco-modified Eagle's Minimum Essential Medium (DMEM; Life Technologies) supplemented with 10% (vol/vol) FBS. 150 ng of dCas9 plasmid DNA, 750 ng of total sgRNA plasmid DNA and 20 ng of one or more plasmid DNAs encoding MCP-FPs, PCP-FPs and/or N22-FPs in the combinations indicated were co-transfected using Lipofectamine 2000 (Life Technologies) and the cells were incubated for another 48 h before imaging.

Fluorescence microscopy. The microscope (Leica DMIRB) was equipped with an EMCCD camera (Andor iXon-897D), 100× oil objective lens (NA 1.4), 2× magnifier, and stage incubation chamber (held at 37 °C). The effective pixel size was 80 nm in image space. RFP (mCherry) was excited at 556/20 nm (wavelength/bandwidth), and its emission was collected in a 630/91 nm channel. GFP (sfGFP) was excited at 470/28 nm, and its emission was collected in a 512/23 nm channel. BFP (mTagBFP2) was excited at 387/11 nm, and its emission collected using a 464/23 nm filter. Images were acquired sequentially and analyzed with MetaMorph software (Molecular Devices). Image size was adjusted to show individual nuclei, intensity thresholds were set manually to display only nuclear focal signals and not background nucleoplasmic fluorescence. Only maximum and minimum levels were adjusted, not contrast and brightness. Composite images were generated by overlaying the individual color images without image registration. For z-stack and live cell images, a custom-built super-registration microscope equipped with 4 lasers

(Coherent) and 4 EMCCD cameras (Andor), 150X oil Apochromat objective lens (Olympus, NA 1.45), 37 °C heater and a chamber with CO₂/air gas mixer (ibidi) was used. The registration images were recorded before experiments using a Geller MRS-4 (Geller MicroAnalytical Laboratory). The RFP was excited by 561 nm laser and the emission was collected through a 615/45 nm filter and EMCCD camera (Andor iXon3-897E). The mNeonGreen was excited by 515 nm laser and the emission was collected through a 534/20 filter and EMCCD camera (Andor iXonUltra-897U). The BFP was excited by 405 nm laser and the emission was collected through a 460/60 nm filter and EMCCD camera (Andor iXon-897D). Images from different colors were acquired simultaneously and recorded in 20 fps, registered by using Matlab software and tracked by using the TrackMate function in Fiji software. Prior to registration, the raw camera data from each channel were subjected to background subtraction. The latter was achieved by calculating the average background intensity value for each pixel using 100 dark frames and subsequently subtracting the average pixel background values from the each frame of the raw experimental data. To achieve subpixel registration accuracy, parameters for shifting, scaling, and rotating camera images were determined by iterative nonlinear (least-squares) curve fitting of Geller images for each slave camera to a master camera target image. Using these parameters, the experimental data from each channel (except from the master camera) were processed through an affine transformation and overlapped in false-color channels for visualization. The image quantification and tracking were performed on raw data after image registration without further image processing. The video in **Supplementary Movie 1** was generated by using Metamorph software. A 3 × 3 kernel was applied for better visibility and the video play rate is 30 fps.

Chromosome-specific repeats. The mining of chromosome-specific repeats was described previously⁴. The chromosome-1-specific locus is situated in the subtelomeric region p36, having the identifier Chr 1: 2581275-2634211 in the human reference genome hg19 at the UCSC genome browser (<http://genome.ucsc.edu>). The chromosome-3-specific locus is situated in the subtelomeric region q29, having the identifier Chr 3: 195199022-195233876. The chromosome-7-specific locus is situated in the subtelomeric region q36, having the identifier Chr 7: 158122661-158135328. The chromosome-13-specific locus is situated in the subtelomeric region q34, having the identifier Chr 13: 112930173-112968847. The chromosome-14-specific locus is situated in the subtelomeric region q32, having the identifier Chr 14: 105695963-105707283. The chromosome-X-specific locus is situated in the subtelomeric region p22, having the identifier Chr X: 76096-94831. The chromosome-9-specific locus is a tandem array of GGAAT repeats that are highly concentrated in the pericentromeric region. The details of all sgRNA sequences are listed in **Supplementary Table 2**.