Supplemental information

Centromeres as universal hotspots of DNA breakage, driving RAD51-mediated recombination during quiescence

Xanita Saayman, Emily Graham, William Nathan, Andre Nussenzweig, and Fumiko Esashi

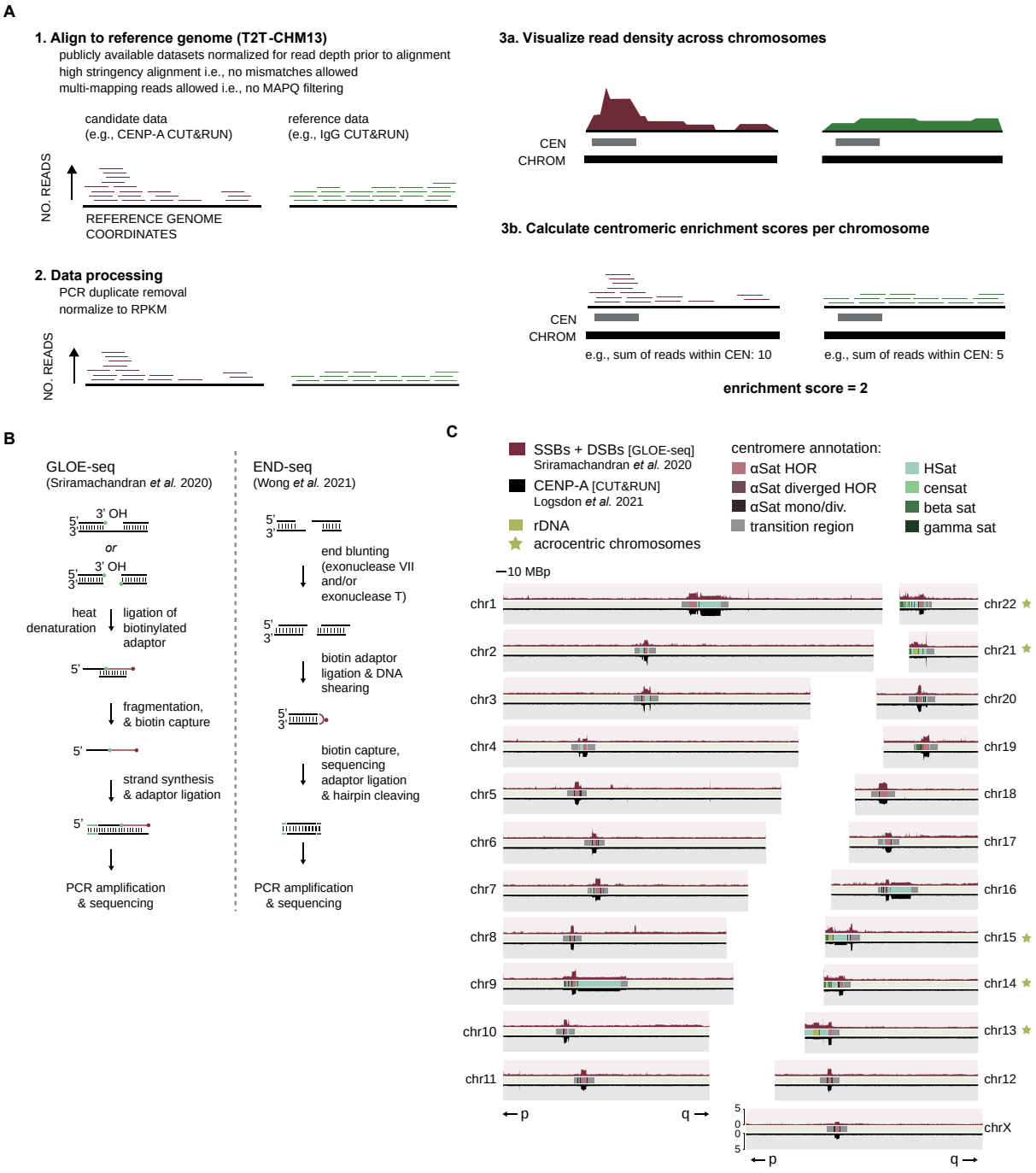

Figure S1 | Next generation sequencing (NGS)-based detection of DNA breaks at centromeres. Related to Figure 1.

(A) Modified computational pipeline for the alignment of short-read sequencing datasets to the T2T-CHM13 reference genome. First, publicly available datasets analysed in this study were trimmed to a consistent read depth after adaptor trimming. Alignments to T2T-CHM13 were then performed with high stringency, allowing no mismatches. Traditional MAPQ filtering was circumvented to retain multi-mapping reads, which were only reported once. PCR duplicates were removed and read scores were normalized to reads per thousand million (RPKM). Following normalization, read scores were visualized or further processed to calculate centromeric enrichment scores. **(B)** Schematic depicting the experimental pipelines developed for GLOE-seq[S1] for the detection of both DSBs and SSBs through free 3'OH and END-seq[S2] for the detection of DSBs specifically. **(C)** Alignment of publicly available next generation sequencing data detecting DNA strand breaks (GLOE-seq[S1]) above and CENP-A localization (CUT&RUN[S3]) below across 22 autosome chromosomes and chromosome X of

the human T2T-CHM13 reference assembly. Centromeric regions are composed of alpha satellites within higher order repeats (αSat HORs), alpha satellites within diverged HORs (αSat diverged HORs), monomeric / diverged alpha satellites (αSat mono/div.), human satellites (HSat), beta satellites (beta sat), gamma satellites (gamma sat), transition regions, other satellites (censat) and ribosomal DNA arrays (rDNA). p and q chromosome arms indicated below.

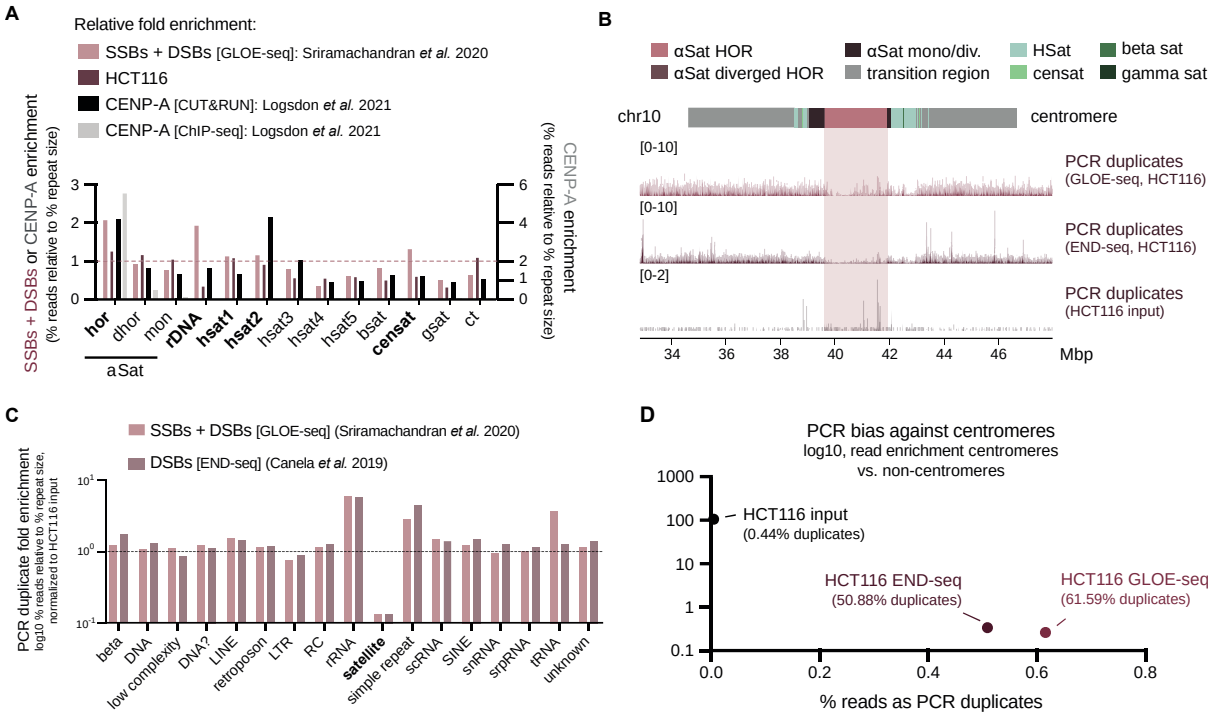

Figure S2 | DNA break accumulation across repeat sequences genome-wide. Related to Figure 1.

(A) Fold enrichment of DNA breaks (GLOE-seq[S1]), HCT116 and CENP-A (CUT&RUN as well as ChIP-seq[S3]) reads across the main centromere sub-categories (excluding tar repeats) as annotated in the T2T-CHM13 reference genome. **(B)** Inset of chr10 centromere depicting PCR duplicates mapped in GLOE-seq[S1], END-seq[S4] and the corresponding input cell line (HCT116 input). After marking duplicates with samtools markdup, reads with SAM flags of 1024 were extracted for visualization and further processing. **(C)** Fold enrichment of PCR duplicates in GLOE-seq and END-seq across all annotated repeat types of the T2T-CHM13 reference sequence. Fold enrichment was calculated as in (D), and enrichment scores were then further normalized to those of the HCT116 input. **(D)** Correlation between the proportion of datasets identified as PCR duplicates and the PCR bias against centromeres. PCR bias was calculated as the proportion of PCR duplicate reads within centromeres relative to the proportion of PCR duplicate reads outside of centromeres.

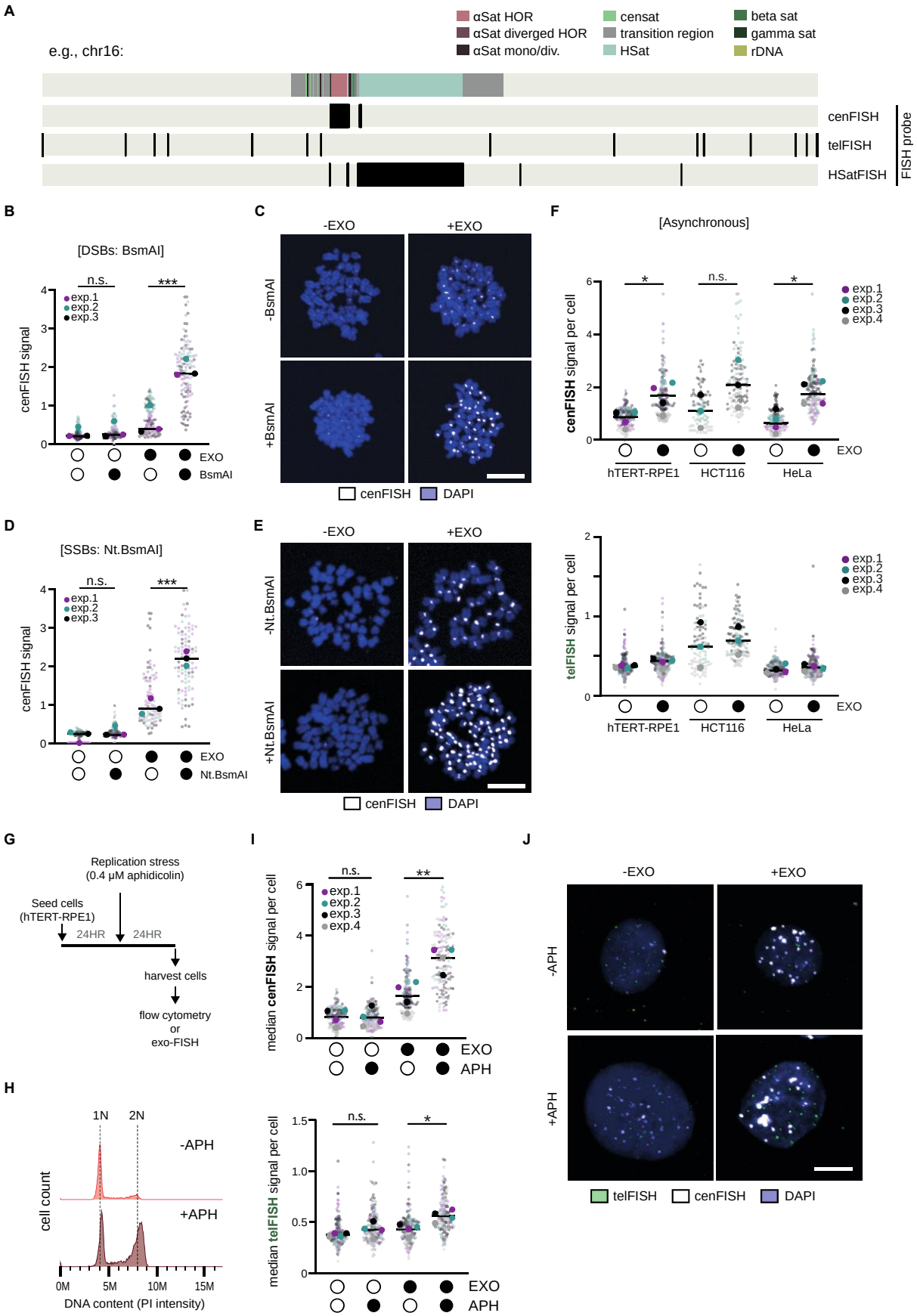
A
αSat HOR
αSat diverged HOR
αSat mono/div.
censat
transition region
HSat
beta sat
gamma sat
rDNA
e.g., chr16:
cenFISH
telFISH
HSatFISH
FISH probe
B
[DSBs: BsmAI]
n.s.
***
exp.1
exp.2
exp.3
cenFISH signal
EXO
BsmAI
C
-EXO
+EXO
-BsmAI
+BsmAI
cenFISH
DAPI
F
[Asynchronous]
*
n.s.
*
exp.1
exp.2
exp.3
exp.4
cenFISH signal per cell
EXO
hTERT-RPE1
HCT116
HeLa
D
[SSBs: Nt.BsmAI]
n.s.
***
exp.1
exp.2
exp.3
cenFISH signal
EXO
Nt.BsmAI
E
-EXO
+EXO
-Nt.BsmAI
+Nt.BsmAI
cenFISH
DAPI
telFISH signal per cell
exp.1
exp.2
exp.3
exp.4
EXO
hTERT-RPE1
HCT116
HeLa
G
Replication stress
(0.4 μM aphidicolin)
Seed cells
(hTERT-RPE1)
24HR
24HR
harvest cells
flow cytometry
or
exo-FISH
H
1N
2N
-APH
+APH
cell count
0M
5M
10M
15M
DNA content (PI intensity)
I
n.s.
**
exp.1
exp.2
exp.3
exp.4
median cenFISH signal per cell
EXO
APH
n.s.
*
median telFISH signal per cell
EXO
APH
J
-EXO
+EXO
-APH
+APH
telFISH
cenFISH
DAPI

Figure S3 | Validation of centromeric DNA break detection using exo-FISH in hTERT-RPE1 cells. Related to Figure 2.

(A) FISH probe hybridization: alignment of FISH probe complementary sequences to chromosome 16 of T2T-CHM13. **(B-E)** Quantification and representative images of DNA break detection by exo-FISH. hTERT-RPE1 cells were arrested in mitosis by STLC, harvested and fixed. Double-strand breaks (DSBs) or single-strand breaks (SSBs) were induced by restriction enzyme digestion using BsmAI or Nt.BsmAI, respectively. **(F)** exo-FISH applied to compare ExoIII (EXO) responsiveness of hTERT-RPE1, HCT116 and HeLa cell lines. **(G)** Experimental schematic for addressing the impact of replication stress (0.4 μM aphidicolin; APH for 24 hours) on spontaneous centromere breaks. **(H)** Flow cytometry verification of replication stress induction by measuring DNA content by propidium iodide (PI) staining on fixed cells, 24 hours after replication stress induction. **(I, J)** Representative images and quantification of exo-FISH following 24 hours replication stress. Filled and empty circles indicate presence and absence, respectively. Quantification of FISH signals were done as described in Figure 2. Scale bar represents 10 μm. FISH signal intensity is X10,000 arbitrary units (A.U.). At least 30 cells were imaged per experimental condition. The medians of each experimental condition were used to perform a two-sided unpaired t-test (*$p<0.05$, **$p<0.01$).

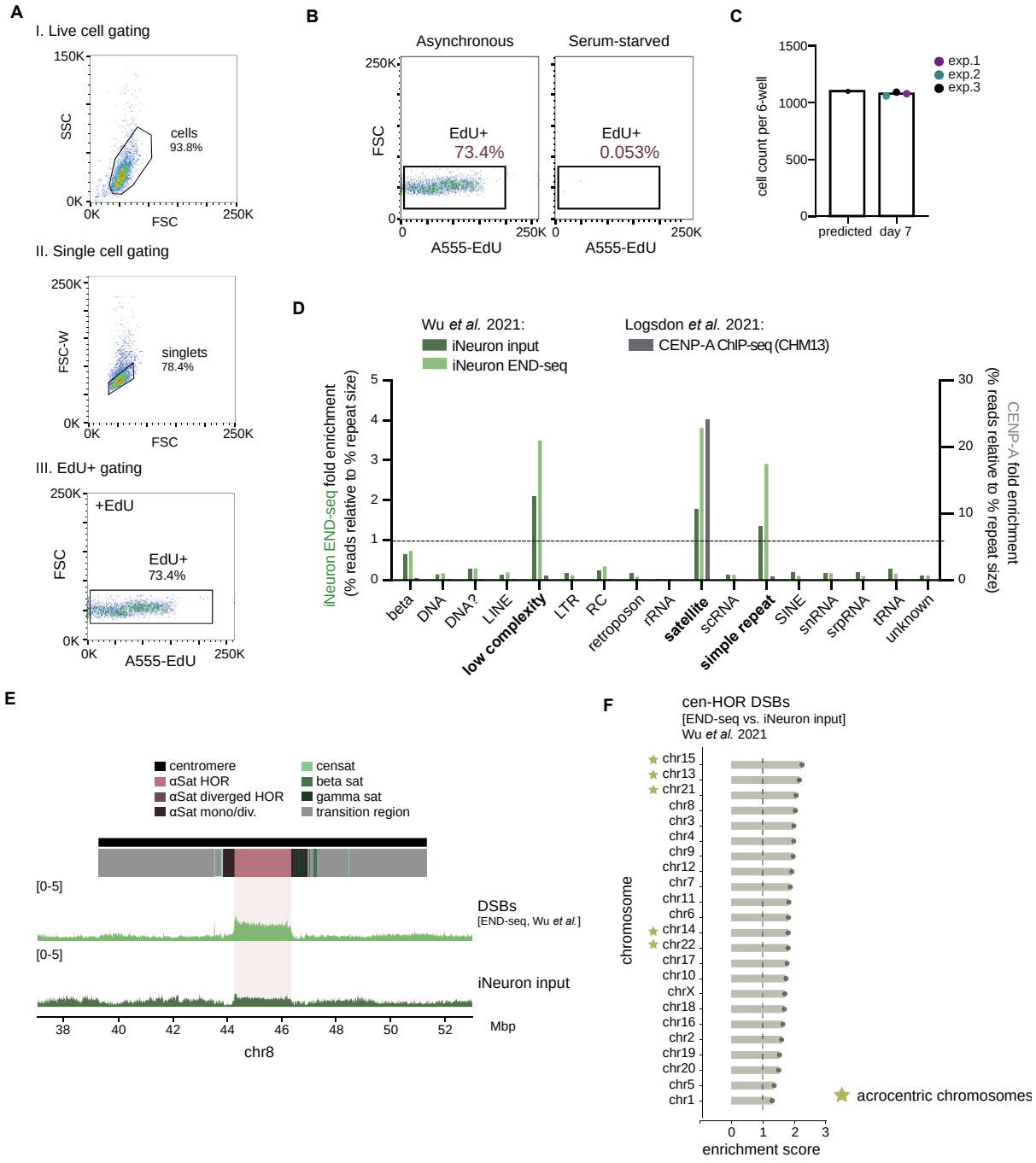

Figure S4 | DNA break accumulation at centromere in post-mitotic iNeurons. Related to Figure 3.

(A) Gating strategy for the selection of single, live cells. First, live cell selection was performed by forward-scattered light (FSC) and side-scattered light (SSC) gating. Second, single cells were selected by gating the width of forward-scatter light (FSC-W) by FSC. Finally, EdU+ populations were determined based on positive controls. **(B)** Validation of serum starvation preventing cell proliferation. hTERT-RPE1 were incubated with serum-starved media (0.1% FBS) for 24 hours and then incubated with EdU for another 24 hours, as proliferating cells will incorporate EdU into DNA during replication. Cells were then fixed and analysed by flow cytometry. Gating to define the EdU+ population was done as described in (A). Forward scatter (FSC) is an indicator of cell size. **(C)** Cell count per 6-well plate after 7 days of quiescence as an indicator of cell proliferation. Predicted cell count was calculated as twice the number of seeded cells (550K cells), as cells were left to grow 24 hours prior to serum starvation. **(D)** Fold enrichment of iNeuron input and END-seq reads[S5] and CENP-A ChIP-seq reads[S3] across all annotated repeat types of the T2T-CHM13 reference sequence. Fold enrichment was calculated as in Figure 1.

(E) Alignment of short-read sequencing data on chromosome 8 detecting DNA DSBs (END-seq) in iNeuron cells[S5]. Inset of chr8 centromere depicts the alignment of END-seq reads with the corresponding input cell line. **(F)** Enrichment scores following END-seq in post-mitotic neuronal cells across human centromere HORs. Enrichment scores were calculated as in Figure 1. Chromosomes are ranked by descending enrichment score.

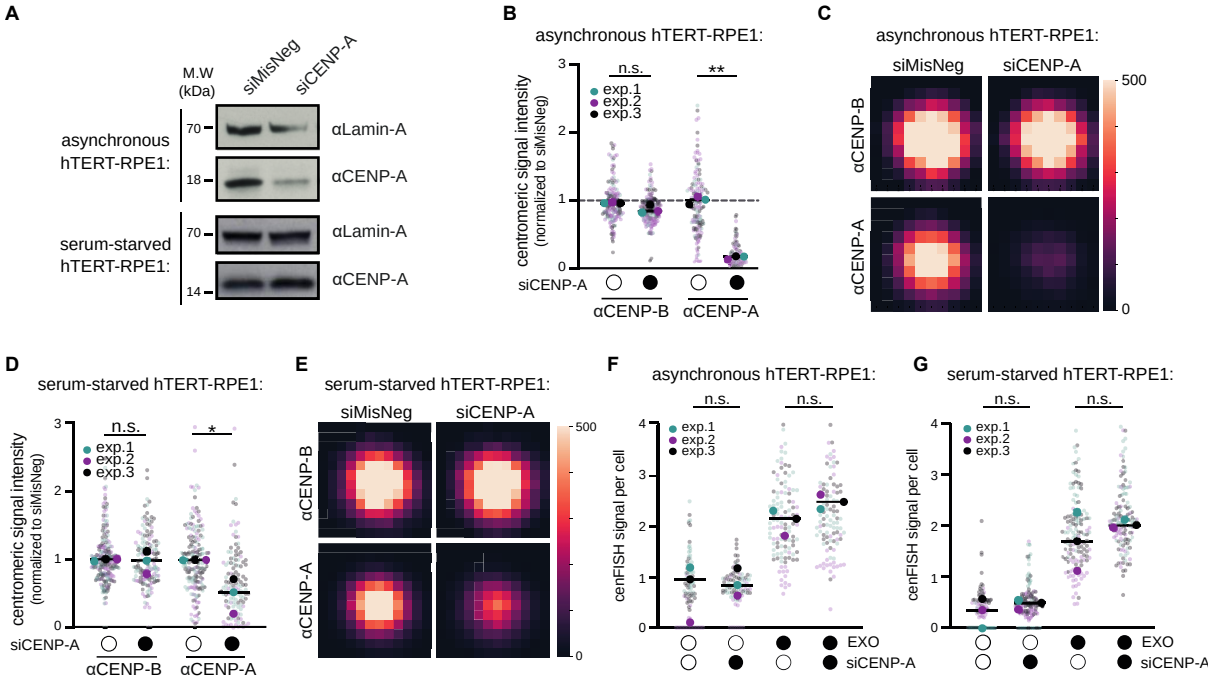

Figure S5 | Investigating the source of centromere HOR DNA breakage in quiescent cells. Related to Figure 4.

(A) Western blot assessing the depletion of CENP-A in asynchronous and serum-starved hTERT-RPE1 cells, 96 hours after siRNA treatment with siMisNeg or siCENP-A. Cells were serum-starved for 24 hours before RNAi treatment. **(B, C)** Quantification and representative median foci of centromeric CENP-A and CENP-B levels in asynchronous hTERT-RPE1 cells. CENP-B foci were used to determine centromere positioning, after which CENP-B and CENP-A fluorescence intensities were measured and quantified as with FISH, described in Figure 2. IF signal intensities were normalized to siMisNeg. **(D, E)** Representative median foci as in (D, E), but for quiescent hTERT-RPE1 cells. **(F)** exo-FISH quantification of asynchronous CENP-A-depleted hTERT-RPE1 cells, with the same experimental design as in (B, C). **(G)** exo-FISH quantification of quiescent CENP-A-depleted hTERT-RPE1 cells, with the same experimental design as in (D, E). Quantification of all exo-FISH experiments were done as described in Figure 2. Scale bar represents 10 μm. FISH signal intensity is X10,000 arbitrary units (A.U.). Each data point represents a median focus intensity value of a cell, with the median of each biological replicate indicated in darker hues. For both immunofluorescence and exo-FISH, at least 30 cells were imaged per experimental condition. The medians of each experimental condition were used to perform a two-sided unpaired t-test (*p<0.05, **p<0.01). Filled and empty circles indicate presence and absence, respectively.

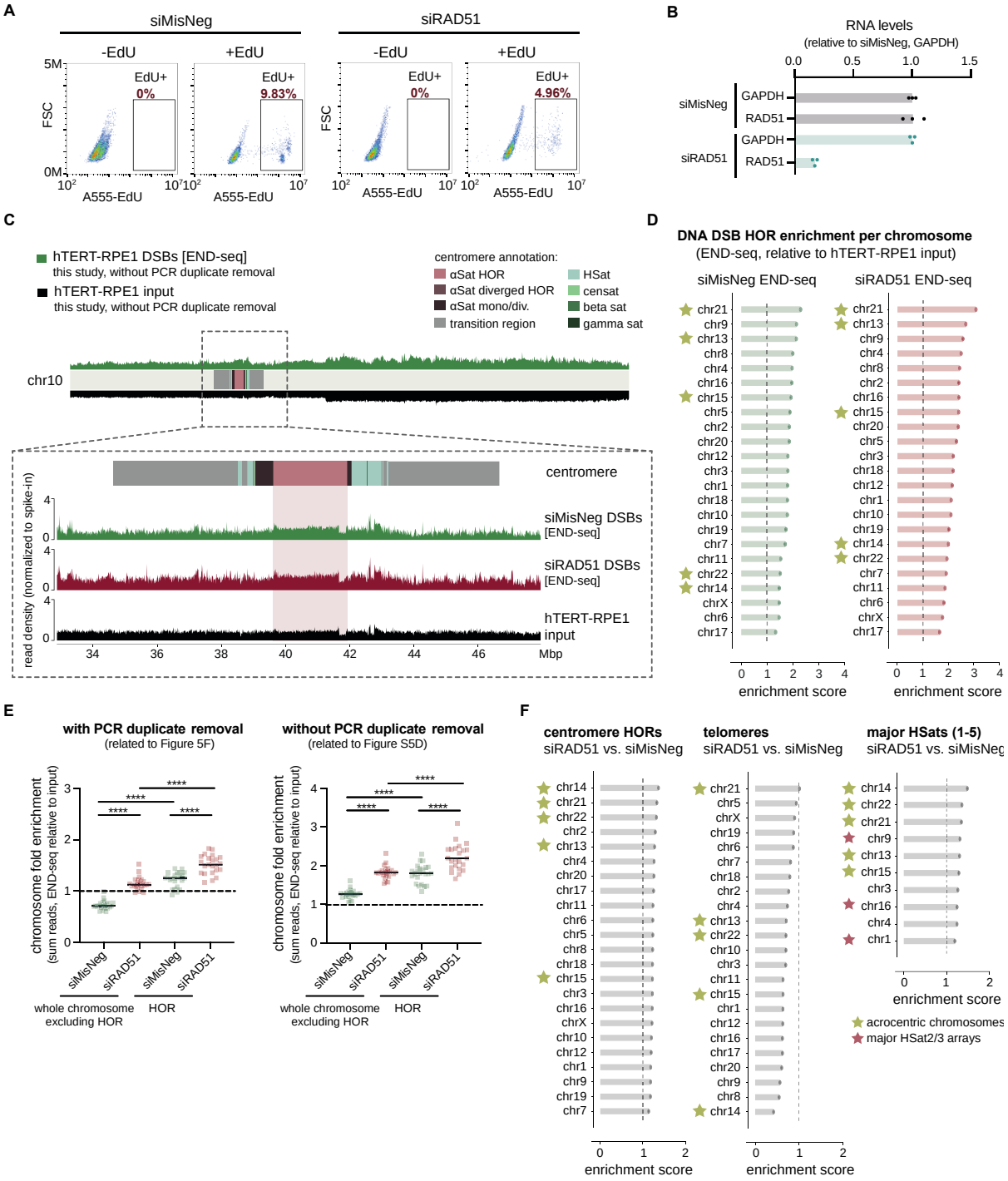

Figure S6 | RAD51 impact on spontaneous centromere DNA breaks. Related to Figure 5.

(A) Impact of RAD51 depletion on cellular proliferation in serum-starved hTERT-RPE1 cells. Cells were incubated with serum-starved media (0.1% FBS) for 24 hours and then treated with siRNA targeting either RAD51 or a universal negative control (siMisNeg). Cells were then incubated for an additional 72 hours in the presence of EdU in serum-starved media, fixed and analysed by flow cytometry to determine what proportion of the population had incorporated EdU. Gating was performed as in Figure S4A and B. Forward scatter (FSC) is an indicator of cell size. **(B)** RT-qPCR of RNA extracted from serum-starved hTERT-RPE1 cells following 96 hours depletion of RAD51 (siRAD51) or a negative control (siMisNeg). **(C)** Inset of chr10 centromere depicting DNA DSBs (END-seq) following 96 hours depletion of RAD51 or the negative control in serum-starved hTERT-RPE1 cells. NGS alignment was performed without PCR duplicate removal and with spike-in normalisation. **(D)**

Enrichment END-seq scores across hTERT-RPE1 centromere HORs. NGS alignment was performed without PCR duplicate removal and with spike-in normalisation, and enrichment scores quantified as in Figure 1. **(E)** Comparison of chromosome enrichment scores for HORs and whole chromosomes excluding HORs, with and without PCR duplicate removal. The dashed line indicates theoretical non-enrichment. To determine statistical significance, a two-sided unpaired t-test was performed (****$p<0.0001$). **(F)** Enrichment scores comparing siRAD51 vs. siMisNeg in serum-starved hTERT-RPE1 cells across centromere HORs, telomeres and major HSat arrays (chr1, 3, 4, 9, 13, 14, 15, 16, 21 and 22). Enrichment scores were calculated as the sum of siRAD51 reads compared to siMisNeg reads across all HOR / telomere / HSat regions of the indicated chromosome, following PCR duplicate removal and spike-in normalisation.

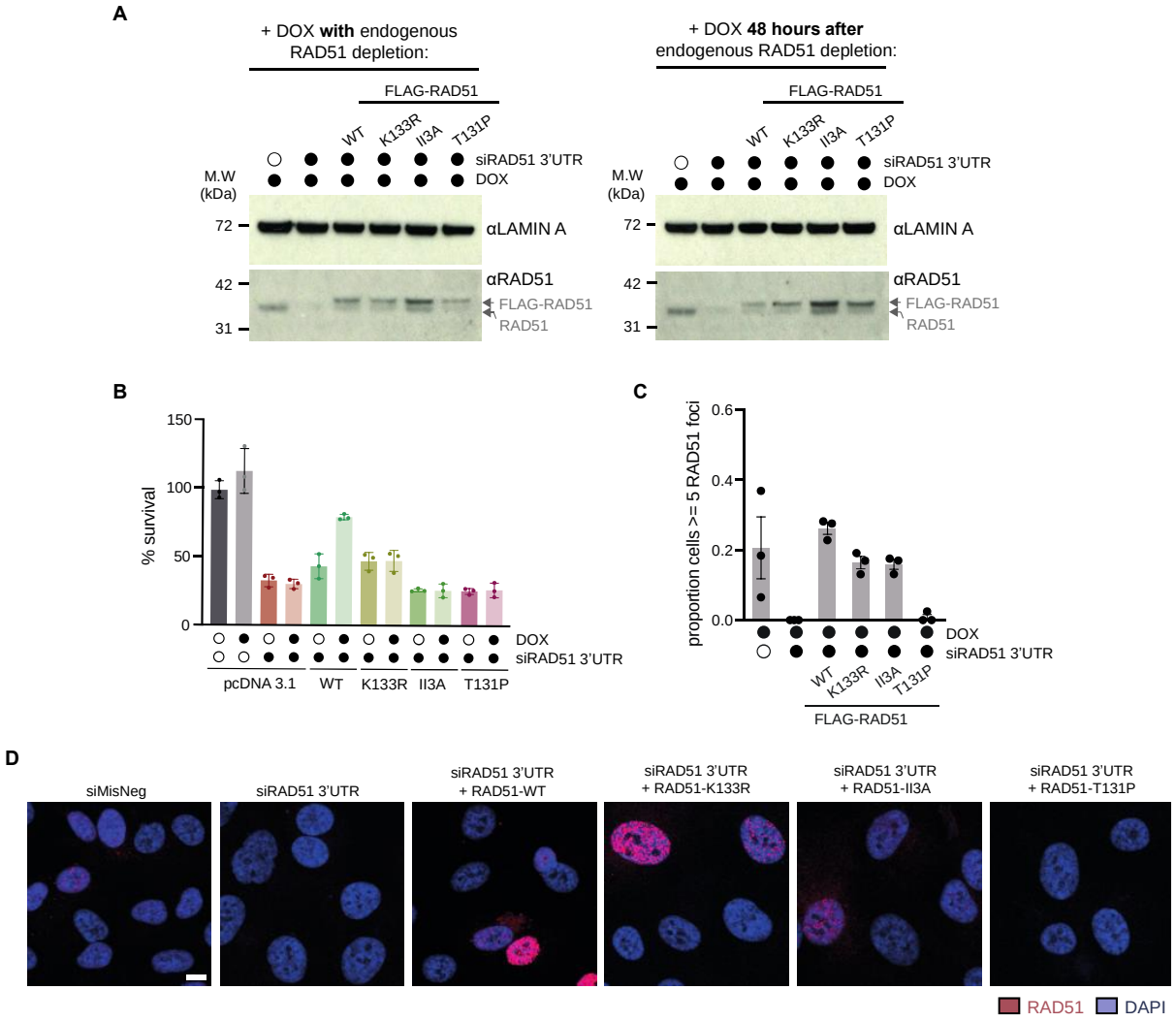

Figure S7 | Validation of RAD51 separation-of-function variants. Related to Figure 6.

(A) RAD51 separation-of-function variant expression, either with doxycycline (DOX) induction at the same time as siRNA treatment, or 48 hours after siRNA treatment. WB samples were harvested 72 hours after siRNA treatment. Lamin-A levels were used as a loading control. **(B)** WST-1 assay measuring cell survival after 120 hours endogenous RAD51 siRNA-mediated depletion alongside expression of the RAD51 separation-of-function variants, as induced by DOX treatment. **(C, D)** Quantification and representative images of RAD51 foci formation in asynchronous hTERT-RPE1 cells. Cells were treated with siRNA targeting either endogenous RAD51 or the universal negative control, followed by expression of RAD51 separation-of-function variants by doxycycline. Scale bar represents 10 μM. Filled and empty circles indicate presence and absence, respectively.

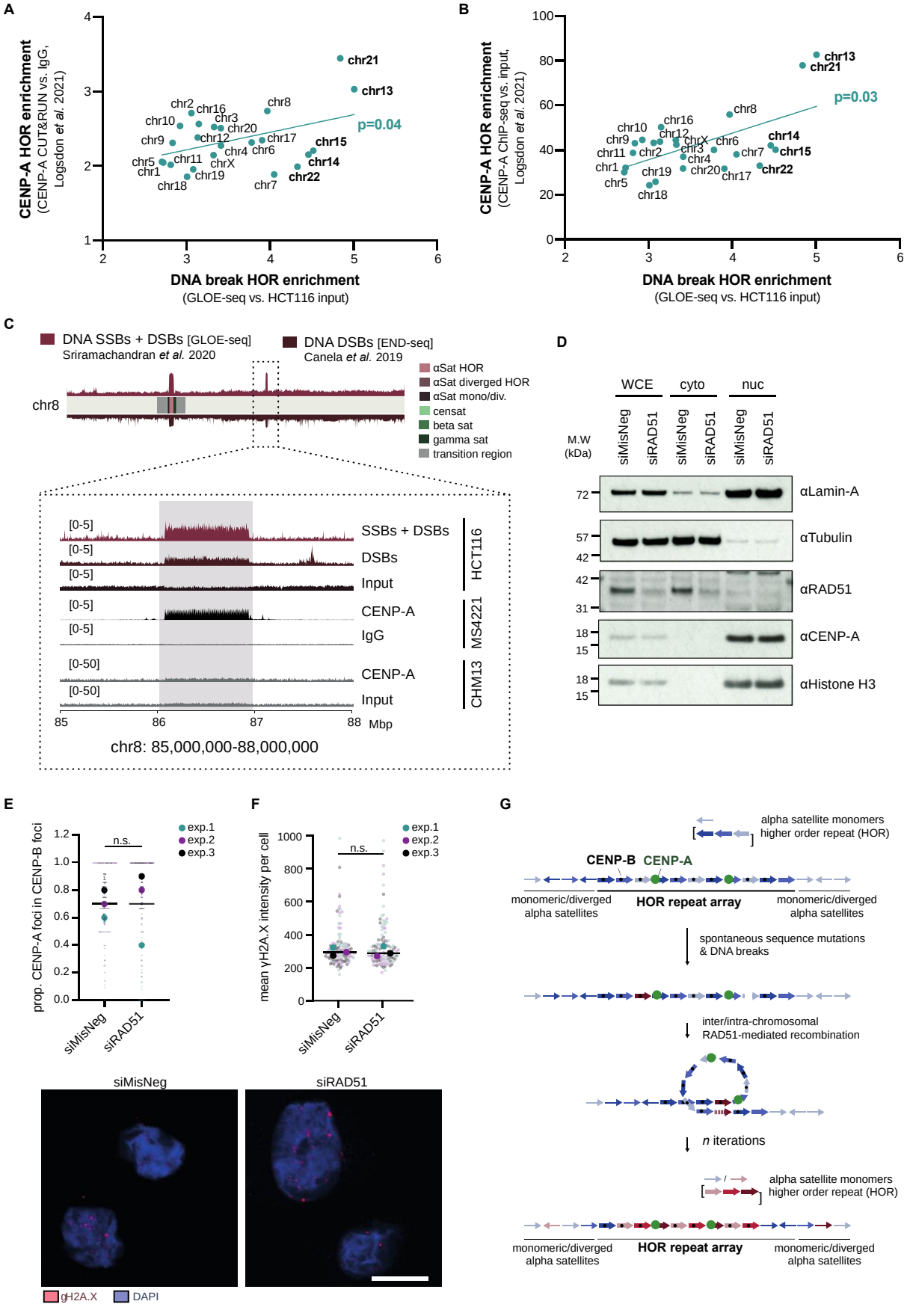
A
CENP-A HOR enrichment
(CENP-A CUT&RUN vs. IgG, Logsdon et al. 2021)
DNA break HOR enrichment
(GLOE-seq vs. HCT116 input)
p=0.04
chr21
chr13
chr15
chr14
chr22
B
CENP-A HOR enrichment
(CENP-A ChIP-seq vs. input, Logsdon et al. 2021)
DNA break HOR enrichment
(GLOE-seq vs. HCT116 input)
p=0.03
C
DNA SSBs + DSBs [GLOE-seq]
Sriramachandran et al. 2020
DNA DSBs [END-seq]
Canela et al. 2019
αSat HOR
αSat diverged HOR
αSat mono/div.
censat
beta sat
gamma sat
transition region
chr8
SSBs + DSBs
DSBs
Input
CENP-A
IgG
CENP-A
Input
HCT116
MS4221
CHM13
Mbp
chr8: 85,000,000-88,000,000
D
WCE
cyto
nuc
M.W (kDa)
siMisNeg
siRAD51
αLamin-A
αTubulin
αRAD51
αCENP-A
αHistone H3
E
prop. CENP-A foci in CENP-B foci
n.s.
exp.1
exp.2
exp.3
siMisNeg
siRAD51
F
mean γH2A.X intensity per cell
n.s.
G
alpha satellite monomers
higher order repeat (HOR)
CENP-B
CENP-A
monomeric/diverged alpha satellites
HOR repeat array
spontaneous sequence mutations & DNA breaks
inter/intra-chromosomal RAD51-mediated recombination
n iterations

Figure S8 | A dual function for RAD51 at human centromeres. Related to Figure 7.

(A, B) Correlation between centromeric enrichments of DNA strand breaks (GLOE-seq[S1]) in HCT116 and CENP-A (CUT&RUN (A) or ChIP-seq (B))[S3] using T2T-CHM13 as a reference genome. Enrichment scores were calculated as in Figure 1, as the sums of mapped reads within centromeric HORs of each chromosome relative to that of the corresponding input control, after normalizing to read depth. Acrocentric chromosomes containing rDNA arrays are marked in bold. p-value was determined using simple linear regression analysis. **(C)** Alignment of short-read sequencing data on chromosome 8 detecting both DNA SSBs and DSBs (GLOE-seq[S1]) or just DNA DSBs (END-seq[S4]) in HCT116 cells. Inset of chr8 depicts the alignment of GLOE-seq and END-seq, as well as CENP-A mapped in CHM13 (CUT&RUN)[S3] or the 8q21 neocentromere cell line MS4221[S6], with the corresponding input or negative controls. **(D)** Cellular fractionation of serum-starved hTERT-RPE1 cells following 96 hours siRNA-mediated depletion of RAD51. **(E)** Quantification of the proportion of CENP-A foci colocalizing with CENP-B foci upon siRNA-mediated depletion of RAD51 in quiescent hTERT-RPE1 cells. Colocalization of foci was counted if the centre of CENP-A foci was within 5 pixels of the centre of CENP-B foci. **(F)** Quantification and representative images of total nuclear γH2A.X intensity in quiescent hTERT-RPE1 cells following 96-hours RNAi-mediated depletion of RAD51. At least 30 cells were imaged per experimental condition. Each data point represents a median value of a cell, with the median of each biological replicate indicated in darker hues. The medians of each experimental condition were used to perform a two-sided unpaired t-test. **(G)** Model for the proposed role for RAD51-mediated recombination in the homogenization of centromeric satellite arrays.

Table S1. Publicly available NGS datasets analyzed in this study. Related to Short-read sequencing analysis (STAR Methods)

Dataset	Source	SRA
HCT116 GLOE-seq	Sriramachandran *et al.*[S1]	SRR9676440
HCT116 END-seq	Canela *et al.*[S4]	SRR8870099
HCT116 input	ENCODE	SRR577511
CHM13 CUT&RUN CENP-A	Logsdon *et al.*[S3]	SRR15395852
CHM13 CUT&RUN IgG	Logsdon *et al.*[S3]	SRR15395848
CHM13 ChIP-seq CENP-A	Logsdon *et al.*[S3]	SRR13278683
CHM13 ChIP-seq input	Logsdon *et al.*[S3]	SRR13278681
iNeuron END-seq	Wu *et al.*[S5]	SRR13764826
iNeuron input	Wu *et al.*[S5]	SRR13764817
MS4221 CENP-A ChIP-seq	Hasson *et al.*[S6]	SRR766738
MS4221 input ChIP-seq	Hasson *et al.*[S6]	SRR766741

Table S2. Mutagenesis primers for construction of RAD51-II3A. Related to Plasmids (STAR Methods)

Target	Construction	Sequence
R130A_F	pDEST-RAD51-II3A	5’ AGAAATGTTTGGAGAATTCGCAACTGGGAAGACCCAGATC 3’
R130A_R	pDEST-RAD51-II3A	5’ GATCTGGGTCTTCCCAGTTGCGAATTCTCCAAACATTTCT 3’
R303A_F	pDEST-RAD51-II3A	5’ AACAACCAGATTGTATCTGGCGAAAGGAAGAGGGGAAACC 3’
R303A_R	pDEST-RAD51-II3A	5’ GGTTTCCCCTCTTCCTTTCGCCAGATACAATCTGGTTGTT 3’
K131A_F	pDEST-RAD51-II3A	5’ GGGGAAACCAGAATCTGCGCAATCTACGACTCTCCCTG 3’
K131A_R	pDEST-RAD51-II3A	5’ CAGGGAGAGTCGTAGATTGCGCAGATTCTGGTTTCCCC 3’

Table S3. siRNA sequences. Related to RNAi depletion (STAR Methods).

Target	Final Conc.	Sequence
siRAD51 #1	20 nmol	5’ GACUGCCAGGAUAAAGCUU 3’
siRAD51 #2	20 nmol	5’ GUGCUGCAGCCUAAUGAGA 3’
siRAD51 3’ UTR #1	20 nmol	5’ GACUGCCAGGAUAAAGCUU 3’
siRAD51 3’ UTR #2	20 nmol	5’ GUGCUGCAGCCUAAUGAGA 3’
siCENP-A	20 nmol	5’ GGACUCUCCAGAGCCAUGAUU 3’

Table S4. Primary Antibodies. Related to Western blotting & chromatin fractionation and Immunofluorescence (IF) (STAR Methods).

Target	Dilution	Catalog No.	Source	RRID
RAD51	1:2000	ab176458	Abcam	AB_2665405
RAD51	1:2000	-	homemade (7946)	-

Lamin-A	1:2000-1:20000	L1293	Sigma-Aldrich	AB_532254
CENP-B	1:1000	IHC-00064	Bethyl Laboratories	AB_669682
CENP-A	1:1000	ab13939	Abcam	AB_300766
CENP-A	1:1000	GTX13939	GeneTex	AB_369391
TOP1	1:1000	ab109374	Abcam	AB_10861978
TOP2A (D10G9)	1:1000	12286	Cell Signaling Technology	AB_2797871
TOP2B	1:1000	GTX102640-GTX-25ul	GeneTex	AB_11169314
TOP3A	1:1000	14525-1-AP	Proteintech	AB_2205881
TOP3B	1:1000	ORB127293	Biorbyt	N/A
Tubulin	1:2000	3873	Cell Signaling Technology	AB_1904178
Histone H3	1:2000	A300-823A-T	Bethyl Laboratories	AB_2118462

Table S5. Secondary Antibodies. Related to Western blotting & chromatin fractionation and Immunofluorescence (IF) (STAR Methods).

Target	**Dilution**	**Catalog No.**	**Source**	**RRID**
anti-rabbit Alexa Fluor 488	1:1000	A11070	Thermo Fisher Scientific	AB_142134
anti-rabbit Alexa Fluor 555	1:1000	A21430	Thermo Fisher Scientific	AB_1500773
anti-rabbit Alexa 647	1:1000	A27040	Thermo Fisher Scientific	AB_2536101
anti-mouse Alexa Fluor 488	1:1000	A11017	Thermo Fisher Scientific	AB_143160
anti-mouse Alexa Fluor 555	1:1000	A21425	Thermo Fisher Scientific	AB_1500751
anti-mouse Alexa 647	1:1000	A21237	Thermo Fisher Scientific	AB_1500743
anti-mouse HRP-conjugated	1:1000	P0447	Agilent	AB_2617137
anti-rabbit HRP-conjugated	1:2000	P0448	Agilent	AB_2617138

Table S6. FISH probe sequences. Related to exo-FISH (STAR Methods).

Target	**Fluorophore**	**Sequence**	**Source**
HOR	Cy3	5’ ATTCGTTGGAAACGGGA 3’	F3009, PNABio (CENPB-RC-Cy3)
telo	Cy5	5’ CCCTAACCCTAACCCTAA 3’	F3009, PNABio (TelC-Cy5)
HSat2	A488	5’ TCGAGTCCATTCGATGAT 3’	Custom, PNABio
HSat3	A488	5’ TCCACTCGGGTTGATT 3’	Custom, PNABio

Table S7. RT-qPCR Primers. Related to RT-qPCR (STAR Methods).

Target	**Input**	**Forward sequence**	**Reverse sequence**	**Source**
RAD51	100 ng	TCTCTGGCAGTGATGTCCTGGA	TAAAGGGCGGTGGCACTGTCTA	Chappell *et al.*[S7]
GAPDH	100 ng	CTGTTGCTGTAGCCAAATTCGT	ACCCACTCCTCCACCTTTGAC	Chappell *et al.*[S7]

Supplementary References

S1. Sriramachandran, A.M., Petrosino, G., Mendez-Lago, M., Schafer, A.J., Batista-Nascimento, L.S., Zilio, N., and Ulrich, H.D. (2020). Genome-wide Nucleotide-Resolution Mapping of DNA Replication Patterns, Single-Strand Breaks, and Lesions by GLOE-Seq. Molecular Cell *78*, 975-985 e977. 10.1016/j.molcel.2020.03.027.

S2. Wong, N., John, S., Nussenzweig, A., and Canela, A. (2021). END-seq: An Unbiased, High-Resolution, and Genome-Wide Approach to Map DNA Double-Strand Breaks and Resection in Human Cells. Methods Mol Biol *2153*, 9-31. 10.1007/978-1-0716-0644-5_2.

S3. Logsdon, G.A., Vollger, M.R., Hsieh, P., Mao, Y., Liskovykh, M.A., Koren, S., Nurk, S., Mercuri, L., Dishuck, P.C., Rhie, A., et al. (2021). The structure, function and evolution of a complete human chromosome 8. Nature *593*, 101-107. 10.1038/s41586-021-03420-7.

S4. Canela, A., Maman, Y., Huang, S.N., Wutz, G., Tang, W., Zagnoli-Vieira, G., Callen, E., Wong, N., Day, A., Peters, J.M., et al. (2019). Topoisomerase II-Induced Chromosome Breakage and Translocation Is Determined by Chromosome Architecture and Transcriptional Activity. Molecular Cell *75*, 252-266 e258. 10.1016/j.molcel.2019.04.030.

S5. Wu, W., Hill, S.E., Nathan, W.J., Paiano, J., Callen, E., Wang, D., Shinoda, K., van Wietmarschen, N., Colon-Mercado, J.M., Zong, D., et al. (2021). Neuronal enhancers are hotspots for DNA single-strand break repair. Nature *593*, 440-444. 10.1038/s41586-021-03468-5.

S6. Hasson, D., Alonso, A., Cheung, F., Tepperberg, J.H., Papenhausen, P.R., Engelen, J.J., and Warburton, P.E. (2011). Formation of novel CENP-A domains on tandem repetitive DNA and across chromosome breakpoints on human chromosome 8q21 neocentromeres. Chromosoma *120*, 621-632. 10.1007/s00412-011-0337-6.

S7. Chappell, W.H., Gautam, D., Ok, S.T., Johnson, B.A., Anacker, D.C., and Moody, C.A. (2015). Homologous Recombination Repair Factors Rad51 and BRCA1 Are Necessary for Productive Replication of Human Papillomavirus 31. J Virol *90*, 2639-2652.