Supplementary Information

Yeast EndoG prevents genome instability by degrading extranuclear DNA species

Yang Yu[1]#, Xin Wang[2,3]#, Jordan Fox[1]#, Ruofan Yu[1,4], Pilendra Thakre[1], Brenna McCauley[4], Nicolas Nikoloutsos[4,5], Yang Yu[2,3], Qian Li[1], P. J. Hastings[1], Weiwei Dang[1,4], Kaifu Chen[2,3*], Grzegorz Ira[1*]

[1] Baylor College of Medicine, Department of Molecular and Human Genetics, One Baylor Plaza, Houston, TX 77030, USA

[2] Department of Cardiology, Boston Children's Hospital, 300 Longwood Avenue, Boston, MA 02115, USA

[3] Department of Pediatrics, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA

[4] Huffington Center on Aging, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

[5] Department of Bioengineering, Rice University, 6500 Main Street, Houston, TX 77030, USA

# - equal contribution

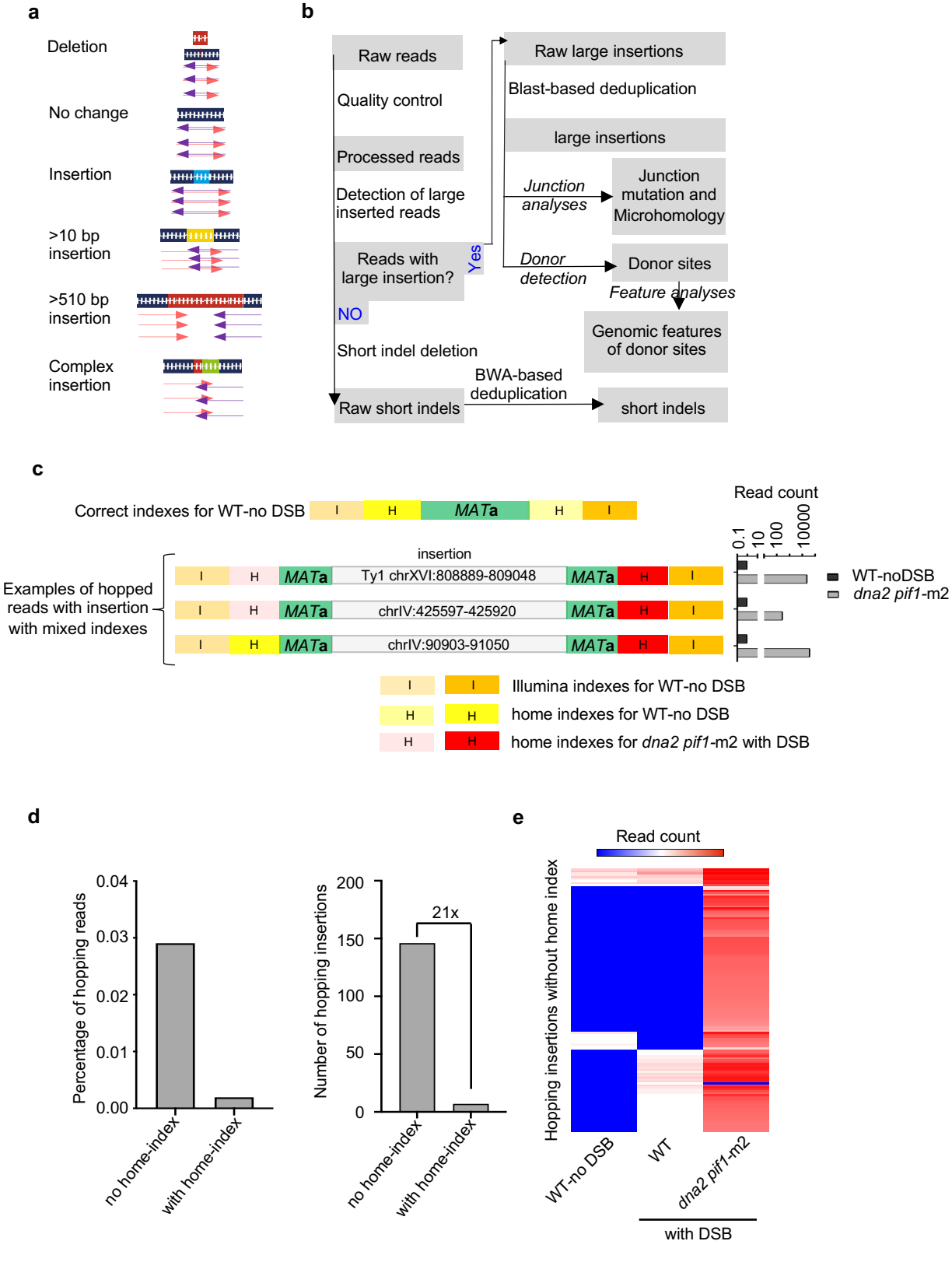
Correspondence:

Grzegorz Ira, e-mail: gira@bcm.edu

Kaifu Chen, e-mail: Kaifu.Chen@childrens.harvard.edu

1. **Supplementary Figures 1-7**
2. **Supplementary Tables 1-2**

Supplementary References


a
Deletion
No change
Insertion
>10 bp insertion
>510 bp insertion
Complex insertion
b
Raw reads
Quality control
Processed reads
Detection of large inserted reads
Reads with large insertion?
Yes
NO
Short indel deletion
Raw short indels
BWA-based deduplication
short indels
Raw large insertions
Blast-based deduplication
large insertions
Junction analyses
Junction mutation and Microhomology
Donor detection
Donor sites
Feature analyses
Genomic features of donor sites
c
Correct indexes for WT-no DSB
I
H
MATa
H
I
Read count
0.1
10
100
10000
insertion
Examples of hopped reads with insertion with mixed indexes
Ty1 chrXVI:808889-809048
chrIV:425597-425920
chrIV:90903-91050
WT-noDSB
dna2 pif1-m2
Illumina indexes for WT-no DSB
home indexes for WT-no DSB
home indexes for dna2 pif1-m2 with DSB
d
Percentage of hopping reads
0.04
0.03
0.02
0.01
0.00
no home-index
with home-index
Number of hopping insertions
200
150
100
50
0
21x
e
Read count
Hopping insertions without home index
WT-no DSB
WT
dna2 pif1-m2
with DSB

Supplementary Figure 1. *Break-Ins* method for analysis of templated insertions

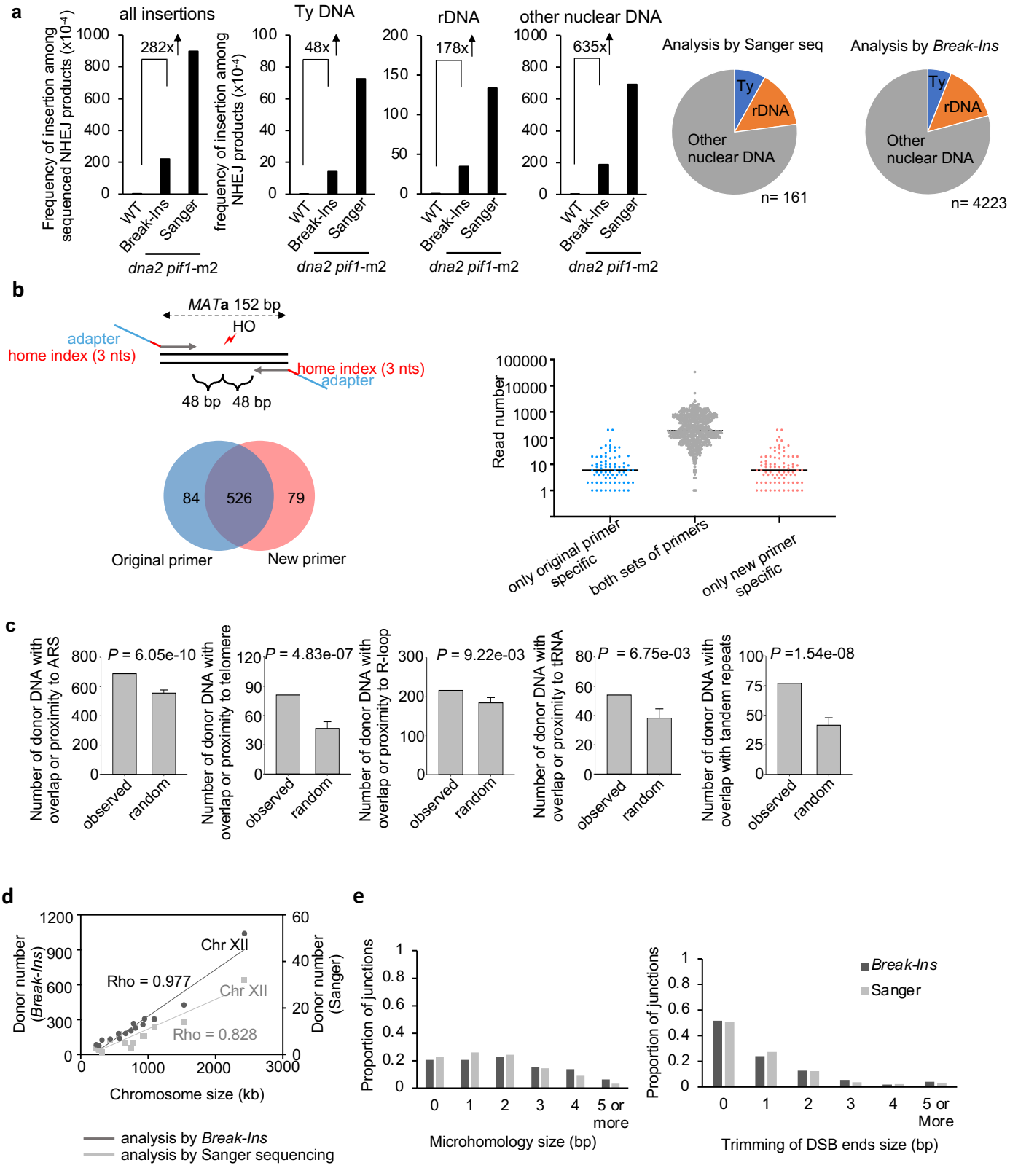
a, Schematic showing different types of sequence variation at DSB.

b, Flowchart showing the computational pipeline for detecting different types of sequence variation at DSB.

c, An example of cross-sample contamination indicated by index hopping during amplicon sequencing by MiSeq. Three examples show reads that have Illumina indexes specific for the DNA sample from a wild-type no DSB control but carry insertions. These reads can be eliminated by analysis of secondary home indexes. Number of false insertion reads in wild-type compared to same insertion reads in *dna2*Δ *pif1*-m2 is shown on the right.

d, Comparison of read hopping identified based on MiSeq amplicon sequencing with and without home indexes.

e, Number of hopped reads without home index in single sequencing run. Source data are provided as a Source Data file.


a
all insertions
Ty DNA
rDNA
other nuclear DNA
Frequency of insertion among sequenced NHEJ products (x10⁻⁴)
frequency of insertion among NHEJ products (x10⁻⁴)
282x
48x
178x
635x
WT
Break-Ins
Sanger
dna2 pif1-m2
Analysis by Sanger seq
Analysis by Break-Ins
Ty
rDNA
Other nuclear DNA
n= 161
n= 4223
b
MATa 152 bp
HO
adapter
home index (3 nts)
48 bp
48 bp
84
526
79
Original primer
New primer
Read number
only original primer specific
both sets of primers
only new primer specific
c
Number of donor DNA with overlap or proximity to ARS
P = 6.05e-10
Number of donor DNA with overlap or proximity to telomere
P = 4.83e-07
Number of donor DNA with overlap or proximity to R-loop
P = 9.22e-03
Number of donor DNA with overlap or proximity to tRNA
P = 6.75e-03
Number of donor DNA with overlap with tandem repeats
P =1.54e-08
observed
random
d
Donor number (Break-Ins)
Donor number (Sanger)
Chr XII
Rho = 0.977
Chr XII
Rho = 0.828
Chromosome size (kb)
analysis by Break-Ins
analysis by Sanger sequencing
e
Proportion of junctions
Microhomology size (bp)
Break-Ins
Sanger
Trimming of DSB ends size (bp)
5 or more
5 or More

Supplementary Figure 2. Comparison of Sanger sequencing and *Break-Ins* analysis

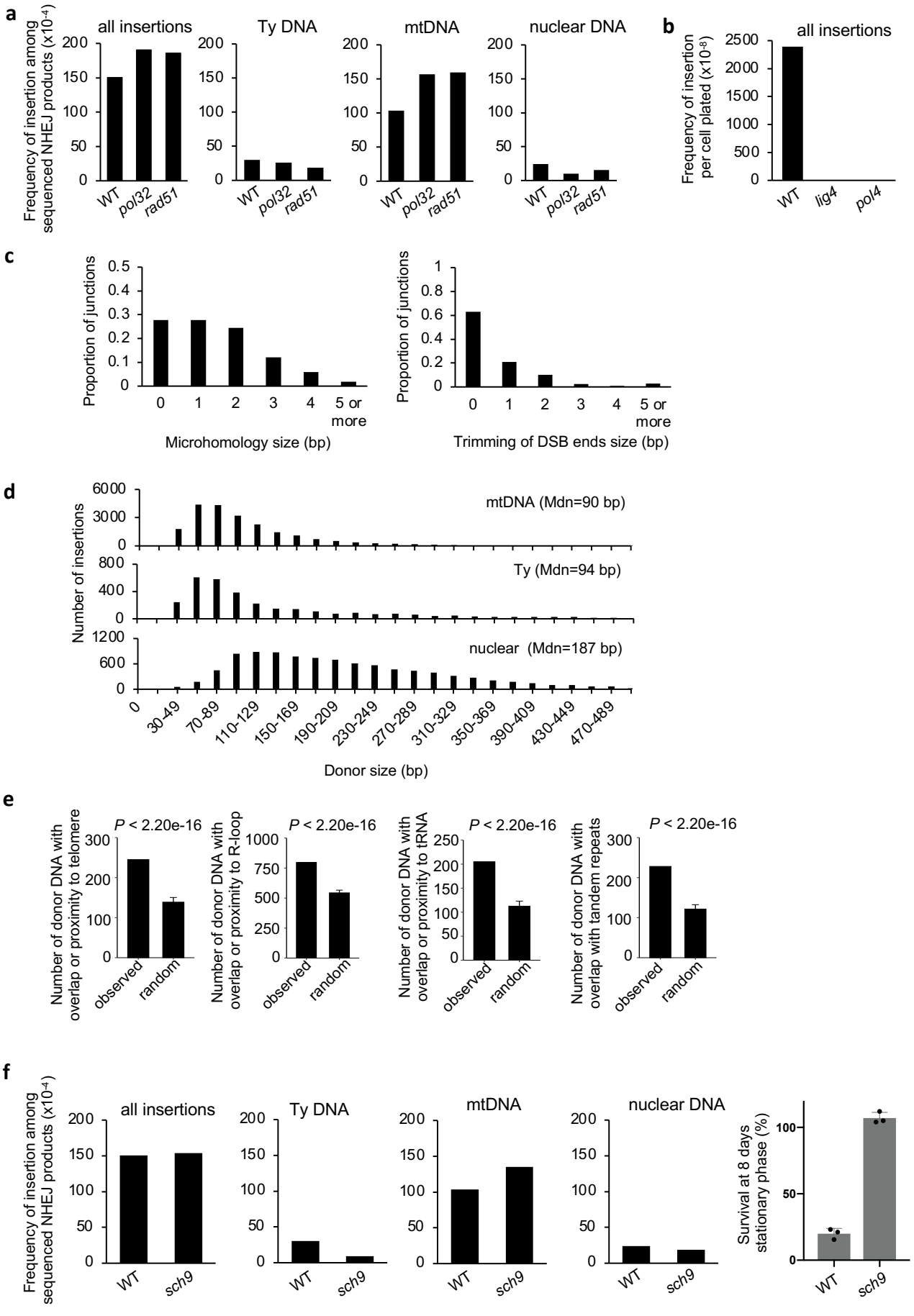
a, Frequency and types of DNA inserted at DSBs in wild-type or *dna2*Δ *pif1*-m2, (n – number of NHEJ products tested is shown in Supplementary Data 1). Sanger sequencing data are taken from previous publication[23].

b, Comparison of *Break-Ins* analysis done with two different primer sets. Scheme showing primer position with respect to DSB ends and number of insertions identified by both or just one set of primers (left panel). Original primer set is shown in Figure **1b**. Most of the unique insertions are represented by low read number (right panel).

c, Analysis of features of DNA inserted from nuclear genome at DSBs in *dna2*Δ *pif1*-m2. P values were calculated using one-sided permutation test. Proximity is defined as sequence within 1 kb from ARS or telomere and within 0.2 kb from tRNA or R-loop.

d, Distribution of donor DNA per chromosome.

e, Analysis of microhomology and DSB ends trimming at insertion junctions in *dna2*Δ *pif1*-m2. Source data are provided as a Source Data file.


a
all insertions
Ty DNA
mtDNA
nuclear DNA
Frequency of insertion among sequenced NHEJ products (x10-4)
WT
pol32
rad51
b
all insertions
Frequency of insertion per cell plated (x10-8)
WT
lig4
pol4
c
Proportion of junctions
Microhomology size (bp)
Trimming of DSB ends size (bp)
5 or more
d
mtDNA (Mdn=90 bp)
Ty (Mdn=94 bp)
nuclear (Mdn=187 bp)
Number of insertions
Donor size (bp)
e
P < 2.20e-16
Number of donor DNA with overlap or proximity to telomere
Number of donor DNA with overlap or proximity to R-loop
Number of donor DNA with overlap or proximity to tRNA
Number of donor DNA with overlap with tandem repeats
observed
random
f
all insertions
Ty DNA
mtDNA
nuclear DNA
Frequency of insertion among sequenced NHEJ products (x10-4)
Survival at 8 days stationary phase (%)
WT
sch9

Supplementary Figure 3. Analysis of templated insertions in mutants affecting DNA repair and aging

a, Frequency and types of DNA inserted at DSBs in *rad51*Δ and *pol32*Δ (n - number of NHEJ products analyzed is shown in Table S1).

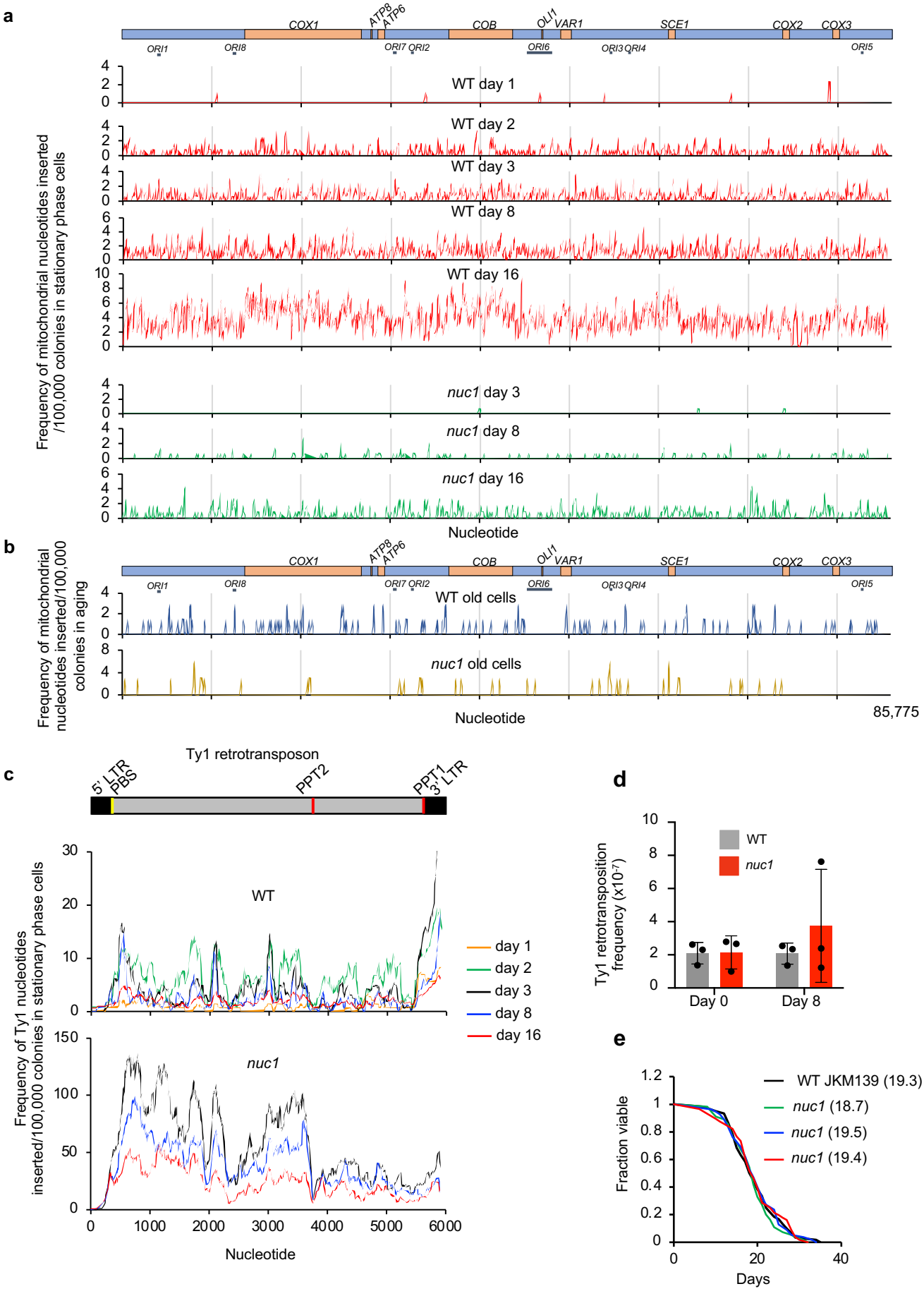
b, Frequency of DNA inserted at DSBs in *lig4*Δ and *pol4*Δ (n - number of cells plated is shown in Table S1).

c, Analysis of microhomology and DSB ends trimming at insertion junctions in 16 days wild-type stationary phase cells.

d, Insertion size analysis originating from mtDNA, Ty DNA, and nuclear genome.

e, Analysis of features of DNA inserted from nuclear genome at DSBs in wild-type stationary phase cells. Insertions observed in all days (3, 8, 16) were combined for this analysis. Insertions coming from Ty retrotransposons, mtDNA, rDNA, *MAT***a** or 2μ plasmid were excluded from these analyses. P values were calculated using one-sided permutation test. Proximity is defined as sequence within 1 kb from ARS or telomere and within 0.2 kb from tRNA or R-loop.

f, Frequency and types of DNA inserted at DSBs in *sch9*Δ at 8 days in stationary phase (n- number of NHEJ products analyzed in shown in Supplementary Data 1). Viability of wild-type and *sch9*Δ cells at 8 days in stationary phase is shown on the right. Source data are provided as a Source Data file.


a
COX1
ATP8
ATP6
COB
OLI1
VAR1
SCE1
COX2
COX3
ORI1
ORI8
ORI7
ORI2
ORI6
ORI3
ORI4
ORI5
WT day 1
WT day 2
WT day 3
WT day 8
WT day 16
nuc1 day 3
nuc1 day 8
nuc1 day 16
Frequency of mitochondrial nucleotides inserted /100,000 colonies in stationary phase cells
Nucleotide
b
WT old cells
nuc1 old cells
Frequency of mitochondrial nucleotides inserted/100,000 colonies in aging
85,775
c
Ty1 retrotransposon
5' LTR
PBS
PPT2
PPT1
3' LTR
WT
nuc1
day 1
day 2
day 3
day 8
day 16
Frequency of Ty1 nucleotides inserted/100,000 colonies in stationary phase cells
Nucleotide
d
WT
nuc1
Ty1 retrotransposition frequency (x10-7)
Day 0
Day 8
e
WT JKM139 (19.3)
nuc1 (18.7)
nuc1 (19.5)
nuc1 (19.4)
Fraction viable
Days

Supplementary Figure 4. Analysis of mtDNA and Ty1 sequences inserted at DSB

a, Analysis of sequences inserted at DSB from mtDNA in wild-type and *nuc1*Δ during stationary phase. A scheme of mtDNA genome, including positions of replication origin (ORI) and major ORFs, is shown.

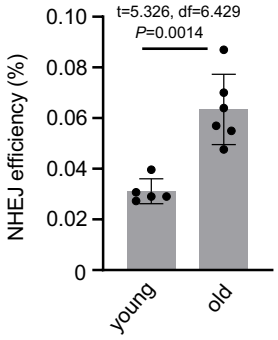
b, Analysis of sequences inserted at DSB from mtDNA in wild-type and *nuc1*Δ in aged cells.

c, Analysis of sequences inserted at DSB from Ty1 in wild-type and *nuc1*Δ during stationary phase.

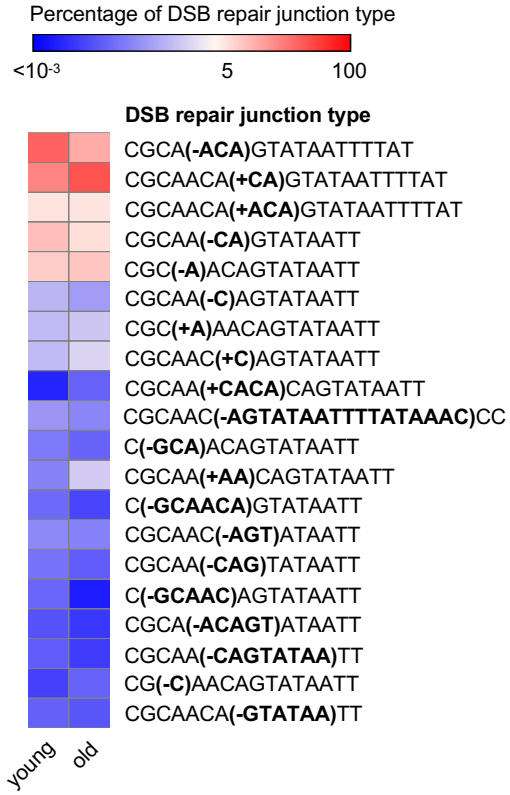
d, Transposition frequency in wild-type and *nuc1*Δ mutant in stationary phase cells, (mean ± SD; n=3; n represents biological repeats.).

e, Analysis of lifespan of wild-type and three independent *nuc1*Δ mutant cells. Average lifespan in each strain is shown in parentheses. Source data are provided as a Source Data file.

a


t=5.326, df=6.429
P=0.0014
NHEJ efficiency (%)
0.10
0.08
0.06
0.04
0.02
0
young
old

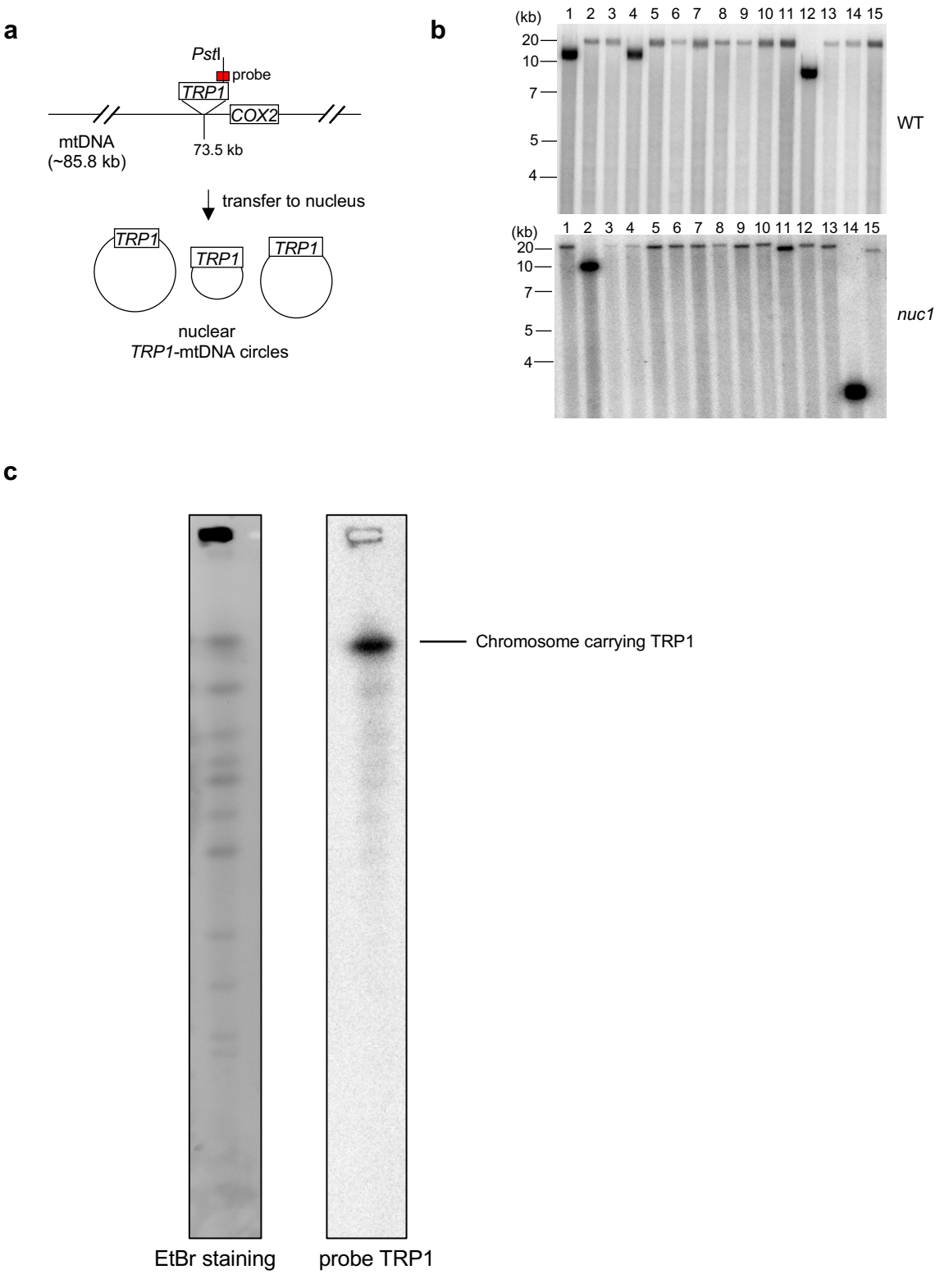
b


Percentage of DSB repair junction type
<10-3
5
100
DSB repair junction type
CGCA(-ACA)GTATAATTTTAT
CGCAACA(+CA)GTATAATTTTAT
CGCAACA(+ACA)GTATAATTTTAT
CGCAA(-CA)GTATAATT
CGC(-A)ACAGTATAATT
CGCAA(-C)AGTATAATT
CGC(+A)AACAGTATAATT
CGCAAC(+C)AGTATAATT
CGCAA(+CACA)CAGTATAATT
CGCAAC(-AGTATAATTTTATAAAC)CC
C(-GCA)ACAGTATAATT
CGCAA(+AA)CAGTATAATT
C(-GCAACA)GTATAATT
CGCAAC(-AGT)ATAATT
CGCAA(-CAG)TATAATT
C(-GCAAC)AGTATAATT
CGCA(-ACAGT)ATAATT
CGCAA(-CAGTATAA)TT
CG(-C)AACAGTATAATT
CGCAACA(-GTATAA)TT
young
old

Supplementary Figure 5. Analysis of indel junctions at repaired DSBs in old mother cells

a, NHEJ efficiency during replicative aging. Mean values are plotted, error bars represent SD; n=5 for young cells and 6 for old cells. n represents biological repeats. P values determined using unpaired two-tailed Welch's t-test.

b, Analysis of top indel junctions among DSB repair products in growing young and old cells. Source data are provided as a Source Data file.


a
PstI
probe
TRP1
COX2
mtDNA
(~85.8 kb)
73.5 kb
transfer to nucleus
TRP1
TRP1
TRP1
nuclear
TRP1-mtDNA circles
b
(kb)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
20
10
7
5
4
WT
(kb)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
20
10
7
5
4
nuc1
c
Chromosome carrying TRP1
EtBr staining
probe TRP1

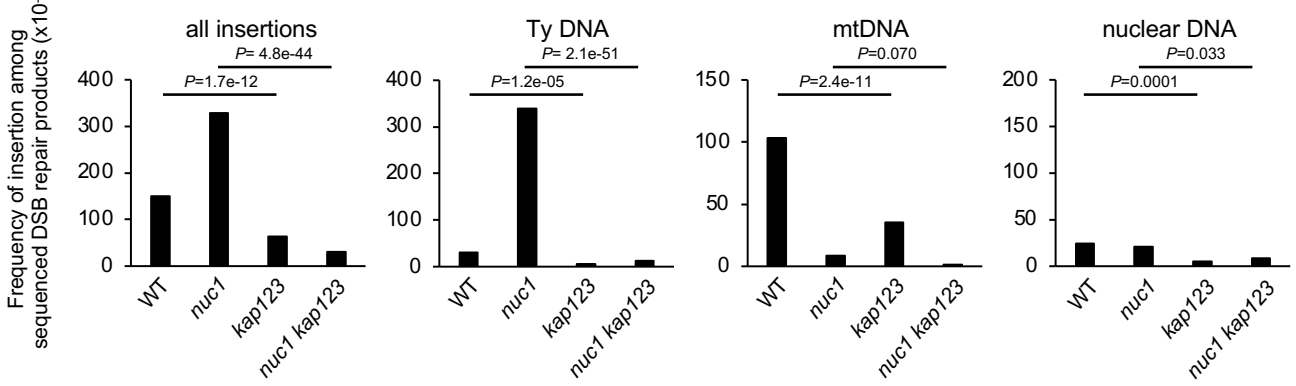
Supplementary Figure 6. Analysis of nuclear *TRP1*-mtDNA

a, Schematic of mtDNA marked with *TRP1* and circular *TRP1*-mtDNA transferred to nucleus. DNA was digested with the *Pst*I restriction enzyme that cuts once 3' to the *TRP1* reporter inserted within mtDNA but not anywhere within the mtDNA itself. Southern blot analysis with a *TRP1*-specific probe is expected to show two fragments if mtDNA-*TRP1* were linear and one fragment if it were circular.

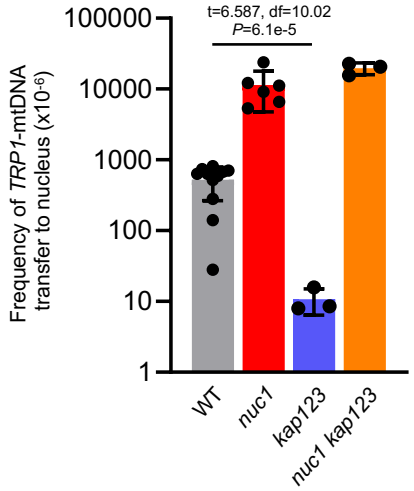
b, Southern blot analysis of nuclear *TRP1*-mtDNA digested with *Pst*I in wild-type and *nuc1*Δ cells. DNA probe location is indicated in **a**. 15 independent *TRP1*-mtDNA for each were analyzed. A single band was observed in all cases.

c, Southern blot analysis of a single colony carrying stable nuclear *TRP1*-mtDNA. Chromosomes were separated by CHEF and gel was stained with EtBr (left). Southern blot analysis of this gel with *TRP1* probe. Source data are provided as a Source Data file.

a


Frequency of insertion among sequenced DSB repair products (x10⁻⁴)
all insertions
P= 4.8e-44
P=1.7e-12
Ty DNA
P= 2.1e-51
P=1.2e-05
mtDNA
P=0.070
P=2.4e-11
nuclear DNA
P=0.033
P=0.0001
WT
nuc1
kap123
nuc1 kap123

b


Frequency of TRP1-mtDNA transfer to nucleus (x10⁻⁶)
t=6.587, df=10.02
P=6.1e-5
WT
nuc1
kap123
nuc1 kap123

Supplementary Figure 7. Analysis of Kap123 in transfer of mtDNA to nucleus

a, Frequency and types of DNA inserted at DSBs in wild-type, *nuc1*Δ, *kap123*Δ and *nuc1*Δ *kap123*Δ at 8 days in stationary phase (P values were determined using two-tailed χ^2 test. n - number of NHEJ products analyzed in shown in Table S1).

b, Frequency of Trp^+ colonies carrying nuclear *TRP1*-mtDNA in wild-type and indicated mutant cells, (mean ± SD; n=11 for wild type, 6 for *nuc1*Δ, 3 for *kap123*Δ or *nuc1*Δ *kap123*Δ mutant. n represents biological repeats.; P values were determined using unpaired two-tailed Welch's t-test). Source data are provided as a Source Data file.

Supplementary Table 1. List of strains used in this study.

Strain name	Parental strain	Genotype	Source
JKM139		DELho *hml*::*ADE1 MAT***a** *hmr*::*ADE1 ade1 leu2-3,112 lys5 trp1*::*hisG ura3-52 ade3*::*GAL10*::*HO*	[1]
yYY363	JKM139	*nuc1::klTRP1*	this study
yYY661	JKM139	*nuc1*-H138A	this study
yYY337	JKM139	*sch9::klTRP1*	this study
yWH475	JKM139	*dna2::kanMX pif1*-m2	[2]
yYY387	JKM139	*pol32::natMX*	this study
yYY466	JKM139	*rad51::kanMX*	this study
yYY399	JKM139	*lig4::klTRP1*	this study
yYY400	JKM139	*pol4::klTRP1*	this study
yYY704	JKM139	*kap123::kanMX*	this study
yYY705	JKM139	*nuc1::klTRP1 kap123::kanMX*	this study
yYY339	JKM139	*spt3::klTRP1*	this study
yYY587	JKM139	*nuc1::kanMX spt3::klTRP1*	this study
PTY44		*MATα ura3-52 lys2 leu2-3,112 trp1-Δ1 [ρ+, TRP1]*	[3]
yYY583	PTY44	*nuc1::kanMX*	this study
yYY660	PTY44	*nuc1*-H138A	this study, constructed using pCORE-UH
yYY706	PTY44	*kap123::kanMX*	this study
yYY707	PTY44	*nuc1::pCORE-UH kap123::kanMX*	this study
yYY664	PTY44	*MAT***a** *ura3-52 lys2 leu2-3,112 trp1-Δ1 [ρ+, TRP1]*	this study, constructed using HO plasmid to switch mating type
yYY671	yYY664	*nuc1::pCORE-UH*	this study
PTY44/ YY664	PTY44	*MAT***a**/Matα *ura3-52/ ura3-52 lys2/lys2 leu2-3,112/ leu2-3,112 trp1-Δ1/ trp1-Δ1 [ρ+, TRP1]*	this study, constructed by mating PTY44 and yYY664
yYY671/ YY583	yYY583	*MAT***a**/Matα *ura3-52/ ura3-52 lys2/lys2 leu2-3,112/ leu2-3,112 trp1-Δ1/ trp1-Δ1 [ρ+, TRP1] nuc1::kanMX/ nuc1::pCORE-UH*	this study, constructed by mating yYY671 and yYY583
DG1657		*MAT***a** *ura3-167 his3Δ-200 trp1-hisG leu2-hisG Ty1-270his3-AI Ty1-588neo Ty1-146[tyb1::lacZ]*	[4]
yJL67	DG1657	*nuc1::klTRP1*	this study

Supplementary Table 2. List of DNA oligos used in this study.

Name	Sequence(5' to 3')	Comments
A. Primers for amplification of the *MAT***a** locus for *Break-In* sequencing		
mata-xxx-F1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGxxxGCATAGTCGGGTTTTTCTTTTAGTTTCAGC	To amplify the *MAT***a** locus for Break-In sequencing. 11/18 bp from HO cut site (Fig. 1b). xxx reprentents home index.
mata-xxx-R1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGxxxCAACCACTCTACAAAACCAAAACCAGGGT	
mata-xxx-F3	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGxxxTGAGATCTAAATAAATTCGTTTTCAATGAT	To amplify the *MAT***a** locus for Break-In sequencing. 48/48 from HO cut site (Supplementary Figure 2b). xxx reprentents home index.
mata-xxx-R3	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGxxxCGTCACCACGTACTTCAGCATAATTATTCG	
B. Primers for amplification of DNA probes used for Southern blot		
TRP1-Fw	GACTGACGCCAGAAAATGTTGG	Supplementary Figure 6
TRP1-Rw	CAAGAATCGGGTCATTGTAGCG	Supplementary Figure 6
Typeak fw	CGCAGTATCCATCATCAGTTG	Fig. 5c probe A
Typeak rv	AAGTGTTATACAAGAAGGTGAGTTC	Fig. 5c probe A
mtDNA F1	CACCACTAATTGAAAACCTGTCTG	Fig. 4d
mtDNA R2	AACCGTACGTGCGACTTTCATC	Fig. 4d
ACT1-fw	TCTTCCCATCTATCGTCGGTAGAC	Fig. 4d, 5c loading control
ACT1 rv	GGTCAATACCGGCAGATTCC	Fig. 4d, 5c loading control
TRA1 fw	GTCCTAATACGACTTTTCAAATTGTCCTTTATGTCCGTCA	Fig. 4d, 5c loading control
TRA1 rv	ATACTTGTAAGCACTCTTCCTGTAGTGAATATCACTTTTG	Fig. 4d, 5c loading control
TOM1 fw	CTTCAAAATTGAAGATCATG	Fig. 4d, 5c loading control
TOM1 rv	CGATTGATTGAGCGATGATG	Fig. 4d, 5c loading control
C. Primers for making point mutants		
Nuc1-pCORE-Fw	CTTATCCACCTACCCAGAAACCTAATAGTAATATTCAATCTCACTCTTTCTTCGTACGCTGCAGGTCGAC	
Nuc1-pCORE-Rw	GCTGTTGGTGCTTCTGCAACAATCAATTTAAAAAAGTGCGTTGGAACAGCCCGCGCGTTGGCCGATTCAT	
Nuc1-F5	ATGTGCAGTAGGATACTCTTGT	
Nuc1-R5	AAGTTCTAGCCCAGTACTTCTC	
D. Primers for amplification of the 84-1000 bp fragments for transformation/insertion analysis (Fig. 4f)		
Lambda-84-Fw	AACACGGTGGGCTCAGAGAATCC	
Lambda-84-Rw	TGTTAGGATGACACTGTACTGACCGT	
Lambda-100-Fw	AACATGGGCACGCTGGAGAC	
Lambda-100-Rw	TGTTCGCGGTATCAGGACGACCAATA	
Lambda-150-Fw	AACATGCTGATTAAGGCAGAGGCTGC	
Lambda-150-Rw	TGTTCGGGCGGCTTCAAGCGCAA	
Lambda-200-Fw	AACAGGCGTTTCCGTTCTTCTTC	
Lambda-200-Rw	TGTTTACGGATACTCGCACCGAA	
Lambda-300-Fw	AACAGGACAAAAATGCGCAGCA	
Lambda-300-Rw	TGTTCGTCAGCAGGGCAGCATGAG	
Lambda-400-Fw2	AACAGTGCTACCCGAACACGGC	
Lambda-400-Rw2	TGTTCGCTGAGCACATCCCACGCC	
Lambda-500-Fw2	AACAACGACGGCACGAACGG	
Lambda-500-Rw2	TGTTGCATTCATTCAGTGTTTCCTGCC	
Lambda-1k-Fw	AACAGGCTTCGCTCACTGTTCAGG	
Lambda-1k-Rw	TGTTCCCTTTCGTTTTCATCCAGTC	
E. Primers for amplification of the 24-84 bp fragments for transformation/insertion analysis (Fig. 4f)		
M13-X-24-Fw	AACACTGGTAAACAAGGGTTAACA	
M13-X-24-Rw	TGTTAACCCTTGTTTACCAGTGTT	
M13-X-34-Fw	AACATTATCGAACACTGGTAAACAAGGGTTAACA	
M13-X-34-Rw	TGTTAACCCTTGTTTACCAGTGTTCGATAATGTT	
M13-X-44-Fw	AACAATTTTGAACATTATCGAACACTGGTAAACAAGGGTTAACA	
M13-X-44-Rw	TGTTAACCCTTGTTTACCAGTGTTCGATAATGTTCAAAATTGTT	

M13-X-54-Fw	AACAGCCTCTAACAATTTTGAACATTATCGAACACTGGTAAACAAGGGTTAACA	
M13-X-54-Rw	TGTTAACCCTTGTTTACCAGTGTTCGATAATGTTCAAAATTGTTAGAGGCTGTT	
M13-X-64-Fw	AACATTTTGCAACAGCCTCTAACAATTTTGAACATTATCGAACACTGGTAAACAAGGGTTAACA	
M13-X-64-Rw	TGTTAACCCTTGTTTACCAGTGTTCGATAATGTTCAAAATTGTTAGAGGCTGTTGCAAAATGTT	
M13-X-74-Fw	AACAGCAAAAAACATTTTGCAACAGCCTCTAACAATTTTGAACATTATCGAACACTGGTAAACAAGGGTTAACA	
M13-X-74-Rw	TGTTAACCCTTGTTTACCAGTGTTCGATAATGTTCAAAATTGTTAGAGGCTGTTGCAAAATGTTTTTTGCTGTT	

Supplementary References

1. Moore, J.K., and Haber, J.E. (1996). Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. Mol Cell Biol *16*, 2164-2173.
2. Zhu, Z., Chung, W.H., Shim, E.Y., Lee, S.E., and Ira, G. (2008). Sgs1 helicase and two nucleases dna2 and exo1 resect DNA double-strand break ends. Cell *134*, 981-994.
3. Thorsness, P.E., and Fox, T.D. (1993). Nuclear mutations in Saccharomyces cerevisiae that affect the escape of DNA from mitochondria to the nucleus. Genetics *134*, 21-28. 10.1093/genetics/134.1.21.
4. Sundararajan, A., Lee, B.S., and Garfinkel, D.J. (2003). The Rad27 (Fen-1) nuclease inhibits Ty1 mobility in Saccharomyces cerevisiae. Genetics *163*, 55-67.