Dna2 nuclease deficiency results in large and complex DNA insertions at chromosomal breaks

Yang Yu1,6, Nhung Pham1,6, Bo Xia2,3,4,6, Alma Papusha1, Guangyu Wang2,3,4, Zhenxin Yan1, Guang Peng5, Kaifu Chen2,3,4,6 & Grzegorz Ira1,6

Insertions of mobile elements1–4, mitochondrial DNA5 and fragments of nuclear chromosomes6 at DNA double-strand breaks (DSBs) threaten genome integrity and are common in cancer7–9. Insertions of chromosome fragments at V(D)J recombination loci can stimulate antibody diversification10. The origin of insertions of chromosomal fragments and the mechanisms that prevent such insertions remain unknown. Here we reveal a yeast mutant, lacking evolutionarily conserved Dna2 nuclease, that shows frequent insertions of sequences between approximately 0.1 and 1.5 kb in length into DSBs, with many insertions involving multiple joined DNA fragments. Sequencing of around 500 DNA inserts reveals that they originate from Ty retrotransposons (8%), ribosomal DNA (rDNA) (15%) and from throughout the genome, with preference for fragile regions such as origins of replication, R-loops, centromeres, telomeres or replication fork barriers. Inserted fragments are not lost from their original loci and therefore represent duplications. These duplications depend on nonhomologous end-joining (NHEJ) and Pol4. We propose a model in which alternative processing of DNA structures arising in Dna2–deficient cells can result in the release of DNA fragments and their capture at DSBs. Similar DNA insertions at DSBs are expected to occur in any cells with linear extrachromosomal DNA fragments.

We analysed DSB repair by NHEJ in yeast cells deficient in the nuclease/ helicase Dna2 and found that approximately 8% of the repair events carried large insertions of about 100–1,500 bp, whereas the remaining repair events were comparable to those found in wild-type cells (Fig. 1a, b, Extended Data Fig. 1, Extended Data Tables 1, 2, Supplementary Information Table 1). In this experimental design11, homothallic switching (HO) endonuclease-induced DSBs at the MATα locus can only be repaired by imprecise NHEJ that alters the HO-cleavage site, preventing further cutting. Analysis was done in cells carrying a suppressor of dna2Δ lethality, the pif1-m2 mutation12. No insertions were found in pif1-m2 or wild-type control cells. The nuclease activity of Dna2, but not its helicase activity, is required to suppress insertions at DSBs (Fig. 1a). We observed similar insertions to those found at the MATα locus at a DSB induced at an artificially introduced ACT1 intron within the URA3 locus or at CRISPR–Cas9-induced DSBs in the LYS2 gene (Extended Data Fig. 1, Extended Data Table 1). Sequencing analysis of about 500 insertions from all Dna2-deficient cells reveals that approximately 15% of events contained 2 to 4 fragments from different chromosomes joined together at the DSB (Fig. 1c, d).

The DNA insertions result in duplications, as none of 25 randomly tested donor DNA fragments were deleted from their original locus, and the number of insertions originating from essential genes (46/222 in all strains tested) was proportional to the number of essential genes in yeast (~20%). The duplicated sequences include short complete genes, replication origins, and fragments of telomeres or centromeres (Supplementary Information Table 1). NHEJ is the primary pathway mediating these insertions, as most of the junctions carried 0–4 nucleotides of microhomology (Fig. 1e). DSB ends were mostly maintained (Extended Data Fig. 1) and deletion of NHEJ components (Ku, Lig4 or Pol4) nearly abolished insertions (Fig. 1a). Single insertion captured in NHEJ-deficient cells shows an increased microhomology and loss of sequences at DSB ends, typical features of alternative end joining. By contrast, deletion of homologous-recombination-specific enzymes in pif1-m2 dna2Δ cells had no effect on the number of insertions in the case of Rad51, and increased the number of insertions in the case of Rad52 (Fig. 1a).

The origins of inserted DNA in pif1-m2 dna2Δ can be grouped into four major categories. First, about 8% of the insertions are fragments of retrotransposons, which comprise about 3% of the 12.1-Mb yeast genome. Second, about 15% of insertions originate from rDNA,
representing about 10% of the genome. Third, about 74% of insertions originate from elsewhere in the genome. Finally, about 3% of insertions originate from 6.3-kb resident 2μ DNA plasmids (approximately 50 copies per cell), reflecting the proportion of these plasmids in nuclear DNA content. Mitochondrial DNA was not inserted. Proximity of insertion donor DNA to the DSB is not important, as few insertions originate from the chromosome carrying the DSB, and the number of donor DNAs from different chromosomes correlates simply with their size, with the exception of the rDNA cluster on chromosome XII, which contains the largest hotspot of insertion donor DNA (Extended Data Fig. 2). Accordingly, analysis of the three-dimensional proximity of the donor DNAs or randomly selected sequences to the locus of DSBs, as measured by chromosome conformation capture (two-tailed) to determine P value; n = 4 independent experiments.

---

**Fig. 2** Dna2 limits mobility of transposable elements. a. Position of inserted DNA originating from yeast transposons. LTR, long terminal repeat. b. Contribution of yeast transposons to insertions. c. Scheme of Ty1 cDNA analysis, Southern blot analysis and quantification of Ty1 cDNA in indicated mutants. Data are mean ± s.d.; n = 3 independent experiments. For gel source data, see Supplementary Figure 1. d. Retrotransposition rates in indicated mutants. Bootstrap resampling (two-tailed) to determine P value; n = 4 independent experiments.

---

**Fig. 3** Origin of inserted DNA at DSBs. a. Top, location of inserted DNA at DSBs originating from chromosome (ch.) XII. Hotspots (HS, numbered), loci of up to 3 kb that provide at least two inserted fragments, are marked in red. Schematic of single rDNA repeat (second from top) and position of inserted DNA originating from rDNA repeats (bottom two). Right, number of insertions with respect to RFB position. b. Examples of hotspots of the origin of insertions with genomic features shown. ORF, open reading frame. c. Inserted DNA originating from 2μ plasmid. FRT, Flp recognition target; ori, origin of replication sequence.
All four active yeast transposons were present among insertions at DSBs (Fig. 2a, b), while only the most abundant Ty1 transposon insertions have previously been reported\(^2\). Deletion of Spt3—which is required for transposon transcription\(^1^\)—in pif1-m2 dna2Δ cells decreases the number of transposon insertions (from 8%) to the proportion of transposons in the genome (approximately 3%) (Extended Data Table 1), indicating that the reverse transcriptase activity of retrotransposons is important for transposon insertion. Ty1 cDNA levels (Fig. 2c) and the rate of retrotransposition measured via Ty1–His3 reporter\(^2\) also increased with DNA2 deletion (Fig. 2d). Increased levels of cDNA in pif1-m2 dna2Δ are not related to increased transcription of transposons, but may result from increased cDNA stability (Extended Data Fig. 3). Together, Dna2 inhibits retrotransposition and insertions of transposon fragments at DSBs.

About 15% of insertions originate from approximately 150 rDNA repeats. Each 9.1-kb repeat contains 55 and 355 genes, origin of replication (ARS) and replication fork blocking (RFB) sequences (Fig. 3a).

Binding of Fob1 protein to RFB sequences prevents head-on collisions between replication forks and 35S transcription bubbles. Most rDNA inserted at DSBs originate from the region between ARS and RFB sequences, in which Dna2 prevents fork stalling\(^6\). This distribution is dependent on Fob1 and, therefore, on fork pausing at RFB (Fig. 3a). Out of 41 donor DNA hotspots providing at least 2 inserted fragments from within a 3-kb region, 34 are located in the vicinity of an ARS. Further, nearly half of 18 insertions from 2μ plasmids come from replication origins (Fig. 3b, c, Extended Data Fig. 4). Genome-wide analysis of the overlap or proximity of insertion donor DNAs to ARS sequences confirms this correlation (Fig. 4a). Donor DNAs were found to be nearer to sites of prominent R-loops, centromeres or telomeres, when compared to randomly selected sequences of equal size and frequency per chromosome (Fig. 4b, c). These features are known to cause fork stalling\(^7\)-\(^9\), and require Dna2 for replication to occur correctly (for example, refs \(^2\)\(^,\)\(^2\)\(^,\)\(^2\)). Finally, treatment of pif1-m2 dna2Δ cells with high dose of hydroxyurea, a drug known to cause fork stalling and reversal\(^10\), results in an approximately twofold increase in insertions events (Fig. 4d). Together, these results show that inserted donor DNAs often originate from fragile genomic regions in which fork stalling is more likely.

As inserted DNA is not deleted from its original locus, it must either be over-replicated and then inserted into a DSB, or originate from a fragmented sister chromatid. We favour the first of these scenarios, because Dna2 has two functions that prevent over-replication: it removes long 5′ flaps during ligation-strand synthesis, and it prevents and/or degrades reversed forks\(^11\)-\(^13\). Long, unprocessed 5′ flaps may contribute to insertions, because deletion of Pol32—the processivity subunit of Pol6— and Pif1 helicase, both of which stimulate the displacement synthesis that generates long flaps\(^14\),\(^15\), reduces insertion frequency by about 50% and decreases the mean size of the insertions (Fig. 4d, e). Moreover, the size range of the insertions observed in these experiments resembles that seen for 5′ flaps in Dna2-deficient mutants\(^16\). Overexpression of Rad52 was previously shown to reduce the level of Dna2 substrates, presumably 5′ flaps\(^17\). Consistent with these results, we found a marked increase in number of insertions in pif1-m2 dna2Δ rad52Δ cells. We note that rad52Δ cells contained rare insertions (1%) of DNA from a 2-kb region on either side of the DSB (Extended Data Fig. 5). Deletion of the nonessential Rad27, which processes much shorter 5′ flaps, does not result in insertions, suggesting that there is efficient alternative processing of the flaps in rad27Δ cells (Fig. 1a). Reversed forks could also contribute, because inserted DNAs originate from genomic regions that are prone to fork stalling (Fig. 4a–c). In the absence of Dna2, unprocessed DNA structures can be cleaved by alternative nucleases, leading to release of DNA fragments that could be subsequently inserted into DSBs. A significant increase of insertions in dna2Δ cells suggests that the nuclease-dead Dna2 may bind and stabilize such DNA structures (Fig. 1a). Deletion of structure-specific nuclease Mus81, which cleaves stalled or reversed forks, reduced the number of insertions by more than half (Fig. 4d) and sensitized dna2Δ mutants to DNA damage (Extended Data Fig. 5). This means that Yen1 can cleave at least some unprocessed structures in dna2Δ cells. Although dna2Δ yen1Δ cannot be constructed, we found that dna2Δ cells carrying constitutively active yen1ON exhibited increased numbers of insertions and complex events

---

**Fig. 4** | **Insertions originate from regions where replication forks stall.** a–c, Plots showing overlap or proximity of observed insertions and control DNA with indicated genomic features. P values for overlap and proximity are determined by bootstrapping and one-tailed Wilcoxon test, respectively. a, b, n = 370; c, n = 370 for centromere proximity analysis; n = 371 for telomere proximity analysis; n represents the number of independent insertions. Experiments with control DNA were repeated 1,000,000 times (a) or 1,000 times (b, c). d, Insertion frequencies in indicated mutants. χ² test is used to determine P values; number of colonies tested per mutant is indicated in Extended Data Table 1. HU, hydroxyurea. e, Insertion length analysis in indicated mutants. P value is calculated by one-tailed, one-sample Wilcoxon test; number of independent insertions per mutant is shown in Extended Data Table 1. Bar graphs: data are mean ± s.d. Box plots: centre line is median, boxes show first and third quartiles, whiskers extend to the most extreme data points that are no more than 1.5 fold of the interquartile range from the box.

---

© 2018 Springer Nature Limited. All rights reserved.
1. Moore, J. K. & Haber, J. E. Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. Nature 383, 644–646 (1996).

2. Teng, S. C., Kim, B. & Gabriel, A. Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. Nature 383, 641–644 (1996).

3. Yu, X. & Gabriel, A. Patching broken chromosomes with extraneous cellular DNA. Mol. Cell 4, 873–881 (1999).

4. Morrish, T. A. et al. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat. Genet. 31, 159–165 (2002).

5. Ricchetti, M., Fairhead, C. & Dujon, B. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. Nature 402, 96–100 (1999).

6. Onozawa, M. et al. Repair of DNA double-strand breaks by templated nucleotide sequence insertions derived from distant regions of the genome. Proc. Natl. Acad. Sci. USA 111, 7729–7734 (2014).

7. Li, Y. et al. Patterns of structural variation in human cancer. Preprint at https://www.biorxiv.org/content/early/2017/08/27/181339 (2017).

8. Ju, Y. S. et al. Frequent somatic mutation of mitochondrial DNA into the nuclear genome of human cancer cells. Genome Res. 25, 814–824 (2015).

9. Henssen, A. G. et al. POGDS5 promotes site-specific oncogenic mutations in human tumors. Nat. Genet. 49, 1005–1014 (2017).

10. Pieper, K. et al. Public antibodies to malaria antigens generated by two LAIR1 insertion modalities. Nature 548, 597–601 (2017).

11. Moore, J. K. & Haber, J. E. 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 (1996).

12. Budd, M. E., Reis, C. C., Smith, S., Myung, K. & Campbell, J. L. Evidence suggesting that PIWI1 helicase functions in DNA replication with the Dna2 helicase/nuclease and DNA polymerase I. Mol. Cell. Biol. 26, 2490–2500 (2006).

13. Belton, J. M. et al. The conformation of yeast chromosome III is mating type dependent and controlled by the recombination enhancer. Cell Reports 13, 1855–1867 (2015).

14. Winston, F., Durbin, K. J. & Fink, G. R. The SPT3 gene is required for normal transcription of Ty elements in S. cerevisiae. Cell 39, 675–682 (1984).

15. Sundararajan, A., Lee, B. S. & Garfinkel, D. J. The Rad27 (Fen-1) nuclease inhibits Ty1 mobility in Saccharomyces cerevisiae. Genetics 163, 55–67 (2003).

16. Weitao, T., Budd, M., Hoopes, L. L. & Campbell, J. L. Dna2 helicase/nuclease causes replicative fork stalling and double-strand breaks in the ribosomal DNA of Saccharomyces cerevisiae. J. Biol. Chem. 278, 22513–22522 (2003).

17. Greenfeder, S. A. & Newton, C. S. Replication forks pause at yeast centromeres. Mol. Cell. Biol. 12, 4056–4066 (1992).

18. Makovets, S., Herskowitz, I. & Blackburn, E. H. Anatomy and dynamics of DNA replication fork movement in yeast telomeric regions. Mol. Cell. Biol. 24, 4019–4031 (2004).

19. Gan, W. et al. R-loop-mediated genomic instability is caused by impairment of replication fork progression. Genes Dev. 25, 2041–2056 (2011).

20. Markiewicz-Potoczny, M., Lisby, M. & Lydall, D. A critical role for Dna2 at unwound telomeres. Genetics 209, 129–141 (2018).

21. Li, Z. et al. hDNA2 nuclease/helicase promotes centromeric DNA replication and genome stability. EMBO J. 20, e96729 (2018).

22. Hu, J. et al. The intra-S phase checkpoint targets Dna2 to prevent stalled replication forks from reversing. Cell 149, 1221–1232 (2012).

23. Thangavel, S. et al. DNA2 drives processing and restart of reversed replication forks in human cells. J. Cell Biol. 208, 545–562 (2015).

24. Liu, B., Hu, J., Wang, J. & Kong, D. Direct visualization of RNA-DNA primer removal from Okazaki fragments provides support for flap cleavage and exonucleolytic pathways in eukaryotic cells. J. Biol. Chem. 292, 4777–4788 (2017).

25. Pike, J. E., Burgers, P. M., Campbell, J. L. & Bambara, R. A. PIWI helicase lengthens some Okazaki fragment flaps necessitating Dna2 nuclease/helicase action in the two-nucleus processing pathway. J. Biol. Chem. 284, 25170–25180 (2009).

26. Stith, C. M., Sterling, J., Resnick, M. A., Gordenin, D. A. & Burgers, P. M. Flexibility of eukaryotic Okazaki fragment maturation through regulated strand displacement synthesis. J. Biol. Chem. 283, 34129–34140 (2008).

27. Lee, M. et al. Rad52/Rad59-dependent recombination as a means to rectify faulty Okazaki fragment processing. J. Biol. Chem. 289, 15064–15079 (2014).

28. Blanco, M., Matos, J. & West, S. C. Dual control of Yen1 nucleolysis activity and cellular localization by Cdk and Cdc14 prevents genome instability. Mol. Cell 54, 94–106 (2014).

29. Ölzner, G. et al. Replication intermediates that escape Dna2 activity are processed by Holliday junction resolase Yen1. Nat. Commun. 7, 13157 (2016).

30. Michel, A. H. et al. Functional mapping of yeast genomes by saturated transposition. elife 6, e23570 (2017).

Acknowledgements We thank A. Gabriel, D. J. Garfinkel, J. Haber, M. G. Blanco and F. Storici for the gifts of strains and plasmids, and J. Haber and P. Hastings for critical reading of the manuscript. This work was funded by grants from the US National Institutes of Health (GM080600 and GM125650 to G.I., GM125632 and HL133254 to K.C.) and the Cancer Prevention Research Institute of Texas (RP140456 to G.I. and G.P., RP150611 to K.C.).

Reviewer information Nature thanks P. Cejka, L. Symington and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author Contributions Y.Y., N.P. and B.X. contributed equally to this work. Y.Y., N.P. and Z.Y. performed all experiments related to insertions at DNA breaks; N.P. carried out experiments on transposition; B.X., G.W. and K.C. performed and described bioinformatics analysis. G.I., Y.Y. and G.P. designed the experiments, discussed the data and wrote the manuscript.

Competing Interests The authors declare no competing interests.

Additional information Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0769-8.

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0769-8.

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to K.C. or G.I.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Media, yeast strains and plasmids. All strains used in this work are derivatives of three strains (i) JK139 to study insertion at MATa locus (DELhho::ADE1 MATα hmr::ADE1 ade1 leu2-3,112 lys5 trp1::HisG ura3-52 ade2::GAL1::His3) (11); (ii) Y′ET2-9 to study insertion at URA3 locus (DELho::ADE1 hphMX hmr::ADE1 ade1 leu2-3,112 lys5 trp1::HisG URA3::act1 intron::HOs ade3::GAL1::HO); and (iii) DG1657 to study retrotransposition and Ty1 cDNA level (MATα ura3-163 his3-D20 200 trp1::His2 leu2-TY1-270his3-AI Tyl-588neo Tyl-146[tyl-1::Laz]) (13), a gift from D. J. Garfinkel. Y′ET379 strain was obtained by replacing the HO cleavage site with hphMX at MATa locus and by replacing URA3 with URA3::ACT1 intron::HOs. URA3::ACT1 intron::HOs cassette was amplified from Agy117 strain (13), a gift from A. Gabriel. A list of all strains is presented in Supplementary Information Table 2. Helicase-dead mutant (dna2(200L)) and nuclease-dead mutant (dna2(200L)) of DNA2 were introduced into the genome using the delitto perfetto approach (13).

HO induction and analysis of NHEJ efficiency. To induce HO endonuclease, cells from an overnight saturated culture in YEPD (yeast extract–peptone–dextrose) (1% yeast extract, 2% peptone, 2% dextrose) were washed twice with YEP–raffinose (1% yeast extract, 2% peptone, 2% raffinose), inoculated into 5 ml YEP–raffinose and incubated overnight at 30 °C. When the density of the culture reached ~1–2 × 10^8 cells/ml, cells were spread on YEP (yeast extract, peptone)–galactose plates (1% yeast extract, 2% peptone, 2% galacatose) and incubated at 30 °C for up to 6 days. As a control, cells were spread onto YEPD plates. The NHEJ efficiency was calculated as the number of colonies on YEP–galactose divided by the number of colonies on YEP. The experiment was repeated at least three times for each mutant. For hydroxyurea treatment, hydroxyurea was added to a final concentration of 80 mM when the density of the culture reached ~1 × 10^8 cells/ml in YEP–raffinose and incubated for 4 h before plating.

Analysis of insertions at MATa locus. Single colonies from YEP–galactose plates were used for colony PCR using the following primers: mata-F (ACTTCAAGTAAGCTTTGATGTTGTTGGTTGTGTTG) (163 bp upstream of HO cleavage site) and mat-Rw (TACTGACAACATTCAGTACTCGAAAG) (165 bp downstream of HO cleavage site). The amfSure PCR Master Mix (GenDePOT, cat. no. P0311) was used for PCR with the following conditions: 94 °C for 5 min; 35 cycles of 94 °C for 10 s, 60 °C for 10 s, 65 °C for 10 s; and 72 °C for 10 s; and 72 °C for 5 min. PCR products were analysed by electrophoresis (1.2% agarose in 1× TBE buffer) at 8 V/cm for 30 min. PCR products having large insertions were cleaned up with the NucleoSpin Kit (Macherey-Nagel, cat. no. 740609) and sequenced by Sanger sequencing. ApE software was used to analyse the microhomology of insertion. SnapGene was used to map the insertion to chromosome, Ty1 cDNA and 2μ plasmid. To determine statistically significant differences, the one-tailed t-tests with Bonferroni correction were used to analyse the distribution.

Analysis of insertion at URA3 locus. Yeast were grown in YEP–raffinose up to a density of ~1–2 × 10^8 cells/ml, and galactose was added to a final concentration of 2% and incubated at 30 °C for 24 h. Cells were plated on 5-fluoroorotic acid (5-FOA) plates and incubated at 30 °C for 6 days. For transient DSB induction, galactose was added at 1, 2 or 4 h to a final concentration of 2% to shut down the expression of galactose-inducible HO and the cells were plated on 5-FOA plates and incubated at 30 °C for 6 days. To screen for insertions at the DSB, primers Act1-Fw (ATATCGTGGTTATTCAGATCGATCA) (165 bp upstream of HO cleavage site) and Ura3-Rw (ATGTTGAAGCGTGTTAAGCAGCAG) (165 bp downstream of HO cleavage site) were used. The sequencing and analysis of inserts was performed as described above for the MATa locus.

Analysis of insertion at LYS2 locus. Plasmids marked with the LEU2 gene containing constitutively expressed gRNA gLYS2-2 and a galactose-inducible Cas9, a gift from J. Haber, were transformed into wild-type and mutant yeast cells. To induce Cas9, cells from an overnight saturated culture in leucine drop-out glucose medium were washed and inoculated into 5 ml YEP–raffinose and incubated overnight at 30 °C. When the density of the culture reached ~1–2 × 10^8 cells/ml, cells were spread on leucine drop-out galactose plates (2% galactose) and incubated at 30 °C for up to 6 days. As a control, cells were spread onto leucine drop-out glucose plates. The NHEJ efficiency was calculated as described (13). To test for large insertions, single colonies from leucine drop-out galactose plates were used for colony PCR using the following primers: Lys2-Fw (TAGACGAGTTCAACATCTATTAGT) (120 bp upstream of Cas9 cleavage site) and Lys2-Rw (CAAGTTCTTAGGTTAGATCAGGT) (122 bp downstream of Cas9 cleavage site). PCR fragments carrying insertions were sequenced and analysed.

Analysis of extrachromosomal DNA. Yeast were grown in YEPD to a density of ~1–2 × 10^8 cells/ml, then, 1 × 10^7 cells were added to the NorthernMax-Gly Kit (Invitrogen cat. no. AM1946). The 3P-labelled DNA probes were made by randomly primed DNA synthesis. Ty1 PvuII-SnaBI fragment of Ty1-H3 was used as a 3P-labelled DNA probe and was prepared as described above. The control PKY1 probe was prepared by PCR using
two primers PYK1-F1 (GTTGTTGCTGTTCTGACCTGAGAA) and PYK1-R1 (TCAAGATACGGAATTCCCTTAGCC). The intensity of bands on Southern blots corresponding to Ty RNA fragments was analysed with ImageQuant TL and normalized to the PYK1 RNA signal.

**Analysis of Ty retrotransposition rates.** The rate of retrotransposition was estimated in strains carrying the Ty1-270his3-AI reporter. Wild-type and mutant cells were streaked for single colonies on YEPD plates. Individual colonies were used to inoculate 5 ml YEPD cultures that were incubated at 24 °C, diluted and grown to 1 × 10^6 cells/ml before plating 20–30 cells per plate on YEPD. Plates were incubated at 24 °C for 6 days. For each strain tested, 10–20 individual colonies were diluted in water and spread on synthetic complete medium lacking histidine and incubated at 24 °C for 3–4 days. His^+ colonies were then counted. To perform statistical comparisons of spontaneous transposition rates between genotypes we used the Drake estimator as previously described. The bootstrap resampling approach was used to determine P values.

**Bioinformatic analysis of genomic features related to insertion sites.** Positions of confirmed origins of replication (ARSes) were downloaded from the OriDB database (http://cerevisiae.oridb.org/). R-loop reference positions were collected from a published source. Hi-C interaction maps were collected from the GEO database (with accession numbers GSM1905067 and GSM1905068). All other genome features were acquired from the SGD database (https://downloads.yeastgenome.org/). Random control insertions were created based on the size and distribution of real insertions. Specifically, for each real insertion a corresponding control insertion of the same size with a random location and on the same chromosome was generated. For analysis of hotspots (multiple insertion donor DNA sites located at the minimum value of maximum value and (Q3 + 1.5 × interquartile range) IQR), and the lower whisker is located at the maximum of minimum value and (Q1 – 1.5 × IQR) (Q1 is first quartile, Q3 is third quartile, IQR is interquartile range). For Hi-C interaction map analysis (Extended Data Fig. 2), insertions of Ty retrotransposons, telomeres, rDNA and 2μ plasmid are excluded from analysis. A two-tailed Wilcoxon test was used to determine whether the frequency of interaction between the HO cleavage site and donor DNA sites is significantly different when compared to the interaction between the HO cleavage site and randomly selected loci. The randomization was repeated 1,000 times and the median P value was used to determine the significance of the difference.

**Code availability.** All codes used in this project are deposited at https://github.com/fagisX/FAID.

**Data availability**

All data supporting the findings of this study are available within the Letter. Sequences of all inserted DNA and sequences of the junctions analysed are provided in Supplementary Table 1. Source gel images are presented in Supplementary Figure 1.

31. Storici, F. & Resnick, M. A. The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzymol.* **409**, 329–345 (2006).
32. Lemos, B. R. et al. CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. *Proc. Natl Acad. Sci. USA* **115**, E2040–E2047 (2018).
33. Church, G. M. & Gilbert, W. Genomic sequencing. *Proc. Natl Acad. Sci. USA* **81**, 1991–1995 (1984).
34. Lee, B. S., Bi, L., Garfinkel, D. J. & Bailis, A. M. Nucleotide excision repair/TFIIH helicases RAD3 and SSL2 inhibit short-sequence recombination and Ty1 retrotransposition by similar mechanisms. *Mol. Cell. Biol.* **20**, 2436–2445 (2000).
35. Mayle, R. et al. DNA repair. *Mol. Cell. Biol.* **20**, 2436–2445 (2000).
36. Siow, C. C., Nieduszynska, S. R., Muller, C. A. & Nieduszynski, C. A. OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Res.* **40**, D682–D686 (2012).
37. Wahba, L., Costantino, L., Tan, F. J., Zimmer, A. & Koshland, D. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev.* **30**, 1327–1338 (2016).
Extended Data Fig. 1 | See next page for caption.
Extended Data Fig. 1 | Insertion analysis at $\textit{MATa}$, $\textit{URA3}$ and $\textit{LYS2}$ loci.

$\textbf{a}$, Experimental system to study insertions at DSBs and PCR analysis of $\textit{MATa}$ locus after DSB repair in wild-type and $\textit{pif1-m2 dna2}$ cells. Analysis was repeated more than 10 times (for gel source data, see Supplementary Figure 1).

$\textbf{b}$, Analysis of change of DSB ends among insertion events.

$\textbf{c}$, Schematic showing experimental system. A HO break is generated at an $\textit{ACT1}$ intron integrated in the $\textit{URA3}$ gene. Insertion of a DNA fragment or large deletion interferes with splicing and generates uracil auxotrophs.

$\textbf{d}$, Analysis of insertions by PCR and agarose gel electrophoresis at $\textit{URA3}$. The experiment was repeated more than three times with similar results. For gel source data, see Supplementary Figure 1.

$\textbf{e}$, Percentage of insertions among 5-FOA resistant colonies. Data are mean ± s.d.; $n = 3$ independent experiments; two-tailed t-test.

$\textbf{f}$, Analysis of origin of DNA inserted at DSB at $\textit{URA3}$ locus in indicated mutants, $n$ represents number of independent insertions from indicated mutants.

$\textbf{g}$, Percentage of insertions among 5-FOA resistant colonies after transient induction of HO break in $\textit{rad51 pif1-m2 dna2}$. Data are mean ± s.d.; $n = 3$ independent experiments.

$\textbf{h}$, Schematic showing experimental system to follow insertions at CRISPR–Cas9-induced DSBs within the $\textit{LYS2}$ locus. Below, percentage of insertions among cells maintaining CRISPR–Cas9 and analysis of origin of DNA inserted at $\textit{LYS2}$. $n$ represents number of independent insertions sequenced in $\textit{pif1-m2 dna2}$ cells. The experiment was repeated four times with similar results.
Extended Data Fig. 2 | Origin of inserted DNA at DSBs. a, Each triangle indicates a single insertion donor DNA; hotspots of insertion donor DNA are marked in red. b, Scatter plot of chromosome size and insertion number. n = 468 independent insertions. Correlation coefficients were calculated based on the Spearman method. c, Contact analysis between MAT locus on chromosome III and loci from which DNA was inserted. For each replicate, 1,000 random sets of DNAs (equal size and number) are compared to experimental inserted DNA. P values are determined by two-tailed Wilcoxon test; n = 358 independent inserted DNAs used for contact analysis.
Extended Data Fig. 3 | Analysis of transposon cDNA stability and Ty1 expression. a, Analysis of Ty1 cDNA stability. The experiment was repeated four times with similar results. b, Analysis of Ty1 expression and its quantification. Data are mean ± s.d. from three independent experiments. For gel source data, see Supplementary Figure 1.
Extended Data Fig. 4 | Hotspots of origin of inserted DNAs. Position of DNAs inserted within DSBs is indicated in red. Blue boxes, origins of replication; yellow circles, centromeres; green boxes, telomeres; open boxes, genes. Hotspots were defined as loci that are the source of at least two inserted DNA fragments separated from each other by no more than 3 kb.
Extended Data Fig. 5 | Genetic interactions between Dna2 and Rad52, Yen1 and Mus81. a, Overall frequency and analysis of origin of DNA inserted at DSB in indicated mutants. n represents the number of independent insertions analysed by sequencing. b, Origin of insertions in rad52Δ mutant cells. c, Active Yen1 rescues non-viability of dna2Δ cells. Tetrad dissection of PIF1/pif1-m2 YEN1/yen1ON DNA2/dna2Δ triple heterozygotes is shown. The experiment was repeated twice. d, Analysis of complex insertions (2 or more DNA fragments inserted at DSB) in Dna2-deficient mutants. Sample size, defined as the number of independent insertions analysed for each mutant is presented in Extended Data Table 1. \( \chi^2 \) test is used to determine the P value. e, DNA damage sensitivity analysis (spot assay, 5× dilution) in indicated mutants. The experiment was repeated twice.
Extended Data Fig. 6 | Model of large insertions at DSBs in Dna2-deficient cells. a, Unprocessed 5′ flaps are processed by alternative nuclease or displaced by synthesis leading to release of over-replicated DNA fragments. b, Stalled and reversed forks, when approached by a converging fork, leave over-replicated DNA that can be released by processing by other nucleases. c, ssDNA can be inserted into DSBs by NHEJ and Pol4.
Extended Data Fig. 7 | Analysis of insertions of transformed DNA at DSBs and analysis of free, short DNA in cells. 

**a**, Analysis of insertions of transformed DNA at DSBs in wild-type and indicated mutant cells. Schematic of the experiment (left) and percentage of cells carrying insertion (right). χ² test was used to determine the P values; n = 160 for dsDNA and n = 320 for ssDNA, and represents the number of colonies tested for the presence of insertion.  

**b, c**, Analysis of inserted DNA after transformation of dsDNA (b) and ssDNA (c). d, Quantitative PCR analysis of short free DNA in indicated mutants. Data are mean ± s.d.; n = 3 independent experiments. Position of the primers used is shown at the top and fold change in DNA amount is shown on the bottom.

© 2018 Springer Nature Limited. All rights reserved.
Extended Data Table 1 | Analysis of NHEJ efficiency and insertion frequency

| Genotype                        | Number of insertions at MATa (%)* | NHEJ efficiency** | # of sequenced events | # of inserts *** | # of complex events*** | transposon fragment insertions (%) *** | rDNA fragment insertions (%) *** | other nuclear genome insertions (%)*** |
|---------------------------------|----------------------------------|-------------------|-----------------------|-----------------|------------------------|----------------------------------------|----------------------------------|---------------------------------------|
| WT                              | 0% (0/644)                      | 0.104±0.058%      | N/A                   |                 |                        |                                        |                                  |                                       |
| pif1-m2                         | 0% (0/160)                      | 0.061±0.013%      | N/A                   |                 |                        |                                        |                                  |                                       |
| pif1-m2 dna2                    | 8.2% (148/1794)                 | 0.105±0.029%      | 142                   | 161             | 18                     | 8.1%                                   | 14.9%                            | 77.0%                                |
| pif1-m2 dna2 R1253Q              | 0% (0/160)                      | 0.116±0.099%      | N/A                   |                 |                        |                                        |                                  |                                       |
| pif1-m2 dna2 E675A              | 15.6% (50/320)                  | 0.123±0.059%      | 3                    |                 |                        |                                        |                                  |                                       |
| rad51 pif1-m2 dna2              | 9.4% (30/320)                   | 0.096±0.020%      | 13                   |                 |                        |                                        |                                  |                                       |
| rad51                           | 0% (0/320)                      | 0.079±0.001%      | N/A                   |                 |                        |                                        |                                  |                                       |
| rad52 pif1-m2 dna2              | 36.1% (52/144)                  | 0.228±0.091%      | 33                   | 43              | 8                      | 4.7%                                   | 25.6%                            | 69.7%                                |
| rad52                           | 0.9% (3/320)                    | 0.167±0.037%      | 3                    |                 |                        |                                        |                                  |                                       |
| rad27                           | 0% (0/320)                      | 0.193±0.165%      | N/A                   |                 |                        |                                        |                                  |                                       |
| yku70 pif1-m2 dna2              | 0% (0/160)                      | 0.011±0.009%      | N/A                   |                 |                        |                                        |                                  |                                       |
| lig4 pif1-m2 dna2               | 0.2% (1/480)                    | 0.007±0.004%      | 1                    |                 |                        |                                        |                                  |                                       |
| pol4 pif1-m2 dna2               | 0% (0/160)                      | 0.014±0.005%      | N/A                   |                 |                        |                                        |                                  |                                       |
| spo3 pif1-m2 dna2               | 6.3% (47/746)                   | 0.117±0.055%      | 47                   | 55              | 7                      | 3.6%                                   | 18.2%                            | 78.2%                                |
| fob1 pif1-m2 dna2               | 6.5% (107/1646)                 | 0.077±0.049%      | 105                  | 118             | 11                     | 9.3%                                   | 16.1%                            | 74.6%                                |
| pif1 pol32 dna2                 | 5.5% (29/527)                   | 0.075±0.023%      | 29                   | 33              | 4                      | 9.1%                                   | 18.2%                            | 72.7%                                |
| slx1 pif1-m2 dna2               | 8.6% (26/304)                   | 0.069±0.009%      | 5                    |                 |                        |                                        |                                  |                                       |
| mus81 pif1-m2 dna2              | 3.6% (16/448)                   | 0.081±0.013%      | 10                   |                 |                        |                                        |                                  |                                       |
| mus81 pif1-m2 dna2 E675A        | 5.0% (8/160)                    | N/D               | N/D                   |                 |                        |                                        |                                  |                                       |
| sgs1                            | 0% (0/240)                      | 0.088±0.012%      | N/A                   |                 |                        |                                        |                                  |                                       |
| exo1                            | 0% (0/240)                      | 0.104±0.023%      | N/A                   |                 |                        |                                        |                                  |                                       |
| yen1<sup>DN</sup>               | 0% (0/160)                      | 0.044±0.005%      | N/A                   |                 |                        |                                        |                                  |                                       |
| yen1<sup>DN</sup> pif1-m2 dna2  | 13.9% (30/216)                  | 0.122±0.043%      | 30                   | 45              | 11                     | 13.3%                                  | 6.7%                             | 80.0%                                |
| yen1<sup>DN</sup> pif1-m2 dna2  | 12.0% (24/200)                  | 0.075±0.015%      | 20                   |                 |                        |                                        |                                  |                                       |
| pif1 dna2 + HU                  | 23.0% (70/304)                  | 0.057±0.026%      | N/D                   |                 |                        |                                        |                                  |                                       |

Table 1b

| Genotype                        | Number of insertions at URA3 among 5-FOA resistant colonies (%) | # of sequenced events | # of inserts | # of complex events | transposon insertions (%) | rDNA fragment insertions (%) | other nuclear genome insertions (%) |
|---------------------------------|-----------------------------------------------------------------|-----------------------|--------------|---------------------|----------------------------|-----------------------------|-----------------------------------|
| rad51                           | 3.8% (5/133)                                                   | 4                     | 4             | 0                   | 100%                       | 0%                          | 0%                                |
| rad51 pif1-m2                   | 2.6% (4/154)                                                   | 3                     | 3             | 0                   | 100%                       | 0%                          | 0%                                |
| rad51 pif1-m2 dna2              | 32.5% (96/295)                                                 | 21                    | 23            | 2                   | 4.3%                       | 26.1%                       | 69.6%                             |

Table 1c

| Genotype                        | Number of insertions at LYS2 (%) | NHEJ efficiency** | # of sequenced events | # of inserts | # of complex events | transposon insertions (%) | rDNA fragment insertions (%) | other nuclear genome insertions (%) |
|---------------------------------|----------------------------------|-------------------|-----------------------|--------------|---------------------|----------------------------|-----------------------------|-----------------------------------|
| WT                              | 0% (0/240)                      | 0.828±0.043%      | N/A                   |              |                     |                            |                            |                                   |
| pif1-m2 dna2                    | 2.9% (7/240)                    | 0.288±0.058%      | 7                     | 8            | 1                   | 37.5%                      | 12.5%                       | 50%                               |

* MATa locus. ** URA3 locus. *** LYS2 locus. * Number of independent insertions and number of colonies tested for presence of insertion are shown; ** data are mean ± s.d.; n ≥ 3 and represents the number of independent experiments; *** the number is shown only for mutants for which at least 25 cases were sequenced.
## Extended Data Table 2 | Sequence analysis of DSB repair survivors in wild type and indicated mutants that do not carry a large insertion

| Mutation             | WT  | *pif1-m2 dna2* | *rad27* |
|----------------------|-----|---------------|---------|
| CGCAACA(+CA)GTA      | 13.3 (6/45) | 26.6 (17/64) | 17.1 (7/41) |
| CGC(+A)AACAGTA       | 4.4 (2/45)  | 1.6 (1/64)   | 4.9 (2/41)  |
| CGCAACA(+ACA)GTA     | 2.2 (1/45)  | 0             | 2.4 (1/41)  |
| CGCAA(+AA)CAGTA      | 0    | 3.1 (2/64)   | 7.3 (3/41)  |
| CGCAA(+C)AGTA        | 0    | 0             | 2.4 (1/41)  |
| CGCA(+C)ACAGTA       | 0    | 0             | 2.4 (1/41)  |
| CGCA(-ACA)GTA        | 44.4 (20/45) | 26.6 (17/64) | 17.1 (7/41) |
| CGCAA(-CA)GTA        | 13.3 (6/45) | 14.1 (9/64)  | 22.0 (9/41) |
| CGC(-A)ACAGTA        | 6.7 (3/45)  | 10.9 (7/64)  | 19.5 (8/41) |
| CGCAACAG(-T)A        | 2.2 (1/45)  | 0             | 0         |
| C(-GCA)ACAGTA        | 2.2 (1/45)  | 0             | 0         |
| CGCAAC(-AGT)A        | 2.2 (1/45)  | 0             | 0         |
| CGCAAC(-A)GTA        | 0    | 1.6 (1/64)   | 0         |
| CGC(-AA)CAGTA        | 0    | 1.6 (1/64)   | 0         |
| CGCAA(-CAG)TA        | 0    | 3.1 (2/64)   | 0         |
| CGCAA(-C)AGTA        | 0    | 0             | 2.4 (1/41) |
| CGCAA(-CAGT+A)A      | 2.2 (1/45)  | 0             | 0         |
| CGCAA(-CAG+AT)A      | 2.2 (1/45)  | 0             | 0         |
| CTCAACAGTA           | 0    | 1.6 (1/64)   | 0         |
| GCAACAGTA            | 0    | 1.6 (1/64)   | 0         |
| > 4 bp deletion      | 4.4 (2/45)  | 7.8 (5/64)   | 2.4 (1/41) |
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- [ ] The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- [ ] An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- [ ] The statistical test(s) used AND whether they are one- or two-sided
- [ ] A description of all covariates tested
- [ ] A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- [ ] A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- [ ] For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
- [ ] Give P values as exact values whenever suitable.
- [ ] For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- [ ] For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- [ ] Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
- [ ] Clearly defined error bars
- [ ] State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

- All yeast genome features are downloaded from SGD database.

- All custom codes are developed by Python with numpy, scipy, pandas. Statistical test is performed by using function within Scipy package. They are deposited into "https://github.com/fagisX/FAID".
- Commercial or free software used:
  - ApE - A plasmid Editor v2.0.53c by Wayne Davis
  - Prism 7.0d by GraphPad Software
  - ImageQuant TL 7.0 by GE
  - SnapGene version 4.2.6 by GSL Biotech LLC

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw data in a form of DNA sequences of insertion are provided within Supplemental Table 1 - excel file. They are grouped by genotype except the last page that has multiple strains but with smaller number of events per genotype. Thus anybody who wishes to repeat the analysis could upload the sequences from Excel file.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The exact number of insertion events and donor DNAs per genotype is provided in Extended Data Table 1. The major observation in the manuscript relates to the frequency of insertions among cells that repaired a DSB. Insertions are not observed in wild type (0/644) and are observed in a dna2 mutant (8%, 148/1794). The sample size (number of colonies tested) that would determine whether derivatives of dna2 mutant or other mutants show similar or significantly different frequency of insertions was determined to be 113; and the lowest sample size tested here is 144. In order to ensure that the 95% confidence interval estimate with 5% margin of error, n = p x (1-p) x (Z/E)^2, where Z for 95% confidence interval is 1.96, E for 5% margin of error is 0.05, p for the proportion of insertions observed in dna2 mutant is 0.08 (148/1794). The estimates of the necessary sample size is 113 to assure a sufficient statistical power to detect a significant result.

Data exclusions

All of the sequences of inserted DNA are provided in Supplemental Table 1. Analysis of the insertion features was done with all events from all dna2 mutants (Fig 1b, 1c, 1e and extended Fig. 1b) except:
- In Extended Data Fig. 2a, 2b, nearly all events are presented. The exceptions are insertions coming from telomeres and transposons because these are repetitive elements and we don’t have exact loci of the origin of inserted DNA. Also 2 micron plasmid insertion events are not shown in Extended Data Fig. 2a, 2b because it is not part of chromosome. These are shown separately in figure 3c.
- All transposon insertion events from all dna2 mutants are presented in Figure 2a and 2b.
- All rDNA insertion events from all dna2 mutants are shown in Figure 3a.
- Insertions from pif1-m2 dna2 rad52 mutant and Cas9 induced insertions were not included in any statistical analysis. These data constitute only a small fraction of all events analyzed and were added at revision step to address specific reviewer questions that are not related to global features of insertions.
- In figures 4a, 4b, and 4c left, all insertions were analyzed with exception of repetitive elements (rDNA insertions, transposons insertions, telomere insertions). Reason - we don’t know exact loci of the origin of inserted repetitive DNA. Also 2 micron plasmid insertion events were not included as the plasmid is not a part of chromosome.
- In figure 4c right, all insertions were analyzed with exception of repetitive elements (rDNA insertions, transposons insertions) and non-chromosomal 2 micron plasmid insertions.

In Extended Data Figure 2c, all insertions at MATa locus were analyzed with exception of rDNA insertions, transposons insertions, 2 micron plasmid insertions and telomere. Reason - we don’t know exact loci of the origin of inserted repetitive DNA and 2 micron is extrachromosomal DNA.

For all bar graphs that show insertion number per mutant strain (1a, 4d) the exact number of colonies tested is shown in Extended Data Table 1. These numbers were used to calculate p values to determine significant differences between mutant strains (chi square, confidence interval 95%).

Replication

We confirmed the major discovery, insertions of DNA fragments in dna2 mutant strains in several ways. First, we observed large insertions in all 3 independent pif1-m2 dna2 strains at MATa locus. Second, we observed this phenotype in yen1ON dna2 strain at HO site at MATa locus. Third, we observed this phenotype in pif1-m2 dna2 strain at HO induced DSB at URA3 locus. Fourth, we observed this phenotype in pif1-m2 dna2 strain at Cas9 induced DSB at LYS2 locus.

To measure the NHEJ efficiency, we repeated the experiment at least three times for each mutant.
Analysis of free DNA, analysis of Ty’s cDNA amount and its stability and RNA amount of Ty1 was repeated 3 times or more. The rate of spontaneous transposition was measured as previously described (Fig 2d), brief description is provided in methods section.

Randomization

To test by PCR the presence of insertions at DSB after repair by nonhomologous end joining we screened all colonies or random colonies grown on YP-GAL plates.

Blinding

All or random colonies from the plate that represent survivors of DSB repair were analyzed. Colonies that carry insertion at MATa or LYS2 loci are not distinguishable from colonies that do not have insertions at DNA break and therefore can not be selected for.
### Reporting for specific materials, systems and methods

| Materials & experimental systems | Methods |
|----------------------------------|---------|
| Involved in the study            | n/a     |
| Unique biological materials      |         |
| Antibodies                       |         |
| Eukaryotic cell lines            |         |
| Palaeontology                    |         |
| Animals and other organisms      |         |
| Human research participants      |         |

| Involved in the study            |
|---|
| ChIP-seq                          |
| Flow cytometry                    |
| MRI-based neuroimaging            |