LETTER

https://doi.org/10.1038/s41586-019-0887-y

Heterozygous mutations cause genetic instability in a yeast model of cancer evolution

Miguel C. Coelho[1,2,3]*, Ricardo M. Pinto[3,4] & Andrew W. Murray[1,2]*

Genetic instability, a heritable increase in the rate of genetic mutation, accelerates evolutionary adaptation[1] and is widespread in cancer[2,3]. In mammals, instability can arise from damage to both copies of genes involved in DNA metabolism and cell cycle regulation[4] or from inactivation of one copy of a gene whose product is present in limiting amounts (haploinsufficiency[5]); however, it has proved difficult to determine the relative importance of these two mechanisms. In *Escherichia coli*[6], the application of repeated, strong selection enriches for genetic instability. Here we have used this approach to evolve genetic instability in diploid cells of the budding yeast *Saccharomyces cerevisiae*, and have isolated clones with increased rates of point mutation, mitotic recombination, and chromosome loss. We identified candidate, heterozygous, instability-causing mutations; engineering these mutations, as heterozygotes, into the ancestral diploid strain caused genetic instability. Mutations that inactivated one copy of haploinsufficient genes were more common than those that dominantly altered the function of the mutated gene copy. The mutated genes were enriched for genes functioning in transport, protein quality control, and DNA metabolism, and have revealed new targets for genetic instability[7–11], including essential genes. Although only a minority (10 out of 57 genes with orthologues or close homologues) of the targets we identified have homologous human genes that have been implicated in cancer[2], the remainder are candidates to contribute to human genetic instability. To test this hypothesis, we inactivated six examples in a near-haploid human cell line; five of these mutations increased instability. We conclude that single genetic events cause genetic instability in diploid yeast cells, and propose that similar, heterozygous mutations in mammalian homologues initiate genetic instability in cancer.

Cancer requires multiple changes in cellular properties, and the accumulation of the mutations that produce these changes can be accelerated by genetic instability. The origins and mechanisms of genetic instability are difficult to study in cancer, making it hard to assess the relative importance of evolutionary trajectories that inactivate both copies of a tumour suppressor gene (Knudson's two-hit model[12]) and those that mutate only one of two copies of a gene, exposing haploinsufficiency[5] or dominantly altering a protein's function. Sequencing the genomes of human tumours reveals mutations that are likely to cause genetic instability[2,13–18], but cannot easily reveal whether these mutations are homozygous or heterozygous. By directly selecting for genetic instability in a more experimentally tractable organism, budding yeast, we assessed the relative importance of the one- and two-hit routes to genetic instability.

Studies on model organisms have primarily sought recessive mutations[7–11] that cause genetic instability. To avoid this restriction, we evolved genetic instability in diploid budding yeast by selecting for sequential inactivation of three sets of growth suppressor genes: genes that can prevent cell proliferation under certain conditions. Because these mutants are recessive, diploids should inactivate both copies of a growth suppressor gene to survive. Formally, the genes that mutate to cause instability and the growth suppressor genes are analogous to two classes of tumour suppressor genes: instability-causing mutations identify genes analogous to tumour suppressors that guard the stability of the genome, such as *TP53*, whereas growth suppressor genes are analogous to tumour suppressor genes that regulate cell growth and proliferation, such as the retinoblastoma gene *RB1*. By selecting for sequential inactivation of growth suppressor genes, we selected for genetic instability; by using random mutagenesis, we avoided the restrictions imposed by systematic collections of deletion mutants; and by using diploids containing homozygous, dominant growth suppressors, we mimicked the selection to inactivate tumour suppressors in cancer evolution.

We searched for genetic instability by selecting clones that survived and proliferated after sequential exposure of diploid yeast to three drugs that suppress growth of wild-type cells (Fig. 1a). We used a reporter cassette containing the *URA3* gene to monitor the level and type of genetic instability in diploid strains (Fig. 1b), distinguishing three classes of mutation: point mutation (local mutations that affect the *URA3* gene), mitotic recombination (the sum of mitotic recombination and chromosome truncation), and chromosome loss (Fig. 1c). We tested the reporter using environmental[19] (Extended Data Fig. 1a) or genetic[10] (Extended Data Fig. 1b) perturbations that cause different types of genetic instability. The triple selection selected more strongly for genetic instability than any of the single selection steps (Extended Data Fig. 2a).

Starting from mutagenized populations, we selected and characterized 89 genetically unstable diploid clones (Fig. 1b, Supplementary Table 1). All except one showed an increase in at least one class of genetic instability (Fig. 1d), 44 showed an increase in two of the three classes of mutation, 27 showed an increase in all three, and 46 showed at least a tenfold elevation in the rate of one class of mutation. Increases in chromosome loss and mitotic recombination were correlated within individual clones (Extended Data Fig. 2b) and all three mutation classes were correlated with the global mutation rate (Extended Data Fig. 2c). We conclude that selecting for growth suppressor inactivation selects for genetic instability, mimicking an important feature of cancer evolution.

We investigated how many events were necessary to trigger genetic instability by examining genetic instability in the four haploid products of meiosis from our diploid, evolved clones; we made the assumption that most mutations that increase genetic instability in diploid strains would also cause instability in haploids (Extended Data Fig. 3a). In 59 out of 61 clones we tested, some of the haploid spores were genetically unstable. We first assessed the minimum number of mutations required for instability (Fig. 2a). In most clones (41 of 59), genetic instability segregated as expected if a single heterozygous mutation was enough to cause instability; in the remaining clones, instability appeared to require the simultaneous presence of two or more heterozygous mutations (16 of 59) or a single homozygous mutation (2 of 59). We also tested which fraction of the spores derived from heterozygous clones had the maximum genetic instability (Extended Data Fig. 3b): 29 of 57 clones

[1]FAS Center for Systems Biology, Harvard University, Cambridge, MA, USA. [2]Molecular and Cell Biology, Harvard University, Cambridge, MA, USA. [3]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [4]Department of Neurology, Harvard Medical School, Boston, MA, USA. *e-mail: mcc150@gmail.com; awm@mcb.harvard.edu

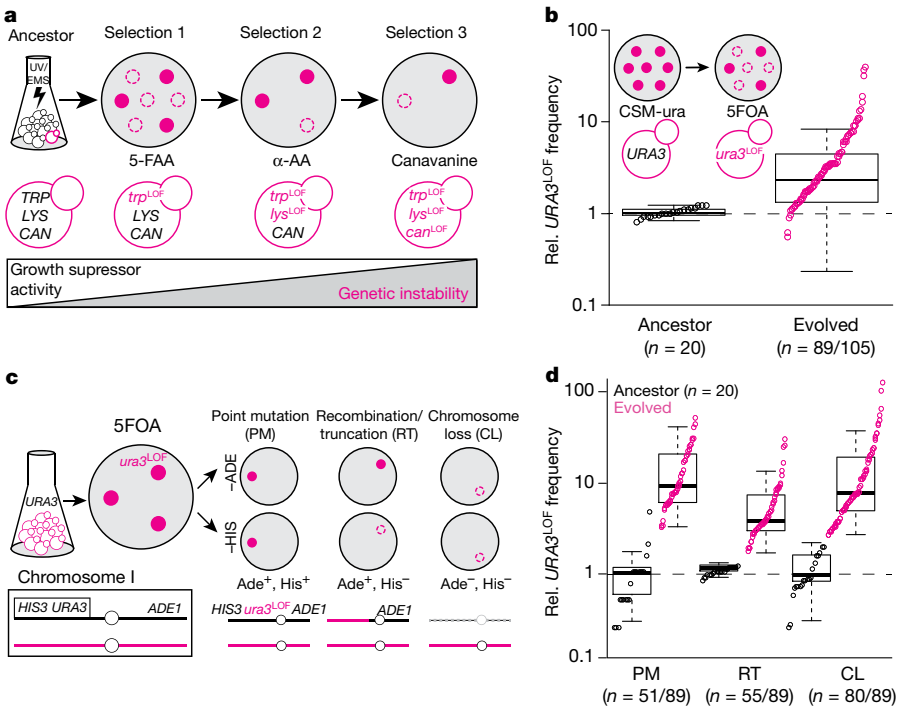


Fig. 1 | Sequential inactivation of growth suppressors selects for the evolution of genetic instability. a, Selection of genetically unstable clones. Clones derived from mutagenized diploid cells survived selection by sequentially disrupting three sets of growth suppressor genes: any one of the five TRP genes (*TRP1*–*TRP5*) (selection 1, growth on 5-fluoroanthranilate (5-FAA)); either *LYS2* or *LYS5* (selection 2, growth on α-aminoadipate (α-AA)); and *CAN1* (selection 3, growth on canavanine). **b**, The frequency of mutation of a single copy of *URA3* (*URA3*^{LOF}) was elevated in evolved diploid clones when compared to ancestor (one-sided Student's *t*-test comparing ancestor and evolved data sets collectively, $P = 3.9 \times 10^{-8}$) growth on 5-fluoroorotic acid (5FOA) (see Methods). Points shown for 89 clones that showed elevation in any one of these mutation rates: global, point mutation, mitotic recombination/chromosome truncation, or chromosome loss. CSM-ura, complete supplement mixture lacking uracil. **c**, Genetic instability reporter in diploid strains consisting of three single-copy genes on chromosome I: *URA3* coupled to the *HIS3* gene 70 kb from the left telomere, and the *ADE1* gene on the opposite side of the centromere (*URA3*, *HIS3* and *ADE1* are inactivated elsewhere in the genome). This architecture allows us to distinguish three different *URA3* inactivation modes in 5FOA-resistant diploids: point mutation (His$^+$ and Ade$^+$), mitotic recombination or loss of the left arm of chromosome I (His$^-$ and Ade$^+$), and chromosome loss (His$^-$ and Ade$^-$). **d**, Frequencies of 5FOA-resistant point mutation (PM), recombination/truncation (RT) and complete chromosome loss (CL) events were significantly elevated in evolved clones (one-sided Student's *t*-test; $P_{PM} = 3.2 \times 10^{-9}$, $P_{RT} = 1.0 \times 10^{-6}$, $P_{CL} = 4.1 \times 10^{-8}$). Values above the ancestor are shown. Mean ± s.d. elevations for the different mutation rates: global = 5.2 ± 6.7, PM = 8.0 ± 10.1, RT = 3.4 ± 4.4, CL = 13.5 ± 20.2. *URA3* inactivation rate (per cell per generation) in ancestor: global = 2.4×10^{-5}, PM = 2.0×10^{-7}, RT = 2.0×10^{-5}, CL = 4.1×10^{-6}. In box plots, centre line is median, upper and lower limits are quartiles and whiskers show 1.5× interquartile range.

contained only one mutation that affected genetic instability, 17 of 57 showed the 1:3 pattern expected if they contained two mutations that affected genetic instability, and 11 of 57 showed the 1:7 pattern expected if they contained three mutations that affected genetic instability.

To identify mutations that correlate with genetic instability, we performed whole-genome sequencing of the 89 evolved clones. For genes that were mutated in four or more independent lineages (25 genes), or carried nonsense mutations (10 genes), we deleted one copy in the ancestor and found this sufficed to cause an increase in genetic instability (Fig. 2b, Extended Data Fig. 4a–d).

To identify less frequent mutations that correlated with genetic instability, we performed bulk segregant analysis[20,21] on the 48 diploid clones that exhibited the highest genetic instability: we pooled and whole-genome sequenced genetically unstable spores, arguing that mutations that caused genetic instability should be at high frequency in these pools, whereas other, non-causative mutations should show a mean frequency of 0.5. Thirty-nine clones had one to three heterozygous alleles with frequencies above the cutoff frequency of 0.67 in the selected, genetically unstable pools (Extended Data Fig. 5a). We identified 74 candidate genes that had mutated to cause genetic instability (Extended Data Fig. 5b, Supplementary Table 2). On the basis of the frequency with which different candidate genes were mutated, we estimate that there are roughly 150 genes that can mutate to produce dominant genetic instability in diploids. Seven genes were identified both as frequently mutated and by bulk segregant analysis.

We were surprised at the prevalence of heterozygous mutations that segregated with genetic instability. We analysed seventeen candidate heterozygous, genetic-instability-causing mutations in more detail. We engineered each mutation, as a heterozygote (mut/+), into the ancestral diploid and measured its effect on mutation rate. For all seventeen genes, engineering one copy of the mutant allele into a diploid strain increased the frequency of at least one form of genetic instability by 2-fold to 105-fold (Fig. 2c, Extended Data Fig. 6a–c).

By comparing the phenotypes of strains with one mutant and one wild-type allele (mut/+) with the phenotypes of strains with one deleted and one wild-type copy of the gene (Δ/+), we can distinguish two ways in which the mutant allele causes genetic instability. The first is haploinsufficiency: if the amount of the gene product is limiting in a diploid cell, removing one copy will cause genetic instability and the phenotypes of the mut/+ and Δ/+ strains should be similar (Fig. 2d, top). In the second, the mutant copy of the gene actively interferes with chromosome metabolism, which predicts that the phenotype of the mut/+ strain should be stronger than that of the Δ/+ strain (Fig. 2d, middle). For 6 out of 17 genes, at least one form of instability exceeded that of the wild-type strain by at least two standard deviations in mut/+ but not in the Δ/+ analogue, indicating that dominant mutations are altering protein function (Fig. 2c, Extended Data Fig. 6a–c).

We tested the effect of restoring the wild-type version of a mutated allele in three types of genetically unstable strain: evolved clones, strains engineered to contain one copy of a mutation shown to cause genetic

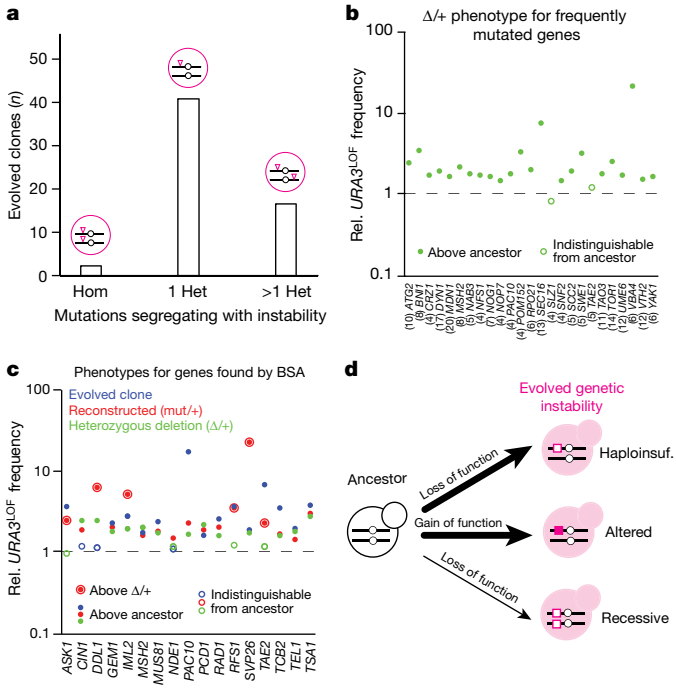


Fig. 2 | A single heterozygous mutation causes genetic instability in diploid cells. a, Diagnosing the genetic basis of instability by following its meiotic segregation. Evolved diploid clones were classified as showing the segregation pattern corresponding to a homozygous (Hom) mutation, a single heterozygous mutation (1 Het) or multiple heterozygous mutations (>1 Het). **b**, Global mutation frequency for heterozygous deletions of each of 25 frequently mutated genes in our evolved diploid clones (mutated in at least four independent clones). Numbers in parentheses show the number of lineages mutated for a given gene. Although heterozygous deletions of *SLZ1* and *TAE2* (also known as *RQC2*) have a global mutation frequency similar to the wild type, they have elevated point mutation and recombination/truncation rates, respectively (see Extended Data Fig. 4a–c). **c**, Analysis of mutated genes identified from bulk segregant analysis. Global mutation frequencies for evolved clones (blue), reconstructed clones where the mutation from the evolved clone was introduced into the ancestor as a heterozygote (mut/+, red), or the heterozygous deletion of the corresponding gene from the ancestor (Δ/+, green). **b**, **c**, Dotted lines, ancestor; filled circles, strains whose instability exceeds that of the ancestor by two s.d. in three biologically independent experiments; outer circles (**c**), cases in which the heterozygous mutation had higher instability than the heterozygous deletion. We also tested a putative causative mutation in *PAC10*, a gene that was frequently mutated. **d**, Scheme representing frequent events (thick arrows) that cause genetic instability: a single heterozygous mutation can cause loss of function (haploinsufficiency) or a dominantly altered function (altered). Homozygous mutations are rare (thin arrow) and cause instability through loss-of-function mutations in genes that are not haploinsufficient (recessive).

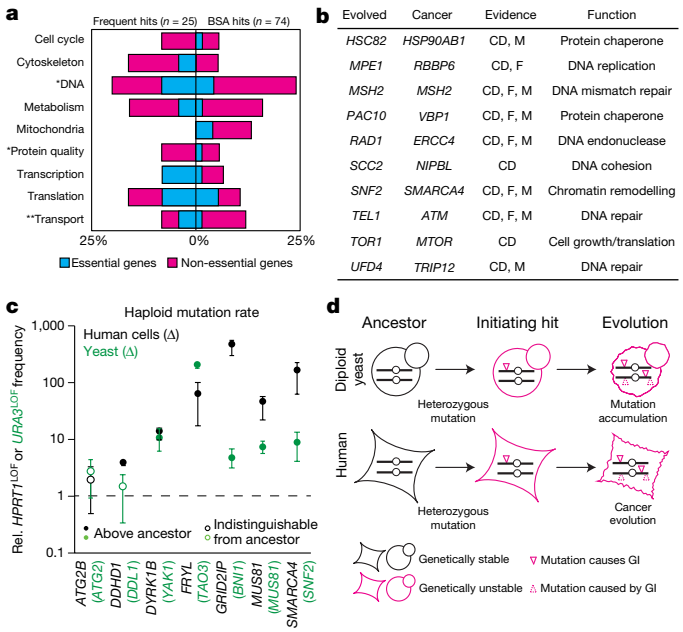


Fig. 3 | Genes selected to cause instability in yeast target different functional classes and homologue inactivation triggers instability in human cells. a, Frequency of functional classes for genes in which we identified mutations that caused genetic instability (left, frequent hits; right, hits enriched in bulk segregant analysis compared to the total number of genes in a given functional category). Seven genes were present in both groups (*ATG2*, *MSH2*, *NFS1*, *NOP7*, *POM152*, *SEC16* and *TAE2*); probabilities were calculated from the binomial distribution. $^{**}P < 0.05$ with Bonferroni correction (ten hypotheses tested); *significant at 2.5% false discovery rate using the Benjamini–Hochberg procedure. **b**, Genes in which we identified mutations that cause genetic instability (either mut/+ or Δ/+ are genetically unstable) and their human homologues that have been implicated in cancer in one of three ways: hereditary cancer predisposition (F), mouse models (M), or the frequency with which they are mutated in cancer patients (cancer driver, CD). **c**, Frequency of *HPRT1* mutation in HAP1 cell lines that have undergone CRISPR-mediated knockout of homologues of yeast instability genes (black dots), compared to frequency of *URA3* mutations in yeast cells with the corresponding homologues deleted (green dots). Mean ± s.d.; closed circles, mutation rate greater than wild type (exceeds that of ancestor by 2 s.d.); $n = 3$ biologically independent experiments. **d**, Scheme comparing evolution of genetic instability in yeast and cancer evolution. Single heterozygous mutations can initiate genetic instability (GI) and accelerate the accumulation of further mutations that alter cellular properties or increase genetic instability.

instability, and strains that contained heterozygous mutations of genes that are haploinsufficient for genetic stability[22]. For each type of strain, we tested three different genes; for eight out of the nine genes tested, restoring homozygosity for the wild-type allele reduced genetic instability, supporting the argument that a single heterozygous mutation is sufficient to cause instability (Extended Data Fig. 6d).

The 92 genes we identified (25 were present in four or more evolved lineages and 74 were identified by bulk segregant analysis, while 7 of these overlap) are associated with nine high-level biological processes (Fig. 3a, Extended Data Fig. 7a): genes associated with transport are significantly enriched, even after Bonferroni correction, and genes associated with protein quality control and DNA metabolism pass a 2.5% false discovery cutoff. Twenty-two per cent of the candidate genes we identified as targets for genetic instability are essential (Supplementary Table 2). There was little overlap between genes identified in deletion screens and our evolution screen (Extended Data Fig. 7b, Supplementary Table 3). Of the yeast genes that we identified, 54 have human orthologues and 3 have close human homologues, and 10 of these 57 genes have been implicated in cancer[2,13–18] (Fig. 3b, Extended Data Fig. 7c–f, Supplementary Table 3).

We investigated whether inactivation of six genes whose homologues lacked known associations with cancer caused genetic instability in a near-haploid human cell line (HAP1)[23]. Two of these yeast genes (*MUS81* and *TAO3*) were identified both in our experiments and in previous systematic screens for genetic instability in yeast, and four were identified only in this study. The products of the human homologue genes participate in recombination (*MUS81*), metabolism (*DYRK1B*), mitochondria (*DDHD1*), the cytoskeleton (*FRYL*, *GRID2IP*), and autophagy (*ATG2B*) in human cells. To monitor genetic instability, we inactivated the hypoxanthine-guanosine phosphoribosyl transferase gene (*HPRT1*) by exposing cells to 6-thioguanine (6TG, Extended Data Fig. 8a). As a control, we inactivated *SMARCA4*, which has been implicated in human cancer and whose mutation has been shown to cause genetic instability[24], and which is a homologue of *SNF2*, a yeast gene that is involved in chromatin

remodelling. Inactivation of *SMARCA4* resulted in a 150-fold increase in mutation rate relative to wild-type cells (Fig. 3c). Individual inactivation of all but one of the other six genes caused a 4- to 440-fold elevation in mutation rate in human cells (Fig. 3c, Extended Data Fig. 8b, c). Inactivation of human homologues of genes identified by selecting for genetic instability in diploid yeast cells causes genetic instability, which suggests that mutations in these genes might contribute to human cancer.

We found many mutations that cause genetic instability as heterozygotes, demonstrating that a single mutational event can accelerate evolutionary adaptation in diploid cells and that strong, directional selection can select for such mutations. The 92 genes we identified as targets for instability include 21 genes involved in DNA and chromosome metabolism; alterations in the functions of these genes are expected to cause genetic instability. The remainder have roles in other essential aspects of cellular behaviour, suggesting that perturbations in these pathways interfere with DNA and chromosome metabolism enough to cause genetic instability. In both yeast and bacteria, strong directional selection leads to the evolution of genetic instability[6,25]. In yeast, selecting for genetic instability yields one-hit, heterozygous, instability-causing mutations much more often than two-hit, homozygous mutations. We argue that the selection for multiple changes in cellular properties that occurs in cancer also selects for genetic instability and that, as they do in yeast, heterozygous mutations can lead to genetic instability. Both in yeast and in cancer, further selection could enhance the initial instability of heterozygous mutations by eliminating the remaining wild-type copy of the genes, or by selecting additional heterozygous, dominant mutations in other genes (Fig. 3d).

Of the 57 genes with human homologues that we identified, 10 have been implicated in human cancer. We hypothesize that the remaining 47 that have close human homologues are candidates for genes in which mutations cause genetic instability in cancer. For five of the six genes we tested, elimination of these genes triggered genetic instability, which suggests that mutations in many of the human homologues of the genes that we identified could play a role in the initiation, progression, and resistance to therapy of human cancer.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-019-0887-y.

Received: 23 December 2017; Accepted: 18 December 2018;
Published online: 30 January 2019

1. Swings, T. et al. Adaptive tuning of mutation rates allows fast response to lethal stress in *Escherichia coli*. *eLife* **6**, e22939 (2017).
2. Lawrence, M. S. et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
3. Drost, J. et al. Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43–47 (2015).
4. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
5. Inoue, K. & Fry, E. A. Haploinsufficient tumor suppressor genes. *Adv. Med. Biol.* **118**, 83–122 (2017).
6. Mao, E. F., Lane, L., Lee, J. & Miller, J. H. Proliferation of mutators in a cell population. *J. Bacteriol.* **179**, 417–422 (1997).
7. Huang, M. E., Rio, A. G., Nicolas, A. & Kolodner, R. D. A genomewide screen in *Saccharomyces cerevisiae* for genes that suppress the accumulation of mutations. *Proc. Natl Acad. Sci. USA* **100**, 11529–11534 (2003).
8. Yuen, K. W. et al. Systematic genome instability screens in yeast and their potential relevance to cancer. *Proc. Natl Acad. Sci. USA* **104**, 3925–3930 (2007).
9. Stirling, P. C. et al. The complete spectrum of yeast chromosome instability genes identifies candidate CIN cancer genes and functional roles for ASTRA complex components. *PLoS Genet.* **7**, e1002057 (2011).
10. Andersen, M. P., Nelson, Z. W., Hetrick, E. D. & Gottschling, D. E. A genetic screen for increased loss of heterozygosity in *Saccharomyces cerevisiae*. *Genetics* **179**, 1179–1195 (2008).
11. Strome, E. D., Wu, X., Kimmel, M. & Plon, S. E. Heterozygous screen in *Saccharomyces cerevisiae* identifies dosage-sensitive genes that affect chromosome stability. *Genetics* **178**, 1193–1207 (2008).
12. Knudson, A. G., Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl Acad. Sci. USA* **68**, 820–823 (1971).
13. Cho, A. et al. MUFFINN: cancer gene discovery via network analysis of somatic mutation data. *Genome Biol.* **17**, 129 (2016).
14. Davoli, T. et al. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **155**, 948–962 (2013).
15. Futreal, P. A. et al. A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
16. Tamborero, D. et al. Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci. Rep.* **3**, 2650 (2013).
17. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
18. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
19. Yin, Y. & Petes, T. D. Genome-wide high-resolution mapping of UV-induced mitotic recombination events in *Saccharomyces cerevisiae*. *PLoS Genet.* **9**, e1003894 (2013).
20. Brauer, M. J., Christianson, C. M., Pai, D. A. & Dunham, M. J. Mapping novel traits by array-assisted bulk segregant analysis in *Saccharomyces cerevisiae*. *Genetics* **173**, 1813–1816 (2006).
21. Segrè, A. V., Murray, A. W. & Leu, J. Y. High-resolution mutation mapping reveals parallel experimental evolution in yeast. *PLoS Biol.* **4**, e256 (2006).
22. Choy, J. S. et al. Genome-wide haploinsufficiency screen reveals a novel role for γ-TuSC in spindle organization and genome stability. *Mol. Biol. Cell* **24**, 2753–2763 (2013).
23. Carette, J. E. et al. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* **477**, 340–343 (2011).
24. Huang, H. T., Chen, S. M., Pan, L. B., Yao, J. & Ma, H. T. Loss of function of SWI/SNF chromatin remodeling genes leads to genome instability of human lung cancer. *Oncol. Rep.* **33**, 283–291 (2015).
25. Sniegowski, P. D., Gerrish, P. J. & Lenski, R. E. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* **387**, 703–705 (1997).

Acknowledgements We thank F. Beça, D. Cabrera, B. Neugeboren and C. Pogliano for developing reagents and conducting preliminary experiments; J. Piper and the BauerCore facility at Harvard University for technical help; G. Wildenberg, J. Koschwanez, N. Wespe and M. Fumasoni for technical help and discussions; and J. Matos for the gift of the *MUS81* deletion CRISPR cell lines used in Fig. 3c and Extended Data Fig. 8. V. Denic, B. Shraiman, M. Desai, M. Fumasoni and L. Bagamery provided comments on the manuscript. M.C.C. received a long-term fellowship from Human Frontiers Science Program (HFSP, LT 000694/2014-L). R.M.P. is a recipient of the Berman/Topper Family HD Career Development Fellowship from the Huntington's Disease Society of America. The work was supported by an NIH/NIGMS award (R01/GM043987) to A.W.M.

Reviewer information *Nature* thanks J. DeGregori, S. Nijman and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions M.C.C., R.M.P. and A.W.M. designed the experiments, M.C.C. and R.M.P. performed the experiments, and M.C.C., R.M.P. and A.W.M. performed data analysis 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/s41586-019-0887-y.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-0887-y.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to M.C.C. or A.W.M.
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 and the investigators were not blinded to allocation during experiments and outcome assessment.

Strains and growth media (yeast). *S. cerevisiae* W303 was used as a base strain to build the genetic instability reporter cassette in our ancestor strains. The strains we used are corrected for the *rad5-535* mutation found in most W303 strains. Knockout and reconstruction of evolved mutations were performed on our ancestor diploid strain (yMC003) or generated by transforming the α-mating type haploid (yMC001) and crossing with the untransformed a-mating type haploid (yMC002, see Supplementary Table 4). Strains were grown in yeast extract–peptone–dextrose medium (YPD; 10 g/l yeast extract, 20 g/l peptone, and 2% (w/v) dextrose) or CSM (10 g/l YNB, 2% (w/v) dextrose)-ura liquid medium, or on CSM-5FOA, CSM-ade or CSM-his 2% agar plates.

HAP1 cell culture and CRISPR–Cas9-mediated gene inactivation (human cells). Near-haploid, HAP1 human cells were obtained directly from Horizon Discovery and cultured according to standard cell culture protocols in Iscove's modified Dulbecco's medium (IMDM), supplemented with 10% FBS and 1% PenStrep. Isogenic cell lines deficient for the human homologues of yeast instability genes (*ATG2B*, *SMARCA4*, *DYRK1B*, *GRID2IP* and *DDHD1*) were generated using CRISPR–Cas9-mediated inactivation. In summary, single-guide RNAs (sgRNAs) were designed against the target genes (Extended Data Fig. 8c), while attempting to maximize on-target activity and minimize off-target activity[26]. Individual sgRNAs were cloned into a spCas9- and GFP-expressing plasmid (pX458)[27] and transfected into wild-type HAP1 cells using Lipofectamine 3000. Individual clones were generated from single GFP-positive cells isolated by FACS (MoFlo XDP Sorter, Beckman Coulter). Sanger sequencing was used to identify isogenic clones for each target gene, carrying small frameshift insertions or deletions in early exons, which were predicted to generate no active protein products (Extended Data Fig. 8c). The Δ*MSH2* and Δ*FRYL* HAP1 lines were obtained directly from Horizon Discovery, while the Δ*MUS81* HAP1 lines[28] were a gift from the Matos laboratory (IBC-ETH, Zurich).

Construction of a genetic instability reporter. We wanted to compare genetic instability between our ancestor and evolved strains. We engineered a *URA3* reporter cassette that could report the global mutation rate and distinguish between point mutation, recombination or truncation and chromosome loss events. A copy of the *HIS3* gene is linked directly to *URA3* on the left arm of chromosome I and the only wild-type copy of the *ADE1* gene lies on the opposite side of the centromere of the chromosome. The phenotypes for growth on -5FOA, -HIS, and -ADE media distinguish the different types of mutational event. To create the reporter, we first transformed a haploid *MAT***a** strain auxotrophic for histidine (*his3-11,15*) by deleting the *URA3* coding region seamlessly *(ura3Δ0)*, and replacing *ADE1* with the Hph:MX cassette (hygromycin resistance) in strain yMC001. In a haploid *MAT*α strain auxotrophic for histidine (*his3-11,15*), we deleted *URA3* seamlessly *(ura3Δ0)*, leaving the *ADE1* locus on the right arm of chromosome I intact (yMC002). Then we inserted a cassette containing the budding yeast *URA3* and heterologous *HIS3* genes into chromosome I of yMC002 in the following arrangement, reading out from the centromere: the endogenous *ARS106* replication origin locus, *URA3* and *KlHIS3* (homologue from *Saccharomyces kluyveri*). We generated the diploid ancestor (yMC003) by crossing yMC001 with yMC002. The diploid is homozygous for all loci that contribute to growth suppression (*TRP1–TRP5*, *LYS2*, *LYS5* and *CAN1*), and heterozygous at chromosome I, where *URA3, KlHIS3* and *ADE1* are present on one copy and the other lacks all three genes (Fig. 1c). Histidine auxotrophs following *URA3* inactivation represent recombination or truncation events, while histidine and adenine double-auxotrophs identify chromosome loss events.

Perturbation experiments to test genetic instability reporter. We perturbed our ancestor with environmental stimuli that increased the frequency of point mutations, recombination and chromosome loss. Cells were treated with 1% ethyl methanesulfonate (EMS) in PBS for 30 min (to induce point mutations), neutralized with 5% sodium thiosulfate (w/v) and allowed to recover for 2 h in CSM-ura, or irradiated with 5.1 mJ/cm^2 of ultraviolet light (UV, λ = 254 nm) on a CSM-ura plate (to induce point mutations and mitotic recombination), or treated with 30 μg/ml Benomyl for 2 h (to induce chromosome loss). We also tested whether mutations known to increase genetic instability[10,29] would generate a detectable signal in our reporter. We transformed our ancestor strain to introduce a *POL3*$^{1566CTG>GAT/+}$ heterozygous mutation (Pol3-L523D, which has an increased point mutation due to inability to proof-read[29]), or deleted both copies of either *RAD27* (whose deletion increases loss of heterozygosity[10]) or *MAD2* (whose deletion increases chromosome loss[30]). Cells were grown in CSM-ura, plated in 5FOA (0.5–1 × 10^{7}cells) and surviving colonies (*URA3* mutants) were replica plated into CSM-his or CSM-ade.

Mutagenesis. To introduce heterozygous mutations and increase the genetic variability in our starting populations, we used UV and EMS mutagenesis (at levels that produced about 200 mutations per diploid ancestor clone). The diploid strain yMC003 was grown to saturation (10^8 cells per ml), plated on YPD plates (10^6 cells) exposed to 5.1 mJ/cm^2 of UV radiation (~90% killing haploids, ~60% killing diploids) on YPD plates, grown in the dark (to prevent photolyase-mediated pyrimidine dimer repair) for two days, and the surviving colonies were washed off with liquid YPD and stored in aliquots at −80 °C in YPD plus 20% glycerol. The same washing and storage procedure was performed following chemical mutagenesis produced by incubating ancestor cells in 1% EMS in 0.1M sodium phosphate buffer, pH 7.0 for 30 min. EMS was inactivated with 5% Na-thiosulfate. Aliquots of 10^7 cells from UV or EMS mutagenesis were used as starting populations to sample the genetic diversity of our evolution experiment.

Selecting for sequential inactivation of growth suppressor genes. To select for inactivation of growth suppressor genes, we used drugs that are toxic to cells expressing the active version of the gene, thus selecting for cells that have inactivated the gene. The *CAN1* gene encodes an arginine permease that allows canavanine, a toxic arginine analogue, to enter yeast cells and block their growth[31]; 5-FAA is converted into the toxic analogue 5-fluoro-tryptophan by the action of the *TRP1–TRP5* genes[32], and α-AA is lethal when used as a sole nitrogen source unless cells have inactivated *LYS2* or *LYS5*[33]. Because these mutants are recessive, diploids should inactivate both copies of a growth suppressor gene to survive. Selecting for repeated inactivation of growth suppressor genes should select for genetic instability[6]. To start our experiment with clones that have functional growth suppressor genes we first plated the mutagenized ancestor in medium lacking tryptophan, and lysine. We pooled the surviving colonies into liquid YPD. The first selection step was performed by growing cells in YPD, then plating on 5-FAA. Colonies that grew on 5-FAA were replica-plated to YPD to increase the number of cells. The second selection step was performed by replica-plating these colonies from YPD onto α-AA. Survivors on this medium were replica-plated to YPD and then these amplified colonies were replica-plated to canavanine (selecting for inactivation of *CAN1*). In the starting diploid, heterozygous deletions for the growth suppressor genes are sensitive to the growth suppressors. At the end of the evolution experiment, approximately 70% of surviving clones ($n = 105$) were resistant to canavanine and unable to grow in CSM–Trp and CSM–Lys media.

Genetic instability quantification (yeast). Clones that survived selection (evolved clones) were tested for their genetic instability. We measured the mutation frequency as the frequency of *URA3* inactivation determined by plating cells on 5FOA-containing medium. To avoid the effects of mutations that occur early in the exponential expansion of a culture (jackpots), cultures were grown in CSM-ura so that cells that inactivated *URA3* would be unable to continue proliferating. Cultures were grown until they reached exponential growth phase (1–5 × 10^7 cells/ml), sonicated, counted using a Coulter counter and plated onto 5FOA plates (5 × 10^6 to 5 × 10^7 cells). To verify the specificity of the 5FOA selection, we took 20 colonies of the ancestor (yMC003) that survived on 5FOA plates, tested their ability to grow on CSM-ura and analysed their *URA3* genes. All 20 failed to grow on CSM-ura; 19 had lost the *URA3* gene and one had a non-synonymous point mutation (475C>A) in *URA3*. To determine how cells became 5FOA-resistant, we used our genetic instability reporter (Fig. 1c) and replica-plated surviving colonies from 5FOA into CSM-his and CSM-ade, thus distinguishing point mutations (His$^+$ Ade$^+$), mitotic recombination (His$^-$ Ade$^+$), and chromosome loss (His$^-$ Ade$^-$).

Genetic instability quantification (human cells). We selected a set of six genes based on how frequently they were mutated in independently evolved yeast lineages, the magnitude of the increase in genetic instability in reconstructed yeast clones, and to sample different functional classes (*ATG2*, mutated in ten lines, fivefold elevation in overall mutation rate, metabolism; *DDL1*, mutated in two lines, threefold elevation in overall mutation rate, mitochondria; *YAK1*, mutated in six lines, twofold elevation in overall mutation rate, cell-cycle; *TAO3*, mutated in eleven lines, 2.5-fold elevation in overall mutation rate, transcription; *BNI1*, mutated in eight lines, ninefold elevation in overall mutation rate, cytoskeleton; *MUS81*, mutated in two lines, sevenfold elevation in overall mutation rate, DNA). To quantify mutation rates in human cells, we used the hypoxanthine-guanine phosphoribosyl transferase (HPRT1) loss-of-function (LOF) frequency assay[34]. In the presence of 6-TG, HPRT1-expressing cells die whereas HPRT1 LOF mutants survive. By comparing the frequency of survival between the CRISPR knockout cell lines and the wild type, we can estimate the relative increase in mutation rate. To eliminate pre-existing *HPRT1* mutations, cells were treated for 2–3 days with hypoxanthine-aminopterin-thymidine (HAT) medium. The optimal concentration of 6-TG in our assay was established using a dose-effect curve on wild-type (6-TG sensitive) and Δ*MSH2* (6-TG resistant) cell lines, with concentrations ranging from 0.1 to 100 μM of 6-TG. The effective concentration of 3.75 μM 6-TG killed all wild-type cells without inducing toxicity in the control Δ*MSH2* line. Cells were counted using a Countess II FL cell counter, and 10^7 cells were seeded in a single-well plate followed by treatment with 3.75 μM 6-TG for 14 days, with medium changed every 1–3 days. Surviving colonies were stained with Hoechst 33342 (NucBlue Live Ready Probes, ThermoFisher) and the entire surface of the

plate (95 cm^2) was imaged at 10× magnification on an inverted fluorescence microscope (Leica DMi8). The total number of colonies per CRISPR knockout cell line ($n = 3$–6 replicates) was counted using the cell counter plugin in Fiji, using a fixed binary threshold and a minimal colony area cutoff (6 cells on day 10 or 12 cells on day 14) for image segmentation.

Genomic DNA preparation. To prepare genomic DNA, cultures were pelleted and resuspended in 50 μl of 1% zymolyase in 0.1 M, pH 8.0 NaEDTA. The cells were incubated for 30 min at 37 °C to digest the cell wall, and then the cells were lysed by adding 50 μl 0.2 M NaEDTA, 0.4 M pH 8.0 Tris and 2% SDS and incubated at 65 °C for 30 min. We added 63 μl of 5 M potassium acetate and incubated the mixture for 30 min on ice. The insoluble residue was then pelleted, and 750 μl of ice-cold ethanol was added to 300 μl of the supernatant to precipitate the DNA. The DNA was pelleted, and the pellet was resuspended in 0.2 mg/ml RNAase A. After 1 h of incubation at 37 °C, 2 μl of 20 mg/ml proteinase K was added, and the solution was incubated for an additional 2 h at 37 °C. The DNA was again precipitated by adding 130 μl isopropanol. The DNA was pelleted, briefly washed with 70% ethanol, repelleted, and resuspended in 100 μl 10 mM Tris, pH 8.0.

Whole-genome sequencing. We used the standard Nextera library preparation kit (Illumina, California), optimized to generate 300–600-bp dsDNA fragments after tagmentation (adapted for small volume library preparation[35]). DNA was analysed using QuBit BR dsDNA quantification, TapeStation (Agilent HS D1000) electrophoretic analysis and qPCR (Agilent BioAnalyzer 2100). Double paired-end 125–150-bp reads were obtained from the HiSeq2000 BauerCore facility sequencer. Evolved clones were sequenced at a depth of 10–20×, while bulk segregant pools were sequenced at a depth of 30–120×.

Bulk segregant analysis of genetic instability. To sporulate the diploid strains, cultures were grown to saturation in YPD, and then diluted 1:50 into YP 2% acetate. The cells were grown in acetate for 12–24 h and then pelleted and resuspended in 2% acetate. After 2–4 days of incubation on a roller drum at 25 °C, sporulation was verified by observing the cells in a microscope. To digest ascii, 1 ml of the sporulated culture was pelleted and resuspended in 50 μl 10% zymolyase (https://www.zymoresearch.com) for 1 h at 30 °C. Water (400 μl) and 10% Triton X-100 (50 μl) were added, and the digested spores were sonicated for 5–10 s to separate the tetrads. Tetrad separation was verified by observing the cells in a microscope. The separated tetrads were then spun down slowly (6,000 r.p.m.) and resuspended in water and plated in YPD. Replica plating to CSM-ura allowed us to confirm correct segregation of the *URA3* marker (2:2). Individual URA^+ haploid segregants were tested (~400 per evolved clone) for *URA3* mutation frequency in 5FOA, and the top 10% of genetically unstable segregants (~40 per evolved clone), which should contain mutations that cause instability, were pooled for gDNA extraction.

Sequence analysis. DNA sequences were aligned to the S288C reference genome r64 (downloaded from the Saccharomyces Genome Database, https://www.yeastgenome.org/) using the Burrows–Wheeler Aligner[36] (http://bio-bwa.sourceforge.net/). The resulting SAM (sequence alignment/map) file was converted to a BAM (binary SAM) file, sorted, indexed, and made into a pileup format file using the samtools software[36] (http://samtools.sourceforge.net/). Variants including single-nucleotide polymorphisms and insertions/deletions were called from the pileup file using Varscan software[37] (http://varscan.sourceforge.net/). To perform the segregation analysis, we wrote a custom sequencing pipeline in Python (https://www.python.org/), using the Biopython (https://biopython.org/) and pysam (https://code.google.com/p/pysam/) modules, that finds sequence variants between the ancestor and clone, classifies each variant as a nonsynonymous coding region, synonymous coding region, or promoter mutation, and ranks each mutation by its segregation frequency[38].

Strain reconstruction (yeast). Alleles were replaced by transformation of the ancestor (yMC003) with a linearized DNA fragment carrying the mutated version of the allele with a selectable marker downstream of the gene and followed by intergenic DNA to direct recombination (reconstructed (mut/+); Kan:MX or Nat:MX). Transformants were selected in appropriate medium (2 × G418 or Clonat). Alleles were disrupted by replacing the full ORF with a Kan:MX or a Nat:MX cassette and selected in a similar way (Δ/+). Colonies that grew were streaked out on YPD, then transferred a second time to the selective medium. The target sequence was PCR amplified and Sanger sequenced to verify correct insertion. DNA polymerase, polynucleotide kinase, and DNA ligase were purchased from New England Biolabs (http://www.neb.com). In eight cases, the reconstructed (mut/+) strain had at least one type of mutation rate (global, PM, RT or CL) that was higher than the evolved clone that harboured the causative mutation (Fig. 2c, Extended Data Fig. 6a–c). Possible explanations include differences in cell physiology due to the accumulation of unselected mutations in the evolved clones, and selection for mutations that partially suppress genetic instability after the inactivation of the last growth suppressor gene.

Venn diagram construction for genetic instability targets comparison. We first grouped the human driver genes from six independent cancer driver studies[2,13–17]. Then we identified by reciprocal blast analysis 273 yeast genes that are orthologous or closely homologous to one of 1,310 human cancer drivers. To be included in this set the yeast and human genes had to satisfy one of three criteria: 1) they were reciprocal best blast hits; 2) the E value for the hit of the yeast gene to the cancer gene was <1E-10 and the difference between the best hit in the human genome and the hit to the cancer gene is <1E10-fold; or 3) the E value for the best hit in the human genome is 0 and the E value for the hit to the cancer gene is <1E-100. We compared three groups of yeast genes: genes whose mutations led to genetic instability in our work (92 genes), genes that were identified as conferring genetic instability in three systematic screens of the yeast gene deletion collection[8–10], and random genes in the yeast genome. To test the overlap between the latter two groups and the cancer-homologous data set, we randomly generated 6 subgroups of 92 genes for the 648 genes that confer instability when deleted, and 6 subgroups for the 5,150 annotated yeast genes (SGD). In this way each group of yeast genes had the same number of genes. The average and s.d. of the overlap between the six subgroups and the genetic instability genes and cancer homologue genes is presented in the Venn diagram in Extended Data Fig. 7b.

Gene ontology enrichment for functional classes. We curated the functions of the genetic instability genes using the SGD database, GO entries, and available literature, and divided the genes into nine major functional classes: cell cycle, cytoskeleton, DNA, metabolism, mitochondria, protein quality control, transcription, translation and transport. We performed two statistical tests to determine whether there was significant enrichment in any of the functional classes, when comparing with the functional distribution of all yeast genes: whether the probability of the number of observed hits was <0.05 using the Bonferroni correction, and whether hits passed a 5% false discovery cutoff.

Figures and data analysis. Unless otherwise noted, data analysis was performed on Excel (Microsoft), Matlab (Mathworks) or in the R programming language (https://www.r-project.org) and plots were generated using the R library ggplot2[39]. Adobe Illustrator was used to format and assemble the plotted data into figures.

Materials availability. The yeast strains (Supplementary Table 4) and HAP1 human cell lines (Extended Data Fig. 8c) created in this study are available from the corresponding authors upon reasonable request.

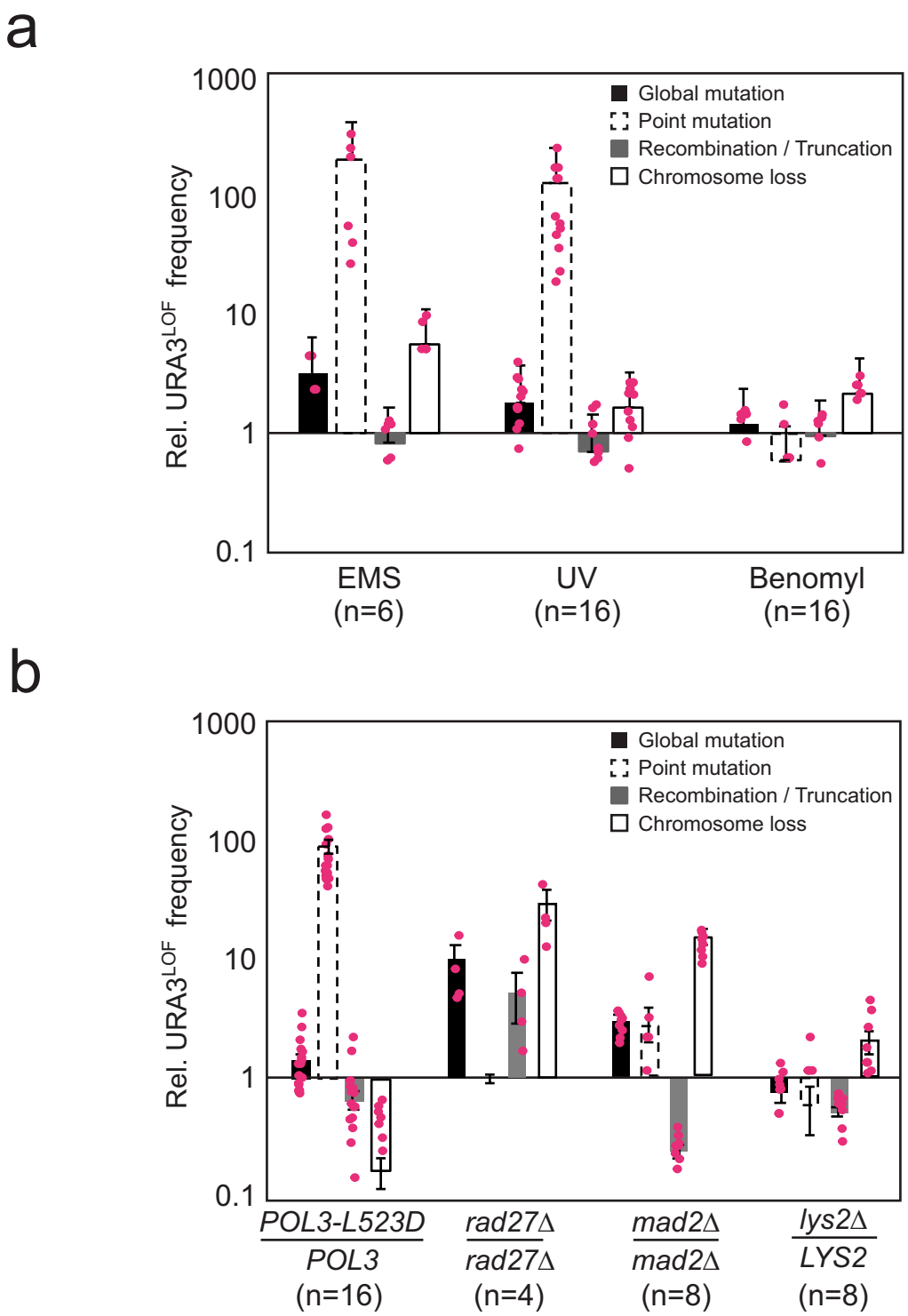
Code availability. The Python scripts used for the analysis of yeast genomic sequences and frequency of segregation are available online (https://github.com/koschwanez/mutantanalysis) or upon request.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

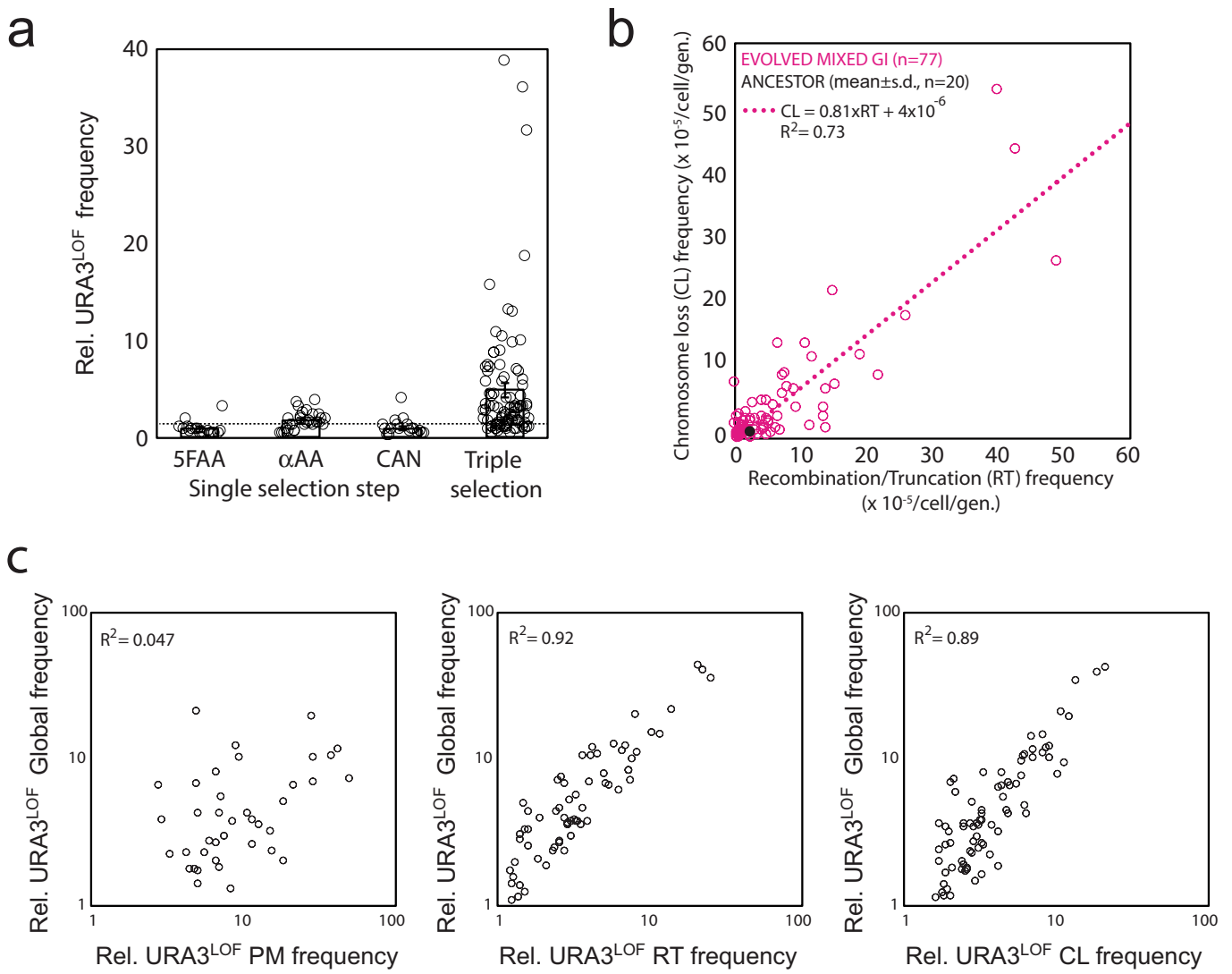
Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files. Source data for all figures are provided with the paper (online). The genome sequence data (in short reads format) have been deposited in the NCBI Bioproject database under accession number PRJNA509936.

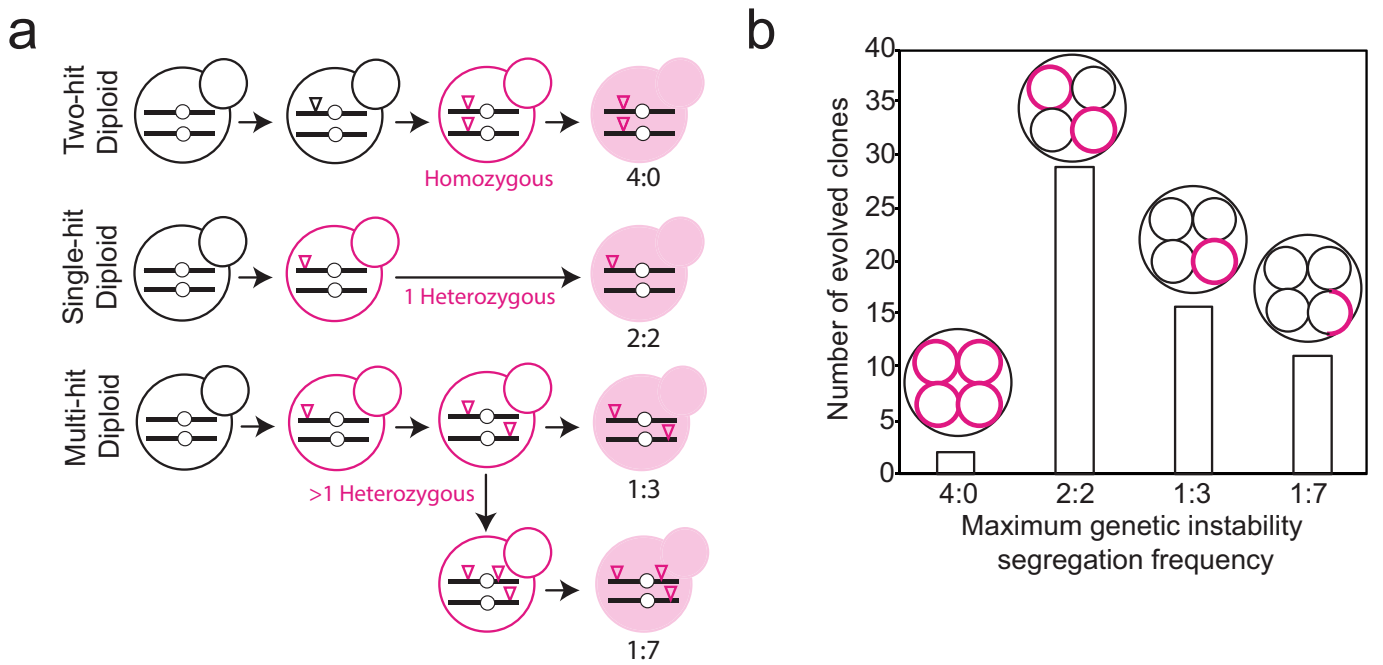
26. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
27. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols* **8**, 2281–2308 (2013).
28. Duda, H. et al. A mechanism for controlled breakage of under-replicated chromosomes during mitosis. *Dev. Cell* **39**, 740–755 (2016).
29. Jin, Y. H. et al. The multiple biological roles of the 3′→5′ exonuclease of *Saccharomyces cerevisiae* DNA polymerase delta require switching between the polymerase and exonuclease domains. *Mol. Cell. Biol.* **25**, 461–471 (2005).
30. Chen, R. H., Shevchenko, A., Mann, M. & Murray, A. W. Spindle checkpoint protein Xmad1 recruits Xmad2 to unattached kinetochores. *J. Cell Biol.* **143**, 283–295 (1998).
31. Rosenthal, G. A. The biological effects and mode of action of L-canavanine, a structural analogue of L-arginine. *Q. Rev. Biol.* **52**, 155–178 (1977).
32. Toyn, J. H., Gunyuzlu, P. L., White, W. H., Thompson, L. A. & Hollis, G. F. A counterselection for the tryptophan pathway in yeast: 5-fluoroanthranilic acid resistance. *Yeast* **16**, 553–560 (2000).
33. Chattoo, B. B. et al. Selection of LYS2 mutants of the yeast *Saccharomyces cerevisiae* by the utilization of α-aminoadipate. *Genetics* **93**, 51–65 (1979).
34. Johnson, G. E. Mammalian cell HPRT gene mutation assay: test methods. *Methods Mol. Biol.* **817**, 55–67 (2012).
35. Baym, M. et al. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One* **10**, e0128036 (2015).
36. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
37. Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
38. Koschwanez, J. H., Foster, K. R. & Murray, A. W. Improved use of a public good selects for the evolution of undifferentiated multicellularity. *eLife* **2**, e00367 (2013).
39. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York, 2009).



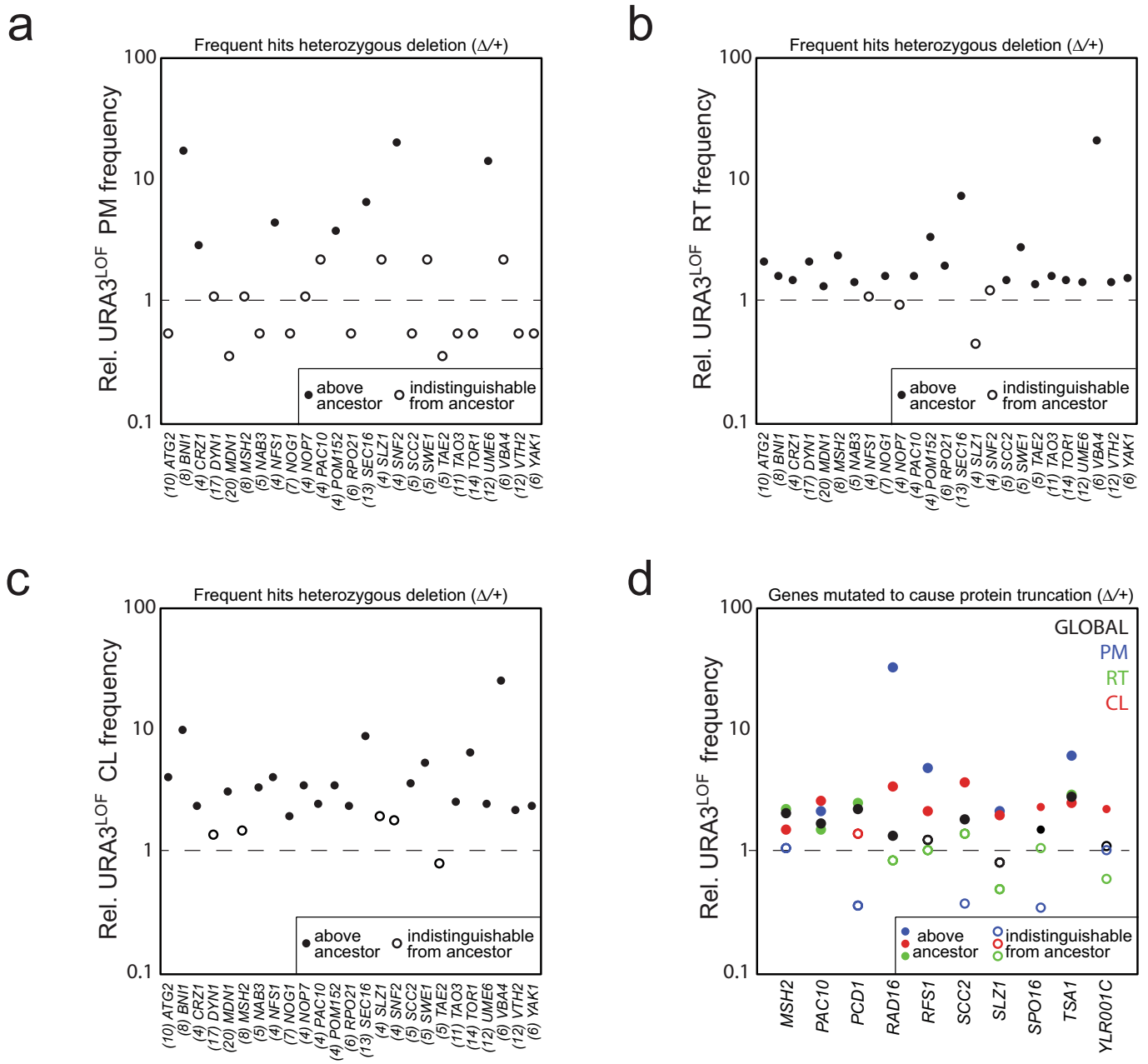
Extended Data Fig. 1 | The genetic instability reporter responds to perturbations that induce point mutations, recombination/truncation events and chromosome loss. a, Cells were treated with three agents: ethyl methanesulfonate (EMS) a chemical mutagen that primarily induces point mutations; ultraviolet radiation (UV), which induces a mixture of events, and benomyl; a microtubule depolymerizing drug, which primarily induces chromosome loss. The graph shows the frequency with which *URA3* was inactivated or lost from our ancestral diploid strain upon each of the treatments. We show the global rate of loss or inactivation of *URA3* as well as the rates for point mutation, mitotic recombination/chromosome truncation, and chromosome loss. **b**, Genetic perturbations were applied to a diploid strain carrying the reporter construct: a proof-reading mutation in polymerase delta (*POL3-L523D/POL3)* primarily induces point mutations, removing *RAD27*, which encodes a 5′ flap endonuclease, primarily elevates mitotic recombination (*rad27*Δ/*rad27*Δ); and removing *MAD2*, which encodes a component of the spindle checkpoint, induces chromosome loss (*mad2*Δ/*mad2*Δ). As a control, inactivation of a single copy of *LYS2*, using the Kanamycin resistance cassette (KanR) did not result in increased instability (*lys2*Δ/*LYS2*). The graph shows the frequency with which *URA3* was inactivated or lost from each of the genetically engineered strains. We show the global rate of loss or inactivation of *URA3* as well as the rates for point mutation, mitotic recombination/chromosome truncation, and chromosome loss (*n* represents biologically independent experiments per condition, bars represent mean and error bars represent s.e.m.). Boxplots display centre line (median), upper and lower limits (quartiles) and whiskers are 1.5× interquartile range. Datapoints = 0 were not plotted, but were used to determine mean and s.e.m.



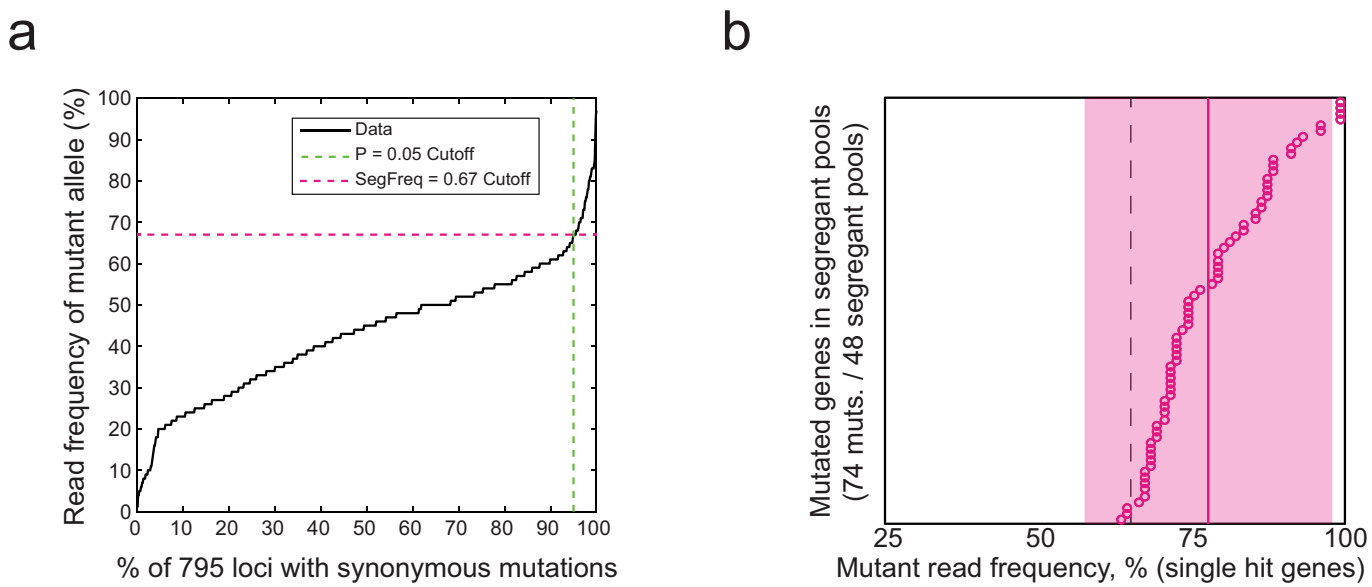
Extended Data Fig. 2 | Triple selection evolved higher levels of genetic instability than single selection and many evolved strains show a correlated increase in recombination and chromosome loss frequencies. **a**, *URA3* mutation frequency in evolved clones after a single selection step for resistance to 5-FAA ($n = 18$ biologically independent samples), α-AA ($n = 29$ biologically independent samples), or canavanine (CAN, $n = 18$ biologically independent experiments), and after triple, sequential selection ($n = 93$ biologically independent experiments) as assayed by testing clones for the frequency with which they gave rise to resistance to 5FOA normalized by the ancestor (mean ± s.e.m.). A single selection step does not promote as much instability as the triple selection (*P* values for one sided Student's *t*-test comparing each single to triple selection pair: $P_{5FAA} = 0.004$, $P_{\alpha\text{-}AA} = 0.009$, $P_{CAN} = 0.002$). **b**, Correlation between recombination/truncation (RT) and chromosome loss (CL) mutation frequency for $n = 77$ mutants that did not show a specific elevation in one class of mutational event (linear quadratic regression). **c**, Correlation between rates of the three classes of mutational events and global mutation rates ($n_{PM} = 51$, $n_{RT} = 54$, $n_{CL} = 80$ biologically independent samples, linear quadratic regression). *R* value, Pearson correlation coefficient.



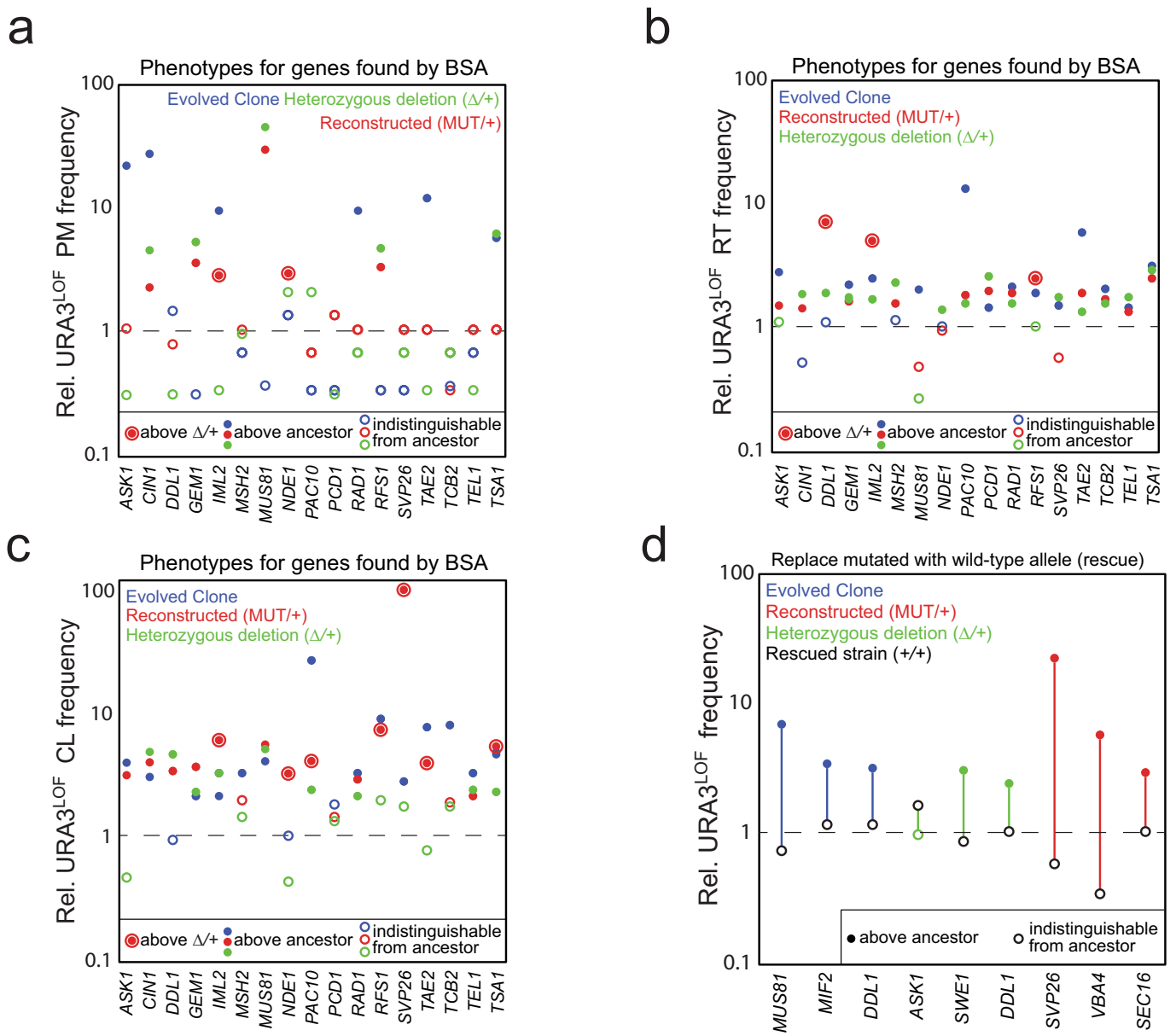
Extended Data Fig. 3 | Meiotic segregation analysis of maximum genetic instability. a, Possible scenarios for the evolution of genetic instability. In diploids, a recessive mutation can cause genetic instability only if both alleles of the gene are inactivated (two-hit diploid), whereas a dominant mutation need occur only in one copy of the gene (single-hit diploid) and some mutators require heterozygous mutations in two or more genes (multi-hit diploid). The number and nature of causative mutations will be reflected in the meiotic segregation pattern, assuming the mutations also cause instability in haploids. **b**, Meiotic segregation for the maximum level of genetic instability (measured as the mutation frequency to 5FOA resistance) in haploid spores derived from different evolved diploid clones. Genetically unstable spores are shown in magenta.



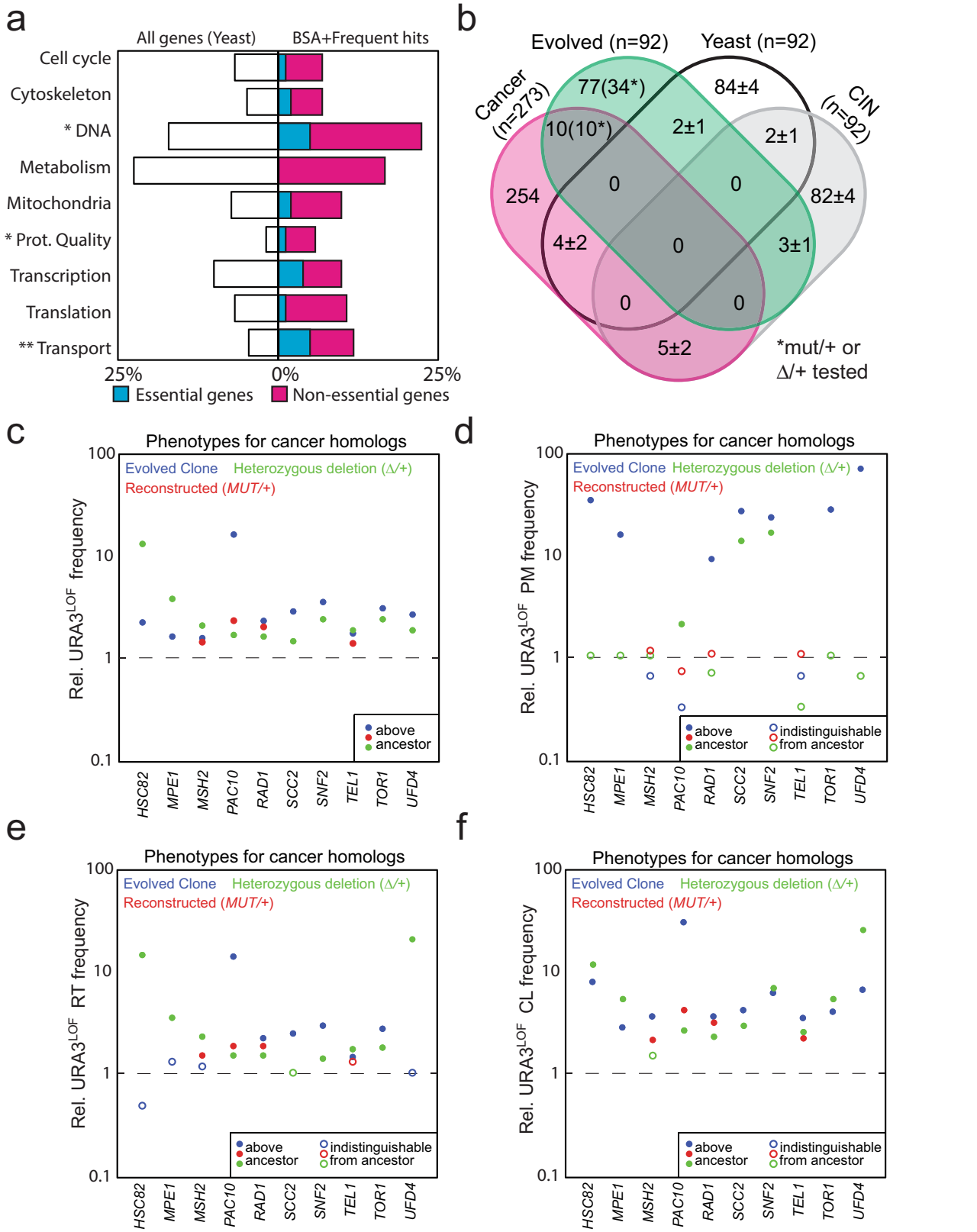
Extended Data Fig. 4 | Heterozygous deletion of frequently mutated genes affects individual genetic instability types. Heterozygous deletions ($n = 25$) were engineered in the ancestral strains (Δ/WT) carrying our reporter construct and used to measure instability. **a**–**c**, Rates of point mutation (**a**), recombination/truncation (**b**) and chromosome loss (**c**) in heterozygous deletion mutants. **d**, Mutation rates for global and individual instability types in strains carrying heterozygous deletions (Δ/+) of genes whose candidate genetic-instability-causing mutations were nonsense mutations leading to the production of truncated proteins, which are likely to be non-functional. Filled circles, strains whose instability exceeds that of the ancestor by 2 s.d. ($n = 3$ biologically independent experiments).



Extended Data Fig. 5 | Mutant allele frequencies in pools of haploid segregants. a, Frequency distribution of synonymous mutations at 795 loci, summed across multiple sequenced pools of segregants ($n = 48$ biologically independent samples). Ninety-five per cent of synonymous mutations have read frequencies of less than 0.67, allowing us to define this as the cutoff for putative causative mutations. Only those loci with 20 or more reads were included in the analysis. **b**, Mutant read frequencies for lineage-specific, putative causative mutations for genetic instability (magenta). The dotted line represents the cutoff frequency (0.67) and the magenta line represents the average frequency of the 74 mutant alleles identified (magenta box shows ± 2 s.d.). Three genes (*RAD1*, *TCB2*, and *CIN1*) that have values slightly below the cutoff were also tested, and in all three cases, reconstruction demonstrated that mutations in these genes increased genetic instability.

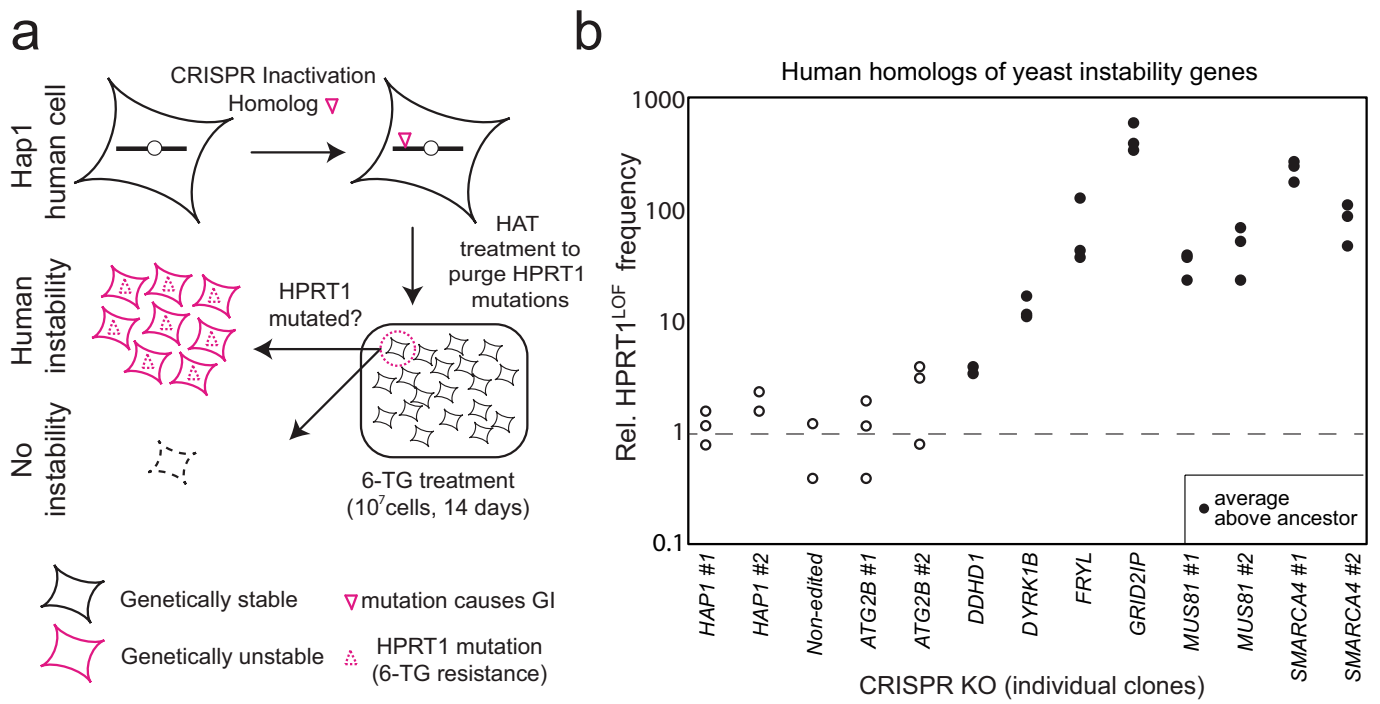


Extended Data Fig. 6 | Effect of engineering causative mutations into the ancestral diploid on point mutation, recombination/truncation and chromosome loss frequencies. Putative causative mutations from genes identified by bulk segregant analysis were engineered into the ancestral diploid strains as heterozygotes by replacing one of the wild type alleles (mut/+) and our reporter construct was used to measure instability compared to that of the evolved clone and a heterozygous deletion for the gene that harboured the putative causative mutation (Δ/+). We also tested a putative causative mutation in *PAC10*, a gene that was frequently mutated. **a**–**c**, Rates of point mutation (**a**), recombination/truncation (**b**) and chromosome loss (**c**). **d**, Global mutation rates for strains in which the causative heterozygous allele (black circle) was replaced with the wild-type allele (green circle, rescue). Filled circles, strains whose instability exceeds that of the ancestor by 2 s.d. ($n = 3$ biologically independent experiments).



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Enrichment in protein quality control and transport genes as heterozygous drivers of instability. a, Frequency of functional classes within genes where we identified mutations that caused genetic instability. Probabilities were calculated from the binomial distribution,∗∗ $P < 0.05$ with Bonferroni correction, 10 hypotheses tested, ∗significant at 2.5% false discovery rate using the Benjamini–Hochberg procedure (left, frequency in yeast genome; right, genes whose mutation we identified as leading to genetic instability, $n = 92$ independent genes). Blue represents the fraction of essential genes, magenta represents the fraction of non-essential genes. **b**, Overlap between genes in which we selected mutations that produced genetic instability (evolved, green), yeast genes homologous to human cancer genes (cancer, magenta), random subsets of yeast genes (yeast, black), and randomly selected subsets of genes previously identified in yeast screens for genetic instability (CIN, grey). **c**–**f**, Quantification of global (**c**), point mutation (**d**), recombination/truncation (**e**) and chromosome loss rates (**f**) in heterozygous point mutants or heterozygous deletion mutants for yeast genetic instability genes with cancer driver homologues. Data for the mut/+ reconstructed strain shown for four strains (*MSH2*, *PAC10*, *RAD1* and *TEL1*). The data for five of these genes (*MSH2*, *PAC10*, *SCC2*, *SNF2* and *TOR1*) are also shown in Extended Data Fig. 4. Filled circles in **c**–**f** correspond to strains whose instability exceeds that of the ancestor by 2 s.d. ($n = 3$ biologically independent experiments).



c

Gene	gRNA*	fs mutation	Sequence (Sanger)
ATG2B #1	GCTCGTACACCTTCTAGCCG	22bp del. exon 8	ACAGCTCGTACACCTTCTAGCCGTGGTAA
ATG2B #2	GCTCGTACACCTTCTAGCCG	10bp del. exon 8	ACCTTCTAGCCGTGGTAAGCAACTCAAAC
DDHD1	GCTTCTGCCAGAGTACGACG	1bp ins. exon 1	TCTGCCAGAGTACGGACGGGGCACGAGC
DYRK1B	GCTGATGAACCAGCATGACA	1bp ins. exon 5	GAACCAGCATGGACACGGAGATGAA
FRYL	GTCAAACAAGGTGCGAAGTA	1bp ins. exon 6	CTCCCTTCCTTACTTTCGCACCTTGTTTG
GRID2IP	CGGGCCTGTCTGGATCGAGT	80bp ins. exon4	CGGCCACGGGT-(76bp)-TCGCCTGTCTGG
MSH2	GCGCCGTATAGAAGTCGCCC	2bp del. exon1	TTTCGACCGGGGCGACTTCTATACGGCG
MUS81 #1,2	TACTGGCCAGCTCGGCACTC	7bp del. exon 4	TACTGGCCAGCTCGGCACTCAGGAGCC
SMARCA4 #1,2	GCAGCAGACAGACGAGTACG	1bp ins. exon10	GCAGACAGACGAGTTACGTGGCTAACCT

Extended Data Fig. 8 | Mutation rates of *HPRT1* in HAP1 human cells carrying inactivated homologues of genes mutated during yeast evolution to trigger genetic instability. a, Targets selected from yeast evolution experiments were tested in human HAP1 cells, by using CRISPR-mediated gene inactivation of the corresponding homologues. HAP1 knockout cell lines were incubated with 6-TG (after treatment with HAT medium to eliminate jackpots) and colony formation was quantified and compared to the wild type. **b**, Inactivation frequency of *HPRT1*, normalized to the wild type for different homologue deletions. HAP1 #1 and HAP1 #2 are the parental wild type; non-edited corresponds to a cell line that went through the transformation protocol for *ATG2B* editing, but did not acquire mutations in *ATG2B*; *ATG2B*#1 and *ATG2B*#2 did not show increased instability. *SMARCA4* was known to increase *HPRT1* inactivation frequency and was used as a positive control for our assay. Values are normalized by average *HPRT1* inactivation rate of wild-type HAP1 cells; filled circles correspond to strains whose instability exceeds that of the ancestor by 2 s.d. ($n = 3$ biologically independent experiments). **c**, Guide RNA target sequences and sequences of the mutated regions of genes after CRISPR-mediated gene inactivation for each of the cell lines tested (fs, frameshift; del., deletion; ins., insertion; red, deleted sequence; green, inserted sequence). Corresponding DNA sequences were cloned into p458x-GFP.

nature research

Corresponding author(s): Miguel Coelho and Andrew Murray

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).

n/a	Confirmed
☐	☒ 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 *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ 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

Data collection	Metamorph,
Data analysis	R, Excel, Matlab, Burrows-Wheeler Aligner, Varscan, Biopython, Samtools, Adobe Illustrator

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

Data availability. The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files. Source data for all figures are provided with the paper (online). Genomic data available, upon reasonable request, from the authors.

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	No statistical methods were used to pre-determine sample size. The number of samples analyzed in the evolution experiment was defined by the time and workload to analyze 100 clones initially, and the subsequent sample sizes for our analysis of instability causative mutations was maximized to analyze as many individual genes as possible.
Data exclusions	No Data was excluded from the analysis.
Replication	All atempts of replication were successfull.
Randomization	No subgroups were used, all samples were treated equally.
Blinding	In this study, blinding was not relevant, since we compared directly the effect of mutants with the wild-type ancestor in both evolved strains or reconstructed strains.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	All unique biological materials are readily available from the authors or standard commercial sources (Horizon Discovery).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HAP1 (Horizon Discovery)
Authentication	Cells were authenticated using Sanger sequencing.
Mycoplasma contamination	Cells were tested negatively for Mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Study did not involve laboratory animals
Wild animals	Study did not involve wild animals
Field-collected samples	Study did not involve field-collected samples