

Strategic targeting of Cas9 nickase expands tandem gene arrays

Takesue, Hiroaki

Department of Biochemistry, Kyushu University Graduate School of Medical Sciences

Okada, Satoshi

Department of Biochemistry, Kyushu University Graduate School of Medical Sciences

Doi, Goro

Department of Biochemistry, Kyushu University Graduate School of Medical Sciences

Sugiyama, Yuki

Department of Biochemistry, Kyushu University Graduate School of Medical Sciences

他

<https://hdl.handle.net/2324/7347460>

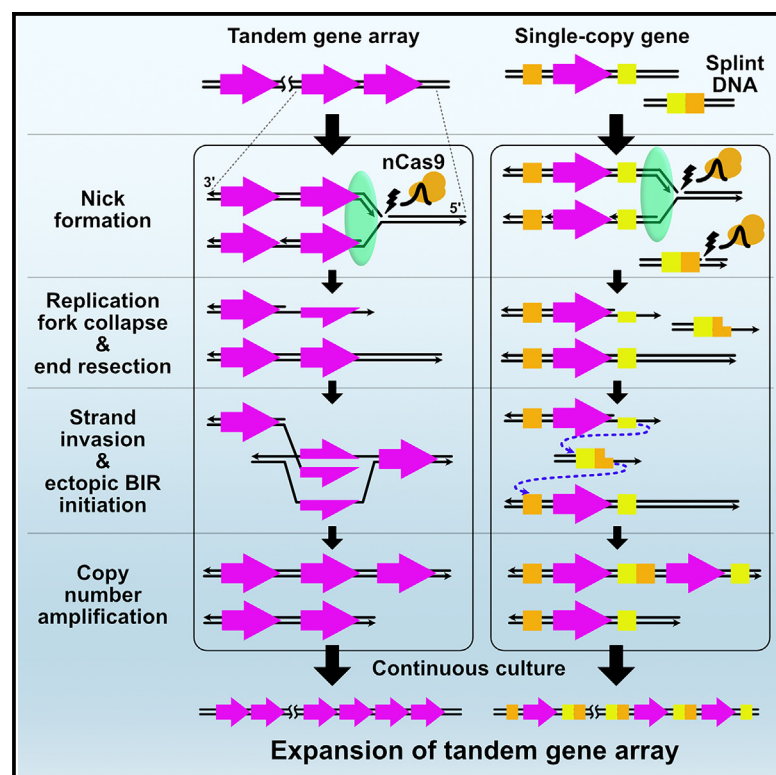
出版情報 : Cell Genomics. 5 (4), pp.100811-, 2025-04-09. Elsevier

バージョン :

権利関係 : © 2025 The Author(s).

Strategic targeting of Cas9 nickase expands tandem gene arrays

Graphical abstract



Authors

Hiroaki Takesue, Satoshi Okada,
Goro Doi, Yuki Sugiyama,
Emiko Kusumoto, Takashi Ito

Correspondence

ito.takashi.352@m.kyushu-u.ac.jp

In brief

Takesue et al. introduce BITREx, a novel genome editing method that strategically directs a single Cas9 nickase adjacent to a target tandem gene array, enabling its efficient expansion. By leveraging replication fork breakage, BITREx drives copy-number amplification, offering a unique approach to genome manipulation.

Highlights

- BITREx uses Cas9 nickase to expand tandem gene arrays by break-induced replication
- BITREx can generate megabase-sized gene arrays in budding yeast
- Splint DNA allows BITREx to form a tandem array *de novo* from a single-copy gene
- BITREx is applicable to mammalian cells



Technology

Strategic targeting of Cas9 nickase expands tandem gene arrays

Hiroaki Takesue,¹ Satoshi Okada,¹ Goro Doi,¹ Yuki Sugiyama,¹ Emiko Kusumoto,¹ and Takashi Ito^{1,2,*}¹Department of Biochemistry, Kyushu University Graduate School of Medical Sciences, Fukuoka 812-8582, Japan²Lead contact*Correspondence: ito.takashi.352@m.kyushu-u.ac.jp<https://doi.org/10.1016/j.xgen.2025.100811>

SUMMARY

Expanding tandem gene arrays facilitates adaptation through dosage effects and gene family formation via sequence diversification. However, experimental induction of such expansions remains challenging. Here, we introduce a method termed break-induced replication (BIR)-mediated tandem repeat expansion (BITREx) to address this challenge. BITREx places Cas9 nickase adjacent to a tandem gene array to break the replication fork that has just replicated the array, forming a single-ended double-strand break. This break is subsequently end-resected to become single stranded. Since there is no repeat unit downstream of the break, the single-stranded DNA often invades an upstream unit to initiate ectopic BIR, resulting in array expansion. BITREx has successfully expanded gene arrays in budding yeast, with the *CUP1* array reaching ~1 Mb. Furthermore, appropriate splint DNAs allow BITREx to generate tandem arrays *de novo* from single-copy genes. We have also demonstrated BITREx in mammalian cells. Therefore, BITREx will find various unique applications in genome engineering.

INTRODUCTION

Gene duplication plays a critical role in adaptation and evolution, as has become increasingly apparent with the advent of comparative and personal genomics. Two major mechanisms of gene duplication are retrotransposition and non-allelic homologous recombination.¹ The latter mechanism initially generates a tandemly duplicated pair of the gene, and this configuration often invites additional recombination events to form a tandem gene array. Expansion and contraction of such arrays result in copy-number variation (CNV) of the gene.

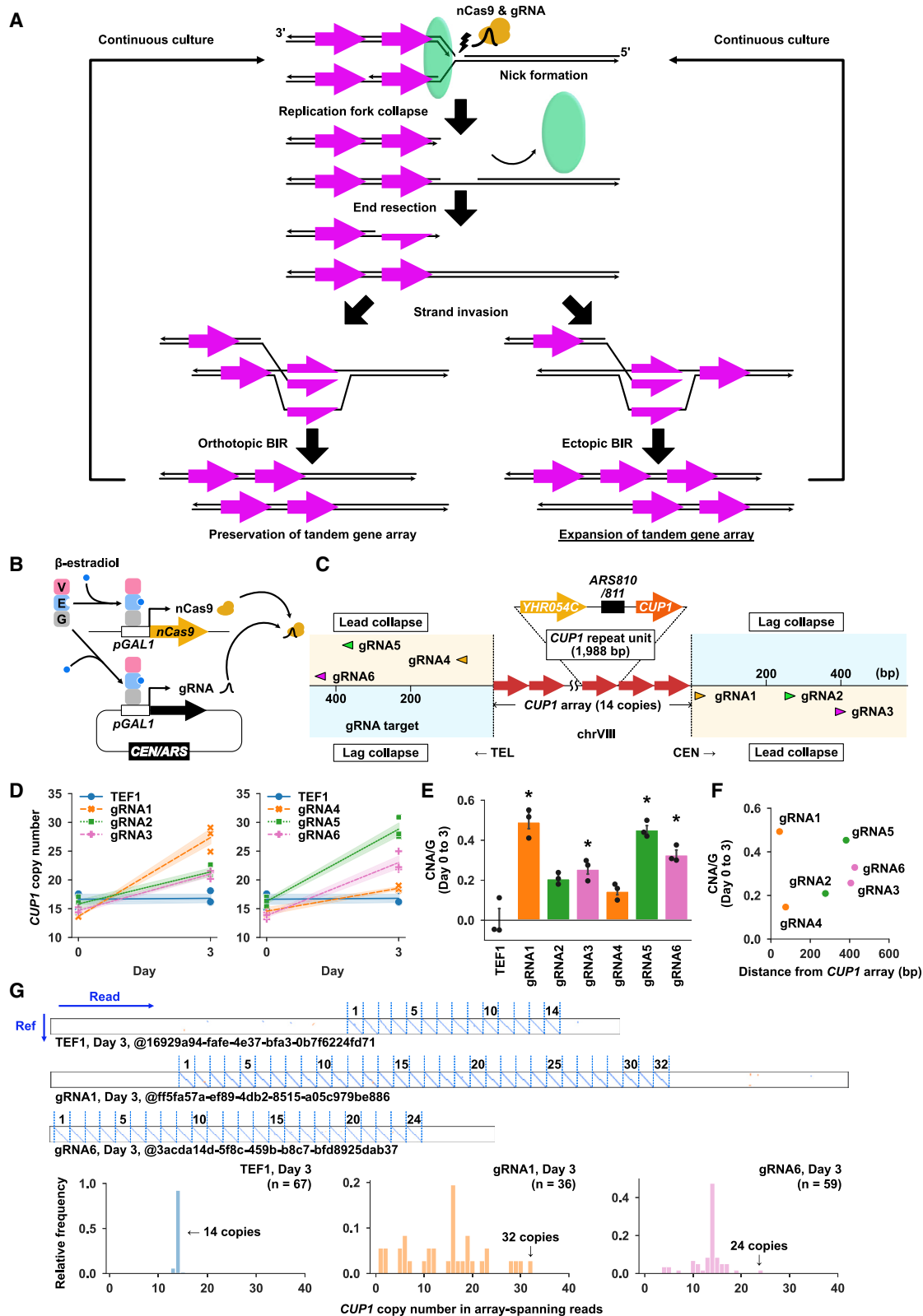
An immediate effect of expanding tandem gene arrays would be dosage effects (Figure S1A). Increased gene dosage often contributes to adaptation to environmental changes. For example, the copy number of an amylase gene, *AMY2B*, shows a notable difference between domestic dogs and wolves, both of which belong to the same species, *Canis lupus*: domestic dogs have more copies than wolves.² This is probably because domestic dogs have adapted to the carbohydrate-rich diet provided by humans, whereas wolves have maintained the ancestral copy number of *AMY2B* without finding any benefit from its high dosage in the wild.

Similarly, the budding yeast *Saccharomyces cerevisiae* shows CNV of the *CUP1* gene, which encodes a copper metallothionein. Different strains have *CUP1* arrays with varying numbers of repeat units, resulting in different levels of copper resistance. Interestingly, the boundaries of the repeat units composing the *CUP1* arrays vary between strains, suggesting that the initial duplication events from a single-copy *CUP1* gene to a two-unit *CUP1* array occurred independently in their ancestors.³ Thus,

the formation of the *CUP1* array is an example of convergent evolution. It is considered to have occurred during domestication because *CUP1* is a single-copy gene in many wild isolates of *S. cerevisiae* and in *S. paradoxus*, a non-domesticated cousin of baker's yeast.⁴ These yeasts are unlikely to have been exposed to high concentrations of copper in the wild, and thus would not have seen any fitness benefit from increased *CUP1* dosage. Expanding tandem gene arrays is a powerful strategy for rapid adaptation to environmental change. Indeed, clones with expanded *CUP1* arrays can be easily obtained in the laboratory by growing yeast cells in the presence of high copper concentrations.⁵

The expansion of tandem gene arrays can also have long-term effects, including sequence diversification that generates paralogs (Figure S1A). This process allows the original tandem gene array to evolve into a genomic locus encoding members of a multigene family. A paradigmatic example of such a locus is the human β -globin locus, which is composed of five genes and one pseudogene. This arrangement allows for the developmental stage-dependent production of three different isoforms—embryonic, fetal, and adult hemoglobins—enabling adaptation to change in oxygen concentration. Extreme examples of gene families with tandemly arrayed paralogs include those for the olfactory receptor (OR), immunoglobulin, and cytochrome P450. For instance, the human and elephant genomes contain ~400 and ~2,000 OR genes, respectively, with nearly as many pseudogenes.⁶ While pseudogenization appears to be an inevitable consequence of the functional differentiation of duplicated/multiplied genes, certain pseudogenes participate in gene conversion, contributing to the immune system, and





(legend on next page)

others play biological roles by generating functional non-coding RNAs.⁷ Tandem gene amplification followed by sequence diversification is therefore a fundamental strategy in evolution by gene duplication.

The recent advent of genome editing has made it possible to manipulate tandem gene arrays. Cleaving the repeat units of a tandem gene array easily induces its contraction through single-strand annealing (SSA) or non-homologous end joining (NHEJ). In contrast, there is currently no method to expand a tandem gene array. Developing such a method could induce dosage effects and serve as a potential strategy for increasing the yield of useful gene products. It could also provide a basis for generating multigene families, serving as a unique tool for experimental evolution to enhance the potential of cells. But how can we expand a tandem gene array?

In this context, it is interesting to note our previous finding that targeting the catalytically inactive variant of Cas9 (dCas9) to the *CUP1* array induces its contraction in the majority of cells and expansion in a minority of cells.⁸ Mechanistically, dCas9 interferes with replication fork progression, and some of the stalled forks likely break, leading to recombinational repair events that inevitably include non-allelic recombination, resulting in *CUP1* CNV. On the other hand, a single-molecule observation study revealed that replisome disassembles upon collision with Cas9 nickase (nCas9).⁹ Based on our considerations of the potential mechanism for dCas9-induced *CUP1* array expansion and the nCas9-induced replisome disassembly observed by others, we conceived the idea of repurposing break-induced replication (BIR)^{10,11}—a mechanism to repair single-ended double-strand breaks (seDSBs) generated upon replication fork breakage—to expand tandem gene arrays by strategically targeting nCas9.

DESIGN

Breakage of a replication fork results in the formation of a seDSB. Subsequent end-resection of the seDSB converts it into a 3'-protruding single-stranded DNA (ssDNA). The ssDNA typically invades the sister chromatid at its allelic position and initiates displacement synthesis, known as BIR (Figure S1B), which continues until it encounters a converging replication fork.^{10–12} When a replication fork collapses within a tandem gene array, the cell can initiate BIR either orthotopically (i.e., at the allelic position) or ectopically (i.e., at non-allelic positions) (Figure S1C). While orthotopic BIR preserves the tandem gene array, ectopic

or out-of-register BIR induces copy-number alterations (CNA) of repeat units. If the ssDNA enters a repeat unit downstream or upstream of the seDSB with respect to the direction of replication fork progression, subsequent BIR will decrease or increase the copy number of the repeat units, resulting in contraction or expansion of the array, respectively.

We hypothesized that, if a replication fork collapses just before completing the replication of a tandem gene array to generate a seDSB within the terminal repeat unit, then the ssDNA derived from the seDSB must invade either the allelic repeat unit or an upstream repeat unit, as there is no downstream repeat unit (Figure 1A). In other words, the array loses the opportunity to contract and thus either remains unchanged or expands. Although we can use nCas9 to induce a seDSB in a replication-dependent manner, we cannot target it selectively to the terminal unit because all repeat units share an identical sequence. However, BIR can also occur, albeit with reduced efficiency, when the invading ssDNA has a 3'-tail sequence that is not homologous to the donor sequence.¹³ Therefore, we hypothesized that targeting nCas9 to the flanking site of a tandem gene array could induce its expansion. We termed this strategy BIR-mediated tandem repeat expansion (BITREx). We first investigated the feasibility of BITREx using the budding yeast *S. cerevisiae* as a model system.

RESULTS

Proof of concept for BITREx by nCas9-induced *CUP1* array expansion

We constructed a strain in which β -estradiol induces the expression of Cas9^{D10A}, the nCas9 that selectively cleaves the target strand hybridized with guide RNA (gRNA). The induction is mediated by the artificial transcription factor GEV, which consists of the Gal4 DNA-binding domain, estrogen receptor, and VP16 transcriptional activation domain.¹⁴ Upon binding to β -estradiol, the GEV migrates to the nucleus and activates *GAL1* promoters to induce the expression of Cas9^{D10A} from the genome and its gRNA from a plasmid (Figure 1B). The *CUP1* array is composed of tandem iteration of a ~2.0-kb DNA segment that contains the *CUP1* gene and the *ARS810/811* origin of replication (referred to as the *CUP1* repeat unit or *CUP1RU*) on chromosome VIII.³ It consists mainly of 14 repeat units in the parental strain used in this study (14×*CUP1RU*, Figures S2A and S2B). We constructed a series of strains that expresses Cas9^{D10A} targeted to the

Figure 1. Proof of concept for BITREx by nCas9-induced *CUP1* array expansion

- (A) Principle of BITREx.
(B) GEV-based system for co-induction of nCas9 and gRNA.
(C) *CUP1* array and target sites of effective gRNAs (gRNA1–gRNA6). A rightward or leftward arrowhead indicates that the gRNA sequence is designed for the top or bottom strand, respectively, with its protospacer adjacent motif (PAM). Consequently, the nCas9–gRNA complex nicks the bottom or top strand, which serves as a template for leading strand synthesis initiated from the *ARS810/811* in the *CUP1* repeat unit (lead collapse).
(D) nCas9-induced CNA of *CUP1*. Expression of nCas9 and gRNAs were induced by the addition of 10 nM β -estradiol on day 0. Population average *CUP1* copy number was determined by qPCR on days 0 and 3. Each point in the line plot indicates the average *CUP1* copy number ($n = 3$ biological replicates), with shading around each line indicating the standard deviation (SD).
(E) nCas9-induced CNA/G of *CUP1*. CNA was divided by the number of cell divisions estimated from the increase in the optical density of the culture. Error bar represents SD ($n = 3$ biological replicates). * $p < 0.05$ (Student's *t* test).
(F) Plot of the CNA/G value versus the distance from the *CUP1* array to each gRNA target sequence.
(G) Nanopore sequencing of *CUP1* arrays. Upper panel, representative dot plots comparing nanopore reads to the *CUP1RU* reference sequence. Lower panel, distribution of *CUP1RU* copy numbers in the specified number of nanopore reads spanning the entire array.

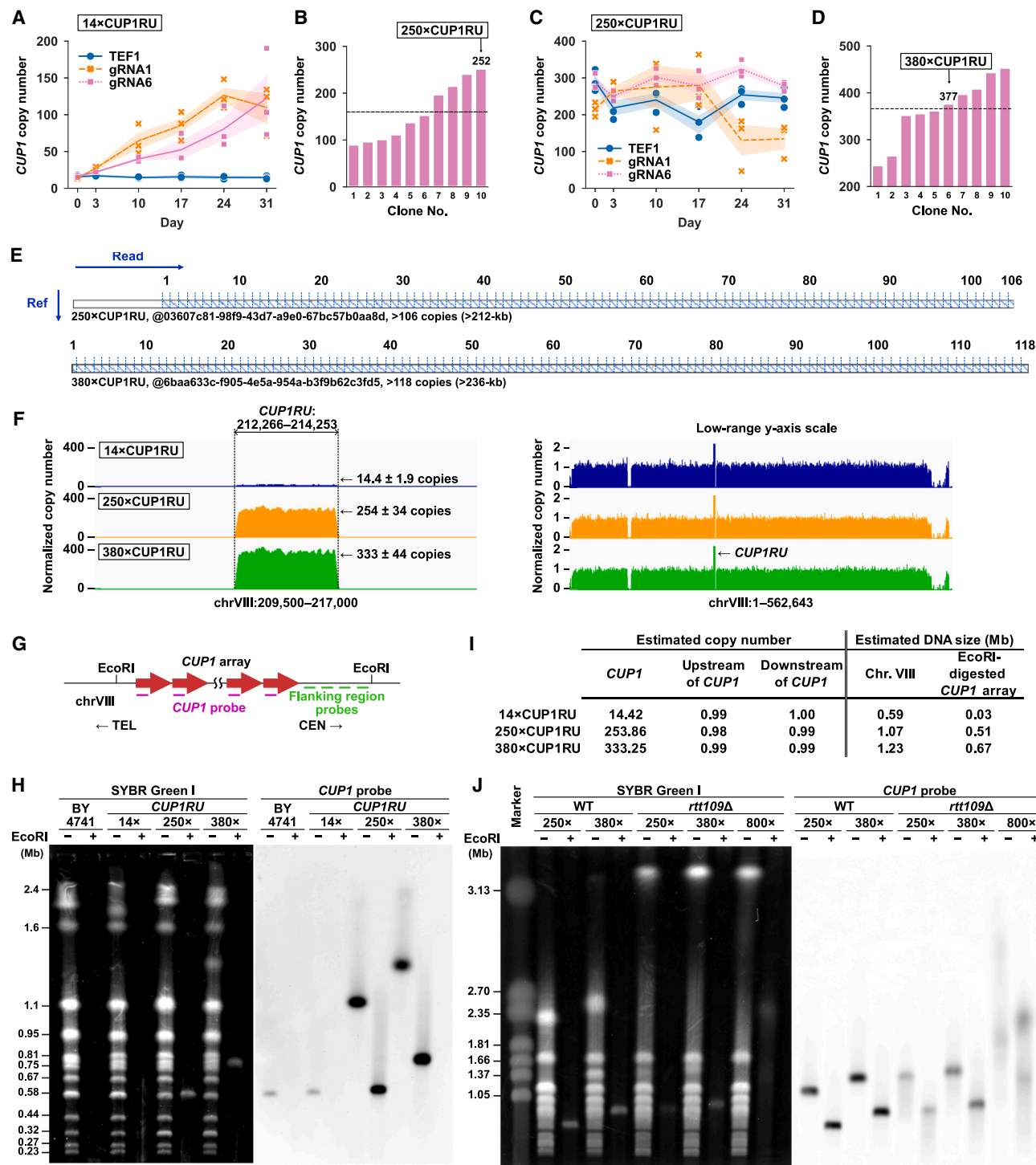


Figure 2. Generation of Mb-sized *CUP1* arrays by long-term BITREx

(A) CNA of *CUP1* over the 31-day BITREx period. Shading, SD ($n = 3$ biological replicates).

(B) *CUP1* copy number of 10 randomly picked clones on day 31 of BITREx using gRNA6. The dashed line indicates the average copy number, which is 160.

(C) CNA of *CUP1* over the second 31-day BITREx period. Shading, SD ($n = 3$ biological replicates).

(D) *CUP1* copy number of 10 randomly picked clones on day 31 of the second cycle of BITREx using gRNA6. The dashed line indicates the average copy number, which is 366.

(E) Representative dot plots comparing nanopore reads from the 250×CUP1RU and 380×CUP1RU strains to the *CUP1RU* reference sequence.

(legend continued on next page)

upstream- and downstream-flanking sites of *CUP1* array (Figure 1C) and a control site in the *TEF1* locus on chromosome XVI.

We grew these strains by daily dilution of the culture with fresh medium, extracted genomic DNA before and 3 days after β -estradiol addition, and measured the population average *CUP1* copy number using qPCR (Figure 1D). There was no change in the *CUP1* copy number in the control strain (Figure 1D). In contrast, an increase in the *CUP1* copy number was evident in 6 out of the 24 strains with Cas9^{D10A} targeted to the flanking regions of the *CUP1* array (Figures 1D, S2C, and S2D). To evaluate the performance of each gRNA, we calculated CNA per generation (CNA/G) (Figure 1E). The 6 effective gRNAs (gRNA1–gRNA6) showed significantly higher CNA/G values than the other 18 gRNAs (gRNAs1–gRNAs18) and the control *TEF1* gRNA (Figures 1E and S2D). Although the CNA/G values varied among the 6 gRNAs, they did not correlate with the distance from the *CUP1* array to the gRNA target sites (Figure 1F). Notably, all 6 gRNAs direct Cas9^{D10A} to nick the template DNA strands for leading strand synthesis initiated by the replication origin *ARS810/811* in the *CUP1RU* (lead collapse). A series of experiments using a strain with an inverted gRNA1 target site (Figures S2E–S2G) and an improved Cas9^{H840A} variant¹⁵ to nick the strand opposite to Cas9^{D10A} (Figures S3A–S3C) confirmed the hypothesized requirement for lead collapse in BITREx and aligned with previous reports demonstrating the superiority of Cas9^{D10A} over Cas9^{H840A}.^{16,17} Accordingly, we used Cas9^{D10A} throughout this study, referring to it hereafter as nCas9 for brevity.

Next, we performed the ligation-based nanopore whole-genome sequencing of DNA extracted from the strains after 3 days of BITREx induction. We selected reads containing both 5'- and 3'-flanking regions of the *CUP1* array and generated dot plots between these reads and the *CUP1RU* reference sequence. The number of diagonal lines in the dot plot indicates the number of repeat units comprising the array (Figure 1G). These analyses identified arrays consisting of ~14, up to 32, and up to 24 copies of *CUP1RU* from strains with *TEF1* gRNA, gRNA1, and gRNA6, respectively (Figure 1G). Note that the apparent discrepancy between the nanopore-based and qPCR-based estimates of *CUP1* copy number arises because the former used only reads that span the entire array, resulting in longer arrays being included less frequently, whereas the latter used all arrays evenly, regardless of their lengths. Intriguingly, nanopore sequencing also revealed the presence of contracted arrays, even though the population average *CUP1* copy number increased (Figures 1G, S4A, and S4B). Additional nanopore read analyses revealed no detectable levels of translocation involving

the *CUP1* locus, ectopic integration of circular DNAs excised from the *CUP1* array, or aneuploidy affecting chromosome VIII (Figures S4C and S4D).

We also examined the effects of mutating genes for essential BIR factors (Rad51, Pol32, and Pif1¹⁸) and a BIR inhibitor (Rtt109¹⁹) on the nCas9-induced *CUP1* array expansion. The expansion was inhibited in mutants of the essential factors and accelerated in the mutant of the inhibitor (Figures S5A–S5D), confirming that BIR mediates BITREx, as originally designed.

Taken together, targeting nCas9 adjacent to the *CUP1* array induced its expansion *in situ*, proving the principle of BITREx. Notably, in contrast to copper-induced expansion,⁵ BITREx expands the *CUP1* array without any selection pressure; it is so efficient that the average copy number of *CUP1RU* increases even in the unselected population.

Long-term BITREx to generate megabase-sized *CUP1* arrays

Theoretically, BITREx occurs every cell cycle to continuously extend the target gene array. To test this possibility, we performed a 31-day continuous BITREx experiment by diluting the yeast cell culture every 3 days for inoculation into fresh medium. Starting from the wild-type strain with 14 copies of *CUP1RU* (14×*CUP1RU*), long-term BITREx using gRNA1 or gRNA6, but not *TEF1* gRNA, increased the *CUP1* copy number (Figure 2A). From the 10 clones randomly isolated from the gRNA6-expressing cell population with an estimated copy number of 160 (Figure 2B), we selected one clone estimated to carry 252 units (250×*CUP1RU*) for further experiments. Following the curing of the gRNA6-expressing plasmid from this clone, we transformed the cells with the *TEF1* gRNA-, gRNA1-, or gRNA6-expressing plasmid and subjected the obtained transformants for another cycle of 31-day BITREx. Intriguingly, the copy number appeared to reach a plateau (~300) during the second cycle with gRNA6 (Figure 2C). We selected a strain estimated to carry 377 units (380×*CUP1RU*) from the resulting cell population for subsequent experiments (Figure 2D). In the gRNA1-expressing strain, the copy number initially appeared to plateau but subsequently declined (Figure 2C). Prolonged nicking over time can sometimes induce mutations at the target sites (e.g., Figure S6A), potentially enabling clones with defective nicking to dominate the population. However, no mutations were detected at the gRNA1 target site in these cell populations (Figure 2C), leaving the cause of the decline unexplained. In the *TEF1* gRNA-expressing control strain, the copy number fluctuated substantially but was largely maintained without showing a consistent decline (Figure 2C).

(F) Normalized read counts in Illumina sequencing. The left panel focuses on the region around *CUP1RU*, while the right panel shows the entire chromosome with a low-range y axis scale. Read counts were normalized to the average counts of genomic regions excluding rRNA, *CUP1RU*, Ty elements, and mitochondrial DNA. Note that, while the *sacCer3* reference genome sequence contains two copies of *CUP1RU*, the second copy is masked with “N” prior to mapping. Consequently, the normalized read count directly reflects the *CUP1RU* copy number. The gaps in read counts on the left arm and adjacent to *CUP1RU* are due to the presence of a Ty4 element and the aforementioned masking, respectively. The dip on the right arm results from the presence of a Ty1 element and the segmental duplication between chromosomes VIII and I.

(G) Strategy of Southern blot hybridization.

(H) PFGE analysis of *CUP1* arrays expanded by long-term BITREx. Left, SYBR Green I stain; right, blot hybridized with the *CUP1* probe.

(I) Deep sequencing-based estimates of the sizes of chromosome VIII and the EcoRI restriction fragment containing the *CUP1* array.

(J) PFGE analysis of *CUP1* arrays expanded in wild-type and *rtt109Δ* cells by long-term BITREx. Left, SYBR Green I stain; right, blot hybridized with the *CUP1* probe. The most slowly migrating bands are chromosome XII, which harbors the rDNA array known to expand remarkably in *rtt109Δ* cells.²⁰

We performed nanopore sequencing to detect the expanded *CUP1* array in the 250×*CUP1RU* and 380×*CUP1RU* strains. Among the reads obtained are those containing more than 106 (>212 kb) and 118 (>236 kb) copies of *CUP1RU* (Figure 2E). However, we failed to obtain such reads that span the full length of the extended array, which must be longer than ~500 and ~760 kb for 250×*CUP1RU* and 380×*CUP1RU*, respectively. Nevertheless, the normalized read counts of *CUP1RU* in Illumina sequencing were 254 ± 34 and 333 ± 44 in the 250×*CUP1RU* and 380×*CUP1RU* strains, respectively, while those of non-*CUP1RU* regions on chromosome VIII were equal to one (i.e., genome-wide average) (Figure 2F). We also confirmed that non-*CUP1RU* portions of *CUP1RU*-containing nanopore reads were derived exclusively from the *CUP1* locus on chromosome VIII (Figure S6B). These results suggested that BITREx can expand the *CUP1* array *in situ* to a sub-Mb size with no gross rearrangement.

To visualize the entire chromosome VIII and the *CUP1* array in these strains, we performed pulsed-field gel electrophoresis (PFGE) followed by Southern blot hybridization using probes derived from either *CUP1RU* or its flanking region (Figure 2G). Both probes hybridized to 0.6-, 1.1-, and 1.4-Mb bands in the intact genomic DNAs derived from the 14×*CUP1RU*, 250×*CUP1RU*, and 380×*CUP1RU* strains, respectively (Figures 2H and S6C). We also examined EcoRI-digested genomic DNA. Since *CUP1RU* has no EcoRI site, the extended arrays appeared as 0.5- and 0.8-Mb bands in the 250×*CUP1RU* and 380×*CUP1RU* strains, respectively (Figures 2H and S6C). These sizes were consistent with those estimated from the Illumina deep sequencing (Figure 2I). We did not detect any aberrant bands indicating translocations or other chromosomal alterations.

To further explore the potential of BITREx in *CUP1* array expansion, we performed a long-term culture experiment using the *rtt109Δ* strain, as it remarkably accelerated the array expansion (Figure S5C). In the absence of Rtt109, BITREx for 31 days increased the average *CUP1* copy number to ~500, resulting in generation of arrays exceeding 1 Mb in length (Figure S6D). We examined individual colonies by qPCR and obtained strains with estimated copy numbers of 274 (*rtt109Δ* 250×*CUP1RU*), 387 (*rtt109Δ* 380×*CUP1RU*), and 819 (*rtt109Δ* 800×*CUP1RU*) for further experiments (Figure S6D). After 3 days of culture without estradiol, the first two strains exhibited only a modest decrease in the *CUP1* copy number as did the wild-type strains with similar copy numbers, while the third strain showed a significant decrease from ~700 to ~350 copies, suggesting an intrinsic instability of extremely long *CUP1* array (Figure S6E). The normalized Illumina read counts in the *CUP1RU* increased significantly, whereas those in the non-*CUP1RU* regions on chromosome VIII did not show any detectable CNAs, ruling out the possibility of aneuploidy (Figure S6F). Non-*CUP1RU* portions of *CUP1RU*-containing nanopore reads showed no evidence of gross rearrangements (Figure S6G). Southern blot hybridization confirmed the elongation of the *CUP1* array to over 1 Mb in the *rtt109Δ* 800×*CUP1RU* strain despite its slow growth (Figures 2J and S6H). However, the hybridization bands in *rtt109Δ* strains appeared fuzzier than those in the wild-type strain, suggesting that *rtt109Δ* cells likely harbor more heterogeneous *CUP1* arrays than their wild-type counterparts. In addition,

hyper-amplification of the rDNA array in *rtt109Δ* cells²⁰ resulted in a marked delay in the migration of chromosome XII (Figures 2J and S6H).

Taken together, BITREx can expand the *CUP1* array to Mb size, especially in the absence of Rtt109. To maintain the extremely long arrays, BITREx likely needs to be continuously induced.

Epigenetic modulation of BITREx

Rtt109 is the sole enzyme responsible for acetylation at Lys-56 of histone H3.²¹ Accordingly, the absence of H3K56ac in *rtt109Δ* strains likely accelerates BITREx while inducing marked heterogeneity, indicative of potential instability (Figures 2J and S6H). Considering that the stability of expanded arrays is crucial for practical applications, we decided to further investigate the role of H3K56ac. For this purpose, we examined the effect of nicotinamide (NAM), which inhibits the NAD⁺-dependent histone deacetylase family consisting of Sir2, Hst1, Hst2, Hst3, and Hst4 in the budding yeast.²² Previous studies have reported that NAM induces *CUP1* copy-number reduction in an Rtt109/H3K56ac-dependent manner.^{8,23} We first confirmed that H3K56ac accumulates in the 14×*CUP1RU* strain after 24 h of NAM exposure (Figure 3A). Notably, the presence of NAM not only suppressed BITREx of the *CUP1* array (Figure 3B), but also led to its gradual contraction, regardless of whether nCas9 was targeted to the *CUP1*-flanking site or the control site (Figure 3C). The effect of NAM on highly extended *CUP1* arrays was remarkable: the 250×*CUP1RU* and 380×*CUP1RU* strains showed a drastic decrease in the copy number during 3 days of NAM exposure (Figure 3D).

These results suggested that the effect of NAM on CNA depends on the initial length of the *CUP1* array. To test this hypothesis, we prepared a series of cell populations with different average *CUP1* copy numbers by temporal sampling from a long-term BITREx culture. Each sample was divided into two subpopulations, which were then cultured in parallel in the presence and absence of NAM for 3 or 4 days (Figure 3E). In all samples, NAM exposure induced a steep decrease in *CUP1* copy number, converging to <30 copies (Figure 3F). There was no difference in CNA/G levels with or without nCas9 induction (Figure 3G), likely because NAM efficiently suppresses BITREx (Figure 3B). The CNA/G appeared to inversely correlate with the initial *CUP1* copy number (Figure 3H), as expected from the theoretical model, which assumes array contraction via homologous recombination between repeat units following second-order kinetics (see STAR Methods for details).

Note that the effects of NAM on BITREx and the *CUP1* array were completely abolished in the absence of Rtt109, the sole H3K56 acetylase (Figures 3A, 3B, and 3D). These results indicate that, although NAM leads to the accumulation of acetylation at multiple Lys residues, its effects are mediated through the accumulated H3K56ac.

General applicability of BITREx

Theoretically, BITREx can extend a tandem array consisting of two or more repeat units but not a single-copy unit. We constructed strains harboring *CUP1* arrays consisting of one, two, and three repeat units (i.e., 1×*CUP1RU*, 2×*CUP1RU*, and 3×

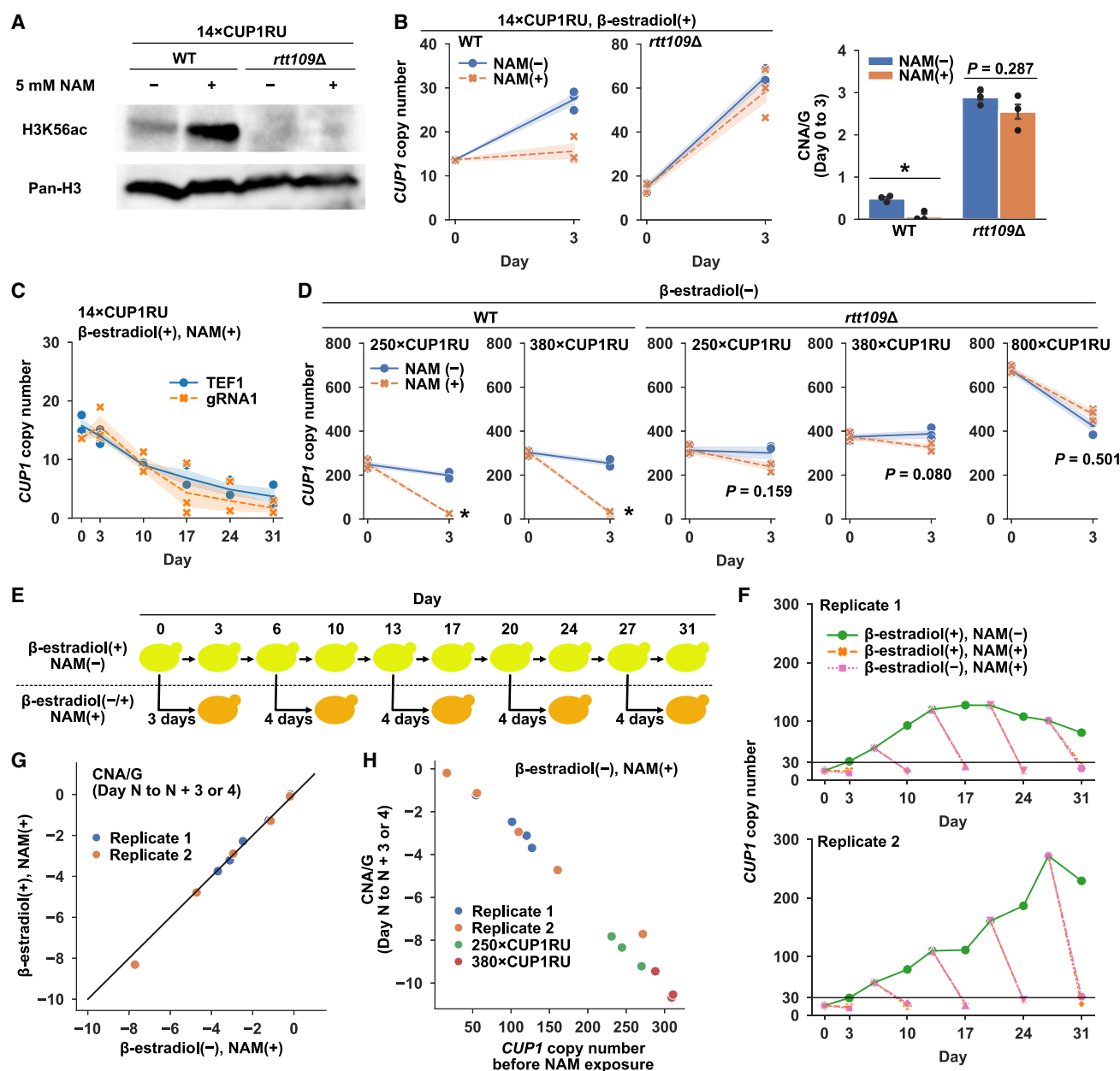


Figure 3. Epigenetic modulation of BITREx

(A) Effects of NAM on H3K56ac. The wild-type and *rtt109Δ* cells were grown in the absence and presence of 5 mM NAM for 3 days and subjected to immunoblot analysis of H3K56ac and total histone H3.

(B) Effects of NAM on BITREx of the CUP1 array using gRNA1. Shading and error bar, SD ($n = 3$ biological replicates). $*p < 0.05$ (Student's t test).

(C) Effects of long-term NAM exposure on the normal CUP1 array in the wild-type strain (14×CUP1RU) with TEF1 gRNA1 and gRNA1. Shading, SD ($n = 3$ biological replicates).

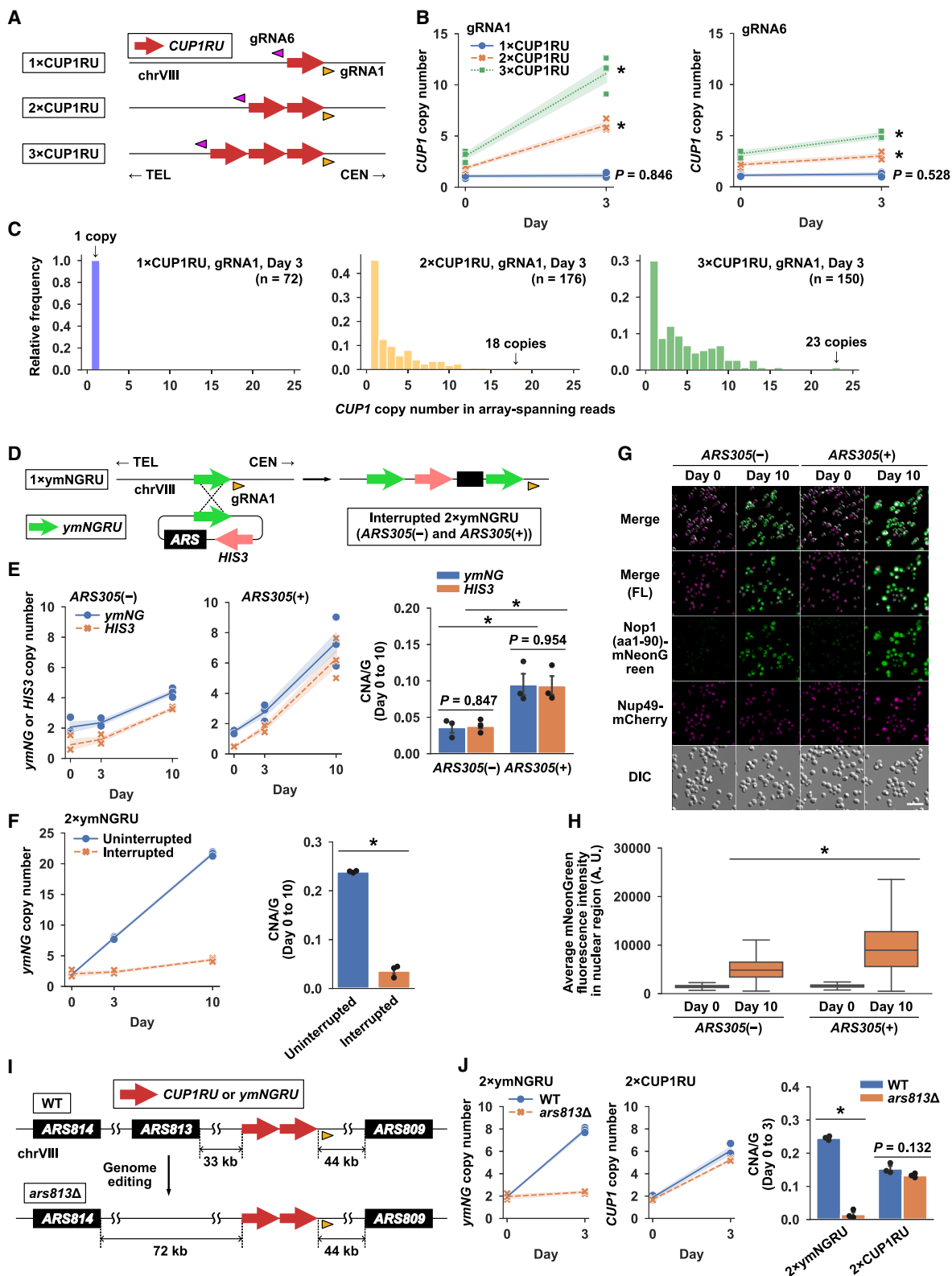
(D) Effects of short-term NAM exposure on the BITREx-extended CUP1 arrays in the presence and absence of Rtt109. Note that BITREx was not induced during the NAM exposure. Shading, SD ($n = 3$ biological replicates). $*p < 0.05$ (Student's t test).

(E) Experimental strategy to examine the effect of initial CUP1 copy number on NAM-induced array contraction.

(F) NAM-induced contraction of variably extended CUP1 arrays. Line plots indicate the actual data following the strategy depicted in (E) ($n = 2$ biological replicates). BITREx was induced in the wild-type strain using gRNA1. The horizontal lines indicate 30 copies.

(G) Effect of BITREx on NAM-induced contraction of extended CUP1 arrays. The CNA/G values are compared between the absence (x axis) and presence (y axis) of BITREx induction. The diagonal line indicates $y = x$.

(H) Plot of CNA/G versus initial CUP1 copy number. Data from (F) and (D) (i.e., 250×CUP1RU and 380×CUP1RU strains) were used.



(legend on next page)

CUP1RU) to determine the minimum number of repeat units required for BITREx (Figure 4A). When targeted to an upstream or downstream flanking site of the *CUP1* array for 3 days, nCas9 increased the *CUP1* copy number in the 2×CUP1RU and 3×CUP1RU strains, but not in the 1×CUP1RU strain (Figure 4B). Nanopore sequencing revealed the expanded *CUP1* arrays in the 2×CUP1RU and 3×CUP1RU strains but not in the 1×CUP1RU strain (Figure 4C). All these results are consistent with the BIR-based mechanism of BITREx.

In addition, we found that BITREx is applicable to the non-*CUP1* natural tandem gene array *ENA1/2/5* encoding P-type ATPase sodium pumps^{24–26} (Figures S7A–S7E); a synthetic two-unit array composed of a fusion gene between *NOP1*²⁷ and *mNeonGreen* (2×*ymNGRU*) at the *CUP1* locus (Figures S7F–S7I); and two-unit arrays, with or without intervening sequences, integrated at loci that do not naturally form tandem repeats (*HO* and *X-2*)^{28,29} (Figures S8A–S8F).

Taken together, BITREx can expand two-unit arrays, whether uninterrupted or interrupted, at different genomic loci. However, the efficiency of expansion varies depending on the locus and the composition of the array.

Effects of ARS on BITREx

The successful expansion of the interrupted two-unit arrays prompted us to explore the possibility of BITREx-mediated amplification of a target sequence embedded between two repeat units, as this configuration can be easily generated using the conventional plasmid integration technique. For example, we integrated a ~5-kb plasmid carrying *ymNGRU* and *HIS3* into the single-copy *ymNGRU* at the *CUP1* locus (Figure 4D) and successfully applied BITREx to the resulting interrupted two-unit array, leading to a simultaneous increase in the copy numbers of *ymNGRU* and *HIS3* (Figure 4E). However, the interrupted 2×*ymNGRU* array showed a significantly lower CNA/G than the uninterrupted 2×*ymNGRU* array (Figure 4F). Given the importance of replication fork directionality (Figures S8A–S8C), we were concerned that the ~3.4-kb plasmid-derived intervening sequence might have decreased the likelihood of achieving the desired fork direction at the nick. To mitigate such adverse effects, we incorporated an autonomously replicating sequence (ARS) into the repeat unit. As expected, incorporating *ARS305*

effectively increased CNA/G and enhanced the fluorescence (Figures 4E, 4G, and 4H).

These results led us to hypothesize that BITREx of a tandem gene array without an internal ARS would be sensitive to the distance between the nick on one side and the nearest ARS on the opposite side of the array. In contrast, BITREx of a tandem gene array with an ARS within the repeat unit would not be affected by this distance. To test this hypothesis by elongating the distance between the nick and the ARS, we generated strains deleted for *ARS813*, the nearest ARS responsible for BITREx using gRNA1 at the *CUP1* locus (Figure 4I). Indeed, BITREx of the 2×*ymNGRU* array, which lacks an internal ARS, depended on *ARS813*, while BITREx of the 2×*CUP1RU* array, which includes *ARS810/811*, did not (Figure 4J). The incorporation of an ARS had no negative effect on BITREx and even enhanced it. Therefore, it is desirable to incorporate an ARS in the repeat unit, especially when no suitable external ARS is available near the target locus or when the repeat unit is long.

Based on these results, we attempted to amplify a multigene array using BITREx. We used plasmid integration to construct strains where the two-unit *CUP1* array at the *CUP1* locus was interrupted by an ~8.5-kb fragment containing yeast codon-optimized coding sequences for four fluorescent proteins (mTagBFP, mRFP682, mCherry, and mNeonGreen) and *HIS3*, with and without *ARS305* (Figure S8G). BITREx successfully extended the array, enhancing gene dosage effects, particularly in the strain with *ARS305* in the repeat unit (Figures S8H–S8J), highlighting the potential of this approach for overexpressing useful gene products.

Splinted BITREx for de novo generation of tandem gene arrays

The minimum requirement for BITREx is two identical sequences to form either an uninterrupted or interrupted two-unit array. However, because BIR frequently switches templates,³⁰ we hypothesized that BITREx could generate a tandem array starting from a single-copy sequence if an engineered DNA serves as a “splint” to mediate the necessary template switching.

To test this idea, we constructed a strain that harbors a repeat unit delimited by *U2* and *LE* (two non-overlapping consecutive fragments of the *LEU2* coding sequence) on chromosome VIII

Figure 4. Core requirements and modulating factors for BITREx

- (A) Schematic of the strains with one to three copies of *CUP1RU*. Orange and magenta arrowheads indicate the target sites of gRNA1 and gRNA6, respectively.
- (B) *CUP1* CNA in the three strains depicted in (A). Shading, SD ($n = 3$ biological replicates). * $p < 0.05$ (Student's *t* test).
- (C) Distribution of *CUP1* copy number in nanopore reads spanning the entire array. Nanopore sequencing was performed on day 3 of BITREx.
- (D) Plasmid integration strategy to generate an interrupted 2×*ymNGRU* array. The intervening sequence was composed of the plasmid backbone containing *HIS3* with or without *ARS305*.
- (E) CNA of *ymNGRU* and *HIS3* in the absence and presence of the embedded *ARS305*. Shading and error bar, SD ($n = 3$ biological replicates). * $p < 0.05$ (Student's *t* test).
- (F) CNA of *ymNGRU* in the uninterrupted and interrupted two-unit arrays. Shading and error bar, SD ($n = 3$ biological replicates). * $p < 0.05$ (Student's *t* test).
- (G) Microscopic images of strains bearing the interrupted 2×*ymNGRU* arrays without and with the embedded *ARS305*. These strains have *NUP49-mCherry* to visualize the nuclei (magenta). FL, fluorescence; DIC, differential interference contrast. Scale bar, 20 μm .
- (H) Quantification of mNeonGreen fluorescence in (G). Boxplots indicate the distribution of the average mNeonGreen fluorescence intensity in the nuclear region. The bottom and top of the box show the first and third quartiles, respectively. The bar in each box represents the median value, and the error bars represent the range of values. * $p < 0.001$ (one-way ANOVA test).
- (I) ARS distribution around the *CUP1* locus in the wild-type and *ars813Δ* strain. Arrowhead, gRNA1 target site.
- (J) Effect of *ARS813* on BITREx of the uninterrupted 2×*ymNGRU* and 2×*CUP1RU* arrays. The 2×*ymNGRU* array lacks internal ARS, while the 2×*CUP1RU* array contains *ARS810/811* within the repeat unit. Shading and error bar, SD ($n = 3$ biological replicates). * $p < 0.05$ (Student's *t* test).

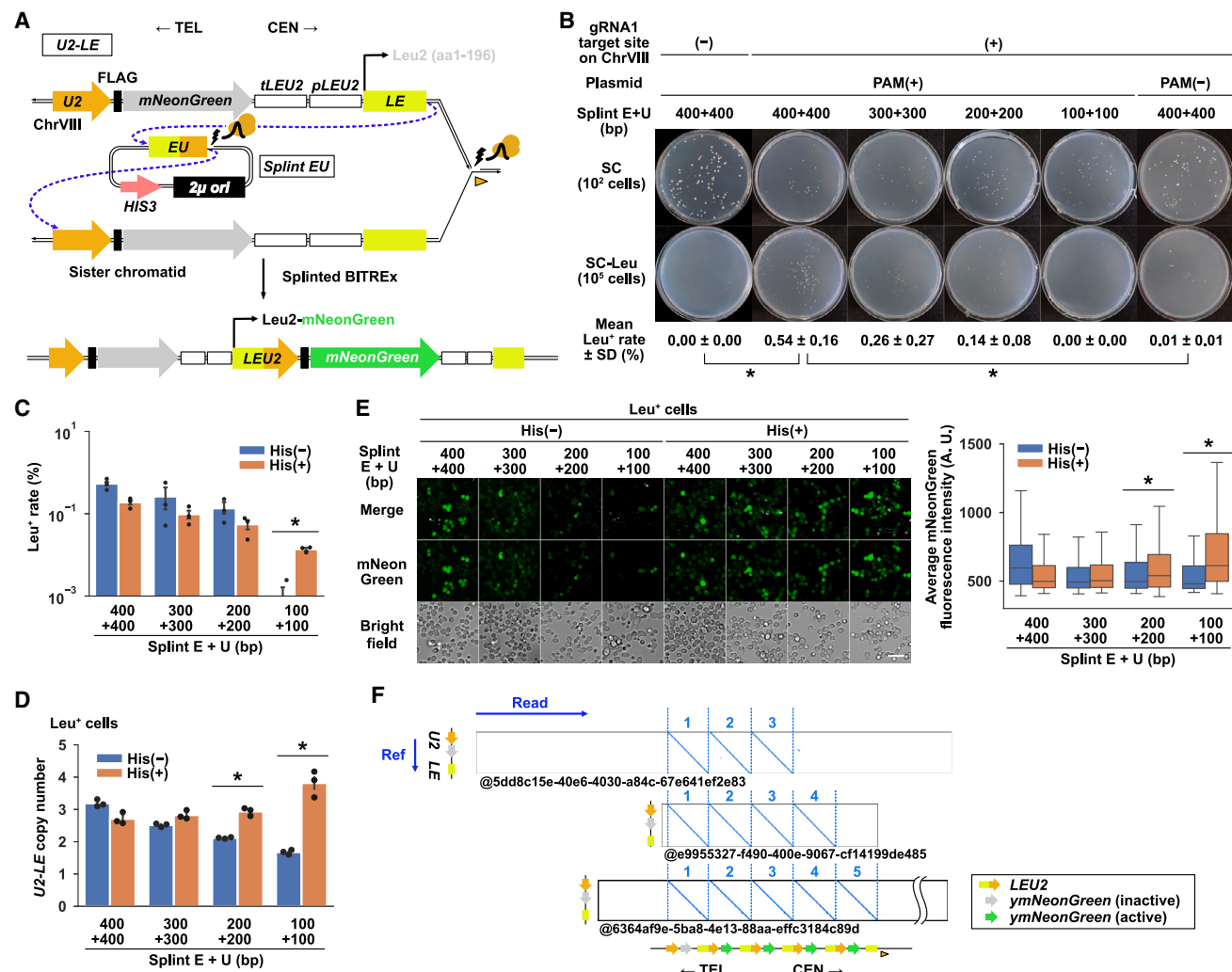


Figure 5. De novo generation of tandem gene array by splinted BITREx

(A) Proof-of-concept experiment for splinted BITREx. A single round of splinted BITREx reconstitutes the *LEU2*-coding sequence, resulting in the expression of a fusion protein composed of Leu2 and mNeonGreen connected by the FLAG tag. *tLEU2*, *LEU2* terminator; *pLEU2*, *LEU2* promoter; orange arrowhead, gRNA1 target site.

(B) Efficiency of splinted BITREx assessed by the appearance of Leu⁺ clones. Splints of four different lengths were tested over 3 days of BITREx. Representative images show colonies formed on SC and SC-Leu agar plates, inoculated with approximately 10² and 10⁵ cells, respectively. The mean Leu⁺ rate with SD is shown at the bottom of the panel ($n = 3$ biological replicates). * $p < 0.05$ (Student's *t* test).

(C) Effects of histidine supplementation from day 2 on the appearance of Leu⁺ colonies. Error bar, SD ($n = 3$ biological replicates). * $p < 0.05$ (Student's *t* test).

(D) Effects of histidine supplementation from day 2 on the copy number of *U2-LE* unit. Error bar, SD ($n = 3$ biological replicates). * $p < 0.05$ (Student's *t* test).

(E) Effects of histidine supplementation from day 2 on the mNeonGreen fluorescence. Left, microscopic images of strains subjected to splinted BITREx. Scale bar, 20 μ m. Right, boxplots showing the distribution of the average mNeonGreen fluorescence intensity in cells. * $p < 0.05$ (Student's *t* test).

(F) Dot plot comparing nanopore reads to the reference sequence of *U2-LE* unit.

and a plasmid carrying the splint fragment *EU* (Figure 5A). Upon the induction of a seDSB by targeting nCas9 to a downstream-flanking site of the genomic *LE*, the ssDNA generated by end-resection of the break site may use its *E* to invade the episomal *EU* for further extension by displacement synthesis. The extended ssDNA may then use the newly acquired *U* to switch its template from the splint plasmid to the genomic *U2*. If these consecutive strand invasion-extensions occur, the *U2-LE* unit duplicates to reconstitute the *LEU2* gene, allowing the growth in the absence of leucine. Indeed, we observed the emergence

of Leu⁺ cells with different efficiencies depending on the length of the splint fragment (Figure 5B). Notably, co-nicking of the chromosome and the splint plasmid was essential for the *LEU2* reconstitution (Figure 5B).

Once the *LEU2* gene is reconstituted, the *LE* ssDNA can initiate BIR either by directly invading the reconstituted *LEU2* or indirectly invading the genomic *U2* or *LEU2* via the splint plasmid (Figure 5A). Since the direct invasion should be much more efficient than the indirect invasion, the splint plasmid may inhibit subsequent cycles of BITREx, even though it is essential

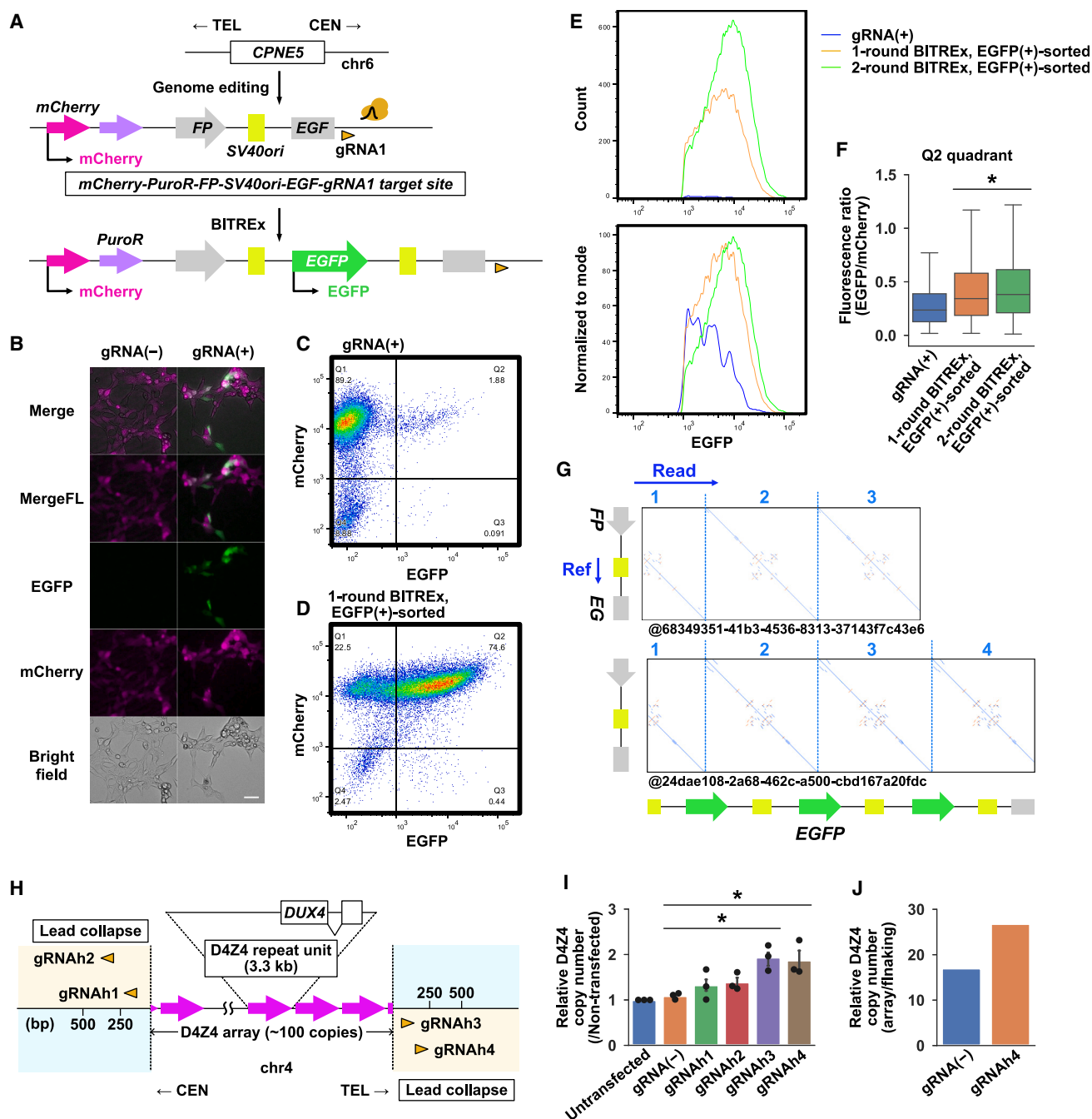


Figure 6. BITREx in mammalian cells

(A) Proof-of-concept experiment for BITREx in HEK293T cells using an EGFP reconstitution reporter integrated into the CPNE5 locus.

(B) Microscopic images of HEK293T cells with the integrated reporter on day 4 after transfection of the nCas9-gRNA1 expression plasmid. Scale bar, 50 μ m.

(C) Flow cytometric analysis of mCherry and EGFP expression on day 3 after transfection of the nCas9-gRNA1 expression plasmid. The cells gated based on light scatter (Figure S9C) displayed mCherry and EGFP expression across four quadrants (Q1, Q2, Q3, Q4), with the percentage distribution indicated in each quadrant.

(D) Flow cytometric analysis of mCherry and EGFP expression in the mCherry⁺/EGFP⁺ cells sorted from quadrant Q2 in (C).

(E) Distribution of EGFP fluorescence. Blue, the total population of cells transfected with the gRNA1-nCas9 co-expression plasmid; orange, flow-sorted EGFP⁺ cells; green, flow-sorted EGFP⁺ cells on day 3 after re-transfection with the co-expression plasmid. The x axis represents EGFP fluorescence intensity, while the y axis represents either the number of cells (top) or the value normalized to mode (bottom).

(F) Boxplots showing the fluorescence ratio of EGFP to mCherry in the cells in (E). **p* < 0.05 (Student's *t* test).

(legend continued on next page)

for the initial reconstitution of *LEU2*. Based on these considerations, we examined the effect of histidine supplementation starting from day 2 to promote the spontaneous loss of the *HIS3*-marked splint plasmid. This supplementation protocol improved the emergence of *Leu*⁺ clones when combined with the shortest splint fragments (Figure 5C). It also increased the copy number of *U2-LE* units and the mNeonGreen fluorescence when combined with the shortest or second shortest splint fragment (Figures 5D and 5E). Nanopore sequencing identified reads spanning a tandem array composed of up to five *U2-LE* units or four *LEU2-mNeonGreen*-fusion genes (Figure 5F).

These results demonstrated that “splinted BITREx” enables *de novo* formation of a tandem gene array from a single-copy sequence.

BITREx in mammalian cells

We next investigated the feasibility of BITREx in mammalian cells. Since BIR has been demonstrated in mammalian cells using the reconstitution of a fluorescent protein gene,³¹ we generated a reporter construct *mCherry-PuroR-FP-SV40ori-EGF-gRNA1 target site* (Figure 6A). In this construct, the *mCherry-PuroR* portion serves as a transfection reporter/ marker, while the *FP-SV40ori-EGF* portion serves as an interrupted two-unit array, in which two *F* fragments are interrupted by the *P-SV40ori-EG* fragment. Therefore, nCas9 with gRNA1 should induce BIR via the *F* fragment to reconstitute the *EGFP* gene, resulting in green fluorescence.

We initially aimed to integrate the construct to the safe harbor locus *AAVS1* of human HEK293T cells, which expresses the large T antigen that activates *SV40ori*, the replication origin of the SV40 virus, using the VIKING method for efficient NHEJ-based knockin (Figure S9A).³² However, genotyping and nanopore sequencing of the puromycin-selected cells revealed that the construct had been integrated not into *AAVS1* but into *CPNE5*, a previously reported off-target site for the *AAVS1* gRNA we used (Figure S9B).³³ Despite this, these cells provided a valuable opportunity to investigate the feasibility of BITREx in mammalian cells. To this end, we transfected them with a plasmid that co-expresses nCas9 and gRNA1 to induce BITREx (Figure S9A). As a fraction of the transfected cells started to show EGFP fluorescence (Figure 6B), we flow-sorted the *mCherry*⁺/*EGFP*⁺ cells (Figures 6C, 6D, and S9C). A subsequent round of transfection with the co-expression plasmid conferred enhanced EGFP fluorescence to the cells (Figure 6E). As expected from the design of the reporter construct, in which BITREx increases the copy number of the *EGFP* gene but not the *mCherry* gene (Figure 6A), the EGFP/*mCherry* fluorescence ratio increased (Figures 6F and S9D). Nanopore sequencing of these cells identified reads containing at least three or four copies of the *FP-SV40ori-EG* unit (Figure 6G). These

results demonstrated the feasibility of BITREx in mammalian cells.

Note that we cannot rule out the possibility that the results described above were influenced by residual plasmids encoding the reporter construct in the puromycin-selected cells (Figure S9E). Therefore, we sought to apply BITREx to gene-sized tandem repeats naturally occurring in the human genome. For this purpose, we focused on the D4Z4 array, which consists of a 3.3-kb repeat unit on chromosome 4q35. This array is significant because its heterozygous contraction causes facioscapulo-humeral muscular dystrophy (FSHD), the third most common type of inherited muscular dystrophy.³⁴ Accordingly, its expansion may have potential implications in FSHD therapeutics. We designed four gRNAs to target nCas9 to the centromeric and telomeric sides of the D4Z4 array (Figure 6H). The results of qPCR consistently demonstrated that targeting nCas9 to the telomeric side increased the D4Z4 copy number (Figure 6I). We subjected the cells with no gRNA and the most effective gRNA (gRNAh4) to nanopore sequencing. The normalized read count indicated that the D4Z4 copy number increased from ~17 to ~27 copies (Figure 6J), consistent with the qPCR results (Figure 6I). Although the complexity of the D4Z4 locus containing many repetitive sequences and the presence of an almost identical locus on chromosome 10q26 precluded the complete characterization of the induced CNAs, these results demonstrated the applicability of BITREx to endogenous tandem gene arrays in mammalian genomes.

DISCUSSION

We have developed BITREx, a method for expanding tandem gene arrays through continuous ectopic BIR induced by strategically targeting nCas9. In BITREx, nCas9 is placed adjacent to a tandem gene array, disrupting the replication fork after it has replicated the array. We previously developed paired nicking-induced amplification (PNAmP), a method for inducing large segmental duplications by paired nicking-induced SSA.³⁵ Similarly, others have developed methods using a pair of prime editors, although the underlying mechanisms remain somewhat unproven.^{36,37} While both PNAmP and BITREx induce structural variations by manipulating replication fork progression, they are mechanistically distinct: PNAmP uses two gRNAs and depends on SSA,³⁵ whereas BITREx uses one gRNA and relies on BIR. The basic requirements for a PNAmP target are an internal ARS and terminal direct repeats, or an interrupted two-unit array with an ARS.³⁵ This is because both nicks at the flanking sites of the target segment must be crossed by outward replication forks. In contrast, BITREx can be applied to an interrupted two-unit array even without an ARS, as it can be designed to work as long as at least one of the flanking sites is crossed by

(G) Dot plots comparing nanopore reads to the reference sequence of the *FP-EG* unit. Nanopore whole-genome sequencing identified 38 reads that included the *FP-EG* unit, 26 of which supported the expected expansion. The upper and lower reads indicate at least triplication and quadruplication, respectively.

(H) Schematic of the D4Z4 array and its flanking regions on human chromosome 4. Arrowheads indicate the target sites of four gRNAs, which guide nCas9 to induce lead collapse of the replication fork moving outward from the D4Z4 array.

(I) D4Z4 copy number quantified by qPCR on day 3 of BITREx with the indicated gRNA. Data from transfected cells were normalized to untransfected control cells. Error bar, SD ($n = 3$ biological replicates). * $p < 0.05$ (Student's *t* test).

(J) Ratio of normalized read counts between the D4Z4 macrosatellite and the centromeric flanking region.

an outward replication fork. Therefore, BITREx is suitable for a broader range of targets compared with PNAmP.

For BITREx to be effective, the replication fork crossing the nick must move from inside to outside the tandem gene array. If the repeat unit lacks an ARS, BITREx depends on replication initiated from an external ARS flanking the tandem gene array on the opposite side of the induced nick. As the array expands, the distance between the ARS and the nick increases, thus decreasing the likelihood that the outward replication fork will reach the nick earlier than the inward fork initiated from the nearest ARS on the same side as the nick. Consequently, the success rate of BITREx per cell cycle declines, reaching a plateau to limit the maximum expansion range. In contrast, if the repeat unit contains an ARS, cells can more reliably maintain the desired replication fork direction. The internal ARS makes BITREx more autonomous and less dependent on an external ARS. Therefore, including an ARS in the repeat unit of a synthetic array is advantageous. Although autonomous BITREx theoretically permits unlimited expansion, the increased instability of highly extended arrays appears to counteract this potential, leading to a plateau where expansion by BITREx is balanced by contraction due to intrinsic instability. Moreover, prolonged nicking can occasionally induce mutations at the target site, leading to progenies that are free from the nCas9-induced seDSB: these progenies should outcompete those with the seDSB in proliferation speed, eventually dominating the cell population. These factors likely impose practical limits on the extent of expansion.

The efficiency of BITREx is influenced by multiple factors, as demonstrated by the CNA/G values for six two-unit arrays integrated at three different loci (Table S1). Arrays embedded at the same locus exhibited varying efficiencies, emphasizing the importance of repeat unit characteristics. Conversely, identical arrays behaved differently depending on their integration loci, highlighting the influence of the local genomic environment. Collectively, these results suggest that the overall efficiency of BITREx is governed by a complex interplay of factors, making precise predictions inherently challenging.

Interestingly, BITREx requires nCas9 to nick the template DNA for leading strand synthesis, but not for lagging strand synthesis. This strand specificity likely reflects the asymmetry observed in the repair of replication-coupled DNA breaks, as revealed in recent studies on mammalian cells: nicks on the leading and lagging strand templates lead to the formation of seDSBs and double-ended DSBs (deDSBs), respectively.^{38,39} The replisome notably bypasses nicks on the lagging strand templates to generate deDSBs directly, or without contribution from the converging replication fork, thereby likely preventing BIR. Intriguingly, one yeast study showed that a nick on a leading strand template induces a seDSB, while another nick on a different leading strand template induces a deDSB, independently of the converging replication fork.⁴⁰ Although the determinants of these differential fates remain elusive, these variations may partly explain why 10 out of the 16 gRNAs designed for lead collapse were ineffective.

While BITREx increased the average copy number of repeat units in the population, nanopore sequencing revealed the presence of contracted arrays. We hypothesized that extensive 5'-to-3' end-resection and/or the "chewing back" of the invading strand by the 3'-to-5' exonuclease activity of Pol δ ⁴¹

contribute to the contraction (Figures S4A and S4B). Therefore, appropriate suppression of these end resection activities may prevent contraction events and improve BITREx. These situations are more likely to occur when the repeat unit is short, meaning BITREx may not be effective in expanding micro- and mini-satellite DNAs, particularly when the repeat number is low. However, BITREx could still be valuable for extending satellite DNA-like repeats, enabling the synthesis of centromere-like DNA—an ongoing challenge in synthetic genomics.

An obvious application of BITREx would be the overexpression of genes of interest. We have successfully applied BITREx to an interrupted two-unit array containing four fluorescent protein genes as the intervening sequence, resulting in their overexpression to exert a dosage effect. These results suggest that when applying to a two-unit array with a biosynthetic gene cluster as the intervening sequence, BITREx can enhance the yield of the biosynthetic pathway's product. In contrast to the serial configuration within a gene cluster, individual genes encoding pathway components can be distributed across multiple loci, where BITREx could act in parallel to increase their copy numbers. BITREx occurs stochastically in each cell cycle. Once it occurs, the subsequent cell division becomes asymmetric regarding unit copy number: one daughter cell inherits the donor chromatid with the original array, while the other inherits the acceptor chromatid with the expanded array. In addition, the efficiency of BITREx varies from one locus to another. Therefore, parallel BITREx would create a cell population with diverse stoichiometry among pathway components, which could help identify an optimal pathway design for maximizing the yield of the pathway's product.

It should be noted that BIR is less accurate than normal S-phase replication.⁴² During BIR, the Pif1 helicase immediately dissociates the newly synthesized leading strand DNA from its template, leaving it single-stranded until lagging strand synthesis occurs (Figure S1B). As a result, the mismatch repair system functions ineffectively in BIR. Furthermore, the exposed nucleobases in ssDNA are much more susceptible to damage compared with those in double-stranded DNA. Indeed, inducing BIR in yeast in the presence of an alkylating agent has led to the formation of mutation clusters similar to those found in cancer genomes.⁴³ Therefore, we hypothesize that BITREx in the presence of mutagens, or mutagenic BITREx, will not only expand a tandem gene array but also diversify its repeat unit sequence, potentially generating an array of paralogs similar to OR gene loci. Moreover, gap repair cloning can be employed to randomly isolate individual paralogs and place them under a promoter, producing a unique population of cells in which each clone expresses a distinct paralog, mirroring the diversity seen in olfactory neurons or lymphocytes. These diverse cell populations hold promise for unique applications in both basic research and applied studies.

It is also noteworthy that BITREx may represent the first method for the targeted expansion of macrosatellite repeats with medical or therapeutic significance, such as D4Z4, DXZ4, and the exon array in the *LPA* gene.^{34,44,45} This approach could advance mechanistic studies of macrosatellite-associated diseases and potentially contribute to the development of regenerative therapies.

We anticipate that BITREx will enable these and other unique applications in genome engineering.

Limitations of the study

BITREx is a replication-coupled process and thus not applicable to non-dividing cells. By its nature, it cannot effectively expand tandem gene arrays located between actively firing nearby replication origins. Currently, predicting the performance of BITREx is challenging due to its reliance on various factors, including gRNA efficacy in inducing seDSB, the local environment of the target locus, and the composition of the repeat unit sequence, as some sequences can hinder BIR.⁴⁶ In addition, there is no method for stably maintaining tandem gene arrays that have been highly expanded by BITREx. Further investigations are required to optimize BITREx in mammalian cells, including analyses of the temporal dynamics of array expansion and stability, which are expected to differ from those observed in yeast.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be addressed to and will be fulfilled by the lead contact, Takashi Ito (ito.takashi.352@m.kyushu-u.ac.jp).

Materials availability

Requests for the generated plasmids and strains in this study should be directed to the lead contact, Takashi Ito (ito.takashi.352@m.kyushu-u.ac.jp).

Data and code availability

- All raw sequencing data used in this study were deposited in the DDBJ BioProject database: PRJDB18647, PRJDB18687, PRJDB18705.
- All original codes used in this study are available from Zenodo at <https://doi.org/10.5281/zenodo.11515696>.

ACKNOWLEDGMENTS

We thank Tetsuya Hayashi and Yasuhiro Gotoh for the PFGE equipment. We appreciate the technical assistance from the Research Support Center of the Research Center for Human Disease Modeling at Kyushu University Graduate School of Medical Sciences, which is partially supported by the Mitsuaki Shiraishi Fund for Basic Medical Research. This work was funded by JST CREST grant no. JPMJCR19S1 and JSPS KAKENHI grant no. 24K02015.

AUTHOR CONTRIBUTIONS

Conceptualization, H.T., S.O., and T.I.; methodology, H.T., S.O., and G.D.; investigation, H.T. and S.O.; writing – original draft, H.T.; writing – review & editing, H.T., S.O., and T.I.; funding acquisition, T.I.; resources, H.T., S.O., G.D., Y.S., and E.K.; supervision, T.I.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES

During the preparation of this work, the authors used ChatGPT to improve the readability of certain sentences. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Yeast strains
 - Yeast plasmids
 - Yeast genome editing
 - Yeast cell culture
 - Quantitative PCR (qPCR)
 - Modeling the contraction of extended *CUP1* array
 - Nanopore sequencing
 - Dot plot analysis of nanopore reads
 - Translocation analysis using nanopore reads
 - Copy number estimation from nanopore reads
 - Illumina sequencing
 - Pulsed-field gel electrophoresis and Southern blot hybridization
 - Immunoblot analysis
 - Fluorescence microscopy and image processing for yeast cells
 - BITREx of *EGFP* reconstitution reporter in HEK293T cells
 - Flow cytometric analysis of transfected HEK293T cells
 - Flow sorting of EGFP-positive HEK293T cells
 - Fluorescence microscopy and image processing for HEK293T cells
 - BITREx of the D4Z4 array in HEK293T cells
 - Nanopore sequencing of genomic DNA prepared from HEK293T cells
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xgen.2025.100811>.

Received: September 20, 2024

Revised: December 20, 2024

Accepted: February 17, 2025

Published: March 20, 2025

REFERENCES

1. Zhang, J. (2003). Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18, 292–298. [https://doi.org/10.1016/s0169-5347\(03\)00033-8](https://doi.org/10.1016/s0169-5347(03)00033-8).
2. Axelsson, E., Ratnakumar, A., Arendt, M.-L., Maqbool, K., Webster, M.T., Perloski, M., Liberg, O., Arnemo, J.M., Hedhammar, Å., and Lindblad-Toh, K. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495, 360–364. <https://doi.org/10.1038/nature11837>.
3. Zhao, Y., Strobe, P.K., Kozmin, S.G., McCusker, J.H., Dietrich, F.S., Koskoska, R.J., and Petes, T.D. (2014). Structures of Naturally Evolved *CUP1* Tandem Arrays in Yeast Indicate That These Arrays Are Generated by Unequal Nonhomologous Recombination. *G3 (Bethesda)* 4, 2259–2269. <https://doi.org/10.1534/g3.114.012922>.
4. Warringer, J., Zörgö, E., Cubillos, F.A., Zia, A., Gjuvsland, A., Simpson, J.T., Forsmark, A., Durbin, R., Omholt, S.W., Louis, E.J., et al. (2011). Trait Variation in Yeast Is Defined by Population History. *PLoS Genet.* 7, e1002111. <https://doi.org/10.1371/journal.pgen.1002111>.
5. Fogel, S., Welch, J.W., Cathala, G., and Karin, M. (1983). Gene amplification in yeast: *CUP1* copy number regulates copper resistance. *Curr. Genet.* 7, 347–355. <https://doi.org/10.1007/BF00445874>.
6. Niimura, Y., Matsui, A., and Touhara, K. (2014). Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Res.* 24, 1485–1496. <https://doi.org/10.1101/gr.169532.113>.
7. Cheetham, S.W., Faulkner, G.J., and Dinger, M.E. (2020). Overcoming challenges and dogmas to understand the functions of pseudogenes. *Nat. Rev. Genet.* 21, 191–201. <https://doi.org/10.1038/s41576-019-0196-1>.

8. Doi, G., Okada, S., Yasukawa, T., Sugiyama, Y., Bala, S., Miyazaki, S., Kang, D., and Ito, T. (2021). Catalytically inactive Cas9 impairs DNA replication fork progression to induce focal genomic instability. *Nucleic Acids Res.* 49, 954–968. <https://doi.org/10.1093/nar/gkaa1241>.
9. Vrtis, K.B., Dewar, J.M., Chistol, G., Wu, R.A., Graham, T.G.W., and Walter, J.C. (2021). Single-strand DNA breaks cause replisome disassembly. *Mol. Cell* 81, 1309–1318.e6. <https://doi.org/10.1016/j.molcel.2020.12.039>.
10. Kockler, Z.W., Osia, B., Lee, R., Musmaker, K., and Malkova, A. (2021). Repair of DNA Breaks by Break-Induced Replication. *Annu. Rev. Biochem.* 90, 165–191. <https://doi.org/10.1146/annurev-biochem-081420-095551>.
11. Epum, E.A., and Haber, J.E. (2022). DNA replication: the recombination connection. *Trends Cell Biol.* 32, 45–57. <https://doi.org/10.1016/j.tcb.2021.07.005>.
12. Mayle, R., Campbell, I.M., Beck, C.R., Yu, Y., Wilson, M., Shaw, C.A., Bjergbaek, L., Lupski, J.R., and Ira, G. (2015). Mus81 and converging forks limit the mutagenicity of replication fork breakage. *Science* 349, 742–747. <https://doi.org/10.1126/science.aaa8391>.
13. Anand, R., Beach, A., Li, K., and Haber, J. (2017). Rad51-mediated double-strand break repair and mismatch correction of divergent substrates. *Nature* 544, 377–380. <https://doi.org/10.1038/nature22046>.
14. McIsaac, R.S., Silverman, S.J., McClean, M.N., Gibney, P.A., Macinskas, J., Hickman, M.J., Petti, A.A., and Botstein, D. (2011). Fast-acting and nearly gratuitous induction of gene expression and protein depletion in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 22, 4447–4459. <https://doi.org/10.1091/mbc.e11-05-0466>.
15. Lee, J., Lim, K., Kim, A., Mok, Y.G., Chung, E., Cho, S.-I., Lee, J.M., and Kim, J.-S. (2023). Prime editing with genuine Cas9 nickases minimizes unwanted indels. *Nat. Commun.* 14, 1786. <https://doi.org/10.1038/s41467-023-37507-8>.
16. Vriend, L.E.M., Prakash, R., Chen, C.C., Vanoli, F., Cavallo, F., Zhang, Y., Jasin, M., and Krawczyk, P.M. (2016). Distinct genetic control of homologous recombination repair of Cas9-induced double-strand breaks, nicks and paired nicks. *Nucleic Acids Res.* 44, 5204–5217. <https://doi.org/10.1093/nar/gkw179>.
17. Bothmer, A., Phadke, T., Barrera, L.A., Margulies, C.M., Lee, C.S., Buquichio, F., Moss, S., Abdulkarim, H.S., Selleck, W., Jayaram, H., et al. (2017). Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat. Commun.* 8, 13905. <https://doi.org/10.1038/ncomms13905>.
18. Schulz, V.P., and Zakian, V.A. (1994). The *saccharomyces* PIF1 DNA helicase inhibits telomere elongation and de novo telomere formation. *Cell* 76, 145–155. [https://doi.org/10.1016/0092-8674\(94\)90179-1](https://doi.org/10.1016/0092-8674(94)90179-1).
19. Che, J., Smith, S., Kim, Y.J., Shim, E.Y., Myung, K., and Lee, S.E. (2015). Hyper-Acetylation of Histone H3K56 Limits Break-Induced Replication by Inhibiting Extensive Repair Synthesis. *PLoS Genet.* 11, e1004990. <https://doi.org/10.1371/journal.pgen.1004990>.
20. Ide, S., Saka, K., and Kobayashi, T. (2013). Rtt109 prevents hyper-amplification of ribosomal RNA genes through histone modification in budding yeast. *PLoS Genet.* 9, e1003410. <https://doi.org/10.1371/journal.pgen.1003410>.
21. Radovani, E., Cadorin, M., Shams, T., El-Rass, S., Karsou, A.R., Kim, H.S., Kurat, C.F., Keogh, M.C., Greenblatt, J.F., and Fillingham, J.S. (2013). The carboxyl terminus of Rtt109 functions in chaperone control of histone acetylation. *Eukaryot. Cell* 12, 654–664. <https://doi.org/10.1128/ec.00291-12>.
22. Tsuchiya, M., Dang, N., Kerr, E.O., Hu, D., Steffen, K.K., Oakes, J.A., Kennedy, B.K., and Kaeblerlein, M. (2006). Sirtuin-independent effects of nicotinamide on lifespan extension from calorie restriction in yeast. *Aging Cell* 5, 505–514. <https://doi.org/10.1111/j.1474-9726.2006.00240.x>.
23. Hull, R.M., Cruz, C., Jack, C.V., and Houseley, J. (2017). Environmental change drives accelerated adaptation through stimulated copy number variation. *PLoS Biol.* 15, e2001333. <https://doi.org/10.1371/journal.pbio.2001333>.
24. Haro, R., Garciadeblas, B., and Rodríguez-Navarro, A. (1991). A novel P-type ATPase from yeast involved in sodium transport. *FEBS Lett.* 291, 189–191. [https://doi.org/10.1016/0014-5793\(91\)81280-I](https://doi.org/10.1016/0014-5793(91)81280-I).
25. Strobe, P.K., Skelly, D.A., Kozmin, S.G., Mahadevan, G., Stone, E.A., Magwene, P.M., Dietrich, F.S., and McCusker, J.H. (2015). The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* 25, 762–774. <https://doi.org/10.1101/gr.185538.114>.
26. Wieland, J., Nitsche, A.M., Strayle, J., Steiner, H., and Rudolph, H.K. (1995). The *PMR2* gene cluster encodes functionally distinct isoforms of a putative Na⁺ pump in the yeast plasma membrane. *EMBO J.* 14, 3870–3882. <https://doi.org/10.1002/j.1460-2075.1995.tb00059.x>.
27. Leslie, D.M., Zhang, W., Timney, B.L., Chait, B.T., Rout, M.P., Wozniak, R.W., and Aitchison, J.D. (2004). Characterization of karyopherin cargoes reveals unique mechanisms of Kap121p-mediated nuclear import. *Mol. Cell Biol.* 24, 8487–8503. <https://doi.org/10.1128/mcb.24.19.8487-8503.2004>.
28. Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., Bell, S.P., and Aparicio, O.M. (2001). Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science* 294, 2357–2360. <https://doi.org/10.1126/science.1066101>.
29. Ronda, C., Maury, J., Jakočiunas, T., Jacobsen, S.A.B., Germann, S.M., Harrison, S.J., Borodina, I., Keasling, J.D., Jensen, M.K., and Nielsen, A.T. (2015). CrEdit: CRISPR mediated multi-loci gene integration in *Saccharomyces cerevisiae*. *Microb. Cell Fact.* 14, 97. <https://doi.org/10.1186/s12934-015-0288-3>.
30. Smith, C.E., Llorente, B., and Symington, L.S. (2007). Template switching during break-induced replication. *Nature* 447, 102–105. <https://doi.org/10.1038/nature05723>.
31. Li, S., Wang, H., Jehi, S., Li, J., Liu, S., Wang, Z., Truong, L., Chiba, T., Wang, Z., and Wu, X. (2021). PIF1 helicase promotes break-induced replication in mammalian cells. *EMBO J.* 40, e104509. <https://doi.org/10.15252/embj.2020104509>.
32. Sawatsubashi, S., Joko, Y., Fukumoto, S., Matsumoto, T., and Sugano, S.S. (2018). Development of versatile non-homologous end joining-based knock-in module for genome editing. *Sci. Rep.* 8, 593. <https://doi.org/10.1038/s41598-017-18911-9>.
33. Tasca, F., Brescia, M., Wang, Q., Liu, J., Janssen, J.M., Szuhai, K., and Gonçalves, M.A.F.V. (2022). Large-scale genome editing based on high-capacity adenovectors and CRISPR-Cas9 nucleases rescues full-length dystrophin synthesis in DMD muscle cells. *Nucleic Acids Res.* 50, 7761–7782. <https://doi.org/10.1093/nar/gkac567>.
34. van der Maarel, S.M., and Frants, R.R. (2005). The D4Z4 repeat-mediated pathogenesis of facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* 76, 375–386. <https://doi.org/10.1086/428361>.
35. Sugiyama, Y., Okada, S., Daigaku, Y., Kusumoto, E., and Ito, T. (2024). Strategic targeting of Cas9 nickase induces large segmental duplications. *Cell Genom.* 4, 100610. <https://doi.org/10.1016/j.xgen.2024.100610>.
36. Jiao, Y., Li, M., He, X., Wang, Y., Song, J., Hu, Y., Li, L., Zhou, L., Jiang, L., Qu, J., et al. (2023). Targeted, programmable, and precise tandem duplication in the mammalian genome. *Genome Res.* 33, 779–786. <https://doi.org/10.1101/gr.277261.122>.
37. Zhang, R., He, Z., Shi, Y., Sun, X., Chen, X., Wang, G., Zhang, Y., Gao, P., Wu, Y., Lu, S., et al. (2024). Amplification editing enables efficient and precise duplication of DNA from short sequence to megabase and chromosomal scale. *Cell* 187, 3936–3952.e19. <https://doi.org/10.1016/j.cell.2024.05.056>.
38. Pavan, R., Tripathi, V., Vrtis, K.B., Zong, D., Chari, R., Callen, E., Pankajam, A.V., Zhen, G., Matos-Rodriguez, G., Yang, J., et al. (2024). Structure and repair of replication-coupled DNA breaks. *Science* 385, eado3867. <https://doi.org/10.1126/science.ado3867>.
39. Elango, R., Nilavar, N.M., Li, A.G., Nguyen, D., Rass, E., Duffey, E.E., Jiang, Y., Abakir, A., Willis, N.A., Houseley, J., and Scully, R. (2025).

- Two-ended recombination at a Flp-nickase-broken replication fork. *Mol. Cell* 85, 78–90.e3. <https://doi.org/10.1016/j.molcel.2024.11.006>.
40. Kimble, M.T., Sane, A., Reid, R.J.D., Johnson, M.J., Rothstein, R., and Symington, L.S. (2025). Repair of replication-dependent double-strand breaks differs between the leading and lagging strands. *Mol. Cell* 85, 61–77.e6. <https://doi.org/10.1016/j.molcel.2024.10.032>.
 41. Choi, J., Kong, M., Gallagher, D.N., Li, K., Bronk, G., Cao, Y., Greene, E.C., and Haber, J.E. (2022). Repair of mismatched templates during Rad51-dependent Break-Induced Replication. *PLoS Genet.* 18, e1010056. <https://doi.org/10.1371/journal.pgen.1010056>.
 42. Deem, A., Keszthelyi, A., Blackgrove, T., Vayl, A., Coffey, B., Mathur, R., Chabes, A., and Malkova, A. (2011). Break-Induced Replication Is Highly Inaccurate. *PLoS Biol.* 9, e1000594. <https://doi.org/10.1371/journal.pbio.1000594>.
 43. Sakofsky, C.J., Roberts, S.A., Malc, E., Mieczkowski, P.A., Resnick, M.A., Gordenin, D.A., and Malkova, A. (2014). Break-induced replication is a source of mutation clusters underlying kataegis. *Cell Rep.* 7, 1640–1648. <https://doi.org/10.1016/j.celrep.2014.04.053>.
 44. Strathmann, E.A., Hölker, I., Tschernoster, N., Hosseinibarkooie, S., Come, J., Martinat, C., Altmüller, J., and Wirth, B. (2023). Epigenetic regulation of plastin 3 expression by the macrosatellite DXZ4 and the transcriptional regulator CHD4. *Am. J. Hum. Genet.* 110, 442–459. <https://doi.org/10.1016/j.ajhg.2023.02.004>.
 45. Schmidt, K., Noureen, A., Kronenberg, F., and Utermann, G. (2016). Structure, function, and genetics of lipoprotein (a). *J. Lipid Res.* 57, 1339–1359. <https://doi.org/10.1194/jlr.R067314>.
 46. Liu, L., Yan, Z., Osia, B.A., Twarowski, J., Sun, L., Kramara, J., Lee, R.S., Kumar, S., Elango, R., Li, H., et al. (2021). Tracking break-induced replication shows that it stalls at roadblocks. *Nature* 590, 655–659. <https://doi.org/10.1038/s41586-020-03172-w>.
 47. Natsume, T., Kiyomitsu, T., Saga, Y., and Kanemaki, M.T. (2016). Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Rep.* 15, 210–218. <https://doi.org/10.1016/j.celrep.2016.03.001>.
 48. Chiang, T.W.W., le Sage, C., Larrieu, D., Demir, M., and Jackson, S.P. (2016). CRISPR-Cas9(D10A) nickase-based genotypic and phenotypic screening to enhance genome editing. *Sci. Rep.* 6, 24356. <https://doi.org/10.1038/s41598-017-18911-9>.
 49. De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34, 2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>.
 50. Noé, L., and Kucherov, G. (2005). YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Res.* 33, W540–W543. <https://doi.org/10.1093/nar/gki478>.
 51. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
 52. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. <https://doi.org/10.1093/gigascience/giab008>.
 53. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
 54. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
 55. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
 56. Brachmann, C.B., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., and Boeke, J.D. (1998). Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* 14, 115–132. [https://doi.org/10.1002/\(SICI\)1097-0061\(19980130\)14:2<115::AID-YEA204>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2).
 57. Guthrie, C., and Fink, G.R. (1991). *Methods in Enzymology: Guide to Yeast Genetics and Molecular Biology* (Academic Press).
 58. DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., and Church, G.M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* 41, 4336–4343. <https://doi.org/10.1093/nar/gkt135>.
 59. Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491. <https://doi.org/10.1016/j.cell.2013.12.001>.
 60. Naito, Y., Hino, K., Bono, H., and Ui-Tei, K. (2015). CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* 31, 1120–1123. <https://doi.org/10.1093/bioinformatics/btu743>.
 61. Okada, S., Doi, G., Nakagawa, S., Kusumoto, E., and Ito, T. (2021). Simple-to-use CRISPR-SpCas9/SaCas9/AsCas12a vector series for genome editing in *Saccharomyces cerevisiae*. *G3 (Bethesda)* 11, jkab304. <https://doi.org/10.1093/g3journal/jkab304>.
 62. Blount, B.A., Driessen, M.R.M., and Ellis, T. (2016). GC Preps: Fast and Easy Extraction of Stable Yeast Genomic DNA. *Sci. Rep.* 6, 26863. <https://doi.org/10.1038/srep26863>.
 63. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. <https://doi.org/10.1038/nbt.1754>.
 64. Rossmann, M.P., and Stillman, B. (2013). Immunoblotting histones from yeast whole-cell protein extracts. *Cold Spring Harb. Protoc.* 2013, 625–630. <https://doi.org/10.1101/pdb.prot067116>.
 65. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. <https://doi.org/10.1038/nmeth.2019>.
 66. Lin, Y.C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., et al. (2014). Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.* 5, 4767. <https://doi.org/10.1038/ncomms5767>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Histone H3 Antibody, CT, pan, clone A3S, rabbit monoclonal antibody	Sigma-Aldrich	Cat# 05-928 RRID:AB_492621
Histone H3K56ac rabbit polyclonal antibody	Active Motif	Cat# 39281 RRID:AB_2661786
Anti-rabbit IgG, HRP-linked Antibody	Cell Signaling Technology	Cat# 7074 RRID:AB_2099233
Bacterial and virus strains		
DH5 α high Champion™ cell	SMOBIO	Cat# CC5202
Chemicals, peptides, and recombinant proteins		
17 β -Estradiol	Nakarai tesque	Cat# 14541-74
Nicotinamide	Nakarai tesque	Cat# 24317-72
Critical commercial assays		
KOD One® PCR Master Mix (Dye-free 2 $\times$ PCR Master Mix)	TOYOBO	Cat# KMM-101
KOD SYBR® qPCR Mix	TOYOBO	Cat# QKD-201
Chelex 100 Chelating Resin, biotechnology grade, 100–200 mesh, sodium form	Bio-Rad	Cat# 1432832
Quick-DNA Fungal/Bacterial Miniprep Kit	ZYMO RESEARCH	Cat# D6005
NucleoSpin Tissue	TaKaRa	Cat# 740952.50
QIAfilter Plasmid Midi Kit	QIAGEN	Cat# 12243
Monarch HMW DNA Extraction Kit for Cell & Blood	NEB	Cat# T3050L
Monarch HMW DNA Extraction Kit for Tissue	NEB	Cat# T3060L
NEB Golden Gate Assembly Kit (BsaI-HF v2)	NEB	Cat# E1601L
NEBuilder HiFi DNA Assembly Master Mix	NEB	Cat# E2621L
ThruPLEX DNA-Seq kit	TaKaRa	Cat# R400674
DNA Single Index kit – 12S Set A	TaKaRa	Cat# R400695
DNA Single Index kit – 12S Set B	TaKaRa	Cat# R400697
MiSeq Reagent Kit v3 (150-cycle)	Illumina	Cat# MS-102-3001
Certified Megabase Agarose	Bio-Rad	Cat# 1613108
CHEF Genomic DNA Plug Kits	Bio-Rad	Cat# 1703491
12% Mini-PROTEAN® TGX™ Precast Protein Gels, 12-wells	Bio-Rad	Cat# 4561045
SYBR Green I Nucleic Acid Gel Stain	Thermo-Fisher	Cat# S7563
Trans-Blot Turbo Mini 0.2 μ m PVDF Transfer Packs	Bio-Rad	Cat# 1704156
iBind™ Solution Kit	Thermo Fisher	Cat# SLF1020
iBind™ Cards	Thermo Fisher	Cat# SLF1010
Clarity Max Western ECL Substrate	Bio-Rad	Cat# 1705062
AlkPhos Direct Labeling Module for 25 labellings	Cytiva	Cat# RPN3680
CDP-Star Detection Reagent for 2,500 cm ² membrane	Cytiva	Cat# RPN3682
AlkPhos Direct Hybridization Buffer for 5,000 cm ² membrane	Cytiva	Cat# RPN3688
Lipofectamine 3000® Reagent	Thermo Fisher	Cat# L3000008
Quick-DNA Microprep Kit	ZYMO RESEARCH	Cat# D3020
Ligation Sequencing Kit	Oxford Nanopore Technologies	SQK-LSK109
Ligation Sequencing Kit	Oxford Nanopore Technologies	SQK-LSK114

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Native Barcoding Expansion 1-12	Oxford Nanopore Technologies	EXP-NBD104
Native Barcoding Expansion 13-24	Oxford Nanopore Technologies	EXP-NBD114
Native Barcoding Kit 96 V14	Oxford Nanopore Technologies	SQK-NBD114.96
MinION Flow Cell (R9.4.1)	Oxford Nanopore Technologies	FLO-MIN106D
PromethION Flow Cell (R10.4.1)	Oxford Nanopore Technologies	FLO-PRO114M
Flongle Flow Cell (R10.4.1)	Oxford Nanopore Technologies	FLO-FLG114
Deposited data		
<i>S. cerevisiae</i> S288C reference genome: sacCer3	Saccharomyces Genome Database	https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000146045.2/
Genome assembly T2T-CHM13v2.0	T2T Consortium	https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_009914755.1/
Raw sequence data	This paper	DDBJ BioProject database: PRJDB18647, PRJDB18687, PRJDB18705
Experimental models: Cell lines		
HEK293T	RIKEN BRC	RBRC-RCB2202
Experimental models: Organisms/strains		
<i>S. cerevisiae</i> : Strain background: BY4741	N/A	N/A
All other synthetic yeast strains used in this paper, listed in Table S2	This paper	N/A
Oligonucleotides		
All oligonucleotides used in this paper, listed in Table S3	This paper	N/A
Recombinant DNA		
pyChrlV545kb_FP-SV40ori-EGF-yChrlV592kb_RFP_puroR	Sugiyama et al. ³⁵	N/A
VKG1-gRNA-pX330	Sawatsubashi et al. ³²	Addgene plasmid #108671
AAVS1 T2 CRISPR in pX330	Natsume et al. ⁴⁷	Addgene plasmid #72833
AIO-Puro	Chiang et al. ⁴⁸	Addgene plasmid #74630
All other plasmids used in this paper, listed in Table S4	This paper	N/A
Software and algorithms		
MinKNOW	Oxford Nanopore Technologies	https://community.nanoporetech.com/downloads?from=support
Guppy v6.5.7	Oxford Nanopore Technologies	https://community.nanoporetech.com/downloads?from=support
Dorado v0.7.3	Oxford Nanopore Technologies	https://community.nanoporetech.com/downloads?from=support
NanoPlot	De Coster et al. ⁴⁹	https://github.com/wdecoster/NanoPlot
YASS	Noé and Kuchero ⁵⁰	https://bioinfo.univ-lille.fr/yass/index.php
Minimap2 v2.17-r941	Li ⁵¹	https://github.com/lh3/minimap2
samtools v1.10	Danecek et al. ⁵²	https://github.com/samtools/samtools
bedtools v2.27.1	Quinlan and Hall ⁵³	https://github.com/arq5x/bedtools2
Read_split_by_target.py	Satoshi Okada	https://doi.org/10.5281/zenodo.11515698
Target_seq_extraction_single.py	Satoshi Okada	https://doi.org/10.5281/zenodo.11515698
Bedgraph_norm_ratio.py	Satoshi Okada	https://doi.org/10.5281/zenodo.11515696
minialign	Hajime Suzuki	https://github.com/ocxtal/minialign
BLAST	Altschul et al. ⁵⁴	https://blast.ncbi.nlm.nih.gov/blast/Blast.cgi
Bowtie2 v2.3.5	Langmead and Salzberg ⁵⁵	https://github.com/BenLangmead/bowtie2

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The budding yeast *Saccharomyces cerevisiae* was used as the primary experimental model in the study. The haploid yeast strain BY4741⁵⁶ was used as the parental strain. As a model of mammalian cells, the human female embryonic kidney-derived cell line

HEK293T was purchased from RIKEN BRC (catalog number RBRC-RCB2202) and cultured in Dulbecco's modified Eagle medium (DMEM) (Gibco, catalog number 11885084) at 37°C.

METHOD DETAILS

Yeast strains

All yeast strains used in this study are derived from BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*)⁵⁶ (Table S2). This study used standard culture media and genetic methods.⁵⁷ We deleted a gene of interest by transforming yeast cells with a DNA fragment composed of a *KanMX* cassette sandwiched by the 5'- and 3'-flanking sequences of the open reading frame of the gene, which was amplified from the corresponding deletant strain in Yeast Deletion Clones *MATa* Complete Set (Invitrogen) using PCR primers listed in Table S3.

Yeast plasmids

All plasmids used in this study are listed in Table S4. All primers for plasmid construction were purchased from Sigma-Aldrich and Eurofins Genomics. Plasmids were constructed by seamless cloning using HiFi DNA Assembly (NEB) or Golden Gate Assembly (NEB).

The integrative plasmid YIplac128-pGAL1-nCas9 (Cas9^{D10A} or Cas9^{H840A/N854A})-tADH1 (*LEU2*) harbors a gene encoding nCas9 derived from *Streptococcus pyogenes* fused with the SV40 nuclear localization signal as described previously⁵⁸ under the control of the *GAL1* promoter. It was used for yeast transformation after *AgeI* digestion to be integrated into the *GAL1* promoter on the genome.

The integrative plasmid pFA6a-pCUP2-yGEV-tADH1-HphMX (HygR) harbors a gene encoding β-estradiol-responsive artificial transcription activator GEV¹⁴ under the control of the *CUP2* promoter. It was used for yeast transformation after *MfeI* digestion to be integrated into the *CUP2* promoter on the genome.

Centromeric plasmids for gRNA expression harbor a gRNA gene under the control of the *GAL1* promoter. The gRNA scaffold sequence contains a base-flip and a stem-loop extension for stable gRNA expression.⁵⁹ To cut off an unnecessary sequence from the 5'-terminal portion of the gRNA-containing transcript, each gRNA sequence is preceded by a hammerhead ribozyme (Table S5). To define the 3'-terminus, each gRNA sequence is followed by the HDV ribozyme on the *GAL1* promoter plasmid (Table S5). For designing gRNAs, CRISPRdirect⁶⁰ was used to select target sites in the yeast genome listed in Table S5.

Yeast genome editing

For constructing the *gRNA1inv*, *pif1-m2*, *rtt109-K290Q*, *cup1ruΔ:ymNG* array, *gRNA1ts-cup1ruΔ::NatMX*, *hoΔ:2×CUP1RU*, *hoΔ:2×ymNGRU*, *X-2Δ:2×CUP1RU*, *X-2Δ:2×ymNGRU* strains, we performed SpCas9 or enAsCas12a-based gene editing as described previously.⁶¹ All gene-editing plasmids used in this study are listed in Table S4.

Yeast cell culture

Yeast cells were grown at 30°C overnight in 2 mL of SC—Ura, SC—His—Ura, SC—Leu—Ura, or SC—His—Leu—Ura medium supplemented with 2% glucose with or without G418 disulfate and/or hygromycin B (Nakalai tesque). On the following day, the OD₆₂₀ of each sample was recorded, and 10–50 μL of the culture diluted up to 1 × 10⁶ times was inoculated into 2 or 5 mL of the fresh medium containing 10 nM β-estradiol, supplemented with or without 5 mM NAM. Genomic DNA was extracted from the remaining culture using the GC prep method for qPCR.⁶² The same process was repeated every 1 to 3 days. The division number per day was calculated from the change of OD₆₂₀.

Quantitative PCR (qPCR)

Genomic DNA was diluted ten times with distilled water before qPCR. Each qPCR solution (20 μL) contained 2 μL of diluted DNA, 10 μL of KOD SYBR qPCR Mix (TOYOBO), 0.04 μL of 50× ROX Reference Dye (TOYOBO), 2 pmol each of the forward and reverse primers. The primers used for qPCR are listed in Table S2. Each qPCR assay was performed in duplicate, using QuantStudio3 (Applied Biosystems) according to the manufacturer's instructions. The amplification condition was initial denaturation at 98°C for 2 min followed by 40 times iteration of a 3-step thermal cycle composed of 98°C for 10 s, 55°C for 10 s, and 68°C for 30 s. All qPCR runs included 10-fold serial dilutions to generate standard curves. The quantity of *CUP1*, *ENA1*, *ymNG*, and *HIS3* was normalized to that of *ACT1*. The copy number of *CUP1*, *ENA1*, *ymNG*, and *HIS3* in the standard curves was calibrated by nanopore sequencing results in the BY4741 strain. The CNA/G for each gene was calculated with the below formula: CNA/G = (Copy number_{Day T} – Copy number_{Day 0})/Division number.

Modeling the contraction of extended *CUP1* array

To interpret the plot between the initial copy number and CNA/G (Figure 3H), we deduced a theoretical plot assuming that contraction occurs via homologous recombination between two *CUP1* repeat units, following second-order kinetics.

Let X be the copy number of repeat units, and k be the rate constant of homologous recombination between the repeat units. The rate equation is given by:

$$\frac{dX}{dt} = -kX^2$$

Let X_0 be the initial repeat unit copy number. Then, solving this differential equation yields:

$$X = \frac{X_0}{1+kX_0t}$$

The copy number as a function of time decreases along a rectangular hyperbola. Since CNA/G is defined as the difference in copy number at $t = T$ and $t = 0$ divided by the generation number G , it can be expressed as:

$$CNA/G = \frac{1}{G} \left(\frac{X_0}{1+kX_0T} - \frac{X_0}{1} \right)$$

Simplifying this:

$$CNA/G = \frac{1}{G} \left(\frac{-kX_0^2T}{1+kX_0T} \right) = -\frac{1}{G} \left(\frac{kTX_0^2}{1+kTX_0} \right)$$

Finally:

$$CNA/G = -\frac{1}{G} \left(\frac{X_0^2}{X_0 + \frac{1}{kT}} \right)$$

If k and T are constants and X_0 is the variable, then CNA/G with respect to X_0 decreases along the sum of a linear function and a rectangular hyperbola. As the initial copy number X_0 increases, the contraction rate CNA/G asymptotically approaches a straight line, as observed in Figure 3H.

Nanopore sequencing

Genomic DNA was extracted using Monarch HMW DNA Extraction Kit for Tissue (NEB). We avoided vortexing to obtain high molecular weight DNA and used mixing by gentle pipetting with a wide-bore tip instead. DNA libraries for nanopore whole-genome sequencing were prepared using the ligation sequencing kit SQK-LSK109, SQK-LSK114 and the native barcoding kit EXP-NBD104, EXP-NBD114, or SQK-NBD114 (Oxford Nanopore Technologies) according to the manufacturer's instructions. We modified the protocol of the ligation sequencing kit as follows: DNA fragmentation, omitted; duration of the enzymatic repair steps at 20°C and 65°C, both extended from 5 min to 30 min; and the duration of the ligation step, extended from 10 to 30 min; incubation time for elution with 0.4× AMPure XP, extended from 10 min to 20 min. The library was sequenced with the flowcell FLO-MIN106D R9.4.1 using the MinION sequencer and FLO-PRO114M R10.4.1 using the PromethION 2 Solo sequencer (Oxford Nanopore Technologies). MinKNOW software was used to control the MinION and PromethION devices. The run time was set to 72 h. Base calling was performed using Guppy v6.5.7 and Dorado v0.7.3. The assessment of sequencing data was performed using NanoPlot.⁴⁹

Dot plot analysis of nanopore reads

We used nanopore sequencing data in FASTA format to draw dot plots using YASS.⁴⁹ We first selected reads spanning the entire array using 1-kb upstream and downstream sequences of the target array as queries of minialign (<https://github.com/ocxtal/minialign>) and then used these reads as the first input sequence for YASS. As the second input, we used the reference sequence of the repeat unit. By manually counting the diagonal lines in each dot plot, we determined the copy number of the repeat unit.

Translocation analysis using nanopore reads

We used nanopore sequencing data in FASTQ format, selected reads containing the *CUP1RU*, divided each read into 5'- and 3'-flanking regions of the *CUP1RU*, and extracted the flanking regions using Target_seq_extraction_single.py and Read_split_by_target.py (<https://doi.org/10.5281/zenodo.11515698>). We mapped reads to the S288c reference genome using Minimap2,⁵¹ SAMtools,⁵² and BEDtools.⁵³ Data were visualized with the Integrative Genomics Viewer (IGV).⁶³

Copy number estimation from nanopore reads

We used nanopore sequencing data in FASTQ format and mapped reads to the S288c reference genome (version R64-2-1, http://sgd-archive.yeastgenome.org/sequence/S288C_reference/genome_releases/S288C_reference_genome_R64-2-1_20150113.tgz) using SAMtools⁵² and BEDtools,⁵³ and then normalized read count of each nucleotide was calculated using Bedgraph_norm_ratio.py (<https://doi.org/10.5281/zenodo.11515696>). Data were visualized with the IGV.⁶³ To eliminate the effect of read clipping and achieve a more accurate estimation of repeat unit number, we collected all reads containing the repeat unit using minialign.

We then used its reference sequence as a query in a BLAST⁵⁴ search against the collected reads and estimated the copy number based on the number of BLAST hits.

Illumina sequencing

Genomic DNA was extracted using *Quick-DNA Fungal/Bacterial Miniprep Kit* (Zymo Research) and then fragmented to 300 bp using the S220 Focused-ultrasonicator (Covaris). DNA libraries with indexing were prepared using the ThruPLEX DNA-Seq kit (TaKaRa). We used the DNA Single Index Kit – 12S Set A or B (TaKaRa) for indexing according to the manufacturer's instructions. The library was sequenced with MiSeq Reagent Kit v3 using the MiSeq instrument (Illumina). We mapped 2 × 75 bp reads to the S288c reference genome using Bowtie2,⁵⁵ SAMtools,⁵² and BEDtools,⁵³ and then normalized read count of each nucleotide was calculated using `Bedgraph_norm_ratio.py` (<https://doi.org/10.5281/zenodo.11515696>). Data were visualized with the IGV.⁶³

Pulsed-field gel electrophoresis and Southern blot hybridization

Agarose-embedded yeast DNA was prepared using CHEF Genomic DNA Plug Kits (BioRad) according to the manufacturer's instructions. DNA digested with or without EcoRI was subjected to 1% and 0.8% Certified Megabase Agarose (Bio-Rad) in 0.5× TBE and 1× TAE, respectively. Pulsed-field gel electrophoresis (PFGE) was performed using CHEF mapper XA (BioRad) according to the manufacturer's instructions. PFGE running conditions were a 60–120 s pulse time, 120° angle, and 6 V/cm for 24 h at 14°C in a 1% agarose gel and 500 s pulse time, 106° angle, and 3 V/cm for 48 h at 14°C in a 0.8% agarose gel. The gel was then stained with SYBR Green I Nucleic Acid Gel Stain (Invitrogen) at room temperature for 30 min with shaking, destained in distilled water for 1 h, and the fluorescence signals were detected with ChemiDocTouch system (Bio-Rad). Transfer to the membrane was performed using Hybond-N+ (Cytiva) according to the manufacturer's instructions. The blot was hybridized with a *CUP1* probe or outside probes at 55°C overnight after UV-crosslinking. The probe was generated by PCR using the primers listed in Table S2, followed by labeling with alkaline phosphatase using the labeling module of the AlkPhos Direct Labeling and Detection System kit (Cytiva). Following appropriate blot washing, chemiluminescent signals were generated using the CDP-Star Detection Reagent in the kit and detected with the ChemiDocTouch system (Bio-Rad). Images were processed with ImageJ software (National Institutes of Health).

Immunoblot analysis

The amount of histone H3 and acetylated histone H3 lysine 56 (H3K56ac) was analyzed by western blotting. Proteins were extracted as described previously,⁶⁴ and separated with sodium dodecyl sulfate-polyacrylamide gel electrophoresis using 12% Mini-PROTEAN TGX Precast Protein Gel (Bio-Rad). Transfer to the membrane was performed with iBind Western System (Thermo Fisher Scientific) according to the manufacturer's protocol. Primary antibodies to detect histone H3 and H3K56ac were Anti-Histone H3 Antibody, CT, pan, clone A3S, rabbit monoclonal antibody (1:500, Sigma-Aldrich), and Histone H3K56ac rabbit polyclonal antibody (1:2500, Active Motif), respectively. The secondary antibody was Anti-rabbit IgG, HRP-linked Antibody (1:2000, Cell Signaling Technology). Following incubation with Clarity Western ECL Substrate (Bio-Rad), chemiluminescent signals were detected with the ChemiDocTouch system (Bio-Rad). Gel images were processed with ImageJ software.

Fluorescence microscopy and image processing for yeast cells

Image acquisitions of yeast cells were performed on a microscope (Ti-E, Nikon, Tokyo, Japan) with a 20× objective lens (CFI Plan Apo Lambda 20X, MRD00205, Nikon), a sCMOS camera (ORCA-Fusion BT, C15440-20UP, Hamamatsu photonics, Hamamatsu, Japan), and a solid-state illumination light source (SOLA SE II, Lumencor, Beaverton, OR, USA). Image acquisition was controlled by NIS-Elements version 5.3 (Nikon). Z-stacks were 7 × 0.9 μm. For imaging of mNeonGreen, a filter set (LED-YFP-A, Semrock, Rochester, NY, USA) was used with excitation light power set at 7% and the exposure time set at 200 msec/frame. For imaging of mCherry, a filter set (LED-TRITC-A, Semrock) was used with excitation light power set at 30% and exposure time set at 300 msec/frame. For imaging of mRFP682, a filter set (LED-Cy5.5-A, Semrock) was used with excitation light power set at 10% and exposure time set at 200 msec/frame. For imaging of TagBFP, a filter set (LED-DAPI-A, Semrock) was used with excitation light power set at 50% and exposure time set at 700 msec/frame. For DIC (differential interference contrast) image acquisition, the exposure time was set at 50 ms/frame. DIC images were captured only at the middle position of the Z-stacks.

Image processing and analysis were performed using Fiji.⁶⁵ To generate 2-dimensional images of fluorescence channels from Z-stacks, background subtraction (sliding paraboloid radius set at 5 pixels with disabled smoothing) and maximum projection using 7 Z-slices were performed. Maximum projected fluorescence images and corresponding smoothed DIC images were superimposed. After global adjusting of brightness and contrast and cropping of the images, sequences of representative images were generated.

BITREx of EGFP reconstitution reporter in HEK293T cells

Human embryonic kidney 293T cell (HEK293T) cells were cultured DMEM (Gibco) supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin–streptomycin (Gibco). Cells were maintained at 37°C in a humidified atmosphere containing 5% CO₂.

The donor vector mCherry-PuroR-FP-SV40 ori-EGF-gRNA1ts was derived from the plasmid pyChrlV545kb_FP-SV40ori-EGF-yChrlV592kb_RFP_puroR.³⁵ We also constructed the nCas9-gRNA1 expression plasmid by inserting the gRNA1-coding sequence into the AIO-Puro plasmid.⁴⁸ Using the Lipofectamine 3000 reagent (Invitrogen), we co-transfected the donor plasmid, the donor

cleavage vector (VKG1-gRNA-pX330),³² and the locus-specific cleavage vector (AAVS1 T2 CRISPR in pX330)⁴⁷ into HEK293T cells to facilitate knock-in of the donor plasmid at the target locus via the VIKING method.³²

Puromycin (0.3 µg/mL) was added to the culture at 24 h after transfection. Following 48 h of cultivation, the cells were transfected with the nCas9-gRNA1 expression plasmid. Puromycin (2.0 µg/mL) was then added to the culture at 24 h after the second transfection, and the cells were grown for an additional 72 h.

We designed a primer pair spanning the cleavage site of the donor plasmid to specifically detect residual donor plasmids while excluding signals from those integrated into the genome. The copy number of residual donor plasmid was estimated from the qPCR results, considering the hypotriploid nature of HEK293 cells.⁶⁶

Flow cytometric analysis of transfected HEK293T cells

Transfected HEK293T cells with the *EGFP* reconstitution reporter were washed twice with phosphate-buffered saline (PBS) and detached from the dishes using 0.25% trypsin-EDTA (Gibco). The cells were then resuspended in PBS containing 0.2% bovine serum albumin (Thermo Fisher Scientific) and filtered through a 50 µm nylon mesh to obtain a single-cell suspension. Cell density was adjusted to 5×10^6 cells/mL. Cells were analyzed using a BD FACSARIA Fusion cell sorter (BD Biosciences). Data acquisition was performed using BD FACSDiva software (BD Biosciences). At least 10,000 events were collected per sample. Data were analyzed using FlowJo software (BD Biosciences), and gates were set based on isotype controls.

Flow sorting of EGFP-positive HEK293T cells

EGFP and mCherry-positive cells were sorted using a BD FACSARIA Fusion cell sorter (BD Biosciences). The excitation wavelength for EGFP was set to 488 nm, and EGFP fluorescence was detected using a 530/30 nm bandpass filter. The excitation wavelength for mCherry was set to 561 nm, and mCherry fluorescence was detected using a 610/20 nm bandpass filter. Untransfected HEK293T cells were used as negative controls to set the gates for GFP and mCherry-positive cells. At least 10,000 events were collected per sample.

Fluorescence microscopy and image processing for HEK293T cells

Image acquisitions of HEK293T cells were performed on an imaging system (EVOS M7000, Thermo Fisher Scientific) with a 20× objective lens (NA 0.70, Olympus, AMEP4765EO). For imaging of EGFP, a filter set (EVOS Light Cube GFP 2.0, Thermo Fisher Scientific, AMEP4951) was used with the exposure time set at 2 msec/frame. For imaging of mCherry, a filter set (EVOS Light Cube Texas Red 2.0, Thermo Fisher Scientific, AMEP4955) was used with the exposure time set at 50 msec/frame. For bright field image acquisition, the exposure time was set at 10 ms/frame. Image processing and analysis were performed using Fiji.⁶⁵ Background subtraction (sliding paraboloid radius set at 5 pixels with disabled smoothing) were performed. Fluorescence images and corresponding bright field images were superimposed. After global adjusting of brightness and contrast and cropping of the images, sequences of representative images were generated.

BITREx of the D4Z4 array in HEK293T cells

We constructed nCas9-gRNA expression plasmids by inserting the coding sequences for gRNAh1, gRNAh2, gRNAh3, and gRNAh4 into the AIO-Puro plasmid.⁴⁸ Each plasmid was transfected into HEK293T cells following the procedure described above.

Nanopore sequencing of genomic DNA prepared from HEK293T cells

Genomic DNA was extracted from HEK293T cells using NucleoSpin Tissue (Macherey-Nagel) and Monarch HMW DNA Extraction Kit for Cells & Blood (Monarch) according to the manufacturer's protocol. The extracted DNA was then used for ligation-based nanopore whole-genome sequencing library preparation. Reads containing the *EGFP* reconstitution reporter were identified using minialign (<https://github.com/ocxtal/minialign>) and subsequently analyzed to generate dot plots with YASS.⁵⁰ Minimap2⁵¹ was used to identify reads containing a unique sequence in the D4Z4 repeat unit (T2T-CHM13v2.0, chr.4:193,434,263–193,435,217) and reads containing a unique region flanking the D4Z4 array (T2T-CHM13v2.0, chr.4:193,427,855–193,437,559). These reads were then used to estimate the D4Z4 copy number, as described above.

QUANTIFICATION AND STATISTICAL ANALYSIS

One-way ANOVA test and Student's *t* test were employed to calculate *p* values, as indicated in the figure legends. In general, results were considered statistically significant when *p* < 0.05.

Cell Genomics, Volume 5

Supplemental information

**Strategic targeting of Cas9 nickase
expands tandem gene arrays**

Hiroaki Takesue, Satoshi Okada, Goro Doi, Yuki Sugiyama, Emiko Kusumoto, and Takashi Ito

Table S1: CNA/G values for six two-unit arrays at three loci, related to Figures 1, 4, S7, and S8

Array type	Interrupting sequence	Repeat unit length (bp)	CNA/G at each locus (top: Day 0 to 3; bottom: Day 3 to 10)		
			<i>CUP1</i> (chr VIII)	<i>HO</i> (chr IV)	X-2 (chr X)
2×CUP1RU	none	1,988	0.153 0.216	—	—
	<i>HIS3</i> 4 fluorescent protein genes plasmid backbone	10,339	0.102 0.004	0.038 0.003	0.051 0.008
	<i>HIS3</i> 4 fluorescent protein genes plasmid backbone <i>ARS305</i>	10,490	0.110 0.141	0.043 0.024	0.032 0.059
2×ymNGRU	none	1,642	0.246 0.237	—	—
	<i>HIS3</i> plasmid backbone	4,994	0.015 0.050	0.039 0.068	0.197 0.001
	<i>HIS3</i> plasmid backbone <i>ARS305</i>	5,072	0.069 0.087	0.038 0.053	0.120 0.144
14×CUP1RU	none	1,988	0.492 0.577	—	—

Note:

The six two-unit arrays were integrated at three genomic loci and subjected to BITREx using gRNA1. CNA/G values were calculated from qPCR data at days 0, 3, and 10 for each indicated two-unit array. The values from day 0 to 3 and day 3 to 10 are shown in the top and bottom lines of each box, respectively. A minus sign indicates that the corresponding strain was not available. As a control, CNA/G values for native *CUP1* array at *CUP1* locus are also included. Notably, arrays embedded at the same locus displayed varying efficiencies, underscoring the importance of repeat unit characteristics, such as sequence composition, length, the presence or absence of an ARS, and the inclusion of intervening sequences. The presence of an internal ARS contributes to maintaining the CNA/G value after day 3. However, during the early stages of BITREx, when array lengths are relatively short, this effect can be overridden by other sequence features. For instance, at the *CUP1* locus, the uninterrupted 2×ymNGRU array outperformed the uninterrupted 2×CUP1RU array from day 0 to day 3, although the former lacks an internal ARS while the latter contains one (*ARS810/811*). This difference diminished between days 3 and 10, likely because the initial advantage of 2×ymNGRU, attributed to its sequence composition, was progressively offset by the influence of the internal ARS in 2×CUP1RU. Moreover, identical arrays demonstrated varying behaviors depending on the locus in which they were embedded, highlighting the impact of the local environment, including the distance from nearby ARS, epigenetic status, and so on. For example, the 2×ymNGRU array interrupted by *HIS3* exhibited a marked CNA/G decline after day 3 at the X-2 locus but not at the *CUP1* or *HO* loci.

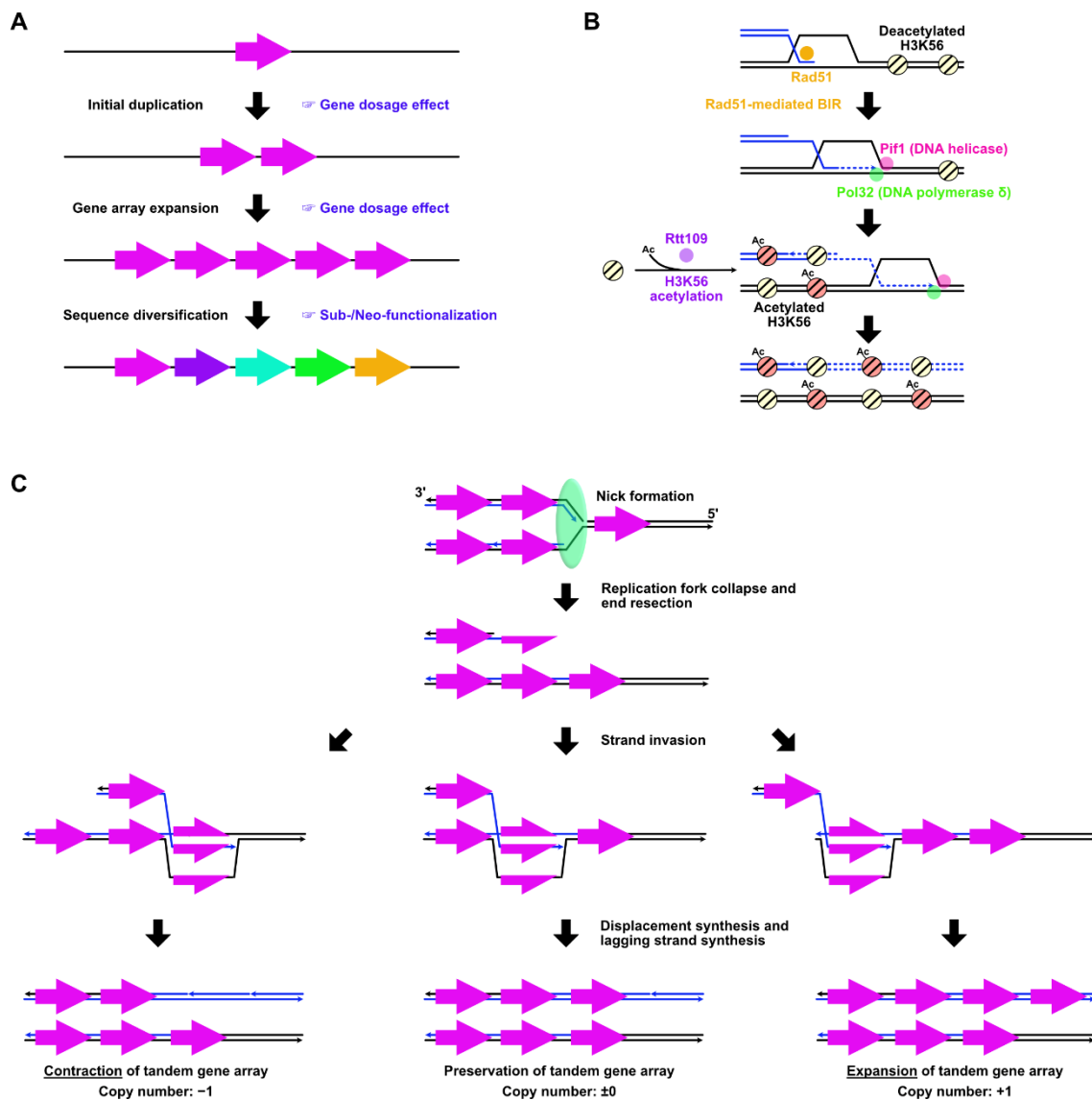


Figure S1: Conceptual foundations of BITREx, related to Figure 1

- (A) Roles of gene duplication and tandem gene array expansion in adaptation and evolution. Initial duplication and subsequent gene array expansion exert gene dosage effect. Sequence diversification leads to sub-/neo-functionalization. See text for detail.
- (B) Mechanism of BIR.^{10,11} The initial step of BIR is the invasion of ssDNA generated by end resection at seDSB into the donor sequence. Rad51 mediates this step in most BIR events. The invaded ssDNA primes the displacement synthesis catalyzed by DNA polymerase δ (Pol δ). Importantly, this step is dependent on Pol32, a Pol δ subunit that is not required for normal S-phase replication. Another essential gene for BIR is PIF1, which encodes the DNA helicase indispensable for BIR fork (D-loop) progression. In addition, BIR is suppressed by Rtt109-catalyzed acetylation at Lys-56 of histone H3 (H3K56ac). The yeast proteins Rad51, Pol32, and Pif1 have mammalian homologs—RAD51, POLD3, and PIF1, respectively. In contrast, Rtt109 is a fungus-specific histone acetylase with no mammalian homolog.
- (C) Outcomes of BIR initiated from an internal unit of a tandem gene array.

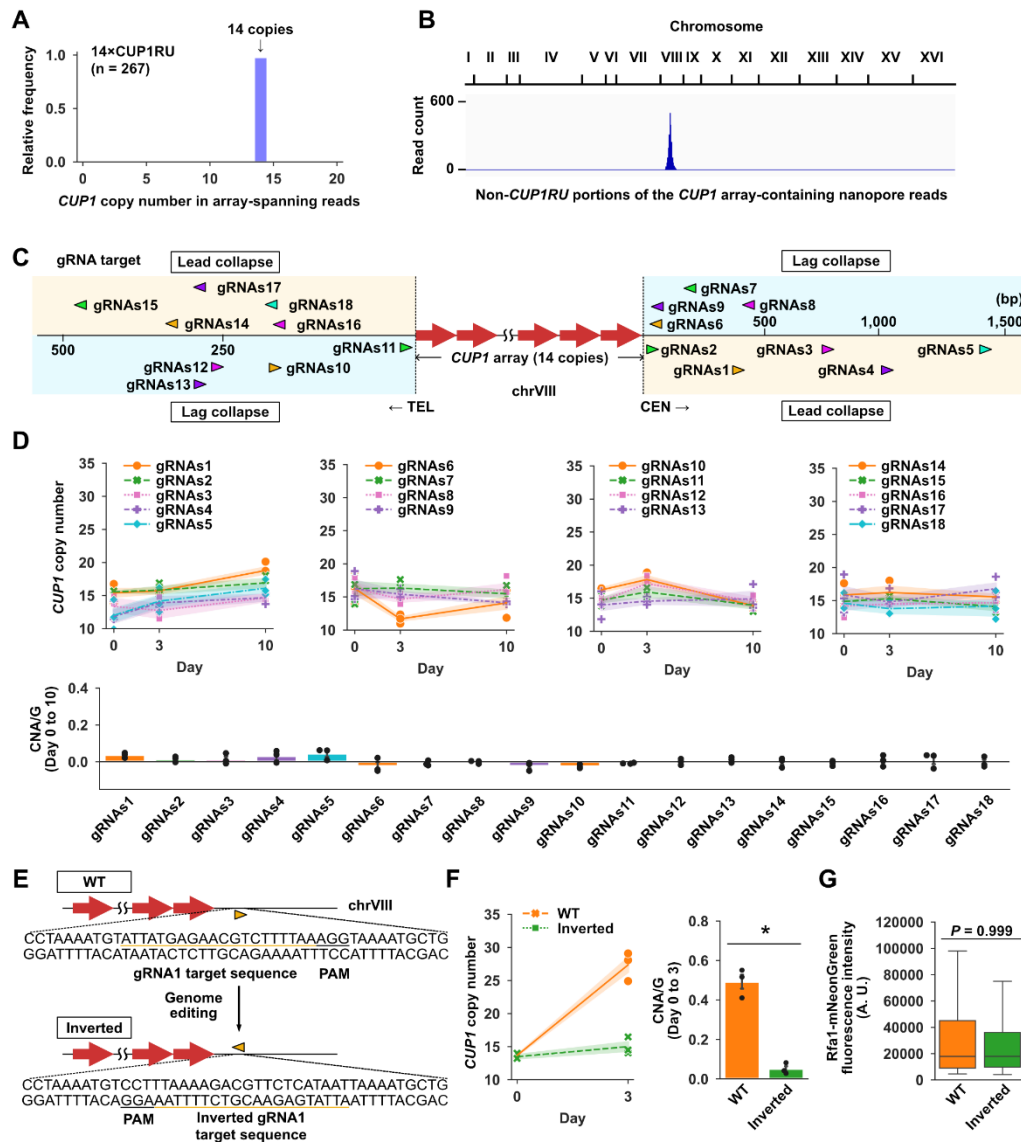


Figure S2: BITREx of *CUP1* array, related to Figure 1

- (A) *CUP1* array length of the parental strain used in this study. Nanopore reads containing both the 5'- and 3'-flanking regions of *CUP1* array were used to determine the distribution of *CUP1RU* copy number.
- (B) Genomic location of *CUP1* array-flanking sequences in the parental strain used in this study. Nanopore reads containing *CUP1RU* were selected and their non-*CUP1RU* portions were mapped to the reference genome sequence.
- (C) *CUP1* array and target sites of ineffective gRNAs (gRNAs1–gRNAs18). Similar to Figure 1C.
- (D) Performance of 18 gRNAs shown in (C). Similar to Figures 1D and 1E. Shading and error bar, SD (n = 3 biological replicates).
- (E) Inversion of the gRNA1 target site. In this strain, gRNA1 should make Cas9^{D10A} induce a nick on the lagging strand template.
- (F) BITREx in the wild-type (WT) and inverted strains. Left, CNA of *CUP1*; right, CNA/G. Shading and error bar, SD (n = 3 biological replicates).
- (G) Box plots showing Rfa1-mNeonGreen fluorescence in the WT and inverted strains with Cas9 and gRNA1. Since Rfa1 accumulates on ssDNA generated by end-resection at DSB sites as a component of the RPA complex, the fluorescence intensity serves as an indicator of gRNA1-guided Cas9 cleavage.

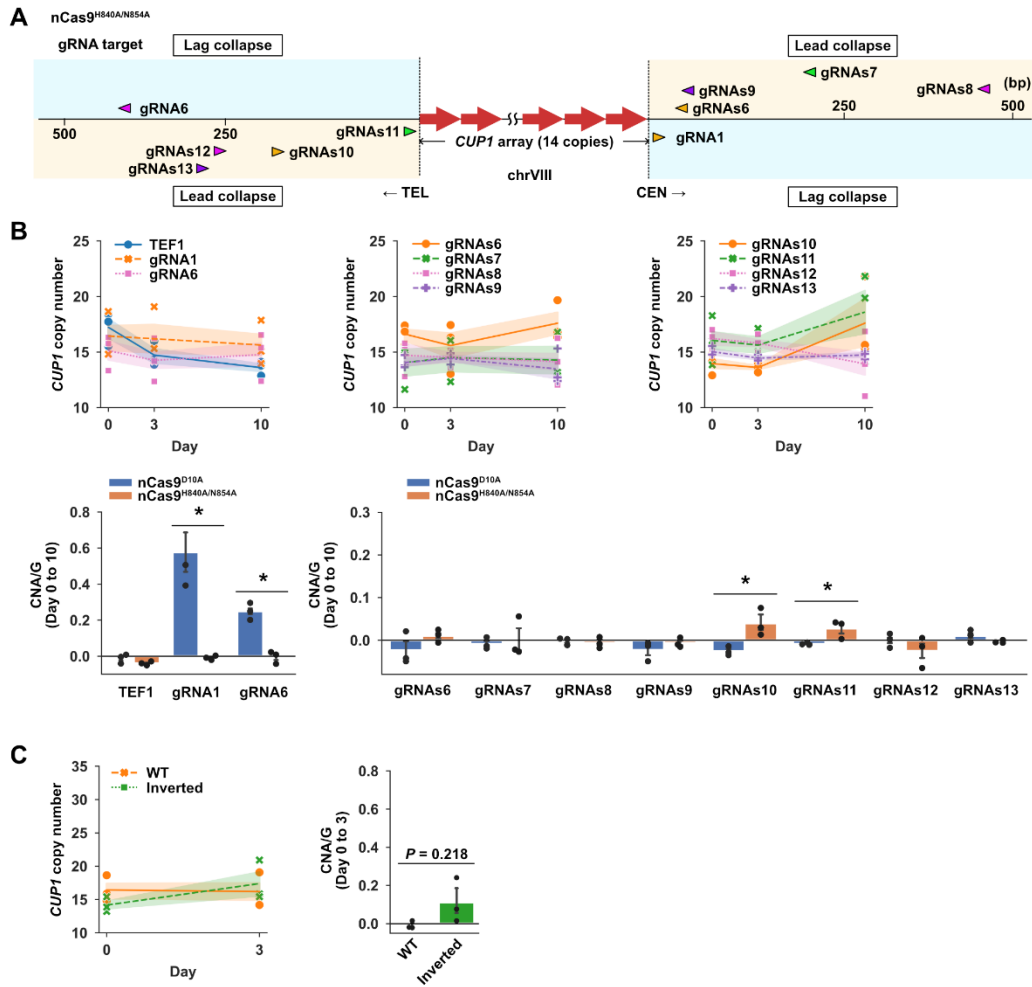
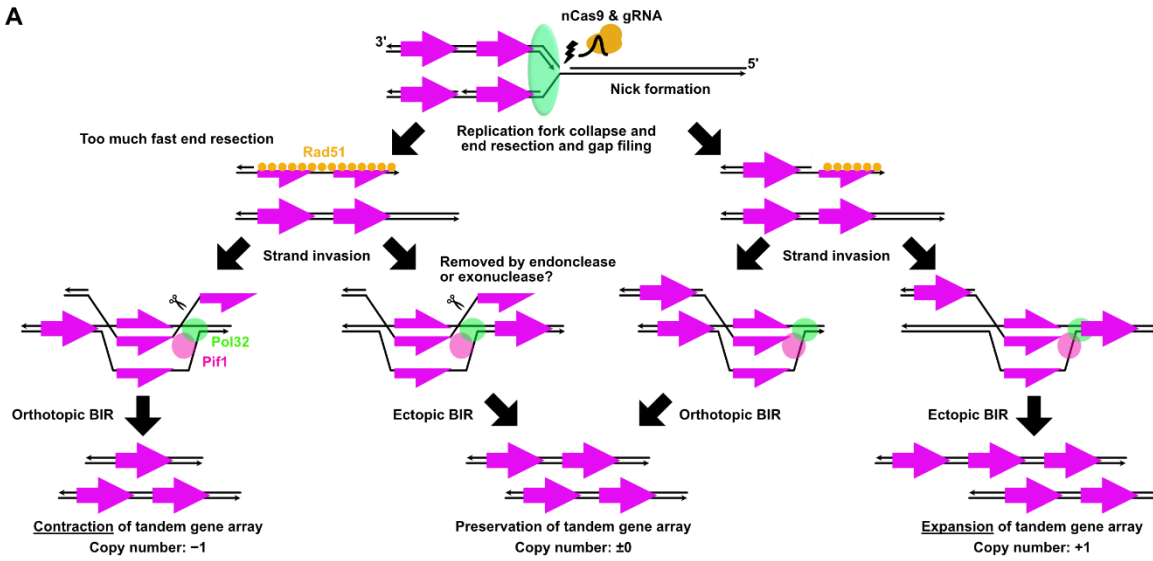


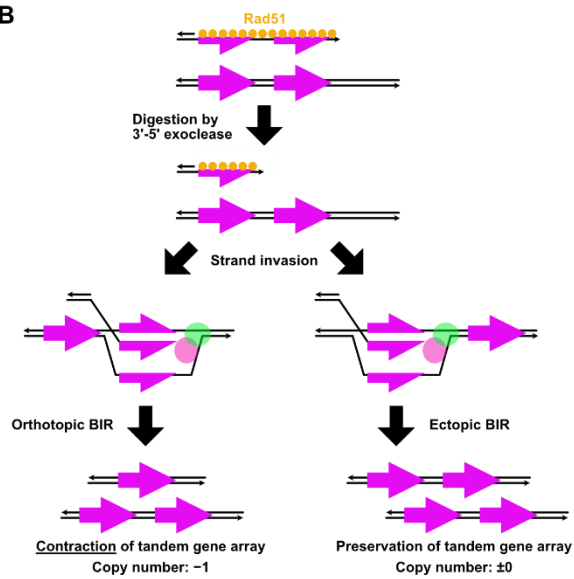
Figure S3: Requirement of lead collapse for BITREx, related to Figure 1

- (A) *CUP1* array and target sites of 10 gRNAs used with nCas9^{H840A/N854A}. Two gRNAs (gRNA1 and gRNA6) were effective when used with nCas9^{D10A} and should induce lag collapse when used with nCas9^{H840A/N854A}. The remaining eight gRNAs (gRNAs6–gRNAs13) were ineffective when used with nCas9^{D10A} (Figure S2D) and should induce lead collapse when used with nCas9^{H840A/N854A}. Note that Cas9^{H840A/N854A} has been reported to surpass Cas9^{H840A} in terms of correct nick formation frequency and reduced occurrence of unwanted indels.¹⁵
- (B) BITREx using nCas9^{H840A/N854A}. Similar to Figures 1D and 1E. Shading and error bar, SD (n = 3 biological replicates). *P < 0.05 (Student's t-test). Notably, when combined with Cas9^{H840A/N854A}, the two gRNAs that performed most efficiently with Cas9^{D10A} (gRNA1 and gRNA6) failed to increase *CUP1* copy number. Conversely, two gRNAs that were ineffective with Cas9^{D10A} (gRNAs10 and gRNAs11) induced a small but significant *CUP1* CNA.
- (C) BITREx using gRNA1 and nCas9^{H840A/N854A} in the inverted strain (Figure S2E). Shading and error bar, SD (n = 3 biological replicates). Note that Cas9^{H840A/N854A} with gRNA1 induced a weak increase in the copy number in the inverted strain.

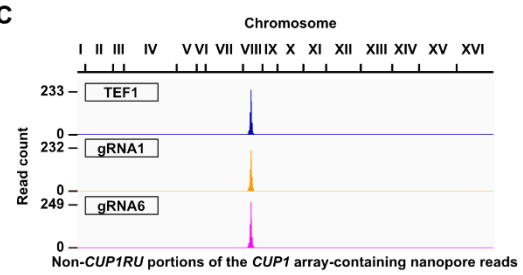
A



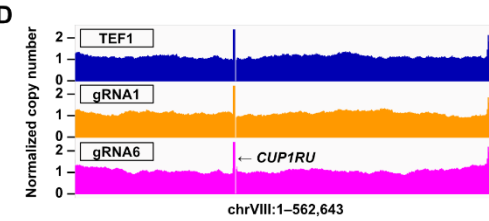
B



C



D



(legend on next page)

Figure S4: Array contraction, gross rearrangement, and target site mutation during BITREx, related to Figure 1

- (A) Array contraction by extensive 5'-to-3' end resection. If extensive 5'-to-3' end-resection converts not only the terminal but also the second terminal repeat unit to ssDNA, the latter ssDNA may hybridize with the terminal repeat unit on the sister chromatid. In this scenario, the terminal repeat unit of the invading ssDNA strand is left behind as a flap. If this flap is degraded by flap endonucleases such as Rad1-Rad10, the invading ssDNA (i.e., the second terminal repeat unit) can initiate BIR, leading to the contraction of the tandem array.
- (B) Array contraction by extensive 3'-to-5' end resection. A similar situation to (A) could occur if the 3'-to-5' exonuclease activity of Pol δ excessively degrades or "chews backs" the invading strand.⁴¹
- (C) Genomic location of *CUP1* array-flanking sequence in the strains subjected to BITREx with the indicated gRNAs for 3 days. Similar to Figure S2B. We collected all *CUP1RU*-containing reads and mapped their non-*CUP1RU* sequences to the reference genome. As expected, almost all of them were derived from the *CUP1* array-flanking regions on chromosome VIII. Each of the few abnormal junctions was unique and appeared only once, likely representing artifacts or chimeric molecules generated during ligation-based library preparation: true translocation junctions, in contrast, would be expected to appear in multiple independent reads, given the >200-fold genome coverage achieved.
- (D) Normalized read counts across chromosome VIII in nanopore sequencing. Read counts were normalized to the average counts of genomic regions excluding rRNA, *CUP1RU*, Ty elements, and mitochondrial DNA. Note that while the *sacCer3* reference genome sequence contains two copies of *CUP1RU*, the second copy is masked with 'N' prior to mapping. Consequently, the normalized read count directly reflects the *CUP1RU* copy number. The gap in read counts adjacent to *CUP1RU* is due to the masking. The copy numbers of *CUP1RU* were estimated to be 17.5, 30.5, and 20.4 in cells subjected to 3-day BITREx using *TEF1* gRNA, gRNA1, and gRNA6, respectively. The average read depth of chromosome VIII was comparable to the genome-wide average, except for the *CUP1* locus and the polymorphic subtelomeric regions.

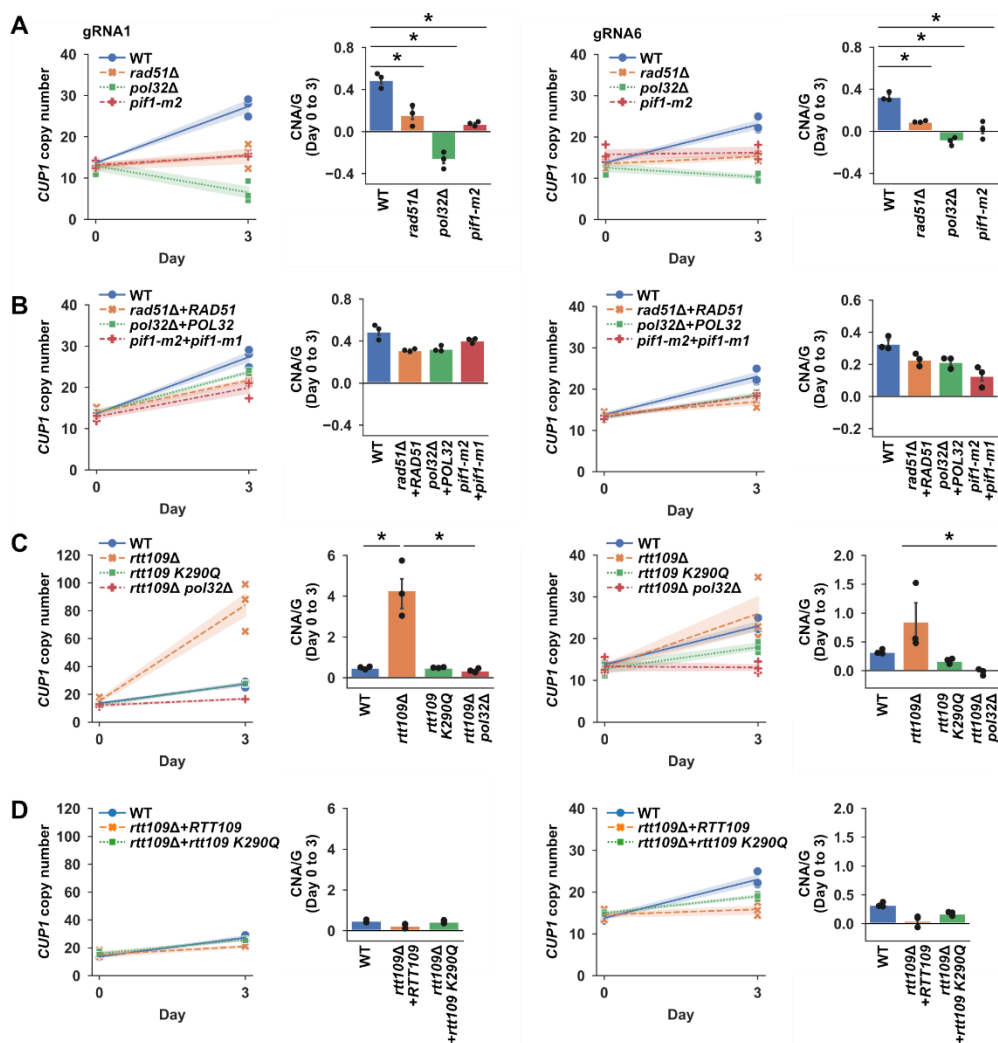
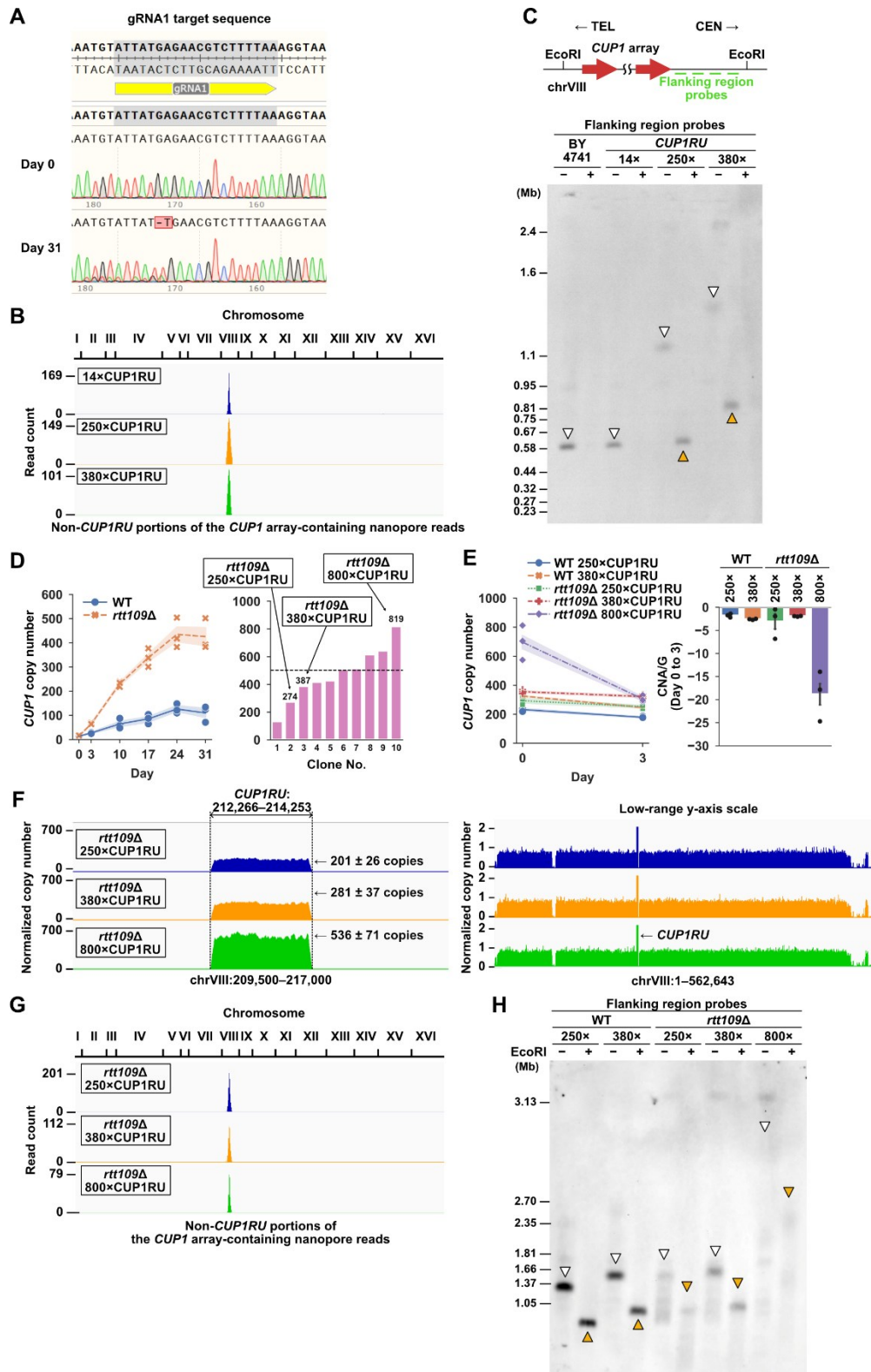


Figure S5: Genetic evidence for BIR to mediate BITREx, related to Figure 1

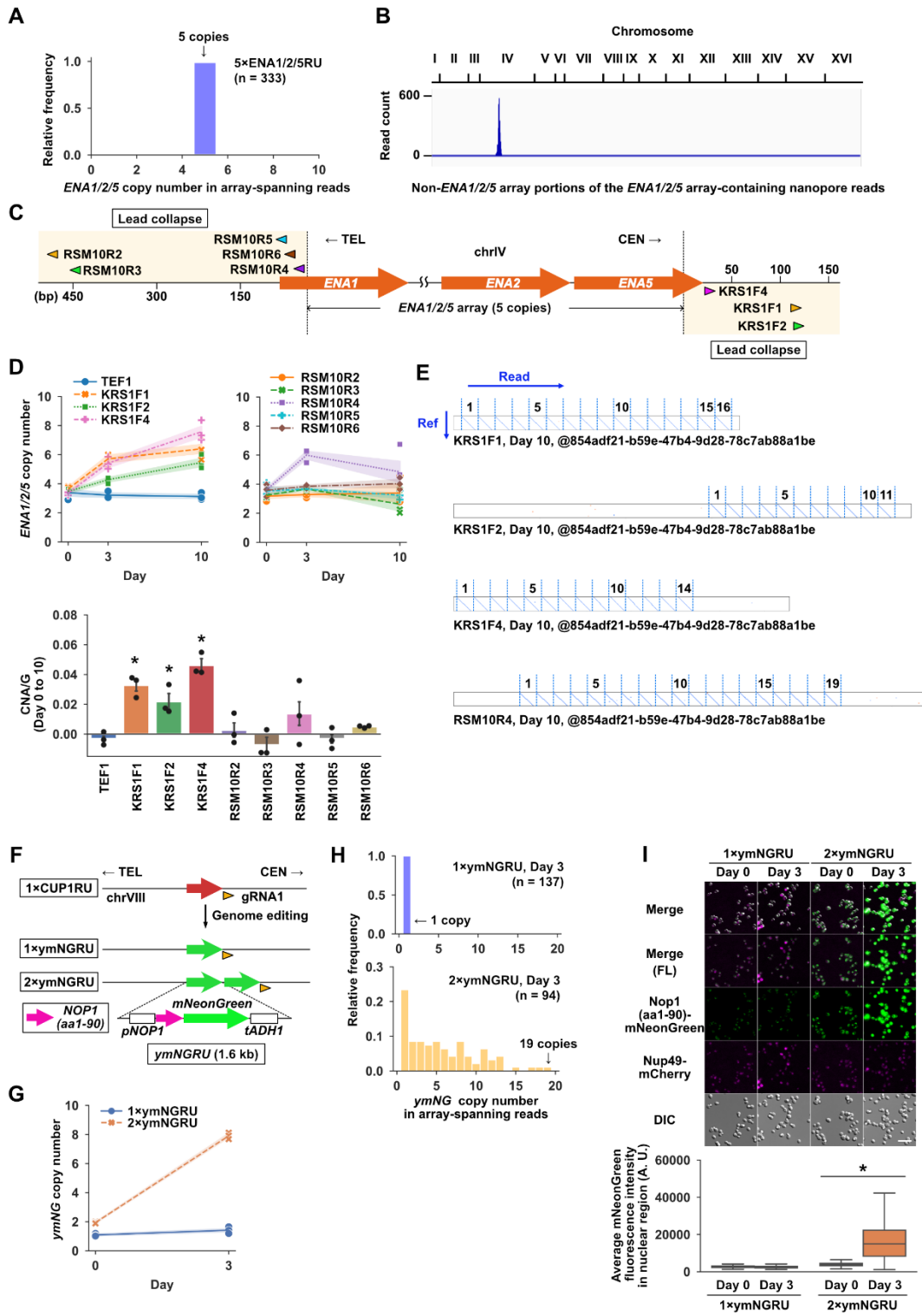
- (A) BITREx defects in *rad51Δ*, *pol32Δ*, and *pif1-m2* cells. Note that *PIF1* contains two initiation Met codons, with the first and second codons directing the synthesis of its mitochondrial and nuclear isoforms, respectively.¹⁷ The *pif1-m2* allele substitutes the second initiation Met codon with an Ala codon, thereby restricting synthesis to the mitochondrial isoform only.¹⁷ The *CUP1* copy number was measured by qPCR on days 0 and 3 to calculate CNA/G in each mutant. Two gRNAs (gRNA1 and gRNA6) were used. Shading and error bar, SD (n = 3 biological replicates). *P < 0.05 (Student's t-test).
- (B) Suppression of BITREx defects in *rad51Δ*, *pol32Δ*, and *pif1-m2* cells by episomal copies of *RAD51*, *POL32*, and *pif1-m1*, respectively. The *pif1-m1* allele substitutes the first initiation Met codon with an Ala codon, thereby restricting synthesis to the nuclear isoform only.¹⁷ Shading and error bar, SD (n = 3 biological replicates). *P < 0.05 (Student's t-test).
- (C) Enhancement of BITREx in *rtt109Δ* cells but not in *rtt109 K290Q* and *rtt109Δ pol32Δ* cells. Rtt109 is the sole enzyme responsible for H3K56ac, but it also contributes to H3K9ac.¹⁸ While no amino acid substitution was known to selectively abolish the H3K56 acetylase activity, K290Q substitution selectively abolishes the H3K9 acetylase activity. Shading and error bar, SD (n = 3 biological replicates). *P < 0.05 (Student's t-test).
- (D) Suppression of BITREx enhanced in *rtt109Δ* cells by episomal copies of *RTT109* and *rtt109 K290Q*. Shading and error bar, SD (n = 3 biological replicates). *P < 0.05 (Student's t-test).



(legend on next page)

Figure S6: Long-term BITREx, related to Figure 2

- (A) Mutation at the gRNA1 target site. Electropherograms are shown for Sanger sequencing of a PCR amplicon including the gRNA1 target site obtained from a cell population exhibiting a notable decline in copy number between days 24 and 31 (Figure 2A).
- (B) Genomic location of *CUP1* array-flanking sequences in the strains subjected to long-term BITREx. Similar to Figure S2B.
- (C) PFGE analysis of *CUP1* arrays expanded by long-term BITREx in wild-type cells. Similar to Figure 2H, but Southern blot hybridization was performed using the flanking region probes shown in the upper panel.
- (D) CNA of *CUP1* over the 31-day BITREx with gRNA1 in *rtt109Δ* cells. Shading, SD (n = 3 biological replicates). *CUP1* copy numbers of 10 randomly picked clones on day 31 are shown in the right panel. The dashed line indicates the average copy number, which is 502.
- (E) Stability of extended *CUP1* arrays. Each strain was cultivated for 3 days without BITREx induction. Shading and error bar, SD (n = 3 biological replicates).
- (F) Normalized read counts in Illumina sequencing of the *rtt109Δ* strains subjected to long-term BITREx. Similar to Figure 2F.
- (G) Genomic location of *CUP1* array-flanking sequences in the *rtt109Δ* strains subjected to long-term BITREx. Similar to Figure S2B.
- (H) PFGE analysis of *CUP1* arrays expanded by long-term BITREx in wild-type and *rtt109Δ* cells. Similar to Figure 2J, but flanking region probes were used for Southern blot hybridization. White and orange arrowheads indicate chromosome VIII and EcoRI-excised *CUP1* array, respectively.



(legend on next page)

Figure S7: BITREx of non-*CUP1* arrays, related to Figure 4

- (A) *ENA1/2/5* array length of the parental strain used in this study. The *ENA1/2/5* array comprises a tandem array of three paralogous genes, namely *ENA1*, *ENA2*, and *ENA5*, on chromosome IV in the S288C reference genome sequence. However, other strains were reported to have four or more paralogs,^{25,26} and the strain used in this study has five paralogs. Nanopore reads containing both the 5'- and 3'-flanking regions of *ENA1/2/5* array were used to determine the distribution of *ENA1/2/5* copy number.
- (B) Genomic location of *ENA1/2/5* array. Nanopore reads containing *ENA1/2/5* were selected, and their non-*ENA1/2/5* portions were mapped to the reference genome sequence.
- (C) *ENA1/2/5* array and target sites of gRNAs tested in this study. Similar to Figure 1C.
- (D) Performance of eight gRNAs shown in (C). Similar to Figures 1D and 1E. Shading and error bar, SD (n = 3 biological replicates). The population average copy number increased to as many as 8.4 copies after 10 days of BITREx.
- (E) Representative dot plots comparing nanopore reads to the reference sequence of *ENA1/2/5* repeat unit. Genomic DNAs prepared from the cells with four gRNAs (KRS1F1, KRS1F2, KRS1F4, or RSM10R4) on day 10 were used for the nanopore sequencing.
- (F) *ymNGRU* arrays generated on chromosome VIII. A single *CUP1RU* at the *CUP1* locus on chromosome VIII was replaced by a single copy or tandemly duplicated copies of *ymNGRU* using genome editing. The *ymNGRU* consists of a yeast codon-optimized coding sequence for the fluorescent protein mNeonGreen (*ymNG*), preceded by the *NOP1* promoter and the coding sequence for the nuclear localization signal-containing domain of Nop1 (amino acid residues 1–90),²⁷ and followed by the *ADH1* terminator.
- (G) CNA of *ymNG* by BITREx with gRNA1 in the strain bearing either a single copy *ymNGRU* (1×*ymNGRU*) or a two-unit *ymNGRU* array (2×*ymNGRU*). Shading, SD (n = 3 biological replicates). The population average copy number increased to as many as 8.1 copies after 3 days of BITREx.
- (H) Distribution of *ymNGRU* copy number in nanopore reads spanning the entire array in the 1×*ymNGRU* and 2×*ymNGRU* strains on day 3 of BITREx.
- (I) Fluorescence microscopic analysis of 1×*ymNGRU* and 2×*ymNGRU* strains. Upper panel, representative microscopic images. These strains have *NUP49-mCherry* to visualize the nuclei (magenta). FL, fluorescence; DIC, differential interference contrast. Scale bar, 20 μm. Lower panel, quantification of fluorescence intensity. Box plots indicate the distribution of the average mNeonGreen fluorescence intensity in the nuclear region. The bottom and top of the box show the first and third quartiles, respectively. The bar in each box represents the median value, and the error bars represent the range of values. *P < 0.001 (one-way ANOVA test).

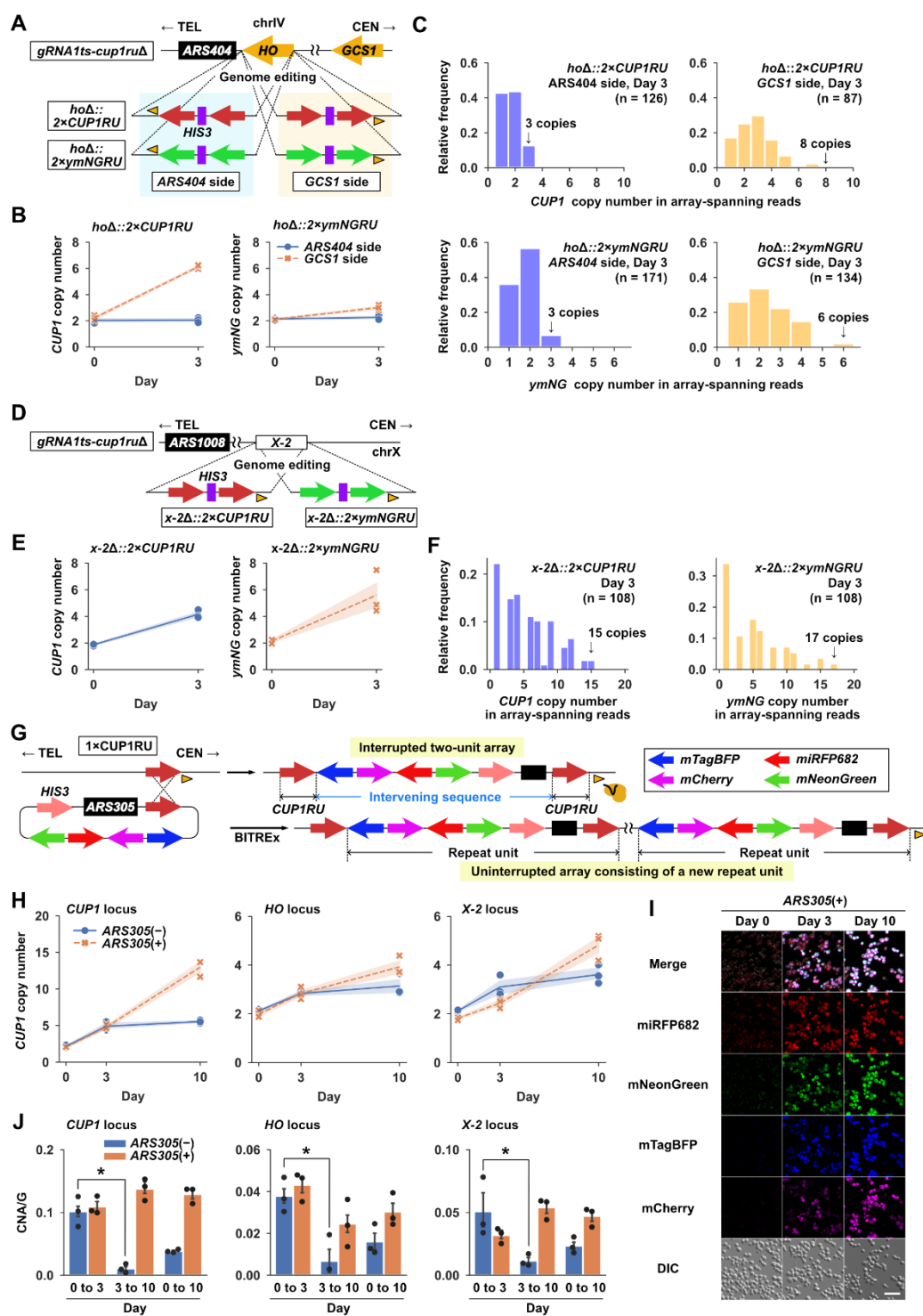
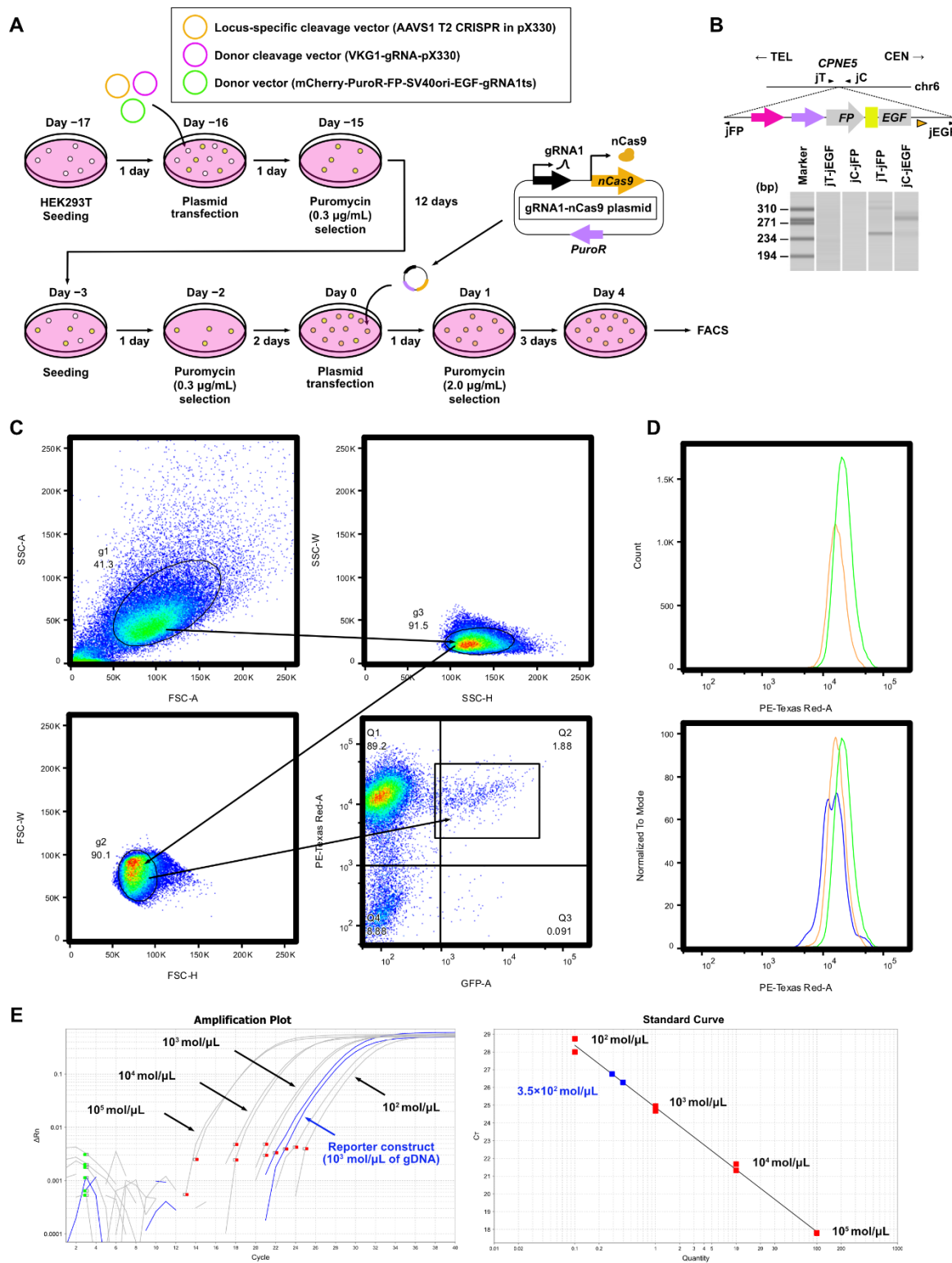


Figure S8: BITREx of interrupted two-unit arrays, related to Figure 4

- (A) Interrupted two-unit arrays of *CUP1RU* and *ymNGRU* inserted at the *HO* locus on chromosome IV. Note that these arrays include *HIS3* between the two repeat units to facilitate strain construction. They can thus be interpreted as interrupted two-unit arrays of *CUP1RU/ymNGRU*. We inserted the interrupted two-unit array in two orientations: in one strain, the nick can be introduced on the *ARS404*-proximal side, and in the other strain, on the *GCS1*-proximal side of the array. Orange arrowheads, gRNA1 target site.
- (B) CNA of *CUP1* and *ymNG* in the *hoΔ::2×CUP1RU* and *hoΔ::2×ymNGRU* strains, respectively. Shading, SD (n = 3 biological replicates).
- (C) Distribution of *CUP1/ymNG* copy number in nanopore reads spanning the entire array obtained on day 3.
- (D) Interrupted two-unit arrays of *CUP1RU* and *ymNGRU* integrated to the *X-2* locus²⁹ on chromosome X. *ARS1008* is the nearest ARS in the side opposite to the gRNA1 target site (orange arrowhead).
- (E) CNA of *CUP1* and *ymNG* in the *x-2Δ::2×CUP1RU* and *x-2Δ::2×ymNGRU* strains, respectively. Shading, SD (n = 3 biological replicates).
- (F) Distribution of *CUP1/ymNG* copy number in nanopore reads spanning the entire array obtained on day 3.
- (G) BITREx of a two-unit *CUP1RU* array interrupted by an intervening sequence containing four fluorescent protein genes (*mTagBFP*, *miRFP682*, *mCherry*, and *mNeonGreen*) and *HIS3*. This array was generated through recombination between a genomic single-copy *CUP1RU* and a plasmid bearing a *CUP1RU* with the intervening sequence. The first cycle of BIR at the interrupted *2×CUP1RU* array generates an uninterrupted two-unit array consisting of a new repeat unit that includes the intervening sequence and *CUP1RU*.
- (H) CNA of *CUP1* in the strains bearing the interrupted *CUP1* array in (G) at *CUP1*, *HO*, or *X-2* loci. Isogenic stains without the embedded *ARS305* were also shown. Shading, SD (n = 3 biological replicates). *P < 0.05 (Student's t-test). Notably, the effects of the embedded *ARS305* were evident between days 3 and 10 but not between days 0 and 3.
- (I) Microscopic images of the strain bearing the interrupted *CUP1* array in (G). Cells were subjected to fluorescence microscopy on days 0, 3, and 10. Red, *miRFP682*; green, *mNeonGreen*; blue, *mTagBFP*; magenta, *mCherry*; grey, DIC. Scale bar, 20 μm.
- (J) Alteration of CNA/G during BITREx. CNA/G was calculated for the periods from day 0 to 3, day 3 to 10, and day 0 to 10. *P < 0.05 (Student's t-test). The CNA/G between days 3 and 10 is significantly lower than that between days 0 and 3 in the absence of *ARS305*, but not in its presence. This is presumably because the desirable replication fork directionality at the nick was similar between ARS-less and ARS-containing arrays while they remained relatively short but could not be maintained in the ARS-less arrays as they expanded, unlike in the ARS-containing arrays.



(legend on next page)

Figure S9: BITREx in mammalian cells, related to Figure 6

- (A) Experimental procedure for BITREx in HEK293T cells. Using the three vectors for the VIKING method, we integrated the reporter construct into the *AAVS1* locus. Following the selection of HEK293T cells with the integrated construct by low concentration of puromycin, the gRNA1-nCas9 co-expression plasmid was transfected and selected by high concentration of puromycin. These cells were used for flow sorting.
- (B) PCR genotyping of the *CPNE5* locus. PCR analysis using four primer combinations shown in the bottom panel consistently confirmed that the *EGFP* reconstitution reporter cassette was integrated as illustrated in the top panel.
- (C) Light scatter-based gating of the cell population depicted in Figure 6C.
- (D) Similar to Figure 6E, except fluorescence of mCherry was detected using the PE-Texas Red A channel. Based on these data and those in Figure 6E, we calculated the fluorescence ratio (EGFP/mCherry) in Figure 6F.
- (E) Quantification of residual donor plasmid by qPCR. A primer pair spanning the cleavage site of the donor plasmid harboring the *EGFP* reconstitution reporter construct cassette, jEGF and jFP in (B), was designed to specifically detect residual donor plasmids while excluding signals from those integrated into the genome. qPCR analysis revealed that 350 moles of the residual donor plasmid were present in genomic DNA equivalent to 1,000 moles of a haploid genome. Considering the hypotriploid nature of HEK293 cells,⁶⁶ each puromycin-selected cell is estimated to contain ~1 copy of the residual plasmid.

Supplemental references

- [S1] Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.V., Doudna, J.A. et al. (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451. <https://doi.org/10.1016/j.cell.2013.06.044>