



A repackaged CRISPR platform increases homology-directed repair for yeast engineering

Deon Ploessl^{1,2}, Yuxin Zhao^{1,2}, Mingfeng Cao^{1,2}, Saptarshi Ghosh^{1,2}, Carmen Lopez^{1,3}, Maryam Sayadi⁴, Siva Chudalayandi⁴, Andrew Severin⁴, Lei Huang¹, Marissa Gustafson¹ and Zengyi Shao^{1,2,3,5,6,7}✉

Inefficient homology-directed repair (HDR) constrains CRISPR–Cas9 genome editing in organisms that preferentially employ nonhomologous end joining (NHEJ) to fix DNA double-strand breaks (DSBs). Current strategies used to alleviate NHEJ proficiency involve NHEJ disruption. To confer precision editing without NHEJ disruption, we identified the shortcomings of the conventional CRISPR platforms and developed a CRISPR platform—lowered indel nuclease system enabling accurate repair (LINEAR)—which enhanced HDR rates (to 67–100%) compared to those in previous reports using conventional platforms in four NHEJ-proficient yeasts. With NHEJ preserved, we demonstrate its ability to survey genomic landscapes, identifying loci whose spatiotemporal genomic architectures yield favorable expression dynamics for heterologous pathways. We present a case study that deploys LINEAR precision editing and NHEJ-mediated random integration to rapidly engineer and optimize a microbial factory to produce (S)-norcoclaurine. Taken together, this work demonstrates how to leverage an antagonizing pair of DNA DSB repair pathways to expand the current collection of microbial factories.

Engineering of microbial factories enables the sustainable production of an array of compounds, from commodity fuels/chemicals to high-value pharmaceuticals. *Saccharomyces cerevisiae* has long served as a broad-range eukaryotic microbial factory, attributable to its thorough characterization and comprehensive engineering toolkit. The convenience of leveraging *S. cerevisiae* has resulted in its overrepresentation as a microbial factory¹. The concurrent discovery and characterization of nonconventional organisms with diverse phenotypes has resulted in a paradigm shift towards identification of pre-existing phenotypes tailored for the desired application². Examples include: *Yarrowia lipolytica* (lipid production), *Kluyveromyces marxianus* (thermotolerant, broad substrate utilization), *Hansenula polymorpha* (thermotolerant, protein production) and *Scheffersomyces stipitis* (xylose utilization, aromatic production)^{3–6}. Precision genome editing is one of the most desirable tools in microbial factory engineering. Unlike *S. cerevisiae*, which primarily employs HDR to repair DSBs, nonconventional organisms primarily employ NHEJ⁷. NHEJ proficiency extends beyond yeast to plants, animals and humans, presenting a ubiquitous obstacle to many researchers who implement CRISPR and related technologies that rely on HDR-mediated genome editing.

Current strategies used to combat NHEJ proficiency involve knockout/knockdown of NHEJ-associated genes or chemical inhibition of gene products^{2,8}. Interfering with a critical DSB repair mechanism negatively impacts cellular fitness/growth, potentially hindering microbial factory capacity⁹. NHEJ, simpler and faster than HDR⁹, grants a cell more capacity to repair DSBs and safeguard its genome.

Working with *S. stipitis* as a model NHEJ-proficient host, we discovered that NHEJ disruption could impede microbial factory performance. Considering the cumbersome nature of engineering

NHEJ-deficient strains and the potential for undesirable downstream performance outcomes, we sought to establish a precision genome-editing workflow without reliance on NHEJ disruption. We identified several shortcomings with conventional yeast CRISPR platforms to develop a revamped CRISPR platform, LINEAR.

Here we show that temporal control of Cas9 expression and repackaging the delivery of CRISPR expression components and donor DNA in a single, linear DNA fragment boosted precise genome editing to efficiencies of 67–100% in several NHEJ-proficient strains. We then leverage the intact NHEJ as a fine-tuning pathway expression tool to illustrate how NHEJ and LINEAR function synergistically in strain engineering, rapidly creating an aromatic compound biosynthesis platform to produce (S)-norcoclaurine, the key precursor for benzyloquinoline alkaloid (BIA) biosynthesis.

Results

NHEJ disruption may diminish microbial factory capacity. To assess NHEJ proficiency and the associated challenge it presents in genome editing, we selected three noteworthy nonconventional yeast species (*Y. lipolytica*, *K. marxianus* and *S. stipitis*) and benchmarked them with *S. cerevisiae*. We cotransformed each species with a digested vector containing a disrupted intense green fluorescent protein (GFP) cassette and a repair fragment containing 30-base-pair (bp) homology arms at the digested ends (Fig. 1a). Restoration of the GFP ORF required HDR-mediated integration of the repair fragment, thereby coupling fluorescence to HDR efficiency. As anticipated, substantially lower ratios of fluorescent colonies to total colony numbers were observed for nonconventional yeasts compared to *S. cerevisiae* (Fig. 1b). Among yeasts surveyed, *S. stipitis* demonstrated the highest propensity (~90%) to repair the vector via NHEJ and/or other error-prone pathways. Predicated

¹Department of Chemical and Biological Engineering, Iowa State University, Ames, IA, USA. ²NSF Engineering Research Center for Biorenewable Chemicals, Iowa State University, Ames, IA, USA. ³Interdepartmental Microbiology Program, Iowa State University, Ames, IA, USA. ⁴The Genome Informatics Facility, Iowa State University, Ames, IA, USA. ⁵Bioeconomy Institute, Iowa State University, Ames, IA, USA. ⁶The Ames Laboratory, Ames, IA, USA. ⁷DOE Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana-Champaign, Urbana, IL, USA. ✉e-mail: zyshao@iastate.edu

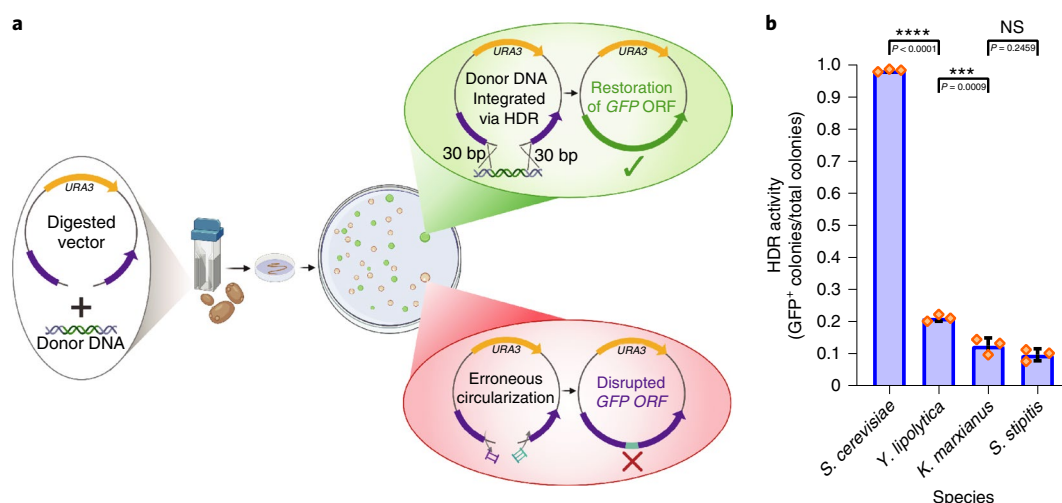


Fig. 1 | Surveying NHEJ preference in industrially relevant yeast species. **a**, Schematic of an assay implemented to assess NHEJ preference. We speculate that this assay, which relies on plasmid-based recombination, still underrepresented NHEJ preference relative to the context of repairing Cas9-induced DSBs on chromosomal DNA, because chromatin accessibility contributes an additional layer of complexity that can suppress HDR. Furthermore, in contrast to restoration of a nonessential *GFP* gene on extrachromosomal DNA, a chromosomal DSB imparts direct pressure on the cell to seal the lesion and safeguard its genome. **b**, HDR activities assessed in four yeasts. Results are presented as mean $\pm$ s.d. of three biological replicates. Statistical analysis was performed using one-way analysis of variance (ANOVA), with the mean of every sample compared with that of every other sample and Tukey's post hoc test for multiple comparisons. Select comparisons are shown. NS, no significance; *** $P < 0.001$, **** $P < 0.0001$.

on our findings, we selected *S. stipitis* as a model NHEJ-proficient organism, ideal for establishing a generalizable nonconventional strain engineering workflow.

The heterodimer of KU70p and KU80p functions in DNA damage response (DDR), committing the cell to NHEJ by capping the exposed ends of a DSB, blocking HDR-associated proteins and recruiting NHEJ-associated proteins to process and ligate the ends. Deletion of *KU70/80* to bias HDR is commonly implemented in yeast genome editing, with our laboratory disrupting *KU70/80* in *S. stipitis*¹⁰. Considering the cumbersome nature of this endeavor (attributable to lowered transformation efficiencies in strains $\Delta ku70$ and $\Delta ku80$ and the arduous process of iterative rounds of genome modification and marker recycling), the associated reduction in vitality/genome instability of the resultant $\Delta ku70\Delta ku80$ strain (consistent with work in mammalian cells^{11,12}) and the NHEJ proficiency of the wild-type strain, we questioned whether NHEJ disruption could result in deleterious side effects that hinder microbial factory performance. Although *KU70/80* can be restored in the final engineered strain, this strategy requires additional rounds of genome modification and marker recycling. An ideal genome-editing workflow should minimize engineering efforts whilst preserving microbial factory performance.

Growth analyses revealed that $\Delta ku70/80$ grew markedly more slowly than the wild type in both selective (basal) and nonselective (rich) media (Fig. 2a). Intriguingly, we observed a greater disparity in growth rate between the wild type and $\Delta ku70/80$ strains in rich medium than in basal medium. Inspection of the semilogarithmic transformed growth curve for $\Delta ku70/80$ in rich medium yielded two distinct linear regimes (Fig. 2b). The 'fast' regime, encompassing the initial 2.75 h, exhibited a growth rate virtually identical to that of the wild type ($\mu = 0.25 \text{ h}^{-1}$). This was immediately proceeded with a 'slow' regime and a drastically reduced growth rate ($\mu = 0.08 \text{ h}^{-1}$). We postulate that the higher metabolic rate from growth in rich medium and the associated increase in DNA-damaging agents such as reactive oxygen species (ROS) exacerbated the weakened DDR of $\Delta ku70/80$. Conversely, lower nutrient quantity in the basal medium slowed metabolic rate, blunting accumulation of ROS and

thus dampening the dependence of cellular proliferation on DDR. We also explored the impact of NHEJ disruption on performance by overexpressing the shikimate accumulation pathway (a reporter metabolite used for assessment of aromatic production potential) in the wild-type and $\Delta ku70/80$ strains. The wild type produced over twofold more shikimate compared to $\Delta ku70/80$ (Fig. 2c), indicating potential negative ramifications of NHEJ disruption in *S. stipitis*. We advise researchers to screen for any deleterious side effects when disrupting NHEJ in their strains.

Exploring a temporal control strategy. After the discovery of nonidealities associated with NHEJ disruption variants and considering the cumbersome nature of deleting *KU70/80*, we sought alternative methods to enhance HDR-mediated genome editing. An extensive review of 200 instances of genome editing in yeasts identified four main platforms (Extended Data Fig. 1), with most (76%) implementing cotransformation of a Cas9/single guide RNA (sgRNA)-expressing plasmid and a linear donor DNA fragment (Platform I). Genomically integrated CRISPR and donor DNA (Platform II), consolidated CRISPR–sgRNA and donor plasmids (Platform IV) and NHEJ-mediated deletion/no donor DNA (Platform V) correspond to the remaining 11, 8 and 5%, respectively. We expanded our review to vertebrates, because NHEJ proficiency also presents a substantial hurdle in genome editing of higher-level eukaryotes. Reviewing an additional 200 instances of vertebrate genome editing, we determined that Platform I represents the consensus platform (Extended Data Fig. 1). Notably, the search in vertebrate genome editing identified a strategy potentially applicable to yeasts: multiple works reported that temporal control of Cas9 expression enhanced HDR in vertebrates^{13–15}.

Repetitive DNA and the associated homologous regions scattered throughout the genomes of NHEJ-proficient organisms increase the probability of a non-sister chromatid incorrectly serving as a repair template for DSBs, potentiating large-scale chromosomal rearrangements¹⁶. HDR activity in NHEJ-proficient organisms is thus constrained to actively dividing cells in the mid-S, G2 and M phases of the cell cycle, when a sister chromatid can serve as a repair template¹⁷.

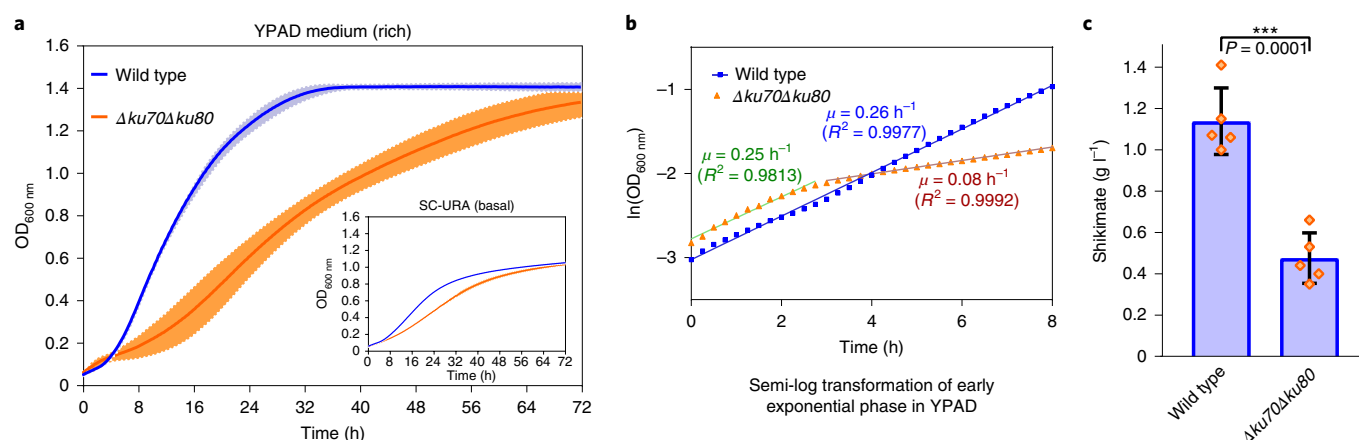


Fig. 2 | NHEJ disruption hinders growth and fermentation performance of *S. stipitis*. **a**, Comparison of growth rates of *S. stipitis* strains with wild-type and $\Delta ku70/80$ genotypes under rich and basal medium conditions. Results are presented as mean $\pm$ s.d. of three biological replicates. **b**, Semi-log transformations of early log phases in rich media presented in **a**. The wild-type genotype grew normally (blue, $\mu = 0.26 \text{ h}^{-1}$), whereas the $\Delta ku70/80$ genotype transitioned from normal growth (green, $\mu = 0.25 \text{ h}^{-1}$) to attenuated growth (red, $\mu = 0.08 \text{ h}^{-1}$) after approximately one doubling time (2.75 h). **c**, Shikimate fermentation conducted in each genotype indicates that NHEJ disruption could inhibit strain performance. Results presented as mean $\pm$ s.d. of five biological replicates. Statistical analysis was performed using a two-sided Welch's t-test. *** $P < 0.001$.

Geminin is an APC^{CDH1} substrate present in multiple model eukaryotes (that is, fly, frog and silkworm)^{18–20} (Fig. 3a). It inhibits DNA replication, and its temporal expression pattern is such that DNA replication occurs exactly once per cell cycle²¹. Fusion of Cas9 with a 110-residue-long peptide corresponding to the N terminus of geminin synchronized Cas9 DSB induction with maximum HDR activity, mediating precision gene editing in human cells²².

Considering the intricacies of cell cycle processes and that some mechanisms utilized by yeast deviate from higher eukaryotes, we questioned whether this human geminin (hGem) tag could confer temporal expression control in yeast (Supplementary Figs. 1–4). To this end, we constructed two blue fluorescent protein (BFP)-expressing *S. stipitis* strains, one expressing BFP and the other expressing a modified BFP with hGem fused to the C terminus. To assess hGem functionality we conducted a time course study, analyzing BFP expression of 25 subcultures from each of the two genotypes (Fig. 3b–e). We designed our assay to eliminate the possibility that attenuated BFP expression in the *BFP-hGem* genotype was not merely an artifact of hGem interfering with posttranslational processing and/or folding. We used the constitutive *TDH2* glycolytic promoter to drive expression of the cassette. Considering the brevity of the G1 phase relative to the remainder of the cell cycle in rapidly proliferating yeast^{23,24}, we anticipated that constitutive transcription would result in a net accumulation of BFP-hGem and a subsequent comparable expression level with that of BFP during the exponential growth phase, provided that hGem did not interfere with BFP function (Fig. 3c). As cells transitioned from exponential to stationary phase, which is associated with a prolonged G1 phase and downregulation of glycolytic transcription in response to depleted glucose levels, these conditions would favor BFP-hGem degradation. This shift towards degradation should become more pronounced once fully transitioning to stationary phase, because senescent cells arrest in G1 phase^{25,26}. As anticipated, we observed appreciable and more drastic reduction in both BFP expression and the percentage of BFP⁺ cells in the *BFP-hGem* genotype compared to the *BFP* genotype as the cultures approached stationary phase (Fig. 3d,e). Moreover, no discernible difference was observed until this transition, corroborating that reduced BFP expression was attributed to APC^{CDH1}-targeted G1 phase degradation as opposed to intrinsic interference from hGem.

Conventional platforms are impotent in the NHEJ-proficient phenotype. To assess whether the temporal dynamics conferred by hGem could mediate HDR, we tagged the C terminus of Cas9 with hGem in our expression plasmid and attempted to delete *XYL2* encoding xylitol dehydrogenase, which is nonessential when glucose is present as a carbon source. We cotransformed *S. stipitis* with a donor DNA consisting of a *GFP* cassette and 500-bp homologies flanking the target cut site with a Cas9-hGem/sgRNA-expressing plasmid (Platform I in Extended Data Fig. 1). Implementation of this most popular Platform I in multiple experimental attempts yielded unexpected genome-editing outcomes with no observable signs of HDR. We also tested a consolidated Cas9-hGem/sgRNA-expressing and donor DNA ‘self-digesting’ plasmid (Platform IV in Extended Data Fig. 1), which yielded similar failures (Extended Data Fig. 2). The inability to obtain precisely edited clones with the aid of a temporal control strategy that succeeded in engineering higher eukaryotes demonstrated the ineffectiveness of the conventional CRISPR platforms in *S. stipitis* (Supplementary Note 1).

Notably, we routinely observed a subset of colonies with visibly intense *GFP* expression relative to standard *GFP*-expressing colonies. We postulated that this phenomenon was the result of an aberrant recombination event whereby the donor DNA circularized with a genome-originating, autonomously replicating sequence (ARS), enabling the *GFP* gene to replicate and propagate to progeny. We conducted an aberrant recombination assay, verifying that aberrant recombination events mediated by NHEJ activity resulted in the formation of ‘pseudoplasmids’ (Extended Data Figs. 3 and 4, Supplementary Fig. 5 and Supplementary Note 1). These results demonstrate the ineffectiveness of the conventional, plasmid-based CRISPR platforms in *S. stipitis* and how NHEJ enables cells to bypass the selection/screening process, reiterating the intrinsically difficult task of precision editing in NHEJ-proficient organisms.

LINEAR enables precise editing in a NHEJ-proficient background. Assessment of failures associated with the traditional CRISPR platforms guided the development of the LINEAR CRISPR platform. LINEAR consists of a single linear DNA molecule encoding the Cas9-hGem/sgRNA expression cassettes and the donor fragment (Fig. 4a and Extended Data Fig. 5). Transformation of a single fragment ensures that transformed cells take up all necessary

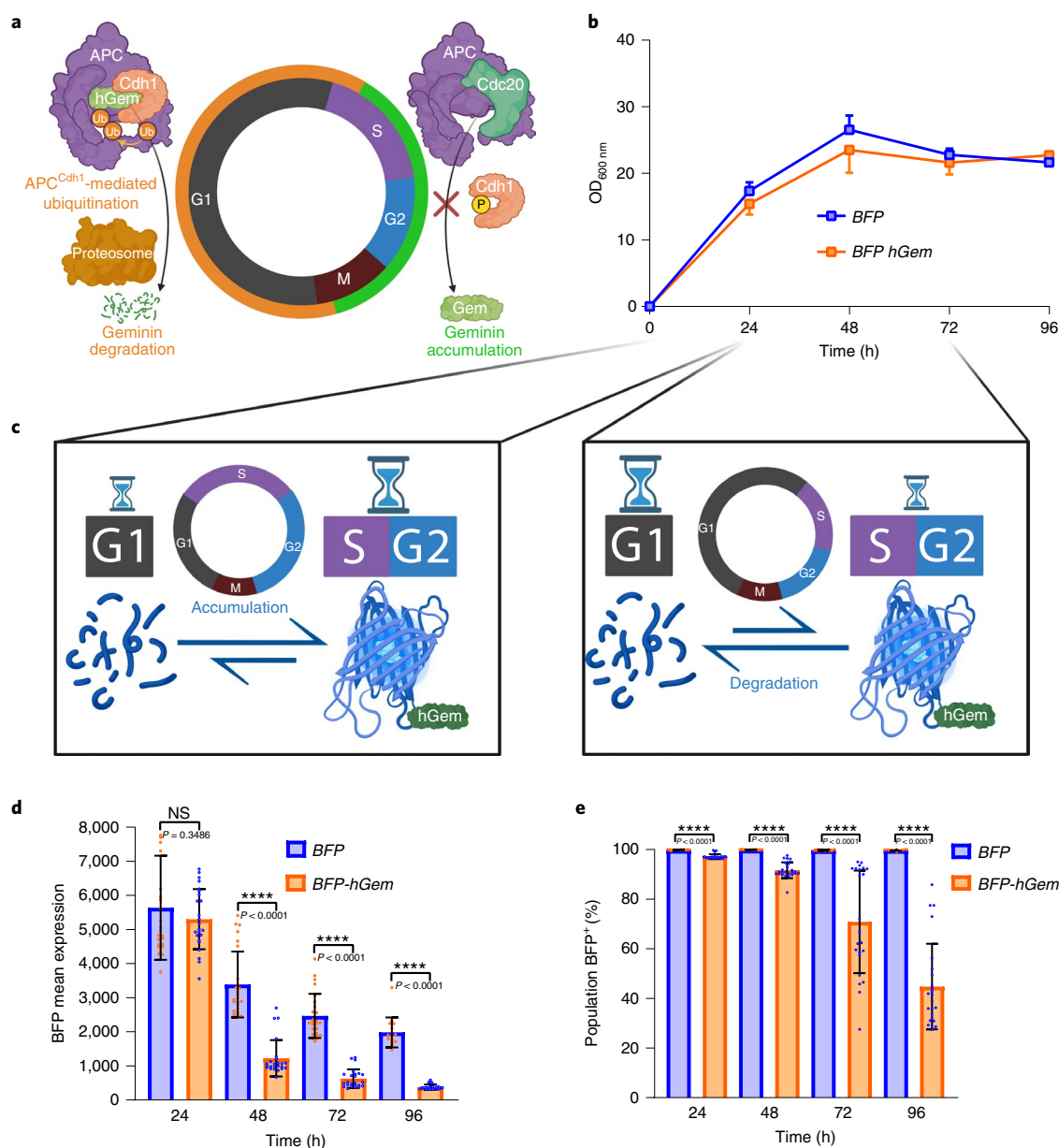


Fig. 3 | Leveraging geminin dynamics to establish a temporal expression control. **a**, APC^{Cdh1} restricts geminin expression to the S/G2/early-M phases of the cell cycle. **b**, Growth curves of strains BFP and BFP-hGem. Results are presented as mean $\pm$ s.d. of five biological replicates (one-sided error bars used to minimize clashing of the overlaid curves). **c**, Net accumulation or degradation of BFP-hGem is contingent on the duration of the G1 phase relative to the remainder of the cell cycle. In rapidly proliferating cells, the relative duration of G1 is reduced compared to that in cells approaching stationary phase. **d**, Comparison of mean BFP expression of strains BFP and BFP-hGem at various time points. **e**, Comparison of population percentage of BFP-expressing cells of BFP and BFP-hGem strains at various time points. Results are presented as mean $\pm$ s.d. of 25 biological replicates. Statistical analysis was performed using a two-sided Welch's *t*-test. The gating methodology is summarized in Methods. **** $P < 0.0001$.

components (Fig. 4b,c). On the donor fragment, LINEAR substitutes a fluorescent reporter with a selection marker, applying selective pressure to employ the HDR pathway²⁷ and facilitating direct screening of precisely edited clones. Selection marker recycling permits iterative rounds of LINEAR editing. The removal of episomal elements (that is, ARS and centromere) from the LINEAR platform prevents propagation of the Cas9-sgRNA expression cassette to progeny, ameliorating the toxicity issues associated with constitutive nuclease expression.

As a proof of concept, we implemented LINEAR for targeted deletion of *XYL2* in *S. stipitidis*. In as little as 2 days, numerous colonies emerged (compared to 4 days and substantially reduced colony

numbers with Platforms I and IV; Extended Data Fig. 6), suggesting that hGem and the removal of episomal elements minimized nuclease toxicity. Sequencing confirmed 100% HDR-mediated precise insertion of donor DNA (From ten randomly picked clones; Supplementary Fig. 6), representing a substantial improvement over the previously reported editing efficiency (Fig. 5). In the initial design, we inserted a GFP cassette between the donor DNA and sgRNA cassette components to assess the frequency of aberrant pseudoplasmid formation (Extended Data Figs. 3 and 4). Among hundreds of colonies, only one visibly expressed GFP (Supplementary Fig. 6a), suggesting an extremely low occurrence of aberrant recombination events with the LINEAR platform. Accordingly, we omitted the

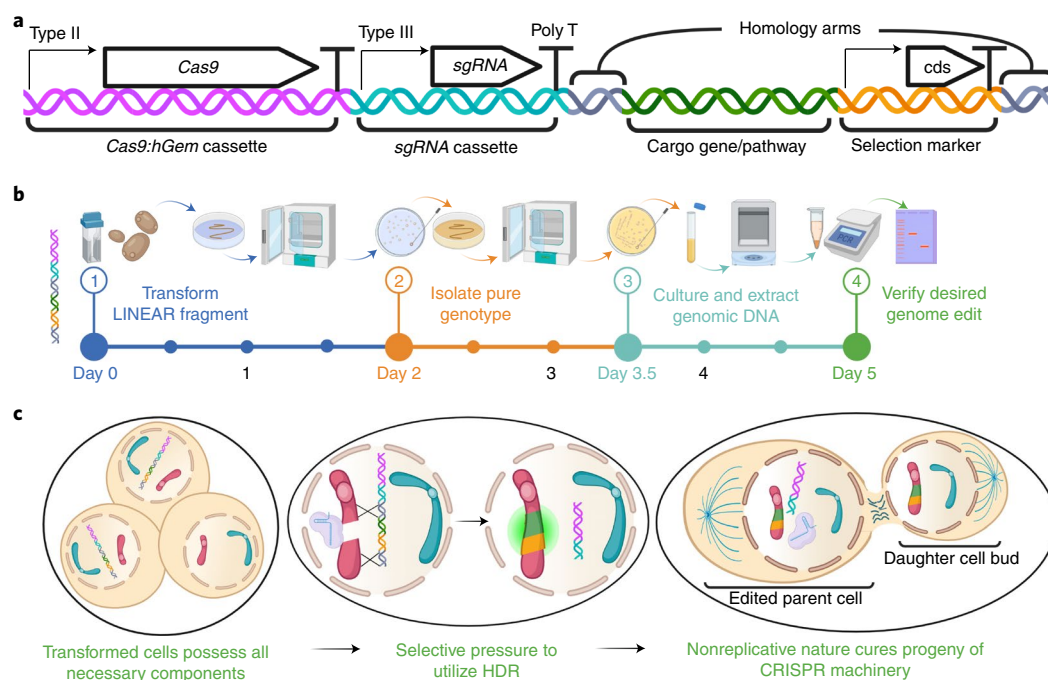


Fig. 4 | LINEAR CRISPR platform. **a**, A Synthetic Biology Open Language visual of the genetic components of LINEAR. **b**, The LINEAR platform expedites experimental workflow from 2 weeks (Extended Data Fig. 2a) to 5 days. **c**, LINEAR addresses the shortcomings associated with the conventional CRISPR platforms, enhancing editing outcomes.

negative reporter *GFP* cassette to simplify subsequent designs. The spontaneous elimination of nonpropagative LINEAR DNA in progeny reduces incubation periods associated with constitutive Cas9 expression and plasmid curing in the conventional platforms, expediting the workflow from 14 days (Extended Data Fig. 2a) to 5 days (Fig. 4b). Having established LINEAR as a precision gene knockout tool, we explored LINEAR's capabilities in achieving more extensive genome modifications. Using LINEAR, we precisely integrated the three-gene, 5.7-kb (*S*)-norocloaurine pathway and the three-gene, 12-kb shikimate pathway into the *XYL2* locus, with HDR efficiency of 80 and 41%, respectively (Supplementary Fig. 7).

To thoroughly characterize the LINEAR platform, we assessed the individual impacts of homology arm length and the hGem tag on editing outcomes. For the *XYL2::URA3* LINEAR design (containing the initial negative reporter *GFP* cassette to assess aberrant recombination rate), we removed hGem and/or truncated the homology arms to 50bp, a length feasibly appended to a primer (thereby modestly simplifying cloning; Extended Data Fig. 5a). Truncation of the homology arms drastically reduced colony number, increased aberrant recombination (that is, *GFP*⁺ clones) and lowered HDR efficiency slightly from 100 to 90% with hGem present in the LINEAR platform and from 90 to 70% with no hGem tag (Extended Data Figs. 7 and 8). Although appreciable HDR was retained with the 50-bp homology arms, it is worth noting that the negative *GFP* reporter was needed to filter *GFP*⁺ clones representing aberrant recombination. In contrast, when 500-bp homology arms were applied, almost no aberrant recombination occurred. Therefore, we recommend researchers use 500-bp homology arms to ensure obtainment of the desired genome modification with minimal screening effort.

In regard to the benefit of the hGem tag, this boosted HDR efficiency from 90 to 100% and from 70 to 90% when using 500- and 50-bp homology arms, respectively, when *XYL2* was replaced by *URA3* (Extended Data Figs. 7 and 8). To verify its contribution to HDR-mediated insertion of a long pathway, we also removed hGem

from the *XYL2::GSAU* LINEAR design and observed a 10% reduction in HDR efficiency (Extended Data Fig. 9). For all three platforms tested, we observed instances of error-prone insertion of the donor DNA when hGem was removed, indicating that inclusion of hGem prevented induction of DSBs in the G1 phase when HDR was inactive (Extended Data Figs. 7d, 8e and 9d).

To test general applicability, we implemented the LINEAR platform in three other industrially relevant NHEJ-proficient yeasts. Previous gene knockout efforts in these species implemented the conventional CRISPR platforms and were reliant on either unpredictable NHEJ-derived indels (*K. marxianus*²⁸ and *H. polymorpha*^{29,30}) or the creation of Δku strains to achieve sufficient HDR (*Y. lipolytica*^{31,32}). Targeting the same loci reported in the previous studies, LINEAR achieved highly efficient HDR in the wild-type background (67–100%; Fig. 5 and Supplementary Fig. 8). We also tested different loci to ensure versatility, achieving efficient precision deletions of *ARO10* in *S. stipitis* and *BAT1* in *Y. lipolytica* (70–90%; Supplementary Fig. 9).

Leveraging NHEJ and LINEAR in strain engineering. Having demonstrated LINEAR as an efficient and robust precision genome-editing tool that functions in the NHEJ-proficient background, we sought to establish heuristics in engineering of NHEJ-proficient yeasts. Heterologous pathway tuning presents a major rate-limiting step in the engineering of industrial strains, requiring stable genomic expression. A common perception persists that integration of pathway genes to accessible, 'safe harbor' regions on the genome represents the best option because transcriptional machinery will be actively recruited and promote high-level pathway transcription. However, this oversimplifies the pathway optimization equation: safe harbor loci are necessary, but not sufficient, for obtaining balanced expression of all pathway genes³³, because the metabolic burden associated with high-level expression can hinder production of the desired compound³⁴. Effective production also demands numerous cross-talks with global metabolism,

Strain	Locus	Previous CRISPR–Cas9 strategies				LINEAR
		CRISPR platform employed	WT or NHEJ-deficient genotype	KO efficiency (%) (HDR or NHEJ)	Reference no.	KO efficiency (%) (HDR in WT genotype)
<i>S. stipitis</i> FLP UC7	<i>XYL2</i>	Cas vector and donor DNA (Platform I)	WT	0 (HDR)	This study	100
<i>K. marxianus</i> YS402	<i>XYL2</i>	Cas vector (Platform V)	WT	NA (HDR) 66 (NHEJ)	28	100
<i>H. polymorpha</i> DL1	<i>ADE2</i>	Cas vector and donor DNA (Platform I)	WT	28 (HDR)	29	67
<i>Y. lipolytica</i> PO1f	<i>PEX10</i>	Consolidated cas/donor vector (Platform IV)	$\Delta ku70\Delta ku80$	28 (HDR)	31	77
		Cas vector and donor DNA (Platform I)	WT	17 (HDR)	32	
			WT	16 (HDR)		
			$\Delta ku70$	86 (HDR)		

Fig. 5 | Comparison of editing efficiency between the LINEAR platform and the conventional CRISPR platforms in various nonconventional species.

LINEAR enhanced HDR editing outcomes in all four nonconventional species tested, demonstrating efficiency and tractability. A select number of colonies were randomly screened for genome-editing verification (*S. stipitis*, $n=10$; *K. marxianus*, $n=9$; *Y. lipolytica*, $n=13$; *H. polymorpha*, $n=9$). WT, wild type. For detailed summaries of downstream verification, refer to Supplementary Figs. 6 and 8.

distributing cellular resources (for example, NAD(P)H, ATP and key metabolites) to maximize productivity. Even with the assistance of genome-scale metabolic models, the metabolic engineering field is still far from resolving the regulation of redox balance and energetics, a prerequisite in determining the rational wiring of a heterologous pathway into the cellular metabolic network³⁵.

The multitude of factors influencing a pathway's performance renders accurate prediction of an ideal expression locus highly improbable, necessitating library generation and screening. NHEJ-mediated random integration enables rapid construction of expression libraries whose members are diversified by the spatiotemporal fluidity of the genome³⁶. Screening of NHEJ-created libraries therefore surveys the landscape of a genome for a locus whose expression dynamics is optimal for the pathway in question. Although NHEJ-mediated random integration is not a new technique for integration of heterologous pathways, we aim to change the associated connotations that NHEJ is not an ideal method and is thus deployed only when other methods such as CRISPR are not viable. Rather, we posit that NHEJ-mediated library creation enables exploration of the spatiotemporal nature of a genome for a favorable expression locus, appending time dimensionality to the current three-dimensional design paradigm in the metabolic engineering field³⁷. Precision editing via LINEAR, in conjunction with diversification of pathway expression dynamics by NHEJ, comprises a serviceable genome-editing toolkit.

To demonstrate the utility of this genome-editing toolkit, we implemented LINEAR and NHEJ-mediated random integration to engineer the plant-sourced (S)-norcoclaurine pathway in *S. stipitis*. (S)-norcoclaurine represents a primary chassis metabolite in the biosynthesis of BIAs, a highly valuable class of bioactive molecules³⁸. We split the seven-gene heterologous pathway into an upstream precursor module and a downstream (S)-norcoclaurine product module (Extended Data Fig. 10). To screen for an enhanced flux towards L-tyrosine production in a high-throughput manner, we adopted a biosensor and an associated colorimetric assay that exploits the spontaneous formation of betaxanthin³⁹, a yellow-pigmented compound. Aro10p catalyzes the first committed step in the degradative Ehrlich pathway, and its deletion has been demonstrated to improve aromatic-derived product titers in yeast⁴⁰. Accordingly, for the first step in the strain engineering process, we integrated the two-gene betaxanthin pathway (comprising *TyrH** and *DOD*) into the ORF of *ARO10* via LINEAR, simultaneously integrating the biosensor and deleting the competitive pathway via HDR (Fig. 6a).

Following isolation and confirmation of a precisely edited clone, we employed Cre-loxP recombination to recycle the *URA3* selection marker. Next, we constructed an NHEJ-mediated random integration library of the L-tyrosine precursor module by transforming this strain, $\Delta aro10::TD^*$, with a DNA fragment containing the four-gene upstream module. The pathway was flanked with *BFP* and *URA3*, with the former serving as a reporter for genomic pathway integration (we abbreviate this cassette to *BSA4* hereafter). We opted for *BFP* as the reporter in this step to avoid interference with the subsequent biosensing step (that is, the GFP fluorophore utilizes the same excitation/emission wavelengths as the biosensor). Genomic *BFP* expression and *Ura*⁺ phenotypes were used to ensure genomic insertion of the entire upstream module⁴¹. As expected, library screening revealed a spectrum of betaxanthin production with the top performer, strain *BSA4* 41, confirmed in a follow-up assay (Fig. 6b,c).

Moving forward with strain *BSA4* 41, we shifted our focus to the downstream product, (S)-norcoclaurine (Extended Data Fig. 10). We implemented LINEAR to replace the *DOD* gene associated with the betaxanthin biosensor with *DODC* and *NCS*, the downstream module genes required for (S)-norcoclaurine biosynthesis (Fig. 6a). After obtaining a precisely edited clone, 41 $\Delta aro10::TDNUloxP$ (that is, *ARO10::TyrH*, *DODC*, *NCS* and *loxP*-flanked *URA3*), we constructed a NHEJ-mediated random integration library for the three-gene downstream module. Although one copy of the downstream pathway already resided in *ARO10*, the genomic context associated with this locus may not result in favorable expression dynamics. This necessitated screening of a second locus that, in conjunction with the copy in *ARO10*, yielded a more optimal solution.

To this end, after recycling *URA3* we transformed strain 41 $\Delta aro10::TDN$ with a fragment containing the three-gene downstream module flanked with *GFP* and *URA3*. Analogous with the upstream module library, initial screening of (S)-norcoclaurine fermentation performance by liquid chromatography–mass spectrometry (LC–MS) revealed a spectrum of (S)-norcoclaurine production (Fig. 6d). Notably, the parent strain 41 $\Delta aro10::TDNU$ produced only $\sim 5 \mu\text{g l}^{-1}$, corroborating that rational selection of expression loci (such as insertion of the downstream module into the *ARO10* locus) will most probably not yield an optimal solution. The top performer produced $75 \mu\text{g l}^{-1}$, an almost 15-fold improvement compared to the parent strain (Fig. 6e).

Lastly, we sequenced the genomes of the top four performing strains by Nanopore sequencing and did not observe any large-scale

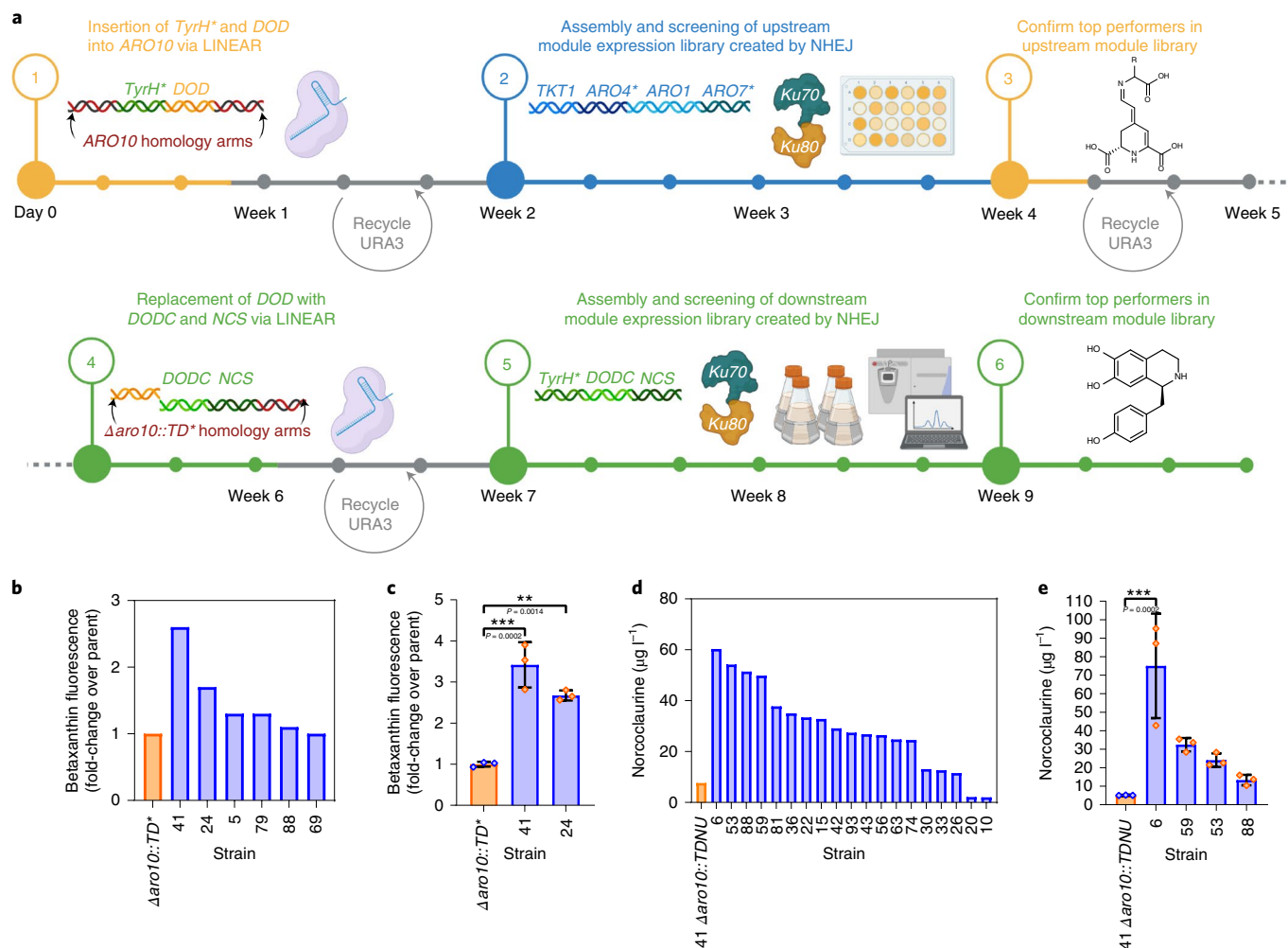


Fig. 6 | Case study demonstrating the utility of LINEAR and NHEJ synergistically in strain engineering. a, Timeline for optimization of a norcoclaurine-producing yeast factory. **b**, Screening the NHEJ-mediated expression library for the upstream module. **c**, Confirming the performance of the top producers identified from the screen in **b**. **d**, Screening the NHEJ-mediated expression library for the downstream module. **e**, Confirming the performance of the top producers identified from the screen in **d**. Strain no. 55 was removed from downstream analysis due to production instability. Results presented as mean $\pm$ s.d. of three biological replicates. Statistical analysis was performed using one-way ANOVA, with the mean of every sample compared with that of every other sample and Tukey's post hoc test for multiple comparisons. Selected comparisons are shown. $^{**}P < 0.01$, $^{***}P < 0.001$.

genomic rearrangements (Supplementary Fig. 10), indicating that the short exposure of the genome to Cas9, iterative marker recycling and NHEJ-mediated pathway integration did not compromise genome integrity in this case. These analyses also confirmed the downstream module inserted into different loci in each strain (for example, strain no. 6 integrated to Chromosome 5 whereas strain no. 59 integrated to Chromosome 4; Supplementary Table 1). These results confirm that NHEJ-mediated random integration can effectively identify loci with favorable expression dynamics, expediting strain engineering efforts.

Discussion

Beyond more easily assessable impacts of NHEJ disruption, such as growth and product titer, we also advise researchers to reconsider NHEJ disruption from the perspective of enzymes. One major advantage of yeasts as prospective microbial factories, compared to simpler prokaryotic organisms, is the former's ability to express membrane-associated, heme-dependent cytochrome P450s capable of performing intricate oxidations in the production of derivatized steroids, fatty acids and other highly sought-after natural products derived from plants⁴². For example, in addition to tyrosine

hydroxylase (encoded by *TyrH*) in the (S)-norcoclaurine pathway, multiple cytochrome P450s catalyze downstream reactions associated with the BIA class of compounds. Expression of cytochrome P450 and other metal cofactor-dependent oxidative enzymes produces appreciable levels of highly oxidative metabolic side products⁴³, imposing oxidative stress and DNA damage^{44,45}. The increased sensitivity of an NHEJ-deficient strain to DNA-damaging agents suggests that heterologous expression of cytochrome P450s may not be well tolerated in a NHEJ-deficient host.

LINEAR presents an alternative to NHEJ disruption to achieve precision genome editing, with the added benefit of not needing a stable episomal plasmid, removing the hurdles associated with identification of an ARS and a centromeric sequence⁴⁶. Currently, many nonconventional hosts lack stable episomal plasmids and the characterization of centromeres presents a major technical hurdle in the development of episomal expression platforms for eukaryotic hosts⁴⁷. Moreover, the extremely long nature of centromeres (40–110 kb) in many of these organisms (for example, the fission yeast *Schizosaccharomyces pombe*) renders them infeasible for plasmid development. The presented case study outlines a workflow that does not require NHEJ disruption or an episomal plasmid,

expediting strain engineering of novel species. With NHEJ intact, the ability of NHEJ to survey the genome could also facilitate the development of gene-trapping strategies developed in mammalian cell lines whereby random insertion and detection of a reporter cassette on the genome sheds light on novel genotype/phenotype relationships⁴⁸. Such a utility enables swift identification of gene knockout candidates to further enhance production.

For those interested in implementing LINEAR in their strains, we have provided a detailed protocol (Supplementary Note 2) and deposited annotated plasmid sequences used for the four strains in this study (Addgene nos. 174837–174840). One may question the necessity of hGem in the LINEAR platform, considering that its impact on editing efficiency was modest (Extended Data Figs. 7 and 8). Indeed, the nonpropagative nature of the LINEAR platform grants substantial control over Cas9 expression within a proliferating population, whereby the additional level of precision hGem offers may not be critical. However, in addition to the complete synchronization of DSB induction with maximal HR activity, the proteolysis of Cas9–sgRNA complexes in the proceeding G1 phase would further reduce toxicity and indels, because Cas9–sgRNA complexes propagated to progeny cells would be rapidly targeted by APC^{CDH1}. The benefit could be more pronounced in other, more difficult, loci or in more challenging organisms. Also, considering that the hGem tag demonstrated no apparent negative consequences in any context, we advocate its inclusion when the LINEAR platform is applied in a new host, to reduce screening efforts. We also explored the prospect of leveraging endogenous APC^{CDH1} substrates, because these could presumably yield higher specificity/more targeted degradation. Yeasts do not possess a geminin homolog, but rather rely on the coordination of multiple cell division cycle (Cdc) proteins to license DNA replication⁴⁹. During the onset of this project, we fused the full-length Cdc5p to Cas9 but observed a drastically reduced number of transformants, attributable to a supraphysiological level of Cdc5p in the Cdc5p–Cas9 chimera disrupting the cell cycle⁵⁰. The heterologous nature of hGem minimizes interference with endogenous cell cycle processes.

In summary, the ability to rapidly perform precision genome edits and tune pathway expression should facilitate the establishment of novel hosts as platform strains in the production of valuable compounds. The new-found genome-editing capabilities in NHEJ-proficient organisms granted by LINEAR indicate that NHEJ-proficient organisms could be potentially handled in a manner similar to that for *S. cerevisiae*. The myriad of tools developed in *S. cerevisiae* that require precise genome modifications can now be implemented in a broader spectrum of hosts more effectively.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-021-00893-5>.

Received: 10 February 2021; Accepted: 9 September 2021;

Published online: 28 October 2021

References

- Hittinger, C. T. et al. Genomics and the making of yeast biodiversity. *Curr. Opin. Genet. Dev.* **35**, 100–109 (2015).
- Thorwall, S., Schwartz, C., Chartron, J. W. & Wheeldon, I. Stress-tolerant non-conventional microbes enable next-generation chemical biosynthesis. *Nat. Chem. Biol.* **16**, 113–121 (2020).
- Goncalves, F. A., Colen, G. & Takahashi, J. A. *Yarrowia lipolytica* and its multiple applications in the biotechnological industry. *ScientificWorldJournal* **2014**, 476207 (2014).
- Fonseca, G. G., Heinzle, E., Wittmann, C. & Gombert, A. K. The yeast *Kluyveromyces marxianus* and its biotechnological potential. *Appl. Microbiol. Biotechnol.* **79**, 339–354 (2008).
- Manfrao-Netto, J. H. C., Gomes, A. M. V. & Parachin, N. S. Advances in using *Hansenula polymorpha* as chassis for recombinant protein production. *Front. Bioeng. Biotechnol.* **7**, 94 (2019).
- Gao, M. et al. Innovating a nonconventional yeast platform for producing shikimate as the building block of high-value aromatics. *ACS Synth. Biol.* **6**, 29–38 (2017).
- Lobs, A. K., Schwartz, C. & Wheeldon, I. Genome and metabolic engineering in non-conventional yeasts: current advances and applications. *Synth. Syst. Biotechnol.* **2**, 198–207 (2017).
- Kuck, U. & Hoff, B. New tools for the genetic manipulation of filamentous fungi. *Appl. Microbiol. Biotechnol.* **86**, 51–62 (2010).
- Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst.)* **7**, 1765–1771 (2008).
- Cao, M., Gao, M., Ploessl, D., Song, C. & Shao, Z. CRISPR-mediated genome editing and gene repression in *Scheffersomyces stipitis*. *Biotechnol. J.* **13**, e1700598 (2018).
- Sishc, B. J. & Davis, A. J. The role of the core non-homologous end joining factors in carcinogenesis and cancer. *Cancers (Basel)* **9**, 81 (2017).
- Ferguson, D. O. et al. The nonhomologous end-joining pathway of DNA repair is required for genomic stability and the suppression of translocations. *Proc. Natl Acad. Sci. USA* **97**, 6630–6633 (2000).
- Yang, H. et al. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154**, 1370–1379 (2013).
- Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766 (2014).
- Janssen, J. M., Chen, X., Liu, J. & Goncalves, M. The chromatin structure of CRISPR-Cas9 target DNA controls the balance between mutagenic and homology-directed gene-editing events. *Mol. Ther. Nucleic Acids* **16**, 141–154 (2019).
- Lieber, M. R. & Karanjawala, Z. E. Ageing, repetitive genomes and DNA damage. *Nat. Rev. Mol. Cell Biol.* **5**, 69–75 (2004).
- Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle* **7**, 2902–2906 (2008).
- Tang, X. F. et al. Two Geminin homologs regulate DNA replication in silkworm, *Bombyx mori*. *Cell Cycle* **16**, 830–840 (2017).
- Kerns, S. L., Torke, S. J., Benjamin, J. M. & McGarry, T. J. Geminin prevents rereplication during xenopus development. *J. Biol. Chem.* **282**, 5514–5521 (2007).
- Quinn, L. M., Herr, A., McGarry, T. J. & Richardson, H. The *Drosophila* Geminin homolog: roles for Geminin in limiting DNA replication, in anaphase and in neurogenesis. *Genes Dev.* **15**, 2741–2754 (2001).
- McGarry, T. J. & Kirschner, M. W. Geminin, an inhibitor of DNA replication, is degraded during mitosis. *Cell* **93**, 1043–1053 (1998).
- Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G. F. & Chin, L. Post-translational regulation of Cas9 during G1 enhances homology-directed repair. *Cell Rep.* **14**, 1555–1566 (2016).
- Slater, M. L., Sharrow, S. O. & Gart, J. J. Cell cycle of *Saccharomyces cerevisiae* in populations growing at different rates. *Proc. Natl Acad. Sci. USA* **74**, 3850–3854 (1977).
- Donaldson, A. D. et al. CLB5-dependent activation of late replication origins in *S. cerevisiae*. *Mol. Cell* **2**, 173–182 (1998).
- Collado, M., Blasco, M. A. & Serrano, M. Cellular senescence in cancer and aging. *Cell* **130**, 223–233 (2007).
- YMotoko Sasaki, Y. N. in *Autophagy: Cancer, Other Pathologies, Inflammation, Immunity, Infection, and Aging* (ed. Hayat, M. A.) pp. 293–303 (Academic Press, 2014).
- Akhmetov, A. et al. Single-step precision genome editing in yeast using CRISPR-Cas9. *Bio. Protoc.* **8**, e2765 (2018).
- Lobs, A. K., Engel, R., Schwartz, C., Flores, A. & Wheeldon, I. CRISPR-Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. *Biotechnol. Biofuels* **10**, 164 (2017).
- Numamoto, M., Maekawa, H. & Kaneko, Y. Efficient genome editing by CRISPR/Cas9 with a tRNA-sgRNA fusion in the methylotrophic yeast *Ogataea polymorpha*. *J. Biosci. Bioeng.* **124**, 487–492 (2017).
- Juergens, H. et al. Contribution of Complex I NADH dehydrogenase to respiratory energy coupling in glucose-grown cultures of *Ogataea parapolymorpha*. *Appl. Environ. Microbiol.* **86**, e00678–20 (2020).
- Gao, S. et al. Multiplex gene editing of the *Yarrowia lipolytica* genome using the CRISPR-Cas9 system. *J. Ind. Microbiol. Biotechnol.* **43**, 1085–1093 (2016).
- Schwartz, C. M., Hussain, M. S., Blenner, M. & Wheeldon, I. Synthetic RNA polymerase III promoters facilitate high-efficiency CRISPR-Cas9-mediated genome editing in *Yarrowia lipolytica*. *ACS Synth. Biol.* **5**, 356–359 (2016).
- Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).

34. Park, S. H. & Hahn, J. S. Development of an efficient cytosolic isobutanol production pathway in *Saccharomyces cerevisiae* by optimizing copy numbers and expression of the pathway genes based on the toxic effect of alpha-acetolactate. *Sci. Rep.* **9**, 3996 (2019).
35. Li, N., Zeng, W., Xu, S. & Zhou, J. Toward fine-tuned metabolic networks in industrial microorganisms. *Synth. Syst. Biotechnol.* **5**, 81–91 (2020).
36. Sato, H., Das, S., Singer, R. H. & Vera, M. Imaging of DNA and RNA in living eukaryotic cells to reveal spatiotemporal dynamics of gene expression. *Annu. Rev. Biochem.* **89**, 159–187 (2020).
37. Chen, R., Yang, S., Zhang, L. & Zhou, Y. J. Advanced strategies for production of natural products in yeast. *iScience* **23**, 100879 (2020).
38. Hagel, J. M. & Facchini, P. J. Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant Cell Physiol.* **54**, 647–672 (2013).
39. DeLoache, W. C. et al. An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. *Nat. Chem. Biol.* **11**, 465–471 (2015).
40. Liu, Q. et al. Rewiring carbon metabolism in yeast for high level production of aromatic chemicals. *Nat. Commun.* **10**, 4976 (2019).
41. Zhao, Y. et al. Leveraging the Hermes transposon to accelerate the development of nonconventional yeast-based microbial cell factories. *ACS Synth. Biol.* **9**, 1736–1752 (2020).
42. Jung, S. T., Lauchli, R. & Arnold, F. H. Cytochrome P450: taming a wild type enzyme. *Curr. Opin. Biotechnol.* **22**, 809–817 (2011).
43. Veith, A. & Moorthy, B. Role of cytochrome P450s in the generation and metabolism of reactive oxygen species. *Curr. Opin. Toxicol.* **7**, 44–51 (2018).
44. Hrycay, E. G. & Bandiera, S. M. Involvement of cytochrome P450 in reactive oxygen species formation and cancer. *Adv. Pharmacol.* **74**, 35–84 (2015).
45. Bendich, A. J. The end of the circle for yeast mitochondrial DNA. *Mol. Cell* **39**, 831–832 (2010).
46. Cao, M. et al. Centromeric DNA facilitates nonconventional yeast genetic engineering. *ACS Synth. Biol.* **6**, 1545–1553 (2017).
47. Cao, M., Seetharam, A. S., Severin, A. J. & Shao, Z. Rapid isolation of centromeres from *Scheffersomyces stipitis*. *ACS Synth. Biol.* **6**, 2028–2034 (2017).
48. Stanford, W. L., Cohn, J. B. & Cordes, S. P. Gene-trap mutagenesis: past, present and beyond. *Nat. Rev. Genet.* **2**, 756–768 (2001).
49. Nguyen, V. Q., Co, C. & Li, J. J. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature* **411**, 1068–1073 (2001).
50. Sopko, R. et al. Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell* **21**, 319–330 (2006).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Strains and media. *Scheffersomyces stipitis* strain FLP-UC7 (*ura3-3*, NRRL Y21448) was gifted from T. W. Jeffries (retired, University of Wisconsin and Xylome Corporation, Madison, WI, USA). *Y. lipolytica* PO1f (*MATA*, *leu2-270*, *ura3-302*, *xpr2-322*, *axp-2*) was purchased from ATCC (Manassas, VA, USA). *K. marxianus* YS402 (CBC6556 Δ ura3)³⁸ was gifted from I. Wheelodon (University of California Riverside, CA, USA). *H. polymorpha* DL1 Δ leu2 was derived from *H. polymorpha* DL1 purchased from ATCC by introduction of a zeocin cassette into the ORF of *LEU2*. *S. cerevisiae* YSG50 (*MATa*, *ade2-1*, *ade3 Δ 22*, *ura3-1*, *his3-11*, *15*, *trp1-1*, *leu2-3*, *112*, *can1-100*) was used as the host for plasmid assembly in this study. All yeast strains were cultured at 30 °C (except for *K. marxianus* YS402, which was cultured at 37 °C) in YPAD, YPAX and synthetic complete (SC) dropout media. YPAD and YPAX media contained 1% w/v yeast extract, 2% w/v peptone, 0.01% w/v adenine hemisulphate and either 2% w/v dextrose or 2% w/v xylose, respectively. SC dropout medium, used to select transformants possessing auxotrophic gene markers, consisted of 2% w/v dextrose, 0.5% w/v ammonium sulfate, 0.17% w/v yeast nitrogen base without amino acids and ammonium sulfate, and 0.083% w/v complete supplement mixture lacking the auxotrophic metabolite of interest (uracil, tryptophan or leucine). SC plus 5-fluoroorotic acid (5-FOA) medium was prepared analogously to the SC medium recipe and supplemented with 1 mg ml⁻¹ 5-FOA⁵¹, which was used for the screening of *ura3-* cells. *Escherichia coli* strain BW25141 (*lacZ* *rrnB*₁₃ Δ lacZ_{W116} Δ phoBR580 *hsdR514* Δ araBAD_{AH33} Δ rhaBAD_{LD78} *galU95* *endA*_{BT333} *uidA*(Δ MluI)::pir⁺ *recA1*) was used for yeast plasmid enrichment, and *E. coli* strains DH5 α and 10B (New England BioLabs) were used as hosts for Gibson Assembly. *E. coli* strains were cultured at 37 °C in Luria Bertani medium supplemented with 100 μ g ml⁻¹ ampicillin. Supplementary Table 2 lists the strains used in this study.

Plasmid construction. Recombinant plasmids were constructed using DNA Assembler⁵² or Gibson Assembly. The genomic DNA of yeasts needed for PCR amplification was extracted using the Wizard gDNA purification kit (Promega). Thermo Fisher Scientific FastDigest enzymes were used for restriction digestion, and PCR amplifications were performed using Q5 high-fidelity DNA polymerase (New England BioLabs). Single-stranded DNA oligos (60–90 nt) for PCR and gBlock gene fragments were synthesized by IDT (Coralville). Primers (20 bp) for PCR and sequencing were synthesized by the ISU DNA Facility. DNA was separated using gel electrophoresis and subsequently purified with the ZymoClean Gel DNA Recovery Kit (Zymo Research). Supplementary Note 2 and Supplementary Figs. 11–15 provide a detailed summary of plasmid construction and vector maps of the constructs, and Supplementary Tables 3–5 list the plasmids and primers used in this study.

For DNA Assembler, the restriction-digested backbone and PCR-amplified/gBlock fragments were cotransformed by electroporation into *S. cerevisiae* YSG50 using a voltage of 0.75 kV mm⁻¹. Transformants were plated on selective plates lacking the amino acid corresponding to the auxotrophic selection marker. Colonies appeared after 48–72 h at 30 °C. Yeast plasmids were isolated using the Zymoprep Yeast Plasmid Miniprep II Kit (Zymo Research) and transformed into *E. coli* BW25141 for enrichment and subsequent downstream applications. For Gibson Assembly, the restriction-digested backbone and insert were assembled using Gibson Assembly Master Mix (New England BioLabs) according to the manufacturer's protocol, with 4 μ l of reaction mixture used for *E. coli* transformation. *E. coli* plasmids were isolated using the QIAprep Spin Plasmid Mini-prep Kit (Qiagen). The constructs were verified by restriction digestion and/or Sanger sequencing (ISU DNA facility).

Yeast strain culture and transformation. All yeast strains were cultured at 30 °C (except for *K. marxianus*, which was cultured at 37 °C) with shaking at 250 r.p.m. All yeast strains in this study were transformed by electroporation as previously described⁵². A voltage of 1.25 V mm⁻¹ was used for all yeast strains, except for *S. cerevisiae*, where 0.75 V mm⁻¹ was used.

Homologous recombination activity assay in yeast. For each yeast strain assayed, each of the three biological replicates was cotransformed with two fragments: (1) 200 ng of a double-digested backbone (disrupting the ORF of *GFP*) treated with alkaline phosphatase to prevent ligation of the compatible ends; and (2) 200 ng of a *GFP*-restoring fragment containing 30-bp homologies flanking the digestion site, which was PCR amplified from the analogous plasmid. For *S. cerevisiae* YSG50, pSC-HR-*GFP* was digested with Eco72I; for *S. stipitis* UC7, pSS-HR-*GFP* was digested with NcoI; for *Y. lipolytica* PO1f, pYL-HR-*GFP* was digested with NcoI; and for *K. marxianus* YS402, pKM-*GFP*-HR was digested with EcoRI and XhoI. Transformed cells were spread on two separate SC-URA plates to prevent lawn formation. After 72 h of incubation, GFP-expressing and GFP-null colonies were counted visually on each plate using a DR46B transilluminator (Clare Chemical Research). To confirm the accuracy of visual discernment between GFP-expressing and GFP-null colonies in the counting procedure, plasmids from five GFP-expressing and five GFP-null colonies (Supplementary Fig. 16) were isolated, transformed and enriched in *E. coli* and sequenced. For *S. cerevisiae*, *S. stipitis* and *Y. lipolytica*, sequencing results confirmed that the five GFP-expressing colonies integrated the restoration fragment via homologous recombination and the five GFP-null colonies repaired the plasmid by error-prone/random mechanisms.

For *K. marxianus*, fluorescence intensity was too weak to differentiate between GFP-expressing and GFP-null colonies by visual inspection on the transilluminator. Alternatively, 100 colonies for each biological replicate were randomly selected and cultured in 2 ml of SC-URA for 48 h for subsequent flow cytometry analysis. Plasmids from five GFP-expressing and five GFP-null colonies were isolated, transformed and enriched in *E. coli* and sequenced. Sequencing results confirmed that the five GFP-null colonies had repaired the plasmid via error-prone/random mechanisms. However, sequencing of the five GFP-expressing colonies revealed that only three had integrated the restoration fragment by homologous recombination. To account for this discrepancy, a factor of 0.6 was applied to the percentage of GFP-expressing clones revealed by flow cytometry when determining the percentage of *K. marxianus* clones that had integrated the restoration fragment via homologous recombination.

Flow cytometry analysis for quantification of GFP and BFP expression. Cell cultures were diluted tenfold with 10 mM PBS (pH 7.4) and analyzed for GFP or BFP fluorescence intensity at 488 or 405 nm, respectively, using a FACSCanto flow cytometer (BD Biosciences). The fluorescence intensity distribution of GFP- and BFP-expressing cells was calculated using BD FACSDiva 8.0.1. Data for forward scatter (FSC, photodiode 488/10 bandpass filter) and side scatter (SSC, photomultiplier 488/10 bandpass filter) were collected in logarithmic mode using instrument voltage settings typical for yeast cell populations. GFP fluorescence was excited using a 488-nm laser and emission collected through a 525/50 bandpass filter; BFP fluorescence was excited using a 405-nm laser and emission was collected through a 450/50 bandpass filter. Fluorescence intensities were quantitatively measured in logarithmic mode, and values were recorded as mean fluorescence intensity for each event. Initial gating to identify yeast cells was based on FSC and SSC data analyzed in dual-parameter dot plots, using data from control samples containing resuspension buffer only (that is, no yeast cells) to distinguish between background/noise and actual yeast cell events. Most background/noise events were eliminated by setting a SSC acquisition threshold. Gated yeast cell fluorescence data were analyzed in dot plots of SSC versus the relevant fluorescent parameter, either GFP (Supplementary Fig. 17) or BFP (Supplementary Fig. 18). Distinctions between 'positive' and 'negative' fluorescence were determined by comparing data from untransformed and transformed yeast; data from untransformed cells were gated as negative. Using the established negative gate as a lower threshold, positive fluorescence in transformed yeast was defined as all events outside the negative gate measured as being progressively higher on the relevant fluorescent scale.

Growth characterization of *S. stipitis* strains UC7 and Δ ku70 Δ ku80. Seed cultures of *S. stipitis* strains UC7 and Δ ku70 Δ ku80 were grown in either 1 ml of YPAD (strains harboring no plasmid) or 1 ml of SC-URA (strains harboring a *URA3*-expressing plasmid) for 24 h to obtain a sufficient density of cells, whilst avoiding cellular senescence to minimize lag time effects (the evident linear trends from the semilogarithmic transformations beginning at the initial time point confirmed negligible lag phases³³). Seed cultures were used to inoculate fresh cultures to a starting OD₆₀₀ of ~0.2, which were then cultured in sterile 96-well plates containing 200 μ l of either YPAD or SC-URA at 30 °C and continuous shaking at 205 r.p.m. OD₆₀₀ was measured every 15 min using a Synergy Eon Microplate Spectrophotometer (BioTek) for 72 h.

Comparison of shikimate production by *S. stipitis* UC7 and Δ ku70 Δ ku80. Seed cultures of *S. stipitis* strains UC7 and Δ ku70 Δ ku80 expressing the shikimate pathway were grown in 3 ml of SC-URA for 36–40 h. Cultures were diluted into 55-ml rimless culture tubes carrying 6 ml of SC-URA supplemented with 4% w/v glucose and cultured at 30 °C with shaking at 250 r.p.m. for 120 h. Next, 200 μ l of culture was pelleted and the supernatant diluted 100-fold with nanopure water. Shikimate production was quantified as previously described⁴¹, then 50 μ l of a solution consisting of 0.5% w/v periodate and 0.5% w/v sodium m-periodate was added to 100 μ l of diluted supernatant and incubated at 37 °C for 45 min. The oxidation reaction was quenched with 100 μ l of a solution consisting of a 3:2 v/v ratio of 1 M NaOH and 56 mM Na₂SO₃. Absorbance was measured at 382 nm using a Synergy HTX multimode reader. Authentic shikimate (Sigma-Aldrich) was used to construct a standard curve (5–40 mg l⁻¹) to correlate absorbance to shikimate titer.

hGem tag characterization in *S. stipitis* via BFP reporting. *BFP-URA3* and *BFP-hGem-URA3* cassettes were constructed via overlap extension PCR. For *BFP-URA3*, a *BFP* expression cassette and a *URA3* expression cassette were PCR amplified from pBFP and pARS/CEN5-500bp-*GFP*, respectively. Amplification of the *BFP* cassette was split into two PCR reactions to append a nuclear localization sequence to the C terminus of the *BFP* ORF as specified by the PCR primers. The three individual PCR fragments were fused via overlap extension PCR. To construct *BFP-hGem-URA3*, an analogous workflow was followed, with the addition of a fourth fragment, the *hGem* tag, which was PCR amplified from pCashGem-Xyl2.

With no reliable means of precision genome editing at this initial stage, we leveraged NHEJ-mediated random integration of *BFP-URA3* and *BFP-hGem-URA3* cassettes to construct individual expression libraries. UC7 was

transformed with 200 ng of each fragment in parallel, plated on SC-URA and incubated for 72 h. To minimize the influence of different integration loci on *BFP* expression, we performed throughput flow cytometry screening to identify a stable, high-*BFP*-expressing mutant from each library. Colonies from each transformation were cultured in 1 ml of YPAD for 24 h and submitted for BFP flow cytometry analysis to identify the highest genome-based *BFP* expression strains, *BFP* no. 15 and *BFP-hGem* no. 28 (Supplementary Fig. 2a).

Intriguingly, during this throughput flow cytometry screening, we observed two candidates in the *BFP-hGem* library that exhibited bimodal expression patterns (Supplementary Fig. 2b). These population distributions mirrored those observed in cell cycle analysis experiments, whereby fluorescent labeling of DNA yields two peaks: one corresponding to cells in G1 (1 C DNA content) and the other corresponding to cells in G2/M (2 C DNA content)³⁴. This observation provided a preliminary support for the temporal control we sought to establish. To investigate hGem-mediated *BFP* temporal degradation characteristics, 25 colonies of each of the two identified high-*BFP*-expressing clones (that is, *BFP* no. 15 and *BFP-hGem* no. 28) were cultured for 96 h in 2 ml of YPAD, with OD₆₀₀ measurement and sampling for BFP flow cytometry conducted every 24 h.

Characterization of aberrant recombination events in *S. stipitis*. pGAOU (Extended Data Fig. 3b) was digested with *SacI* and *XhoI* to obtain a linear cassette containing *E. coli Amp* and *Ori* elements flanked by *S. stipitis GFP* and *URA3* cassettes (GAOU; the acronyms used in naming of a plasmid or a fragment correspond to the first letter of each encoding element). Next, 1.2 µg of the digested GAOU fragment was transformed into UC7, plated on SC-URA and incubated for 72 h. Colonies expressing GFP were identified using a DR46B transilluminator. Ten strongly visibly intense green colonies (Extended Data Fig. 3c) were cultured in 3 ml of SC-URA for 36 h and submitted for GFP flow cytometry analysis; 70% of colonies (7/10) showed broad peaks (Extended Data Fig. 3d), suggesting *GFP*-episomal expression. Five of these strains (nos. 1–4 and 8) exhibiting broad peaks were cultured in 3 ml of SC-URA medium to isolate the pseudoplasids. The five isolated yeast plasmids were each transformed into *E. coli* strain BW25141 for plasmid enrichment and functional verification (Extended Data Fig. 4). Two colonies from each LB + Amp agar plate were randomly selected for plasmid isolation and subsequent enzyme digestion verification. The isolated plasmids were digested with *NdeI* and *EcoRI*, where a two-fragment pattern consisting of a 4,038-bp band and a 1,088-bp band corresponds to the product of *in vivo* self-ligation of the digested GAOU fragment (Extended Data Fig. 3b). Three of the five broad-peak groups (nos. 1–3) contained larger plasmids, with some unknown sequences inserted (the two technical replicates were consistent among each of these plasmids; Extended Data Fig. 4b). The two relatively small plasmids from nos. 2 and 3 were selected for DNA sequencing to identify the unknown inserted sequences. After blasting against the genome sequence of *S. stipitis*, we determined that the inserts from nos. 2 and 3 originated completely from the genome of *S. stipitis*. The insert from no. 2 corresponded to the 5,427-bp region from Chr1:1825584–1831010 whereas that from no. 3 corresponded to the 6,013-bp region from Chr3:781885–787897 (Extended Data Fig. 4c). To confirm the propagative function enabled by the inserted sequences, plasmid nos. 2 and 3 were retransformed into UC7. Many colonies appeared on both plates and exhibited uniform green fluorescence (Extended Data Fig. 4d), indicating that the inserted sequences of nos. 2 and 3 contained an ARS sequence that conferred propagation of these plasmids in *S. stipitis*. The digested GAOU fragment was also transformed to strain $\Delta ku70\Delta ku80$. No colonies were obtained, confirming that the aberrant recombination was attributable to NHEJ (Supplementary Fig. 5).

LINEAR CRISPR editing in yeast. CRISPR guide sequences were designed using the CHOPCHOP online tool⁵⁵. LINEAR cassettes were obtained via restriction digestion of their associated plasmids. Yeast strains were transformed with 200–400 ng of a gel-purified fragment, plated on selective medium corresponding to the selection marker on the LINEAR fragment and incubated until sizable colonies were obtained (~48 h). Colonies were randomly selected and plated on YPAD to isolate a pure genotype, filtering out false positives associated with extrachromosomal-based selection marker expression. After 36 h of incubation, single colonies from the YPAD plate were cultured in liquid YPAD for 24 h for gDNA isolation. Loci of interest were PCR amplified and gels visualized using Image Lab 5.2.1 (BioRad). The gels used for determination of the reported genome-editing efficiencies are summarized in Supplementary Figs. 6–9, with unprocessed gel images detailed in Supplementary Figs. 19–22. Purified PCR products were sequenced using an Applied Biosystems 3730xl DNA Analyzer (Thermo Fisher Scientific) to confirm the desired genome modifications. Primers used for verification are listed in Supplementary Table 4.

Selection marker curation for iterative rounds of genome editing. Recovery of the *URA3* selection marker to permit subsequent genome edits in *S. stipitis* via LINEAR or NHEJ leveraged Cre recombination, adapted from a previous procedure⁴¹. Strains were transformed with plasmid pJML545 (ref. ⁵⁶), plated on YPAD plus zeocin and incubated for 72 h. Five transformants were inoculated into 1 ml of YPAX medium and cultured for 24 h to express Cre recombinase driven by the *XYL1* promoter. Then, 50 µl of each culture was spread on a SC + URA + 5-FOA

plate and incubated for 96–120 h to isolate a *ura3*[−] genotype. One colony from each streak was inoculated into 1 ml of YPAD and cultured for 24 h to isolate gDNA for confirmation of the desired recombination and removal of *URA3* via PCR.

Construction of the NHEJ-mediated expression library. pBSA4-*URA3-loxP* and pGTDN-*URA3-loxP* were digested with *NotI* and *SgsI* to liberate the upstream *BSA4UloxP* and downstream *GTDNUloxP* expression cassettes, respectively. To construct the upstream module expression library, *S. stipitis* $\Delta ara10:TD^*loxP$ was transformed with 6 µg of the linear *BSA4UloxP* fragment, plated on SC-URA and incubated for 120 h; 100 colonies were randomly selected and cultured in 1 ml of YPAD for 24 h. Cultures were sent for flow cytometry analysis to identify genomic *BFP*-expressing *Ura*⁺ mutants. To construct the downstream module expression library, *S. stipitis* *BSA4 loxP* 41 was transformed with 3 µg of the linear *GTDNUloxP* fragment, plated on SC-URA and incubated for 120 h. With GFP as the fluorescent reporter, 100 GFP⁺ colonies were identified using a DR46B transilluminator (Clare Chemical Research) and cultured in 1 ml of YPAD for 24 h. Cultures were sent for flow cytometry analysis to identify genomic *GFP*-expressing *Ura*⁺ mutants.

Quantification of betaxanthin production via biosensing. Screening for high-producing betaxanthin mutants in the upstream module library was conducted as previously described⁴¹. Mutants exhibiting genomic *BFP* expression profiles from flow cytometry analyses were inoculated in 1.5 ml of SC-URA medium with an initial OD₆₀₀ of ~0.1 and grown for 24 h. The OD₆₀₀ of the cultures was measured to normalize the fluorescence intensity of each. In parallel, 200 µl of each culture was centrifuged and cell pellets were resuspended in 200 µl of 10 mM PBS (pH 7.4). The cells were pipetted onto a 96-well, black polystyrene plate and their fluorescence intensity measured with a Synergy HTX multimode reader (excitation, 485/20 nm; emission, 516/20 nm; gain, 50)³⁹. Fold change over background fluorescence was calculated by normalizing the signal to the average fluorescence of the control parent strain, *S. stipitis* $\Delta ara10:TD^*loxP$.

Quantification of (S)-norcoclaurine production via LC–MS analysis. Screening for high-producing (S)-norcoclaurine mutants in the downstream module library was conducted as previously described⁴¹. Briefly, mutants exhibiting genomic *GFP* expression profiles from flow cytometry analyses were inoculated in 3 ml of SC-URA for 48 h. Saturated seed cultures were used to inoculate 6 ml of 2× concentrated SC-URA (without glucose), supplemented with 6% glucose and 0.5% L-ascorbic acid (Sigma-Aldrich) to mitigate oxidation of (S)-norcoclaurine. After 120 h of fermentation, 1 ml of each culture was pelleted and the supernatant was filtered for subsequent liquid chromatography–tandem mass spectrometry analysis⁴¹.

Nanopore sequencing of (S)-norcoclaurine-producing strains. *Sample and library preparation.* Seed cultures were inoculated into 20 ml of YPAD and grown to OD₆₀₀ ~ 1.0. High-molecular-weight DNA was isolated using Nanobind CBB Big DNA and Tissue Big DNA kits (Circulomics) according to the manufacturer's extraction protocol (*S. cerevisiae* High Molecular Weight DNA Extraction Protocol v.0.5a). The presequencing mix was quantified by Qubit (Qubit dsDNA BR assay Kit). Each flow cell (no. FLO-MIN106, Oxford Nanopore Technologies) was primed according to the manufacturer's guidelines (kit nos. SQK-LSK109, EXP-NBD114). The flow cell was mounted on a GridIONx5 (Oxford Nanopore Technologies) for sequencing with the MinNOW v.20.10.6 and Guppy 4.2.3 script.

Genome assembly and elucidation of NHEJ-mediated pathway insertions.

A detailed methodology is available on Zenodo: <https://doi.org/10.5281/zenodo.5544230>. Briefly, the nextflow⁵⁷ workflow, NanoQCtrm, was used for quality control and trimming of the reads. This workflow implements NanoPlot⁵⁸ (v.1.32.0) for quality control and trims adapters using Downpore (v.0.3.3). Flye⁵⁹ (v.2.8.2-b1691) was used for de novo assembly of each of the sequenced genomes. Minimap2 (ref. ⁶⁰) (v.2.2-r409) was used for all alignments performed in this study. The inserted pathway module sequence was aligned to each assembled genome. The 10,000-bp sequence upstream of each aligned insert to the assembled genome was extracted and then aligned to the reference genome (no. ASM20916v1) to identify the insertion location in the reference genome.

The desktop app, re-DOT-able (Babraham Bioinformatics), was used to visualize pairwise comparisons between sequences and generate interactive plots. Pairwise comparisons between the reference genome and each of the assembled genomes were conducted to identify the insertion locations of the pathway modules (cross-validated with the alignment method) and to determine whether a genome had undergone large-scale rearrangements. To verify the insertion locations, each assembly was also plotted against the relevant insert sequences.

Statistics. Statistical analyses and construction of the resultant graphical figures were performed using GraphPad Prism v.9.2.0 for Windows (GraphPad Software). Results are reported as means ± s.d., with individual data points overlaid. The nature of the statistical tests employed is stated in the individual figure legends. Statistical significance was determined using a threshold value of *P* < 0.05.

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

Data availability

Source data are provided for the results presented in Main and Extended Data figures. Vectors containing LINEAR constructs for *Y. lipolytica*, *K. marxianus*, *H. polymorpha* and *S. stipitis* are available from Addgene (plasmid nos. 174837–174840). Raw reads and assembled genome sequences for the final engineered norcoclaurine-producing strains have been deposited in NCBI under BioProject ID PRJNA753835. Additional data are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Code availability

All code used for processing and assembly of genome sequencing reads (along with an accompanying method for how the code was implemented) is available on Zenodo: <https://doi.org/10.5281/zenodo.5544230>.

References

51. Boeke, J. D., Trueheart, J., Natsoulis, G. & Fink, G. R. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol.* **154**, 164–175 (1987).
52. Shao, Z., Zhao, H. & Zhao, H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* **37**, e16 (2009).
53. Herricks, T. et al. One-cell doubling evaluation by living arrays of yeast, ODELAY! *G3 (Bethesda)* **7**, 279–288 (2017).
54. Calvert, M. E., Lannigan, J. A. & Pemberton, L. F. Optimization of yeast cell cycle analysis and morphological characterization by multispectral imaging flow cytometry. *Cytometry A* **73**, 825–833 (2008).
55. Labun, K. et al. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* **47**, W171–W174 (2019).
56. Laplaza, B. R. T. JoseM., Jin, Yong-Su & Jeffries, ThomasW. Sh ble and Cre adapted for functional genomics and metabolic engineering of *Pichia stipitis*. *Enzyme Microb. Technol.* **38**, 741–747 (2006).
57. Di Tommaso, P. et al. Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **35**, 316–319 (2017).
58. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).
59. Kolmogorov, M. et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat. Methods* **17**, 1103–1110 (2020).
60. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

Acknowledgements

This work was supported by the National Science Foundation Grant (no. 1716837 to D.P., Y.Z., M.C., S.G., C.L., M.S., S.C., A.S., L.H. and Z.S.), the NSF Graduate Research Fellowship Program (to D.P.) and Griswold Internship (to M.G.). We thank T. W. Jeffries (professor emeritus, University of Wisconsin-Madison and the Founder of Xylome Corporation) for sharing *S. stipitis* FLP-UC7 (*ura3-3*, NRRL Y21448) and the plasmid pJML545. We thank I. Wheeldon, University of California at Riverside, for sharing *K. marxianus* YS402 (CBC6556 Δ ura3); J. Dueber, University of California at Berkeley for sharing the *S. cerevisiae* plasmid containing *TyrH** and *DOD*; and S. Yang, Chinese Academy of Science, for sharing the *Y. lipolytica* plasmids pCas1-yl and pCAS2-yl. We also thank S. M. Rigby for flow cytometry and L. Showman and K. Narayanaswamy at the W. M. Keck Metabolomics Research Laboratory of ISU for the detection of (S)-norcoclaurine. Finally, we thank T. Murtha at the ISU DNA facility for preparing gDNA samples for Nanopore sequencing.

Author contributions

D.P. and Z.S. conceptualized the initial idea, designed the experiments, performed troubleshooting and wrote the manuscript. Y.Z. contributed to aberrant recombination characterization and hGem tag characterization in *S. stipitis*. M.C. contributed to the initial design of the hGem tag. S.G. contributed to *Y. lipolytica* *PEX10* deletion. C.L. contributed to the construction of *H. polymorpha* DL1 Δ leu2. M.S., S.C., and A.S. contributed to genome sequence analysis. M.G. and L.H. aided in cloning efforts and LINEAR CRISPR editing efficiency determination.

Competing interests

The authors declare no competing interests.

Additional information

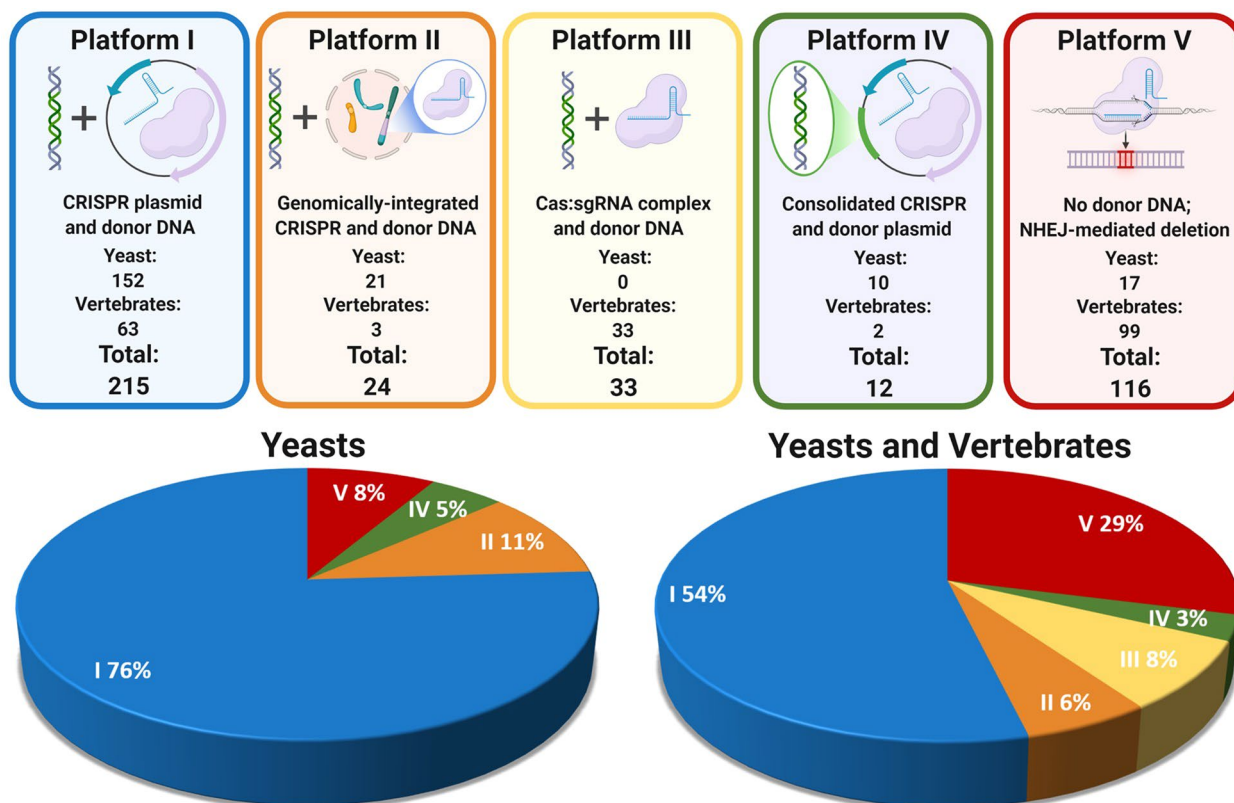
Extended data is available for this paper at <https://doi.org/10.1038/s41589-021-00893-5>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41589-021-00893-5>.

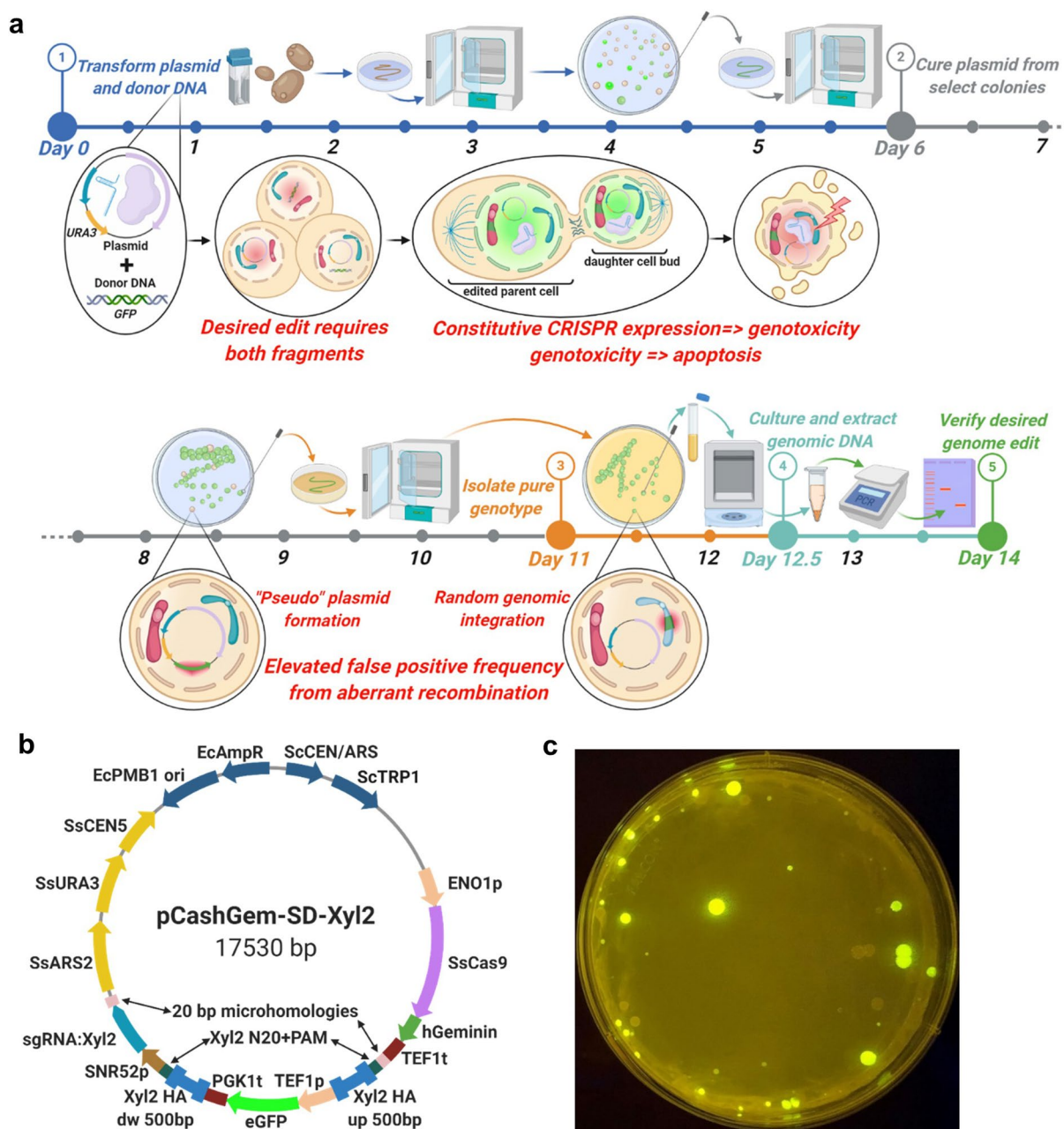
Correspondence and requests for materials should be addressed to Zengyi Shao.

Peer review information *Nature Chemical Biology* thanks Zihui Liu, Eric Young and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

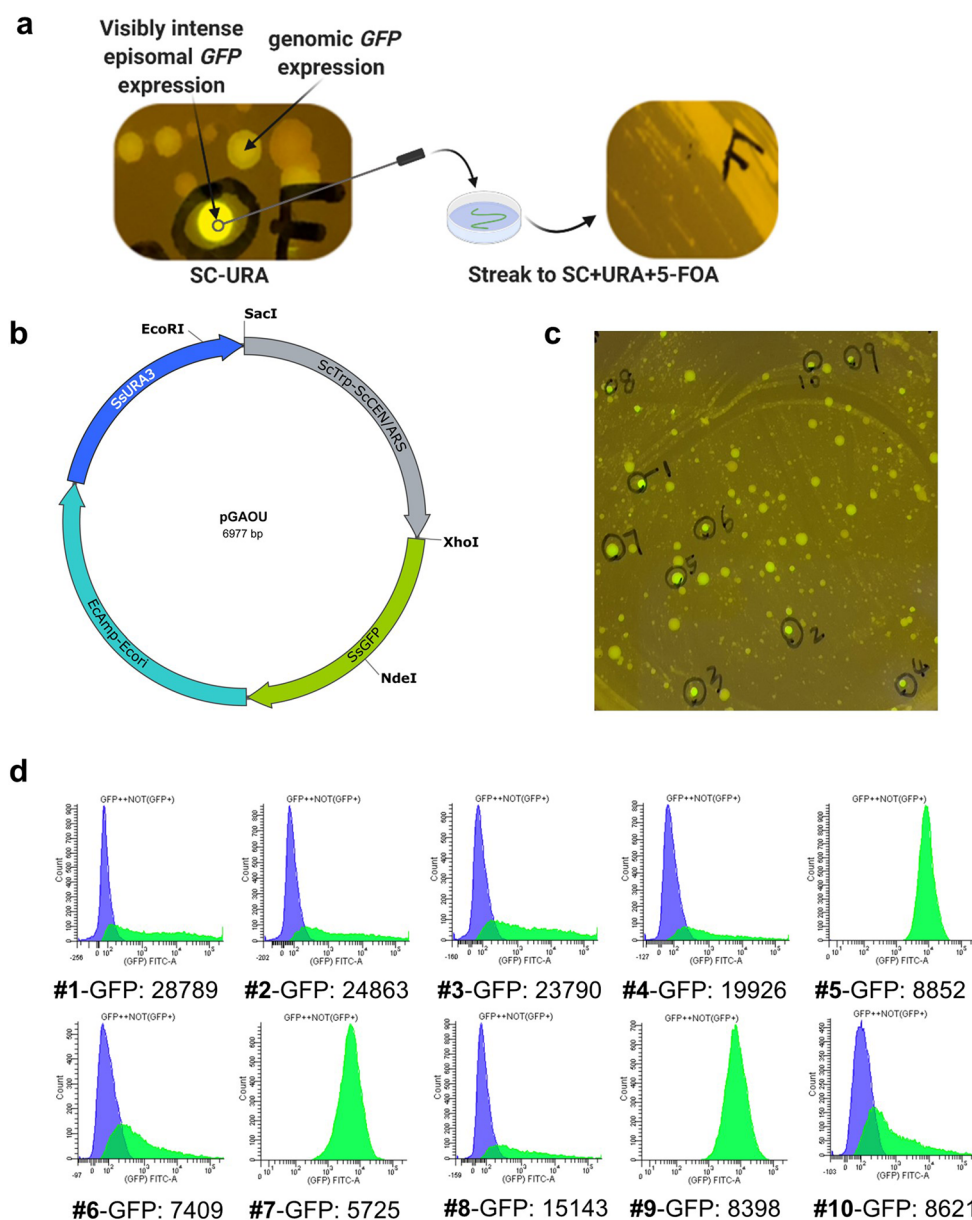
Reprints and permissions information is available at www.nature.com/reprints.



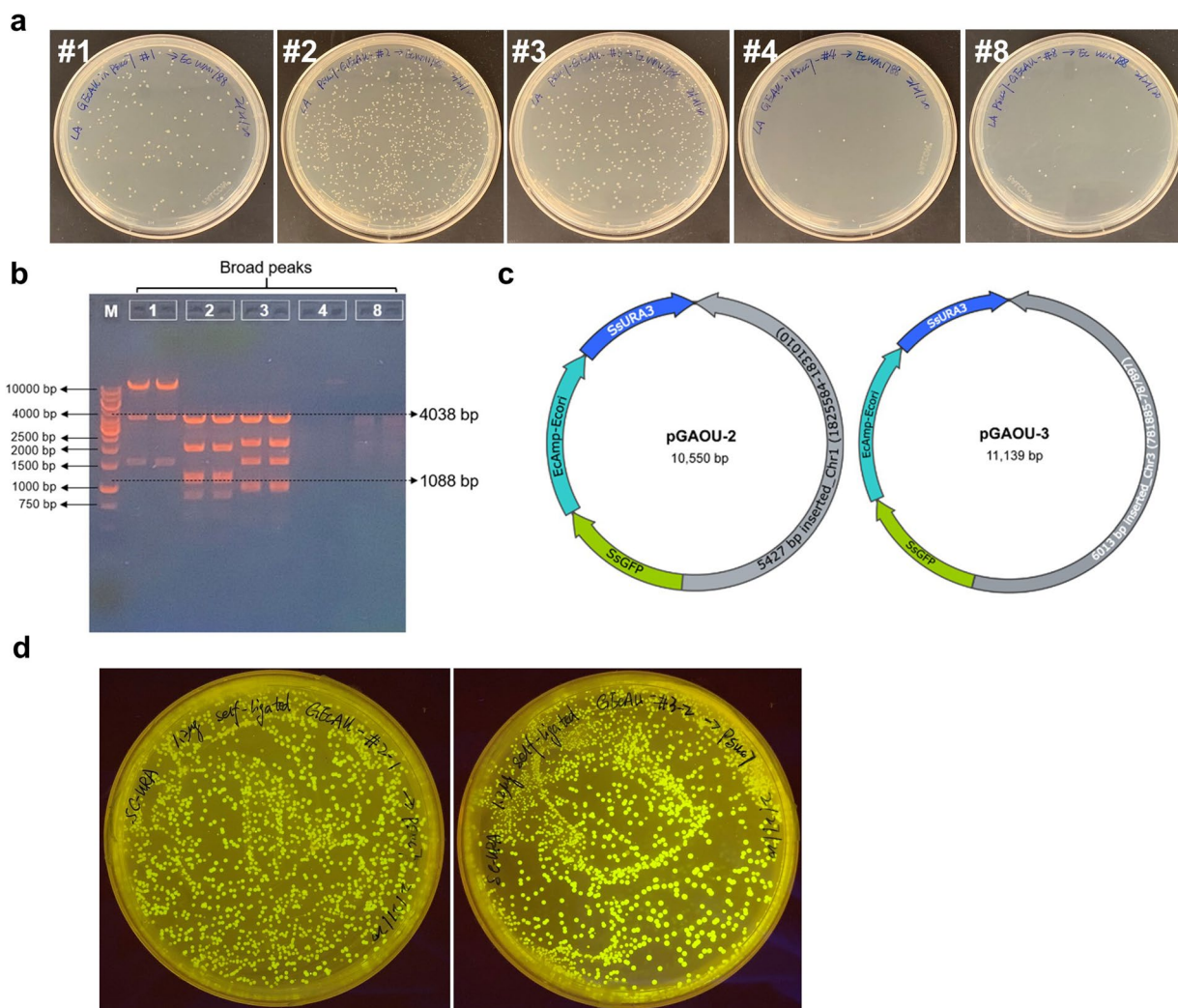
Extended Data Fig. 1 | Literature survey to determine the most implemented CRISPR platform. A compilation of 400 instances where studies deployed CRISPR for genome editing revealed that most researchers implement the conventional CRISPR plasmid and donor DNA Platform I. Studies were identified on PubMed using 'Yeast CRISPR' and 'Mammal CRISPR' in the search query.



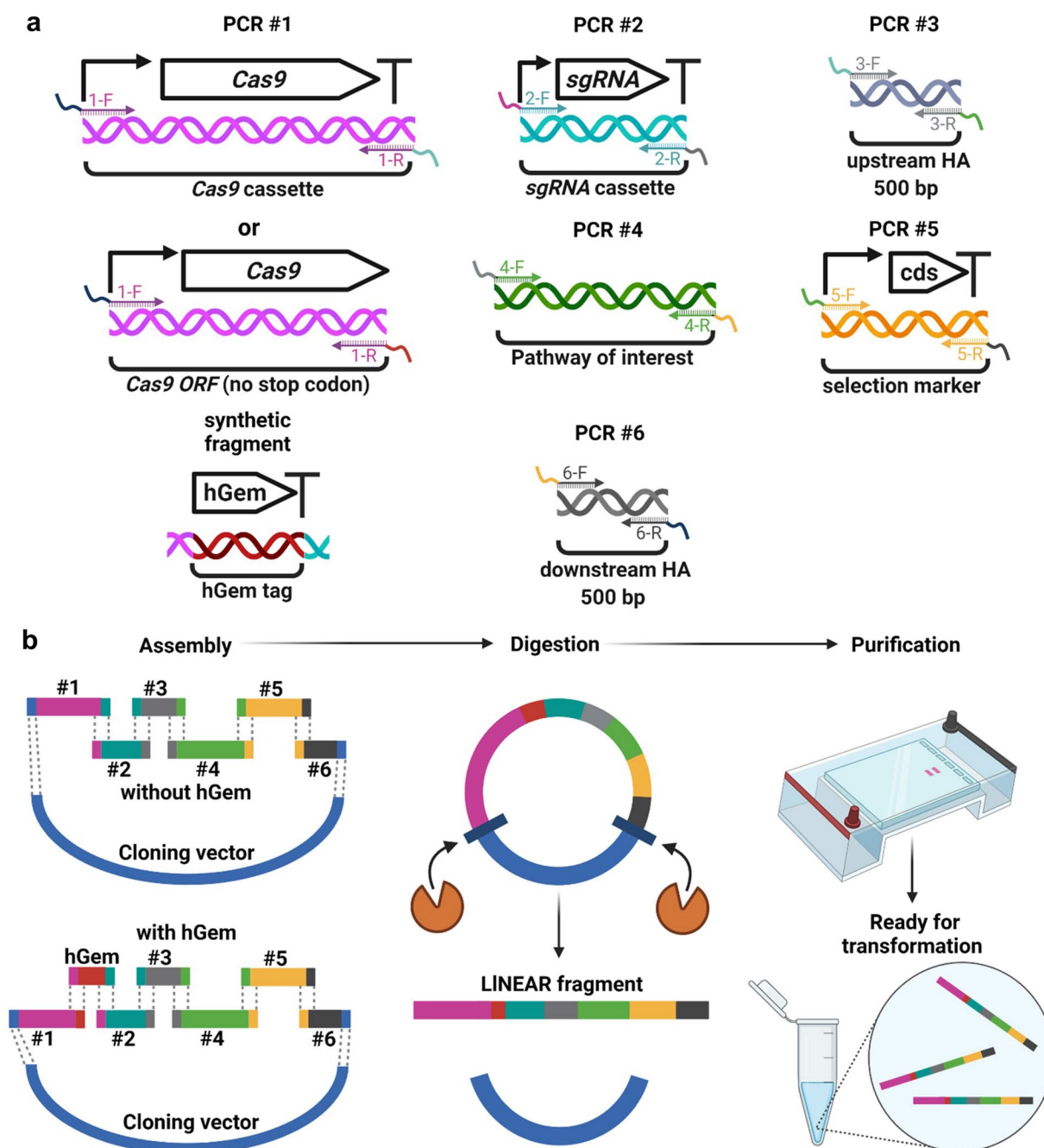
Extended Data Fig. 2 | The shortcomings associated with the conventional CRISPR platforms in *S. stipitis*. **a**, The workflow for the conventional CRISPR plasmid and donor DNA platform (Platform I in Extended Data Fig. 1) deployed in *S. stipitis*, depicting various shortcomings our group has encountered that impeded obtaining correctly edited cells. **b**, The vector map of the consolidated 'self-digesting' plasmid, pCashGem-SD-Xyl2, highlighting the key design components (Platform IV in Extended Data Fig. 1). **c**, Transformation of UC7 with the pCashGem-SD-Xyl2 plasmid failed to address the shortcomings, with almost all the colonies displaying an episomal GFP expression profile.



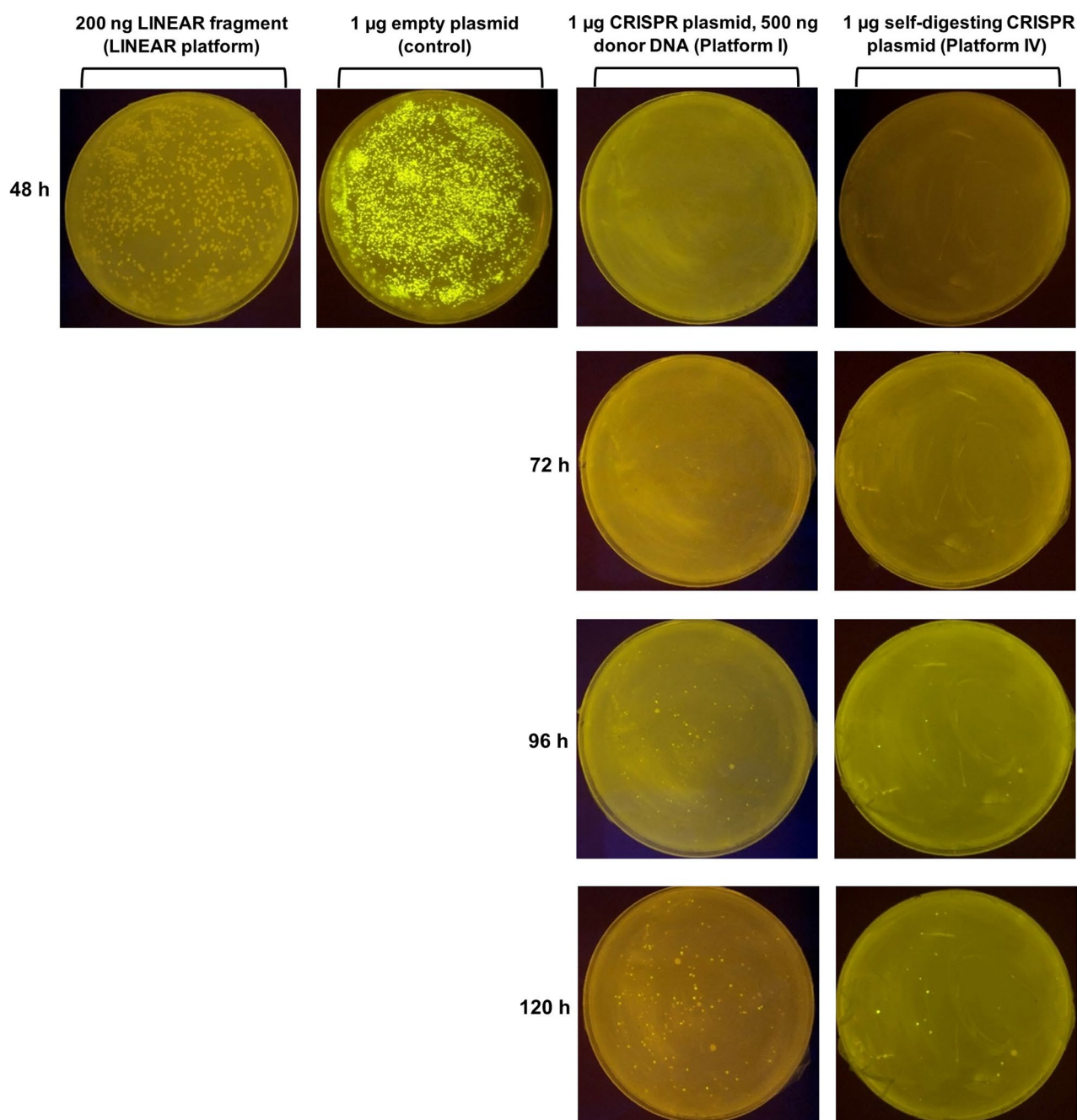
Extended Data Fig. 3 | Aberrant recombination characterization in *S. stipitis*. **a**, Visual depiction of visibly intense GFP expression and subsequent loss of expression **b**, Vector map of the pGAOU plasmid, harboring the GAOU fragment flanked by *SacI* and *XhoI* restriction sites. The digested GAOU fragment was transformed into UC7. **c**, Ten strongly intense green colonies observed under the DR46B transilluminator. **d**, The GFP expression intensity of the ten variants was evaluated by flow cytometry after 36 h of cultivation. Numbers correspond to the mean GFP expression intensity. Among these ten variants, only #5, #7, and #9 exhibited genomic expression of *GFP* (sharp peaks). The gating methodology was summarized in Supplementary Information.



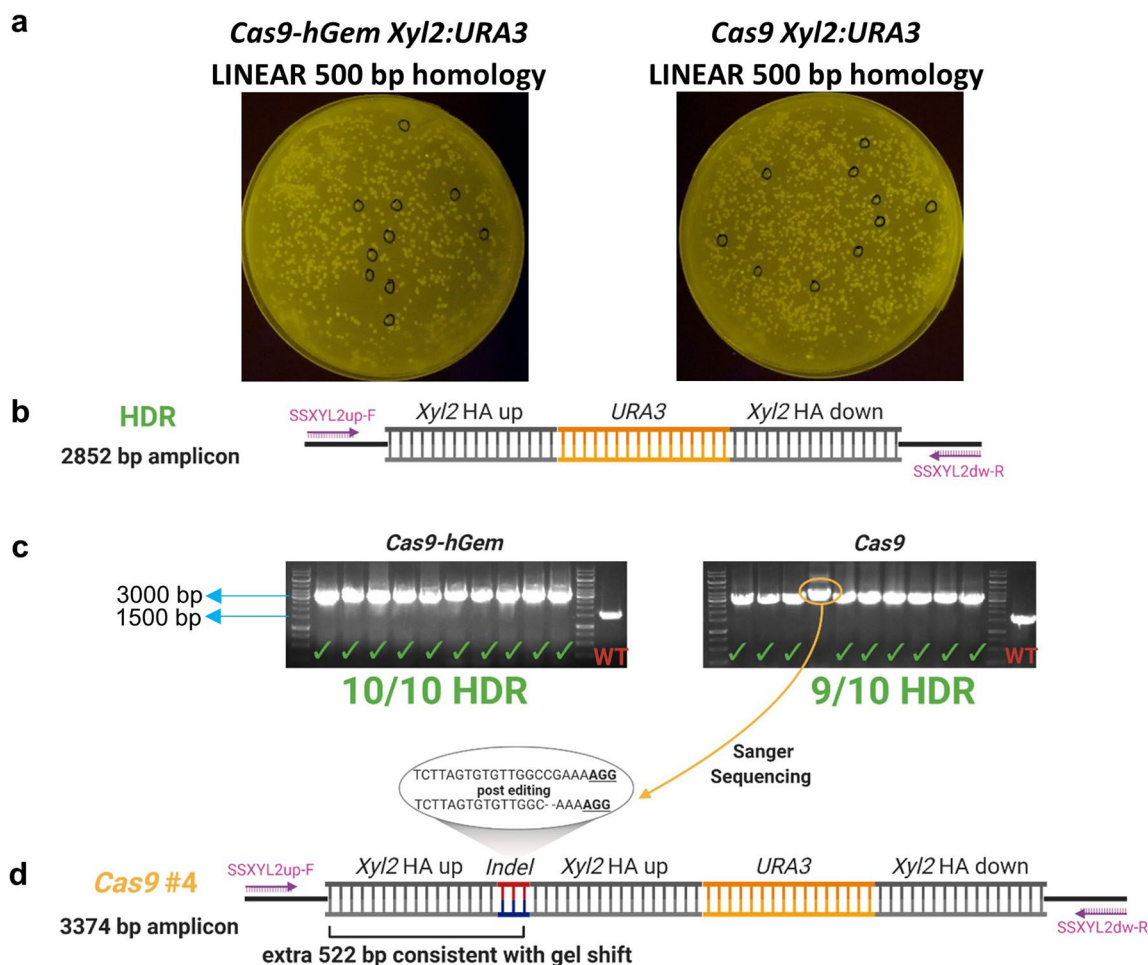
Extended Data Fig. 4 | Verification of 'pseudo' plasmids formed by aberrant recombination events in *S. stipitis*. **a**, Transformation of *E. coli* BW25141 with the five plasmids isolated from *S. stipitis* variants. **b**, Enzyme digestion verification of pGAOU-X vectors isolated from *E. coli* DH5 α . Lane 1: M, GeneRuler 1kb DNA Ladder. Lanes 2 and 3: pGAOU-1 isolates. Lanes 4 and 5: pGAOU-2 isolates. Lanes 6 and 7: pGAOU-3 isolates. Individual restriction digestions for each biological sample were performed once. **c**, Vector maps of the isolated pGAOU-X plasmids. **d**, Transformation of UC7 with pGAOU-2 and pGAOU-3. Plates were observed under a DR46B transilluminator.



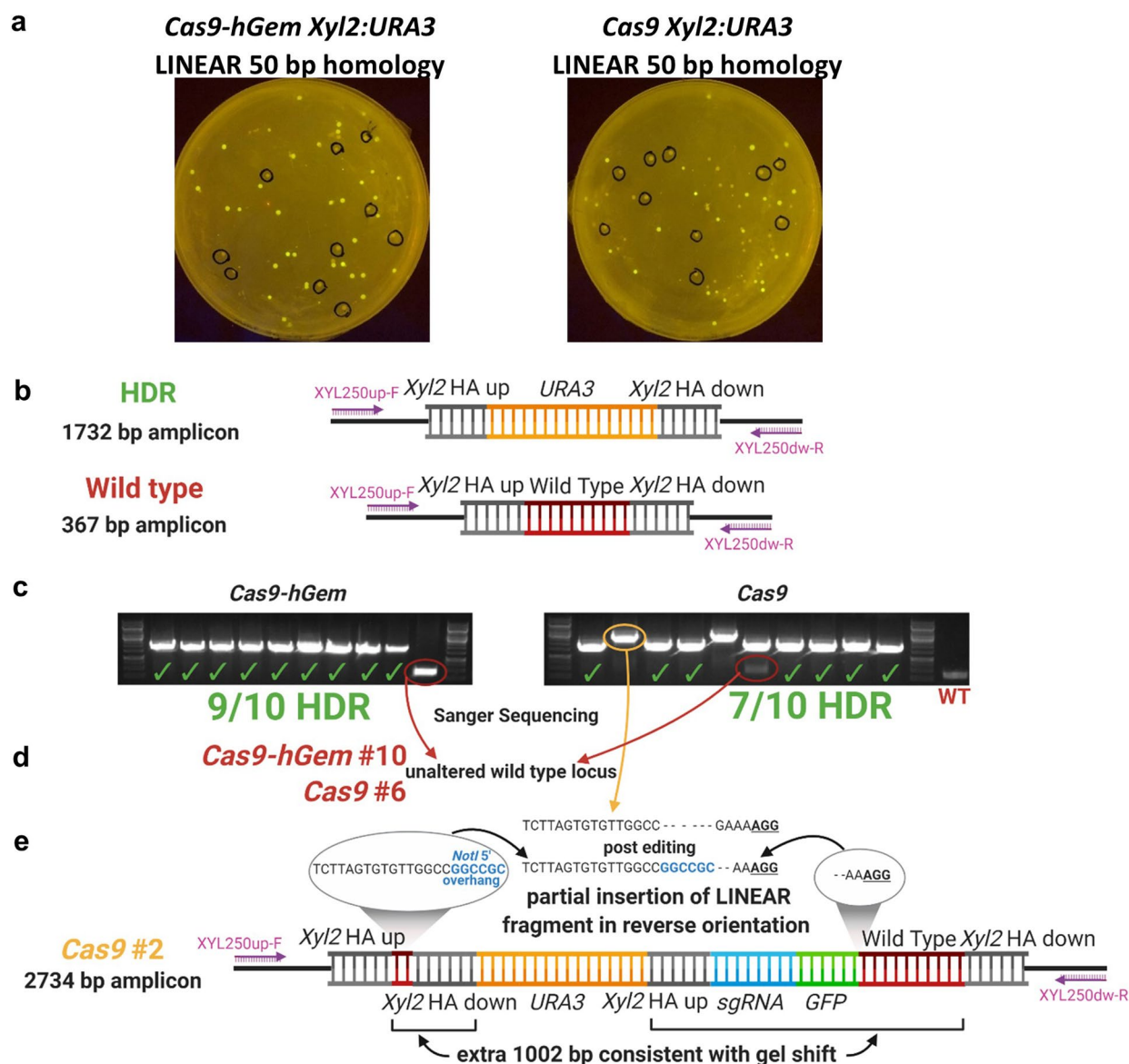
Extended Data Fig. 5 | Design protocol for the LINEAR platform. a, Cloning of individual LINEAR components. **b**, Assembly of LINEAR components into a cloning vector. Digestion and gel electrophoresis of the assembled vector yields the LINEAR fragment used in genome editing.



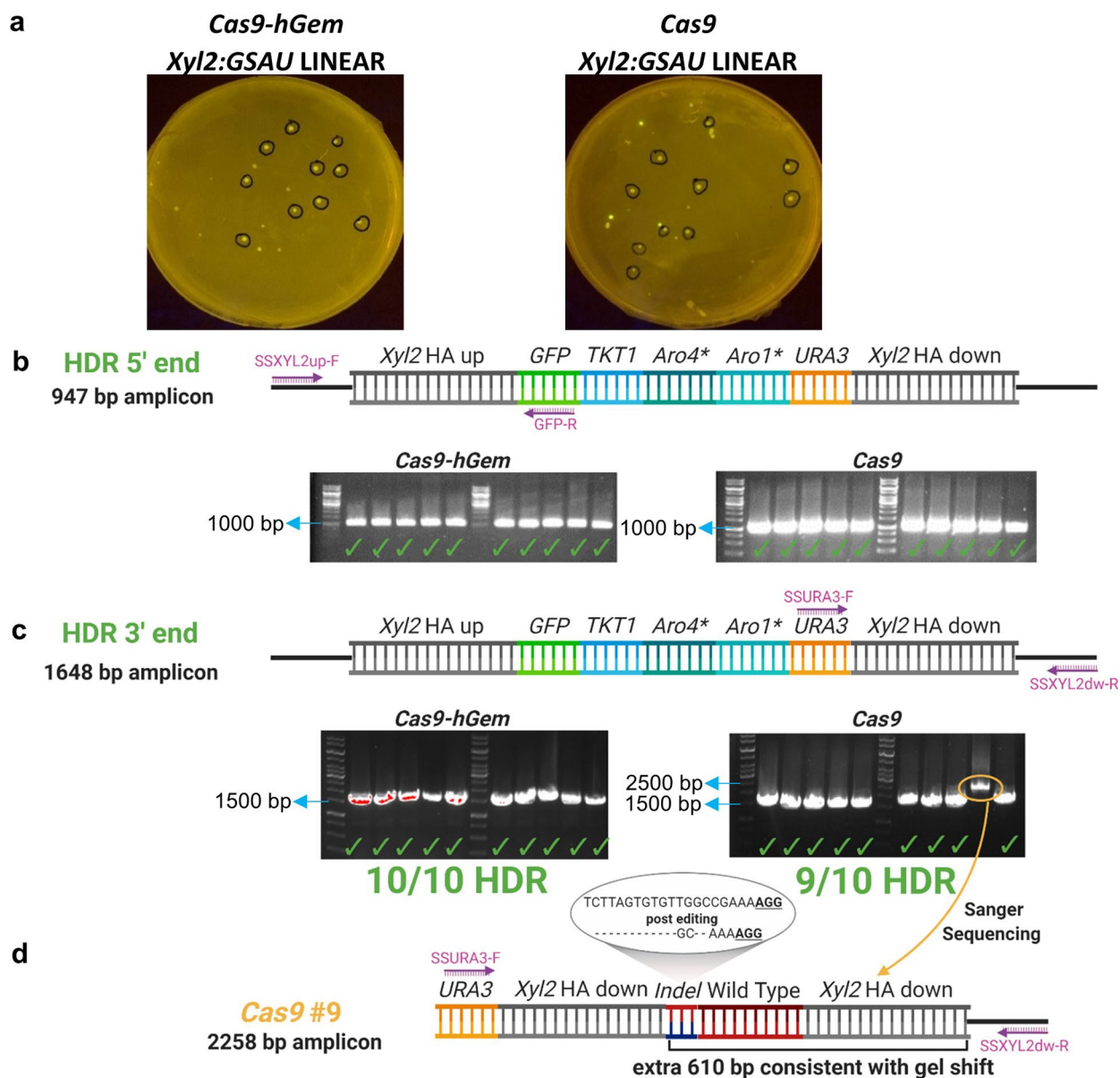
Extended Data Fig. 6 | Comparison of the conventional CRISPR platforms versus LINEAR highlights growth deficiency associated with constitutive Cas9 expression. All platforms utilized hGem tagged-Cas9. Plate pictures were taken at various timepoints post-transformation.



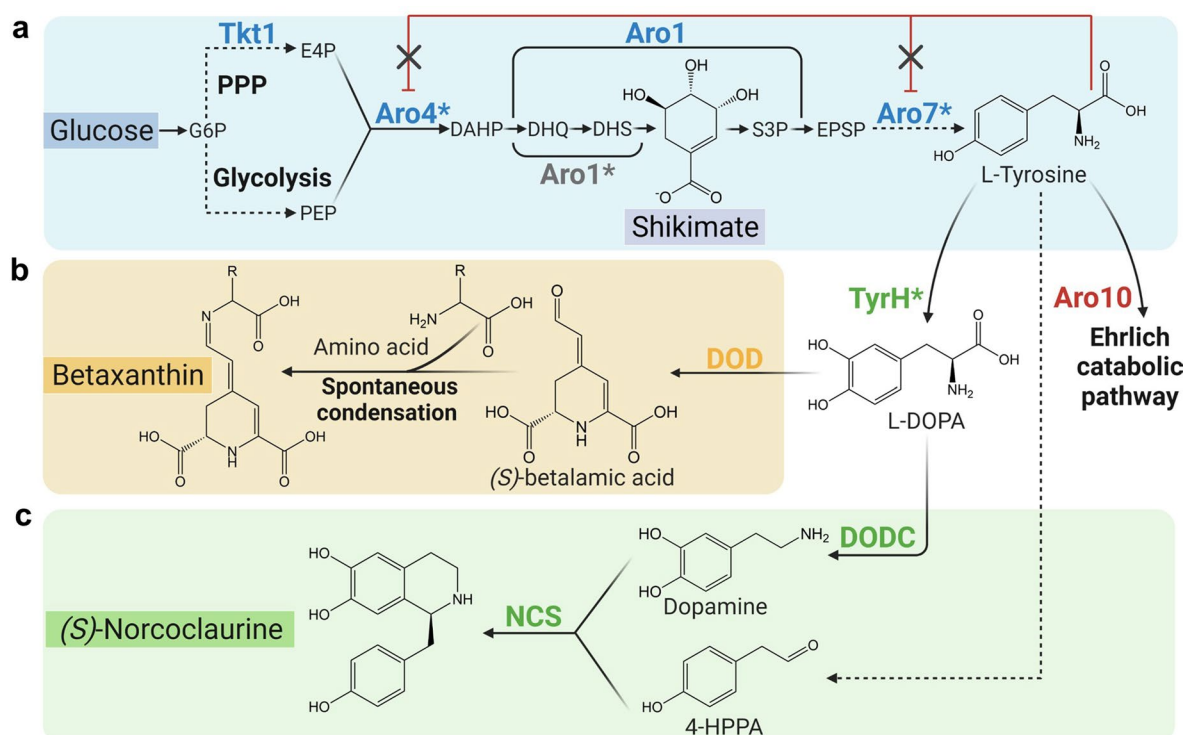
Extended Data Fig. 7 | Assessing the contribution of the hGem tag on *Xyl2:URA3* LINEAR editing efficiency with 500-bp homology arms. **a, Equimolar quantities of each LINEAR fragment were transformed to *S. stipitis* UC7. Plate pictures were taken at 48 h post-transformation. Ten colonies were selected at random (circled) for downstream analysis. **b**, The expected edit outcome on *Xyl2* locus. **c**, PCR products obtained from the primer pair in **b** using the genomic DNA isolated from the colonies circled in **a** as templates. Individual PCR reactions for each biological sample were performed once. **d**, Sequencing of the amplicon associated with *Cas9* #4 revealed an erroneous insertion of the donor template at the 5' end of *Xyl2*.**



Extended Data Fig. 8 | Assessing the contribution of the hGem tag on *Xyl2:URA3* LINEAR editing efficiency with 50-bp homology arms. a, Equimolar quantities of each LINEAR fragment were transformed to *S. stipitis* UC7. Plate pictures were taken at 48 h post-transformation. Ten colonies selected at random (circled) for downstream analysis. Note that *GFP* cassette was placed in between *sgRNA* cassette and the donor DNA as a negative reporter to assess aberrant recombination. The colonies with strong *GFP* expression levels were false positives. **b**, The expected edit outcome on *Xyl2* locus and the intact locus. **c**, PCR products obtained from the primer pair in **b** using the genomic DNA isolated from the colonies circled in **a** as templates. Individual PCR reactions for each biological sample were performed once. **d**, Sequencing of the amplicons associated with *Cas9-hGem* #10 and *Cas9* #6 confirmed the presence of the intact wild type *Xyl2*. **e**, Sequencing of the amplicon associated with *Cas9* #2 revealed an erroneous insertion of the donor template along with a portion of the upstream region of the LINEAR fragment.



Extended Data Fig. 9 | Assessing the contribution of the hGem tag on *Xyl2:GSAU LINEAR* editing efficiency. **a**, Equimolar quantities of each LINEAR fragment were transformed to *S. stipitis* UC7. Plate pictures were taken at 48 h post-transformation. Ten colonies that displayed genomic *GFP* expression as visualized on transilluminator selected (circled) for downstream analysis. Slightly different from the design used to achieve *Xyl2:URA3 LINEAR* editing with 50-bp homology arms, *GFP* cassette was included in the donor DNA as a positive reporter. **b**, PCR analysis of the 5' end of the *Xyl2* locus. Due to the large size of the inserted pathway, two reactions were designed to assess the nature of integration at each 500-bp homology arm. PCR products were obtained from the primer pair using the genomic DNA isolated from the colonies circled in **a** as templates. Individual PCR reactions for each biological sample were performed once. **c**, PCR analysis of the 3' end of the *Xyl2* locus. Individual PCR reactions for each biological sample were performed once. **d**, Sequencing of the amplicon associated with *Cas9* #9 revealed an erroneous insertion of the downstream portion of the donor template at the 3' end of *Xyl2*. Wild type stands for a ~100-bp sequence originating from the wild type *Xyl2* locus between the upstream and the downstream homology arms.



Extended Data Fig. 10 | Pathway map depicting the aromatic amino acid pathway and the *de novo* synthesis of (S)-norcoclaurine. **a**, Glycolysis and the pentose phosphate pathway (PPP) channel carbon fluxes into the aromatic amino acid (AAA) pathway. Shikimate can serve as a reporter for assessing AAA pathway flux. Tkt1, Aro4*, Aro1, and Aro7* comprise the four-enzyme, upstream L-tyrosine precursor module. The red lines depict L-tyrosine feedback inhibition on the native Aro4 and Aro7 enzymes. **b**, (S)-betalamic acid, a product of the biosensing module of TyrH* and DOD, spontaneously condenses with endogenous amino acids to produce yellow pigmented betaxanthins. **c**, TyrH*, DODC, and NCS comprise the three-enzyme, downstream (S)-norcoclaurine module. **Metabolite abbreviations:** G6P, glucose-6-phosphate; PEP, phosphoenolpyruvate; E4P, erythrose 4-phosphate; DAHP, 3-deoxy-D-arabino-heptulosonate 7-phosphate; DHQ, 3-dehydroquinic acid; DHS, 3-dehydroshikimate; S3P, shikimate-3-phosphate; EPSP, 5-enolpyruvylshikimate-3-phosphate; L-DOPA, L-3,4-dihydroxyphenylalanine; 4-HPPA, 4-hydroxyphenylacetaldehyde. **Enzyme abbreviations:** Tkt1, transketolase; Aro4*, DAHP synthase feedback inhibition-resistant mutant (Aro4_{K220L}); Aro1, pentafunctional Aro1 polypeptide; Aro1*, Aro1 mutant (Aro1_{D900A}) with the shikimate kinase domain inactivated, enabling shikimate accumulation; Aro7*, chorismate mutase feedback inhibition-resistant mutant (Aro7_{G139S}); TyrH*, tyrosine hydroxylase mutant (TyrH_{W13L_F309L}) from *Berberis vulgaris* with low L-DOPA oxidase activity; DOD, L-DOPA dioxygenase from *Mirabilis jalapa*; DODC, L-DOPA decarboxylase from *Pseudomonas putida*; NCS, norcoclaurine synthase from *Papaver somniferum*.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| ☐ | ☒ | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☒ | ☐ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Image Lab 5.2.1 (BioRad), MinNOW v. 20.10.6 (Oxford Nanopore Technologies), Guppy 4.2.3 (Oxford Nanopore Technologies)

Data analysis Prism 9.2.0 (Graphpad), BD FACSDiva 8.0.1 (BD Biosciences), NanoPlot v. 1.32.0, Downpore v. 0.3.3, Minimap2 v. 2.2-r409 (GitHub: <https://github.com/ISUGenomics/Zengyi-2021-ScheffersomycesStipitis>), re-DOT-able (Babraham Bioinformatics)

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

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Source data (growth curve, flow cytometry, shikimate and betaxanthin absorbance assays, norcoclaurine LCMS) is provided for the results presented in the Main and Extended Data Figures. Vectors containing LINEAR constructs for *Y. lipolytica*, *K. marxianus*, *H. polymorpha*, and *S. stipitis* are made available via Addgene (plasmids nos. 174837-174840, which can be searched upon the acceptance of the manuscript). The raw reads and the assembled genome sequences for the final engineered norcoclaurine-producing strains have been deposited in NCBI under BioProject ID PRJNA753835, which will become available as soon as the article is accepted. Additional data is available from the corresponding author upon reasonable request.

Field-specific reporting

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

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

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

Sample size	Sample sizes were selected to best adhere with standard practices in yeast engineering whilst being consistent with throughput of the methods, such as flow cytometry and genomic DNA isolation. For PCR verification of CRISPR edited loci, we selected n=10, as that represents a reasonable number of colonies to screen in a practical application and yields clean percentages for reporting purposes. This is in alignment with the analysis conducted in the H. polymorpha study we cite in Figure 5. From 10.1016/j.jbiosc.2017.06.001: "In each donor DNA case, ten Ade*null*HygR transformants were randomly chosen and analyzed by PCR amplification and DNA sequencing of the OpADE12 target locus." For K. marxianus and H. polymorpha, we reported n=9. This was an artifact of one failed genomic DNA extraction (out of 10) for each of the two strains. For Y. lipolytica, the sample size was higher, n=13, because the individual performing this experiment needed a PEX10 deletion for downstream application, thus opting to screen more colonies to maximize identification of the desired mutant.
Data exclusions	The norcoclaurine production of Strain 55 in the preliminary screening summarized in SourceData_Fig6 was not included for further characterization/genome sequencing as the other strains in Figure 6e, as this particular strain did not yield reproducible production titers. The SD associated with production exceeded the y-axis of Figure 6e, rendering this sample statistically insignificant from the pre-established criterion of $p < .05$ to warrant downstream analysis.
Replication	All biochemical experiments were performed using at least three independent biological replicates. A biological replicate in the context of our experimentation is one yeast colony (not one individual culture from the same inoculum, as this represents a technical replicate). For genetic experiments, we performed one round of CRISPR editing experiments for each locus, but selected multiple colonies (n=9-13) for downstream verification. We report genome editing efficiencies for 8 separate genome modifications across 4 strains, which provides cross-validation for the efficacy and reliability of LINEAR. In S. stipitis, we performed one round of initial NHEJ-mediated library screening, as we planned beforehand to confirm the top-producing strains in a follow-up assay. Where applicable, all attempts at replication were successful.
Randomization	Randomization was applied when selecting colonies transformed with the LINEAR fragment. For example, when performing XYL2 deletions in S. stipitis and H. polymorpha and ADE2 gene deletion in H. polymorpha, we picked colonies randomly, not relying on genetic techniques to identify desired mutants to skew results (i.e., identifying mutants incapable of utilizing xylose as sole carbon source for XYL2 deletion and red/white screening for ADE2 deletion).
Blinding	Blinding was not performed as the nature of the genetic and biochemical experiments and the associated data generated does not potentiate human bias influencing the final conclusions.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	S. cerevisiae YSG50 (requested from Dr. Huimin Zhao at UIUC) S. stipitis FPL-UC7 (requested from Dr. Thomas W. Jeffries, professor emeritus at University of Wisconsin) K. marxianus CBS6556 (requested from Dr. Ian Wheeldon at UC-Riverside) Y. lipolytica PO1f (ATCC)
---------------------	---

	H. polymorpha DL1 (ATCC)
Authentication	Cell lines were used without further authentication.
Mycoplasma contamination	Cells lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	N/A

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Yeast cultures were diluted 10-fold with 10 mM PBS (pH 7.4) and stored in a 4 C fridge for no longer than 30 minutes before being analyzed.
Instrument	BD FACSCanto (BDBiosciences, San Jose, CA)
Software	The fluorescence intensity distrubtion of GFP or BFP-expressing cells was calculated using BD FACSDiva (v8.0.1, BD Biosciences).
Cell population abundance	We did not perform cell sorting in this study.
Gating strategy	Data for forward scatter (FSC, photodiode 488/10 bandpass filter) and side scatter (SSC, photomultiplier 488/10 bandpass filter) were collected in logarithmic mode using instrument voltage settings typical for yeast cell populations. GFP fluorescence was excited using a 488 nm laser and emission collected through a 525/50 bandpass filter. BFP fluorescence was excited using a 405 nm laser and emission collected through a 450/50 bandpass filter. Fluorescence intensities were quantitatively measured in logarithmic mode and values recorded as mean fluorescence intensities (MFI) for each event. Initial gating to identify yeast cells was based on FSC and SSC data analyzed in dual-parameter dot plots, using data from control samples containing resuspension buffer only (i.e., no yeast cells) to distinguish between background/noise and actual yeast cell events. Most background/noise events were eliminated by setting a SSC acquisition threshold. Gated yeast cell fluorescence data was analyzed in dot plots of SSC versus the relevant fluorescent parameter, either GFP or BFP. Distinctions between "positive" and "negative" fluorescence were determined by comparing data from untransformed and transformed yeast. Data from untransformed cells was gated as "negative" fluorescence. Using the established negative gate as a lower threshold, "positive" fluorescence in transformed yeast was defined as all events outside the negative gate measured as being progressively higher on the relevant fluorescent scale.

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.