Identification of genome integration sites for developing a CRISPR-based gene expression toolkit in *Yarrowia lipolytica*

Xiaoqin Liu, Zhiyong Cui, Tianyuan Su, Xuemei Lu, Jin Hou* and Qingsheng Qi**
State Key Laboratory of Microbial Technology, Shandong University, Qingdao, 266237, China.

Summary

With the rapid development of synthetic biology, the oleaginous yeast *Yarrowia lipolytica* has become an attractive microorganism for chemical production. To better optimize and reroute metabolic pathways, we have expanded the CRISPR-based gene expression toolkit of *Y. lipolytica*. By sorting the integration sites associated with high expression, new neutral integration sites associated with high expression and high integration efficiency were identified. Diverse genetic components, including promoters and terminators, were also characterized to expand the expression range. We found that in addition to promoters, the newly characterized terminators exhibited large variations in gene expression. These genetic components and integration sites were then used to regulate genes involved in the lycopene biosynthesis pathway, and different levels of lycopene production were achieved. The CRISPR-based gene expression toolkit developed in this study will facilitate the genetic engineering of *Y. lipolytica*.

Introduction

Synthetic biology aims to apply engineering principles to biological systems by using a design-build-test-learn cycle to obtain strains with optimized production (Larroude *et al*., 2018a). The construction of efficient cell factories usually requires the integration of multi-step pathway genes and the rewiring of cellular metabolism. Therefore, it is necessary to characterize the genetic elements for gene expression and integration and develop editing tools for genome engineering.

Recently, the non-conventional yeast *Yarrowia lipolytica* has emerged as an excellent organism for the production of lipid-based chemicals (Blazeck *et al*., 2014), organic acids (Papanikolaou *et al*., 2008; Cui *et al*., 2017) and isoprenoids (Larroude *et al*., 2018b) due to its unique physiological and metabolic advantages, such as extensive substrate utilization (Ledesma-Amaro and Nicaud, 2016), low pH tolerance (Cui *et al*., 2017), high lipid production (Friedlander *et al*., 2016), high acetyl-CoA flux and tricarboxylic acid cycle flux (Markham and Alper, 2018). However, the engineering of *Y. lipolytica* is often inefficient (Wagner and Alper, 2016); therefore, various synthetic biology tools have been developed to facilitate the engineering process.

The expression of heterologous genes mainly relies on episomal vectors and genome integration in yeast (Lian *et al*., 2018). In *Y. lipolytica*, episomal expression is based on a chromosome replication system–centromere autonomously replicating sequences (CEN/ARS) (Vernis *et al*., 1997; Dulermo *et al*., 2017); however, CEN/ARS plasmids have the disadvantages of genetic instability and low copy numbers. In our previous study, we developed a new episomal plasmid system based on the mitochondrial replication origin in *Y. lipolytica*. This system exhibited good genetic stability in the Ku70 deletion strain (Cui *et al*., 2021a).

Integration of genes into the genome is a more stable approach for gene expression. Canonically, organisms employ two basic mechanisms to repair DNA double-strand breaks (DSBs): homologous recombination (HR) and non-homologous end joining (NHEJ) (Lieber, 2010). CIRPSR-mediated insertion generally requires homologous recombination to insert DNA into Cas9-induced DSBs, and it inserts DNA in a targeted site. Differently, NHEJ-mediated insertion integrates DNA during spontaneous DSB generation, and therefore, the integration site is random. Similar to other non-conventional yeasts, NHEJ is the dominant pathway for repairing DSBs in *Y. lipolytica* (Wagner and Alper, 2016); therefore, NHEJ-mediated integration was often used for heterologous

Received 13 January, 2022; revised 18 March, 2022; accepted 24 March, 2022.
For correspondence. *E-mail houjin@sdu.edu.cn; Tel. +86 532 58632401; Fax +86 532 58632401. **Email qiqingsheng@sdu.edu.cn; Tel. +86 532 58632580; Fax +86 532 58632580.
Microbial Biotechnology (2022) **15**(8), 2223–2234
doi:10.1111/1751-7915.14060
Funding information
This work was supported by the National Key R&D Program of China (2021YFC2100500), the Key R&D Program of Shandong Province (2020CXGC010602) and the National Natural Science Foundation of Shandong province (ZR2020YQ18).

gene expression (Palmer *et al*., 2020). In our previous work, NHEJ-mediated random integration was used to create a modular expression library to optimize the metabolic pathway in *Y. lipolytica* (Cui *et al*., 2019). In addition, we also developed a CRISPR/NHEJ-mediated targeted genome integration tool in *Y. lipolytica*, enabling targeted genome integration of the DNA fragment without homologous arms (Cui *et al*., 2021b). However, NHEJ-mediated random genome integration requires selection marker for gene expression and the random integration of DNA fragments may disrupt important endogenous genes, and NHEJ-mediated targeted integration displayed relatively low integration efficiency (Verbeke *et al*., 2013; Jang *et al*., 2018). Therefore, additional genetic tools and integration sites with higher integration efficiency for gene expression are desired.

Various tools have been designed for gene editing and genome integration. Schwartz *et al*. (2016) used synthetic RNA polymerase III promoters to improve the efficiency of CRISPR/Cas9-mediated genome editing and developed a CRISPR/Cas9-based tool to achieve standardized markerless gene integration in *Y. lipolytica* (Schwartz *et al*., 2017). Gao *et al*. developed a single plasmid-based CRISPR/Cas9 system using pCASyl that enabled efficient, scarless, single or multiple-gene editing of *Y. lipolytica* (Gao *et al*., 2016). Larroude *et al*. (2019) developed a modular Golden Gate toolkit containing nine promoters, five terminators, six markers and one random integration site (ZETA sequence) as well as three targeted genome integration sites (*LIP2, GSY1* and *MFE*) for the rapid sequential construction of multiple elements. A set of modular cloning vectors compatible with the BioBrick standard has been developed, called YaliBricks, which allows the assembly of multigene pathways in *Y. lipolytica* (Wong *et al*., 2017). The EasyCloneYALI genetic toolkit integrated gene expression vectors into target sites without markers using CRISPR/Cas9 technology with an efficiency exceeding 80%. This study identified 11 intergenic sites and evaluated the compatibility of 12 promoters (Holkenbrink *et al*., 2018).

Although these advances have greatly expanded the available toolkits in *Y. lipolytica*, the number of neutral integration sites for stable gene expression and high integration efficiency without significant effects on cellular physiology and metabolism is still relatively few. The genomic location, histone modification and the distance to adjacent genes may influence the expression of integrated genes and cellular physiology (Chen *et al*., 2013; Arnone, 2020); therefore, the genomic sites for gene expression must be carefully selected.

Previously, we mapped the distribution of NHEJ-mediated integration and demonstrated that it randomly inserts DNA into chromosomes (Liu *et al*., 2022). The integration efficiency is higher in intergenic regions than in intragenic regions. We also found that the expression of genes integrated via NHEJ-mediated random integration varies due to the difference in genomic locations (Cui *et al*., 2019). Rapid random genomic insertion through NHEJ-mediated integration provides the possibility to construct a random expression library, which also lays a foundation for screening the integration sites with high gene expression.

In this study, through fluorescence-activated cell sorting (FACS) of an NHEJ-mediated random *GFP* expression library, sequencing of the high expression strains and charactering the potential integration sites, we obtained new neutral integration sites in *Y. lipolytica* that can achieve high gene expression and high integration efficiency using CRISPR/Cas9 gene editing. Asides from it, 18 promoters and 12 terminators were also characterized to expand the expression range. They were then combined to regulate the lycopene biosynthesis pathway as a proof-of-concept. These neutral integration sites, promoters and terminators can be used as a synthetic biology toolkit for constructing cell factories in *Y. lipolytica*.

Results

Identification of potential neutral integration sites by constructing an NHEJ-mediated GFP *expression library*

Our previous studies revealed that NHEJ-mediated random genome integration generates variation in the locations of the inserted fragments, resulting in expression differences among the integrated genes (Cui *et al*., 2019), and it is also a powerful approach to construct a genome-scale trackable insertional mutagenesis library (Liu *et al*., 2022). To identify potential neutral integration sites with relatively high expression, an *hrGFP* expression cassette (*hrGFP* under the control of a *UAS1B8-TEF1* promoter [*pUT8*] (Madzak *et al*., 2004, Shabbir Hussain *et al*., 2016) and *CYC1* terminator [*CYC1 t*]) was transformed into the *Y. lipolytica* Po1f strain to construct an *hrGFP* random expression library. Through FACS, low and high fluorescence expression libraries were obtained (Fig. 1A and B). The mutant strains were then randomly selected from the expression libraries for fluorescence quantification. As shown in Fig. 1C, the fluorescence of mutants from the high fluorescence expression libraries was approximately 2–5-fold higher than that of mutants from the low fluorescence expression library. Low fluorescence expression may be caused by gene silencing or partial insertion of *hrGFP* expression cassettes (Chen and Zhang, 2016; Wu *et al*., 2017; Arnone, 2020). The mutants from the high fluorescence expression library were used for neutral integration sites screening (Fig. 1A–C). To determine the

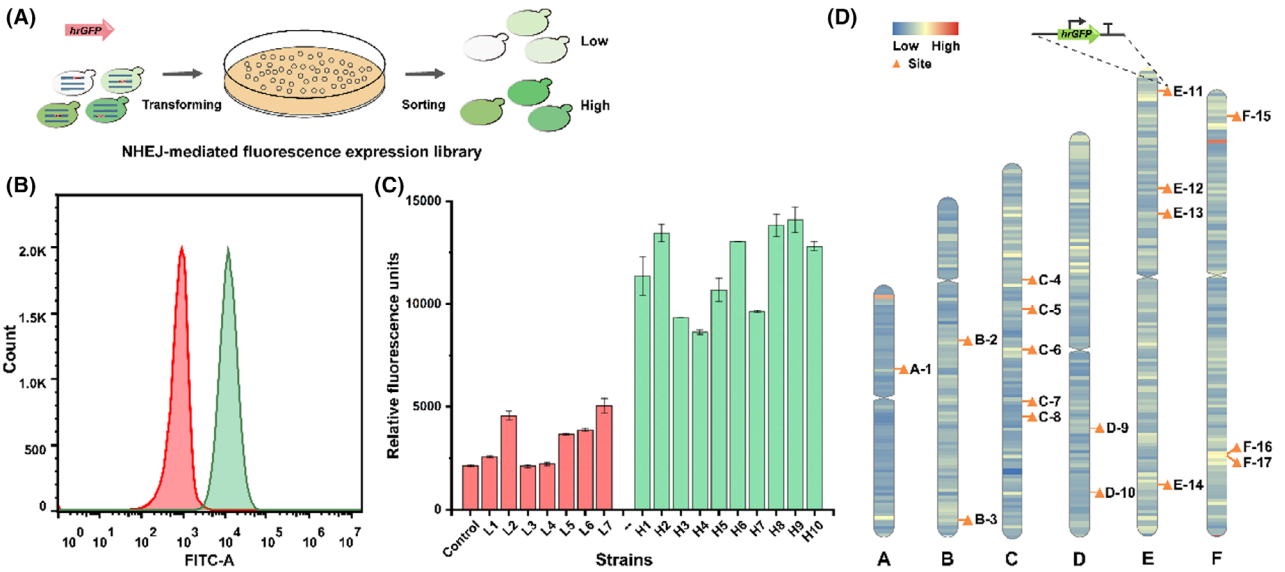

Fig. 1. Screening of integration sites in *Yarrowia lipolytica*.
A. Construction of a non-homologous end joining (NHEJ)-mediated fluorescence expression library. The fluorescence expression library was constructed via random genome integration of an *hrGFP* expression cassette via NHEJ-mediated integration in *Y. lipolytica*. The high fluorescence expression mutants were then obtained by flow cytometry sorting.
B. Fluorescence-activated cell sorting (FACS) results of the sorted high and low fluorescence libraries.
C. The fluorescence intensity of mutant strains randomly selected from high and low fluorescence expression libraries. The control was the strain without *hrGFP* expression.
D. The location of candidate neutral integration sites in each chromosome. The yellow triangle represents the location of sites in the chromosomes. The data represent the mean ± standard deviation of three biological triplicates.

integration sites, ligase-mediated intramolecular circularization followed by sequencing was employed (Michel *et al*., 2017; Liu *et al*., 2022). The integration sites in the intergenic regions and 600 bp from the coding region were selected as candidate sites. Finally, 17 candidates from the high fluorescence expression library distributed across the six chromosomes were selected for further verification (Fig. 1D).

Determining the integration efficiency of CRISPR/Cas9-mediated integration

CRISPR/Cas9-mediated targeted integration is often affected by NHEJ-mediated random integration and the targeting efficiency of sgRNA. To obtain efficient genome integration sites, the integration efficiency of the candidate sites was determined. Previously, we integrated the *Cas9* expression cassette at YALI1_E15321g located in chromosome E of Po1f (Po1f-Cas9) and found that the cutting efficiency was higher than that in the strain in which Cas9 was expressed in the plasmid (Cui *et al*., 2021b). Po1f-Cas9 was, therefore, used for CRISPR-mediated integration. The donor plasmid with a reporter gene expression cassette flanked by 1-kb homology regions and a sgRNA expression plasmid were constructed. Unique restriction sites (NdeI and BstBI) on the flanks of the homology region were introduced to facilitate the cloning of other genes of interest (Fig. 2A). A pair of primers were used to amplify the sgRNA expression plasmids where the reverse primers carrying a 20-bp designed crRNA sequence (Table S1). The linearized plasmids were directly transformed into *E. coli* DH5α without the external assembly enzyme system to obtain sgRNA plasmids with the new crRNA sequence. The donor plasmid and sgRNA plasmid were co-transformed into Po1f-Cas9, and the rate of targeted integration of *hrGFP* expression cassettes was evaluated by colony PCR.

The integration efficiency of 17 tested integration sites ranged from 8.3 to 87.5%. The integration efficiency of test sites including C-7, B-3, F-17, A-1, E-13, D-9, D-10, C-4, E-14 and E-12 exceeded 50%, indicating that these locations could be potential sites for genome integration (Fig. 2B).

The expression stability of the integration sites

To test the expression levels of these integration sites, *hrGFP* and *α-amylase* were used as reporter genes for verification. The expression levels were affected by the integration sites. For *hrGFP* expression, the difference was less than onefold under regulation by different promoter/terminator combinations of pUT8-CYC1 t and pGPD-1-CYC1 t (Fig. 3A and Fig. S1). Compared with

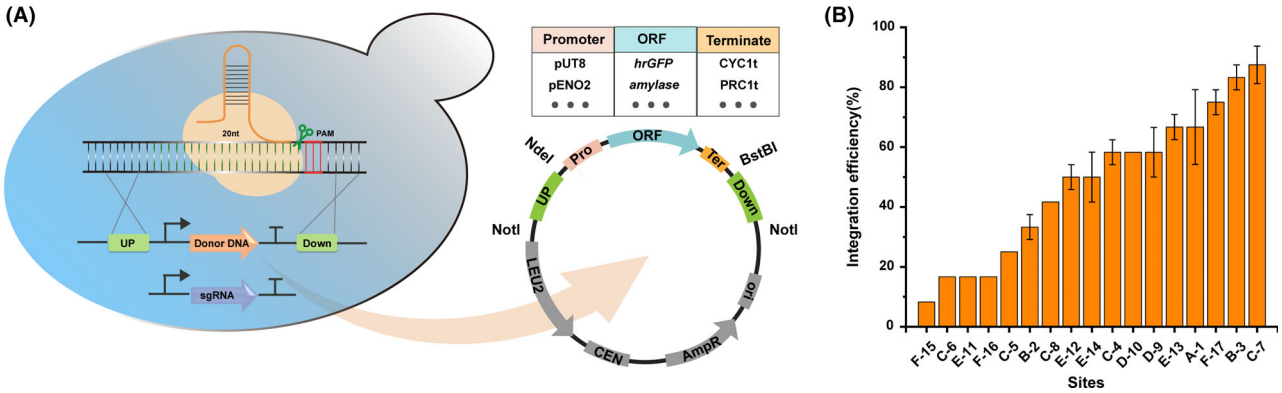

Fig. 2. Design of CRISPR-based toolkit for programming gene expression in *Y. lipolytica*.
A. Schematic representation for CRISPR/Cas9-mediated markerless gene integration.
B. Integration efficiency of the integration sites. The integration efficiency was quantified by colony PCR and sequencing of 12 random picked transformants from three separate transformations. The data represent the mean ± standard deviation of three biological triplicates.

the fluorescence intensity, α-amylase activity displayed a similar trend in general, but the differences were bigger, such as the threefold expression difference between sites C-7 and F-17 (Fig. 3B). It was speculated that the inconsistency of expression differences may be attributable to the differences of the cultivation conditions (deep well plate for fluorescence vs. shake flasks for α-amylase activity). In addition, most strains had higher relative fluorescence and α-amylase activity than the A08 strain (Schwartz *et al.*, 2017), the integration site described in previous studies. Overall, most of the integration sites met the requirement for high expression.

To test the adaptability of our selected integration sites in different conditions, we measured the fluorescence and growth of strains in different media and growth stages (Fig. 3C–F and Fig. S2). Nutrient-rich yeast extract peptone dextrose (YPD) medium and yeast extract peptone glycerol (YPG) medium and nutrient-poor synthetic dextrose (SD) medium supplemented with suitable amino acid dropout mixes were selected for the analysis. Overall, the trend of the fluorescence intensity of integration sites was similar under different conditions. Among them, the expression level of the fluorescent protein was the highest in YPD medium followed by YPG medium, whereas the fluorescence intensity in SD medium was much lower (Fig. 3C–E). The fluorescence intensity increased with increasing cultivation time in YPD and YPG medium, whereas in SD medium, the fluorescence intensity decreased in the later cultivation stage (Fig. 3E). The growth rate of all integrated strains was similar to that of control strains in YPD, YPG and SD media (Fig. 3F and Fig. S2), demonstrating that all tested integration sites can stably express the reporter gene without affecting growth.

In summary, neutral integration sites A-1, B-3, D-10, E-13, E-14, and F-17 (highlighted using bold font in Table S1) displayed high gene expression level and integration efficiency; therefore, they can be used for future genome engineering. In addition, neutral integration sites with differences in expression levels can also be selected according to the requirements of gene expression level.

Characterization of different promoters

The promoter and terminator regions are important elements for controlling gene expression. When driving recombinant gene expression to optimize metabolic pathways, the promoter strength is usually the most important parameter considered in research. In addition to the promoter, the terminator also plays an important role in regulating gene expression. We, therefore, characterized a series of promoters and terminators for their effects on gene expression using integration site B-3. In total, 15 endogenous and 3 hybrid promoters were selected, and their activity was quantified using the fluorescent reporter gene (*hrGFP*) controlled by the *CYC1* terminator. The hybrid promoters included *UT8*, *HP4D* (Madzak *et al.*, 2004) and *UAS1B-TEF1* (a copy of UAS1B was inserted upstream of the *TEF1* promoter). UAS1B is a 90-bp region of the upstream activation sequence (UAS1) of the *XPR2* promoter (Shabbir Hussain *et al.*, 2016). The selected promoters span a wide range of transcriptional strengths in different media (Fig. 4A and B). Among them, the *UT8* and *UAS1B-TEF1* promoters displayed the highest activity in YPD and YPG media, followed by the *TEF1* and *GPD-2* promoters. We found that the *CYC1* promoter, which was not often used in *Y. lipolytica*, also exhibited relatively

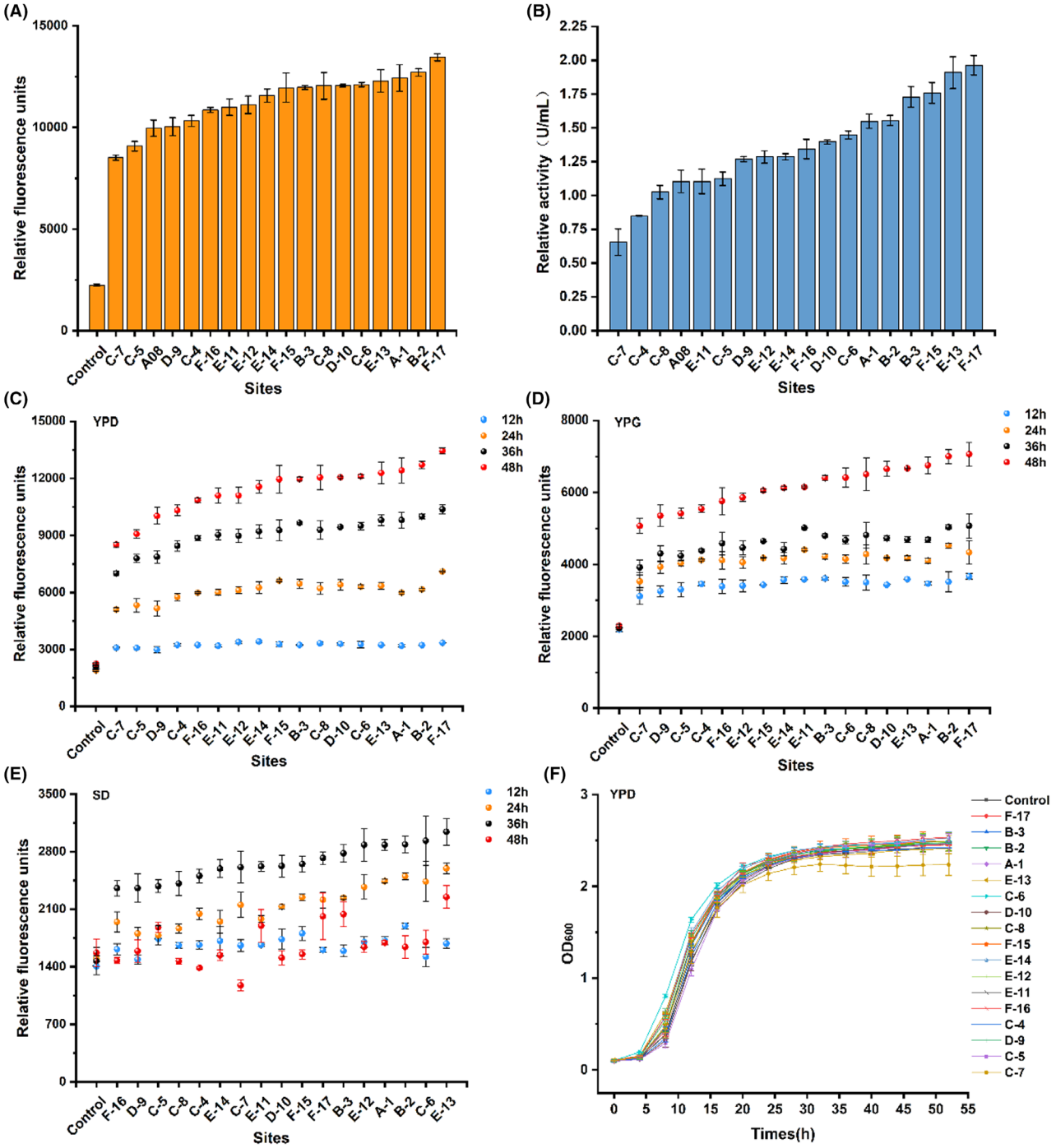

Fig. 3. Expression stability of the integration sites.
A. Fluorescence intensity of *hrGFP* at different integration sites.
B. α-amylase activity at different integration sites. The A08 was the integration site described in previous studies (Schwartz *et al.*, 2017).
C. The fluorescence intensity in YPD medium.
D. The fluorescence intensity in YPG medium.
E. The fluorescence intensity in SD medium.
F. The growth curve in YPD medium. The data represent the mean ± standard deviation of three biological triplicates.

high activity in YPD and YPG media. In addition, we found that differences in promoter length resulted in different activity. For the *GPD* promoter, when we increased the length from 1000 bp (−1000 bp to −1 bp upstream of the coding region, *pGPD-1*) to 1231 bp (−1231 bp to −1 bp upstream of the coding region,

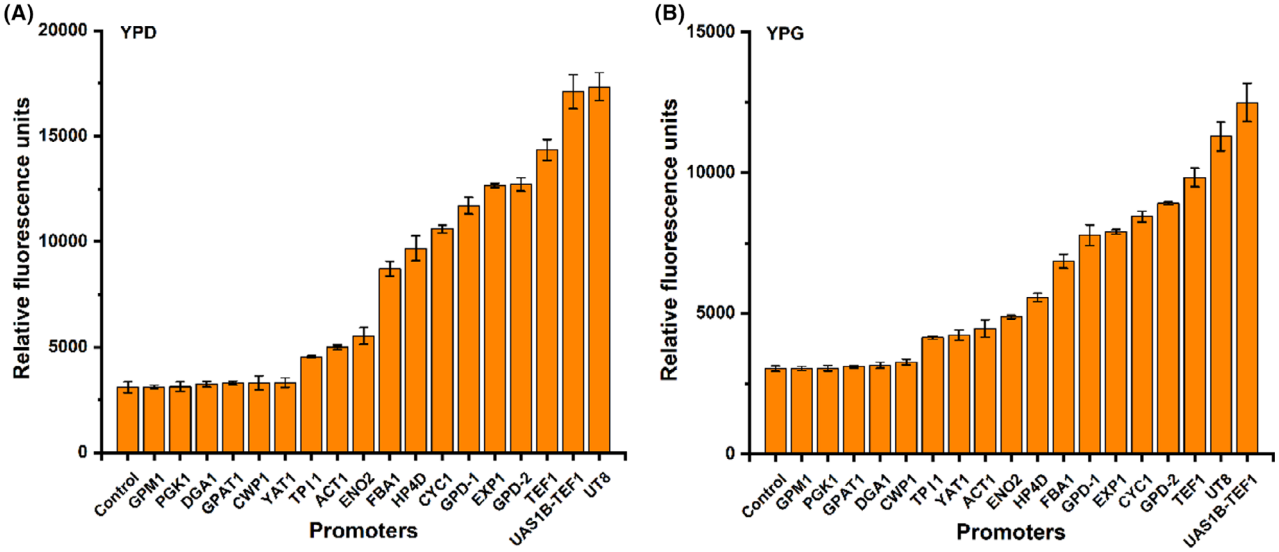

Fig. 4. Characterization of promoters in different medium.
A. The fluorescence intensity of strains expressing *hrGFP* controlled by various promoters in YPD medium.
B. The fluorescence intensity of strains expressing *hrGFP* controlled by various promoters in YPG medium. *pGPD-1*, −1000 bp to −1 bp upstream of the coding region. *pGPD-2*, −1231 bp to −1 bp upstream of the coding region. The data represent the mean ± standard deviation of three biological triplicates.

pGPD-2) of the upstream sequence, the activity increased by 14.64% (Fig. 4A and B).

Identification and characterization of different terminators

In *Y. lipolytica*, relatively few terminators have been characterized, and we identified nine terminators based on the homologous genes from *Saccharomyces cerevisiae* and considered the downstream sequences of the genes as terminators. Three previously reported terminators were also included. The terminators were constructed downstream of the *hrGFP* reporter gene under the control of the endogenous *EXP1* promoter, and the activity of the terminators was assessed. We found that the different terminators were exhibited large variations in expression, and the activity of the strongest terminator *PRC1* was sixfold and threefold higher than those of *VHS1* and *CYC1*, respectively (Fig. 5A and B). In YPD

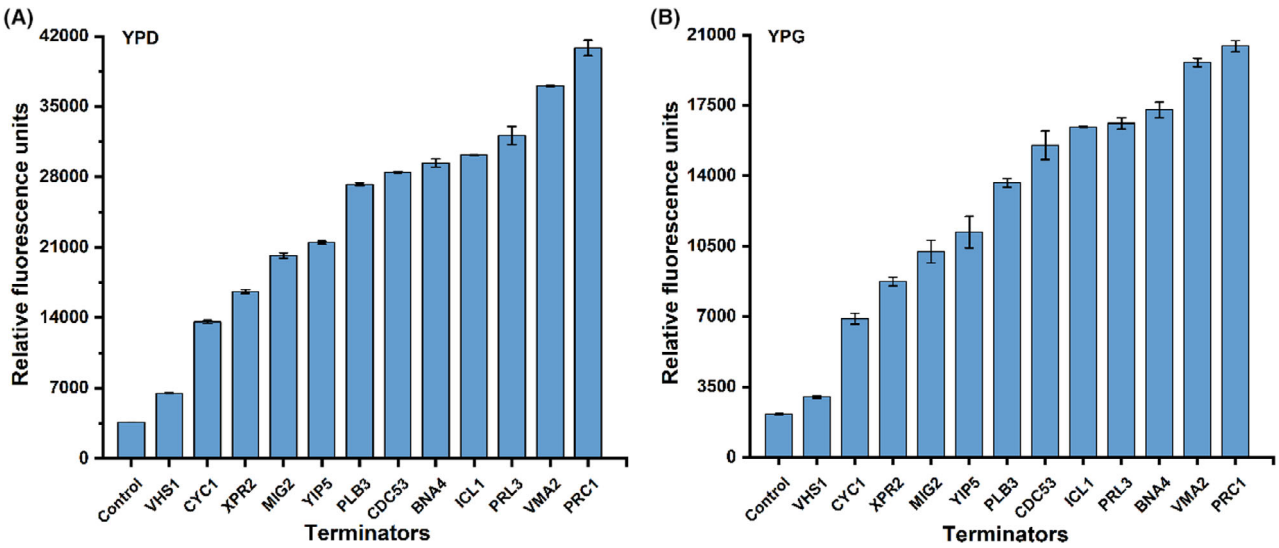

Fig. 5. Characterization of terminators in different medium.
A. The fluorescence intensity of strains expressing *hrGFP* controlled by various terminators in YPD medium.
B. The fluorescence intensity of strains expressing *hrGFP* controlled by various terminators in YPG medium. The data represent the mean ± standard deviation of three biological triplicates.

medium, the activity decreased in the order of *PRC1 t* > *VMA2 t* > *PRL3 t* > *ICL1 t* > BNA4 t > *CDC53 t* > *PLB3 t* > *YIP5 t* > *MIG2 t* >*XPR2 t* > *CYC1 t* > *VHS1 t* (Fig. 5A and B). Notably, the selected terminators exhibited even stronger gene regulation than the promoters. Previous studies illustrated that terminators can both terminate transcription and affect protein expression by altering the abundance of mRNA and half-life of mRNA (Ito *et al*., 2013, 2020; Curran *et al*., 2015). In this study, we newly identified nine terminators and demonstrated their extensive impact on gene expression. Therefore, promoters and terminators with a wide range of activities also constitute important parts of our toolkit, providing elements for optimizing pathways and constructing efficient cell factories.

Construction of the lycopene biosynthesis pathway using the identified neutral integration sites, promoters and terminators

We identified the neutral integration sites, promoters and terminators that can drive gene expression levels to a wide range. To detect the combinatorial effects of these elements, as a proof-of-concept, we incorporated the lycopene biosynthesis pathway into *Y. lipolytica*, which required the co-expression of three heterogeneous genes, containing the *CrtE*, *CrtB* and *CrtI* (GenBank: M87280.1) from *Pantoea agglomerans* (Fig. 6A). We selected promoters (*pENO2*, *pFBA1*, *pUT8* and *pEXP1*), terminators (*PRC1 t*, *ICL1 t* and *CYC1 t*) and neutral integration sites (B-3 and F-17) to fine-tune gene

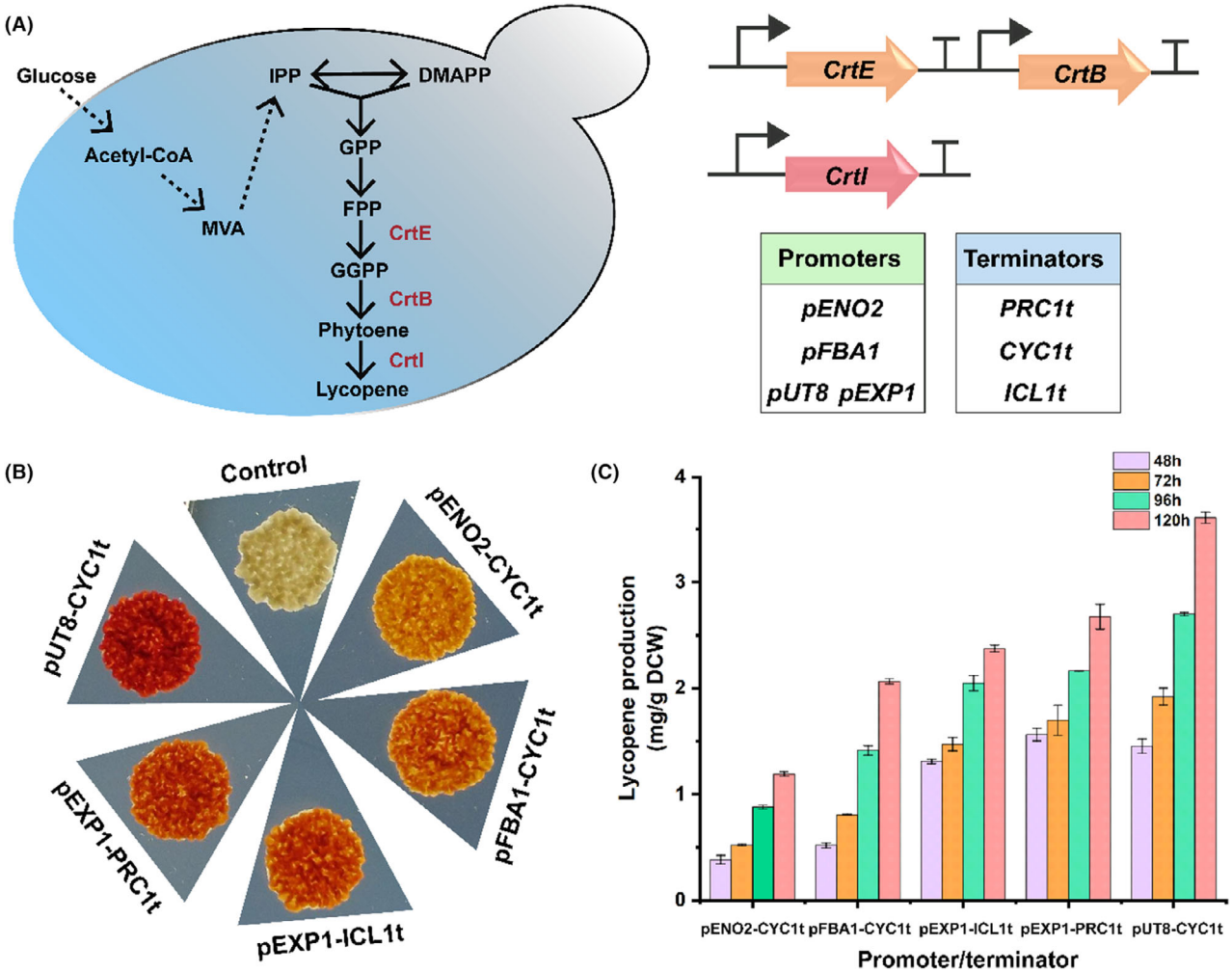

Fig. 6. Engineering lycopene biosynthesis with characterized neutral integration sites and promoters and terminators in *Y. lipolytica*.
A. Construction and integration of lycopene biosynthesis pathway in *Y. lipolytica* through a CRISPR-based toolkit. *CrtE* and *CrtB* were expressed at neutral integration site F-17. Different promoter and terminator combinations were used to express *CrtI* at neutral integration site B-3.
B. Phenotypes of strains with different lycopene production.
C. The lycopene production of the strains expressing *CrtI* controlled by various promoter and terminator combinations at different growth stages. The data represent the mean ± standard deviation of three biological triplicates.

expression. *CrtE* and *CrtB* were expressed under the control of the *UT8* and *GPD-1* promoters and *CYC1* and *XPR2* terminators and integrated into neutral integration site F-17. *CrtI* was integrated into neutral integration site B-3 and controlled by different combinations of promoters and terminators (Fig. 6A). As presented in Fig. 6B, strains with differences in lycopene production as indicated by differences in colour (darker colours indicated greater lycopene production) were obtained. Then, lycopene production was further assessed by high-performance liquid chromatography (HPLC). The expression of *CrtI* regulated by different combinations of promoters and terminators resulted in threefold variation of lycopene production (1.19–3.61 mg/g DCW) in YPD medium (Fig. 6C). These results indicated the feasibility of the combined regulation of gene expression and pathway optimization.

Discussion

Due to the rapid development of synthetic biology tools and the metabolic plasticity of *Y. lipolytica*, it has quickly become one of the most attractive unconventional yeasts for chemicals production (Markham and Alper, 2018). Stable gene expression is considered to be a decisive factor in the construction of a robust cell factory for industrial production. Therefore, genome integration is preferred in strain development versus expression based on relatively unstable plasmid systems (Yamane *et al.*, 2008; Liu *et al.*, 2014). Although CRISPR/Cas9 gene editing tools have been established and advances have been made for *Y. lipolytica* (Gao *et al.*, 2016; Schwartz *et al.*, 2016, 2017), the limited number of well-characterized genome integration sites and genetic expression components hinders the development of strain engineering.

Some studies have used auxotrophic markers for gene integration, such as *LEU2, TRP1* and *URA3*, but they may affect cell fitness (Juretzek *et al.*, 2001; Nicaud *et al.*, 2002; Yu *et al.*, 2021). Repetitive genome sequences such as the rDNA region (the DNA sequence encoding ribosomal RNA) and ZETA region (the long terminal repeat sequence of the *Y. lipolytica* retrotransposon Ylt1) were also commonly selected for multi-copy integration (Ledall *et al.*, 1994; Juretzek *et al.*, 2001; Bordes *et al.*, 2007). However, this might cause the integration of multiple genes in an unpredictable manner and result in instability (Wang *et al.*, 2018). Therefore, it is necessary to characterize the integration sites to construct a metabolic pathway featuring multiple genes in *Y. lipolytica*.

The abundance of neutral integration sites and genetic components for fine-tuning gene expression can strongly support extensive metabolic engineering in *Y. lipolytica*. In our study, we developed a synthetic biology toolkit that permitted tunable gene expression (neutral integration sites, promoters and terminators) and targeted integration (CRISPR/Cas9) for genetic engineering in *Y. lipolytica*. A fluorescence expression library was constructed via NHEJ-mediated random genome integration in *Y. lipolytica*. Integration sites with high expression level were obtained through sorting the expression library. Combined with determining the CRISPR-mediated integration efficiency, six neutral integration sites with high expression and high integration efficiency were obtained.

The expansion of gene expression component libraries is also a critical part of metabolic engineering. Many natural, synthetic and inducible promoters have been characterized, and they exhibited a wide range of transcriptional activities (Ogrydziak and Scharf, 1982; Muller *et al.*, 1998; Madzak *et al.*, 2004). Generally, promoters were the first choice for regulating gene expression in metabolic engineering, and the effect of terminators on transcription was often underestimated. Terminators can regulate gene expression by altering processes including the cleavage of 3'-mRNA, the addition of poly(A) and post-transcriptional regulation, and they influence mRNA abundance, mRNA stability and post-transcriptional interactions with endogenous *trans*-acting factors (Ito *et al.*, 2013, 2020; Curran *et al.*, 2015). Compared with the promoters, the terminators in *Y. lipolytica* have not been fully characterized. In our study, the characterized endogenous terminators exhibited large variation of expression levels. In addition, different combinations of neutral integration sites, promoters and terminators make it possible to regulate biosynthetic pathways. Through this multi-factor regulation strategy, we obtained different levels of lycopene production.

In summary, we developed a CRISPR-based gene expression toolkit that describes the expression of neutral integration sites, promoters and terminators under different culture conditions, and verified the utility of the toolkit through lycopene production. The toolkit provides a powerful tool for creating efficient cell factories.

Experimental procedures

Strains and growth conditions

All yeast strains are listed in Table S2. The *Y. lipolytica* strains were routinely cultivated at 30°C with shaking at 220 rpm in yeast extract peptone dextrose (YPD) medium (20 g/L glucose, 20 g/L tryptone and 10 g/L yeast extract), yeast extract peptone glycerol (YPG) medium (20 g/L glycerol, 20 g/L tryptone and 10 g/L yeast extract) or synthetic dextrose (SD) medium (20 g/L glucose, 1.7 g/L yeast nitrogen base and 5 g/L $(NH_4)_2SO_4$) supplemented with suitable amino acid drop-out mixes as indicated. *Escherichia coli* DH5α was

grown in Luria–Bertani broth (LB) supplemented with ampicillin (50 mg/L) at 37°C for routine plasmid construction and subcloning. For yeast shake flask fermentation, the seed culture was transferred to shake flasks containing 50 mL YPD fermentation medium (20 g/L glucose, 20 g/L tryptone and 10 g/L yeast extract).

Plasmid construction and transformation

The plasmids and primers used in this study are listed in Tables S2 and S3.

hrGFP and the terminator *CYC1* were amplified from pKi-1-hrGFP, fused by overlap extension PCR and then assembled via Gibson Assembly to obtain the plasmid YLEP-pUT8-hrGFP-CYC1 t-LEU. The pUT8-hrGFP-CYC1 t expression cassette was amplified and assembled with the JMP114 plasmid to generate the integrative plasmid JMP-pUT8-hrGFP-CYC1 t-LEU, which was used for the subsequent construction of the fluorescence expression library. The pUT8-hrGFP-CYC1 t expression cassette and the resulting backbone were assembled with 1-kb homology arms using Gibson Assembly to generate the homology donor plasmid Site-pUT8-hrGFP-CYC1 t-LEU. Unique restriction sites such as NotI and NdeI, BstBI and NotI were added on both sides of the homology arms. To test the gene regulation of other promoter/terminator combinations at different integration sites, we replaced the promoter *UT8* with *GPD-1*. The sgRNA expression plasmid YLEP-URA-sgRNA (Cui *et al*., 2021b) was amplified by a pair of reverse primers carrying a 20-bp designed crRNA sequence and transformed into *E. coli* DH5α to construct the sgRNA plasmid with specific sgRNA.

α-amylase was amplified from pTH-Amys (Tang *et al*., 2015). The LIP2 signal peptide sequence was amplified from the genomic DNA of *Y. lipolytica* Po1f. The DNA fragment containing *α-amylase*, the LIP2 signal peptide and the terminator *CYC1* was obtained by overlap extension PCR and assembled with the linearized fragment YLEP-LEU-pUT8 by Gibson Assembly to obtain the plasmid YLEP-pUT8-LIP2-amylase-CYC1 t-LEU. The plasmid YLEP-pUT8-LIP2-amylase-CYC1 t-LEU and homology donor plasmid Site-pUT8-hrGFP-CYC1 t-LEU were digested by SwaI and BstBI to produce the fragment LIP2-amylase-CYC1 t and homology plasmid backbone Site-pUT8-LEU respectively. Then, a series of homology donor plasmids Site*-pUT8-LIP2-amylase-CYC1 t-LEU were generated by enzyme digestion and ligation.

All endogenous promoters and terminators were amplified from the genomic DNA of *Y. lipolytica* Po1f. To cover a relatively wide range of expression, representative strong, medium and weak promoters and terminators were selected. Most of the tested promoters have been reported in *Y. lipolytica*, and the unreported promoters (i.e. *pENO2*, *pTPI1*, *pACT1*) were constructed on the basis of the approximately 1000-bp region upstream of the gene that was homologous to *S. cerevisiae ENO2*, *TPI1* and *ACT1*. In addition, we identified nine terminators based on the homologous genes from *S. cerevisiae* and considered the downstream 800-bp sequence of the genes as the terminator. The estimated activity of terminators mainly referred to those reported in *S. cerevisiae*. The sequences are listed in Tables S4 and S5. The hybrid promoters *HP4D* and *UT8* were amplified from pKi-2 and YLEP-LEU respectively. The homology donor plasmid B-3-pUT8-hrGFP-CYC1 t-LEU was digested by SwaI and NdeI to produce the fragment B-3-hrGFP-CYC1 t-LEU, which was assembled with different promoters by Gibson Assembly to obtain a series of plasmids B-3-Pro*-hrGFP-CYC1 t-LEU. Similarly, the homology donor plasmid B-3-pEXP1-hrGFP-CYC1 t-LEU was digested by AatII and BstBI to produce the fragment B-3-pEXP1-hrGFP-LEU, which was assembled with different terminators by Gibson Assembly to obtain a series of plasmids B-3-pEXP1-hrGFP-Ter*-LEU.

CrtE, *CrtB* and *CrtI* from *P. agglomerans* were optimized according to the codon preference of *Y. lipolytica*, synthesized by General Biosystems (Anhui, China), and ligated with YLEP-URA and JMP-hyg-GPD to generate the plasmids YLEP-pUT8-CrtB-CYC1 t-URA and JMP-pGPD-CrtE-XPR2 t-hyg respectively. The fragments pUT8-CrtB-CYC1 t-URA and pGPD-CrtE-XPR2 t were amplified by PCR and ligated with the homology donor plasmid F-17-pUT8-hrGFP-CYC1 t-LEU.

The synthesized plasmid B-3-pEXP1-CrtI-PRC1 t-LEU was digested and assembled with the terminator *ICL1* by Gibson Assembly to obtain the plasmid B-3-pEXP1-CrtI-ICL1 t-LEU. *CrtI* was amplified by PCR and use to replace *hrGFP* in the plasmids B-3-pENO2-hrGFP-CYC1 t-LEU, B-3-pFBA1-hrGFP-CYC1 t-LEU and B-3-pUT8-hrGFP-CYC1 t-LEU to generate plasmids B-3-pENO2-CrtI-CYC1 t-LEU, B-3-pFBA1-CrtI-CYC1 t-LEU and B-3-pUT8-CrtI-CYC1 t-LEU respectively.

These homology donor plasmids and YLEP-URA-sgRNA plasmids were transformed into the Po1f-Cas9 strain using a lithium acetate protocol (Chen *et al*., 1997; Gao *et al*., 2016). YPD medium was used for *URA3* and *LEU2* marker plasmid removal. The positive strains were selected by plate screening and colony PCR.

Construction of the fluorescent expression library

The *Y. lipolytica* Po1f strain was used for fluorescence expression library construction. The *hrGFP* (Cui *et al*., 2021b) expression cassette and *LEU2* marker were used as an integration fragment and transformed into *Y. lipolytica* Po1f via lithium acetate transformation.

Transformants were cultivated on SD-LEU selection plates at 30°C for 48 h. The colonies were obtained and used for subsequent flow cytometry sorting.

Flow cytometry

The cell library was cultured overnight at 30°C and 220 rpm. Flow cytometry sorting experiments were performed using the BD FACSAria™ Fusion flow cytometer. Cells were sorted at a rate of approximately 200 events/s. GFP was excited at 478 nm, and emission was detected at 530 nm. Cells were separated according to the fluorescence intensity, and cell libraries with high and low fluorescence intensity were obtained. Cells lacking *hrGFP* expression were used as autofluorescence control. Fluorescence analysis was performed using a BD Accuri™ C6 flow cytometer. A total of 30000 events were recorded from high and low fluorescence libraries for analysis respectively. Fluorescence data were analysed using FlowJo V10 software.

Sequencing and determination of the genomic location of integration sites

To track the location of the *hrGFP* expression cassette, the genomic DNA of the colonies was extracted and digested by DpnII, which can recognize tetrameric sequences (5′-GATC-3′), followed by intramolecular circularization. Circular DNA was amplified by inverse PCR using *hrGFP*-specific outward-facing primers (Table S3), and the integration locations were then determined by sequencing the PCR products.

Integration efficiency determination

As described previously, *Cas9* from the plasmid pCAS1yl-trp (Gao *et al*., 2016) was integrated into the YALI1_E15321g location of the genome to obtain a high cutting efficiency (Cui *et al*., 2021b). Po1f-Cas9 was co-transformed with 0.5 μg of the YLEP-URA-sgRNA plasmid and 1 μg of the homology donor plasmid. The transformants were spread on SD-URA-LEU agar plates and incubated for 2 days. The integration efficiency of CRISPR/Cas9-mediated genome editing was determined by PCR and sequencing analysis using primers paired with regions outside the flanking 1-kb homology region.

Fluorescence determination and growth assays

The strains were inoculated into YPD, YPG or SD medium supplemented with suitable amino acid drop-out mixes at 30°C and 220 rpm for 24 h. The seed cultures were collected, transferred to fresh medium in a 2% (v/v) inoculation volume and grown at 30°C and 220 rpm in a deep well plate. The optical density at 600 nm (OD_{600}) of the cells and hrGFP fluorescence (excitation, 485 nm; emission, 528 nm) were simultaneously detected using a Cytation microplate reader (BioTek). All fluorescence measurements were normalized to OD_{600}. All experiments were performed in triplicate.

Enzyme activity assay of α-amylase

To measure α-amylase activity, the seed culture was transferred to shake flasks containing 50 mL of YPD fermentation medium and cultured at 30°C and 220 rpm for 48 h. The supernatant was centrifuged to determine α-amylase activity. α-amylase activity was determined using a Ceralpha Assay kit (Megazyme, Ireland).

Extraction and analysis of lycopene

The pre-cultured seed of *Y. lipolytica* lycopene-producing strains was transferred to shake flasks containing 50 mL of YPD fermentation medium and cultured at 30°C and 220 rpm for 5 days. A 200-μL aliquot of the fermentation broth was centrifuged to collect the cells. The cells were resuspended in 0.7 mL of dimethyl sulfoxide, incubated at 55°C for 10 min and then incubated at 50°C for 10 min with an equal volume of acetone. Finally, the samples were centrifuged at 12,000 × *g* for 10 min. The supernatant containing lycopene was analysed using a Shimadzu LC-20 AT high-performance liquid chromatograph equipped with a 450-nm variable wavelength detector and an XDB-C18 column (Eclipse, USA). The mobile phase was methanol, acetonitrile and dichloromethane (42:42:16), the flow rate was 1.0 mL/min, and the column temperature was 30°C.

Acknowledgements

This work was supported by the National Key R&D Program of China (2021YFC2100500), the Key R&D Program of Shandong Province (2020CXGC010602) and the National Natural Science Foundation of Shandong province (ZR2020YQ18).

Conflict of interest

The authors declare no competing interest.

References

Arnone, J.T. (2020) Genomic considerations for the modification of *Saccharomyces cerevisiae* for biofuel and metabolite biosynthesis. *Microorganisms* **8:** 321.

Blazeck, J., Hill, A., Liu, L., Knight, R., Miller, J., Pan, A., *et al*. (2014) Harnessing *Yarrowia lipolytica* lipogenesis to create a platform for lipid and biofuel production. *Nat Commun* **5:** 3131.

Bordes, F., Fudalej, F., Dossat, V., Nicaud, J.M., and Marty, A. (2007) A new recombinant protein expression system for high-throughput screening in the yeast *Yarrowia lipolytica*. *J Microbiol Meth* **70:** 493–502.

Chen, D.C., Beckerich, J.M., and Gaillardin, C. (1997) One-step transformation of the dimorphic yeast *Yarrowia lipolytica*. *Appl Microbiol Biot* **48:** 232–235.

Chen, M., Licon, K., Otsuka, R., Pillus, L., and Ideker, T. (2013) Decoupling epigenetic and genetic effects through systematic analysis of gene position. *Cell Rep* **3:** 128–137.

Chen, X., and Zhang, J. (2016) The genomic landscape of position effects on protein expression level and noise in yeast. *Cell Syst* **2:** 347–354.

Cui, Z.Y., Gao, C.J., Li, J.J., Hou, J., Lin, C.S.K., and Qi, Q.S. (2017) Engineering of unconventional yeast *Yarrowia lipolytica* for efficient succinic acid production from glycerol at low pH. *Metab Eng* **42:** 126–133.

Cui, Z.Y., Jiang, X., Zheng, H.H., Qi, Q.S., and Hou, J. (2019) Homology-independent genome integration enables rapid library construction for enzyme expression and pathway optimization in *Yarrowia lipolytica*. *Biotechnol Bioeng* **116:** 354–363.

Cui, Z., Zheng, H., Zhang, J., Jiang, Z., Zhu, Z., Liu, X., *et al*. (2021b) A CRISPR/Cas9-mediated, homology-independent tool developed for targeted genome integration in *Yarrowia lipolytica*. *Appl Environ Microbiol* **87:** e02666-20.

Cui, Z.Y., Zheng, H.H., Jiang, Z.N., Wang, Z.X., Hou, J., Wang, Q., *et al*. (2021a) Identification and characterization of the mitochondrial replication origin for stable and episomal expression in *Yarrowia lipolytica*. *ACS Synth Biol* **10:** 826–835.

Curran, K.A., Morse, N.J., Markham, K.A., Wagman, A.M., Gupta, A., and Alper, H.S. (2015) Short synthetic terminators for improved heterologous gene expression in yeast. *ACS Synth Biol* **4:** 824–832.

Dulermo, R., Brunel, F., Dulermo, T., Ledesma-Amaro, R., Vion, J., Trassaert, M., *et al*. (2017) Using a vector pool containing variable-strength promoters to optimize protein production in *Yarrowia lipolytica*. *Microb Cell Fact* **16:** 31.

Friedlander, J., Tsakraklides, V., Kamineni, A., Greenhagen, E.H., Consiglio, A.L., MacEwen, K., *et al*. (2016) Engineering of a high lipid producing *Yarrowia lipolytica* strain. *Biotechnol Biofuels* **9:** 77.

Gao, S.L., Tong, Y.Y., Wen, Z.Q., Zhu, L., Ge, M., Chen, D.J., *et al*. (2016) Multiplex gene editing of the *Yarrowia lipolytica* genome using the CRISPR-Cas9 system. *J Ind Microbiol Biot* **43:** 1085–1093.

Holkenbrink, C., Dam, M.I., Kildegaard, K.R., Beder, J., Dahlin, J., Domenech Belda, D., and Borodina, I. (2018) EasyCloneYALI: CRISPR/Cas9-based synthetic toolbox for engineering of the yeast *Yarrowia lipolytica*. *Biotechnol J* **13:** e1700543.

Ito, Y., Terai, G., Ishigami, M., Hashiba, N., Nakamura, Y., Bamba, T., *et al*. (2020) Exchange of endogenous and heterogeneous yeast terminators in Pichia pastoris to tune mRNA stability and gene expression. *Nucleic Acids Res* **48:** 13000–13012.

Ito, Y., Yamanishi, M., Ikeuchi, A., Imamura, C., Tokuhiro, K., Kitagawa, T., and Matsuyama, T. (2013) Characterization of five terminator regions that increase the protein yield of a transgene in *Saccharomyces cerevisiae*. *J Biotechnol* **168:** 486–492.

Jang, I.S., Yu, B.J., Jang, J.Y., Jegal, J., and Lee, J.Y. (2018) Improving the efficiency of homologous recombination by chemical and biological approaches in *Yarrowia lipolytica*. *PLoS One* **13:** e0194954.

Juretzek, T., Le Dall, M., Mauersberger, S., Gaillardin, C., Barth, G., and Nicaud, J. (2001) Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica*. *Yeast* **18:** 97–113.

Larroude, M., Celinska, E., Back, A., Thomas, S., Nicaud, J.M., and Ledesma-Amaro, R. (2018b) A synthetic biology approach to transform *Yarrowia lipolytica* into a competitive biotechnological producer of beta-carotene. *Biotechnol Bioeng* **115:** 464–472.

Larroude, M., Park, Y.K., Soudier, P., Kubiak, M., Nicaud, J.M., and Rossignol, T. (2019) A modular Golden Gate toolkit for *Yarrowia lipolytica* synthetic biology. *Microb Biotechnol* **12:** 1249–1259.

Larroude, M., Rossignol, T., Nicaud, J.M., and Ledesma-Amaro, R. (2018a) Synthetic biology tools for engineering *Yarrowia lipolytica*. *Biotechnol Adv* **36:** 2150–2164.

Ledall, M.T., Nicaud, J.M., and Gaillardin, C. (1994) Multiple-copy integration in the yeast *Yarrowia lipolytica*. *Curr Genet* **26:** 38–44.

Ledesma-Amaro, R., and Nicaud, J.M. (2016) Metabolic engineering for expanding the substrate range of *Yarrowia lipolytica*. *Trends Biotechnol* **34:** 798–809.

Lian, J.Z., Mishra, S., and Zhao, H.M. (2018) Recent advances in metabolic engineering of *Saccharomyces cerevisiae*: new tools and their applications. *Metab Eng* **50:** 85–108.

Lieber, M.R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* **79:** 181–211.

Liu, L.Q., Otoupal, P., Pan, A., and Alper, H.S. (2014) Increasing expression level and copy number of a *Yarrowia lipolytica* plasmid through regulated centromere function. *FEMS Yeast Res* **14:** 1124–1127.

Liu, X., Liu, M., Zhang, J., Chang, Y., Cui, Z., Ji, B., *et al*. (2022) Mapping of nonhomologous end joining-mediated integration facilitates genome-scale trackable mutagenesis in *Yarrowia lipolytica*. *ACS Synth Biol* **11:** 216–227.

Madzak, C., Gaillardin, C., and Beckerich, J.M. (2004) Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: a review. *J Biotechnol* **109:** 63–81.

Markham, K.A., and Alper, H.S. (2018) Synthetic biology expands the industrial potential of *Yarrowia lipolytica*. *Trends Biotechnol* **36:** 1085–1095.

Michel, A.H., Hatakeyama, R., Kimmig, P., Arter, M., Peter, M., Matos, J., *et al*. (2017) Functional mapping of yeast genomes by saturated transposition. *eLife* **6:** e23570.

Muller, S., Sandal, T., Kamp-Hansen, P., and Dalboge, H. (1998) Comparison of expression systems in the yeasts *Saccharomyces cerevisiae*, *Hansenula polymorpha*,

Klyveromyces lactis, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Cloning of two novel promoters from *Yarrowia lipolytica*. *Yeast* **14:** 1267–1283.

Nicaud, J.M., Madzak, C., van den Broek, P., Gysler, C., Duboc, P., Niederberger, P., and Gaillardin, C. (2002) Protein expression and secretion in the yeast *Yarrowia lipolytica*. *FEMS Yeast Res* **2:** 371–379.

Ogrydziak, D.M., and Scharf, S.J. (1982) Alkaline extracellular protease produced by *Saccharomycopsis lipolytica* CX161-1B. *J Gen Microbiol* **128:** 1225–1234.

Palmer, C.M., Miller, K.K., Nguyen, A., and Alper, H.S. (2020) Engineering 4-coumaroyl-CoA derived polyketide production in *Yarrowia lipolytica* through a beta-oxidation mediated strategy. *Metab Eng* **57:** 174–181.

Papanikolaou, S., Galiotou-Panayotou, M., Fakas, S., Komaitis, M., and Aggelis, G. (2008) Citric acid production by *Yarrowia lipolytica* cultivated on olive-mill wastewater-based media. *Bioresour Technol* **99:** 2419–2428.

Schwartz, C.M., Hussain, M.S., Blenner, M., and Wheeldon, I. (2016) Synthetic RNA polymerase III promoters facilitate high-efficiency CRISPR-Cas9-mediated genome editing in *Yarrowia lipolytica*. *ACS Synth Biol* **5:** 356–359.

Schwartz, C., Shabbir-Hussain, M., Frogue, K., Blenner, M., and Wheeldon, I. (2017) Standardized markerless gene integration for pathway engineering in *Yarrowia lipolytica*. *ACS Synth Biol* **6:** 402–409.

Shabbir Hussain, M., Gambill, L., Smith, S., and Blenner, M.A. (2016) Engineering promoter architecture in oleaginous yeast *Yarrowia lipolytica*. *ACS Synth Biol* **5:** 213–223.

Tang, H., Bao, X., Shen, Y., Song, M., Wang, S., Wang, C., and Hou, J. (2015) Engineering protein folding and translocation improves heterologous protein secretion in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* **112:** 1872–1882.

Verbeke, J., Beopoulos, A., and Nicaud, J.M. (2013) Efficient homologous recombination with short length flanking fragments in Ku70 deficient *Yarrowia lipolytica* strains. *Biotechnol Lett* **35:** 571–576.

Vernis, L., Abbas, A., Chasles, M., Gaillardin, C.M., Brun, C., Huberman, J.A., and Fournier, P. (1997) An origin of replication and a centromere are both needed to establish a replicative plasmid in the yeast *Yarrowia lipolytica*. *Mol Cell Biol* **17:** 1995–2004.

Wagner, J.M., and Alper, H.S. (2016) Synthetic biology and molecular genetics in non-conventional yeasts: current tools and future advances. *Fungal Genet Biol* **89:** 126–136.

Wang, L., Deng, A., Zhang, Y., Liu, S., Liang, Y., Bai, H., *et al*. (2018) Efficient CRISPR-Cas9 mediated multiplex genome editing in yeasts. *Biotechnol Biofuels* **11:** 277.

Wong, L., Engel, J., Jin, E., Holdridge, B., and Xu, P. (2017) YaliBricks, a versatile genetic toolkit for streamlined and rapid pathway engineering in *Yarrowia lipolytica*. *Metab Eng Commun* **5:** 68–77.

Wu, X.L., Li, B.Z., Zhang, W.Z., Song, K., Qi, H., Dai, J.B., and Yuan, Y.J. (2017) Genome-wide landscape of position effects on heterogeneous gene expression in *Saccharomyces cerevisiae*. *Biotechnol Biofuels* **10:** 189.

Yamane, T., Sakai, H., Nagahama, K., Ogawa, T., and Matsuoka, M. (2008) Dissection of centromeric DNA from yeast *Yarrowia lipolytica* and identification of protein-binding site required for plasmid transmission. *J Biosci Bioeng* **105:** 571–578.

Yu, W., Gao, J., Zhai, X., and Zhou, Y.J. (2021) Screening neutral sites for metabolic engineering of methylotrophic yeast *Ogataea polymorpha*. *Synth Syst Biotechnol* **6:** 63–68.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary Material