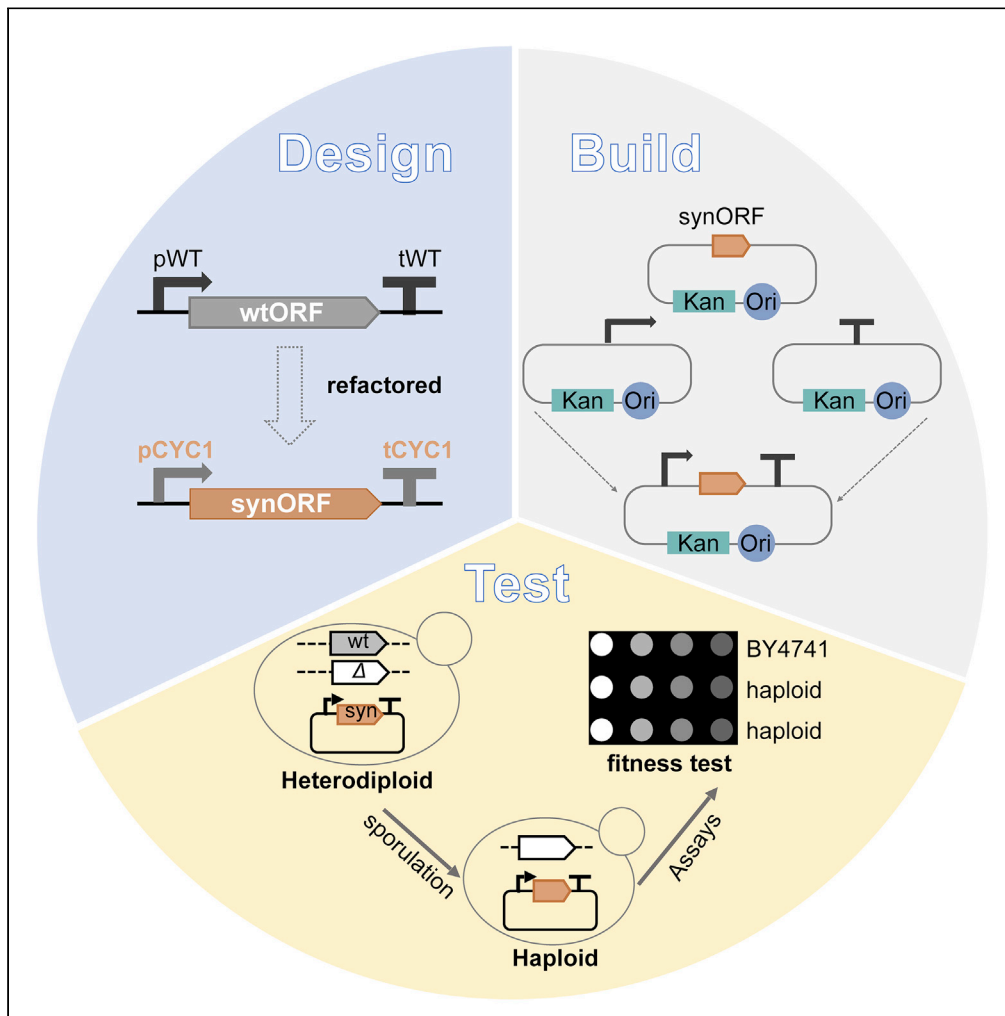**Article**

Synthetic refactor of essential genes decodes functionally constrained sequences in yeast genome

Zhenzhen Liang,
Zhouqing Luo,
Weimin Zhang, ...,
Shihui Yang,
Yingxin Ma,
Junbiao Dai

luozq@xmu.edu.cn (Z.L.)
junbiao.dai@siat.ac.cn (J.D.)

Highlights

Altering expression of most essential genes from *chrXIIL* does not affect function

Refactoring native sequences facilitates identification of novel regulatory elements

The N-terminal region of *GPI13* regulates its own translation besides protein coding

Coding sequence of *GRC3* contains essential components of *PRP19*'s promoter

Article

Synthetic refactor of essential genes decodes functionally constrained sequences in yeast genome

Zhenzhen Liang,[1,6] Zhouqing Luo,[1,2,6,*] Weimin Zhang,[1,3,6] Kang Yu,[1] Hui Wang,[1,2] Binan Geng,[4] Qing Yang,[4] Zuoyu Ni,[1] Cheng Zeng,[1] Yihui Zheng,[5] Chunyuan Li,[5] Shihui Yang,[4] Yingxin Ma,[1] and Junbiao Dai[1,5,7,*]

SUMMARY

The relationship between gene sequence and function matters for fundamental and practical reasons. Here, yeast essential genes were systematically refactored to identify invariable sequences in the coding and regulatory regions. The coding sequences were synonymously recoded with all optimal codons to explore the importance of codon choice. The promoters and terminators were swapped with well-characterized *CYC1* promoter and terminator to examine whether a specialized expression is required for the function of a specific gene. Among the 10 essential genes from *Chr.XIIL*, this scheme successfully generated 7 refactored genes that can effectively support wild-type-like fitness under various conditions, thereby revealing amazing sequence plasticity of yeast genes. Moreover, different invariable elements were identified from the remaining 3 genes, exampling the logics for genetic information encoding and regulation. Further refactoring of all essential genes using this strategy will generate comprehensive understanding of gene sequence choice, thereby guiding its design in various applications.

INTRODUCTION

As the basic physical and functional units of heredity, genes are continuously evolving in their sequences to meet the functional requirements of an organism (Carvunis et al., 2012). Functional interpretation of the various elements embedded in the coding and regulation region of genes, such as the triplet codon (Crick et al., 1961), TATA-box (Lifton et al., 1978), poly-A tail (Darnell et al., 1971; Edmonds et al., 1971; Lee et al., 1971), and many others, has significantly helped the researchers to understand and engineer gene sequence. *De novo* gene synthesis can enable the convenient creation and modification of the gene sequences, thus providing unprecedently materials to elucidate the mechanisms about how the gene functions are encoded and regulated by its sequence (Lajoie et al., 2013; Patwardhan et al., 2012).

The degeneracy of genetic code allows most amino acids a choice of optimal and non-optimal codons. Although there is no change in the protein sequence, synonymous coded sequences might have very different impact on the protein expression, conformation, and function (Hanson and Coller, 2018; Plotkin and Kudla, 2011; Tuller et al., 2010). There are almost 50 different diseases that are currently known to be associated with synonymous mutations, and recent estimates suggest that 5%–10% of human genes contain at least one region where synonymous mutations could be effectively harmful (Sauna and Kimchi-Sarfaty, 2011). Moreover, enabled by the synthetic genomics approaches, the effects of genome-wide recoding were directly probed recently in *Escherichia coli (E. coli)*. 16 out of the 42 highly expressed essential genes in *E.coli* failed to be recoded while removing 13 rare codons in them (Lajoie et al., 2013), but another recoding scheme successfully removed 3 codons in *E.coli* genome (Fredens et al., 2019; Wang et al., 2016), thus revealing the importance of codon choice. In the synthetic yeast genome project-Sc2.0, synonymous recoding was used to generate PCRtags, a DNA stretch about 28 bp that were used to distinguish the synthetic and wild-type sequences (Dymond et al., 2011; Luo et al., 2018b). Some of these PCRtags have been found to lead to significant impairment of gene functions (Mitchell et al., 2017; Shen et al., 2017; Wu et al., 2017; Zhang et al., 2017). A more radical synonymous recoding has not been previously reported in yeast, therefore limiting the understanding of codon choices in eukaryotes and the designing of codon compressed yeast genomes, an aim of the Sc3.0 project (Dai et al., 2020).

[1]CAS Key Laboratory of Quantitative Engineering Biology, Guangdong Provincial Key Laboratory of Synthetic Genomics and Shenzhen Key Laboratory of Synthetic Genomics, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

[2]State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Faculty of Medicine and Life Sciences, Xiamen University, Xiamen 361102, China

[3]Institute for Systems Genetics and Department of Biochemistry and Molecular Pharmacology, New York University Langone Medical Center, New York, NY 10011, USA

[4]Hubei Collaborative Innovation Center for Green Transformation of Bio-resources, Environmental Microbial Technology Center of Hubei Province, Hubei Key Laboratory of Industrial Biotechnology, College of Life Sciences, Hubei University, Wuhan 430062, China

[5]Key Laboratory for Industrial Biocatalysis (Ministry of Education) and Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China

[6]These authors contributed equally

[7]Lead contact

*Correspondence: luozq@xmu.edu.cn (Z.L.), junbiao.dai@siat.ac.cn (J.D.)

https://doi.org/10.1016/j.isci.2022.104982

While the "genetic code" determines a protein's amino acid sequence, other genomic regions can determine when, where, and how many of these proteins are produced according to the various "regulatory codes". Many specific functional elements, such as enhancer, transcription factor binding site, TATA box, transcription start site, and poly(A) site, are involved in these regulatory codes (Maston et al., 2006). Regrettably, how to match the regulatory code to gene function has been only explored to a limited extent in the synthetic genomics, as no synthetic genomes have been constructed with significant alterations in the regulatory sequences. Some well-characterized regulatory sequences, such as the promoter and terminator of *CYC1* gene, have been widely applied in exogenous gene expression during metabolic engineering (Curran et al., 2013; Da Silva and Srikrishnan, 2012), but has not yet been applied genome-widely for the expression of endogenous genes. Regulatory sequence swapping by replacing the native promoters and terminators with those well-characterized ones can potentially identify the regulatory codes, thereby it is important for our understanding about the regulatory codes and can guide the accurate design of gene expression behaviors in further synthetic genomics practice.

As described above, synthetic strategies have been demonstrated to be meaningful for dissecting the logics of gene sequence choice and guiding the design of new genes with desired functions, but the related practice is still limited. Here, synthetic genes were designed according to a radical refactor scheme to identify invariable elements in yeast essential genes. Open reading frames of essential genes were synonymously recoded using the "one amino acid encoded by only one codon" rule; promoters and terminators of them were swapped to the previously well-characterized *CYC1* promoter and terminator. All the 10 essential genes in the left arm of chromosome XII (*Chr.XIIL*) were used as the demos for this strategy. Our results showed that 7/10 of the refactored genes can support viability and showed normal functions under various conditions, thus revealing the high plasticity of yeast genes in codon choice and expression regulation. The reasons for non-complementation of the remaining 3 refactored genes were explored, thereby identifying the 150 bp promoter before *SFI1*'s start codon as a necessity for both its transcription and translation. Moreover, the N-terminal half of *GPI13* coding sequence was identified as an essential translation regulator beyond its coding capacity, and the coding sequence of *GRC3* as an activator sequence for *PRP19*'s transcription. These findings not only proved the possibility of using an actually different sequence to encode the similar functions on the premise of the same amino acid sequence and basal expression level but also clearly exampled the power of gene sequence refactor in decoding the novel functional elements.

RESULTS
Strategies to refactor the yeast essential genes

Promoter, coding sequence, and terminator are the three essential parts of yeast genes, and only limited yeast genes (280 out of the totally over 6,000 genes) contain introns (Figure 1A) (Parenteau et al., 2019). To find the possible sequence constraints in yeast genes, a radical refactor scheme was designed here that could cause alterations in the nucleotide sequence of all these four parts but without affecting the translated amino acid sequence (Figures 1B and 1C). Essential genes were chosen as the first stage targets because any defects in their functions will result in lethality or severe growth defects.

Firstly, we removed introns and replaced each codon of the wild-type open reading frame with its corresponding optimal one in yeast (Figure 1C). This refactoring process was done using the online tool Codon Juggling : http://54.235.254.95/cgi-bin/gd/gdCodJug.cgi, in which the highest relative synonymous codon usage (RSCU) value of the highly expressed genes in the yeast genome was used to define the optimal codon for each amino acid (Richardson et al., 2006; Sharp et al., 1988). The RSCU value was calculated by dividing the observed number of codons by the total number expected if all codons for that amino acid were used equally (Sharp et al., 1986). The difference in sequence between the native open reading frame (*wtORF*) and the recoded one (*synORF*) can result in significant changes in functions if there are some functionally constrained elements in these sequences, which can potentially aid to reveal the logics of synonymous codon choices in eukaryotes.

Secondly, *CYC1* promoter (*pCYC1*) and terminator (*tCYC1*) were used to drive the expression of *synORF* (Figure 1B). This is primarily because not only these are one of the most extensively characterized and widely used regulatory elements in yeast (Guarente et al., 1984; Guarente and Mason, 1983; Pfeifer et al., 1987) but also *CYC1* gene shows considerable differences with essential genes in expression profiles (Table S1). In Table S1, the genes in yeast genome were ranked by scores indicating the similarity of their expression profiles to that of *CYC1,* a score that was generated by the online tool SPELL : https://spell.
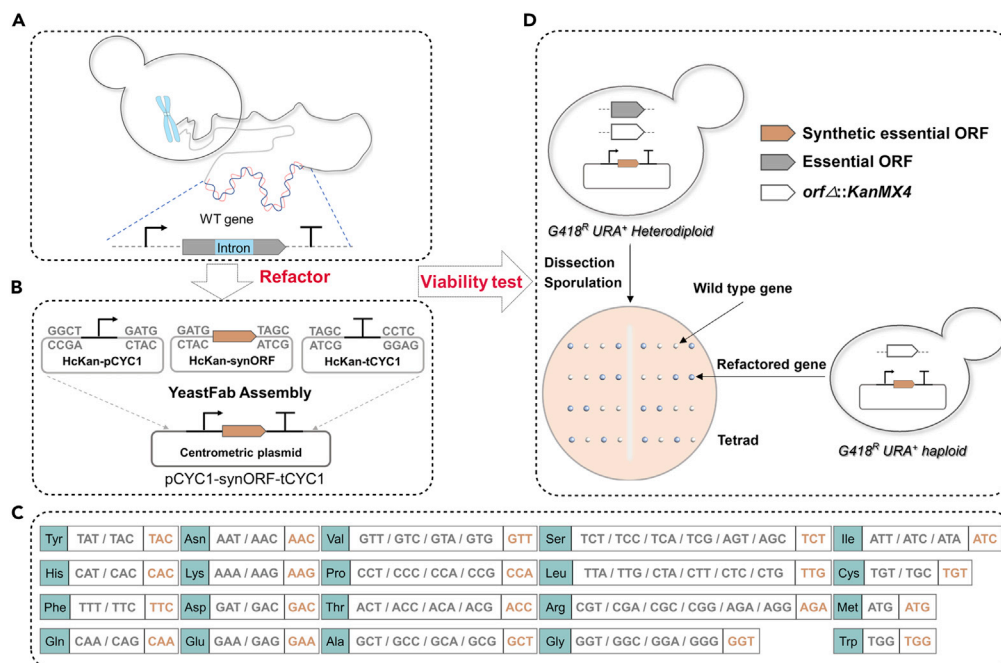
Figure 1. Strategies to refactor the yeast essential genes

(A) Gene is the basic physical and functional units of yeast genome, which are composed of the promoter, ORF, terminator, and intron (only a few genes have introns).

(B) *CYC1* promoter (denoted as *pCYC1*), recoded ORF (denoted as *synORF*), and *CYC1* terminator (denoted as *tCYC1*) were synthesized into the HcKan vector to facilitate the efficient One-pot assembly of *pCYC1-synORF-tCYC1* transcription unit into the centromeric plasmid carrying *URA3* marker gene, a method known as YeastFab assembly (Guo et al., 2015). The sticky ends have been shown in gray.

(C) The three letter names of 20 amino acids (filled by turquoise) were shown with their corresponding codons, in which the optimized ones used in this study were labeled by brown. The optimization process was done using the online tool Codon Juggling : http://54.235.254.95/cgi-bin/gd/gdCodJug.cgi.

(D) The plasmids carrying *pCYC1-synORF-tCYC1* were transformed into the corresponding heterodiploid strains with one copy essential gene being deleted by *KanMX4* for sporulation and tetrad dissection. Both the presence of the viable tetrads and the $G418^R$ URA^+ phenotype of generated haploid strains indicated that the refactored gene was functional to support the viability. Please also refer to Table S1.

yeastgenome.org, currently the largest gene expression microarray dataset for yeast containing roughly 2,400 experimental conditions (Hibbs et al., 2007). The lower the score, the greater the difference and the bigger the rank. The ranks of these ten essential genes were from 314 to 4694, indicating there are 313 genes with more similar expression profiles to *CYC1* than them. The difference between the native regulatory sequence and *CYC1* promoter/terminator can result in defects if there is a gene specific expression profile important for function.

Thirdly, *pCYC1*, *synORF*, and *tCYC1* were synthesized into the HcKan backbone as described before (Guo et al., 2015) to facilitate the efficient assembly of intact refactored gene onto the centromeric plasmid by one-pot reaction (Figure 1B). The modularity of this method allows fast replacement of either part back to the wild-type sequence, which can provide significantly help during the debugging process to reveal the invariable elements. The plasmid location excludes the interference from original chromosome environments, thus helping to reveal the constraints come from chromosomal location.

Finally, functionality of the refactored gene was analyzed through tetrad analysis of heterodiploid strains carrying the corresponding centromeric *pCYC1-synORF-tCYC1* plasmids (Figure 1D). The tetrads can generate four colonies with similar sizes if the refactored genes are with wild-type-like functions under optimal growth condition (YPD at 30°C). Debugging was performed to define the invariable parts and to decipher the mechanisms of gene sequence choice. Further tests can be conducted for the haploid strains using the refactored gene for viability to determine whether defects can be detected under stress conditions.
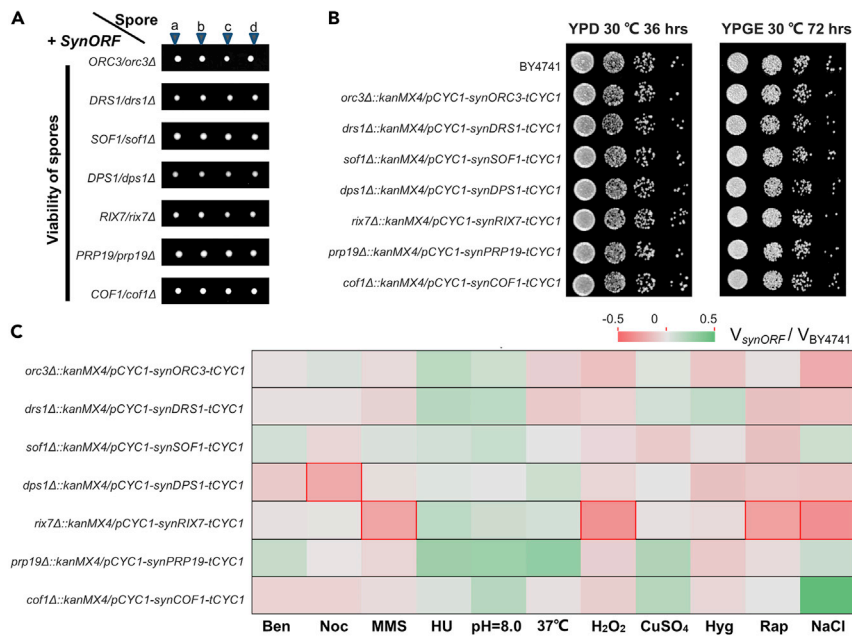
Figure 2. Most essential genes in *Chr.XIIL* could be reprogrammed

(A) The tetrad dissection results of the 7 heterodiploids carrying the *pCYC1-synORF-tCYC1* constructs (denoted as *synORF* in this figure). Spore a, b, c, and d were from the same ascus.

(B) The indicated haploids in log phase were 10-fold serially diluted onto the YPD and YPGE plates and incubated at 30°C for indicated time intervals to monitor their growth.

(C) The maximum growth rates of indicated haploids under various conditions were measured using the method as described before (Lin et al., 2019) and normalized to the maximum growth rate of BY4741 under that specific condition. The statistically significant differences (2-tailed Student's t test, $p < 0.05$) were marked with red borders. 10 μg/mL benomyl (Ben); 5 μg/mL nocodazole (Noc); 0.002% methyl methanesulfonate (MMS); 10 mM hydroxyurea (HU); 1.5 mM hydrogen peroxide (H_2O_2); 6 mM $CuSO_4$; 35 μg/mL hygromycin B (Hyg); and 3.5 ng/mL rapamycin (Rap); 1 M NaCl. Please also refer to Figure S1 and Table S2.

Most essential genes in *Chr.XIIL* could be reprogrammed

As a pilot project, all the 10 essential genes in *Chr.XIIL* (Table S2), which play vital functions in cell cycle, replication, transcription, translation, glycosylphosphatidylinositol (GPI) anchor biosynthesis, and actin filament depolarization were refactored using this strategy.

In average, 21% of the wild-type coding sequences of these 10 genes were recoded (Table S2) and the differences were distributed over entire coding sequence (Figure S1A). Sequence alignments against *pCYC1* done by standard NCBI nucleotide blast program (Zhang et al., 2000) using the upstream 500 bp sequence of these 10 ORFs revealed no significant similarity. Similar results were obtained for the downstream 200 bp sequences of the 10 ORFs. The expression profiles of these 10 genes were also very different from *CYC1*, as there are 313 genes with more similar expression profiles to *CYC1* than them (Table S1). These data revealed considerable variations, including not only the sequence variations in promoter, ORF, and terminator but also the variations in expression profiles, have been introduced by our refactor scheme.

Thereafter, essentiality of these genes was reconfirmed by tetrad analysis of the heterodiploids with each of them being deleted (Figure S1B). The *pCYC1-synORF-tCYC1* plasmid was then transformed into the corresponding strain for sporulation and dissection. Interestingly, 7 out of the 10 *pCYC1-synORF-tCYC1* constructs were found to be able to support the full viability of tetrads (Figure 2A), thereby indicating that essential functions of these genes were retained in these refactored genes. The similar sizes of the four colonies obtained from the same ascus (Figure 2A) and the similar growth of *pCYC1-synORF-tCYC1 orf*Δ haploid strains compared to the wild-type BY4741 strain (Figure 2B) both suggested the wild-type-like functions of these refactored genes. The microscope imaging and flow cytometry analysis also showed these haploid "*pCYC1-synORF-tCYC1 orf*Δ" strains exhibited no significantly morphological alterations when compared to the wild-type BY4741 strain with *pRS416* (Figures S1C and S1D).
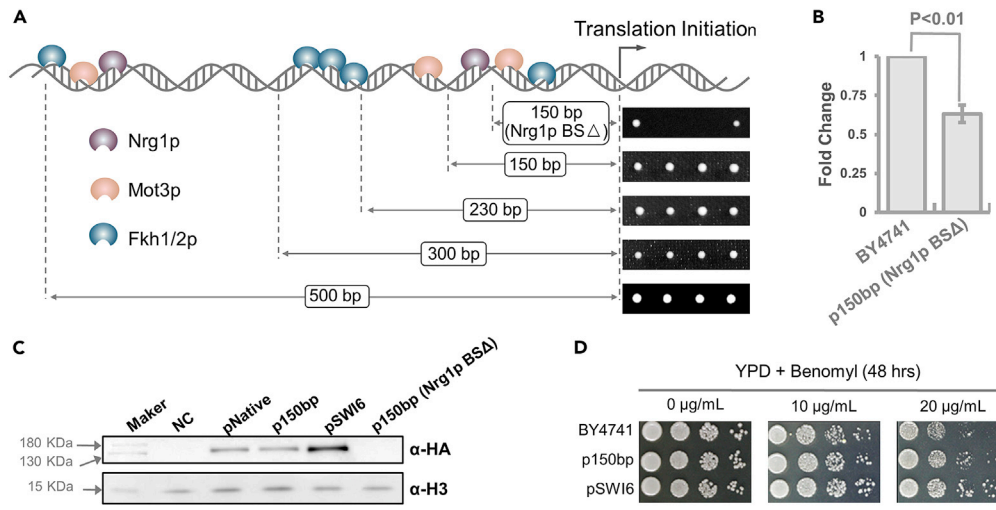
Figure 3. The -150 bp sequences were required for promoter function in *SFI1*

(A) Transcription factor binding sites were predicted using the Yeastract online tool (Monteiro et al., 2020) and only the three transcription factors having binding sites residing in the *p150bp* were shown along the *SFI1* promoter. The dissection results of heterodiploid strains (*SFI1/sfi1Δ*) carrying truncated promoters have been shown at right. Nrg1p BSΔ deletion of the predicted binding site of the transcription factor Nrg1p.

(B) *p150bp (Nrg1p BSΔ)* promoter produced significantly lower *wtSFI1* mRNA level. The cells in the log phase were collected for total mRNA extraction. The detected mRNA level (mean $\pm$ SD) was normalized to the level of actin (*ACT1*). Three biological replicates were measured, P-value was calculated using two-tailed Student's *t* test.

(C) The predicted Nrg1p binding sequence in the 150 bp promoter was found to be important for Sfi1p expression. The haploid strains carrying the indicated promoters and 6HA tags at N-terminus were cultured to the log phase and the expression of Sfi1p was determined by immunoblotting. Histone protein H3 was used as the loading control. NC, BY4741 strain without 6HA-tagged Sfi1p.

(D) *SFI1* expression driven by *pSWI6* enhanced benomyl tolerance. The cells in the log phase were serially diluted and spotted onto YPD medium or YPD medium with indicated concentration of benomyl. Please also refer to Figure S2.

The differential regulation of the gene expression was often important for stress tolerance. To further explore whether any regulatory elements important for stress response were also inactivated in the refactored genes, a high-throughput, semi-quantitative phenotype assay for evaluating the fitness of synthetic yeasts under representative stress conditions was carried out as described previously (Lin et al., 2019). To our surprise, as shown in Figure 2C, no significant differences between the haploid strains *pCYC1-synORF-tCYC1 orfΔ* and the wild-type BY4741 strain were observed under most conditions, thereby suggesting robust functions of these refactored genes, although some important roles of these genes in some of the conditions used have already been reported (Hoepfner et al., 2014; Svensson et al., 2011; van Pel et al., 2013). Only one condition for *DPS1* and four conditions for *RIX7* were identified to cause significant growth delays of the corresponding refactored strains, thus indicating previously undiscovered functions of the two genes in these conditions.

The maintenance of wild-type-like functions in 7 out of the 10 refactored genes clearly reflected the amazing plasticity of yeast essential genes to sequence variations and expression profile changes, thus providing the feasibility to radically engineer yeast genome sequence. For the three genes that could not be refactored, *SFI1*, *GPI13*, and *GRC3*, further experiments were performed to dissect the essential elements in their wild-type sequences.

The −150 bp sequences were required for promoter function in *SFI1*

For *SFI1*, we found that the recoded ORF driven by native promoter sequence (500 bp upstream ATG, pNative) was completely active to support viability (Figure 3A). Further truncation analysis of the native promoter according to the clusters of possible transcription factor binding sites analyzed by Yeastract (Monteiro et al., 2020) suggested that the 150 bp sequence upstream ATG was the minimal requirement for *SFI1*'s function (Figure 3A, *p150bp*). Deletion of the transcription factor Nrg1p-predicted binding site in *p150bp* resulted in about 40% reduction in mRNA level (Figures 3B and S2A–S2C) and leads to an

undetectable Sfi1p-HA level (Figure 3C), and of course, the lethality of the corresponding strain (Figure 3A), thereby suggesting the involvement of this sequence in both transcriptional and translational regulation of *SFI1*. Moreover, deletion of Mot3p-predicted binding site or Fkh1p/Fkh2p-predicted binding site in *p150bp* also resulted in lethality (Figure S2D), but no synthetic negative effect was observed when these four transcription factors were individually deleted in the *p150bp* strain (Figure S2E), thus suggesting the transcription factor-independent importance of this region for regulating the function of *SFI1*. No binding of Fkh1p, Fkh2p, Mot3p, or Nrg1p on the 150 bp promoter of *SFI1* was detected by chromatin immunoprecipitation (Figure S2F). The result of Fkh1p and Fkh2p was consistent with the previous reports (MacIsaac et al., 2006; Mondeel et al., 2019; Ostrow et al., 2014; Venters et al., 2011) and further supported the transcription factor-independent function of this region.

As Sfi1p is a spindle pole body (SPB) protein required for SPB duplication during cell cycle progression (Jaspersen and Winey, 2004), we hypothesized that a cell-cycle-regulated promoter with similar expression profile might support its function. The *SWI6* promoter was selected by searching SPELL database (Hibbs et al., 2007) for the genes with expression profile similar to *SFI1* and can directly regulate cell cycle progression (Dirick et al., 1992). The expression profiles of *SFI1*, *CYC1*, and *SWI6* during cell cycle were extracted from Cyclebase (Santos et al., 2015) and shown in Figures S2G–S2I. To our surprise, the *pSWI6-synSFI1-tCYC1* construct grew not only comparable to BY4741 under optimal condition but even demonstrated better than BY4741 under high benomyl concentration (Figure 3D), thus suggesting a more robust function of the refactored gene, which was possibly due to the increased mRNA (Figure S2J) and protein levels (Figure 3C) of *SFI1* in the refactored strain.

Besides the well-established phosphorylation-mediated functional control of Sfi1p (Avena et al., 2014; Cavanaugh and Jaspersen, 2017), our study also identified its promoter as a vital regulator of its transcription, translation, and function, although the detailed mechanism needs more additional explorations. The robust function of *pSWI6-synSFI1-tCYC1* indicates the possibility of artificially designed genes to outperform the wild-type ones, which is a long-cherished goal of synthetic biology but has encountered many difficulties.

The ORF of *GPI13* has important function in translation regulation besides protein coding

For *GPI13*, we found that the wild-type ORF regulated by *CYC1* promoter and terminator was completely active to support the viability (Figure 4A). Further dissection experiments using the chimeric ORFs with half wild-type sequence and half synthetic sequence revealed that the N-terminal half of *GPI13* ORF was essential for supporting the viability and could not be recoded (Figure S3A). To further dissect the exact defects of recoded ORF, we constructed two diploids that contained one copy wild-type *GPI13* (*wtGPI13*) and one copy recoded *GPI13* (*synGPI13*) in the native *GPI13* loci, either *wtGPI13* or *synGPI13* was tagged with 6HA tag to facilitate the protein level detection (Figure 4B). No obvious mRNA level differences between *wtGPI13* and *synGPI13* were observed in these two strains as indicated by the RT-qPCR results (Figures 4C, S3B, and S3C), but the Western blot results suggested that no *synGPI13* was translated (Figure 4D), thus indicating that this synonymous recoding largely affected the protein level but not the mRNA level of *GPI13*.

To find the mechanisms underlying this defect, several hypothesis-driven experiments were done based on current knowledge about *GPI13* translation regulation. Gpi13p is a protein that acts at ER for protein glycosylation (Orlean and Menon, 2007). Since the seven distinct GCAU elements in the *wtGPI13* mRNA, which are important for interaction with Whi3p (Colomina et al., 2008), were all synonymously recoded in the *synGPI13* mRNA (Figure S3D), we first hypothesized that whether the loss of Whi3p-mediated targeting of *GPI13* mRNA to ER in *synGPI13*, which can play important roles in the regulation of local translation, is the reason for the lack of viability. However, the restoration of these GCAU elements in two new synthetic versions of *GPI13* did not effectively generate viable tetrads (Figure S3E), thereby suggesting this was not the primary reason for loss-of-function in *synGPI13*.

Gpi13p is heavily glycosylated at its ER luminally localized N-terminal part as reported previously by (Flury et al., 2000) and supported by our Western blot results in Figure 4D. In addition, local slowdown of translation by non-optimal codon clusters at N-terminal has been shown to promote nascent-chain recognition by signal-recognition particle (Pechmann et al., 2014), which is important for the process of glycosylation and folding of nascent chain. The *wtGPI13* ORF also contains such N-terminal non-optimal codon clusters, which were recoded in
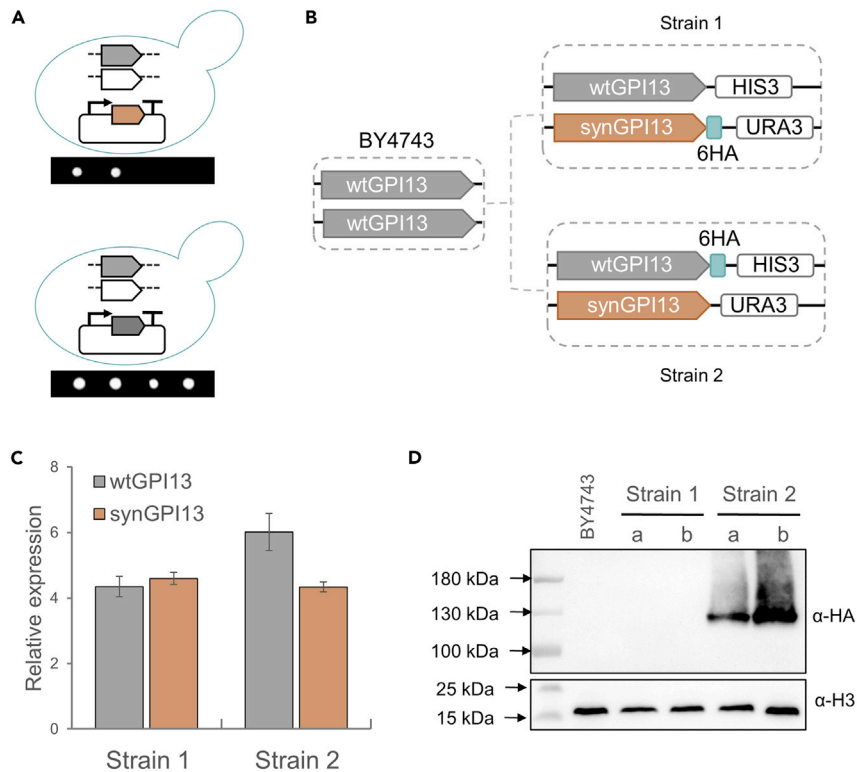
Figure 4. The ORF of *GPI13* has important function in translation regulation besides protein coding

(A) The *pCYC1-wtGPI13-tCYC1* construct could generate viable tetrads. Brown stands for *synGPI13*, gray indicates for *wtGPI13*, and white represents for *gpi13* deletion.

(B) Illustration of the two diploid strains used in Figures 4C and 4D, which were constructed from BY4743.

(C) The ORF recoding did not significantly affect *GPI13* mRNA level. The cells in the log phase were collected for the total mRNA extraction. The detected mRNA level (mean $\pm$ SD) was normalized to the level of actin (*ACT1*). Three biological replicates were performed and the slight differences in *GPI13* mRNA levels between these two strains were possibly caused by HA tagging.

(D) Indicated strains were cultured to the log phase and the expression of Gpi13p was determined by immunoblotting. Histone protein H3 was used as the loading control. a, b represents two individual colonies used. The corresponding protein was detected in Western blots as a major band at about 130 kDa plus a heterogeneously glycosylated smear. Please also refer to Figures S3 and S4.

synGPI13 (Figure S4A). Thus, a translocation assay was performed at first as described previously (Dalley et al., 2008) to evaluate whether the recoded N-terminal 540 bp of *GPI13* possesses the same translocation ability as the wild-type one (Figure S4B). In this assay, the N-terminal sequence of wt*GPI13* was fused to the N-terminal of *URA3* coding sequence (denoted as WT-URA) and directed the translocation of Ura3p to the ER lumen where it cannot access its substrate orotidine-5′-phosphate (OMP) to produce uridine monophosphate (UMP), which resulted in uracil auxotrophy. As shown in Figure S4C, the Syn-URA (the N-terminal sequence of syn*GPI13* was fused to the N-terminal of *URA3* coding sequence) also generated uracil auxotrophy, indicating the recoded N-terminal sequence does not affect the co-translational ER translocation process. Secondly, recoding the N-terminal 540 bp of wt*GPI13* generated viable strains and showed no growth defects on YPD medium at 30°C, although it actually reduced the function of *GPI13* as indicated by the increased sensitivity to calcofluor white (Figure S4D), a common phenotype of *gpi* mutants (Richard et al., 2002). However, replacing the N-terminal 540 bp of syn*GPI13* with wild-type sequence did not produce detectable protein (Figures S4E and S4F) and did not generate fully viable tetrads (Figure S4G). Together, these results suggested that the recoded N-terminal 540 bp was partially functional and other recoding events in the remained N-terminal half also contribute to the lethality.

In summary, all these findings revealed an essential role of the N-terminal half wild type sequence in the post transcriptional regulation of *GPI13* besides its inherent coding capacity. The non-optimal codon clusters in the N-terminal 540 bp can contribute to but is not the only element involved in this regulation

process. Among the currently known 5 genes that could not be recoded [*GPI13*, *MMM1* (Zhang et al., 2017), *TSC10* (Shen et al., 2017), *PRE4* (Mitchell et al., 2017), and *FIP1* (Wu et al., 2017)], the first four are ER-related, thereby indicating the importance of codon choice for ER-related proteins, but the underlying mechanisms need further exploration.

The coding sequence of *GRC3* contains essential components of *PRP19*'s promoter

For *GRC3*, we first found that both the recoded ORF and promoter strength were not the reason for non-viability (Figure S5A). Thereafter, we investigated that whether using a native promoter could be useful. However, only the upstream 1.6 kbp but not the upstream 666 bp could generate fully viable tetrads (Figure 5A), which was much longer than the common 500 bp length of yeast promoters (Guo et al., 2015) and covers the major coding region of another essential gene, *PRP19* (Figure 5B). Given the head-to-head arrangement of *GRC3* and *PRP19*, we hypothesized the deletion of *GRC3* ORF can markedly affect the expression of *PRP19*.

To test our hypothesis, we constructed two different diploids using the similar strategy as *GPI13* (Figure 5C). The Western blot results confirmed our hypothesis that *grc3* deletion can largely deplete the expression of Prp19p (Figure 5D). Further co-expression of *synGRC3* and *synPRP19* in the *GRC3/grc3Δ::KanMX4* diploids or expression *synGRC3* in the *GRC3/grc3* (mATG, with ATG start codon being mutated) diploids was found to generate fully viable tetrads (Figure 5E). The *PRP19* mRNA levels in the BY4741 or *grc3Δ::KanMX4* strains, both of which were co-transformed with *pCYC1-synGRC3-tCYC1* and *pCYC1-synPRP19-tCYC1*, were detected by RT-qPCR and suggested that *grc3* deletion significantly reduced *PRP19*'s mRNA level (Figures S5B and S5C), thus suggesting the coding sequence of *GRC3* contained a transcription activation element to enhance the transcription of *PRP19*, which was essential for its viability supporting function. These results rendered us to recheck the essentiality of *GRC3* by expressing the *PRP19* gene in the *GRC3/grc3Δ::KanMX4* diploid strain (Figure S5D), and the inability to produce four viable spores confirmed the essentiality of *GRC3*.

Two-micron plasmids (pJD1621, pJD1622, and pJD1623) were constructed in which the intergenic region between *GRC3* and *PRP19* or the reserve sequence or random sequence was used to drive the expression of eGFP to check whether it can function as a bidirectional core promoter. Compared to the BY4742 strain containing the negative control plasmid (the random sequence as the promoter of eGFP), the fluorescence intensity of eGFP has increase in strains containing the plasmid pJD1621 or pJD1622 (Figures S5E and S5F). This suggested the 323 bp sequence can drive the expression of genes in both directions.

Our results suggested the intergenic region between *GRC3* and *PRP19* function as a bidirectional core promoter for base-level transcription of these two genes. This finding demonstrates that the gene pairs regulated by the bidirectional promoters can be asymmetrically transcribed, which was meaningful for the bidirectional promoter design and application in metabolic engineering (Vogl et al., 2018).

DISCUSSION

Clarifying the biological significance of every nucleotide in a genome is an ultimate dream of current genome biology. However, we are far from achieving this goal. Even the well-studied model genomes, for example, the *Saccharomyces cerevisiae* genome, appear to be complicated for us. The Yeast genome Deletion Project and further functional profiling (Costanzo et al., 2010, 2016; Giaever et al., 2002; Kuzmin et al., 2018; Winzeler et al., 1999) have depicted a functional blueprint of yeast genes. Moreover, systematic targeted mutations have detailed the importance of each amino acid for regulating the functions of a protein (Dai et al., 2008; Jiang et al., 2017, 2019). However, at the nucleotides level, how the synonymous codons, promoters, and terminators are selected to make a functional gene has not been systematically studied, thereby largely limiting our understanding about the genome sequence and our ability to engineer desired functions. The radical refactor scheme in this study depicts that how the sequence engineering ability enabled by synthetic biology could be functionally applied to explore these questions.

Firstly, when a new protein coding gene is evolved from the already existing ones or built from scratch, two important aspects should be assured: one is that it should be able to encode a defined amino acids sequence, the other is that it should be transcribed and translated. In this simple sense, any promoter and terminator could be used for the gene regulation and only 20 codons plus stop codons can be enough for protein coding. Our results showed that for the 10 essential genes in *Chr.XIIL*, 7 can be radically refactored while maintaining these two important properties, thus claiming an amazing robustness and
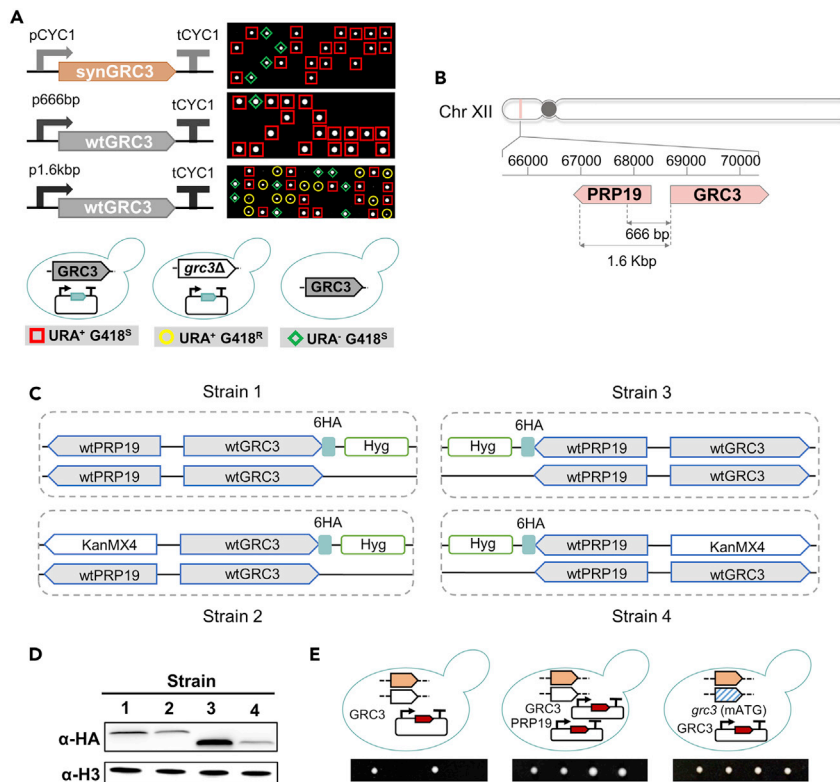
Figure 5. The coding sequence of *GRC3* contains essential components of *PRP19*'s promoter

(A) Only the 1.6 kbp sequence upstream of *GRC3* coding sequence (denoted as p1.6kbp) could be used to drive the expression of *GRC3* to produce the viable tetrads. The indicated plasmids were transformed into the *GRC3/grc3Δ::KanMX4* heterodiploid. Then, sporulation and tetrad analysis were performed for each strain. URA$^+$ means the spore contained the plasmid shown on the left; URA$^-$ means the spore did not contain the plasmid; G418^R referred to the spore in which the ORF of *GRC3* was deleted by *KanMX4*; G418^S referred to spore containing the ORF of *GRC3*. Turquoise denoted the ORF of *wtGRC3* or *synGRC3*.

(B) Illustration of the genomic location of *GRC3* and *PRP19*.

(C) Illustration of the various diploid strains constructed for the protein level detection in Figure 5D.

(D) Deletion of *GRC3* coding sequence largely depleted the expression of *PRP19*. Histone protein H3 was used as the loading control.

(E) Co-expression of *pCYC1-synGRC3-tCYC1* and *pCYC1-synPRP19-tCYC1* in the *GRC3/grc3Δ::KanMX4* diploids or expression *pCYC1-synGRC3-tCYC1* in the *GRC3/grc3* (mATG, with ATG start codon being mutated) diploids generated the completely viable tetrads. Please also refer to Figure S5.

plasticity of yeast genome to sequence variations and expression disturbs. After refactoring all essential genes, we will be able to evaluate how many genes in the yeast genome can be encoded by the only 20 codons and tolerate drastic changes in expression behavior.

Secondly, the invariable sequences residing in the 3 genes that cannot be refactored are quite different, demonstrating the power of our refactoring scheme to identify the different functionally constrained elements in yeast genome. No invariable element in terminator was identified in this study, but a case was reported in our previous study in which identifying the 3'UTR of *ACE2* as an important regulator for ethanol tolerance (Luo et al., 2018a). A genome-wide refactoring will be needed to elucidate some rules for gene encoding and regulation, for example, whether the ER-related proteins require a specified choice of codons. Additional time is needed to analyze the biological mechanism behind these newly identified elements, which will inevitably revolutionize our understanding about genome sequences.

Finally, the results in this study suggested that a much more compacted codon table and quite different regulation sequences may be suitable for a functional yeast genome, which is of interest to be tested by synthetic genomics approaches, just as the Sc3.0 project we proposed recently (Dai et al., 2020).

Limitations of the study

While we have demonstrated the power of synthesis-based gene sequence refactor in the functional elements decoding using the 10 essential genes in *Chr.XIIL*, further expanding of our scheme to all yeast genes will be meaningful, although some functional assays should be specifically designed to understand the functions of the non-essential genes. We anticipate that refactoring all essential genes in the next step might elucidate some underlined mechanisms about how promoter, codon choice, and gene arrangement can be effectively shaped by the evolution process for fine-tuning the function of a gene.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Strains and growth media
- METHOD DETAILS
  - YeastFab assembly of refactored genes
  - Tetrad analysis
  - Flow cytometry analysis
  - Cell morphology observation by microscope
  - RT-qPCR analysis
  - Western blot analysis
  - ChIP-qPCR
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.104982.

AUTHOR CONTRIBUTIONS

J.B.D. conceived the study; J.B.D. and Z.Q.L. designed the study; Z.Q.L., Z.Z.L., W.M.Z., K.Y., H.W., B.N.G., Q.Y., Z.Y.N., C.Z., Y.H.Z., and C.Y.L. performed experiments; Z.Q.L., Z.Z.L., S.H.Y., Y.X.M., and J.B.D. made the figures and wrote the manuscript. All authors contributed to revising the manuscript and approved the final submission.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Avena, J.S., Burns, S., Yu, Z., Ebmeier, C.C., Old, W.M., Jaspersen, S.L., and Winey, M. (2014). Licensing of yeast centrosome duplication requires phosphoregulation of Sfi1. PLoS Genet. 10, e1004666. https://doi.org/10.1371/journal.pgen.1004666.

Carvunis, A.R., Rolland, T., Wapinski, I., Calderwood, M.A., Yildirim, M.A., Simonis, N., Charloteaux, B., Hidalgo, C.A., Barbette, J., Santhanam, B., et al. (2012). Proto-genes and de novo gene birth. Nature 487, 370–374. https://doi.org/10.1038/nature11184.

Cavanaugh, A.M., and Jaspersen, S.L. (2017). Big Lessons from Little yeast: budding and fission yeast centrosome structure, duplication, and function. Annu. Rev. Genet. 51, 361–383. https://doi.org/10.1146/annurev-genet-120116-024733.

Colomina, N., Ferrezuelo, F., Wang, H., Aldea, M., and Garí, E. (2008). Whi3, a developmental regulator of budding yeast, binds a large set of mRNAs functionally related to the endoplasmic reticulum. J. Biol. Chem. 283, 28670–28679. https://doi.org/10.1074/jbc.M804604200.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H.M., Koh, J.L.Y., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. Science 327, 425–431. https://doi.org/10.1126/science.1180823.

Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D., et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. Science 353. https://doi.org/10.1126/science.aaf1420.

Crick, F.H., Barnett, L., Brenner, S., and Watts-Tobin, R.J. (1961). General nature of the genetic code for proteins. Nature 192, 1227–1232. https://doi.org/10.1038/1921227a0.

Curran, K.A., Karim, A.S., Gupta, A., and Alper, H.S. (2013). Use of expression-enhancing terminators in Saccharomyces cerevisiae to increase mRNA half-life and improve gene expression control for metabolic engineering applications. Metab. Eng. 19, 88–97. https://doi.org/10.1016/j.ymben.2013.07.001.

Da Silva, N.A., and Srikrishnan, S. (2012). Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae. FEMS Yeast Res. 12, 197–214. https://doi.org/10.1111/j.1567-1364.2011.00769.x.

Dai, J., Boeke, J.D., Luo, Z., Jiang, S., and Cai, Y. (2020). Sc3.0: revamping and minimizing the yeast genome. Genome Biol. 21, 205. https://doi.org/10.1186/s13059-020-02130-z.

Dai, J., Hyland, E.M., Yuan, D.S., Huang, H.L., Bader, J.S., and Boeke, J.D. (2008). Probing nucleosome function: a highly versatile library of synthetic histone H3 and H4 mutants. Cell 134, 1066–1078. https://doi.org/10.1016/j.cell.2008.07.019.

Dalley, J.A., Selkirk, A., and Pool, M.R. (2008). Access to ribosomal protein Rpl25p by the signal recognition particle is required for efficient cotranslational translocation. Mol. Biol. Cell 19,

2876–2884. https://doi.org/10.1091/mbc.E07-10-1074.

Darnell, J.E., Wall, R., and Tushinski, R.J. (1971). An adenylic acid-rich sequence in messenger RNA of HeLa cells and its possible relationship to reiterated sites in DNA. Proc. Natl. Acad. Sci. USA 68, 1321–1325. https://doi.org/10.1073/pnas.68.6.1321.

Dirick, L., Moll, T., Auer, H., and Nasmyth, K. (1992). A central role for SWI6 in modulating cell cycle Start-specific transcription in yeast. Nature 357, 508–513. https://doi.org/10.1038/357508a0.

Dymond, J.S., Richardson, S.M., Coombes, C.E., Babatz, T., Muller, H., Annaluru, N., Blake, W.J., Schwerzmann, J.W., Dai, J.B., Lindstrom, D.L., et al. (2011). Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature 477, 471–U124. https://doi.org/10.1038/nature10403.

Edmonds, M., Vaughan, M.H., Jr., and Nakazato, H. (1971). Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: possible evidence for a precursor relationship. Proc. Natl. Acad. Sci. USA 68, 1336–1340. https://doi.org/10.1073/pnas.68.6.1336.

Flury, I., Benachour, A., and Conzelmann, A. (2000). YLL031c belongs to a novel family of membrane proteins involved in the transfer of ethanolaminephosphate onto the core structure of glycosylphosphatidylinositol anchors in yeast. J. Biol. Chem. 275, 24458–24465. https://doi.org/10.1074/jbc.M003844200.

Fredens, J., Wang, K., de la Torre, D., Funke, L.F.H., Robertson, W.E., Christova, Y., Chia, T., Schmied, W.H., Dunkelmann, D.L., Beránek, V., et al. (2019). Total synthesis of Escherichia coli with a recoded genome. Nature 569, 514–518. https://doi.org/10.1038/s41586-019-1192-5.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature 418, 387–391. https://doi.org/10.1038/nature00935.

Guarente, L., Lalonde, B., Gifford, P., and Alani, E. (1984). Distinctly regulated tandem upstream activation sites mediate catabolite repression of the CYC1 gene of S. cerevisiae. Cell 36, 503–511. https://doi.org/10.1016/0092-8674(84)90243-5.

Guarente, L., and Mason, T. (1983). Heme regulates transcription of the CYC1 gene of S. cerevisiae via an upstream activation site. Cell 32, 1279–1286. https://doi.org/10.1016/0092-8674(83)90309-4.

Guo, Y., Dong, J., Zhou, T., Auxillos, J., Li, T., Zhang, W., Wang, L., Shen, Y., Luo, Y., Zheng, Y., et al. (2015). YeastFab: the design and construction of standard biological parts for metabolic engineering in Saccharomyces cerevisiae. Nucleic Acids Res. 43, e88. https://doi.org/10.1093/nar/gkv464.

Hanson, G., and Coller, J. (2018). Codon optimality, bias and usage in translation and mRNA decay. Nat. Rev. Mol. Cell Bio. 19, 20–30. https://doi.org/10.1038/nrm.2017.91.

Hibbs, M.A., Hess, D.C., Myers, C.L., Huttenhower, C., Li, K., and Troyanskaya, O.G. (2007). Exploring the functional landscape of gene expression: directed search of large microarray compendia. Bioinformatics 23, 2692–2699. https://doi.org/10.1093/bioinformatics/btm403.

Hoepfner, D., Helliwell, S.B., Sadlish, H., Schuierer, S., Filipuzzi, I., Brachat, S., Bhullar, B., Plikat, U., Abraham, Y., Altorfer, M., et al. (2014). High-resolution chemical dissection of a model eukaryote reveals targets, pathways and gene functions. Microbiol. Res. 169, 107–120. https://doi.org/10.1016/j.micres.2013.11.004.

Jaspersen, S.L., and Winey, M. (2004). The budding yeast spindle pole body: structure, duplication, and function. Annu. Rev. Cell Dev. Bi. 20, 1–28. https://doi.org/10.1146/annurev.cellbio.20.022003.114106.

Jiang, Q., Zhang, W., Liu, C., Lin, Y., Wu, Q., and Dai, J. (2019). Dissecting PCNA function with a systematically designed mutant library in yeast. J. Genet. Genomics 46, 301–313. https://doi.org/10.1016/j.jgg.2019.03.014.

Jiang, S., Liu, Y., Wang, A., Qin, Y., Luo, M., Wu, Q., Boeke, J.D., and Dai, J. (2017). Construction of comprehensive dosage-matching core histone mutant Libraries for Saccharomyces cerevisiae. Genetics 207, 1263–1273. https://doi.org/10.1534/genetics.117.300450.

Kuzmin, E., VanderSluis, B., Wang, W., Tan, G., Deshpande, R., Chen, Y., Usaj, M., Balint, A., Mattiazzi Usaj, M., van Leeuwen, J., et al. (2018). Systematic analysis of complex genetic interactions. Science 360. https://doi.org/10.1126/science.aao1729.

Lajoie, M.J., Kosuri, S., Mosberg, J.A., Gregg, C.J., Zhang, D., and Church, G.M. (2013). Probing the Limits of genetic recoding in essential genes. Science 342, 361–363. https://doi.org/10.1126/science.1241460.

Lee, S.Y., Mendecki, J., and Brawerman, G. (1971). A polynucleotide segment rich in adenylic acid in the rapidly-labeled polyribosomal RNA component of mouse sarcoma 180 ascites cells. Proc. Natl. Acad. Sci. USA 68, 1331–1335. https://doi.org/10.1073/pnas.68.6.1331.

Lifton, R.P., Goldberg, M.L., Karp, R.W., and Hogness, D.S. (1978). The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications. Cold Spring Harb Symp Quant Biol. 42, 1047–1051. https://doi.org/10.1101/sqb.1978.042.01.105.

Lin, Y., Zou, X., Zheng, Y., Cai, Y., and Dai, J. (2019). Improving chromosome synthesis with a semiquantitative phenotypic assay and refined assembly strategy. ACS Synth. Biol. 8, 2203–2211. https://doi.org/10.1021/acssynbio.8b00505.

Luo, Z., Wang, L., Wang, Y., Zhang, W., Guo, Y., Shen, Y., Jiang, L., Wu, Q., Zhang, C., Cai, Y., et al. (2018a). Identifying and characterizing SCRaMbLEd synthetic yeast using ReSCuES. Nat. Commun. 9, 1930. https://doi.org/10.1038/s41467-017-00806-y.

Luo, Z., Yang, Q., Geng, B., Jiang, S., Yang, S., Li, X., Cai, Y., and Dai, J. (2018b). Whole genome

engineering by synthesis. Sci. China Life Sci. *61*, 1515–1527. https://doi.org/10.1007/s11427-018-9403-y.

Maclsaac, K.D., Wang, T., Gordon, D.B., Gifford, D.K., Stormo, G.D., and Fraenkel, E. (2006). An improved map of conserved regulatory sites for Saccharomyces cerevisiae. BMC Bioinf. *7*. https://doi.org/10.1186/1471-2105-7-113.

Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional regulatory elements in the human genome. Annu. Rev. Genom. Hum G *7*, 29–59. https://doi.org/10.1146/annurev.genom.7.080505.115623.

Mitchell, L.A., Wang, A., Stracquadanio, G., Kuang, Z., Wang, X., Yang, K., Richardson, S., Martin, J.A., Zhao, Y., Walker, R., et al. (2017). Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. Science *355*. https://doi.org/10.1126/science.aaf4831.

Mondeel, T., Holland, P., Nielsen, J., and Barberis, M. (2019). ChIP-exo analysis highlights Fkh1 and Fkh2 transcription factors as hubs that integrate multi-scale networks in budding yeast. Nucleic Acids Res. *47*, 7825–7841. https://doi.org/10.1093/nar/gkz603.

Monteiro, P.T., Oliveira, J., Pais, P., Antunes, M., Palma, M., Cavalheiro, M., Galocha, M., Godinho, C.P., Martins, L.C., Bourbon, N., et al. (2020). YEASTRACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts. Nucleic Acids Res. *48*, D642–D649. https://doi.org/10.1093/nar/gkz859.

Orlean, P., and Menon, A.K. (2007). Thematic review series: lipid posttranslational modifications. GPI anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and love glycophospholipids. J. Lipid Res. *48*, 993–1011. https://doi.org/10.1194/jlr.R700002-JLR200.

Ostrow, A.Z., Nellimoottil, T., Knott, S.R., Fox, C.A., Tavaré, S., and Aparicio, O.M. (2014). Fkh1 and Fkh2 bind multiple chromosomal elements in the S. cerevisiae genome with distinct specificities and cell cycle dynamics. PLoS One *9*, e87647. https://doi.org/10.1371/journal.pone.0087647.

Parenteau, J., Maignon, L., Berthoumieux, M., Catala, M., Gagnon, V., and Abou Elela, S. (2019). Introns are mediators of cell response to starvation. Nature *565*, 612. https://doi.org/10.1038/s41586-018-0859-7.

Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.I., Cooper, G.M., et al. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. Nat. Biotechnol. *30*, 265. https://doi.org/10.1038/nbt.2136.

Pechmann, S., Chartron, J.W., and Frydman, J. (2014). Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. Nat. Struct. Mol. Biol. *21*, 1100–1105. https://doi.org/10.1038/nsmb.2919.

Pfeifer, K., Arcangioli, B., and Guarente, L. (1987). Yeast HAP1 activator competes with the factor RC2 for binding to the upstream activation site UAS1 of the CYC1 gene. Cell *49*, 9–18. https://doi.org/10.1016/0092-8674(87)90750-1.

Plotkin, J.B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. *12*, 32–42. https://doi.org/10.1038/nrg2899.

Richard, M., de Groot, P., Courtin, O., Poulain, D., Klis, F., and Gaillardin, C. (2002). GPI7 affects cell-wall protein anchorage in Saccharomyces cerevisiae and Candida albicans. Microbiology (Read.) *148*, 2125–2133. https://doi.org/10.1099/00221287-148-7-2125.

Richardson, S.M., Wheelan, S.J., Yarrington, R.M., and Boeke, J.D. (2006). GeneDesign: rapid, automated design of multikilobase synthetic genes. Genome Res. *16*, 550–556. https://doi.org/10.1101/gr.4431306.

Santos, A., Wernersson, R., and Jensen, L.J. (2015). Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. Nucleic Acids Res. *43*, D1140–D1144. https://doi.org/10.1093/nar/gku1092.

Sauna, Z.E., and Kimchi-Sarfaty, C. (2011). Understanding the contribution of synonymous mutations to human disease. Nat. Rev. Genet. *12*, 683–691. https://doi.org/10.1038/nrg3051.

Sharp, P.M., Cowe, E., Higgins, D.G., Shields, D.C., Wolfe, K.H., and Wright, F. (1988). Codon usage patterns in Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster and Homo sapiens; a review of the considerable within-species diversity. Nucleic Acids Res. *16*, 8207–8211. https://doi.org/10.1093/nar/16.17.8207.

Sharp, P.M., Tuohy, T.M., and Mosurski, K.R. (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. *14*, 5125–5143. https://doi.org/10.1093/nar/14.13.5125.

Shen, Y., Wang, Y., Chen, T., Gao, F., Gong, J., Abramczyk, D., Walker, R., Zhao, H., Chen, S., Liu, W., et al. (2017). Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. Science *355*. https://doi.org/10.1126/science.aaf4791.

Svensson, J.P., Pesudo, L.Q., Fry, R.C., Adeleye, Y.A., Carmichael, P., and Samson, L.D. (2011).

Genomic phenotyping of the essential and non-essential yeast genome detects novel pathways for alkylation resistance. BMC Syst. Biol. *5*, 157.

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell *141*, 344–354. https://doi.org/10.1016/j.cell.2010.03.031.

van Pel, D.M., Stirling, P.C., Minaker, S.W., Sipahimalani, P., and Hieter, P. (2013). Saccharomyces cerevisiae genetics predicts candidate therapeutic genetic interactions at the mammalian replication fork. G3 *3*, 273–282. https://doi.org/10.1534/g3.112.004754.

Venters, B.J., Wachi, S., Mavrich, T.N., Andersen, B.E., Jena, P., Sinnamon, A.J., Jain, P., Rolleri, N.S., Jiang, C., Hemeryck-Walsh, C., et al. (2011). A comprehensive genomic binding map of gene and chromatin regulatory proteins in Saccharomyces. Mol. Cell *41*, 480–492. https://doi.org/10.1016/j.molcel.2011.01.015.

Vogl, T., Kickenweiz, T., Pitzer, J., Sturmberger, L., Weninger, A., Biggs, B.W., Köhler, E.M., Baumschlager, A., Fischer, J.E., Hyden, P., et al. (2018). Engineered bidirectional promoters enable rapid multi-gene co-expression optimization. Nat. Commun. *9*, 3589. https://doi.org/10.1038/s41467-018-05915-w.

Wang, K., Fredens, J., Brunner, S.F., Kim, S.H., Chia, T., and Chin, J.W. (2016). Defining synonymous codon compression schemes by genome recoding. Nature *539*, 59–64. https://doi.org/10.1038/nature20124.

Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., et al. (1999). Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science *285*, 901–906. https://doi.org/10.1126/science.285.5429.901.

Wu, Y., Li, B.Z., Zhao, M., Mitchell, L.A., Xie, Z.X., Lin, Q.H., Wang, X., Xiao, W.H., Wang, Y., Zhou, X., et al. (2017). Bug mapping and fitness testing of chemically synthesized chromosome X. Science *355*. https://doi.org/10.1126/science.aaf4706.

Zhang, W., Zhao, G., Luo, Z., Lin, Y., Wang, L., Guo, Y., Wang, A., Jiang, S., Jiang, Q., Gong, J., et al. (2017). Engineering the ribosomal DNA in a megabase synthetic chromosome. Science *355*. https://doi.org/10.1126/science.aaf3981.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. J. Comput. Biol. *7*, 203–214. https://doi.org/10.1089/10665270050081478.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse mono-cloned α-HA	Sigma-Aldrich	Cat#H3663; RRID:AB_262051
Histone H3 Mouse Monoclonal Antibody	Beyotime	Cat#AF0009; RRID:AB_2715593
Goat anti-mouse IgG(H&L)-HRP Conjugated	Easybio	Cat#BE0102; RRID:AB_2923205
Monoclonal ANTI-FLAG® M2 antibody	Sigma-Aldrich	Cat# F1804; RRID:AB_262044
Chemicals, peptides, and recombinant proteins		
BsmBI	NEB	Cat#R0580
10× T4 ligase Buffer	Thermo scientific	Cat#B69
T4 DNA ligase	Thermo scientific	Cat#EL0011
Trizol reagent	Invitrogen	Cat#10296028
Triton X-100	Sigma-Aldrich	Cat#T8787
Benomyl	Sigma-Aldrich	Cat#17804-35-2
Nocodazole	Macklin	Cat# N863524
Methyl methanesulfonate (MMS)	Millipore	Cat#820775
Hydroxyurea(HU)	Sigma	Cat# H8627
Hygromycin B, Streptomyces sp.	Millipore	Cat#400052
Rapamycin	Aladdin	Cat#S293790
Nourseothricin	Solarbio®	Cat#N9210
Formaldehyde	Sigma	Cat#F8775-500ML
Micrococcal nuclease	NEB	Cat# M0247S
Protein A/G Magnetic Beads	MCE	Cat#HY-K0202
Proteinase K	Solarbio®	Cat#1245680100-100mg
Rnase A	Solarbio®	Cat#R1030
Calcofluor white	MaoKangbio	Cat# MM1011
Critical commercial assays		
PrimeScript™ RT reagent Kit with gDNA Eraser	TaKaRa	Cat#RR047A
DNA Clean & Concentrator®-5	Zymo research	Cat#D4014
TransStart® Top Green qPCR SuperMix	Transgen	Cat#AQ132-11
Experimental models: Organisms/strains		
S. cerevisiae, strain background: BY4741 and BY4743	EUROSCARF	http://www.euroscarf.de/index.php?name=News
Oligonucleotides		
For all PCR Primers, see Table S5	This paper	N/A
Recombinant DNA		
For all plasmids, see Table S4	This paper	N/A
Other		
SPELL database	(Hibbs et al., 2007)	https://spell.yeastgenome.org
Yeastract	(Monteiro et al., 2020)	http://www.yeastract.com/
Cyclebase	(Santos et al., 2015)	https://cyclebase.org/CyclebaseSearch
Codon Juggling	(Richardson et al., 2006)	http://54.235.254.95/cgi-bin/gd/gdCodJug.cgi

RESOURCE AVAILABILITY

Lead contact

Further information and request for resources should be directed to and will be fulfilled by the lead contact, Junbiao Dai (junbiao.dai@siat.ac.cn).

Materials availability

This study did not generate any new unique reagents. All the requests for the generated plasmids and strains should be directed to the lead contact and will be made available on request after completion of a Materials Transfer Agreement.

Data and code availability

- The published article includes all datasets generated or analyzed during this study.

- This paper does not report any original code.

- Any additional information required to reanalyze the data reported in this paper is available form the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Strains and growth media

The yeast strains used in this study were derivates of BY4741(*MATa his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*) and BY4743 (*MATa/alpha his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 ura3Δ0/ura3Δ0 met15Δ0/MET15 lys2Δ0/LYS2*). Standard methods for yeast culture, gene disruption and transformation were applied. The strains used in this study were list in Table S3.

METHOD DETAILS

YeastFab assembly of refactored genes

The assembly process was performed as described previously (Guo et al., 2015). In brief, the three synthetic plasmids including HcKan-pCYC1, HcKan-synORF and HcKan-tCYC1 were mixed with the POT receiving vector together with 5 U of Bsm*BI* in 1× T4 DNA ligase buffer and incubated at 55°C for 1 h. Thereafter, 0.5 U of T4 DNA ligase was added into the mixture and the mixture was incubated at 25°C for additional 1 h. After reaction, the enzymes were inactivated by incubating at 50°C for 5 min and 80°C for 10 min. Then the transformation was done and properly assembled plasmids were identified by colony PCR and restriction enzyme digestion.

Tetrad analysis

The diploids were patched onto an appropriate medium and grown at 30°C overnight. The fresh cells were scratched and washed using 1 mL sterile water twice. Thereafter, the washed cells were added into 2 mL sporulation medium (10 g/L potassium acetate, 0.05 g/L zinc acetate dihydrate) to a final optical density of 1.0 OD/ml. The solution was incubated at 25°C for 3–10 days. The asci were then dissected onto the YPD plate. After the growth at 30°C for appropriate time, the cells were replicated onto the various selective media to identify their auxotrophs and mating types.

Flow cytometry analysis

Inoculated the strains into SC-URA medium and grew at 30°C overnight. Diluted the culture with fresh SC-URA medium to OD_{600} = 0.1 and further grew for 5–6 h with shaking. Collected the cells by centrifugation at 6000 rpm for 1 min. Resuspended the pellets with 70% ethanol at 4°C overnight. Pelleted cells by centrifugation and resuspended them with 50 mM sodium citrate (pH = 7.0), sonicated the suspension on ice. Repeated the wash process, resuspended the cells with 50 mM sodium citrate with 0.25 mg/mL RNase A and incubated at 37°C overnight. Pelleted cells by centrifugation, washed with 50 mM sodium citrate (pH = 7.0) again and resuspended the cells with 50 mM sodium citrate (pH = 7.0) supplemented with propidium iodide, incubated at room temperature for 30 min.

The promoter activity of the intergenic region between *GRC3* and *PRP19* was measured by flow cytometry. Inoculated the strains into SC-URA medium and grew at 30°C overnight. Diluted the culture with fresh

SC-URA medium to OD_{600} = 0.1 and further grew for 5–6 h with shaking. The eGFP fluorescent intensity were detected by the 488 nm laser.

Cell morphology observation by microscope

Inoculate the strains into SC-URA medium and grew at 30°C overnight. Diluted the culture with fresh SC-URA medium to OD_{600} = 0.1 and further grew for 5–6 h with shaking. Collected the cells by centrifugation at 3000 rpm for 1 min. Washed twice with sterile water. Then the cells were visualized using the Nikon A1 confocal microscope under an oil immersion 60× objective.

RT-qPCR analysis

Total RNA was extracted using TRIzol reagent according to the manufacturer's instruction. To eliminate genomic DNA contamination and obtain cDNA, 1μg of total RNA were pretreated with gDNA Eraser and then reverse-transcribed into cDNA using PrimeScript^{TM} RT reagent Kit with gDNA Eraser. Quantitative Real-time PCR was performed with TransStart® Top Green qPCR Super-Mix. *ACT1* was used as an internal control for normalization. The standard curves for each primer pair were generated using a gradient dilution of a mixture sample cDNAs as the template. The amplification efficiency of each primer pair (Eff) was determined by standard curves using the formula 10(-1/slope). The differences in primer pair amplification efficiencies between the target and reference genes were considered when the relative expression levels of genes were calculated. The relative expression levels of *SFI1* and *PRP19* were determined using the following equation

$$N = \frac{(1 + \text{Eff target gene})^{\Delta Ct(\text{target gene})}}{(1 + \text{Eff reference gene})^{\Delta Ct(\text{reference gene})}}.$$

The relative expression levels of *GPI13* were determined using the equation

$$N = 10000 \times \frac{(1 + \text{Eff target gene})^{Ct(\text{target gene})}}{(1 + \text{Eff reference gene})^{Ct(\text{reference gene})}}.$$

Western blot analysis

A single colony was picked and placed into 5 mL proper liquid medium, cultured at 30°C with shaking at 220 rpm overnight. The overnight culture was diluted into a total 5 mL culture to OD_{600} = 0.1. The diluted medium was cultured at 30°C for another 8 h and the cells were collected by 3000 rpm centrifugation for 3 min. The cells were resuspended using 1 mL sterile water and the suspension was transferred to a 1.5 mL EP tube, 12000 rpm centrifugation for 1 min. For the detection of Gpi13p level, the supernatant was discarded, 50 μL sterile water and 50 μL 0.2 M NaOH was added, and the solution was mixed by vortexing. The solution was maintained at the room temperature for 5 min, followed by 12000 rpm centrifugation for 1 min, after which the supernatant was discarded. Thereafter, added 100 μL SDS sample buffer with 1% triton X-100, followed by addition of 50 μL glass beads, broken at 2500 rpm for 1 min 30 s in gDNA prep machine, then placed the tube onto ice immediately for 5 min, and repeated the procedure for at least 3 times. The mixture was then centrifuged at 14680 rpm for 10 min and the supernatant was aliquoted as the total protein extraction. Usually, 10 μL per lane for 10-lanes gel was used to detect Gpi13p-HA, 2.5 μL per lane for 10-lanes gel to detect H3.

For protein level detection of Sfi1p, procedure similar to the Western blot analysis of Gpi13p was carried out, except for the cells suspended by 100 μL SDS sample were broken directly by boiling.

For protein level detection of Grc3p and Prp19p, procedure similar to the Western blot analysis of Gpi13p, except for no 1% triton X-100 was used and the cells suspended by 100 μL SDS sample were broken directly by boiling.

ChIP-qPCR

The strains were cultured in YPD medium at 30°C overnight. Diluted the culture with fresh YPD medium to OD_{600} = 0.1 and further grew for 5–6 h with shaking until the OD_{600} = 0.8–1.0. Then the cells were cross-linked by formaldehyde (1% final) at 25°C for 25min. Glycine (0.15 M final) was added to quench the formaldehyde. 85 OD of cells were collected and broken by glass beads. The chromatin sonicated by 0.5 uL micrococcal nuclease (MNase), 1/9 of the sonicated chromatin was used as the input and 8/9 were

incubated with 2 uL anti-flag antibody overnight. The antibody-protein-DNA complex was put down by Protein A/G Magnetic beads. Reversed the protein-DNA crosslinks by 2.5 uL 10 mg/mL Rnase A and 5 uL 20 mg/mL Proteinase K. The DNA was purified by the kit DNA Clean & Concentrator-5. ChIP-qPCR was performed to determine the protein occupancy. All the occupancy data were shown as the percentage of enrichment at target loci normalized by negative control sequence (the ORF region (+2499/+2716) of *POL1*) and calculated using the following equation:

$$N = \frac{2^{\Delta Ct\ (\text{target gene})}}{2^{\Delta Ct\ (\text{negative gene})}}, \quad \Delta Ct = Ct\ (\text{input}) - Ct\ (\text{IP}).$$

QUANTIFICATION AND STATISTICAL ANALYSIS

Bars and error bars represent mean $\pm$ SD. The data was analyzed using MS-Excel. Two-tailed student t-tests were used to compare between the different groups. Mean differences were considered as statistically significant at p value $\leq$ 0.05.