RNAi expression tuning, microfluidic screening, and genome recombineering for improved protein production in *Saccharomyces cerevisiae*

Guokun Wang[a], Sara M. Björk[b,c], Mingtao Huang[a,d], Quanli Liu[a], Kate Campbell[a,d], Jens Nielsen[a,d,e], Haakan N. Joensson[b,c,1], and Dina Petranovic[a,d,1]

[a]Department of Biology and Biological Engineering, Chalmers University of Technology, SE41296 Gothenburg, Sweden; [b]Science for Life Laboratory, Division of Protein Science, School of Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, SE17165 Solna, Sweden; [c]Novo Nordisk Foundation Center for Biosustainability, KTH Royal Institute of Technology, SE17165 Solna, Sweden; [d]Novo Nordisk Foundation Center for Biosustainability, Chalmers University of Technology, SE41296 Gothenburg, Sweden; and [e]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK2800 Kongens Lyngby, Denmark

The cellular machinery that supports protein synthesis and secretion lies at the foundation of cell factory-centered protein production. Due to the complexity of such cellular machinery, the challenge in generating a superior cell factory is to fully exploit the production potential by finding beneficial targets for optimized strains, which ideally could be used for improved secretion of other proteins. We focused on an approach in the yeast *Saccharomyces cerevisiae* that allows for attenuation of gene expression, using RNAi combined with high-throughput microfluidic single-cell screening for cells with improved protein secretion. Using direct experimental validation or enrichment analysis-assisted characterization of systematically introduced RNAi perturbations, we could identify targets that improve protein secretion. We found that genes with functions in cellular metabolism (*YDC1*, *AAD4*, *ADE8*, and *SDH1*), protein modification and degradation (*VPS73*, *KTR2*, *CNL1*, and *SSA1*), and cell cycle (*CDC39*), can all impact recombinant protein production when expressed at differentially downregulated levels. By establishing a workflow that incorporates Cas9-mediated recombineering, we demonstrated how we could tune the expression of the identified gene targets for further improved protein production for specific proteins. Our findings offer a high throughput and semirational platform design, which will improve not only the production of a desired protein but even more importantly, shed additional light on connections between protein production and other cellular processes.

droplet microfluidic screening | genome recombineering | protein production | RNA interference | *Saccharomyces cerevisiae*

Protein synthesis and secretion are central cellular processes that involve polypeptide formation, protein folding, modification, trafficking, and export to the extracellular space (1). These processes are carried out in several cellular compartments or complexes, such as on ribosomes [bound to the endoplasmic reticulum (ER)], in the lumen of the ER, in vesicles transporting to the Golgi apparatus, in the lumen of the Golgi apparatus, in vesicles transporting to the plasma membrane, and at the plasma membrane where proteins are released. As such, these processes closely interact with other biological processes in the cell, including those that supply energy and precursors, and/or coordinate redox homeostasis and signaling (2). Alterations in protein synthesis and secretion may lead to a complex array of dysfunctions such as protein accumulation and aggregation, and cell death, and consequently are associated with many human diseases, including neurodegeneration (3, 4). Protein synthesis and secretion are at the base of recombinant protein production and have been exploited for the development of various cell factories, such as yeast (5), filamentous fungi (6), and mammalian cells (7). To construct cell factories with optimized protein production, it is important to identify nodes in cellular function that can be reengineered for this purpose. In doing so, protein production can be enhanced, and the link between protein synthesis and the secretion process of the cellular machinery can be better understood.

Due to its dual function as a model eukaryal organism and as a cell factory for protein production, the budding yeast *Saccharomyces cerevisiae* has been comprehensively characterized (8–11), including its protein production (12–14). However, due to the complexity of the cellular interaction network, the linkage between gene function and protein production cannot be reconciled via overexpression (15) or deletion (16) of single genes using hypothesis-driven or random gene targeting alone. For example, the role of modulation of gene expression in the cellular machinery for protein production is still largely unknown, even though its potential has been harnessed in strain improvement via mutagenesis (14), which introduces various hard-to-track causal and noncausal genetic mutations. Gene attenuation, an alternative approach, reduces the function of a specific gene product and in turn may perturb its gene interaction network. Therefore,

Significance

Recombinant proteins are broadly used in many applications, as medicines or enzymes, etc. One of the most common cell factories that produces many proteins is yeast *Saccharomyces cerevisiae*. Synthesis, posttranslational modifications, and secretion of the protein are crucial steps for optimization but rational design is often challenging due to the complexity of the system and interactions between many cellular processes. Here we dissect the genetic basis of gene down-regulation for improved protein production in yeast, via microfluidics-assisted tracking of systematically introduced RNAi perturbations, and establish a workflow for identification and combinatorial engineering of favorable gene targets. The knowledge and the recombineering approach will aid the understanding of protein-producing machinery and enable strain improvement.

unlike enhancement or complete abolition of gene function, gene attenuation allows for dose-dependent modulation which may exert a positive effect on protein production, and can thus be of interest in pursuing superior cell behavior for improved protein production and identifying beneficial down-regulation targets, including essential genes.

Here we employ droplet microfluidics-based single-cell analysis to evaluate ~243,000 knockdown effectors in *S. cerevisiae*, with the enhanced secretion of the model protein α-amylase as an indicator of improved recombinant protein production (Fig. 1). Through systematic analysis, trackable causal genetic perturbations were identified to have functions in diverse biological processes, including metabolism, protein modification and degradation, and cell cycle, and were shown to have different impacts on protein production when their expression was varied. Using Cas9-mediated recombineering, the expression of identified genes was fine tuned to generate strains with improved protein production capabilities (Fig. 1).

Results

Gene Down-Regulation by RNA Interference Can Modify Protein Production in a Yeast Cell Factory.

To employ gene expression attenuation for improving protein production, we first verified the functionality of RNA interference (RNAi) for enhancing protein production in *S. cerevisiae*. We introduced genes *AGO1* and *DCR1* from *Saccharomyces castelli*, into the genome of *S. cerevisiae* CEN.PK113-11C strain to reconstitute the RNAi machinery in this RNAi-negative strain. We then validated the RNAi functionality in this strain by testing the down-regulation of green fluorescent protein (GFP) (*SI Appendix*, Fig. S1) and reconstructed the strain with high amylase expression levels by genomic integration of the amylase gene from *Aspergillus oryzae* into Ty4 long terminal repeat retrotransposon sites [delta (δ) sites], after identifying the stabilization of gene expression in these δ-sites by the RNAi machinery (*SI Appendix*, Fig. S2). Amylase from *A. oryzae* is a three-domain protein with 478 amino

Fig. 1. Schematic workflow of the microfluidic droplet screening process for yeast RNA interference libraries and genome recombineering for improved protein production. Yeast RNAi libraries with different knockdown levels were generated by introducing plasmid mixtures carrying high or low knockdown cassette libraries into strains expressing α-amylase and reconstructed RNAi machinery. Single cells from these yeast libraries were encapsulated to form microfluidic droplets, which were then subjected to fluorescence-based sorting. Functional RNAi targets within the sorted yeast individuals were characterized through evaluation by functional enrichment and reverse validation. The identified targets were then employed in the semirational engineering by Cas9-mediated genome-scale multiplexing for improved protein production.

acids and four disulfide bonds and serves as a model recombinant protein for production of larger secreted proteins with post-translational modifications in yeast (13). With the 11C-GK3 strain integrated with ~14 copies of the amylase gene (as estimated by qPCR, Fig. 2*A*), we examined the feasibility of perturbing protein producing capability via RNAi, by manipulating expression of several genes, which were reported previously to improve protein production, in particular by overexpression [protein disulfide isomerase *PDI1* (17), basic leucine zipper transcription factor *HAC1* (18)] or deletion [histone deacetylase *HDA2* (19) and a putative oxidoreductase involved in protein transport *TDA3* (20)]. Knockdowns of *PDI1* and *TDA3*, via the complete reverse coding sequence-assisted pattern (21), significantly decreased (55.7%) and intensified (90.7%) amylase production, respectively, while down-regulations of *HAC1* or *HDA2* had no effect (Fig. 2*B*). Furthermore, by using another construct previously reported to have a lower knockdown efficiency (21), we generated two additional strains with low level/moderate down-regulations of *TDA3*. Unlike the promotive effect conferred by the high level knockdown of *TDA3*, the low level/moderate down-regulations of *TDA3* did not have any impact on amylase production (*SI Appendix*, Fig. S3), with expression levels validated using qPCR in a primer/DNA region-dependent manner (*SI Appendix*, Fig. S4). These results suggest that the RNAi machinery can be used to alter protein producing capability of yeast in a dosage-associated manner and that RNAi-driven gene down-regulation potentially can be harnessed to identify novel gene targets with favorable effects on protein production.

Library Construction and Screening for RNAi Gene Targets for Improved Protein Production.

Library design and construction. Considering the dosage effect, we constructed two RNAi cassette libraries expected to have different knockdown efficiencies (Fig. 1): low knockdown library (LKD-lib) and high knockdown library (HKD-lib), and generated the yeast libraries YeaLKD-lib and YeaHKD-lib accordingly, to achieve high coverage on functional gene targets. The LKD and HKD constructs always yielded a comparative difference in the knockdown efficiency for a specific gene. In the YeaLKD-lib, the short double-stranded RNA formed under control of convergent promoters was processed by AGO1 and DCR1 proteins to generate smaller antisense RNA or miRNA, for further RNA targeting and digestion, resulting in modest down-regulation of target genes in yeast cells (21). However, in the YeaHKD-lib, driven by a constitutive promoter, the complete reverse coding sequence was transcribed and the reverse mRNA was bound to the mRNA for digestion, thus down-regulating expression of specific genes at a high level (21).

We constructed the LKD-lib and HKD-lib with 200,000 and 40,300 individuals, respectively, and evaluated the coverage by sequencing the plasmids from random transformants. Sequence hits of 40 (LKD-lib) and 20 (HKD-lib) individuals (Dataset S1, sheets 1 and 2) showed a good coverage on 14/16 and 9/16 chromosomes, respectively.

Droplet microfluidic screening. To enrich yeast variant cells with improved amylase production from the RNAi libraries, we next employed droplet microfluidics, a single-cell analysis technology, focusing on extracellular characteristics (secreted amylase) to

Fig. 2. Down-regulation of gene expression alters amylase production in yeast. The amylase copy number (*A*) and the effect of various high-level gene down-regulations on amylase production (*B*) were characterized in the control 11C-GK3 strain with reconstructed RNAi mechanism and amylase expression cassettes. The amylase production performance of individual cells from the sorted YeaLKD-lib (*C*) or YeaHKD-lib (*D*) is shown. Each dot represents data from an individual strain and dots in black/gold/orange indicate those selected for reverse validation, in which gold (*C*) and orange (*D*) dots are functional hits. Improved amylase production by manipulation of the identified RNAi targets (*E*) were validated with 11C-GK3 as a control. Gold bars represent genes characterized from the YeaLKD-lib and orange ones from the YeaHKD-lib. Data shown are mean values ± SDs of triplicates. Statistical difference between control and indicated strains was determined by two-tailed Student's *t* test. *$P < 0.05$; **$P < 0.01$. DCW, dry cell weight.

analyze and sort individual cells from the YeaLKD-lib and YeaHKD-lib at very high throughput. Single variant cells were encapsulated in 20-pL droplets together with a fluorogenic boron-dipyrromethene-starch substrate, which was hydrolyzed by the secreted amylase to generate a fluorescent signal dependent on the amylase concentration in the droplet (*SI Appendix*, Fig. S5*A*).

The in-droplet assay time to achieve sufficient signal resolution for the RNAi libraries was first determined to be 7 h by comparing the control and high *TDA3* knockdown strains (*SI Appendix*, Fig. S5*B*). To maximize the coverage and account for single-cell variability, microfluidic sorting was performed on 1,000,000 cells (5-fold coverage on variant number) or 500,000 cells (>10-fold) from the YeaLKD-lib or YeaHKD-lib, respectively, in three sequential rounds with 4%, 2%, and 2% most fluorescent droplets sorted in each round (*SI Appendix*, Fig. S5*C*). In doing this, yeast variant cells with potentially improved amylase production were enriched from the RNAi libraries.

We then sought to identify the RNAi targets by direct experimental validation. We first verified the improved amylase production of the postsorted cell fraction (*SI Appendix*, Fig. S6) and individual strains (16 from each group) (*SI Appendix*, Fig. S7) through cultivation by tube fermentation and an amylase assay. Following the same process, we evaluated the amylase production of 340 and 334 postsorted individual cells from the YeaLKD-lib and YeaHKD-lib (Fig. 2 *C* and *D*) and selected 25 and 27 of these strains, respectively, which had the best performance in both titer and yield for amylase. RNAi cassettes from these selected strains were reconstructed into plasmids and transformed into the fresh parent strain (11C-GK3) for analysis of amylase production (*SI Appendix*, Fig. S8). Subsequent sequencing (Dataset S1, sheets 3 and 4) validated six targets from the YeaLKD-lib (*VPS73*/*NUP188*, *YDC1*, *KTR2*, *YPR148C*/*STB5*, *AAD4*, and *CDC39*) and three from the YeaHKD-lib (*SSA1*, *TEF2*, and *FIT3*) that demonstrated favorable effects on amylase production (Fig. 2*E*, *SI Appendix*, Fig. S9, and Dataset S1, sheets 5 and 6), with improvements of up to 24.9% (YPR148C/STB5). The screens hit not only genes involved in processes known to be related to improved protein production such as protein modification and degradation (*VPS73* and *KTR2*), but also genes involved in cellular metabolism (*YDC1* and *AAD4*), as well as the essential gene *CDC39* required for cell cycle regulation.

Frequency comparison-driven identification of RNAi gene targets. The repeated sequence hits on functional targets, *YDC1* and *YPR148C*/*STB5* (Dataset S1, sheet 3), indicated a possible correlation between promotive target impact and the high frequency of selection in the postsorted pool of candidates containing the same gene function. As result of this correlation, we proceeded to identify potentially useful targets through a frequency-based approach. We sequenced the RNAi gene targets from pre- and postsorted yeast knockdown libraries, the initially constructed library and enriched high-amylase-producing variant yeast cell pool, respectively, and analyzed the gene enrichment. Out of a total of 6,433 annotated genes, 93% (5,980) and 89% (5,749) were detected in the presorted YeaLKD-lib and YeaHKD-lib, respectively (Dataset S1, sheets 7 and 8), confirming the sufficient coverage of the libraries constructed for this study. The sequencing of the postsorted YeaLKD-lib and YeaHKD-lib identified 1,833 and 2,571 genes, respectively (Dataset S1, sheets 7 and 8), demonstrating enrichment of target genes for modulation affecting amylase production (*SI Appendix*, Fig. S10). We mapped the 50 most enriched genes onto the gene interaction network to visualize any functional clusters. The genes were mainly concentrated in protein modification and transport [vesicle traffic, glycosylation, protein folding, multivesicular body (MVB) sorting], protein turnover, and mitochondrial activity (cellular metabolism and respiration) (*SI Appendix*, Fig. S11). Based on both frequency and function information, we validated the favorable effects from down-regulations of *ADE8* (LKD-lib), *CNL1* (HKD-lib), and *TIM17* (HKD-lib) involved in nucleotide biosynthesis, vesicle transport, and

import channel in mitochondria membrane, respectively, demonstrating up to 18% increased amylase secretion (Fig. 3 *C* and *D*).

To gain more insights from the enriched genes from the screen, we manually set cutoff values of 3.3 and 2 on the $\log_2$ enrichment score (corresponding to the values of 10 and 4 on the frequency ratio) to fraction the ~10% of the top enriched genes for gene ontology analysis (Dataset S1, sheets 9 and 10). The enriched biological processes again identified genes associated with protein modification and transport (*SI Appendix*, Fig. S12); nevertheless, consensus-enriched processes also identified genes related to the metabolism of lipid and carbohydrates (Fig. 3*E*). These indicated the potential of improving protein production by mitigating metabolism of other cellular organic compounds. Subsequently, we identified favorable effects on protein production from down-regulation of *SDH1* (Fig. 3*D*), which encodes succinate dehydrogenase required for the tricarboxylic acid (TCA) cycle.

Relevance of identified targets and their effects depending on knockdown levels. To test the general relevance of identified RNAi targets, we evaluated their effects using a strain that expresses amylase in a different plasmid expression system with *POT1* from *Schizosaccharomyces pombe* as the selection marker (13), using the strain 581-GK4. The strain 581-GK4 expresses Cas9, the RNAi machinery, and has *CAN1* deleted, but, importantly, none of these genetic manipulations impact amylase production (*SI Appendix*, Fig. S13). The 581-GK4 strain was transformed with p416-based plasmids with *URA3* as the selection marker (different to p413 plasmids with *HIS3* as the marker for the screens) carrying the reconstructed RNAi targets. An amylase assay of the resulting strains showed that, except for *SSA1* and *AAD4*, the identified RNAi targets (seven of nine) from the screens enhanced the amylase production upon down-regulation of their expression in the 581-GK4 strain by up to 46.2% (Fig. 4*A*). This would suggest that the RNAi targets are highly robust with respect to their effects on the protein expression system and strain genotype (including auxotrophy).

We next constructed strains with low knockdown, high knockdown, and deletion of RNAi targets to test the impact pattern (except for essential gene *CDC39*, wherein only low and high knockdown were constructed). Increased amylase production was detected upon knockdown of *TEF2* and *FIT3*, while deletions had no promotive impact (Fig. 4*B* and *SI Appendix*, Fig. S9). In contrast, both knockdown and knockout of *YDC1* and *KTR2* elevated the amylase production in a dosage-unassociated and dosage-related manner, respectively (Fig. 4*B*). Unlike the low knockdown which served as the optimal condition for *KTR2* (further decrease in *KTR2* expression reduced improvement margin of amylase production), increasingly favorable effects on amylase production was observed with the strengthening of the knockdown levels of *CDC39* (Fig. 4*B* and *SI Appendix*, Fig. S9). These results demonstrated the variable impacts of gene expression levels on protein production, indicating the importance of tuning gene expression for protein production.

Genome recombineering for tuning gene expression. To tune the expression of a set of genes for optimized protein production, we established a Cas9-mediated genome integration approach to simultaneously enhance and reduce different genes' expression levels by integrating expression cassettes and/or RNAi cassettes. We first tested the efficiency of obtaining transformants and the feasibility of the approach. The efficiency was evaluated with manipulations of two or four genomic sites, by transforming the plasmid generating short guide RNA (sgRNA plasmid) and flanking homologous region-bounded expression cassettes (repair fragments) into a strain with genome integrated Cas9. With the optimized transformation procedures, greater than 10,000 colonies (up to 100,000) were obtained for two integrations, and greater than 100 colonies (up to 1,046) were obtained for four
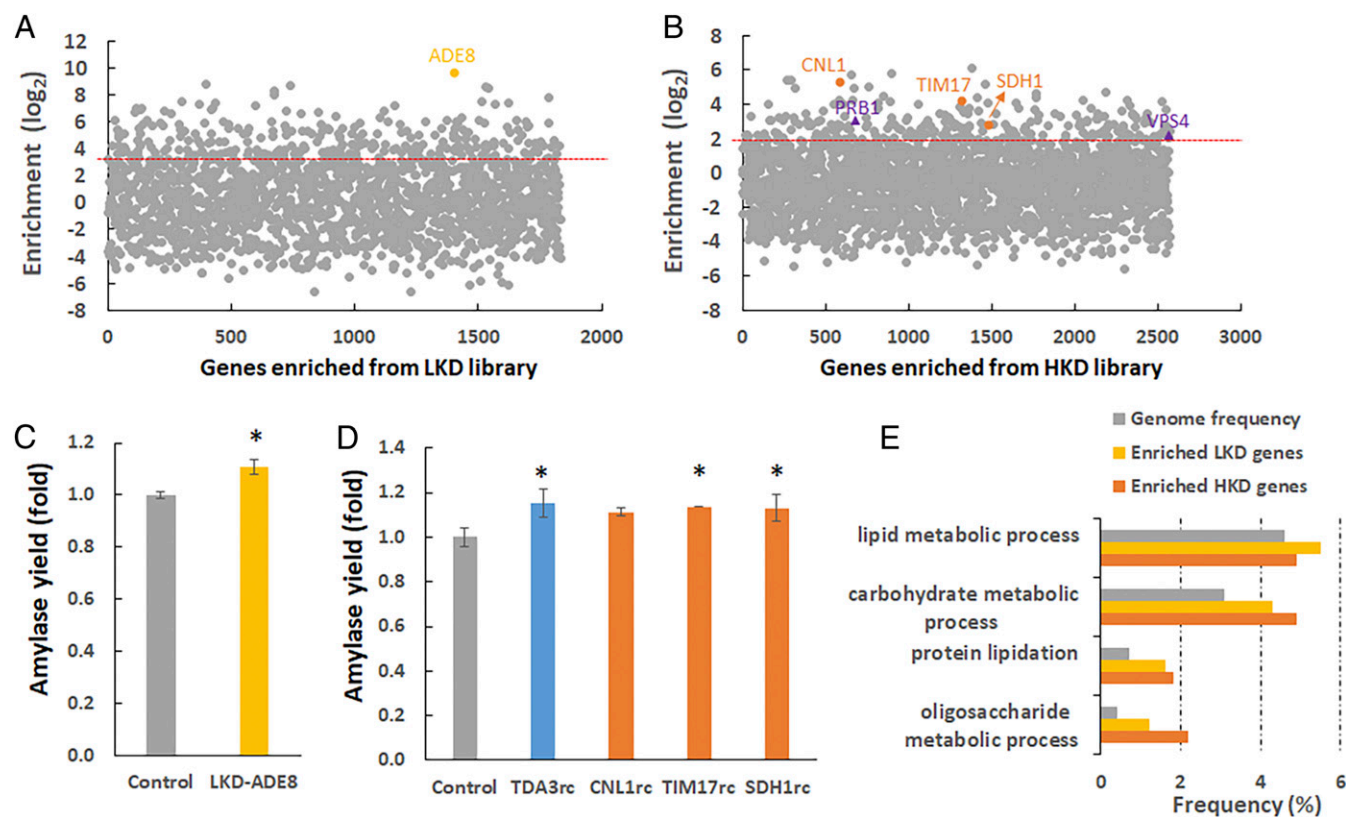
Fig. 3. Gene enrichment-assisted analysis of gene targets from RNAi screening. Count-weighted enrichment scores presented for each gene covered in the sorted YeaLKD-lib and YeaHKD-lib, with each dot corresponding to one gene. The red line marks the enrichment score of 3.32 (10-fold enrichment) and 2 (4-fold enrichment) for LKD (*A*) and HKD (*B*), respectively, to fraction the top 10% enriched genes. Dots in gold (*A*) and orange (*B*) represent functional RNA interference targets in this study and dots in purple are favorable targets on protein production upon deletion. Improved amylase production by manipulation of the identified RNAi targets from the YeaLKD-lib (*C*) and YeaHKD-lib (*D*) validated with 11C-GK3 as control. Data shown are mean values ± SDs of triplicates. Statistical difference between control and indicated strains was determined by two-tailed Student's *t* test. *$P < 0.05$. The biological processes enriched in both sorted YeaLKD-lib and YeaHKD-lib are shown in *E*.

integrations (*SI Appendix*, Fig. S14). The feasibility of expression modulations of multiple genes via cassette number-dependent dosage control was further demonstrated. By integrations of expression and/or interference cassettes of fluorescent proteins in different numbers, strains with overexpressed red fluorescence protein (RFP)/GFP (SOERG) (*SI Appendix*, Fig. S15), or overexpressed RFP and/or down-regulated GFP (SORDG) (*SI Appendix*, Fig. S16) were obtained.

Since protein production is a complex phenotype, global tuning of cellular processes is needed for optimization of this trait (19). Current methods to engineer complex traits involving rational design and individual building and testing of each variant are labor intensive. Strain identification through testing of variants from a semirationally built library, which covers a greater array of gene tuning possibilities thereby serves as a faster and higher throughput alternative to simplify design–build processes. To confirm the feasibility of constructing such a library that covers diverse variants through Cas9-mediated genome integrations, we generated two libraries for SOERG and SORDG by preparing and transforming mixed repair fragments targeting two different genomic sites which harbor RFP/GFP expression cassettes or GFPi cassettes. Testing of individuals from the two libraries showed that all three tuning possibilities for each library were obtained, respectively (*SI Appendix*, Figs. S15 and S16).

We then applied this strategy to manipulate two favorable targets, *PDI1*, via its overexpression, and *TDA3*, via its downregulation, using genomic integrations into four sites, to identify the best combination/expression tuning for protein production.

The genetic manipulations of all 15 combinations of *PDI1* overexpression and *TDA3* down-regulation were manually generated and evaluated. The P2T2 strain, with two copies of the *PDI1* expression cassette and two copies of the *TDA3* interference cassette, as well as the P1T2 and P1T3 were identified as the three most favorable combinations (Fig. 4*C*). Two rounds of evaluations of the 70 individual variants of strain library generated with mixed *PDI1* expression cassettes and *TDA3* interference cassettes (Fig. 4*D* and *SI Appendix*, Fig. S18) identified the most productive strain with the 2.4-fold improvement, in which the *PDI1* expression cassette and *TDA3* interference cassette were integrated into the genome with two copies present for each (Fig. 4*D*). These results indicated that tuning gene expression through the semirational recombineering of favorable gene targets with the Cas9-mediated genome integrations approach acted as a feasible tool for optimized protein production.

Improved protein production by recombineering. We further applied the approach for recombineering of strains for improved protein production with the nine identified RNAi targets (Dataset S1, sheets 3 and 4) alongside previously reported overexpression targets. Two rounds of recombineering for eight genomic integrations (four site integrations for each round) were performed to generate two separate libraries containing either (*i*) RNAi targets (RNAi) or (*ii*) RNAi and overexpression targets (RNAi+OE). With two rounds of evaluations on 228 and 129 variants from RNAi and RNAi+OE (*SI Appendix*, Figs. S18 and S19), strains with gradual improvements in amylase production were identified
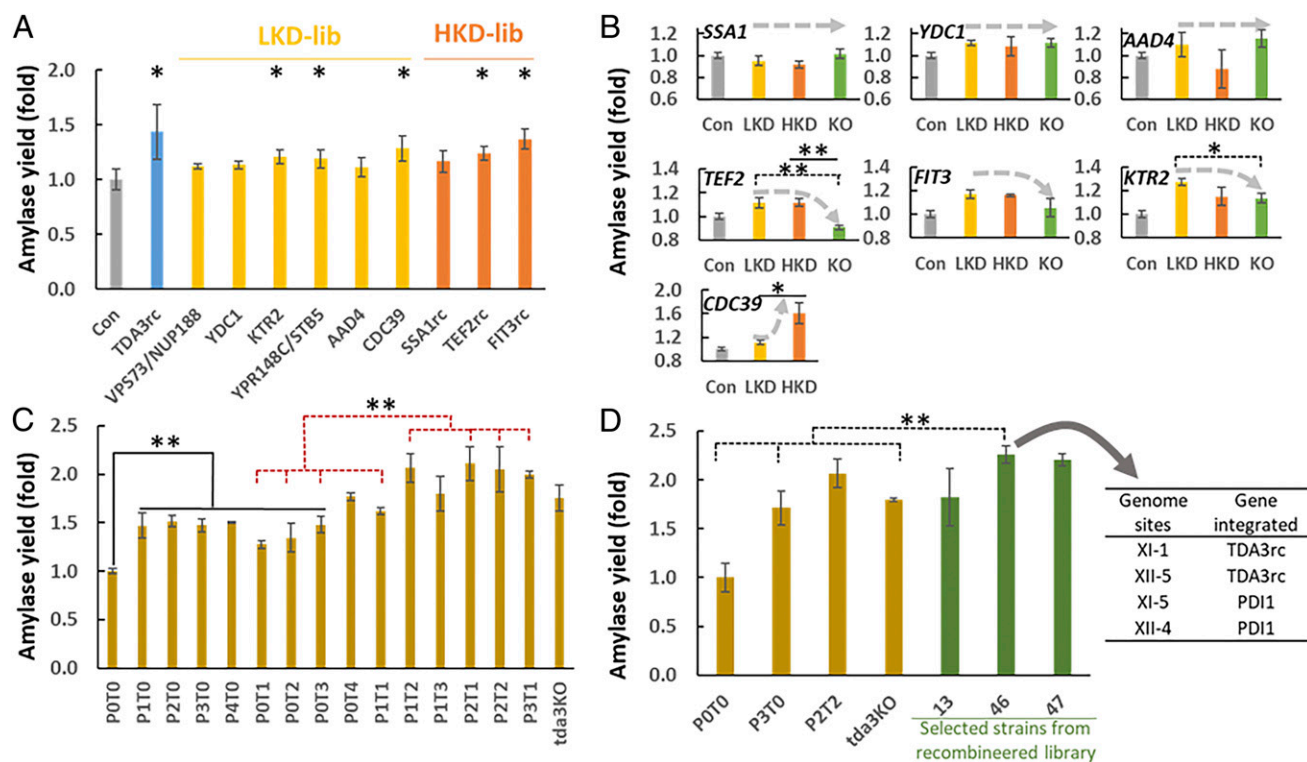
Fig. 4. Impact patterns from tuning gene expression levels and further genome recombineering for optimized protein production. (*A*) The independence of the identified favorable RNAi target effects from the protein expression system was characterized by implementing knockdowns from LKD and HKD in 581-GK4 (Con), a strain carrying a different protein expression system for amylase expression. (*B*) Distinct impacts on amylase production from knockdown of genes at various levels were evaluated in low knockdown (LKD), high knockdown (HKD), and knockout (KO) of RNAi target genes. (*C*) The amylase production of strains with fine-tuned expression of *PDI1* and *TDA3* was quantified in strains with various copies (number) of *PDI1* expression (P) and *TDA3* interference (T) cassettes integrated. (*D*) Reassessment of amylase yield from strains with fine-tuned expression and selected strains from a recombineered PDI1-TDA3rc library (13, 46, 47). The integration sites and integrated segments from the PDI1-TDA3rc library strain with the highest amylase production were characterized. Data shown are mean values ± SDs of triplicates. Statistical difference between control and indicated strains (*A*) or between indicated strains (*B*–*D*) was determined by two-tailed Student's *t* test. *$P < 0.05$; **$P < 0.01$.

(Fig. 5*A* and *SI Appendix*, Fig. S20), with a maximal improvement of 2.2-fold (strain R22-31 from the strains recombineered with RNAi targets). The time-course amylase titer and yield were validated using shake flask fermentation and showed an improved capacity in amylase production of these selected strains (*SI Appendix*, Fig. S21).

To test the compatibility of these recombineered strains with other targets known to impact protein production, we introduced *PDI1* overexpression into strains R22-31 and R22-96. The *PDI1* overexpression enhanced amylase production further by 17.8–60.1% (Fig. 5*B*). Additionally, to test the generality of protein production improvements, endo-1,4-beta-xylanase was introduced into R22-31 and R22-96. These strains showed superior capability of xylanase production (of up to 122%), compared with the control strain (Fig. 5*C*). These results suggest that tuning gene expression via Cas9-mediated genome-scale integrations could aid the generation of better producer strains.

Discussion

Here we apply droplet microfluidic screening to identify RNAi targets favorable for recombinant protein production in *S. cerevisiae* through direct experimental validation and enrichment analysis-assisted characterization and establish a framework for tuning the targets for strain optimization, using a semirational, systematic, and high-throughput systems biology approach.

The wide-range down-regulation (0.2–0.95, *SI Appendix*, Figs. S4 and S9) and high gene coverage of the YeaLKD-lib (31-fold in terms of gene number) and YeaHKD-lib (6.3-fold) allow for

the comprehensive analysis of RNAi modulation effectors in protein production. Screens of such libraries necessitate high-throughput sorting by extracellular characteristics with adequate sorting resolution. Droplet microfluidics, a high-throughput screening platform (22–25) which can distinguish secretion capacity of cells encapsulated in droplets (26), was used for screens of yeast RNAi libraries. The utilization of droplet microfluidics permits thorough analysis of variants, demonstrated by the number of cells screened for the YeaLKD-lib and YeaHKD-lib exceeding the number of library variants by 5- and 10-fold, respectively.

In terms of detection sensitivity and sorting accuracy, the stability of droplet microfluidic systems is crucial, as this enables reproducible enrichment of superior protein-producing strains. Connected to this, it is equally important to manipulate the library coverage and frequencies of individual RNAi constructs, to maximize the target existence and chance for isolation when aiming for beneficial targets of specific biological processes. An unnormalized cDNA library was used for the HKD library in this work as highly expressed genes were focused; however, this could have reduced the frequencies of gene targets with low expression levels. A normalized cDNA library or customized cDNA from cells responsive to special conditions could be options for more even or purposive-biased screens.

RNAi-mediated gene down-regulation functions to alter protein production in a dosage-associated manner, which is consistent with reports on distinct effects from knockout and gene down-regulation in stress tolerance (21, 27) and cancer control (28). This could be a partial reason for why beneficial knockout targets of vacuolar protein sorting (VPS) (29, 30) were seldomly
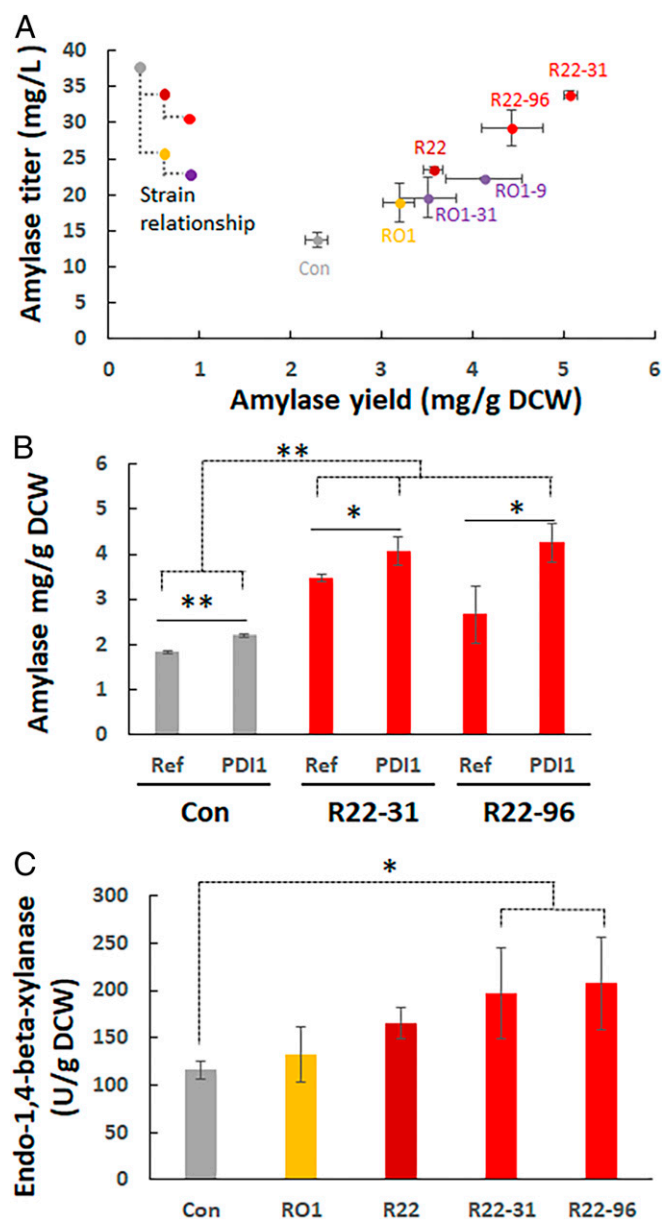
Fig. 5. Versatility of strains recombineered for improved amylase production. Strains with improved amylase production were isolated via two rounds of recombineering and evaluation (*A*). Strains were recombineered by integration of various favorable RNAi targets (R strains) or RNAi and overexpression targets (RO strains) through a Cas9-mediated approach, and strains with improved amylase productions were selected. (*B*) Compatibility with another favorable target [PDI1 overexpression (PDI1) versus reference control strain (Ref)] of the recombineered strains was determined. (*C*) The versatility of the recombineered strains was further demonstrated by enhancing endo-1,4-beta-xylanase production. Data shown are mean values ± SDs of triplicates. Statistical difference was determined by two-tailed Student's *t* test. *$P < 0.05$; **$P < 0.01$.

hit here. Based on the effect difference, RNAi screens were pursued in this work for the identification of targets which can enhance protein production with intermediate expression levels, unlike screens using overexpression (15) and knockout libraries (16), allowing for the identification of essential genes, such as *CDC39* and *TIM17*, identified in this work as candidates for protein production optimization. In addition, our screens identify targets via the characterization of causal trackable perturbations, which are generally inaccessible to random mutagenesis screens,

aiding in understanding the optimized cellular machinery for protein production from a bottom-up perspective.

Genes that were hit in our RNAi screens participate in diverse cellular processes, which support the idea that protein production is a complex process and that global optimization of cellular machinery is required for optimized protein production (19). Both validated RNAi targets and enriched genes from the YeaLKD-lib and YeaHKD-lib primarily associate with cellular processes of metabolism and protein modification, transport, and degradation. Targeted metabolic processes include sphingolipid metabolism (*YDC1*), nucleotide biosynthesis (*ADE8*), and central carbon metabolism (*SDH1*). Down-regulations of whole central carbon metabolism, including *SDH1* down-regulation, were previously characterized in a set of protein hyper-producing yeast strains through systems biology analysis (19). These results, together with a recent report on reorganization of carbon metabolism for optimized lipid yield (31), indicate the engineering potential of redirecting cellular resources toward protein synthesis and secretion. Our screens also emphasize the importance of protein modification and turnover processes in strain improvement for optimized protein production, which have been demonstrated by disruption of glycosylation (32, 33), protein sorting (34), and degradation (16, 35). A beneficial effect from down-regulation of chaperone *SSA1*, unlike that from overexpression of *SSA4* in *Pichia pastoris* (36), indirectly stressed SSA1's functional specificity in the recognition of misfolded protein for ER-associated degradation (ERAD) (37) and destabilization of correctly folded multimeric protein (38). The functioning of *TEF2* down-regulation in particular may be associated with the proteome mass reallocation from translation elongation factors (~0.81%, SI Appendix, Fig. S22) and/or coordinated translation initiation and elongation vital for efficient protein translation (39–41). Down-regulation of the essential gene *CDC39* involved in the cell cycle, conditional mutation of which resulted in G1 arrest (42), also enhanced the protein yield, confirming the universal applications of sustaining G0/G1 phase (43) and attenuating cell growth for improved protein production (44). Our findings provide targets for improved protein production and help to expand our understanding of the protein-producing and -secreting machinery. Furthermore, insights into connections between cell functions and protein production could be exploited further, such as in the case of *TEF2*, wherein other proteins with disproportionate allocation of proteome mass could be targeted for improved recombinant protein production, a strategy recently referred to as proteome streamlining (45).

As gene dosage affects protein production (46), optimized protein production necessitates combinatorial manipulation of expression levels of multiple gene effectors. This can be achieved by Cas9-mediated genome integration of customized modules, including both favorable overexpression and knockdown cassettes. Library construction with such a recombineering approach was preferred above oligo-mediated recombination such as multiplex automated genome engineering (MAGE) (47) trackable multiplex recombineering (TRMR) (48), or yeast oligo-mediated genome engineering (YOGE) (49), due to the low oligo recombination efficiency and the limitations in the tuning range of these approaches in yeast. Our recombineering approach offers opportunities to create platform strains with optimized combinations of favorable targets that enhance protein production by each introduced gene modification contributing respectively to increased amylase production. The harnessing of concrete separate genomic sites, other than disperse repetitive δ-sites (50), also allows the integrated targets to be easily trackable and stably expressed. Based on four genomic sites, this recombineering approach realizes the simultaneous expression tuning of favorable targets of certain numbers (10–20) due to the capacity limit on the library construction. Therefore, the throughputs of the diverse combinatorial possibilities (10^4–20^4), the easy-to-achieve relatively high efficiency of strain creation (100–1,000), and the

microfluidic droplet screening capability (10^5–10^6) (14) are comparable, it is hence possible to generate large libraries and still test all of the individuals. The limitation of such two-step strategies for improved protein production, including identification of contributing genes and combinatorial optimization is, however, that it is difficult to realize the global combinatorial optimization of all synergistic beneficial genetic determinants, due to the capacity of strain creation and individual screening. This could be mitigated through the development of higher throughput methods for the build-and-test phase of the design–build test–learn cycle or by iterative recombineering.

Strains obtained through the identification and recombineering of genes has proven to be a versatile and robust approach for improved protein production; furthermore, it offers a platform for engineering new and known targets for further improvement. Knowledge acquired in this work toward understanding protein-producing machinery as well as the framework for identifying promising candidates and recombineering can also be extended to other species that have a wide application in protein production, like filamentous fungi and mammalian cells, all of which retain the RNAi machinery. The additional requirement is the development of efficient site-specific integration and elimination of nonhomologous end joining.

In summary, we investigate the genetic basis of gene down-regulation for improved protein production in the yeast *S. cerevisiae*, via microfluidics-assisted tracking of systematically introduced RNAi perturbations, and establish a workflow for identification and combinatorial engineering of favorable gene targets. These down-regulation targets with functions in metabolism, protein modification, and degradation as well as cell cycle are isolated as genetic determinants. The general indication from these targets and the recombineering pipeline will aid in the understanding of the protein production machinery and potentially enable the creation of better cell factories.

Materials and Methods

All details on materials and methods associated with the strains, plasmids, media, and microfluidic devices can be found in *SI Appendix, Materials and Methods* and Dataset S1, sheets 11 and 12. Cloning for plasmids, RNAi cassette libraries, and DNA fragments for genome recombineering, yeast transformation, strain cultivation and protein quantification, fluorescence observation and quantification, quantitative PCR, manufacturing and operation of microfluidic devices for droplet sorting, sample preparation and next-generation sequencing (NGS), NGS data processing, and analysis were carried out according to the procedures described in *SI Appendix, Materials and Methods*.

1. Hou J, Tyo KE, Liu Z, Petranovic D, Nielsen J (2012) Metabolic engineering of recombinant protein secretion by *Saccharomyces cerevisiae*. *FEMS Yeast Res* 12:491–510.
2. Wang G, Huang M, Nielsen J (2017) Exploring the potential of *Saccharomyces cerevisiae* for biopharmaceutical protein production. *Curr Opin Biotechnol* 48:77–84.
3. Wang M, Kaufman RJ (2016) Protein misfolding in the endoplasmic reticulum as a conduit to human disease. *Nature* 529:326–335.
4. Gitler AD, et al. (2008) The Parkinson's disease protein alpha-synuclein disrupts cellular Rab homeostasis. *Proc Natl Acad Sci USA* 105:145–150.
5. Nielsen J (2013) Production of biopharmaceutical proteins by yeast: Advances through metabolic engineering. *Bioengineered* 4:207–211.
6. Punt PJ, et al. (2002) Filamentous fungi as cell factories for heterologous protein production. *Trends Biotechnol* 20:200–206.
7. Dinnis DM, James DC (2005) Engineering mammalian cell factories for improved recombinant monoclonal antibody production: Lessons from nature? *Biotechnol Bioeng* 91:180–189.
8. Giaever G, Nislow C (2014) The yeast deletion collection: A decade of functional genomics. *Genetics* 197:451–465.
9. Hartwell LH (1974) *Saccharomyces cerevisiae* cell cycle. *Bacteriol Rev* 38:164–198.
10. Münch T, Sonnleitner B, Fiechter A (1992) The decisive role of the *Saccharomyces cerevisiae* cell cycle behaviour for dynamic growth characterization. *J Biotechnol* 22:329–351.
11. Lahtvee PJ, et al. (2017) Absolute quantification of protein and mRNA abundances demonstrate variability in gene-specific translation efficiency in yeast. *Cell Syst* 4:495–504.e5.
12. Tang H, et al. (2015) Engineering protein folding and translocation improves heterologous protein secretion in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 112:1872–1882.
13. Liu Z, Tyo KE, Martínez JL, Petranovic D, Nielsen J (2012) Different expression systems for production of recombinant proteins in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 109:1259–1268.
14. Huang M, et al. (2015) Microfluidic screening and whole-genome sequencing identifies mutations associated with improved protein secretion by yeast. *Proc Natl Acad Sci USA* 112:E4689–E4696.
15. Wentz AE, Shusta EV (2007) A novel high-throughput screen reveals yeast genes that increase secretion of heterologous proteins. *Appl Environ Microbiol* 73:1189–1198.
16. Kitagawa T, et al. (2011) Identification of genes that enhance cellulase protein production in yeast. *J Biotechnol* 151:194–203.
17. Schultz LD, et al. (1994) Using molecular genetics to improve the production of recombinant proteins by the yeast *Saccharomyces cerevisiae*. *Ann N Y Acad Sci* 721:148–157.
18. Valkonen M, Penttilä M, Saloheimo M (2003) Effects of inactivation and constitutive expression of the unfolded- protein response pathway on protein production in the yeast *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 69:2065–2072.
19. Huang M, Bao J, Hallström BM, Petranovic D, Nielsen J (2017) Efficient protein production by yeast requires global tuning of metabolism. *Nat Commun* 8:1131.
20. Huang M, Wang G, Qin J, Petranovic D, Nielsen J (2018) Engineering the protein secretory pathway of *Saccharomyces cerevisiae* enables improved protein production. *Proc Natl Acad Sci USA* 115:E11025–E11032.
21. Si T, Luo Y, Bao Z, Zhao H (2015) RNAi-assisted genome evolution in *Saccharomyces cerevisiae* for complex phenotype engineering. *ACS Synth Biol* 4:283–291.
22. Huebner A, et al. (2007) Quantitative detection of protein expression in single cells using droplet microfluidics. *Chem Commun (Camb),* 1218–1220.
23. Mazutis L, et al. (2013) Single-cell analysis and sorting using droplet-based microfluidics. *Nat Protoc* 8:870–891.
24. Sjostrom SL, et al. (2014) High-throughput screening for industrial enzyme production hosts by droplet microfluidics. *Lab Chip* 14:806–813.
25. Bjork SM, Sjostrom SL, Andersson-Svahn H, Joensson HN (2015) Metabolite profiling of microfluidic cell culture conditions for droplet based screening. *Biomicrofluidics* 9:044128.
26. Wang BL, et al. (2014) Microfluidic high-throughput culturing of single cells for selection based on extracellular metabolite production or consumption. *Nat Biotechnol* 32:473–478.
27. Crook N, Sun J, Morse N, Schmitz A, Alper HS (2016) Identification of gene knockdown targets conferring enhanced isobutanol and 1-butanol tolerance to *Saccharomyces cerevisiae* using a tunable RNAi screening approach. *Appl Microbiol Biot* 100:10005–10018.
28. Lin A, Giuliano CJ, Sayles NM, Sheltzer JM (2017) CRISPR/Cas9 mutagenesis invalidates a putative cancer dependency targeted in on-going clinical trials. *eLife* 6:e24179.
29. de Ruijter JC, Jurgens G, Frey AD (2017) Screening for novel genes of *Saccharomyces cerevisiae* involved in recombinant antibody production. *FEMS Yeast Res* 17:fow104.
30. Marsalek L, et al. (2017) Disruption of genes involved in CORVET complex leads to enhanced secretion of heterologous carboxylesterase only in protease deficient *Pichia pastoris*. *Biotechnol J* 12:1600584.
31. Ajjawi I, et al. (2017) Lipid production in *Nannochloropsis gaditana* is doubled by decreasing expression of a single transcriptional regulator. *Nat Biotechnol* 35:647–652.
32. Hoshida H, Fujita T, Cha-aim K, Akada R (2013) N-Glycosylation deficiency enhanced heterologous production of a Bacillus licheniformis thermostable α-amylase in *Saccharomyces cerevisiae*. *Appl Microbiol Biotechnol* 97:5473–5482.
33. Tang H, et al. (2016) N-hypermannose glycosylation disruption enhances recombinant protein production by regulating secretory pathway and cell wall integrity in *Saccharomyces cerevisiae*. *Sci Rep* 6:25654.
34. Zhang B, Chang A, Kjeldsen TB, Arvan P (2001) Intracellular retention of newly synthesized insulin in yeast is caused by endoproteolytic processing in the Golgi complex. *J Cell Biol* 153:1187–1198.
35. Tomimoto K, et al. (2013) Protease-deficient *Saccharomyces cerevisiae* strains for the synthesis of human-compatible glycoproteins. *Biosci Biotechnol Biochem* 77:2461–2466.
36. Gasser B, Sauer M, Maurer M, Stadlmayr G, Mattanovich D (2007) Transcriptomics-based identification of novel factors enhancing heterologous protein secretion in yeasts. *Appl Environ Microbiol* 73:6499–6507.
37. Han S, Liu Y, Chang A (2007) Cytoplasmic Hsp70 promotes ubiquitination for endoplasmic reticulum-associated degradation of a misfolded mutant of the yeast plasma membrane ATPase, PMA1. *J Biol Chem* 282:26140–26149.
38. Delic M, et al. (2013) The secretory pathway: Exploring yeast diversity. *FEMS Microbiol Rev* 37:872–914.
39. Shah P, Ding Y, Niemczyk M, Kudla G, Plotkin JB (2013) Rate-limiting steps in yeast protein translation. *Cell* 153:1589–1601.

APPLIED BIOLOGICAL SCIENCES

40. Weinberg DE, et al. (2016) Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep* 14: 1787–1799.

41. Schlesinger O, et al. (2017) Tuning of recombinant protein expression in *Escherichia coli* by manipulating transcription, translation initiation rates, and incorporation of noncanonical amino acids. *ACS Synth Biol* 6:1076–1085.

42. de Barros Lopes M, Ho JY, Reed SI (1990) Mutations in cell division cycle genes *CDC36* and *CDC39* activate the *Saccharomyces cerevisiae* mating pheromone response pathway. *Mol Cell Biol* 10:2966–2972.

43. Du Z, et al. (2015) Use of a small molecule cell cycle inhibitor to control cell growth and improve specific productivity and product quality of recombinant proteins in CHO cell cultures. *Biotechnol Bioeng* 112:141–155.

44. Li S, et al. (2016) Enhanced protein and biochemical production using CRISPRi-based growth switches. *Metab Eng* 38:274–284.

45. Valgepea K, Peebo K, Adamberg K, Vilu R (2015) Lean-proteome strains: Next step in metabolic engineering. *Front Bioeng Biotechnol* 3:11.

46. Bao J, Huang M, Petranovic D, Nielsen J (2017) Moderate expression of *SEC16* increases protein secretion by *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 83: e03400-16.

47. Wang HH, et al. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460:894–898.

48. Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LBA, Gill RT (2010) Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat Biotechnol* 28:856–862.

49. DiCarlo JE, et al. (2013) Yeast oligo-mediated genome engineering (YOGE). *ACS Synth Biol* 2:741–749.

50. Si T, et al. (2017) Automated multiplex genome-scale engineering in yeast. *Nat Commun* 8:15187.