Improving heterologous protein production in yeast with massively parallel CRISPR genome editing

Introduction

Protein production plays an important role in the growing bioeconomy industry. With various applications in biotechnology and biopharma - from enzymes and therapeutics to meat and materials alternatives[1-3] - the demand for large-scale protein production creates an ongoing challenge to optimize the synthesis of heterologous proteins in industrial microorganisms. Yeast *Saccharomyces cerevisiae* provides a good platform for protein expression due to the ease of genetic manipulation and cultivation, ability to express plant-derived and mammalian proteins, and possibility of post-translational modifications[4,5]. When the goal is to maximize the amount of protein produced, both the protein expression and the host strain engineering approaches need to be considered[6,7]. Protein expression engineering is typically limited to codon usage optimization, increasing the gene copy number, and promoter and ribosomal binding site (RBS) substitution, while strain optimization is done using crude methods like random mutagenesis. However, accumulating mutational load can potentially weaken the strain by increasing the probability of disrupting important physiological processes or inducing epistatic effects.

Here we demonstrate a highly multiplexed, automated CRISPR-based genome editing approach using the Onyx™ platform that can target both the protein expression cassette and the host strain genomic background for improving production of industrially important cellobiohydrolase I enzyme (CBH1). We engineer the strain with the protein integrated into the chromosome to account for the genomic context of the host. We apply a combination of informed and exploratory approaches to create genome-wide targeted libraries containing a wide variety of edit types, from gene knockouts to site saturation mutagenesis and promoter engineering. After just one round of engineering, we identify variants with a significant increase in CBH1 production across multiple functional targets, offering insights into the mechanisms of strain improvement for protein production.

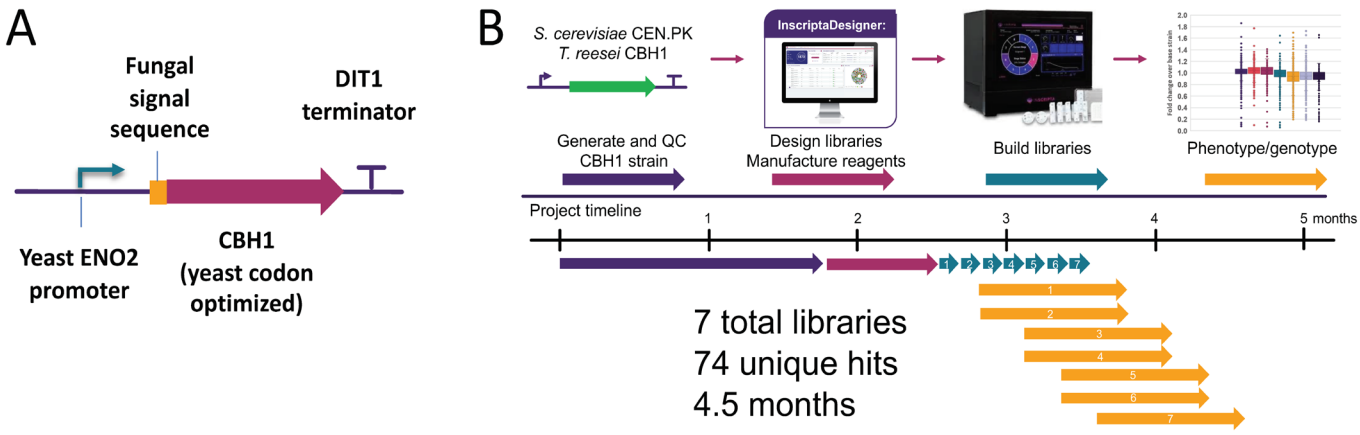


Figure 1: A) Schematic of the *cbh1* expression cassette used to construct the strain. B) Edited strain library construction workflow and timeline.

Target selection and design of libraries

Cellobiohydrolases are an important class of enzymes responsible for cellulose degradation. They are actively explored for engineering industrial microorganisms that can efficiently utilize cellulosic carbon sources for production of valuable chemicals (consolidated bioprocessing). Heterologous variants of CBH1 have been successfully expressed in yeast, albeit at low titers[8,9]. In our base strain, a single copy of the *cbh1* gene from *Taloromyces emersonii* was introduced into the chromosome of *S. cerevisiae* CEN.PK strain downstream of the LEU2 locus. Chromosomal expression eliminates copy number variation and allows unbiased assessment of activity improvement. The expression cassette contained a native pENO2 constitutive promoter, a native *T. emersonii* signal sequence, and a DIT1 terminator (**Figure 1A**). The integration was confirmed by PCR, the strain sequenced, and expression of CBH1 assessed by the CBH1 activity assay with chromophoric substrate pNP-beta-lactopyranoside, which shows characteristic absorbance at 405 nm. The assay was used to establish baseline activity in the CBH1 expressing base strain (average absorbance of 0.26, with a coefficient of variance of 1.7%). No activity was detected in the parent CEN.PK strain.

Previous attempts to improve CBH1 production and activity in yeast identified stress response pathways and high plasmid copy number as factors limiting production[9] along with other potential causes, such as hyper glycosylation or protein misfolding[10,11]. Based on this knowledge, we designed targeted editing libraries for a diverse range of cellular functions, including knockouts of genes previously shown to improve heterologous protein production in yeast; knockouts of known glycosylation and protein degradation genes to improve protein stability; promoter diversification through transcription factor binding site (TFBS) shuffling; a synonymous codon library for *cbh1* gene. In addition, we also included untargeted genome-wide knockouts (GW KO) and terminator libraries as a genomic background diversification strategy. The library sizes varied from 200 to several thousands, and the total number of edits designed was >14,000 (**Table 1**). The libraries were constructed over 7 runs using the Onyx instrument, sequenced and assessed using the InscriptaResolver™ software to determine the screener's score metric for calculating the number of variants to be screened in each library (**Figure 1B**).

Library Concept	Library size	Isolates Screened	Est Library coverage	# of retest hits	# of unique hits	Hit rate
GW KO	5133	2024	18%	40	26	2.0%
1-59bp GW deletions	1142	1104	26%	11	6	1.0%
GW terminator	5145	1104	4%	13	8	1.2%
Targeted KO	201	920	68%	21	7	2.3%
pENO2 TFBS ins/del	280	1104	48%	22	4	2.0%
pENO2 TFBS swap	1588	1104	9%	23	19	2.1%
CBH1 Alternate codon	1370	1104	32%	6	4	0.5%
Total	**14859**	**8464**		**136**	**74**	

Table 1: Summary of edit design libraries size, screening throughput, and hit rate.

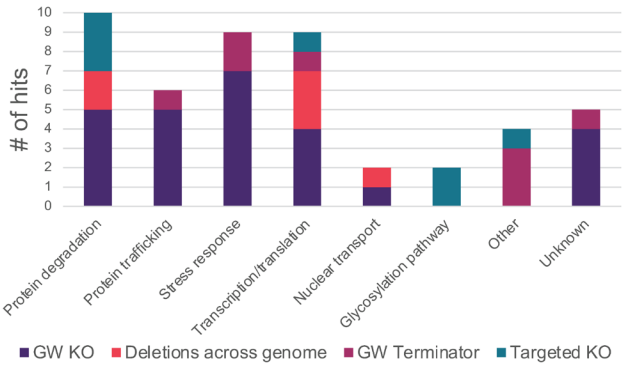
Screening of genome-wide exploratory libraries

We designed and constructed three different genome-wide libraries as a diversity generation strategy, which included gene knockouts, terminators, and short (1-59 bp) deletions throughout the entire genome. These libraries were analyzed using the Onyx Barcode Diversity Assay and screened for improvement in CBH1 activity. The assay was performed on cell supernatant after 24 hours of cultivation in 50mM acetate buffer (pH 5) using 1mM pNP-beta-lactopyranoside at 42º C. After ~2 hours, the reaction was quenched with 1M carbonate. The assay was conducted in 96-well plate format using 300 µL cell supernatant and the output measured using a SpectraMax iD3 plate reader. The threshold for activity improvement was set at 10% based on the signal sensitivity, variability between samples and background noise.

The genome-wide knockout library contained a total of 5,133 edit designs for all *S. cerevisiae* genes and was achieved by introducing a triple stop codon insertion at the 15th amino acid of the coding sequence. From this library, 2,024 isolates were selected for screening and 40 showed a significant improvement in activity. These 40 variants were retested (in quadruplicates) and sequenced to reveal 26 unique edits that conferred improved CBH1 activity (**Table 1**). The genome-wide terminator library consisting of 5,145 edit designs was screened at 1,104 isolates, identifying 13 hits (8 unique) with improved CBH1 activity. Finally, the short deletions library containing 1,142 edit designs was more deeply screened at 1,104 isolates, yielding 11 hits (6 unique). The different screening strategies were used to assess the screening effort to success rate and showed that shallow screening of large libraries (such as the >5,000-variants terminator library) can yield a similar number of hits as deeper screening of smaller libraries (i.e., short deletions).

The hits identified during screening of the genome-wide knockout libraries fell across a wide spectrum of gene functions: protein degradation and secretion, stress response, transcription and translation, nuclear transport, and some uncharacterized genes (**Figure 2A**). Similarly, short deletions disrupted genes of various function, such as transcriptional repressor SUM1, which was also identified in the knockout screen (**Figure 2B**). Since loss of function offers limited insight into the mechanism of action, other edit types that result in modulating gene expression were included. Terminator sequences are known to impact transcript stability and abundance and we identified 8 unique hits spanning diverse gene functions, such as protein trafficking, stress response, transcription, and other, within the terminator library.

A



B

Gene	Library	Fold improvement
UBP8	Deletions across genome	1.17
UBP8	Targeted KO	1.19
STM1	GW Terminator	1.12
STM1	GW KO	1.16
SUM1	GW KO	1.24
SUM1	Deletions across genome	1.25

C

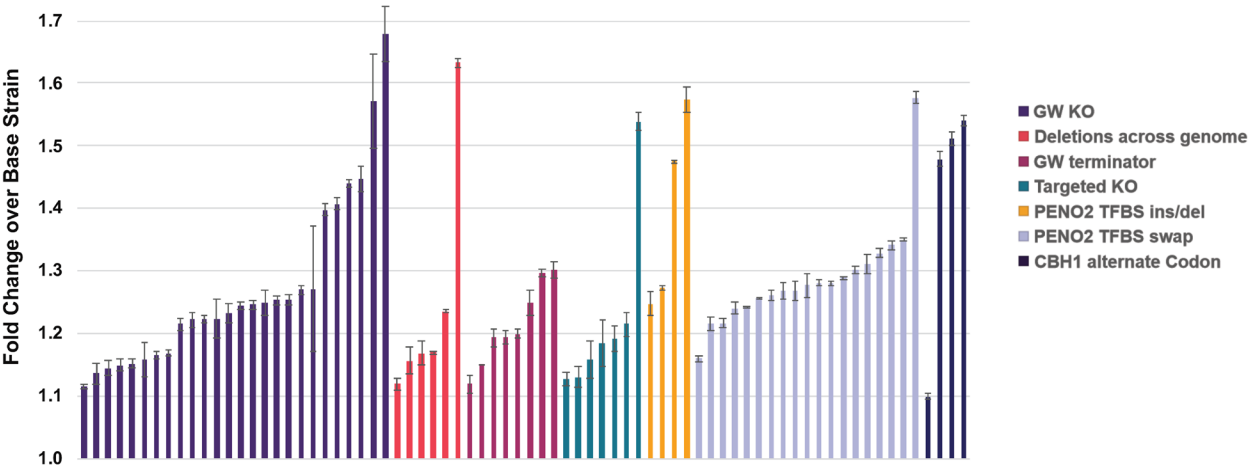


Figure 2: A) Number of unique variants with improved CBH1 activity sorted by gene function. B) Specific gene targets that were identified in different library types and their corresponding fold improvement. C) Activity improvement in edited strains across different libraries.

Targeted libraries for improved CBH1 production

The targeted libraries represented about 25% of the total designs and constituted smaller pools, allowing to sample them at a deeper level. These libraries included engineering the ENO2 promoter via TFBS insertions, deletions and substitutions; gene knockout of known proteases and other genes implicated in improving protein production; an alternate codon library for CBH1 (**Figure 2C**). The targeted knockout library focused on the genes previously shown to improve production of heterologous proteins, including known proteases and glycosylation pathway proteins[7,12,13]. The library contained only 201 members, with 920 isolates sampled resulting in 21 hits. These contained 7 unique variants, including GCN5, a catalytic subunit of histone acetyltransferase complexes previously shown to improve the yield of a recombinant protein in yeast, which showed up to ~1.2-fold improvement in activity[13]. Another hit from the targeted knockout library, a ubiquitin-specific protease UBP8, was also identified in the genome-wide knockout library and showed a ~1.18-fold improvement.

The remaining two libraries focused on CBH1 expression. Previous works have shown that the

spacing, location, and combinations of TFBSs are all important for expression[14,15]. Known TFBS sequences in the ENO2 promoter were selected and either deleted, extra copies introduced, or substituted with alternate TFBSs creating a diverse promoter library of 1,868 total variants. The libraries were screened and 45 variants identified, including 23 unique edits, linked to improved CBH1 activity. Another strategy for improving protein expression is to optimize codon usage. Typically, codon optimization algorithms are used for recombinant protein expression, as was done for *cbh1*. However, these computational approaches do not account for factors such as translation rate and tRNA pools depletion. We created a 1,370-member alternate codon library and screened 1,104 isolates. This library yielded 6 hits with up to ~1.5-fold improvement, indicating that codon usage could still be improved in a gene that had been initially codon-optimized. Since the library contained only single-codon edits, these results could potentially be combined for further optimization of the protein expression.

Conclusions

Here we demonstrate that production of heterologous proteins can be successfully improved using different genome-wide editing approaches. The entire process, from strain construction through testing and validation, took about 4.5 months, with the library construction stage taking less than a month. We demonstrate that shallow screening of large libraries can be fruitful in identifying hits and this strategy can be advantageous with limited phenotyping capacity. Furthermore, we show the utility in screening several diverse libraries which ultimately identified 74 unique hits. Our targeted libraries focused on functions specific to protein transcription (TFBS libraries), translation (alternate codon and transcript stability), and protein stability (targeted knockout), yielding hits across all categories (**Figure 3**). In addition, untargeted genome-wide edit libraries identified 40 unique variants with significantly improved activity. Because these libraries target such a wide range of cellular functions, the beneficial edits likely can be successfully recombined to generate further improvement in protein production. The short design and construction times make the Onyx workflow exceptionally well suited for iterative genome engineering and rapid combinatorial strain optimization.

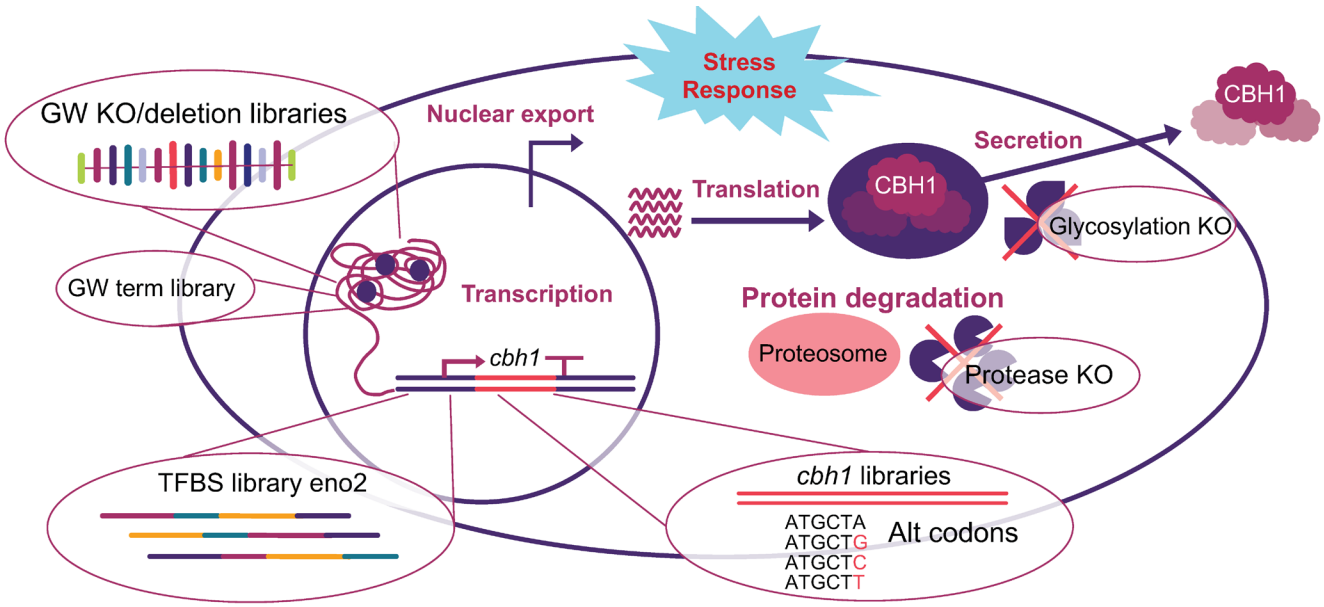


Figure 3: Edit library design with gene targets across many functional classes.

REFERENCES

1. Wang, G., Huang, M., and J. Nielsen. "Exploring the potential of *Saccharomyces cerevisiae* for biopharmaceutical protein production." Current opinion in biotechnology 48 (2017): 77-84.
2. Waltz, E. Appetite grows for biotech foods with health benefits. Nature biotechnology 37, 573–480, 2019.
3. Whittall, D. R., et al. "Host Systems for the Production of Recombinant Spider Silk." Trends in Biotechnology (2020).
4. Gerngross, T. U. (2004). Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. Nature biotechnology, 22(11), 1409-1414.
5. Walsh, G., & Jefferis, R. (2006). Post-translational modifications in the context of therapeutic proteins. Nature biotechnology, 24(10), 1241-1252.
6. Tripathi, N. K., & Shrivastava, A. (2019). Recent developments in bioprocessing of recombinant proteins: expression hosts and process development. Frontiers in bioengineering and biotechnology, 7, 420.
7. Sinha, J., et al. "Causes of proteolytic degradation of secreted recombinant proteins produced in methylotrophic yeast *Pichia pastoris*: Case study with recombinant ovine interferon–τ." Biotechnology and bioengineering 89.1 (2005): 102-112.
8. Den Haan, R., et al. "Functional expression of cellobiohydrolases in *Saccharomyces cerevisiae* towards one-step conversion of cellulose to ethanol." Enzyme and Microbial Technology 40.5 (2007): 1291-1299.
9. Ilmén, M., et al. "High level secretion of cellobiohydrolases by *Saccharomyces cerevisiae*." Biotechnology for biofuels 4.1 (2011): 1-15.
10. Penttilä, M. E., et al. "Efficient secretion of two fungal cellobiohydrolases by *Saccharomyces cerevisiae*." Gene 63.1 (1988): 103-112.
11. Boer, H., Teeri, T. T., and A. Koivula. "Characterization of *Trichoderma reesei* cellobiohydrolase Cel7A secreted from *Pichia pastoris* using two different promoters." Biotechnology and bioengineering 69.5 (2000): 486-494.
12. Huang, M., et al. "Engineering the protein secretory pathway of Saccharomyces cerevisiae enables improved protein production." Proceedings of the National Academy of Sciences 115.47 (2018): E11025-E11032.
13. Bonander, N., et al. "Altering the ribosomal subunit ratio in yeast maximizes recombinant protein yield." Microbial cell factories 8.1 (2009): 1-14.
14. Reményi, A., Schöler, H. R., and M. Wilmanns. "Combinatorial control of gene expression." Nature structural & molecular biology 11.9 (2004): 812-815.
15. Blazeck, J., and H. S. Alper. "Promoter engineering: recent advances in controlling transcription at the most fundamental level." Biotechnology journal 8.1 (2013): 46-58.

Learn more at **INSCRIPTA.COM**