

Genome-wide engineering for synthetic auxotrophy biocontainment in yeast

Yizhi Cai

yizhi.cai@manchester.ac.uk

The University of Manchester <https://orcid.org/0000-0003-1663-2865>

Stefan Hoffmann

Systems and Synthetic Biology, Wageningen University & Research

Tarryn Miller

Inscripta Inc. <https://orcid.org/0000-0003-2274-2083>

David Leeming

Manchester Institute of Biotechnology, The University of Manchester

Mark McCullough

Manchester Institute of Biotechnology, The University of Manchester

Raymond Wan

Manchester Institute of Biotechnology, The University of Manchester <https://orcid.org/0000-0003-3202-7008>

Bryan Leland

Inscripta Inc.

Article

Keywords:

Posted Date: August 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9118282/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Genome-wide engineering for synthetic auxotrophy biocontainment in yeast

Stefan A. Hoffmann^{*1}, Tarryn E. Miller², David Leeming³, Mark McCullough³, Raymond Wan³, Bryan Leland², Yizhi Cai^{*3}

1 Systems and Synthetic Biology, Wageningen University & Research, Wageningen 6708 WE, NL

2 Inscripta Inc., 5500 Central Ave STE 220, Boulder, CO 80301, USA

3 Manchester Institute of Biotechnology, The University of Manchester, Manchester M1 7DN, UK

Abstract

Comprehensive safety measures need to be developed and put into place for the safe use of genetically engineered microorganisms for open-environment applications. Synthetic auxotrophy for non-canonical amino acids (ncAAs) can be a genetic safeguard, leveraging a translation system incorporating ncAAs at UAG codons. These stop codons are placed in the reading frame of essential genes such that in the presence of the ncAA functional essential protein is made through stop codon suppression, whereas its absence results in a truncated non-functional gene product. We have performed a large-scale screening for suitable proteins and positions for TAG codon placement to create synthetic auxotrophy in *Saccharomyces cerevisiae*, testing ~4000 edits across 397 essential proteins using a pooled, barcoded editing library. We determined barcode frequency changes over time when growing with and without the ncAA O-methyl-tyrosine, looking for barcodes being depleted without the ncAA, but maintained in its presence. This screening provided information about the suitability of essential genes, sites, and sequence context for this containment approach. We combined multiple TAG edits aiming to achieve stringent containment while maintaining high fitness in permissive condition, being most successful with introducing two select TAG edits in a single gene (*PRP5*), with an escape frequency of 2.3×10^{-5} . Further, we layered a synthetic auxotrophy system with a single TAG edit with another containment system based on conditional essential protein stability, achieving extremely stringent containment (determined escape frequency 7.5×10^{-11}) and nearly wild-type like growth in permissive medium. These findings lay the groundwork for reliable, high-stringency eukaryotic biocontainment systems that balance environmental safety with functional performance.

Introduction

Synthetic biology offers tremendous potential to develop novel technological solutions to key issues humanity is facing. However, it is critical to ensure the safety of synthetic biology-based applications and prevent unintended consequences for the environment and human health. This is particularly challenging in prospective open-environment applications of engineered microorganisms due to their high replicative potential and propensity for horizontal gene transfer. Tackling this problem requires a multi-layered safety concept and the development of suitable technologies. A central role likely will be played by genetically encoded biocontainment systems, which permit survival only under well-defined permissive conditions, for instance in the presence or absence of selected small molecules. Accordingly, there has been substantial interest in developing such systems, with a variety of approaches having been explored (reviewed in [1-3]).

One of these approaches leverages orthogonal translation systems for conditional suppression of amber stops placed in essential genes. These systems specifically incorporate non-canonical amino acids (ncAAs) at TAG stop codons, suppressing the stop codon only if a suitable ncAA is being supplied. Conceptually, by introducing these stop codons in suitable positions of essential genes, the expression of the functional essential gene product, and thus organism survival, is made dependent on the presence of the ncAA. This approach has mostly been investigated in bacteria and has allowed very stringent containment in *E. coli* strains devoid of TAG stop codons and release factor 1 [4, 5]. Evidently, using orthogonal translation can be efficiently leveraged to make organisms strictly dependent on amino acids not commonly available in nature. Recently, it has been demonstrated that synthetic auxotrophy for ncAAs can also be created in yeast applying the same genetic code expansion-based principle, hosting the orthogonal translation system on a plasmid and testing biocontainment in defined media [6]. However, plasmid-based encoded heterologous systems are generally considered less robust than genomically hosted ones [7], and from a biocontainment perspective have the drawback that they can be subject to horizontal gene transfer [8]. Further, most relevant bioindustrial or environmental applications requiring biocontainment are likely to take place in complex rather than defined media.

Therefore, we aimed to develop a fully genomically encoded synthetic auxotrophy system and applying it in rich medium. To this end, we comprehensively investigated the suitability of essential target proteins and positions for biocontainment through genetic code expansion-based synthetic auxotrophy in yeast. Besides providing stringent containment in the restrictive environment, genetic containment systems also should not impose a fitness burden in the permissive environment. This dual requirement presents an engineering problem. When leveraging control over an essential gene for biocontainment, the choice of the target gene is a decisive parameter for the behavior of the implemented containment system. In previous work we have screened large numbers of essential genes to find suitable target genes for approaches based on conditional transcription [9] and conditional protein stability [10] to create safeguards offering both high stringency and minimal fitness impact under permissive conditions.

However, inserting suppressible stop codons into essential genes to create dependency on a ncAA not only requires identifying a suitable protein target, but also a suitable residue to substitute. Firstly, absence of the ncAA must result in a truncation that results in a non-functional gene product. This condition generally should be fulfilled for positions not too close to the C-terminus. Secondly, in the presence of the ncAA, sufficient functional essential gene product must be made to fully support organism viability. For this condition to be fulfilled the chosen protein position must a) functionally tolerate substitution with the ncAA and b) the orthogonal translation system must have enough stop suppression activity at that position. Typically, orthogonal translation systems are considerably less efficient than endogenous ones [11, 12], and their activity appears to be dependent on the sequence context [13, 14].

Here, we combined structure-guided prediction with a highly parallel genome editing approach to screen for suitable TAG edit sites across 397 essential genes. For permissible edits resulting in ncAA dependence, overall expression strength of target genes appeared to matter less than their biological function and sequence context of edits. Targeted follow-up on the strongest candidates showed that combining multiple TAG edits, including within a single gene, produced strains with stringent ncAA dependence, though often at a fitness cost. Finally, pairing synthetic auxotrophy with an orthogonal, conditional protein stability-based containment system yielded a strain combining extremely stringent containment (escape frequency 7.5×10^{-11}) with wild-type-like growth.

Results

Comparing ectopic and genomic expression

As a proof of concept, we applied the approach of creating synthetic auxotrophy to non-canonical amino acids on *PDII*, an essential protein disulfide isomerase required for correct protein maturation. In a BY4742 background (*MAT α* *his3 Δ 1* *leu2 Δ 0* *lys2 Δ 0* *ura3 Δ 0*) in which the used orthogonal translation system (OTS) LeuOmeRS was supplied on a centromeric plasmid (pRS315-LeuOmeRS), a TAG stop codon (*) was introduced at two different sites, either substituting K341 or D375. These two positions were manually chosen based on structural considerations: both are located in loop regions and highly solvent exposed, facing away from the protein core. The resulting strains showed growth dependence on O-methyl-tyrosine (OMeY) (Figure 1), a non-canonical amino acid that is efficiently incorporated by the LeuOmeRS system in yeast [12, 15]. Curiously, the K341* and D375* mutants clearly differed in their stringency of OMeY dependence, in that the D375* mutant showed some growth without OMeY, but K341* did not show any growth without. However, both mutants had a good growth recovery in this background when supplied with 2 mM OMeY. However, in a background where the orthogonal translation system was genomically inserted at the *ho* locus (strain SHy081), only the D375* mutant could be generated, whereas the more stringent K341* mutant could not be made. Apparently, the OTS had a lower activity when genomically integrated compared to when hosted on a centromeric plasmid.

We made similar observations with TAG mutants *ALG1* F250* and *CDC27* R14*, which we chose for comparison to previous work [6]. Both had been reported to recover to wild-type fitness with OMeY when the essential gene with the nonsense mutation and the OTS were being expressed from centromeric plasmids. We first generated strains with the respective genomic mutations, providing the OTS on the centromeric plasmid pRS315-LeuOmeRS. The resulting strains showed the desired behavior of a clear ncAA dependence, but had a moderate growth defect at 2 mM OMeY (Supplementary Figure S1). However, as observed for *PDII* K341*, attempts at generating these mutations in SHy081 were not successful, indicating OTS activity was too low to rescue those nonsense mutations in this background.

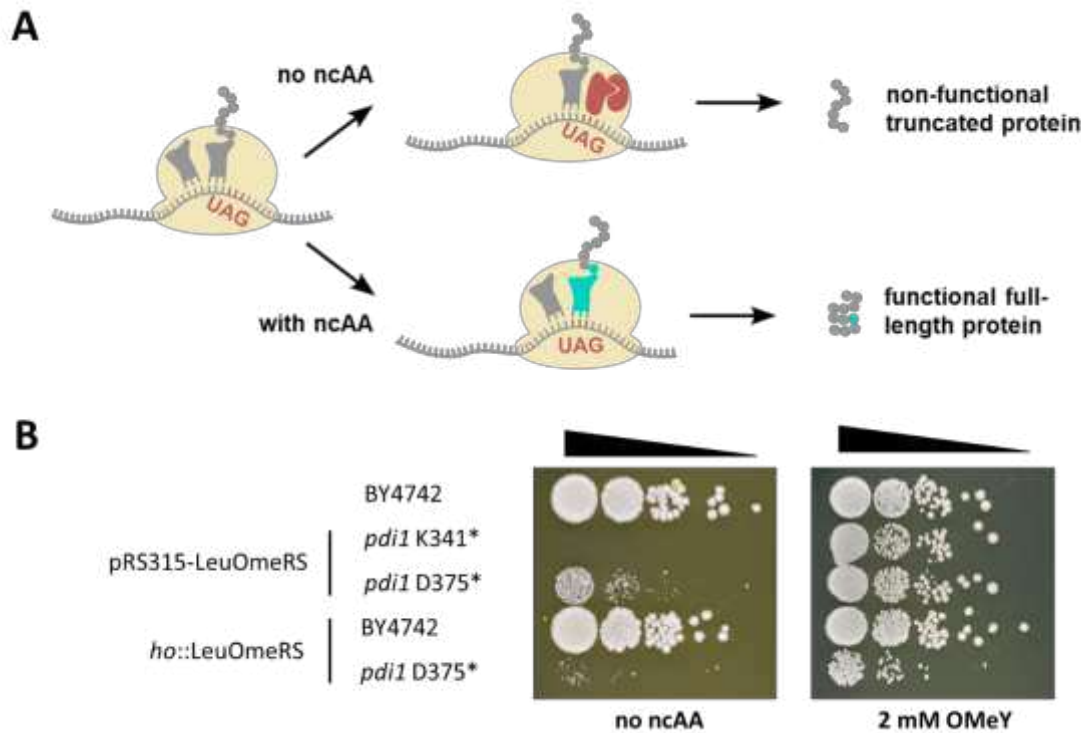


Figure 1: Principle of synthetic auxotrophy based on genetic code expansion and proof of principle. **A)** Schematic of logic of synthetic auxotrophy engineering. In a host with an orthogonal translation system (OTS), a suitable codon within an essential gene is replaced with a UAG stop codon (red triplet). With a non-canonical amino acid (ncAA) being recognized by the OTS (in teal), it can be incorporated at the UAG stop, thus suppressing it and leading to translation of a full-length essential gene product. However, in the absence of the ncAA, translation is terminated at the engineered-in stop codon. The created truncated gene product is not functional. **B)** Demonstration of principle with *PDII*. The codons corresponding to the protein positions K341 and D375 were substituted by a TAG codon, with the OTS LeuOmeRS either supplied on a centromeric plasmid (pRS315) or genomically integrated at the *ho* locus. In the background with genomic OTS integration, only the *PDII* D375* mutation could be generated.

Computational prediction of tolerant positions and library design

We aimed to computationally pre-select positions in essential proteins likely to tolerate substitution by non-canonical amino acids to restrict the queried sequence space. There are tools to predict whether a given amino acid substitution is likely to produce a functional protein or not (e.g. SIFT[16], SuSPect[17], PredictSNP[18]). Rovner et al. [5] have used the sequence alignment-based algorithm SIFT to predict residues tolerating substitutions. For a given queried protein sequence, SIFT's web service takes about 10-15 min to run predictions. However, our use case does not require accurate classification of all residues of a protein into tolerant or intolerant of substitution, but only identification of several positions that have a very high likelihood of tolerating the incorporation of a bulky non-canonical amino acid. We opted to employ a structure-based prediction, since a relatively large number of experimentally determined structures of essential *S. cerevisiae* proteins is available (401 at the time we performed this analysis).

Formalizing the way we selected tolerant sites manually in *Pdi1*, we developed a computationally inexpensive algorithm to automate this task using ChimeraX (Figure 2A). The most surface exposed residues of a given protein structure are substituted for tryptophan and arginine while preserving the side chain angles, counting sterical clashes for both substitutions. Positions were ranked by number of clashes and surface exposure, assuming high surface

exposure and no or few (≤ 3) total clashes upon substitution indicate positions able to accommodate a variety of bulky side chains (Figure 2). This calculation took only about 1-2 min per protein to run locally on a consumer-grade computer. For verification, the script was run on a protein structure of the lac repressor (PDB 1LBI) across the four chains of the homotetramer, and compared to experimentally assessed substitution tolerance [19]. The heuristic identified residues that had been shown to tolerate substitutions well (Supplementary Table S1).

The algorithm was run on 401 available *S. cerevisiae* essential protein structures from PDB as of May 2021. Solvent accessible residues were ranked from least to most total clashes upon substitution with arginine and tryptophan, excluding positions within 20 amino acids from the C-terminus. A total of 4,109 edits in 398 of those essential proteins were included in the library, in each targeting about ten of the highest-ranking residues for TAG substitution. Additionally, 50 designs were included which target a subset of these positions, but encode synonymous substitutions of the original residues. These 50 intended edits should not be harmful, and thus allow to assess baseline editing efficiency of the respective essential genes. As further controls, 98 TAG edits in 10 non-essential genes were added to assess efficiency of stop codon insertion in the absence of the need for read-through. Also, the two previously manually characterized *PDII* TAG edits K341* and D375* were included with a total of 82 delivery designs. A list of all designs in the library can be found in Supplementary Table S2.

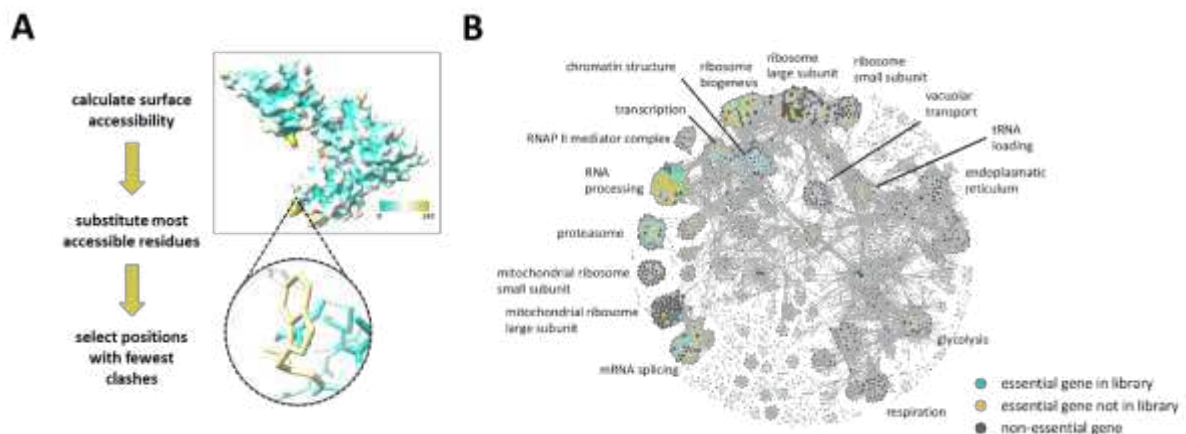


Figure 2: Structure-guided library design. **A)** ChimeraX was used to identify residues likely to tolerate substitutions by bulky ncAAs. Firstly, the surface exposure of all residues was calculated (shown in $\text{\AA}^2$), and the 20% most exposed residues were substituted with tryptophan and arginine, counting steric clashes for both substitutions. Residues were ranked from least to most clashes. **B)** Yeast protein interactome with essential proteins sampled in the library in teal and essential proteins not included in olive. Large, dense interaction clusters are marked with their respective biological function or cellular structure. Interactome data was taken from [20], displaying only interactions with correlation scores of at least 4.

Library creation and screening

The edit designs were delivered as a plasmid library into *S. cerevisiae* BY4742 with the LeuOmeRS orthogonal translation system inserted at the *ho* locus (strain designated SHy081) using the Onyx platform, a device for massively parallel genome editing. Each plasmid species of the library encodes an edit design, carrying the donor and the guide for the *Eubacterium rectale* MAD7 nuclease, as well as a unique barcode [21].

To assess the necessity of including OMeY during library creation, it was done both without and with supply of OMeY to culture media of the Onyx. Accordingly, two yeast libraries were obtained. From each, 48 independent colonies were isolated and grown in the presence of

OMeY for isolation of their genomic DNA and whole genome sequencing to determine whether the desired edit had been introduced (Figure 3B). Cells grew for several generations during their treatment on the Onyx platform. Therefore, it was expected that TAG edits in essential genes are being depleted unless complemented by the ncAA. Supplying OMeY on the platform clearly yielded a higher frequency of correct edits, and the resulting yeast library was used for subsequent selection experiments.

The editing plasmids were maintained throughout selection experiments via their hygromycin resistance marker. The yeast library was grown in triplicate and cultured with and without OMeY for a total of about 18 generations (Figure 3A). Samples were collected at the start of differential treatment, and after 6, 12, and 18 generations. DNA in the samples was isolated and the barcodes were amplified and sequenced. Resulting barcode counts at the different time points (Supplementary Table S3, Supplementary Figure S2) were used for enrichment analysis. Using an Enrich2 [22] based algorithm, competition coefficients were calculated for each barcode (Supplementary Table S4, Supplementary Figure S2). Using barcode counts from generations 0, 12 and 18 for this analysis yielded a clear separation of barcodes associated with OMeY-dependent growth from a background not showing this OMeY dependence (Figure 3C). Competition coefficients in both environments along with their estimate standard errors were aggregated to a Z-score. We had assumed that genes amenable for a differential response to OMeY would have comparably low expression levels, as to not overwhelm capacity of the orthogonal translation system. Indeed, target genes with barcodes linked to clear or strong OMeY dependence (Z-score >3 or >6) displayed lower median protein expression compared to all essential genes represented in the library. However, both groups still included fairly highly expressed genes (Figure 3E). In addition, the biological function appeared to be relevant, as genes involved in mRNA processing and RNA splicing were clearly enriched (Figure 3F).

The dataset also allowed investigating the preferred mRNA sequence context of amber stop suppression by the orthogonal translation system in *S. cerevisiae*. We looked at the positions flanking the UAG codon of edits with highly OMeY-dependent competition coefficients (Z-score >6, total of 98 edits in 54 genes) (Figure 3G). There was a clear sequence bias for the nucleotide directly downstream of the stop codon, with 51% of sequences having guanine at that position and only 4% uracil. From the same set of genes, this bias was not found for edits not displaying OMeY dependence, where present (Z-score <0, total of 79 edits in 41 genes) (Figure 3H). Similarly, analyses of context dependence of amber stop suppression in *E. coli* have shown a strong sequence bias of that position [14, 23], dubbed the ‘4th nucleotide effect’ [13].

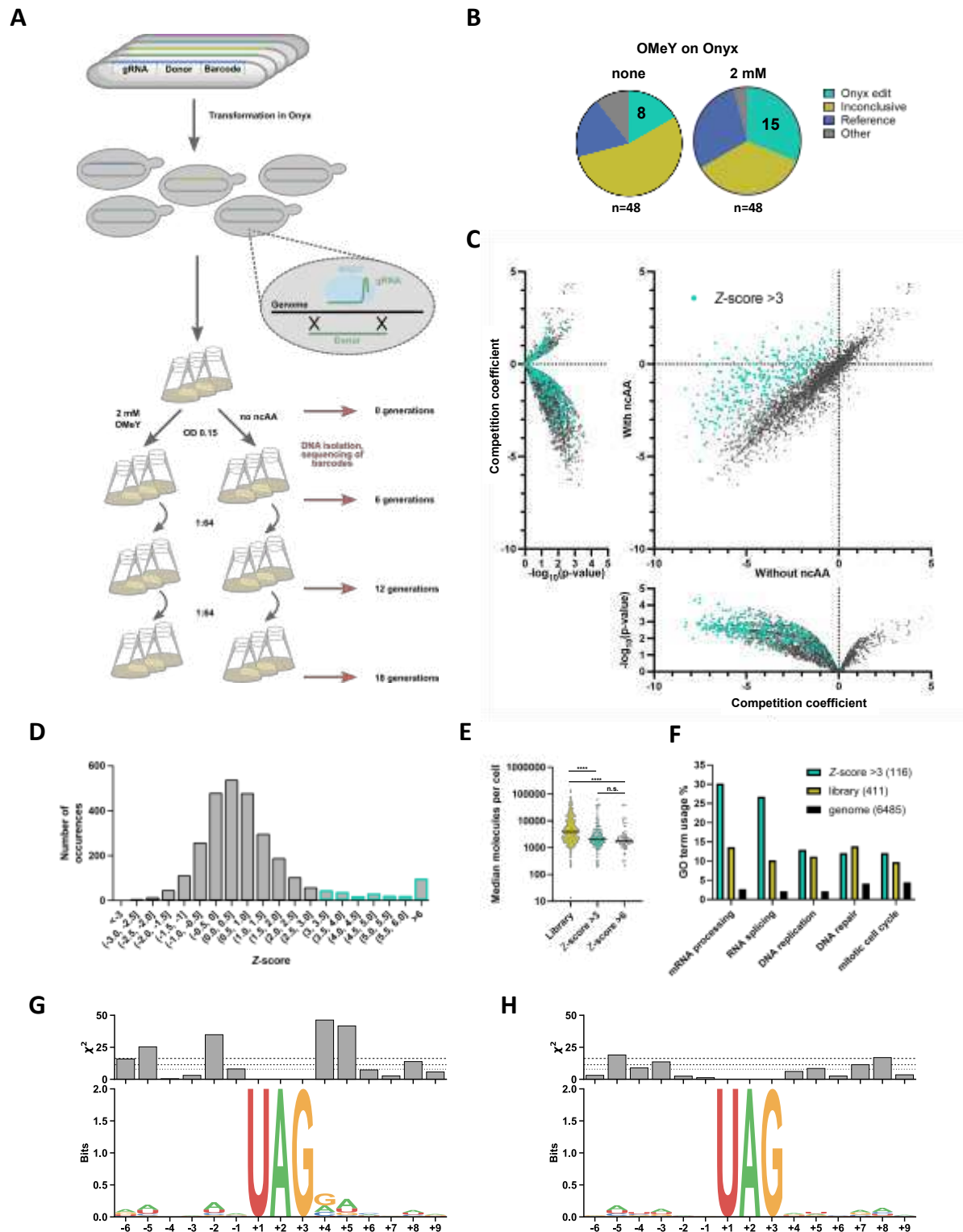


Figure 3: Library generation and screening. **A)** shows the workflow of library selection. Each plasmid species encoding a specific edit contains a unique barcode. The designed plasmid library for edit delivery was transformed into yeast on the Onyx platform. The resulting yeast library was grown in triplicate serial cultures with and without OMeY, with dilutions chosen to control the number of generations. DNA was isolated from cultures, and barcodes were amplified and sequenced. **B)** It was assessed whether it is necessary to supply OMeY during library preparation on the Onyx. The yeast library was once built on the machine without the ncAA present, and once with 2 mM OMeY in reagents used. Whole genome sequencing of 48 individual isolates from each built library showed that the fraction of correctly edited cells ('Onyx edit') was substantially higher with OMeY supplementation. Partial edits or edits with additional mutations were classified as 'Inconclusive', whereas the

lack of edits as ‘Reference’. **C)** The central scatter plot shows competition coefficients of edit barcodes in the absence of OMeY against coefficients in the presence. Volcano plots at its axes show coefficient estimates in the respective environment against $-\log_{10}$ of the p -values of parameter estimation. Barcodes with an overall Z-score >3 for the difference between competition coefficient with and without ncAA are shown in teal. **D)** Histogram of Z-value distribution of barcodes. Bars with Z-score >3 are shown with a teal outline. **E)** Shown are median protein molecules per cell (from [24]) for the essential genes sufficiently represented in the library ($n=397$), as well as for genes for which at least one edit barcode had a Z-score of >3 ($n=116$) and >6 ($n=54$). Dots represent individual genes and the horizontal dashes the median of each category. Difference between each of the groups were tested with Mann-Whitney test. *** indicates $p<0.001$. **F)** shows Biological Process GO term usage of the genes from **E** and from the whole genome. Shown are the five most common GO terms among genes associated with a Z-score >3 . **G)** Sequence logo [25] of the sequence context of edits showing highly OMeY-dependent competition coefficients (Z-score >6) edits, which are a total of 98 across 54 genes. Two codons up- and downstream of introduced UAG codons are shown. Relative to a uniform nucleotide distribution, χ^2 values were calculated from the nucleotide frequencies at respective positions, and are displayed above the sequence logo with indication of significance thresholds for $\alpha = 0.05$ (dotted line), $\alpha = 0.01$ (line with short dashes), and $\alpha = 0.001$ (line with medium dashes) (each with $df = 3$). **H)** For comparison, the same sequence context analysis was carried out on edits not linked to OMeY dependence (Z-score <0) in the set of genes from **G** that had highly OMeY-dependent edits (Z-score >6), which comprised 79 edits across 41 genes.

Assaying top hits from large scale screen

The 48 associated with the highest Z-score in the Onyx screen were chosen to be made individually with a CRISPR/Cas9 based approach. Some of these designs included helper edits in addition to the intended TAG edits, serving to immunize the correctly edited locus against cutting by the CRISPR/Cas system. Any helper edits in the Onyx delivery designs for MAD7-based editing were not replicated, but where necessary, unrelated helper edits for Cas9-based editing were included. A list of the Cas9-mediated TAG edit designs is given in Supplementary Table S5. Of those TAG edits, 46 were successfully introduced in individual strains, whereas two of the 48 highest-ranking edits, both in the gene *DBF4*, could not be made. The 46 strains were assayed for growth with and without OMeY (Figure 4A). The area under curve of the difference between growth with and without OMeY was used as a measure for OMeY dependence of growth ($AUC_{\text{difference}}$), and the area under curve of the growth with OMeY as a comparable metric for fitness under permissive conditions ($AUC_{+\text{OMeY}}$).

Strains that showed both strongly OMeY-dependent growth and high fitness recovery were characterized more closely to investigate the suitability of their TAG edits as genetic safeguards. For each strain, growth in liquid culture was assessed under varying OMeY concentrations. Further, saturated cultures were washed and plated without OMeY (restrictive condition) to gauge their afforded containment stringency. Unexpectedly, strong growth suppression without the ncAA observed in growth assays did not necessarily translate into low escape frequencies, as for instance seen for *MOB1* Y167* and *CDC5* Y608* (Figure 4E and G).

When plated on restrictive medium for most of the strains we observed the early appearance (after 2-4 days) of normally growing colonies, and the later appearance of slowly growing colonies. To investigate the nature of these different kinds of escapees, we picked a large, a medium and two small escapee colonies derived from *PRP24* E184* after 10 days of incubation into permissive medium. Performing growth assays with and without OMeY showed that the strains derived from small colonies were still partially dependent on OMeY supplementation for growth recovery, and had accrued a fitness defect even when OMeY was available (Supplementary Figure S3). For these escapees, no causative mutations could be identified. Likely, they have acquired a $[PSI^+]$ state. These prions of the translation termination factor Sup35p are known to increase unspecific translational read-through [26], and can cause fitness defects of varying severity [27]. Whole genome short-read sequencing revealed a definitive

causative mutation only for the fast growing escapee; it had acquired a mutation in the anticodon loop (c.34C>T) of the tRNA^{Trp} gene tW(CCA)J, turning it into a UAG stop decoder. The occurrence of amber suppressors through mutation of anticodons in endogenous tRNA genes has been observed to be the most common escape mode in previous work on creating ncAA dependence in yeast [6]. It has been shown that this escape route can be blocked by CRISPR/Cas targeting of frequently observed suppressors.

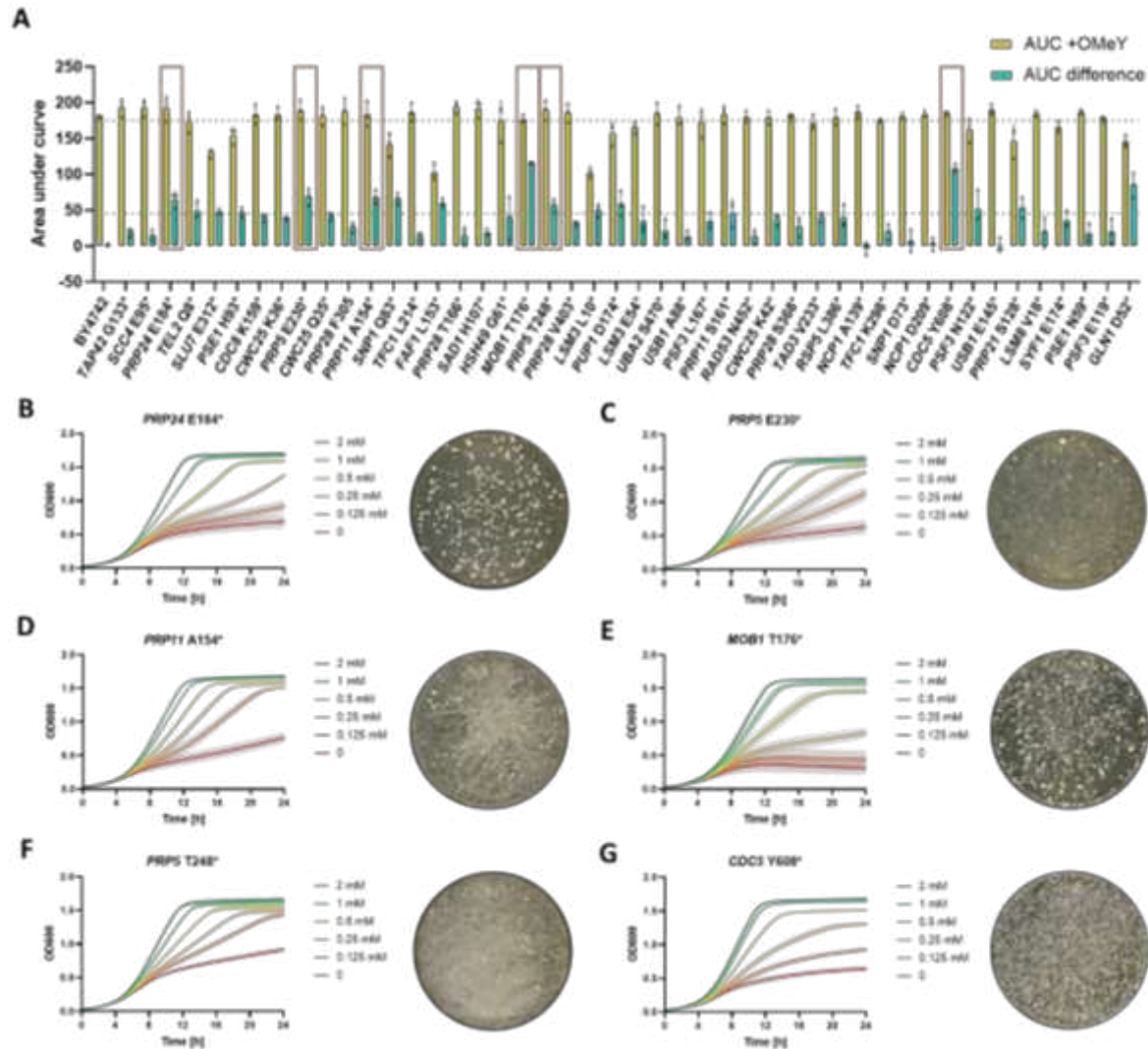


Figure 4: Assaying individual TAG edits from library screening. A) 46 of the 48 TAG edits with the highest Z-scores in the library screening were individually made and screened for growth dependence on OMeY ($AUC_{\text{difference}}$), and reconstitution of high fitness in its presence at 2 mM ($AUC_{+\text{OMeY}}$) ($n=3$, dots represent individual values, bars represent means $\pm$ SD). B-G) 6 strains with a strong growth response to the ncAA ($AUC_{\text{difference}} > 50$) and high fitness in its presence ($AUC_{+\text{OMeY}} > 175$) were characterized for their OMeY response by growth assays in YPD with OMeY concentrations between 0 and 2 mM ($n=3$, shown are means $\pm$ SD as shading). Further, dense overnight cultures were washed, plated on restrictive medium (YPD agar) and grown for 5d at 30°C to gauge containment stringency.

Combining TAG edits

None of the selected TAG edits individually provided stringent containment, with estimated escape frequencies in the order of 10^{-3} to 10^{-2} . Thus, we explored combinations of TAG edits to increase stringency while maintaining fitness (Figure 5, Supplementary Figure S4). Despite its good growth recovery with OMeY, we excluded *CDC5 Y608** from further scrutiny due to its low stringency. Generally, combination of multiple TAG edits led to a decrease of fitness

with 2 mM OmeY, an effect that was particularly evident with combinations involving *MOBI* T176*.

In contrast, combinations with *PRP24* E184* seemed to not decrease fitness when combined with select TAG edits in *PRP5* or *PRP11*. Interestingly, the gene product of each of these three genes are part of the spliceosome or involved in its formation. Spliceosome-related genes were strongly overrepresented in the edits associated with the strongest ncAA dependence. The combination of one TAG edit in each of these genes (*PRP24* E184*, *PRP5* T248* and either *PRP11* A154* or *PRP11* S161*) afforded fairly stringent containment with 10-day escape frequencies of less than 7×10^{-6} on rich media. However, these triple TAG strains exhibited a moderate, but clear growth defect in the permissive condition (2 mM OMeY) (Supplementary Figure S5). Introduction of another copy of *EctRNA*_{CUA}^{Leu}, encoding the cognate tRNA of the used orthogonal translation system, at the *ura3* locus indeed partially rescued this phenotype, and shifted OMeY requirements to lower concentrations. However, it also led to a marked loss of containment stringency, with too many clones appearing after 5 days to be countable when plating an undiluted dense culture on restrictive medium (Figure 5F).

In addition to combining TAG edits distributed across different genes, we were interested in whether introducing multiple TAG edits in a single essential gene would afford more stringent containment. Inserting two TAG stops in *PRP5* and *PRP11* appeared to be lethal in the base strain (SHy081) with a single OTS inserted in the *ho* locus, as no such mutants could be generated. However, in a strain with a second orthogonal tRNA inserted in the *ura3* locus, creating double TAG mutants *PRP5* E230* T248* and *PRP11* A154* S161* was possible (Supplementary Figure S6). The *PRP11* A154* S161* strain showed a substantial growth defect with 2 mM OMeY supplementation. In contrast, *PRP5* E230* T248* showed only a slight growth defect at 2 mM OMeY, and afforded containment with an average 10-day escape frequency of 2.3×10^{-5} on restrictive medium.

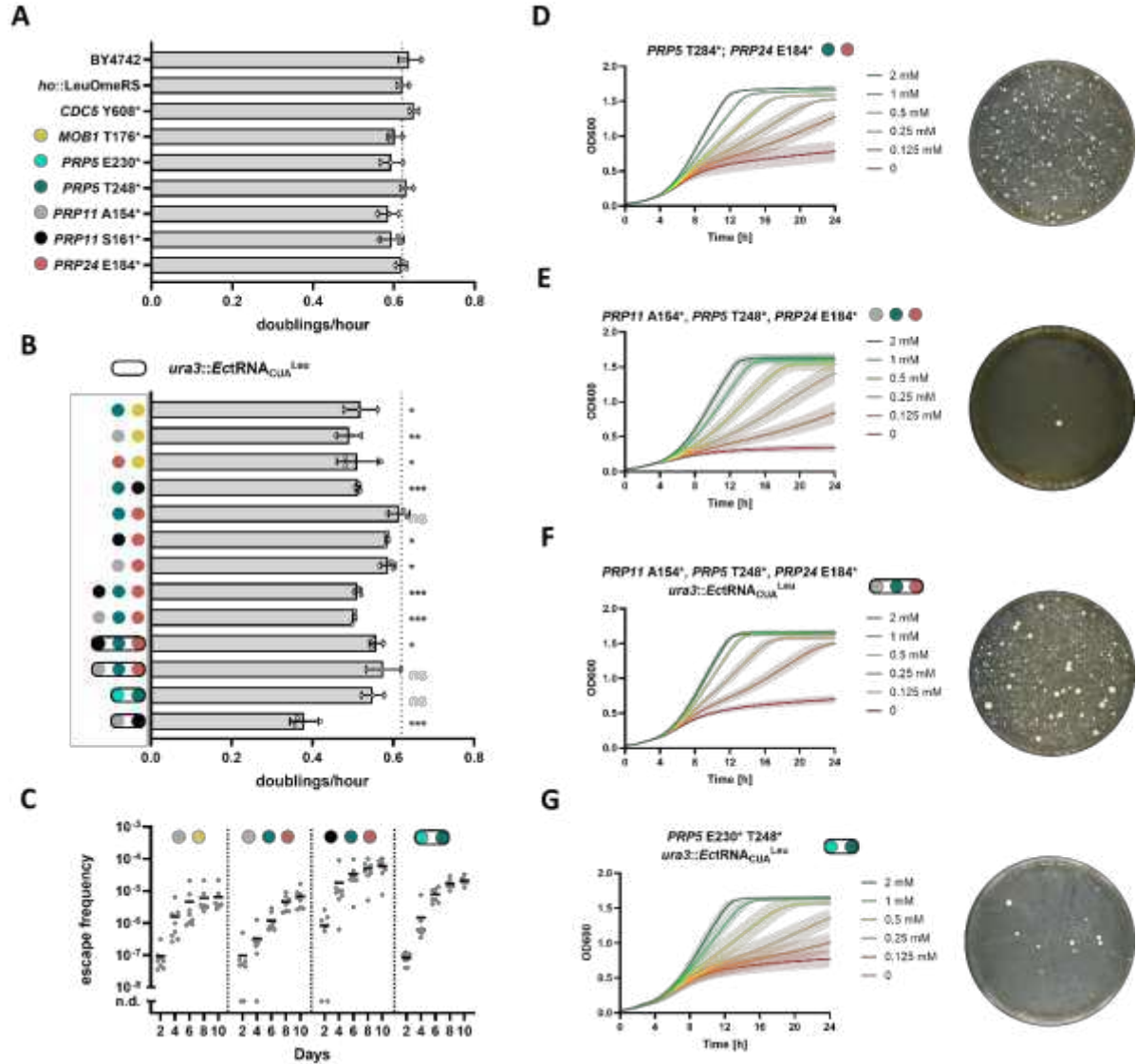


Figure 5: Combination of TAG edits to increase containment stringency. **A)** Growth rates of unedited BY4742, the strain with the genomically integrated orthogonal translation system and derived strains with single TAG edits, all with 2 mM OMeY ($n=3$, dots represent individual values, bars represent means $\pm$ SD). **B)** Growth rates of strains with multiple TAG edits with 2 mM OMeY ($n=3$, dots represent individual values, bars represent means $\pm$ SD). Strains with an additional orthogonal tRNA are marked with a black frame. Asterisks indicate significance in two-sided t-test comparing to growth rates of the base strain SHy081 (*: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$). **C)** Escape frequencies of select contained strains over 10 days at 30°C ($n=8$) with individual values shown as gray dots, and their means as black bars. The two strains with three TAG edits (without second tRNA) had about the same growth rate as *MOB1 T176**; *PRP11 A154** with 2 mM OMeY, but about 10 times lower escape frequencies. *PRP5 E230** *T248** only had a slight growth defect, and more stringent containment than *MOB1 T176**; *PRP11 A154**. **D-G)** Growth of strains with multiple select TAG edits in essential genes in liquid culture in YPD with a range of OMeY concentrations ($n=3$, shown are means $\pm$ SD as shading), and growth on permissive medium YPD agar (containment escape) observed after 5d at 30°C. Addition of a third TAG edit increased stringency, but caused a slightly slower growth with 2 mM OMeY supplementation. Introduction of another copy of the orthogonal tRNA largely removed the growth defect and shifted the response to OMeY to lower concentrations, but abolished containment stringency. Combining two TAG edits in *PRP5* was possible by adding a second copy of the orthogonal tRNA, and resulted in only a slight growth defect, but afforded fairly stringent containment.

Combination with protein stability containment system

In the genomic background used here, stringent containment through multiplexing of TAG edits appeared to come with a trade-off with fitness in the permissive condition, YPD with 2

mM OMeY. We explored whether the combination of synthetic auxotrophy with another biocontainment system results in increased containment stringency compared to each system individually without compromising fitness. Thus, we combined individual fitness-neutral TAG edits with a genetic biocontainment system based on estradiol-dependent stability of essential proteins we previously developed [10]. This approach leverages a destabilizing domain decon derived from the ligand-binding domain of the human estrogen receptor. It has been engineered such that in the absence of its ligand it is in an unordered conformation, while in the presence of estradiol it is stable. We used this differential behavior to convey conditional, estradiol-dependent stability to essential proteins, and screened for the most suited protein targets.

Here, we introduced a C-terminal *ERdd* to the *SPC110* gene, which we had found to be an effective target to create stringent estradiol dependance [10]. This was done for the *ho::LeuOmeRS* base strain SHy081, as well as the derived ncAA auxotroph strains with *PRP5* T248* and *PRP24* E184* TAG edits, respectively (Figure 6A). All three resulting strains showed wild-type-like growth in permissive medium containing both 2 mM OMeY and 1 μ M β -estradiol. As expected, the two strains with both the *ERdd* and the TAG edit were now dependent on both estradiol and OMeY, and showed lower escape frequencies in restrictive medium than afforded by the *SPC110-ERdd* edit alone, which showed an escape frequency of 8.7×10^{-8} (Figure 6B). The combination with the *PRP24* E184* TAG edit in particular resulted in a notably increased containment stringency, with escape undetectable in our standard escape assay with a detection limit for escape frequency of about 3×10^{-9} . Accordingly, we carried out an escape assay to detect low frequency escape events. Plating a total of 1.3×10^{10} cells on restrictive medium, replica plating after 2 days, and incubating for another 10 days, we obtained a single escapee, amounting to an escape frequency of 7.5×10^{-11} for the *PRP24* E184*; *SPC110-ERdd* strain.

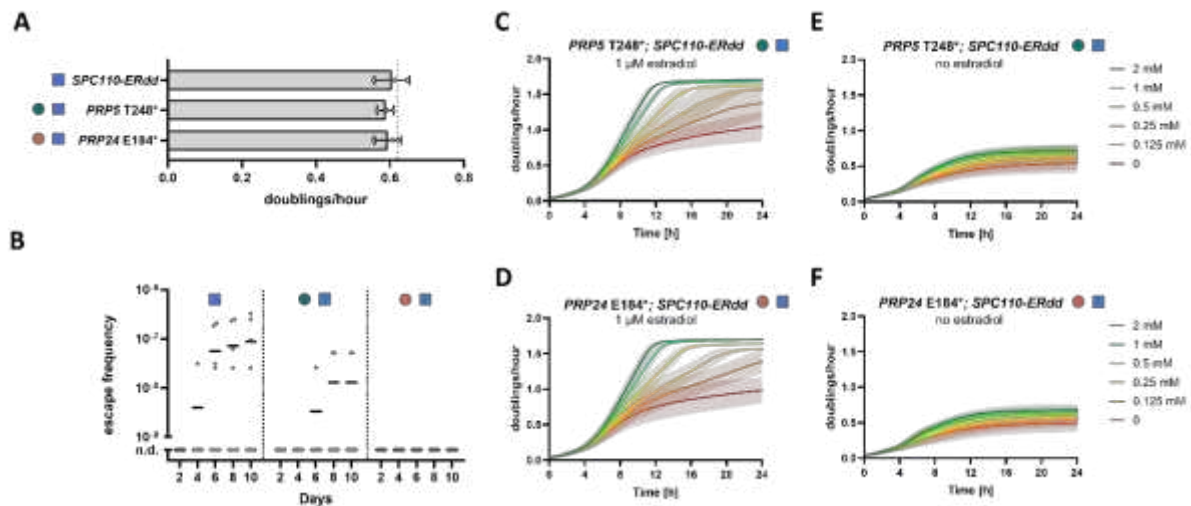


Figure 6: Combination of synthetic auxotrophy with ERdd-based genetic biocontainment system. Based on the strain SHy081, an ERdd tag was integrated as a C-terminal fusion to *SPC110*, either alone (*SPC110-ERdd*), or in combination with either the TAG edit *PRP5* T248* or *PRP24* E184*. **A)** Calculated growth rates in permissive medium (YPD + 2 mM OMeY + 1 μ M β -estradiol) (n=3, dots represent individual values, bars represent means $\pm$ SD). **B)** Assessed escape frequencies in standard escape assay (n=8) with individual values shown as gray dots, and their means as black bars. For the strain with the ERdd tag and *PRP24* E184*, not a single escapee was observed here. **C-F)** Growth of strains with combined containment systems in liquid culture in YPD with a range of OMeY concentrations and with (C and D) and without (E and F) estradiol (n=3, shown are means $\pm$ SD as shading).

Discussion

Synthetic biology holds immense promise to tackle central issues humanity is facing, for instance in resource management, environment, and health care. For open-environment release of engineered microorganisms, suitable safety measures need to be developed and put into place to prevent unintended consequences and gain regulatory and public approval. We believe an important safety layer will be genetically encoded, intrinsic biocontainment, in which microorganisms have been engineered to allow control over their survival [3]. In previous work, we have engineered systems that enable small-molecule control of survival through essential genes at different points along the central dogma, interfering with the presence of essential genes [28], their transcription [9, 28], and stability of their gene products [10]. We engineered these systems with modularity in mind, aiming to create mutually orthogonal methods to allow their eventual combination for extremely stringent containment. An important consideration is that each individual system allows the contained organism to grow as well as the wild type in permissive environments. Incurred fitness defects are likely to compound when combining multiple systems, creating evolutionary incentives to inactivate biocontainment systems. Further, since slower growth of a microbial chassis reduces its productivity, such fitness defects would also pose economic disincentives for the implementation of these systems.

Here, we sought to obtain control of the translation of essential genes, creating dependence on a non-canonical amino acid (ncAA). The principle on the targeted introduction of TAG stop codons in essential genes, and suppression of these stop codons by an orthogonal translation system engineered only in the presence of suitable ncAAs. This general approach has been used in *E. coli* to create very stringent containment [4, 5]. It has recently also been applied to yeast [6], screening several genes and positions leveraging the temperature-sensitive essential gene collection [29]. Here, the OTS was supplied on a centromeric plasmid, as was the queried TAG edited gene for screening. Two genes, *CDC4* and *CDC27*, were chosen for genomic TAG introduction, being rescued by an OTS still hosted on a centromeric plasmid. Strain growth and containment characteristics were determined in defined media. We aimed to create biocontainment based on synthetic auxotrophy with entirely genomically encoded components, and assessing performance in rich medium.

However, when we integrated the OTS system genomically a drastic change in system behavior was observed. Most genomic edits that we had characterized with the OTS on a plasmid, both in our proof-of-concept target *PDII* (K341*), as well as TAG edits previously reported to accommodate a very high degree of rescue through the same OTS (*ALG1* F250* and *CDC27* R14*) [29], could not be generated in this background. The edit we could generate, *PDII* D375*, now showed a severely compromised system behavior, with poor growth in the presence of the ncAA. A possible workaround could be genomic introduction of multiple copies of the OTS. However, we reasoned this might be a source of genetic instability and containment escape, for instance due to facilitating copy number change through homologous recombination or increasing the frequency of mutations in OTS genes that allow incorporation of endogenous amino acids. Instead, we opted for a large-scale screen for genes suitable for the containment approach in this background with apparently lower OTS activity.

Since the possible sequence space of introducing TAG edits into essential proteins is very large, it must be restricted computationally before experimental assessment. An important characteristic of a target position in a given essential protein is that it must be able to accommodate the ncAA sidechain. In previous work, this problem had been tackled by looking

at residue conservation through multiple sequence alignments [5, 6]. Tools used for this purpose such as SIFT [16] have been developed to discriminate functionally neutral and deleterious mutations. While relatively computationally expensive, its classification into substitution-tolerant and intolerant residues based on analysis of sequence conservation proved a tractable way to choose sites for recoding. However, we reasoned that the present use case does not actually require an exact classifier of individual positions, but solely a prediction of a handful of positions that are tolerant specifically to bulky substitutions for each essential protein. We developed a simple structure-based algorithm to identify positions very likely to be tolerant to substitution with bulky amino acids, and ran it locally with available experimentally determined structures of yeast essential proteins.

For experimental assessment of about 4000 shortlisted positions, we made use of Inscripta's Onyx platform. Its library generation and screening approach rests on barcoded editing plasmids that are being maintained throughout to track depletion or enrichment of barcodes associated to individual encoded edits; the same principle can be employed without this proprietary platform. In our case, we compared barcode competition coefficients in the presence and absence of the ncAA. The majority of barcodes did not show significant OMeY dependence. A main reason is likely that many TAG edits cannot be rescued in this background, in line with observations in the proof-of-concept experiments. Respective cells may maintain the editing plasmids through partial edits immunizing against Cas mediated cutting, or through mutational inactivation of Cas system components on the plasmid, or the delivery designs being ineffective at targeting. Looking at individual clones of the library, only about a third contained the intended edit encoded on their editing plasmid.

In addition to identifying the most promising edits, the high-throughput screening based on pooled library generation and barcode tracking allowed general inferences on suitable target genes and positions. Expression strength of the target gene appeared to be less important than we had assumed, with edits in highly expressed genes being among OMeY-dependent edits. Still, average expression strength of proteins with OMeY-dependent edits was significantly lower in OMeY-dependent edits compared to the overall pool. Curiously, essential proteins involved in RNA processing were very strongly overrepresented among proteins with OMeY-dependent edits. Accordingly, promising edits that merited detailed characterization were found in proteins functionally related in pre-mRNA processing, namely *PRP5*, *PRP11* and *PRP24*. The mechanistic basis for the observed amenability of RNA processing genes to this approach remains unclear.

Regarding suitable positions within a given gene, it is known that the sequence context influences the efficiency of OTS. In *E. coli*, the nucleotide directly downstream of the target codon has been shown to be an important factor for amber stop suppression efficiency, with purines being the most favorable [13]. Interestingly, we observed a very similar pattern here, with the +4 position exerting the strongest effect and a strong bias for guanine or adenosine among the strongly OMeY-dependent edits in this position. This similarity to the context preference in *E. coli* is notable, particularly because it contrasts with observations in mammalian cells, where the +6 position was reported to be the strongest determinant and cytosine was favored at the +4 position [30]. While selection of edit sites in our initial library design was solely structure-based, preferred sequence contexts for stop codon suppression should ideally be considered in future library designs.

As mentioned before, we aim to generate containment systems that do not impair fitness in the permissive condition. In the present background, individual TAG edits that fulfilled this condition did not offer stringent containment. For most of the respective strains, we saw early appearance of some fast-growing escapee colonies, and delayed appearance of usually very many slow-growing escapees, which seemed to still be partially dependent on OMeY. Combination of multiple of these edits usually led to improved stringency, but also mild fitness impairments increasing with the number of consolidated TAG edits. A maximum of two select TAG edits (*PRP5* T248*; *PRP24* E184*) could be combined without a noticeable fitness defect. Reasoning that expression of the orthogonal tRNA is a likely bottleneck for ncAA incorporation, we added a second copy to the genome. This shifted system behavior, improving fitness for triple TAG strains, and made the combination of two TAG edits in a single essential gene possible in the first place. However, in the former case, it also degraded containment efficiency. We assume that a background with higher OTS activity could yield systems that perform better, provided that OTS misincorporation activity is not increased. While the genomic introduction of multiple OTS copies has likely drawbacks, as discussed earlier, other avenues to enhancing genetic code expansion could enable improved synthetic auxotrophy systems.

Ourselves and others have sought to increase ncAA incorporation efficiency in yeast by changing its genetic makeup [31-34], but there is likely a ceiling to possible improvements through knocking out select genes or changing their expression. The aforementioned highly stringent ncAA dependence that had been engineered in *E. coli* was underpinned by highly engineered host strains devoid of endogenous TAG stop codons and the non-essential release factor RF1 [4, 5]. A comparable eukaryotic host system does not yet exist. However, a TAG stop codon free yeast is being built within the Sc2.0 project which aims to create the world's first synthetic eukaryotic genome. Synthetic versions of all chromosomes and an accompanying neochromosome housing tRNA genes relocated from their original loci [35] have been created, and are now to be consolidated in a single strain [36]. This presents an impending opportunity to advance applications leveraging genetic code expansion in yeast, including its use for improved synthetic auxotrophy systems. However, for efficient incorporation of ncAAs through amber stop suppression, competition with translation termination presents another hurdle in yeast. In *E. coli*, release factor RF1 could be removed because of functional redundancy. Its removal allowed efficient recoding of TAG codons [37]. In contrast, eukaryotic release factor eRF1 is an essential gene responsible for translation termination of all three stop codons. Engineering a eukaryotic release factor that does not recognize TAG stop codons is a promising route to increase ncAA incorporation efficiency and with it the achievable performance of amber suppression-based biocontainment systems.

Overall, we demonstrate a generalizable pipeline to create synthetic auxotrophy for ncAAs. The structure-based approach of pre-selection of positions can feasibly be implemented for other organisms. For non-model organisms there will be fewer available experimentally determined essential protein structures. However, the high performance of AlphaFold on most proteins [38] should allow using its structure predictions instead. Therefore, the approach only requires the sequences of essential genes to be known for a given organism. The large-scale screening method we have developed for experimentally assessing thousands of edits does not rely on specific genetic resources typically only available just for model organisms, such as strain collections, but solely requires genetic tractability with a CRISPR-Cas editing system. We have shown that our workflow allows identifying suitable edits from the large sequence space of possible positions in hundreds of target genes. We chose edits conveying ncAA dependence while allowing full growth recovery in its presence. While individually not

offering stringent containment, we showed the resulting system with the best performing TAG edits can be effectively combined with an orthogonal containment system to increase containment stringency. This demonstrates that the layering of different kinds of genetic biocontainment systems is a promising route to creating extremely stringent containment while maintaining high fitness in permissive conditions.

Methods

Prediction of tolerant residues

The RCSB Protein Data Bank (rcsb.org) was searched for structures of essential yeast proteins (rdrr.io/bioc/SLGI/man/essglist.html) using their Search API. Suitable structures were collated and retrieved using RCSB's Data API. We developed an algorithm to rapidly predict positions that tolerate substitution with bulky amino acids without disrupting protein function. It is packaged as an add-on to ChimeraX and all processes are executed locally (github.com/the-cai-lab/Cai-lab-ChimeraX-script). Within the downloaded PDB files, peptide chains corresponding to targeted essential proteins were located and the chains collated into a list. For each protein, the following analysis was performed. Firstly, the solvent accessible surface area of each residue was calculated with a radius of 1.4Å. Following this, each residue was sequentially substituted with tryptophan and arginine while original rotamer angles were retained and any clashes with adjacent atoms are recorded for each substitution. Once all the residues have been analyzed, the positions were validated against a FASTA file, referencing the amino acid identity of each residue in the structural file with the corresponding one in the genomic coding sequence. Only residues with matching identities between the PDB and FASTA file are included in the output file, any residues that do not match are not included. Positions with a total number of clashes of less than 3 and a solvent accessible surface area greater than 20Å² are assumed to be highly likely to be tolerant to substitutions.

Point mutation editing library design and generation

Point mutation editing libraries were constructed using high-throughput CRISPR-based editing provided by the Onyx Digital Genome Engineering platform. Briefly, repair templates were designed using Inscripta's Designer software so that specific amino acids would be replaced with the amber codon TAG or synonymous codons to serve as controls. In addition to the desired edit, if needed additional synonymous edits were included to prevent re-cutting by eliminating the PAM site and/or introducing edits that interfere with cutting. For synonymous recoding in these helper edits, the most frequently used available codon was chosen. A specific genetic barcode is associated with each design.

Cell library construction

Edited library pools were generated on the Onyx Digital Genome Engineering Platform, a commercial benchtop instrument by Inscripta, Inc. using the MAD7 nuclease, a type V CRISPR nuclease from *Eubacterium rectale*, to generate multiplexed genome engineered libraries. Genome editing was performed using developmental reagents and protocols optimized for *S. cerevisiae* S288C (Onyx™ Engineering Handbook *E. coli* and *S. cerevisiae*. 2022. 1001178). A single colony of SHy081 was isolated from a YPD agar plate and grown overnight in YPD with 100 µg/mL carbenicillin and 2 mM of O-methyl-L-tyrosine (Sigma Aldrich product #158259) prepared in an alkaline solution (25 mM NaOH) sterile filtered using a 0.2 µm filter. The overnight culture was adjusted to an OD₆₀₀ of 5.0 in the same medium, and 2 mL of this adjusted culture was loaded onto the instrument for editing using the Onyx® *S. cerevisiae* Genome Engineering kit (Cat. #GEN-SC-1012) with the program 'Onyx *S. cerevisiae* S288C Engineering Protocol v1.1.0'. Onyx growth and recovery media (provided as part of the Genome Engineering Kit) were supplemented with 2 mM of O-methyl-L-tyrosine prepared as above.

Briefly, the instrument diluted and transferred the input cells to a cell growth cuvette, where growth was monitored until reaching an OD₆₀₀ of 1.5, as measured by the instrument. After

this initial outgrowth, cells were turned electrocompetent. The MAD7-containing editing plasmid as well as the linear gRNA/repair template/barcode-containing DNA were transformed into cells by electroporation. Following electroporation, cells recovered in a second cell growth cuvette for recovery. The resulting normalized pool of edited cells was collected as multiple tubes from the instrument. Approximately 5 mL of edited cell library were collected. Cells were spun down and resuspended in YPD containing 15% glycerol, 200 µg/ml hygromycin B and 2 mM O-methyl-L-tyrosine and immediately stored at -80°C. Depending on cell growth, total run time on the instrument was around 72 hours.

Illumina sequencing for detection of mutations in saturation editing libraries

Genomic DNA for the library build and selection samples were isolated according to standard Inscripta protocols (Onyx™ Genotyping Handbook *E. coli* and *S. cerevisiae*. 2022. 1001182 RevB.), which uses the Wizard SV Genomic DNA Purification System (Promega, Cat. #A2365/A2360/A2361). PCR amplification of the barcode was performed using Q5 Hot-Start PCR Master Mix with a total volume of 50 µL. The reaction was carried out in a BioRad T100 thermal cycler as follows: 98°C for 2 min; 20 cycles of 98°C for 10s, 72°C for 2 minutes 30s, with a final extension at 72°C for 5 min. PCR fragments were prepared for sequencing according to the Inscripta Onyx™ Genotyping Handbook (Onyx™ Genotyping Handbook *E. coli* and *S. cerevisiae*. 2022. 1001182).

Designs were quantified using Inscripta's proprietary genomic amplicon edit detection pipeline. Briefly, this approach performed competitive alignment to determine the origin of each read. Each read was aligned to an augmented reference, which contained the original unedited reference genome, supplemented by a set of alternative contigs. For each design, the alternative contigs consisted of the repair template (corresponding to the Onyx edit) in its genomic context and in the cassette backbone context. Alignment was performed using BWA-MEM [38]. Ambiguously mapped reads were remapped to the unedited reference. Each read was categorized as either providing evidence for the complete intended edit ('Onyx edit' reads), no edit ('reference' reads), or insufficient evidence to make a call. Importantly, an 'Onyx edit' consists of both the intended non-synonymous variant as well as any required additional 'immunizing edits' that prevent MAD7 recutting by modifying the PAM sequence with synonymous edits (if the PAM occurs in a gene CDS). Thus, false positive 'Onyx edit' calls are unlikely to result from rare Illumina sequencing errors, as the error would have to occur in multiple, non-contiguous nucleotides.

Library screening experiment and analysis

Library screening experiments were done with three biological replicates. For each, from a glycerol stock of the yeast library, 25 ml YPD with 2 mM OMeY and 200 µg/ml hygromycin B were inoculated with about 2×10^7 colony forming units and grown at 30°C for 8 h. Cells were then washed once with 1 ml YPD and used to inoculate two 20 ml YPD cultures with hygromycin B to an OD₆₀₀ of about 0.15, so that cells could undergo roughly 6 doublings before reaching stationary phase. The rest of the washed cell suspension was split into two, pelleted and stored at -20°C for DNA isolation. One culture was supplemented with 2 mM OMeY, the other contained no ncAA (but 2 mM NaOH as vehicle control). Cultures were incubated for 24 h at 30°C. From each culture, 2x200 µl culture were pelleted and stored at -20°C for DNA isolation. Cultures were backdiluted 1:64 in 20 ml of the respective medium, grown for another 24 h at 30°C, undergoing 6 generations. Samples were taken as before, and cultures back diluted 1:64, grown and sampled again once more, such that cells underwent a total of 18 generations with and without OMeY.

From the stored cell samples, DNA was isolated and barcodes amplified as described above. Barcode amplicons from the cell population were sequenced to quantify changes in the abundance of the edits over time. DNA isolation, barcode amplification and sequencing was once for the initial cultures and twice for each subsequent sample, giving a total of three biological replicates with two technical replicates for each of the latter. Only barcodes with a minimal total count of 120 across both conditions and all time points (average of 20 counts per technical replicate) were considered. A weighted least squares log-linear model was fit to model the change in abundance of each mutation using Enrich2 [22]. Regression weights were calculated by Enrich2 according to the Poisson variance of each mutation's read counts across replicates, accounting for variable measurement error rates that depend on sequencing depth. The slope of the linear fit was taken and normalized to obtain a mutation's 'competition coefficient' as a quantitative measure of its relative fitness in the condition. Competition coefficients were normalized to the total read counts in the sample and adjusted such that a specific subset of control edits assumed to be fitness-neutral (median of synonymous edits in non-essential genes, n=50) was set to 0. Competition coefficient values greater than 0 indicate improved growth, whereas scores less than 0 indicate impaired growth in the respective condition.

To identify edits leading to OMeY-dependent growth, Z-scores were calculated from the competition coefficients and their respective standard errors according to Eq. (1) as follows.

$$Z = \frac{\text{comp.coeff}_{+\text{OMeY}} - \text{comp.coeff}_{-\text{OMeY}}}{\sqrt{\text{SE}_{+\text{OMeY}}^2 + \text{SE}_{-\text{OMeY}}^2}} \quad (1)$$

The 48 TAG edits with the highest Z-scores were chosen to be individually made with a CRISPR/Cas9 based editing system.

Manual genome editing

Top ranked TAG edits from the primary Onyx screening as well as other select TAG edits and insertions of OTS components were made with a yeast CRISPR/Cas9 tool kit. For bulk design of edits, targets were chosen and donor constructs designed using a custom CRISPR design script (github.com/the-cai-lab/High-throughput-CRISPR-design), which introduces synonymous, usage frequency-matched helper edits when necessary.

For creation of the gRNA constructs, pairs of oligonucleotides containing the respective target sequence and bases to create required overhangs were annealed. Subsequently, they were cloned into the gRNA entry vector pWS082 (Addgene #90516) in a Golden Gate reaction (30 cycles of 1 min at 37°C, 1 min at 16°C and heat inactivation for 5 min at 80°C) with 40 nM of annealed insert, 50 ng of pWS082, 5U Esp3I and 20 cohesive end units T4 DNA ligase in 10 µl T4 DNA ligase buffer.

Editing of yeast strains carrying the orthogonal translation system (either supplied on a plasmid or integrated in the genome) was carried out by LiOAc transformation of the CRISPR components. Exponentially growing yeast cells were harvested, washed first with water and then with 100 mM LiOAc. The transformation mix contained cells from 5 ml culture in 400 µl with final concentrations of 33.3% PEG 3350, 100 mM LiOAc, 0.3 mg/ml boiled herring sperm DNA (Promega) and about 500 ng pWS174 (Addgene #90961) digested with Esp3I (Cas9 plasmid), 600 ng of the respective guide construct digested with EcoRV and 1000 ng of the respective donor DNA (Twist Bioscience). After thorough mixing, transformation cultures were incubated for 60 min at 30°C at 200 rpm and then for 20 min at 42°C. Cultures were spun

down and resuspended in 500 μ l YPD with 2 mM OMeY when needed. After a transformation outgrowth period of 1 h at 30°C under shaking, 100 μ l were plated on YPD agar with 100 μ g/ml nourseothricin and 2 mM OMeY and incubated for 3d at 30°C. Strains with multiple TAG edits were created with the same protocol, introducing one additional TAG edit per transformation round if edits were on different loci. Correctly edited colonies were identified by OMeY-dependent growth and confirmed by colony PCR and Sanger sequencing. An overview of individually generated strains is given in Supplementary Table S6.

Assaying of OMeY dependence

Manually edited strains representing the top ranking edits from the library screening which were screened individually in liquid culture. Strains were cultured in deep well plates in 1 ml YPD with 2 mM OMeY each overnight at 30°C under rigorous shaking and washed once with 1 ml YPD. In a 96 well F-bottom microtiter plate with lid, for each strain a sample of 200 μ l YPD with 2 mM OMeY and a sample of YPD with 2 mM NaOH (vehicle control) was set up and inoculated 1:100 with the respective washed cell suspension. The assay was run for 24 h at 30°C in a BioTek Synergy H1 plate reader, recording the optical density at 600 nm every 10 min while shaking between the measurement cycles. Readings were blank subtracted. The area under the curve was calculated with OMeY as a measure of fitness, and for the difference curve (with OMeY minus without OMeY) as a measure of OMeY dependence for each strain. Values from three independent runs were averaged. Strains with high values for both metrics were more closely characterized.

For detailed characterization of OMeY dependence, strains were cultured in YPD with 2 mM OMeY overnight at 30°C. Cultures were washed once in YPD. Cell suspensions were adjusted to an OD₆₀₀ 0.1 in YPD and used both for liquid culture assays and spot assays. Liquid culture assays were performed in microtiter plates as described above, using final concentrations of OMeY at 2 mM, 1 mM, 0.5 mM, 0.25 mM, 0.125 mM and no OMeY (all with a final NaOH concentration of 2 mM) and each in technical duplicates. For spot assays, log₁₀ dilution series of the OD₆₀₀ 0.1 cell suspensions to OD₆₀₀ 10⁻⁷ were prepared and 5 μ l of each dilution were spotted on YPD agar at the different OMeY concentrations used (2 mM, 1 mM, 0.5 mM, 0.25 mM, 0.125 mM and no OMeY), incubated at 30°C and imaged with a Phenobooth after two and three days.

Escape assays

Escape assays were carried out as previously described [10]. In short, 8 individual overnight cultures were washed and 100 μ l plated of undiluted on restrictive semisolid medium (YPD agar) to count escapee colonies. For reference, 10⁻⁵ dilutions were plated on permissive medium (YPD agar with 2 mM OmeY and 1 μ M estradiol where applicable). Restrictive plates were incubated for a total of 10 days at 30°C, and escapee colonies counted on days 2, 4, 6, 8 and 10. To assess escape frequencies below the detection limit of this standard escape assay (< 10⁻⁸), 50 ml of overnight cultures were washed twice, concentrated to 5 ml, and plated onto 10 petri dishes with restrictive medium. After two days of 30°C incubation, they were replicated onto fresh restrictive medium and incubated for 10 more days. Escape frequencies were calculated using estimates for total plated colony forming units obtained from colony counts on permissive medium (plating 100 μ l of 10⁻⁶ dilution).

References

1. Wright, O., G.B. Stan, and T. Ellis, *Building-in biosafety for synthetic biology*. Microbiology (Reading), 2013. **159**(Pt 7): p. 1221-1235.
2. Lee, J.W., et al., *Next-generation biocontainment systems for engineered organisms*. Nat Chem Biol, 2018. **14**(6): p. 530-537.
3. Hoffmann, S.A., et al., *Safety by design: Biosafety and biosecurity in the age of synthetic genomics*. iScience, 2023. **26**(3): p. 106165.
4. Mandell, D.J., et al., *Biocontainment of genetically modified organisms by synthetic protein design*. Nature, 2015. **518**(7537): p. 55-60.
5. Rovner, A.J., et al., *Recoded organisms engineered to depend on synthetic amino acids*. Nature, 2015. **518**(7537): p. 89-93.
6. Chang, T., et al., *A robust yeast biocontainment system with two-layered regulation switch dependent on unnatural amino acid*. Nat Commun, 2023. **14**(1): p. 6487.
7. Mikkelsen, M.D., et al., *Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform*. Metab Eng, 2012. **14**(2): p. 104-11.
8. Cheng, Y.Y., et al., *Efficient plasmid transfer via natural competence in a microbial co-culture*. Mol Syst Biol, 2023. **19**(3): p. e11406.
9. Agmon, N., et al., *Low escape-rate genome safeguards with minimal molecular perturbation of *Saccharomyces cerevisiae**. Proc Natl Acad Sci U S A, 2017. **114**(8): p. E1470-E1479.
10. Hoffmann, S.A. and Y. Cai, *Engineering stringent genetic biocontainment of yeast with a protein stability switch*. Nat Commun, 2024. **15**(1): p. 1060.
11. Gan, R., et al., *Translation system engineering in *Escherichia coli* enhances non-canonical amino acid incorporation into proteins*. Biotechnol Bioeng, 2017. **114**(5): p. 1074-1086.
12. Stieglitz, J.T., et al., *A Robust and Quantitative Reporter System To Evaluate Noncanonical Amino Acid Incorporation in Yeast*. ACS Synth Biol, 2018. **7**(9): p. 2256-2269.
13. Chemla, Y., et al., *Context effects of genetic code expansion by stop codon suppression*. Curr Opin Chem Biol, 2018. **46**: p. 146-155.
14. Xu, H., et al., *Re-exploration of the Codon Context Effect on Amber Codon-Guided Incorporation of Noncanonical Amino Acids in *Escherichia coli* by the Blue-White Screening Assay*. Chembiochem, 2016. **17**(13): p. 1250-6.
15. Wu, N., et al., *A genetically encoded photocaged amino acid*. J Am Chem Soc, 2004. **126**(44): p. 14306-7.
16. Ng, P.C. and S. Henikoff, *SIFT: Predicting amino acid changes that affect protein function*. Nucleic Acids Res, 2003. **31**(13): p. 3812-4.
17. Yates, C.M., et al., *SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features*. J Mol Biol, 2014. **426**(14): p. 2692-701.
18. Bendl, J., et al., *PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations*. PLoS Comput Biol, 2014. **10**(1): p. e1003440.
19. Suckow, J., et al., *Genetic studies of the Lac repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure*. J Mol Biol, 1996. **261**(4): p. 509-23.
20. Michaelis, A.C., et al., *The social and structural architecture of the yeast protein interactome*. Nature, 2023. **624**(7990): p. 192-200.
21. Garst, A.D., et al., *Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering*. Nat Biotechnol, 2017. **35**(1): p. 48-55.

22. Rubin, A.F., et al., *A statistical framework for analyzing deep mutational scanning data*. Genome Biol, 2017. **18**(1): p. 150.
23. Stormo, G.D., T.D. Schneider, and L. Gold, *Quantitative analysis of the relationship between nucleotide sequence and functional activity*. Nucleic Acids Res, 1986. **14**(16): p. 6661-79.
24. Ho, B., A. Baryshnikova, and G.W. Brown, *Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome*. Cell Syst, 2018. **6**(2): p. 192-205 e3.
25. Schneider, T.D. and R.M. Stephens, *Sequence logos: a new way to display consensus sequences*. Nucleic Acids Res, 1990. **18**(20): p. 6097-100.
26. Derkatch, I.L., et al., *Genesis and variability of [PSI] prion factors in Saccharomyces cerevisiae*. Genetics, 1996. **144**(4): p. 1375-86.
27. McGlinchey, R.P., D. Kryndushkin, and R.B. Wickner, *Suicidal [PSI+] is a lethal yeast prion*. Proc Natl Acad Sci U S A, 2011. **108**(13): p. 5337-41.
28. Cai, Y., et al., *Intrinsic biocontainment: multiplex genome safeguards combine transcriptional and recombinational control of essential yeast genes*. Proc Natl Acad Sci U S A, 2015. **112**(6): p. 1803-8.
29. Li, Z., et al., *Systematic exploration of essential yeast gene function with temperature-sensitive mutants*. Nat Biotechnol, 2011. **29**(4): p. 361-7.
30. Bartoschek, M.D., et al., *Identification of permissive amber suppression sites for efficient non-canonical amino acid incorporation in mammalian cells*. Nucleic Acids Res, 2021. **49**(11): p. e62.
31. Zackin, M.T., J.T. Stieglitz, and J.A. Van Deventer, *Genome-Wide Screen for Enhanced Noncanonical Amino Acid Incorporation in Yeast*. ACS Synth Biol, 2022. **11**(11): p. 3669-3680.
32. Wen, J.R., et al., *Exogenous and endogenous optimization for efficient noncanonical amino acids incorporation in*. Chemical Engineering Journal, 2025. **508**.
33. Sanders, J.E.J., et al., *Synthetic yeast genome SCRaMbLEing uncovers a new role for ribosomal proteins in genetic code expansion*. bioRxiv, 2025.
34. Sanders, J., et al., *New opportunities for genetic code expansion in synthetic yeast*. Curr Opin Biotechnol, 2022. **75**: p. 102691.
35. Schindler, D., et al., *Design, construction, and functional characterization of a tRNA neochromosome in yeast*. Cell, 2023. **186**(24): p. 5237-5253 e22.
36. Zhao, Y., et al., *Debugging and consolidating multiple synthetic chromosomes reveals combinatorial genetic interactions*. Cell, 2023. **186**(24): p. 5220-5236 e16.
37. Johnson, D.B., et al., *RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites*. Nat Chem Biol, 2011. **7**(11): p. 779-86.
38. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold*. Nature, 2021. **596**(7873): p. 583-589.
39. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform*. Bioinformatics, 2009. **25**(14): p. 1754-60.

Acknowledgements

This work was supported by an EPSRC fellowship awarded to Y.C. (EP/V05967X/1), and grants from the BBSRC (BB/W014483/1 and BB/Y007743/1) and the Volkswagen Foundation “Life? Initiative” (Ref. 94,771).

The authors thank Dr. James Sanders for help preparing the yeast interactome map, and Dr. Aaron Brooks for helpful discussions on library data analysis.

Data Availability

Strains and plasmid are available from the authors upon request. Barcode sequencing data and whole genome sequencing data for escapee analysis has been deposited to SRA with the BioProject number PRJNA1431978.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTableS4.xlsx](#)
- [SupplementaryTableS2.xlsx](#)
- [SupplementaryTableS3.xlsx](#)
- [SupplementaryInformationSyntheticAuxotrophy.docx](#)
- [SupplementaryTableS6.xlsx](#)
- [SupplementaryTableS5.xlsx](#)