Host-Informed Expression of CRISPR Guide RNA for Genomic Engineering in Komagataella phaffii

HHS Public Access

Author manuscript

ACS Synth Biol. Author manuscript; available in PMC 2021 January 19.

Published in final edited form as:

ACS Synth Biol. 2020 January 17; 9(1): 26–35. doi:10.1021/acssynbio.9b00372.

Host-informed expression of CRISPR guide RNA for genomic engineering in *Komagataella phaffii*

Neil C. Dalvie[1,2], **Justin Leal**[1,2], **Charles A. Whittaker**[2], **Yuchen Yang**[1,2], **Joseph R. Brady**[1,2], **Kerry R. Love**[1,2], **J. Christopher Love**[1,2,*]

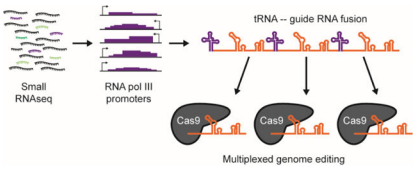
[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

[2]The Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 01239, United States

Abstract

There is growing interest in the use of non-model microorganisms as hosts for biopharmaceutical manufacturing. These hosts require genomic engineering to meet clinically relevant product qualities and titers, but the adaptation of tools for editing genomes, like CRISPR-Cas9, has been slow for poorly characterized hosts. Specifically, a lack of biochemical characterization of RNA polymerase III transcription has hindered reliable expression of guide RNAs in new hosts. Here, we present a sequencing-based strategy for the design of host-specific cassettes for modular, reliable, expression of guide RNAs. Using this strategy, we achieved up to 95% gene editing efficiency in the methylotrophic yeast *Komagataella phaffii*. We applied this approach for the rapid, multiplexed engineering of a complex phenotype, achieving humanized product glycosylation in two sequential steps of engineering. Reliable extension of simple gene editing tools to non-model manufacturing hosts will enable rapid engineering of manufacturing strains tuned for specific product profiles, and potentially decrease the costs and timelines for process development.

Graphical Abstract



*Correspondence to: clove@mit.edu.

Author contribution

N.C.D., K.R.L, and J.C.L. developed the concepts and designed the study. N.C.D. and J.L. performed the experiments. N.C.D., K.R.L., and J.C.L. wrote the manuscript. N.C.D. and C.A.W. performed bioinformatics analyses. Y.Y. and J.R.B. generated constructs for glycosylation engineering and testing.

Supporting Information

Attached: Plasmid sequences (see Table S4)

Attached: *K. phaffii* small RNA data (expression, trimness, and BLAST results)

Keywords

Pichia pastoris; RNA Pol-III; tRNA; recombinant protein; alternative host

Model microorganisms like *Escherichia coli* and *Saccharomyces cerevisiae* are commonly used for manufacturing complex molecules. Recently, interest has risen in the development of non-model microorganisms, including bacteria, yeast, and filamentous fungi as hosts for biomanufacturing and chemical processing, owing to specialized phenotypes like unique metabolic chemistries or high capacity for pathway engineering.[1] The lack of broad gene editing tools, however, impedes the agile development of new potential hosts for producing high-value molecules.[2] To date, efforts to elucidate and engineer complex phenotypes have required excessive trial and error using slow, inefficient methods of genetic disruption. Examples include the engineering of cyanobacteria to produce high-value chemicals from CO_2,[3] and the humanization of glycoproteins in *Komagataella phaffii.*[4]

CRISPR-Cas9 has emerged as the technique of choice for both functional genetics and genome engineering in microbial hosts, enabling parallel, targeted genomic disruptions without the use of selection markers.[5] Cas9 is often expressed using common designs for genomic vectors, but single guide RNAs (sgRNAs) require nuclear localization and 5'−3' trimming. These features are characteristic of RNA polymerase-III (Pol-III) mediated transcription. Model organisms boast reliable, well-characterized RNA Pol-III promoters, like the U6 promoter in mammalian cells[6] and the *snr52* promoter in *S. cerevisiae*[7]. The small size and modularity of RNA Pol-III cassettes allow parallel cloning for sgRNAs to enable multiplexed engineering[8–10] and genome-wide phenotypic screens. RNA Pol-III promoters, however, are often intragenic and unique to the host's genome with few conserved motifs,[11] making transfer of designs and strategies for expression sgRNAs difficult among species.[12]

Microorganisms' small genomes are highly amenable to sequencing-based analysis.[13] In model microorganisms, sequencing paired with robust genetic manipulation techniques has enabled nearly complete functional characterization and annotation of genes.[14–16] Here, we present a simple strategy for development of Cas9-mediated gene editing in microorganisms that uses sequencing for the identification of RNA Pol-III expression cassettes. We demonstrated this strategy in *K. phaffii* (*Pichia pastoris*), a non-model yeast of interest as a cellular factory for biopharmaceuticals and vaccines.[17] In this host, engineering complex phenotypes like secretion and post-translational modification would benefit from simplified genetic tools and functional genetics.[18] To date, editing with CRISPR-Cas9 in *K. phaffii* has relied on RNA Pol-II based expression of sgRNAs, which increases the difficulty of multiplexed and batched cloning and expression of sgRNAs.[19] RNA Pol-III activity in *trans* has been observed with substantial variance in efficiency and only with substantial copies of the encoding plasmid.[20] Using small RNA sequencing, we identified native promoters for RNA Pol-III and tested candidates with three sgRNAs to knockout a gene with a clear phenotypic effect. We found that precise fusion of sgRNA to the 3' end of several native tRNAs provided efficiencies in gene knockout of up to 95%. These efficient, orthogonal

tRNA-sgRNA cassettes allowed multiplexed integration of a three gene pathway to remodel and homogenize glycoforms appended on expressed proteins.

Results and Discussion

Current methods for editing genes in *K. phaffii* with CRISPR-Cas9 rely on sgRNAs expressed using an RNA Pol-II cassette, with self-cleaving ribozyme sequences appended to the 5' and 3' ends.[19] This cassette produces a functional sgRNA, but homology requirements of the 5' hammerhead ribozyme extend the variable region for a sgRNA from 20 bp to 69 bp.[21] These long variable regions increase the cost of large-scale synthesis of sgRNAs (~10^5 for genome wide screens), and preclude the expression of multiple sgRNAs on a single cassette for multiplexed editing. Several strategies exist from other organisms that could circumvent the lack of a consensus promoter if deployed in *K. phaffii*. A recently reported T7 bacterial polymerase cassette performs effectively across several yeasts, but requires expression of an additional polymerase and modification of sgRNAs to achieve robust 5' trimming.[22] Another simple strategy uses native RNA Pol-III transcripts by fusing a sgRNA to a native tRNA. Fusion to native tRNA is effective and sequence independent, but requires accurate annotation of the host's endogenous tRNA, often obtained through biochemical knowledge of RNA processing.[23] Among organisms, tRNAs are highly conserved in structure, and sequences are identified by homology. Exact annotation of tRNAs and other small RNAs requires, however, thorough biochemical knowledge of RNA processing. This strategy has been commonly used in plants,[24] but in few microorganisms.[25] Here, we instead obtained exact RNA annotation from transcriptomic and genomic data, which can be easily collected in microbial hosts like *K. phaffii*.[26]

Sequencing and characterization of native RNA Pol-III transcripts

We assayed the RNA Pol-III transcriptome in *K. phaffii* using small RNA sequencing (Figure 1A). We cultivated wild type *K. phaffii* NRRL Y-11430 in common laboratory conditions, collected total RNA, and sequenced RNA less than ~250 bp long, resulting in a dataset of 577 small RNAs. Most small RNAs were <100 bp, with an average length of 87±61 bp (Figure 1B). We performed a BLAST search against the RNA Central Small RNA dataset to classify RNA types including transfer RNAs (tRNAs), small nucleolar RNAs (snoRNAs), and ribosomal RNAs (rRNAs) (Figure 1C). We detected 5 of the 6 conserved small nuclear RNAs (snRNAs),[27] and both the RNAse for mitochondrial RNA processing (mrpRNA) and the signal recognition particle RNA (srpRNA). We observed a broad range of small RNA expression over five orders of magnitude (Figure 1D), with snoRNAs and tRNAs comprising the most highly expressed genes (Figure 1E). As expected, few small RNAs exhibit sequence homology to common yeasts (Figure 1F), further motivating an unbiased *de novo* search for a platform for expression of sgRNA.

Next, we hypothesized that we could identify the 3' end of an active RNA Pol-III promoter region by identifying the 5' edge of a RNA Pol-III transcribed small RNA. We created a metric we call "edge resolution" to systematically identify individual small RNAs that were clearly defined by the sequencing reads (Methods). High edge resolution indicated sequencing reads that created a clean RNA footprint with clearly defined 5' and 3' edges

(Figure 1H). Both snoRNAs and tRNAs had high edge resolution, while rRNAs, which are likely fragments of larger molecules, had lower edge resolution (Figure 1G). After sorting for highly resolved small RNAs, we inspected the 5' ends of mapped sequencing reads to identify the 3' ends of potential RNA Pol-III promoters. This annotation was used to inform the construction of cassettes for expression of guide RNAs.

A promoter-tRNA-sgRNA fusion cassette yields highly efficient knockout of genes

We constructed four cassettes using two tRNA promoters, one snoRNA promoter, and the srpRNA promoter for expressing single guide RNAs (Figure 2A). Promoters were placed directly in front of a reporter sgRNA and poly T termination sequence. Because the 5' edge of the corresponding small RNA was used to predict the transcription start site, these promoters were chosen based on RNA class diversity, and high 5' edge resolution. RNA Pol-III promoters are often intragenic, meaning the small RNA sequence is essential to promoter function, so we created an additional cassette with the first tRNA promoter that included the native tRNA sequence directly fused to the sgRNA. In this case, the 3' edge of the sequencing footprint was used to place the sgRNA instead of the 5' edge. All five RNA Pol-III cassettes C1-C5 (Table 1) were tested in an autonomously replicating plasmid in *K. phaffii*. We expressed *Streptococcus pyogenes* Cas9 using the *K. phaffii eno1* promoter after determining that promoter strength had little effect on Cas9 activity (Figure S1). We used sgRNAs targeting *gut1*, a glycerol kinase that results in limited growth on glycerol media when knocked out (Figure S2). Cells were transformed with circular plasmid and plated on a glycerol agar plate to screen for growth deficiency (Figure S2).

Cassette C2 (promoter-tRNA-sgRNA) enabled knockout activity with up to 93% efficiency, with all other cassettes exhibiting essentially no observed knockouts. Interestingly, C1 exhibited no knockout activity despite containing the same RNA Pol-III promoter as C2, but without the tRNA sequence. This result suggests that tRNA promoter activity indeed depends on the presence of the tRNA gene. A similar strategy for expression using a promoter-tRNA-sgRNA cassette has been demonstrated in organisms spanning kingdoms.[23,25,28–30] Processing of the 5' end of the sgRNA is only dependent on proper identification of the 3' end of the tRNA, as the protein complex RNase Z recognizes folded tRNA structures, releasing any appended sequence indiscriminately.[31] Since tRNA processing occurs in the nucleus, efficiency may be improved over the ribozyme cassette by confining sgRNA to the nucleus. tRNA fusion provided robust gene knockout across several sgRNAs, so efforts to identify an extragenic RNA Pol-III promoter were not pursued further.

Identification of tRNA edge location determines sgRNA efficiency

After establishing a reliable design principle for expression of sgRNAs, we explored the generality of using small RNA sequencing to identify RNase Z cut sites, and subsequently sgRNA fusions. We visualized the expression and edge resolution of all small RNAs in our dataset (Figure 3A). As expected from our analyses, tRNA1 from cassette C2 is both highly expressed and has well-defined 5' and 3' edges. We then selected tRNA3—a tRNA that is highly expressed but has poorly resolved edges from sequencing (reads shown in Figure 1H). Upon inspection of the 3' end, it was unclear where the tRNA is trimmed by RNase Z (Figure 3B). We created 13 variants of cassette C6, where the sgRNA is fused at different

locations at the 3’ end of tRNA3. The knockout efficiency was highest at a specific 3’ location, presumably the RNase Z cut site (Figure 3C). When the sgRNA is fused too distant from the tRNA structure, efficiency likely decreases due to excess sequence attached to the 5’ end of the sgRNA. When fused too close, RNase Z may cleave the sgRNA, reducing targeting specificity. The low efficiency of sgRNAs with only one or two appended base pairs may be specific to the sequences of this tRNA-sgRNA fusion, and is mitigated when the sgRNA is properly cleaved. We therefore hypothesized that accurate tRNA annotation is necessary to create an efficient sgRNA expression cassette from any native tRNA.

To test this idea, we selected four additional tRNAs, tRNA4 - tRNA7, that exhibited a wide range of both expression level and edge resolution (Figure 3A). By manual inspection of sequencing read edges, we intuited locations of RNase Z cut sites and generated promoter-tRNA-sgRNA cassettes based on our best interpretation of sequencing-derived annotations. All four additional cassettes (C7-C10) yielded highly efficient gene knockout (Figure 3D). It is unclear why sgRNA B is not functional on cassette 7; this is currently under investigation. Interestingly, the native level of expression for each tRNA promoter did not correlate with the efficiency of gene knockout, perhaps due to high plasmid copy number and lack of *cis* regulation. Our success in creating expression cassettes for sgRNAs from multiple promoter-tRNA pairs suggests that small RNA sequencing can provide a useful small transcriptome annotation for accurate identification of tRNA 3’ edges. Edge resolution can be used to quickly identify the 3’ edges of tRNAs for assembly of sgRNA cassettes. Fusions of tRNA and sgRNA have been demonstrated in a range of organisms, and given the ease of sequencing microorganisms, we posit this algorithm can also facilitate developing elements to express sgRNA in previously uncharacterized hosts.

Efficient, multiplexed engineering of glycosylation

To demonstrate the efficiency of our platform for sgRNA expression, we next sought to engineer a strain with modified glycosylation with minimal rounds of editing. Uniform humanized glycosylation has been demonstrated to reduce antigenicity and increase efficacy of biologics manufactured in non-human hosts.[32] Prior engineering of glycosylation in *K. phaffii* has relied on repeated recycling of a limited number of auxotrophic markers – a time and labor intensive process.[33] For example, knockout of the native mannosyl transferase *och1* using traditional selection required screening of at least 1,000 transformants.[34] We devised a strategy for pathway engineering involving the knockout of the native mannosyl transferase *och1* and integration of three heterologous genes involved in glycosylation (*Caenorhabditis elegans* mannosidase (*mns1*), *Kluyveromyces lactis* UDP-GlcNAc transporter (*mnn2*), and human GlcNAc transferase (*gnt1*)) in a multiplexed manner (Figure 4A).

First, we targeted and knocked out *och1*. We observed a highly efficient knockout despite the expected morphological differences and slow growth (Figure 4B), comparable to other reported knockouts of *och1* using Cas9.[19] Next, we constructed an RNA Pol-III cassette to express three sgRNAs simultaneously for the integration of *mns1*, *mnn2*, and *gnt1*. By precise annotation of not only the 3’ edge of tRNAs but also the 5’ edge, sgRNAs can be multiplexed on a single RNA Pol-III transcript, as has been recently reported in *S.*

cerevisiae.[35] We identified more than three tRNA promoter cassettes by sequencing, so we could construct multiplexed sgRNAs without repeating tRNA sequences (Figure 4C); this design reduces cassette homology and improves compatibility with cloning techniques like Golden Gate Assembly. We transformed the multiplexed plasmid along with linear DNA fragments of each heterologous gene into wild type cells, Δ*och1* cells, and Δ*ku70* cells, a knockout previously reported to reduce the frequency of non-homologous end joining (Figure S3).[36] Complete multiplexing was achieved in all three strains with efficiency up to 40% (Figure 4D). Amongst all loci screened, nearly 50% of possible edits were successful, with only small frequencies of indels, off-target integrations, or mixed colonies (Figure 4E). Efficiency of integration declined with each gene, and may reflect effects due to the order of expression of sgRNA (Table S6).[35,37] Interestingly, Δ*ku70* and wild type cells behaved similarly with respect to integration efficiency. We therefore elected to proceed with a multiplexed integrated Δ*och1* strain with an intact *ku70* gene.

We then expressed the Kringle 3 domain of human plasminogen (K3) to test the glycoengineered strain because it exhibits a single, highly accessible motif for glycosylation at position 41 (N-R-T, Table S5).[34] We transformed K3 into a wild type strain, a Δ*och1* strain, and our glycoengineered strain, cultivated clones in batch culture, and evaluated the glycosylation on secreted K3 by mass spectrometry. The glycoengineered strain produced K3 with uniform glycosylation, markedly different from the yeast's native, heterogeneous hypermannosylation (Figure 4F). Analysis of K3 by LC-MS before and after treatment with PNGase revealed a humanized five-mannose glycan structure (Figure 4G). These data show the successful engineering of a complex phenotype of interest in a specialized microorganism in only two steps. Equivalent modifications with traditional tools would require four transformations, using either four orthogonal or recycled auxotrophic markers—a time and resource intensive endeavor that would leave genomic scars or alter amino acid metabolism.

In conclusion, we have reported a simplified, multiplexed cassette for expression of sgRNAs in *K. phaffii.* In addition, we demonstrated a method to establish RNA Pol-III mediated expression of sgRNA, which we believe is broadly extensible to other microorganisms. Finally, we demonstrated our CRISPR platform to rapidly engineer a complex host phenotype, including concurrent knockout and knock-in of genes. To our knowledge, these results also show the first demonstration of multiplexed editing by heterologous gene insertion in *K. phaffii*.[38] These techniques can now be quickly applied to a broad range of applications in *Pichia pastoris* or other microorganisms, including introduction of mammalian chaperones to enhance folding of complex molecules, generation of protease-deficient strains to improve yields of full-length product,[39] reduction and redirection of vacuolar and endoplasmic reticulum-associated protein degradation pathways,[40] enhancement of lipid synthesis and vesicular machinery,[41,42] and introduction of novel metabolic pathways.[43] Simplified tools for genomic editing that are compatible with multiplexing and parallelization will be critical to the widespread adoption of alternative hosts for manufacturing, eventually leading to engineered fit-to-purpose strains for unique classes of biologic medicines.[17] Expansion of tools to new organisms with unique phenotypes will enable manufacturers to continue to expand repertoires of complex products while benefiting from the simplicity of microbial hosts in cultivation.[44]

Methods

Small RNA sequencing and analysis

Wild type *Komagataella phaffii* strain NRRL Y-11430 was cultivated in 5 mL of rich defined medium (RDM)[45] with 4% glycerol by volume. Five identical cultures independently cultivated were used as biological replicates. Cultures were inoculated at OD600 = 0.1, and grown at 30°C for 24 hours until OD600 ≈ 16. Total RNA was extracted using the Qiagen miRNA extraction kit. RNA was converted to cDNA libraries using the New England Biolabs Small RNA Kit and sequenced on an Illumina NextSeq. Reads were aligned to the *K. phaffii* genome[46] using Salmon.[47] Small RNAs were defined as continuous regions of >50 read depth. Expression of each small RNA was calculated using transcripts per million (tpm), normalized by sample.

To quantify edge resolution of each small RNA, we defined each edge of the RNA as the 20 bp upstream and downstream of the annotated boundary. Over this region, we calculated the change in read depth at each base pair. We defined edge resolution as the ratio of the maximum change in read depth to the maximum read depth.

$$Edge\ resolution = \frac{\max(\Delta read\ depth)}{\max(read\ depth)}$$

The edge resolution score was then calculated as the mean of the 5' and 3' edge resolution for each small RNA. An edge resolution of 1 indicates that all paired end reads start and end at the same 5' and 3' base pairs.

Plasmid construction

Complete plasmid and primer sequences are available, see Supplemental Information. PCR was performed using Q5 Hotstart High Fidelity Master Mix (NEB) according to the manufacturer's instructions. Fragment assembly was performed using Gibson Assembly Master Mix (NEB) according to the manufacturer's instructions. Plasmids were stored and propagated in Top10 *Escherichia coli* (Invitrogen). Primers were synthesized by Integrated DNA Technologies (IDT) or Genewiz Inc. Double stranded DNA fragment synthesis was performed by IDT.

Spacer sequences were generated randomly and inserted to separate plasmid components and assist in Gibson assembly. The *K. phaffii* replication origin (PARS) sequence was synthesized separately. Human codon optimized *S. pyogenes* Cas9 was obtained from Addgene plasmid #43802. *K. phaffii* RNA Pol-II promoters and terminators P_{tef1}, P_{eno1}, and TT_{tef1} were PCR amplified from the genome. RNA Pol-II promoters were defined as the 1000 bp upstream of the gene coding sequence, or shorter in the presence of an adjacent coding sequence. RNA Pol-III promoters were defined as 300 bp upstream of a small RNA and synthesized. For tRNA fusions, tRNAs were fused to the 3' end of the 300 bp promoter. Ribozyme sequences were obtained from Weninger et al. and synthesized.[19] Guide RNAs were designed targeting the *gut1* gene coding sequence using the ATUM gRNA Designer tool (www.atum.bio). A list of sgRNAs used in the study is available, see Table S1.

Gut1 knockout screening

Competent cells were prepared as described elsewhere[48] and transformed with 100 ng of circular plasmid DNA. Cells were allowed to recover for 3h at 30°C without shaking before plating on YPD agar (BD Difco) plates supplemented with 800 μg/mL G418 Geneticin (Thermo Fisher Scientific). Single colonies were used to inoculate 200 μL YPD cultures cultivated overnight with shaking at 1000 rpm in 96 well plates. Cells from liquid culture were then stamped onto minimal glycerol media (6.7 g/L BD Yeast Nitrogen Base w/o Amino Acids, 10 mg/L L-histidine, 20 mg/L L-methionine, 20 mg/L L-tryptophan, 10 mL/L glycerol) plates without selection. Plates were incubated for 2 days at 30°C before examination for deficient growth (Figure S2). Three biological replicates of each tested construct were generated via three independent transformations. Sixteen colonies were picked at random from each plate. Knockout efficiency was calculated as the average percentage of complete knockout across 3 biological replicates. For a subset of transformations, DNA was extracted as previously described[49] and purified using the MagJET gDNA Kit (Thermo Fisher Scientific). The *gut1* locus was PCR amplified and Sanger sequenced to corroborate the glycerol growth assay.

Multiplexed engineering

Guide RNAs targeting *och1* and *ku70* were cloned into an RNA Pol-III cassette as described above. For *ku70* knockout screening, colonies were screened using genomic DNA and extraction as described above. Cells transformed with plasmids targeting *och1* were grown on solid media for two weeks prior to examination for Δ*och1* colonies. Knockout efficiency was calculated by counting the number of morphologically different colonies on plates in three biological replicates. Representative Δ*och1* colonies identified by altered morphology were confirmed by Sanger sequencing.

Glycosylation genes *mns1*, *mnn2*, and *gnt1* were targeted to intragenic regions of the genome adjacent to genes GQ67_04576, PFK1, and ROX1, respectively using 500 bp flanking sequences for homologous recombination. Genes were codon optimized for *K. phaffii* (*Pichia pastoris*), synthesized, and subcloned for storage. All genes were placed between the *K. phaffii* AOX1 promoter, and the *S. cerevisiae* CYC1 terminator. Linear inserts were amplified using PCR and column purified. The multiplexed guide RNA cassette was synthesized in three sections separated by tRNA to avoid self-homology of the structural component of sgRNAs. Fragments were assembled into a multiplexed RNA Pol-III cassette using Golden Gate assembly (NEB). The multiplexed RNA Pol-III cassette was then cloned into the previously constructed Cas9 expression vector using Gibson assembly.

Wild type, Δ*och1*, and Δ*ku70* cells were transformed in duplicate with a mixture of 100 ng of plasmid pND413 and 1.5 μg of each linear insert encoding each glycosylation gene. Sixteen colonies were randomly selected for genomic DNA extraction. All three targeted loci were amplified from each genomic DNA sample and Sanger sequencing was performed to confirm integration of linear inserts at the correct loci. Off-target integrations were identified by amplification using heterologous gene specific primers in samples with an otherwise unedited target locus.

Glycosylation testing

Wild type, Δ*och1*, and Δ*och1:mnn2:mns1:gnt1* cells were transformed[48] with a vector for multicopy expression of human K3 plasminogen peptide, appended with a C-terminal 6-His tag (Table S5). Strains were grown in 24-well deep well plates (25°C, 600 rpm) using glycerol-containing media (BMGY-Buffered Glycerol Complex Medium, Teknova) supplemented to 4% (v/v) glycerol. After 24h of biomass accumulation, cells were pelleted and resuspended in BMMY (Buffered Methanol Complex Medium, Teknova) containing 3% (v/v) methanol. After 24 hours of production, supernatant was harvested for analysis by 16% Tricine SDS-PAGE (Thermo Fisher Scientific). K3 peptide was purified using HisTrap columns (GE) and deglycosylated with PNGase F (NEB). To verify glycosylation, 1 μL of sample was mixed with 1 μL of sinapinic acid matrix and run on a Bruker Microflex MALDI TOF in positive, linear mode. Data was collected for intact masses from 5–30 kD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We thank D.E. Root for fruitful discussions. We thank the Koch Institute Swanson Biotechnology Center for technical support, specifically the Bioinformatics & Computing and Genomics core facilities.

Funding sources

This work was supported by the Defense Advanced Research Projects Agency (DARPA) and SPAWAR Systems Center Pacific (SSC Pacific) (contract no. N66001-13-C-4025), and by the AltHost Consortium. The content is solely the responsibility of the authors and does not necessarily represent the official views of DARPA, SSC Pacific, or the AltHost Consortium.

References

(1). Matthews CB, Wright C, Kuo A, Colant N, Westoby M, and Love JC (2017) Reexamining opportunities for therapeutic protein production in eukaryotic microorganisms. Biotechnol. Bioeng 114, 2432–2444. [PubMed: 28688187]

(2). Shapiro RS, Chavez A, and Collins JJ (2018) CRISPR-based genomic tools for the manipulation of genetically intractable microorganisms. Nat. Rev. Microbiol 16, 333–339. [PubMed: 29599458]

(3). Johnson TJ, Gibbons JL, Gu L, Zhou R, and Gibbons WR (2016) Molecular genetic improvements of cyanobacteria to enhance the industrial potential of the microbe: A review. Biotechnol. Prog 32, 1357–1371. [PubMed: 27604481]

(4). Hamilton SR, Bobrowicz P, Bobrowicz B, Davidson RC, Li H, Mitchell T, Nett JH, Rausch S, Stadheim TA, Wischnewski H, Wildt S, and Gerngross TU (2003) Production of complex human glycoproteins in yeast. Science 301, 1244–6. [PubMed: 12947202]

(5). Raschmanová H, Weninger A, Glieder A, Kovar K, and Vogl T (2018) Implementing CRISPR-Cas technologies in conventional and non-conventional yeasts: Current state and future prospects. Biotechnol. Adv

(6). Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, and Zhang F (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol 31, 827–832. [PubMed: 23873081]

(7). DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, and Church GM (2013) Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res 41, 4336–4343. [PubMed: 23460208]

(8). Si T, Chao R, Min Y, Wu Y, Ren W, and Zhao H (2017) Automated multiplex genome-scale engineering in yeast. Nat. Commun 8, 15187. [PubMed: 28469255]

(9). Jakočiūnas T, Bonde I, Herrgård M, Harrison SJ, Kristensen M, Pedersen LE, Jensen MK, and Keasling JD (2015) Multiplex metabolic pathway engineering using CRISPR/Cas9 in Saccharomyces cerevisiae. Metab. Eng 28, 213–222. [PubMed: 25638686]

(10). Horwitz AA, Walter JM, Schubert MG, Kung SH, Hawkins K, Platt DM, Hernday AD, Mahatdejkul-Meadows T, Szeto W, Chandran SS, and Newman JD (2015) Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. Cell Syst.

(11). Geiduschek EP, and Tocchini-Valentini GP (1988) TRANSCRIPTION BY RNA POLYMERASE III. Annu. Rev. Biochem 873–914. [PubMed: 3052292]

(12). Zheng X, Zheng P, Zhang K, Cairns TC, Meyer V, Sun J, and Ma Y (2018) 5S rRNA Promoter for Guide RNA Expression Enabled Highly Efficient CRISPR/Cas9 Genome Editing in Aspergillus niger. ACS Synth. Biol acssynbio.7b00456.

(13). Ziemert N, Alanjary M, and Weber T (2016) The evolution of genome mining in microbes – a review. Nat. Prod. Rep 33, 988–1005. [PubMed: 27272205]

(14). Kim D-U, Hayles J, Kim D, Wood V, Park H-O, Won M, Yoo H-S, Duhig T, Nam M, Palmer G, Han S, Jeffery L, Baek S-T, Lee H, Shim YS, Lee M, Kim L, Heo K-S, Noh EJ, Lee A-R, Jang Y-J, Chung K-S, Choi S-J, Park J-Y, Park Y, Kim HM, Park S-K, Park H-J, Kang E-J, Kim HB, Kang H-S, Park H-M, Kim K, Song K, Song K. Bin, Nurse P, and Hoe K-L (2010) Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol 28, 617–623. [PubMed: 20473289]

(15). Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, and Bartel DP (2009) RNAi in Budding Yeast. Science (80-.) 326, 544–550.

(16). Tong AHY (2001) Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants. Science (80-.) 294, 2364–2368.

(17). Love KR, Dalvie NC, and Love JC (2018) The yeast stands alone: the future of protein biologic production. Curr. Opin. Biotechnol 53, 50–58. [PubMed: 29277062]

(18). Maccani A, Landes N, Stadlmayr G, Maresch D, Leitner C, Maurer M, Gasser B, Ernst W, Kunert R, and Mattanovich D (2014) Pichia pastoris secretes recombinant proteins less efficiently than Chinese hamster ovary cells but allows higher space-time yields for less complex proteins. Biotechnol. J 9, 526–537. [PubMed: 24390926]

(19). Weninger A, Hatzl AM, Schmid C, Vogl T, and Glieder A (2016) Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast Pichia pastoris. J. Biotechnol 235, 139–149. [PubMed: 27015975]

(20). Gu Y, Gao J, Cao M, Dong C, Lian J, Huang L, Cai J, and Xu Z (2019) Construction of a series of episomal plasmids and their application in the development of an efficient CRISPR/Cas9 system in Pichia pastoris. World J. Microbiol. Biotechnol 35, 79. [PubMed: 31134410]

(21). Gao Y, and Zhao Y (2014) Self-processing of ribozyme-flanked RNAs into guide RNAs *in vitro* and *in vivo* for CRISPR-mediated genome editing. J. Integr. Plant Biol 56, 343–349. [PubMed: 24373158]

(22). Morse NJ, Wagner JM, Reed KB, Gopal MR, Lauffer LH, and Alper HS (2018) T7 Polymerase Expression of Guide RNAs *in vivo* Allows Exportable CRISPR-Cas9 Editing in Multiple Yeast Hosts. ACS Synth. Biol 7, 1075–1084. [PubMed: 29565571]

(23). Port F, and Bullock SL (2016) Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. Nat. Methods 13, 852–854. [PubMed: 27595403]

(24). Xie K, Minkenberg B, and Yang Y (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc. Natl. Acad. Sci. U. S. A 112, 3570–5. [PubMed: 25733849]

(25). Schwartz CM, Hussain MS, Blenner M, and Wheeldon I (2016) Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR-Cas9-Mediated Genome Editing in Yarrowia lipolytica. ACS Synth. Biol 5, 356–359. [PubMed: 26714206]

(26). Vervoort Y, Linares AG, Roncoroni M, Liu C, Steensels J, and Verstrepen KJ (2017) High-throughput system-wide engineering and screening for microbial biotechnology. Curr. Opin. Biotechnol 46, 120–125. [PubMed: 28346890]

(27). Wise JA, Tollervey D, Maloney D, Swerdlow H, Dunn EJ, and Guthrie C (1983) Yeast contains small nuclear RNAs encoded by single copy genes. Cell 35, 743–751. [PubMed: 6197183]

(28). Nødvig CS, Hoof JB, Kogle ME, Jarczynska ZD, Lehmbeck J, Klitgaard DK, and Mortensen UH (2018) Efficient oligo nucleotide mediated CRISPR-Cas9 gene editing in Aspergilli. Fungal Genet. Biol 115, 78–89. [PubMed: 29325827]

(29). Shiraki T, and Kawakami K (2018) A tRNA-based multiplex sgRNA expression system in zebrafish and its application to generation of transgenic albino fish. Sci. Rep 8, 13366. [PubMed: 30190522]

(30). Numamoto M, Maekawa H, and Kaneko Y (2017) Efficient genome editing by CRISPR/Cas9 with a tRNA-sgRNA fusion in the methylotrophic yeast Ogataea polymorpha. J. Biosci. Bioeng 124, 487–492. [PubMed: 28666889]

(31). Xie K, Minkenberg B, and Yang Y (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc. Natl. Acad. Sci 112, 3570–3575. [PubMed: 25733849]

(32). Zhang N, Liu L, Dan Dumitru C, Cummings NRH, Cukan M, Jiang Y, Li Y, Li F, Mitchell T, Mallem MR, Ou Y, Patel RN, Vo K, Wang H, Burnina I, Choi BK, Huber H, Stadheim TA, and Zha D (2011) Glycoengineered Pichia produced anti-HER2 is comparable to trastuzumab in preclinical study. MAbs 3, 289–298. [PubMed: 21487242]

(33). Hamilton SR, and Gerngross TU (2007, 10 1) Glycosylation engineering in yeast: the advent of fully humanized yeast. Curr. Opin. Biotechnol Elsevier Current Trends.

(34). Choi B-K, Bobrowicz P, Davidson RC, Hamilton SR, Kung DH, Li H, Miele RG, Nett JH, Wildt S, and Gerngross TU (2003) Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast Pichia pastoris. Proc. Natl. Acad. Sci. U. S. A 100, 5022–7. [PubMed: 12702754]

(35). Zhang Y, Wang J, Wang Z, Zhang Y, Shi S, Nielsen J, and Liu Z (2019) A gRNA-tRNA array for CRISPR-Cas9 based rapid multiplexed genome editing in Saccharomyces cerevisiae. Nat. Commun 10, 1053. [PubMed: 30837474]

(36). Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, and Glieder A (2012) Deletion of the Pichia pastoris KU70 Homologue Facilitates Platform Strain Generation for Gene Expression and Synthetic Biology PLoS One (Riezman H, Ed.) 7, e39720. [PubMed: 22768112]

(37). Deaner M, and Alper HS (2019) Enhanced scale and scope of genome engineering and regulation using CRISPR/Cas in Saccharomyces cerevisiae. FEMS Yeast Res. 19.

(38). Liu Q, Shi X, Song L, Liu H, Zhou X, Wang Q, Zhang Y, and Cai M (2019) CRISPR–Cas9-mediated genomic multiloci integration in Pichia pastoris. Microb. Cell Fact 18, 144. [PubMed: 31434578]

(39). Jiang H, Horwitz AA, Wright C, Tai A, Znameroski EA, Tsegaye Y, Warbington H, Bower BS, Alves C, Co C, Jonnalagadda K, Platt D, Walter JM, Natarajan V, Ubersax JA, Cherry JR, and Love JC (2019) Challenging the workhorse: Comparative analysis of eukaryotic micro-organisms for expressing monoclonal antibodies. Biotechnol. Bioeng 116, 1449–1462. [PubMed: 30739333]

(40). Marsalek L, Puxbaum V, Buchetics M, Mattanovich D, and Gasser B (2019) Disruption of vacuolar protein sorting components of the HOPS complex leads to enhanced secretion of recombinant proteins in Pichia pastoris. Microb. Cell Fact 18, 119. [PubMed: 31269943]

(41). Huang M, Wang G, Qin J, Petranovic D, and Nielsen J (2018) Engineering the protein secretory pathway of Saccharomyces cerevisiae enables improved protein production. Proc. Natl. Acad. Sci. U. S. A 115, E11025–E11032. [PubMed: 30397111]

(42). de Ruijter JC, Koskela EV, and Frey AD (2016) Enhancing antibody folding and secretion by tailoring the Saccharomyces cerevisiae endoplasmic reticulum. Microb. Cell Fact 15, 87. [PubMed: 27216259]

(43). Li P, Sun H, Chen Z, Li Y, and Zhu T (2015) Construction of efficient xylose utilizing Pichia pastoris for industrial enzyme production. Microb. Cell Fact 14, 22. [PubMed: 25889970]

(44). Wagner JM, and Alper HS (2016) Synthetic biology and molecular genetics in non-conventional yeasts: Current tools and future advances. Fungal Genet. Biol 89, 126–136. [PubMed: 26701310]

(45). Matthews CB, Kuo A, Love KR, and Love JC (2018) Development of a general defined medium for Pichia pastoris. Biotechnol. Bioeng 115, 103–113. [PubMed: 28865117]

(46). Love KR, Shah KA, Whittaker CA, Wu J, Bartlett MC, Ma D, Leeson RL, Priest M, Borowsky J, Young SK, and Love JC (2016) Comparative genomics and transcriptomics of Pichia pastoris. BMC Genomics 17.

(47). Patro R, Duggal G, Love MI, Irizarry RA, and Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417–419. [PubMed: 28263959]

(48). Lin-Cereghino J, Wong WW, Xiong S, Giang W, Luong LT, Vu J, Johnson SD, and Lin-Cereghino GP (2005) Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. Biotechniques 38, 44–48. [PubMed: 15679083]

(49). Lõoke M, Kristjuhan K, and Kristjuhan A (2011) Extraction of genomic DNA from yeasts for PCR-based applications. Biotechniques 50, 325–328. [PubMed: 21548894]

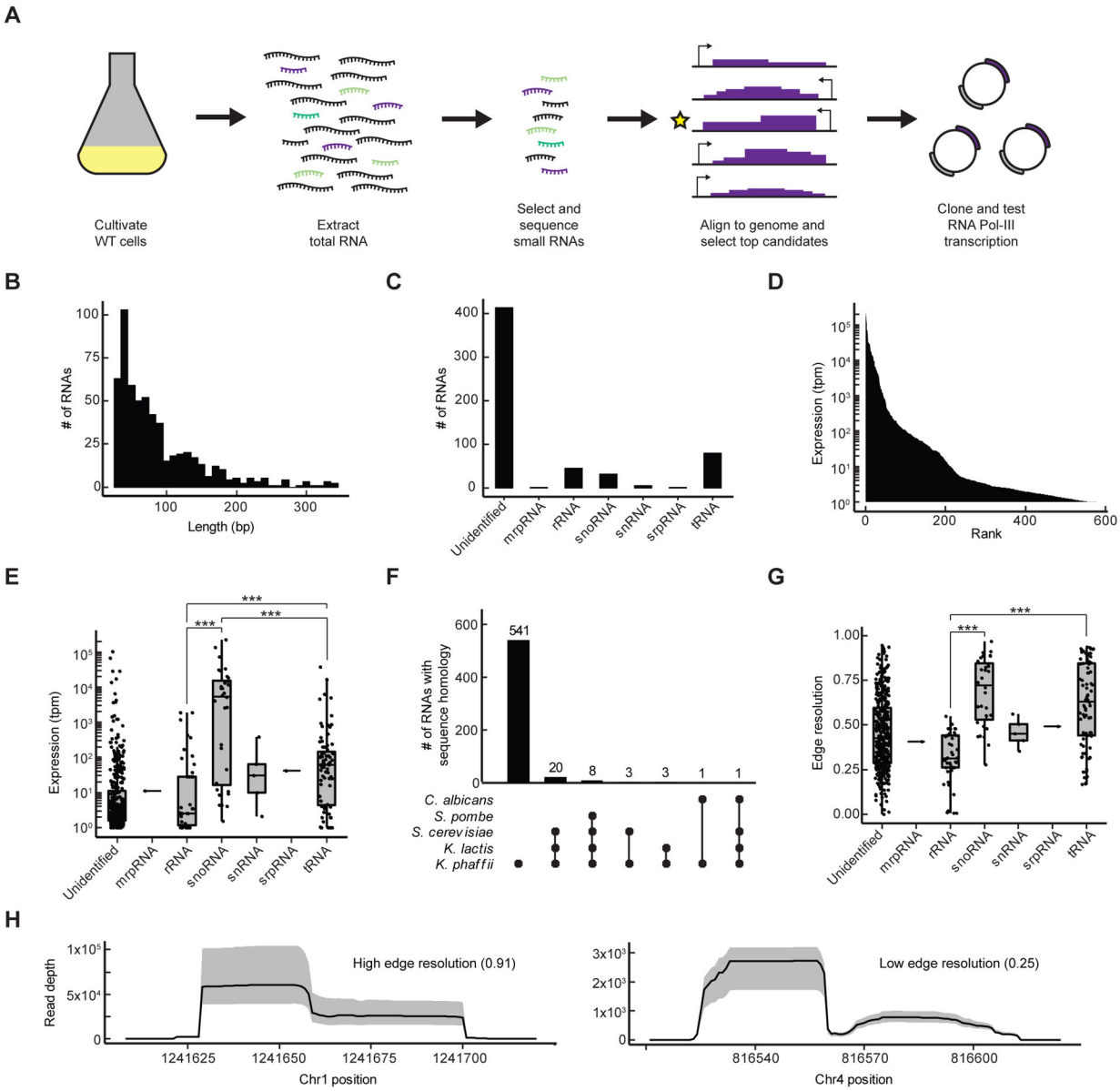


Figure 1:

RNA sequencing for small RNA annotation. a) Pipeline of cultivation and sequencing of wild type (WT) cells to select candidate RNA Pol-III promoters. b) Histogram of captured transcript lengths. Transcripts were filtered to be between 30 and 350 bp. c) Classification of *K. phaffii* transcripts using BLAST against RNA-Central database. d) Histogram of small RNA expression. e) Expression of each class of small RNA. Three asterisks indicate $p<0.001$ by the Mann-Whitney-Wilcoxon test. f) Homology of *K.phaffii* small RNAs to common yeasts. g) Edge resolution of each class of small RNA, defined as the ratio of change in read depth to read depth at the edge of each called transcript. h) Examples of high (tRNA1) and low (tRNA3) edge resolution. Average sequencing read depth (black line) drops steeply at the edges of the tRNA with high resolution, and declines gradually with low

resolution. Gray region represents maximum and minimum read depths across 5 biological replicates.

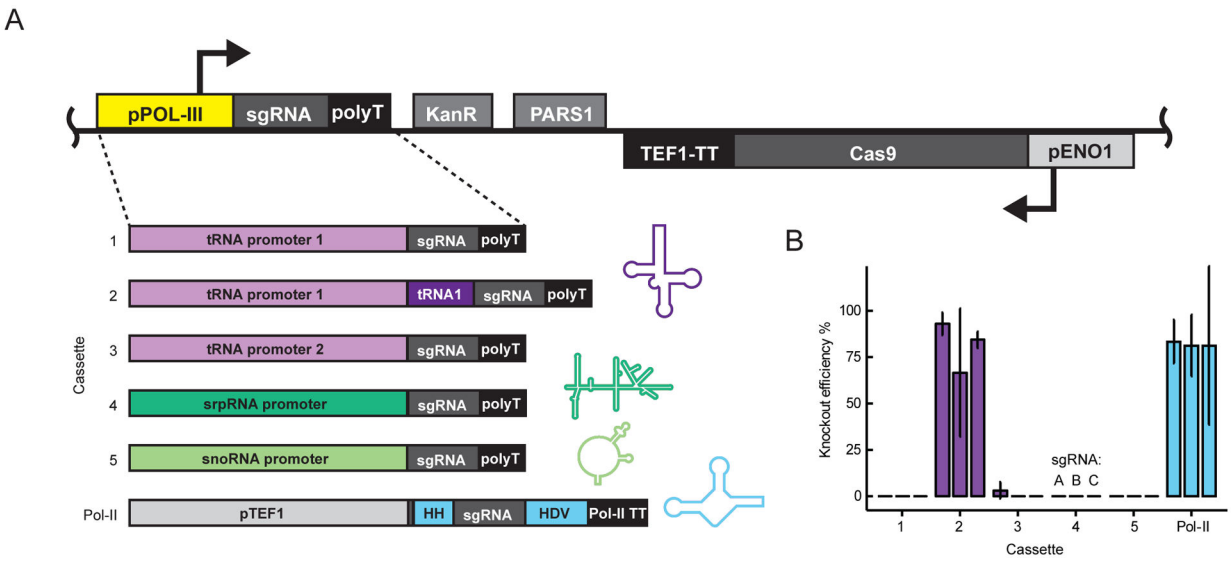


Figure 2:

RNA Pol-III expression of guide RNA. a) Diagram of CRISPR plasmid for testing RNA Pol-III transcription of sgRNA. Expanded regions are five RNA Pol-III sgRNA expression cassettes, compared to a control RNA Pol-II cassette. Self-cleaving ribozyme sequences are shown in light blue. Sequence lengths are not to scale. b) Knockout efficiency of cassettes for three sgRNAs targeting the *gut1* reporter gene.

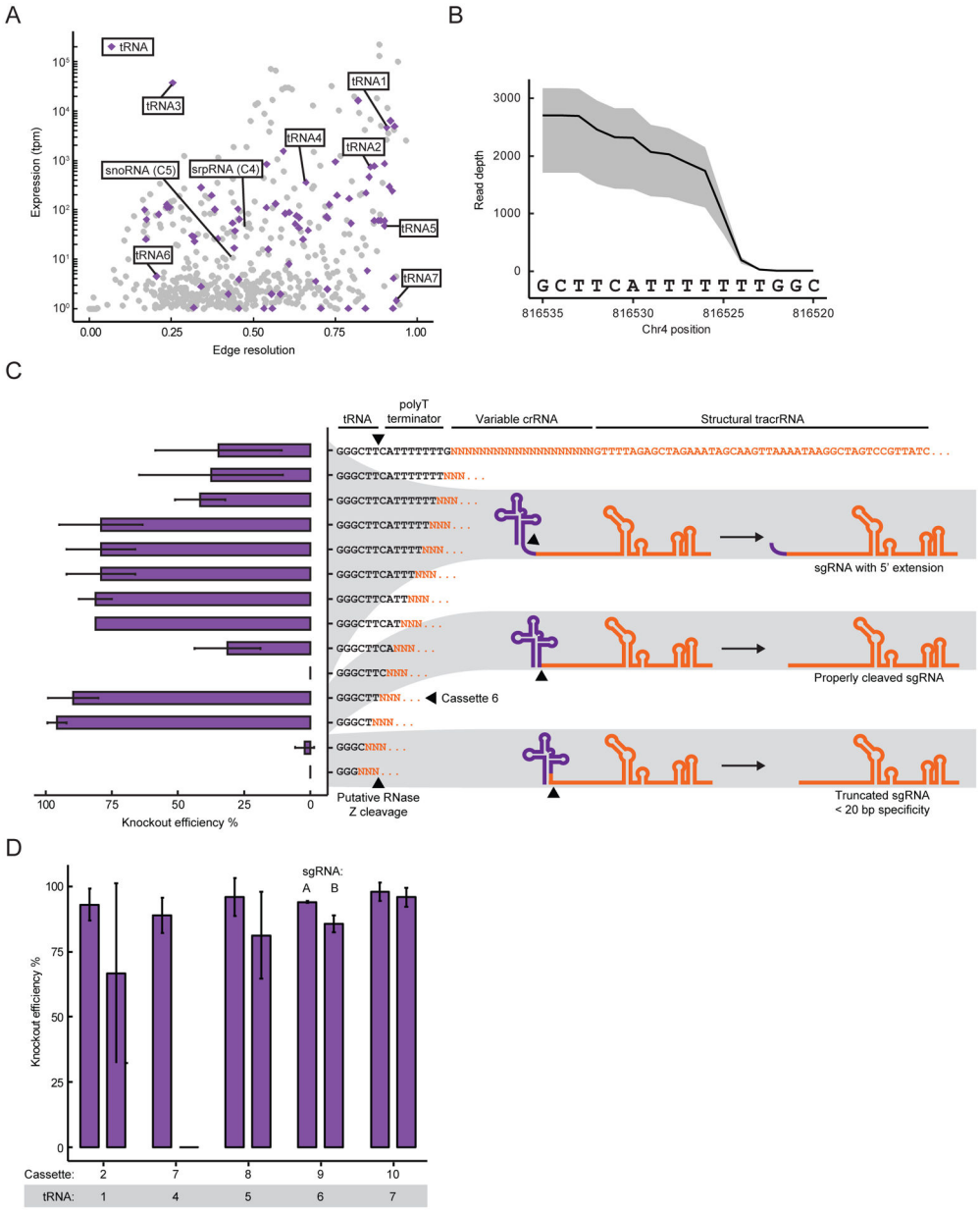


Figure 3:

Location of sgRNA-tRNA fusion impacting knockout efficiency. a) Expression against trimness for all *K. phaffii* small RNAs (tRNAs are shown in purple). RNAs present in all RNA Pol-III cassettes are labeled. b) Sequencing read depth at the 3' end of tRNA3. c) Knockout efficiency of varying fusion sites of tRNA3 in cassette C6. Shown on the right is the fusion location, and hypothesized mechanisms for reduced efficiency. Efficiency was tested using sgRNA A. d) Knockout efficiency of each additional RNA Pol-III expression cassette with two sgRNAs.

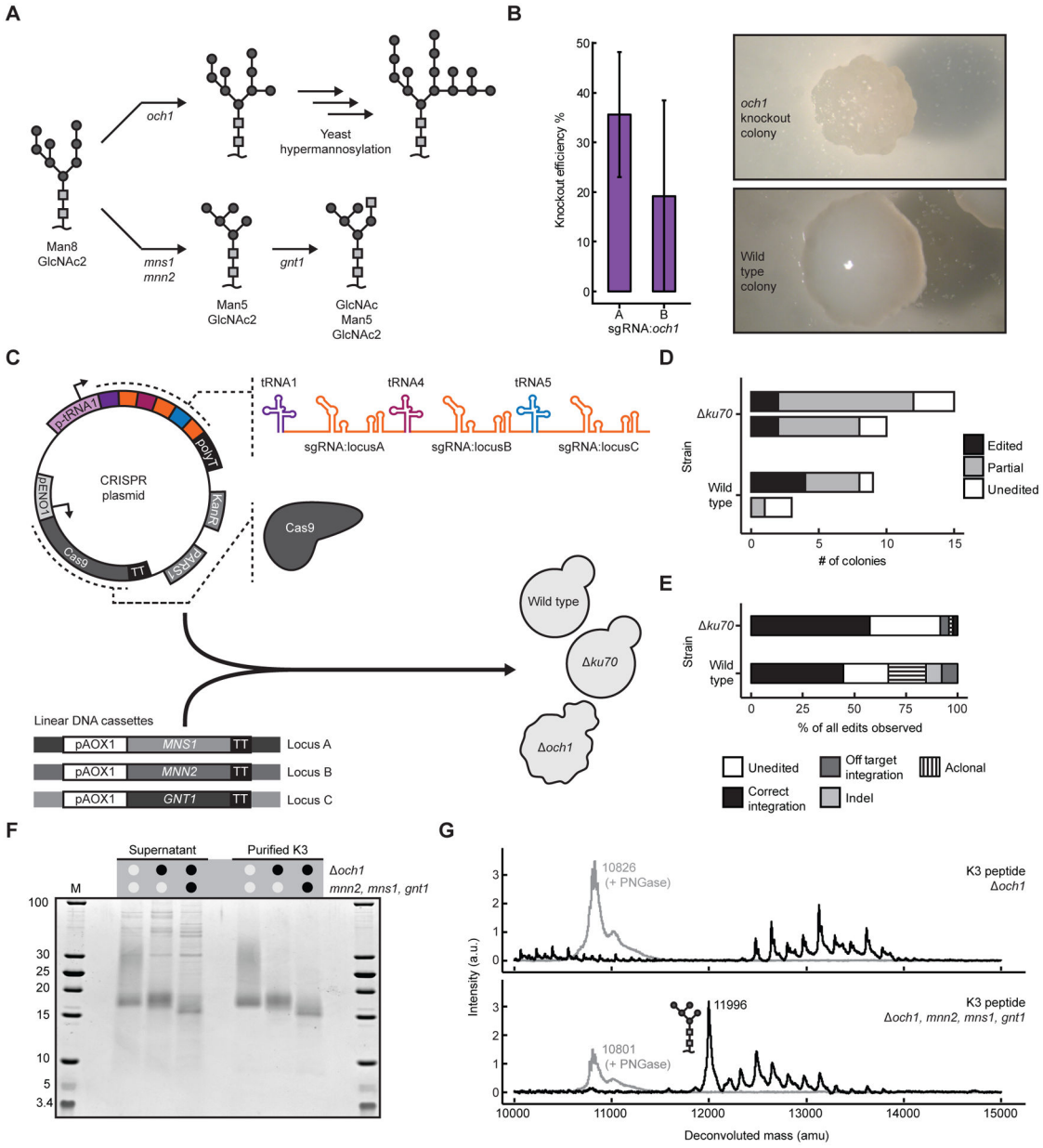


Figure 4:

Editing of the yeast glycosylation pathway using knockout and multiplexed gene insertion. a) Diagram of the native yeast glycosylation pathway (top) and the previously demonstrated engineered pathway to mammalian $GlcNAcMan_5GlcNAc_2$ structures (bottom). b) Knockout efficiency of the native mannosyltransferase *och1*, and morphology of the resulting colonies. c) Schematic of multiplexed genome editing. Three sgRNAs and Cas9 are expressed from a plasmid, co-transformed with three linear integration fragments. d) Multiplexed genome editing efficiency. Edited colonies exhibited all three edits, partial colonies exhibited a non-zero fraction of edits, and unedited colonies exhibited no editing or indel genotypes without insertion. e) Error modes of all screened edits in wild type and Δ*ku70* strains. Aclonal refers to colonies that exhibit multiple genotypes. f) SDS-PAGE of the K3 peptide with native

yeast glycans and with human glycosylation following multiplexed glycosylation editing. g) Mass spectra of K3 peptide expressed in Δ*och1* (top) and engineered (bottom) cells. Identical samples treated with PNGase are shown in gray.

Table 1.

Cassettes for expression of sgRNAs based on native small RNAs

Cassette structure		Corresponding small RNA		
ID	**Promoter--fusion**	**Expression (log2tpm)**	**Edge resolution**	**Locus**
C1	P_{tRNA1}	12.2	0.91	chr1:1241621–1241706
C2	P_{tRNA1}--tRNA1	12.2	0.91	chr1:1241621–1241706
C3	P_{tRNA2}	9.5	0.86	chr3:953065–953139
C4	P_{srpRNA}	5.4	0.49	chr1:1523775–1524047
C5	P_{snoRNA}	3.4	0.44	chr2:635515–635661
C6	P_{tRNA3}--tRNA3	15.2	0.25	chr4:816524–816610
C7	P_{tRNA4}--tRNA4	8.5	0.66	chr4:1297483–1297568
C8	P_{tRNA5}--tRNA5	5.6	0.90	chr1:1609353–1609426
C9	P_{tRNA6}--tRNA6	2.2	0.20	chr1:2379876–2379912
C10	P_{tRNA7}--tRNA7	0.5	0.94	chr2:1606792–1606868