Validating genome-wide CRISPR-Cas9 function in the non-conventional yeast *Yarrowia lipolytica*

Cory Schwartz[1], Jan-Fang Cheng[2], Robert Evans[2], Christopher A. Schwartz[3], James M. Wagner[4], Scott Anglin[5], Adam Beitz[5], Weihua Pan[6], Stefano Lonardi[6], Mark Blenner[5], Hal S. Alper[4,7], Yasuo Yoshikuni[2], and Ian Wheeldon[1]*

1. Chemical and Environmental Engineering, University of California Riverside, Riverside, CA 92521
2. DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, CA 94598
3. Department of Civil and Mechanical Engineering, United States Military Academy, West Point, NY 10996
4. McKetta Department of Chemical Engineering, The University of Texas at Austin, 200 E Dean Keeton St. Stop C0400, Austin, TX 78712
5. Chemical and Biomolecular Engineering, Clemson University, Clemson, SC 29634
6. Computer Science and Engineering, University of California Riverside, Riverside, CA 92521 USA
7. Institute for Cellular and Molecular Biology, The University of Texas at Austin, 2500 Speedway Avenue, Austin, TX 78712

*Correspondence: iwheeldon@engr.ucr.edu

Abstract

Genome-wide mutational screens are central to understanding the genetic underpinnings of evolved and engineered phenotypes. The widespread adoption of CRISPR-Cas9 genome editing has enabled such screens in many organisms, but identifying functional sgRNAs still remains a challenge. To address this limitation, we developed a methodology to quantify the cutting efficiency of each sgRNA in a genome-scale library in the biotechnologically important yeast *Yarrowia lipolytica*. Screening in the presence and absence of native DNA repair enabled high-throughput quantification of sgRNA function leading to the identification of high efficiency sgRNAs that cover 94% of genes. Library validation enhanced the classification of essential genes by identifying inactive guides that create false negatives and mask the effects of successful disruptions. Quantification of guide effectiveness also creates a dataset from which functional determinants of CRISPR-Cas9 can be identified. Finally, application of the library identified mutations that led to high lipid accumulation and eliminated pseudohyphal morphology.

Introduction

A critical challenge in CRISPR-based library screens is the inability to separate active from inactive guides to ensure and quantify genome-wide coverage. Typically, multiple guide RNAs are designed to target each gene of interest with the hypothesis that some guides may not be functional[1, 2]. This strategy maximizes the likelihood of full genome coverage through redundant targeting of each gene and has been successful in identifying new phenotypes[3-5], but the presence of inactive guides can obscure screening results and create false negatives[6]. While *in silico* design and activity predictions are emerging, complete training sets that correlate CRISPR endonuclease activity, guide sequence, and local genetic context are not yet available.

Herein, we validate the genome-wide function of a synthetized library useful for CRISPR-Cas9 screens in the oleaginous yeast *Yarrowia lipolytica*. We selected this non-conventional yeast because it has value as a bioprocessing host for the conversion of biomass derived sugars and industrial waste streams (*e.g.*, glycerol, alkanes, and fatty acids) into value added chemicals and fuels[7-10]. Unlike the model yeast *Saccharomyces cerevisiae*, DNA repair in *Y. lipolytica* and most other eukaryotes is dominated by nonhomologous end-joining (NHEJ)[11, 12]. We reasoned that the effectiveness of each guide RNA in a high coverage CRISPR library could be quantified by comparing library evolution in an NHEJ-proficient strain and an NHEJ-deficient strain. In a strain lacking NHEJ and a homologous template (*i.e.,* NHEJ-disrupted haploid *Y. lipolytica*), double strand break repair is severely limited, and the most likely outcome of a Cas9-induced DNA double strand break is cell death[13]. Using this approach, CRISPR-Cas9 activity can be coupled to cell viability thus providing a facile, quantitative metric of guide efficiency.

Validated CRISPR libraries promise to generate more accurate and robust genetic screens with a drastically reduced false negative rate compared with libraries that are naïve to the cutting efficiency of each guide. We demonstrate this by identifying essential genes and by screening for the industrially relevant phenotypes of increased lipid accumulation and elimination of pseudohyphal cell morphology. In addition, when coupled with genome structure analysis (*i.e.*, nucleosome occupancy), validation experiments can provide insights into the biological effects that dictate guide RNA effectiveness and genome accessibility. Defining a list of validated guide RNAs for any given organisms is valuable in and of itself, and is also an important step in moving towards high coverage, combinatorial screens that simultaneously target multiple genes in the genome[14].

Results

Genome-wide CRISPR-Cas9 library design, construction, and analysis

We designed a library of single guide RNAs (sgRNAs) to target 7,854 coding sequences in *Y. lipolytica* PO1f (Figure 1a, Supplementary Fig. 1). Unique sgRNAs were designed to target the first 300 exon base pairs in each ORF, then scored and ranked based on their predicted on-target cutting efficiency[15]. The complete library was subsequently designed to contain the 6 highest scoring sgRNAs for each ORF, along with 480 nontargeting controls. The library was

synthesized and subsequently cloned into an expression vector with sgRNA expression driven by a synthetic RNA polymerase III (Pol III) promoter[16,17]. Sequencing revealed that over 97% of the designed sgRNAs were well represented in the library (Supplementary Fig. 2).

Growth screens in three different *Y. lipolytica* strains enabled us to experimentally determine two unique metrics for each sgRNA (Figure 1a). A fitness score (FS) was defined as the $\log_2$ ratio of the abundance of any given sgRNA in a Cas9 expressing strain (PO1f Cas9) to the abundance of the same sgRNA in the negative control (PO1f). Each sgRNA was also given a cutting score (CS) by calculating the same $\log_2$ ratio, but in this case comparing a Cas9 expressing NHEJ deficient strain (PO1f Cas9 *ku70*) to the control strain. FS is a measure of how harmful a CRISPR-Cas9 mediated indel was to cell growth, whereas CS is a measure of how efficient an sgRNA was for inducing a Cas9-mediated DNA double strand break. A pilot scale experiment with a nontargeting sgRNA and known functional sgRNAs targeting an essential gene and a nonessential gene confirmed that the outcomes of the FS and CS experiments were consistent with this scoring methodology (Supplementary Fig. 3).

The evolution of the full genome-wide library over three subculture cycles (6 days in total) and corresponding CS and FS values are shown in Figure 1b. Only after two growth cycles (day 4) were significant changes in the library distribution observed. The nontargeting control population behaved as expected, shifting to the upper right quadrant of the FS/CS plot indicating nonfunctional sgRNAs that resulted in no change in cell fitness. The day 4 and day 6 data indicate that the majority of the library was able to create double strand breaks. This is apparent from the shift toward the lower two quadrants. The lower left quadrant is indicative of sgRNAs that both cut (negative CS) and have a negative effect on growth (negative FS). The lower right quadrant also contains sgRNAs that cut effectively but have a neutral or positive effect on cell growth (positive FS). The upper left quadrant is only sparsely populated around the origin, likely because this quadrant represents instances with observed negative growth effects in the absence of a functional sgRNA (positive CS). Overall, three important trends are apparent: 1) the library skews strongly towards negative CS values, indicating that many of the sgRNAs are functional; 2) sgRNAs that do not produce a cut in the genome can be identified in the upper right and left quadrants; and, 3) a broad range of FS values results from outgrowth-based screens, suggesting that the library can be used to select for a wide variety of phenotypes.
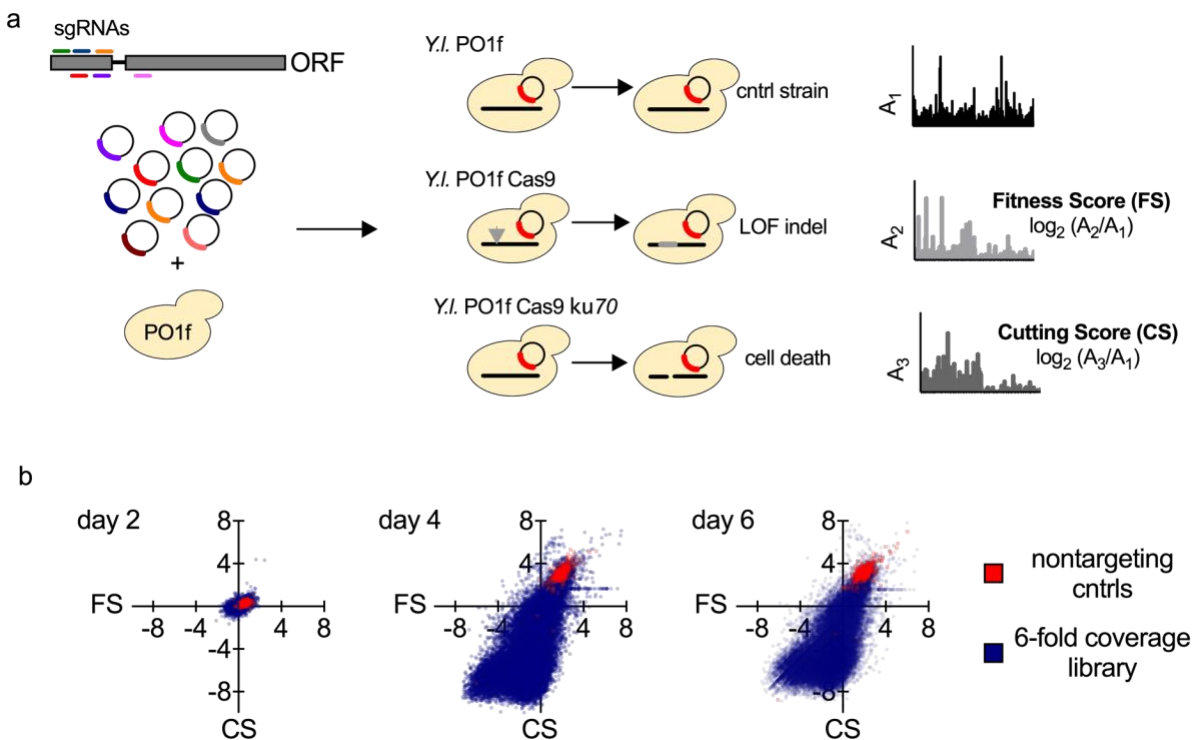
Figure 1. Genome-wide CRISPR-Cas9 validation and screening in *Yarrowia lipolytica*. (a) Schematic representation of sgRNA library design and workflow for genome-wide validation and screening. *Y. lipolytica* PO1f was used as the base strain for all experiments. Fitness score (FS) experiments used PO1f with constitutive expression of Cas9 from *Streptococcus pyogenes* (PO1f Cas9). Cutting score (CS) experiments used PO1f Cas9 with functionally disrupted nonhomologous end-joining (NHEJ) by inactivation of *KU70* (PO1f Cas9 *ku70*). LOF indel indicates loss of function insertion or deletion, and cntrl strain indicates the negative control strain that does not express Cas9 but does contain the sgRNA library. (b) Comparison of FS to CS for each sgRNA after 2, 4, and 6 days of growth on defined minimal media containing 2% glucose. Transformations were done in biological triplicates with at least 100-fold coverage of the library in each replicate. FS and CS values shown are the mean of the three biological replicates.

sgRNA validation and analysis of functional determinants

Based on the average CS of the nontargeting controls, we set a value of 2 as the threshold to identify inactive sgRNAs (*i.e.* noncutters). At the other end of the spectrum, we set a CS value of -5 (a 32-fold reduction in sgRNA abundance) as the lower limit of sgRNAs considered to be excellent cutters. This threshold was set based on a subpopulation of sgRNAs whose abundance was reduced to zero after two subculture cycles (4 days). Two intermediate categories were also defined: poor (-1<CS< 2) and moderate (-5<CS<-1) cutters. Figure 2a shows the CS distribution of the full library as well as six subpopulations ranked from best sgRNA per gene (1st) to worst sgRNA per gene (6th). Each subpopulation represents 1-fold coverage of the genome. The best subpopulation contained an excellent cutter for 94.6% of the targeted genes, while 82.6% of the genes have at least two excellent cutters. Histograms of each subpopulation, and the number of excellent cutters per gene are presented (Figures 2b and c, and Supplementary Figs. 4 and 5).
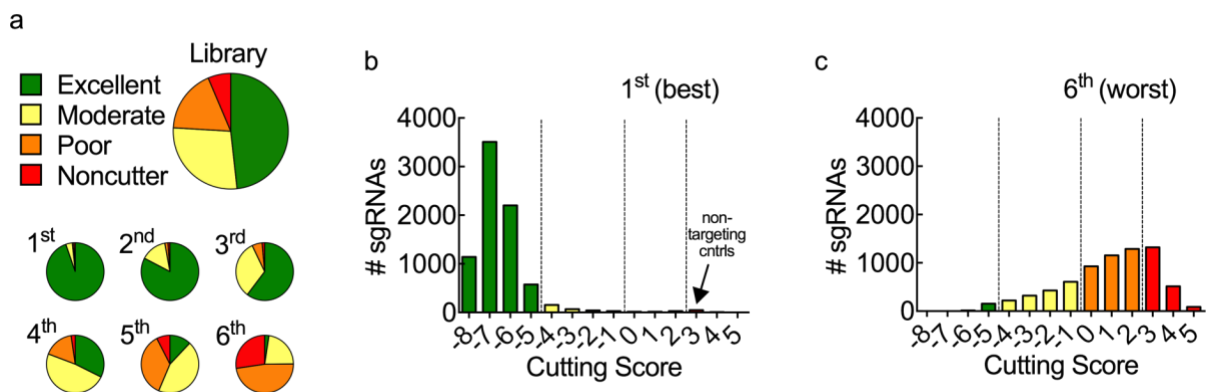
Figure 2. Genome coverage and distribution of CRISPR-Cas9 cutting efficiency. (a) Library and genome coverage by sgRNA classification: Excellent (CS<-5; green; 48% of full library), Moderate (-5<CS<-1; yellow; 28%), Poor (-1<CS<2; orange; 18%), and Noncutter (>2; red; 6%). The sgRNAs for each gene were ranked by CS from 1st to 6th and separated into six subpopulations each representing 1-fold genome coverage. (b,c) Histograms of the 1st and 6th sgRNA subpopulations. Mean CS values were determined after two growth cycles (4 days) from triplicate biological samples.

This CS analysis revealed that almost half of the library (48%) falls into the excellent category. Another 5% (2,828) were found to be inactive and were designated as noncutters (Figure 3a). To better understand the determinants of CRISPR-Cas9 function in *Y. lipolytica*, we looked for trends in the noncutter population. PolyT motifs (4 or more consecutive T's) are known terminator sequences for RNA Pol III promoters, and while 1,808 sgRNA sequences in the full library contained this motif, only 492 were identified as noncutters. The CS values of these subpopulations were significantly above that of the full library (Figure 3b). The variability of CS within the polyT-containing guides suggests that while this motif can negatively impact CRISPR function, its presence does not inherently exclude the possibility of efficient cutting. sgRNAs that target near the ends of chromosomes were also found to be enriched with noncutters. The CS values of sgRNAs proximal to the end of each chromosome (731 in total; see Methods) were significantly increased relative to the library, and 463 of these were noncutters (Figure 3b and Supplementary Fig. 6).

The remaining 1,873 noncutters showed no obvious correlation with internal terminator sequence, chromosomal position, or predicted sgRNA secondary structure (Supplementary Fig. 7). To further evaluate these sgRNAs, we also mapped CS values to experimental estimates of nucleosome occupancy across the genome[18]. As shown in Figure 3c, excellent cutters had a lower occupancy than the library as a whole, while the unexplained noncutters had a significant increase in occupancy. Overall, nucleosome occupancy correlated well with CS (Supplementary Fig. 8). This data reveals that occupancy provides a stronger correlation to CRISPR-Cas9 activity (Pearson's, $r = 0.223$) than the algorithm used to design the library[15] ($r = 0.0525$) and other algorithms[19, 20] ($r = 0.0511$; $r = 0.1431$; Supplementary Fig. 9).
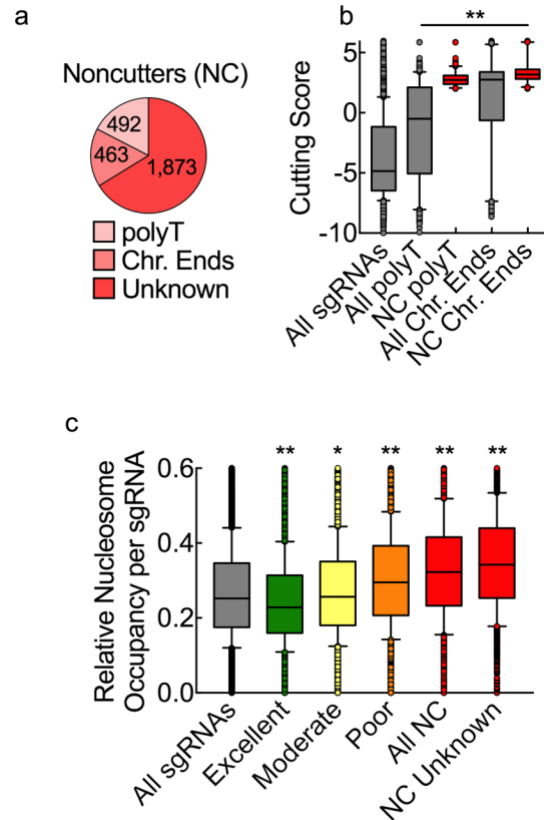
Figure 3. Identifying determinants of CRISPR-Cas9 function. (a) Classification and distribution of noncutting (NC) sgRNAs. (b) Cutting score (CS) of all sgRNAs (n = 46,234), sgRNAs that contain a polyT motif in the spacer sequence (All polyT; n = 1,808), noncutter sgRNAs with a polyT sequence (NC polyT; n = 492), sgRNAs proximal to chromosomal ends (All Chr. Ends; n = 731), and noncutter sgRNAs near the end of each chromosome (NC Chr. Ends; n = 463). Statistical significance was determined by one-way ANOVA with Dunnett's multiple comparison (F-value = 1,588; P-values to all sgRNAs < 0.0001 for all subpopulations). (c) Average nucleosome occupancy per sgRNA for all sgRNAs (n = 46,234), excellent (n = 21,898), moderate (n = 12,592), poor (n = 7,993) and noncutter sgRNAs (n = 2,828), and noncutter sgRNAs that do not contain a polyT motif and are not proximal to chromosomal ends (NC Unknown; n = 1,873). Statistical significance was determined by one-way ANOVA with Dunnett's multiple comparison (F-value = 544; P-values to all sgRNAs < 0.0001 for excellent, poor, and NC subpopulations, and p = 0.0041 for the moderate sgRNAs). Statistical significance is shown graphically with * p<0.01 and ** p<0.0001. Boxes extend from the 25th to 75th percentiles. The box line is plotted at the median. The whiskers extend from the 10th to the 90th percentiles.

Activity validated CRISPR-Cas9 library improves essential gene analysis

Inactive sgRNAs in CRISPR libraries can produce false negatives in essential and nonessential gene classification[6]. This problem arises when one or more poor cutting sgRNAs mask the effect of successful disruptions on the FS average for a particular gene. The second dimension of data provided by CS experiments can eliminate this issue by focusing analysis on the validated cutters and excluding inactive sgRNAs. Figure 4a shows the rank-ordered FS calculated from the full library, which is naïve to the effectiveness of each sgRNA (the naïve

library). Twelve nonessential genes and twelve genes that are known to be essential across eukaryotes are shown as two subpopulations (Figure 4 and Supplementary Table 1). The subpopulations partially overlapped and had closely related FS distributions when calculated from the naïve library (Figures 4a and b). Three of the genes are notable; *ACT1*, *MYO1*, and *FOL2* are genes critical to eukaryotic cell viability, but were not distinguishable from nonessential genes, suggesting that the presence of poor cutting sgRNAs artificially increased their FS. Analysis with only excellent cutters (the validated library) resulted in a clear separation between the subpopulations: *ACT1*, *MYO1*, and *FOL2* clustered as expected with other essential genes, and the difference between FS of the essential and nonessential subpopulations increased significantly (Figures 4c and d).

One method of evaluating sgRNA effectiveness is to compare multiple targets to the same known essential gene[19, 21]. By definition, disruptions to essential genes are fatal (with or without intact DNA repair), which leads to a hypothesis similar to that of the CS experiments. That is, double strand breaks lead to cell death, therefore viability in a growth screen can be used as a measure of cutting efficiency. This hypothesis is supported by the observed trends in the essential and nonessential subpopulations (Figures 4e and f). The regression line of CS/FS data for all sgRNAs targeting the selected essential gene subpopulation has a slope of 0.997 (r = 0.82), suggesting a quantitative correlation between FS and CS, and providing experimental evidence to support CS as a metric for CRISPR-Cas9 activity. Appropriately, this relationship does not hold for the nonessential gene subpopulation.
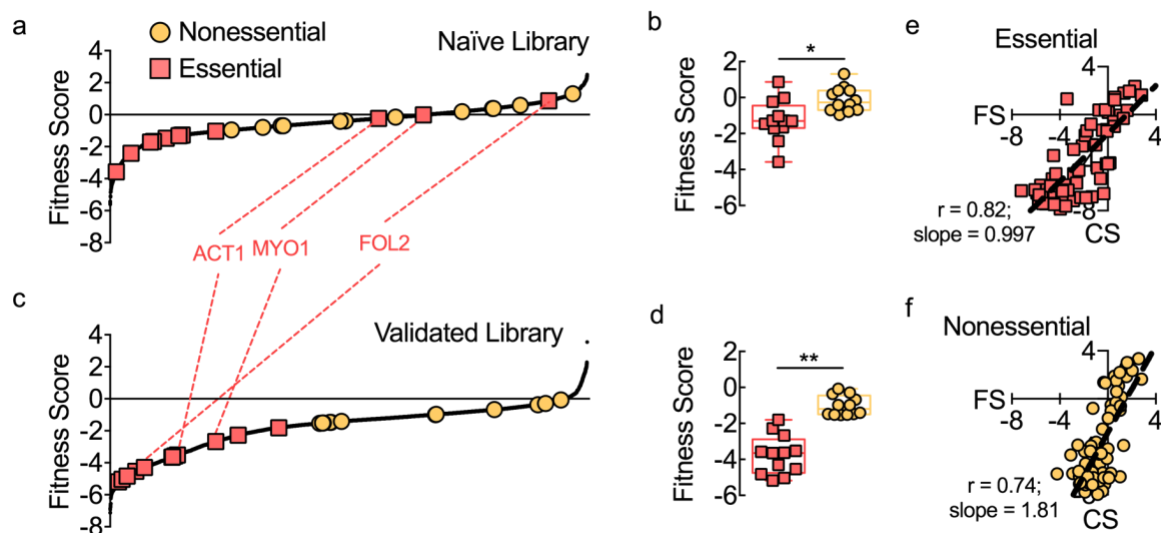

Figure 4. Library validation enhances essential gene identification. (a) Rank-ordered fitness score (FS) for each gene calculated using the 6-fold sgRNA library (naïve library). Twelve nonessential (yellow) and putatively essential (red) genes are shown. (b) FS for essential and nonessential genes using the naïve library. Boxes extend from the 25th to 75th percentiles, the box line indicates the median, and the whiskers extend from the 10th to the 90th percentiles, n = 12 for both nonessential and essential. Comparison between the means was accomplished by two-sided T-test (t = 3.029 and p = 0.0062). Statistical significance is shown graphically with * indicating p<0.01 and ** indicating p<0.0001. (c) Rank-ordered FS for each gene calculated using only the validated excellent sgRNAs (validated library). For the 5.4% of genes that did not have an excellent sgRNA, the best cutter was included. (d) FS for essential and nonessential gene subpopulations using the validated library (t = 7.764 and p < 0.0001). (e,f) Comparison of FS to CS for each sgRNA targeting the essential and nonessential subpopulations.

Pearson's coefficients from linear regression (r) and slopes are shown (n = 72). FS and CS values are the mean of three biological replicates, the standard deviations associated with the means are presented in Supplementary Fig. 10.

The validated library identified 1,450 ORFs as essential in *Y. lipolytica*, which represents 18.4% of coding sequences (Figure 5). Using the same statistical tests, the naïve library identified only 5.4% of the genes (424 ORFs) as essential. Genome-wide knockout collections in model yeasts suggest that ~20% of yeast genes are essential (19.5% or 1,268 ORFs in *S. cerevisiae*[22] and 26.1% or 1,260 ORFs in *Schizosaccharomyces pombe*[23]), indicating that the naïve library produced a high number of false negatives. A recent transposon-based analysis in *Y. lipolytica* supports our results[24]. The transposon screen identified 22.4% of *Y. lipolytica* genes as essential, an increase in 513 ORFs over the CRISPR-based analysis. The validated CRISPR library implicitly accounts for genomic targets that may be inaccessible due to heterochromatin structure, a known experimental limitation of transposon-based screens that can cause false positives[25]. The difference between these data sets is also likely due to challenges in transposon targeting of short ORFs (Supplementary Fig. 11).
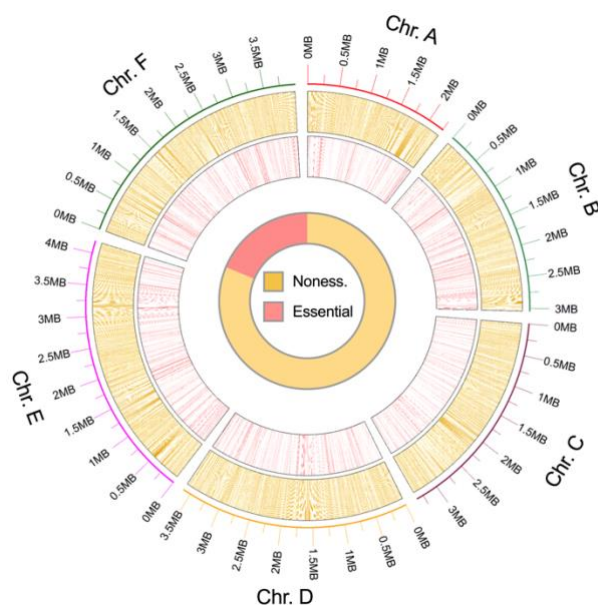

Figure 5. *Y. lipolytica* genome map. *Y. lipolytica* genome map with all nonessential and essential genes when grown on synthetic minimal media with 2% glucose at 30 °C are shown by chromosomal position. Nonessential genes (n = 6,404; yellow) are shown in the outer ring, and essential genes (n = 1,450; red) are shown in inner ring.

Of the 1,450 ORFs identified as essential for growth of *Y. lipolytica*, 1,003 had a homolog in *S. cerevisiae*, 502 of which were found to be essential in both organisms (Supplementary Fig. 12). 156 genes essential in *Y. lipolytica* are conditionally essential in *S. cerevisiae* for respirative growth. As *Y. lipolytica* is an obligate aerobe, respiration is essential. 98 more ORFs are duplicated in the *S. cerevisiae* genome, and 71 are involved in amino acid or nucleotide biosynthesis. GO term analysis[26] was also performed on all genes that had a *S.*

9

cerevisiae homolog. Genes involved in transcription, translation, cellular organization, and cell cycle were found to have significantly lower FS (Supplementary Fig. 13).

Genome-wide screening for growth and nongrowth associated phenotypes

To demonstrate the utility of the CRISPR library, we performed a series of growth and nongrowth associated screens to identify genes linked to new phenotypes. The classic genetic test for canavanine resistance was successful. Ninety-four percent of the sequenced colonies that survived the canavanine challenge (50 mg/L over 2 days of growth) harbored an expression vector encoding a *CAN1* targeting sgRNA (Figure 6a). All identified sgRNAs were excellent cutters.

In a second screen, *Y. lipolytica* strains with increased lipid staining were isolated from the CRISPR library using a fluorescent lipid dye and fluorescence activated cell sorting (Figure 6b). A total of four unique genes associated with a high lipid staining phenotype were identified in this screen. Two of the four identified genes have known roles in lipid turnover and growth on fatty acids in other yeast species[27, 28]. A third gene is likely involved in the enzymatic processing of acyl-CoA species[29], a function that is often intertwined with lipid metabolism (*e.g.*, β-oxidation). The final gene in the set has no known function or similarity to a well-annotated sequence, but two different excellent sgRNAs targeting the gene were isolated from the high lipid staining subpopulation. This redundant targeting supports the validity of the result and demonstrates the utility of the screen to identify novel genotype to phenotype relationships.
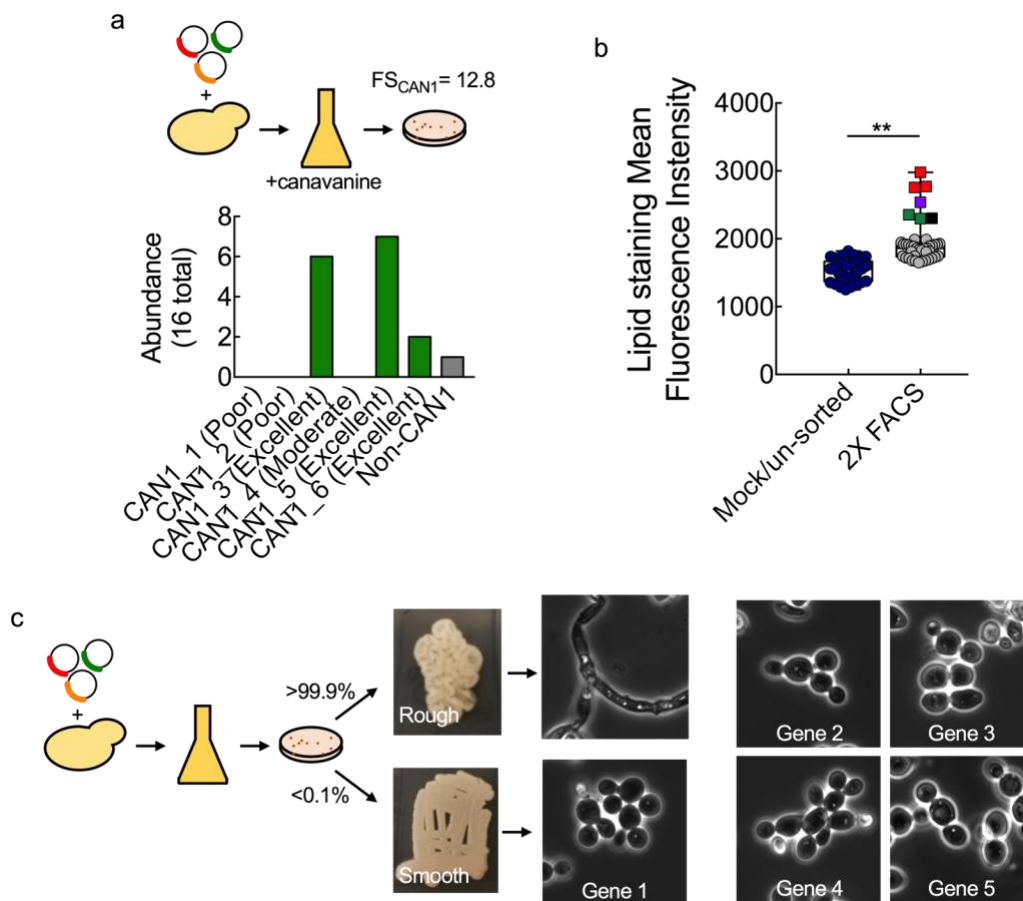
10

Figure 6. Genome-wide CRISPR-Cas9 screens for growth and non-growth associated phenotypes.
(a) Growth screen for canavanine resistance. Cells were grown in the presence of 50 mg L^{-1} canavanine for 3 days, plated and sequenced. (b) Two populations were stained and analyzed using fluorescence activated cell sorting (PO1f with library and PO1f Cas9 with library). The PO1f Cas9 strain with library was sorted twice for high lipid producers to yield a high lipid producing population (2x FACS Pool). Several strains from the high lipid producing population were isolated and compared to a mock/un-sorted control consisting of cells subjected to the same produce but without a fluorescence sorting cutoff. Comparison between the means was accomplished by two-sided T-test (n = 34, 41; t = 6.713; and, p < 0.0001). Statistical significance is shown graphically with * indicating p<0.01 and ** indicating p<0.0001. (c) Non-growth associated screen for altered cell morphology. Plated colonies were visually inspected for a smooth colony appearance, indicating a lack of pseudohyphal cell morphology. 5 of ~50,000 colonies were identified with a smooth phenotype. Rough and smooth colony appearance along with brightfield microscopy images are shown.

A third screen tested the ability of the library to screen for a nongrowth associated phenotype. *Y. lipolytica* is a dimorphic yeast, where the transition from yeast to pseudophyphal morphology is often precipitated by environmental stress or nutrient limitations[30]. Multiday growth on solid media is often sufficient to trigger this transition, which can be readily observed by visual inspection of agar plates (Figure 6c). To identify strains deficient in pseudohyphal formation, we plated the library, identified colonies with altered morphology, confirmed cellular phenotype by microscopy, and sequenced. A total of 5 unique genes were identified as necessary

11

for the pseudohyphal transition. These results demonstrate the utility of this library to rapidly identify novel targets for phenotypes of interest.

Discussion

With genome-scale mutagenesis technologies such as MAGE, TRMR, Tn-seq and other similar approaches, it is now possible to generate and track genome edits and their effects on native and engineered phenotypes[31-35]. Genome-wide CRISPR systems are complementary to these approaches[3-5, 36, 37] and can have the added functionality of genome targeted site saturation mutagenesis[38-41]. However, these techniques have not yet been widely translated to non-model and other industrially and medically relevant organisms. The widespread adoption of CRISPR–Cas9 promises to enable genome-wide engineering and screening for a vast array of hosts[3, 5, 12, 42-45]. Here, we designed and applied such a system for functional genomic screening and non-native phenotype engineering in the yeast *Y. lipolytica*. Key to the success of the system was our ability to functionally disable native nonhomologous DNA repair, which allowed quantification of the cutting efficiency of each sgRNA-Cas9 complex in parallel and yielded a list of validated sgRNAs for nearly every gene in the *Y. lipolytica* genome.

The additional genome-wide data created by cutting score (CS) experiments also generates new information that can be used to identify the determinants of CRISPR-Cas9 function, as well as improve screen accuracy. Our experiments in *Y. lipolytica* revealed that 48% of guides in the 6-fold coverage library were highly active, and that 94.6% of protein coding sequences were targeted by at least one excellent cutting sgRNA (Figure 2). This library validation also substantially improved essential/nonessential gene classification (Figures 4 and 5) by significantly reducing false negatives that arise in the naïve library. Nongrowth associated screens were also successful in identifying novel gene disruptions for the industrially-relevant phenotypes of high lipid accumulation and elimination of pseudohyphal growth (Figure 6). Correlation of CS data with nucleosome occupancy and several sgRNA scoring algorithms revealed that occupancy was a stronger predictor of function than *in silico* design (Figure 3; see Refs. [15, 18-20, 46]). We also observed that chromosomal ends were difficult to target. Current scoring algorithms have been limited by training set quality and size, relying on subsets of essential genes and/or high expression reporters to learn about the relationships between guide RNA sequence and resultant Cas9 activity[15, 19, 21]. The genome-wide activity measurements obtained in the CS experiments provide a substantially richer data set for correlating guide sequence parameters to system-level characteristics. We envision that our experiments in *Y. lipolytica* may also be possible in other eukaryotes by genetically disrupting or chemically inhibiting native nonhomologous DNA repair[47]. As such, the approach described here not only resulted in a new validated screening tool for *Y. lipolytica*, but also represents a generalizable methodology for enhancing CRISPR-based screens in other organisms.

Author Contributions
CS and IW conceived the study. CS, CAS, and IW designed the library. JFC, RE, and YY constructed the library. CS, WP, SL, and IW performed library validation. CS, SA, AB, MB, and

IW designed, completed and analyzed the essential gene experiments. CS, JW, HA, and IW completed and analyzed phenotypic screens. All authors wrote and edited the manuscript.

Acknowledgements

Methods

Strains and media

All strains were derived from *Yarrowia lipolytica* strain PO1f (MatA, leu2-270, ura3-302,xpr2-322, axp-2)[48]. The PO1f Cas9 strain was constructed by markerless integration of a UAS1B8-TEF(136)-Cas9-CycT expression cassette into the A08 site[17, 49]. The PO1f Cas9 *ku70* strain was constructed by disrupting *KU70* using CRISPR-Cas9 as previously described[16].

Yeast culturing was done at 30 °C, unless specifically noted otherwise. Yeast strains were grown in YPD (1% Bacto yeast extract, 2% Bacto peptone, 2% glucose) for nonselective growth. Cells transformed with sgRNA-expressing plasmids were selected in media deficient in leucine (SD-leu; 0.67% Difco yeast nitrogen base without amino acids, 0.069% CSM-leu (Sunrise Science, San Diego, CA), and 2% glucose). For canavanine resistance screening, 50 µg/mL was added to SD-leu. For lipid overproduction experiments, SDL-leu (0.67% Difco yeast nitrogen base without amino acids, 0.069% CSM-leu, and 8% glucose) was used for outgrowth.

Plasmid construction

For integration of Cas9, pIW1009 (UAS1B8-Cas9 integration into A08 site) was constructed. pHR_A08_hrGFP (Addgene #84615) was digested with BssHII and NheI, and Cas9 was inserted via Gibson Assembly after PCR via Cr_1250 and Cr_1254 from pCRISPRyl (Addgene #70007). To facilitate cloning of the library of sgRNAs, pIW771 (SCR1'-tRNA-AvrII site) was generated by digesting pCRISPRyl with BamHI and HindIII and circularizing.

sgRNA library design

Custom Matlab scripts were used to design each sgRNA in the library based on a recent assembly of the *Y. lipolytica* PO1f parent strain genome, W29[50]. sgRNAs were designed for all non-redundant protein coding sequences. The top 6 sgRNAs, as ranked by on target efficiency score[15] that met the following criteria were selected for each gene: (1) uniqueness of target sequence and PAM (final 12 bp + NGG) in genome, (2) located within first 300 bp of gene, and (3) target sequence present in both genome and mRNA sequence (to avoid targeting introns). 480 sgRNAs of random sequence, confirmed to not target in the genome, were included as nontargeting controls.

sgRNA library cloning

Four oligo pools consisting of 25% of the designed sgRNAs each were ordered from Twist Bioscience. Equal volumes of the oligo pool (1 µM) and a complementary primer (5 µM) were mixed together and annealed using the Thermo annealing advanced protocol and the calculated Tm of the complementary primers using IDT's OligoAnalyzer 3.1. For libraries 1 and 3, second strand synthesis reactions were completed using T4 DNA polymerase (NEB) and sgRNA-Rev2, gel extracted, and purified using Zymo Research Zymo-Spin 1 columns. Libraries 2 and 4 were amplified via Q5 DNA polymerase (NEB) using 60mer_pool-F and spacer-AarI.rev or pLeu-mock-sgRNA.fwd and sgRNA-Rev2 with 0.2 picomoles of DNA as template for 7 cycles and column purified.

For libraries 1, 3, and 4, Gibson Assembly HiFi HC 1-step Master Mix (SGI-DNA) were used to clone into pIW771 digested with AarI. Library 2 was digested with AarI and cloned into pIW771 digested with AarI with a GoldenGate reaction using T4 DNA ligase (NEB). Cloned DNA was transformed into NEB 10-beta *E. coli* and plated. Sufficient electroporations were performed for each library to yield a >10-fold excess colonies for the number of library variants, and the plasmid library was isolated.

Yeast transformation and screening

Transformation of *Y. lipolytica* with the sgRNA plasmid library was done using a previously described method, with slight modifications[16]. 2 mL of YPD was inoculated with a single colony of the strain of interest and grown in a 14 mL tube with shaking at 200 RPM for 24-30 h. 1.5 mL of cells were washed with transformation buffer (0.1 M LiAc, 10 mM Tris (pH=8.0), 1 mM EDTA) via centrifugation at 4,000 *g* and resuspended in 600 μL transformation buffer. To these cells, 18 μL of ssDNA mix (8 mg/mL Salmon Sperm DNA, 10 mM Tris (pH=8.0), 1 mM EDTA), 90 μL of β-mercaptoethanol mix (5% β-mercaptoethanol, 95% triacetin), and 4 μg of plasmid library were added, mixed via pipetting, and incubated. After incubation at room temperature for 30 min, 900 μL of PEG mix (70% w/v PEG (3,350 MW)) was added and mixed via pipetting, and incubated at room temperature for 30 min. Cells were then subjected to heat shock for 25 min at 37 °C, washed with 10 mL H_2O, and used to inoculate 25 mL of SD-leu media for outgrowth experiments. Dilutions of the transformation were plated on solid SD-leu media to calculate transformation efficiency. Transformations were pooled as necessary to ensure adequate diversity to maintain library representation (100x coverage, 5 x 10^6 total transformants per biological replicate).

Outgrowth was allowed to proceed in 25 mL of liquid media in a 250 mL baffled flask. After 2 days, cells reached confluency ($OD_{600} > 10$), and approximately 150 μL (at least 200-fold library coverage) were used to inoculate 25 mL of fresh media as desired. Also at each timepoint, 1 mL of culture was taken, DNase I (New England Biolabs) and the corresponding buffer were added, and the mix was incubated for 1 h at 30 °C to digest any plasmid DNA in the media. Pellets were then frozen and stored at -80 °C for future analysis. Each growth cycle from inoculation to sampling represented approximately 7 cell doublings.

Library isolation and sequencing

Pellets were thawed and resuspended in 1 mL H_2O. Cells were split into 5 samples of 200 μL, and plasmid was isolated using a Zymo Yeast Miniprep Kit (Zymo Research). Samples from a single pellet were pooled, and plasmid copy number was quantified using RT-qPCR with M13_F and M13_R and SsoAdvanced Universal SYBR Green Supermix (Biorad) and verified to be higher than 1 x 10^7.

To prepare samples for next generation sequencing, isolated plasmid was subjected to PCR using forward (Cr_1665-1668) and reverse primers (Cr_1669-1673 and Cr_1709-1711) containing all necessary sequences and barcodes[1] (Supplementary Table 2). At least 1 x 10^7 plasmids were used as template, and PCR reactions were not allowed to go to completion to avoid biased amplification. PCR product at 250 bp was gel extracted, samples were pooled at

equimolar amounts, and submitted for sequencing on a NextSeq500 at the UCR IIGB core facility.

Next generation sequencing results were demultiplexed and mapped to each sgRNA using custom scripts. A total of 453 sgRNAs were not present in the sequencing data. In addition, 440 sgRNAs had only a very low number of reads (less than 10) in the untransformed library or were erroneously designed, and were excluded from downstream analysis. Pairwise comparison between normalized read abundances for biological replicates were done to verify consistency (Supplementary Fig. 14).

Cutting and fitness score analysis

To calculate fitness score (FS) and cutting score (CS) for each replicate, the number of reads for each sgRNA in each sequencing sample was normalized to the total number of reads, with a pseudo-count of 0.5 added when no reads were mapped to a given sgRNA for a given biological replicate. Normalized read counts for biological triplicates were then averaged together. The $\log_2$ for each abundance was then taken, and then the FS was calculated by subtracting the PO1f $\log_2$ abundance from the PO1f Cas9 $\log_2$ abundance. CS was calculated in analogous way, but with PO1f Cas9 replaced with PO1f Cas9 *ku70*.

sgRNA analysis

Analysis of sgRNA characteristics was done using a range of publicly available tools and data. Nucleosome occupancy was determined by mapping sgRNAs to previously published nucleosome occupancy data[18], adding the occupancy for each base pair in the target sequence, and dividing by 20. The scoring for predicting on-target activity for each sgRNA was carried out using previously published [15, 19, 20]. Secondary structure was predicted using the default settings for RNA using Quikfold on the DINAMelt web server (http://unafold.rna.albany.edu/)[51]. polyT sequences were identified by searching each sgRNA sequence for 4 consecutive "T". sgRNAs at chromosome ends were annotated as the 50 sgRNAs nearest either end of each chromosome, except for the end of chromosome A (84 sgRNAs), the beginning of chromosome C (126 sgRNAs), and the end of chromosome E (71 sgRNAs). Ends of chromosomes determinations were informed by where sgRNA cutting scores increased.

Essential gene identification

Essential genes were identified based on FS. A two-sided T-test was used to determine which genes had a FS lower than the highest of the 12 putative essential genes plus two standard deviations (*POL1* (YALI_E14644), FS<-1.38 for validated library). Samples with p values < 0.05 were considered significant. A second two-sided T-test was performed to identify which genes had a FS lower than the 12 putative nonessential genes (p<0.05), with a Bonferroni correction for multiple comparisons. The same series of tests were also applied to the FS obtained from analysis with the 6-fold coverage library (the naïve library). In this case, the initial FS cut off was from *FOL2* (YALI1_A10254, FS<1.53).

Essential gene comparison to *S. cerevisiae*

16

Mapping of all *Y. lipolytica* genes to *S. cerevisiae* genes was carried out using BLAST (E<10^{-30}). Essential genes in *S. cerevisiae* were those annotated as being "inviable" null mutants[52].

Canavanine screen

Glycerol stocks of the transformed library taken from day 2 cultures were grown in SD-leu for 2 days, and then used to inoculate SD-leu with 50 μg/mL canavanine. Cultures were then allowed to grow to confluency and plated on solid SD-leu. Single colonies were subjected to colony PCR using Cr_1742 and Cr_1743 to identify the sgRNA expressed in each colony.

Morphology screen

Glycerol stocks of the transformed library taken from day 2 cultures were grown in SD-leu for 2 days, and then plated on solid SD-leu. Approximately 50,000 colonies were examined for a smooth morphology after 4 days of growth on solid media. Colonies were restreaked, and smooth morphology confirmed. All colonies with confirmed smooth morphology were observed using an Olympus BX51 microscope (UPlanFL 100X 1.30 oil-immersion objective lens), and images were captured with a Q-Imaging Retiga Exi CCD camera and processed using Cellsense Dimension 1.7 software.

LipidTOX Deep Red Staining, Fluorescence Activated Cell Sorting (FACS), and Flow Cytometry Confirmation

CRISPR libraries were cultivated for FACS in media and conditions that were previously established for *Y. lipolytica* lipid characterization[10]. Specifically, cultures were started from frozen glycerol stocks (PO1f with CRISPR library and PO1f Cas9 with CRISPR library, starting from day 2 stocks) by inoculating 50 mL of SD-leu media in a 250 mL baffled shake flask, followed by 2 days of outgrowth at 28°C. The resulting inoculum cultures were then used to start 50 mL lipid production cultures in SDL-leu media (initial optical density of 0.1) with 80 g/L glucose in 250 mL shake flasks. These shake flask cultures were grown using air-permeable plugs by shaking for 3 days at 28°C and 225 RPM in a shaking incubator (New Brunswick Scientific I 26). HCS LipidTOX Deep Red (Thermo Fisher Scientific) was then used to fluorescently stain lipids according to the manufacturer-recommended protocol: cells were pelleted, all culture media was removed, and then the cell pellet was re-suspended in 100 μL of a 1:200 dilution of lipid stain in phosphate buffered saline (PBS). Two successive rounds of fluorescence activated cell sorting (FACS) were performed on a BD FACS Aria III using the APC laser/filter set. For high lipid sorting, the top 0.3% of 3×10^6 APC$^+$ cells and the top 3% of 1×10^5 APC$^+$ cells were collected. After the first APC sort, the collected cells were outgrown in SD-leu to saturation (3 days) and glycerol stocked. Fresh cultures were then started from the frozen glycerol stock by inoculating 50 mL of SD-leu media in a 250 mL baffled shake flask, followed by 2 days of outgrowth at 28°C. The resulting inoculum culture was then used to start a 50 mL lipid production culture in SDL-leu media with 80 g/L glucose (initial optical density of 0.1) in a 250 mL shake flask. This shake flask culture was grown using air-permeable plugs by shaking for 3 days at 28°C and 225 RPM in a shaking incubator (New Brunswick Scientific I 26) until ready for the second sort. After the second APC sort, the collected cells were outgrown in SD-leu to saturation (3 days) and glycerol stocked again. A mock-sorted / un-sorted pool was also generated in parallel throughout this double sorting process by passing the same pool of

17

cells through the same FACS process/conditions without applying the sorting gate (i.e. all cells were collected, regardless of APC fluorescence).

From the 2X sorted (2X FACS) glycerol stock and mock/un-sorted glycerol stock, individual colonies were then isolated on SD-leu plates to isolate individual strains. 48 colonies from the 2X FACS and 48 colonies from the mock/un-sorted plates were then subjected to colony PCR and Sanger sequencing of the sgRNA cassette. Clones with sgRNAs that were successfully colony PCR amplified, sequenced, and identified as cutters (see cutting and fitness score analysis above) were subsequently re-screened to establish a lipid staining mean fluorescence intensity for individual clones (41 from 2X FACS, 34 from mock/unsorted) using the same procedure as for FACS, but with a BD LSRII Fortessa analytical flow cytometer for quantifying APC channel fluorescence.

References

1. Joung, J. et al. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc* **12**, 828-863 (2017).
2. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* **15** (2014).
3. Chen, S.D. et al. Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis. *Cell* **160**, 1246-1260 (2015).
4. Koike-Yusa, H., Li, Y.L., Tan, E.P., Velasco-Herrera, M.D. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature biotechnology* **32**, 267-273 (2014).
5. Sidik, S.M. et al. A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes. *Cell* **166**, 1423-+ (2016).
6. Ong, S.H., Li, Y., Koike-Yusa, H. & Yusa, K. Optimised metrics for CRISPR-KO screens with second-generation gRNA libraries. *Sci Rep-Uk* **7** (2017).
7. Xue, Z.X. et al. Production of omega-3 eicosapentaenoic acid by metabolic engineering of Yarrowia lipolytica. *Nature Biotechnology* **31**, 734-+ (2013).
8. Qiao, K.J., Wasylenko, T.M., Zhou, K., Xu, P. & Stephanopoulos, G. Lipid production in Yarrowia lipolytica is maximized by engineering cytosolic redox metabolism. *Nature Biotechnology* **35**, 173-177 (2017).
9. Schwartz, C., Frogue, K., Misa, J. & Wheeldon, I. Host and Pathway Engineering for Enhanced Lycopene Biosynthesis in Yarrowia lipolytica. *Front Microbiol* **8** (2017).
10. Wagner, J.M., Williams, E.V. & Alper, H.S. Developing a piggyBac Transposon System and Compatible Selection Markers for Insertional Mutagenesis and Genome Engineering in Yarrowia lipolytica. *Biotechnol J* **13**, e1800022 (2018).
11. Mao, Z.Y., Bozzella, M., Seluanov, A. & Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA repair* **7**, 1765-1771 (2008).
12. Lobs, A.K., Schwartz, C. & Wheeldon, I. Genome and metabolic engineering in non-conventional yeasts: Current advances and applications. *Synthetic and Systems Biotechnology*, 1-10 (2017).
13. Schwartz, C., Frogue, K., Ramesh, A., Misa, J. & Wheeldon, I. CRISPRi repression of nonhomologous end-joining for enhanced genome engineering via homologous recombination in Yarrowia lipolytica. *Biotechnology and Bioengineering* **114**, 2896-2906 (2017).
14. Shen, J.P. et al. Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. *Nat Methods* **14**, 573-+ (2017).
15. Doench, J.G. et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature biotechnology* **32**, 1262-U1130 (2014).
16. Schwartz, C.M., Hussain, M.S., Blenner, M. & Wheeldon, I. Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR-Cas9-Mediated Genome Editing in Yarrowia lipolytica. *ACS synthetic biology* **5**, 356-359 (2016).
17. Schwartz, C., Shabbir-Hussain, M., Frogue, K., Blenner, M. & Wheeldon, I. Standardized Markerless Gene Integration for Pathway Engineering in Yarrowia lipolytica. *ACS synthetic biology* **6**, 402-409 (2017).
18. Tsankov, A.M., Thompson, D.A., Socha, A., Regev, A. & Rando, O.J. The Role of Nucleosome Positioning in the Evolution of Gene Regulation. *Plos Biol* **8** (2010).
19. Doench, J.G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology* **34**, 184-+ (2016).
20. Labuhn, M. et al. Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications. *Nucleic Acids Research* **46**, 1375-1385 (2018).

21.   Xu, H. et al. Sequence determinants of improved CRISPR sgRNA design. *Genome Res* **25**, 1147-1157 (2015).

22.   Cherry, J.M. The Saccharomyces Genome Database: Advanced Searching Methods and Data Mining. *Cold Spring Harbor protocols* **2015**, pdb prot088906 (2015).

23.   Kim, D.U. et al. Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. *Nature biotechnology* **28**, 617-623 (2010).

24.   Patterson, K. et al. Functional Genomics for the Oleaginous Yeast Yarrowia Lipolytica. *Metab Eng* (2018).

25.   Gangadharan, S., Mularoni, L., Fain-Thornton, J., Wheelan, S.J. & Craig, N.L. DNA transposon Hermes inserts into DNA in nucleosome-free regions in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21966-21972 (2010).

26.   Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).

27.   Henry, S.A., Kohlwein, S.D. & Carman, G.M. Metabolism and regulation of glycerolipids in the yeast Saccharomyces cerevisiae. *Genetics* **190**, 317-349 (2012).

28.   Strijbis, K. et al. Identification and characterization of a complete carnitine biosynthesis pathway in Candida albicans. *Faseb J* **23**, 2349-2359 (2009).

29.   Sherman, D., Durrens, P., Beyne, E., Nikolski, M. & Souciet, J.L. Genolevures: comparative genomics and molecular evolution of hemiascomycetous yeasts. *Nucleic acids research* **32**, D315-318 (2004).

30.   Barth, G. & Gaillardin, C. Physiology and genetics of the dimorphic fungus Yarrowia lipolytica. *Fems Microbiol Rev* **19**, 219-237 (1997).

31.   Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-898 (2009).

32.   van Opijnen, T., Bodi, K.L. & Camilli, A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* **6**, 767-772 (2009).

33.   Warner, J.R., Reeder, P.J., Karimpour-Fard, A., Woodruff, L.B.A. & Gill, R.T. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nature biotechnology* **28**, 856-U138 (2010).

34.   Si, T. et al. Automated multiplex genome-scale engineering in yeast. *Nature communications* **8** (2017).

35.   Si, T., Luo, Y., Bao, Z. & Zhao, H. RNAi-assisted genome evolution in Saccharomyces cerevisiae for complex phenotype engineering. *ACS synthetic biology* **4**, 283-291 (2015).

36.   Hart, T. et al. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163** (2015).

37.   Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096-1101 (2015).

38.   Garst, A.D. et al. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nature biotechnology* **35**, 48-55 (2017).

39.   Roy, K.R. et al. Multiplexed precision genome editing with trackable genomic barcodes in yeast. *Nature biotechnology* (2018).

40.   Bao, Z. et al. Genome-scale engineering of Saccharomyces cerevisiae with single-nucleotide precision. *Nature biotechnology* **36**, 505-508 (2018).

41.   Guo, X., Chavez, A., Tung, A., Chan, Y., Kaas, C., Yin, Y., Cecchi, R., Garnier, S.L., Kelsic, E.D., Schubert, M., DiCarlo, J.E., Collins, J.J., Church, G.M. High-throughput creation and functional profiling of DNA sequence variant libraries using CRISPR-Cas9 in yeast. *Nature biotechnology* **36**, 540-546 (2018).

42. Lobs, A.K., Engel, R., Schwartz, C., Flores, A. & Wheeldon, I. CRISPR-Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in Kluyveromyces marxianus. *Biotechnol Biofuels* **10** (2017).

43. Shan, Q.W. et al. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature biotechnology* **31**, 686-688 (2013).

44. Nodvig, C.S., Nielsen, J.B., Kogle, M.E. & Mortensen, U.H. A CRISPR-Cas9 System for Genetic Engineering of Filamentous Fungi. *Plos One* **10** (2015).

45. Lee, J.S., Grav, L.M., Lewis, N.E. & Kildegaard, H.F. CRISPR/Cas9-mediated genome engineering of CHO cell factories: Application and perspectives. *Biotechnol J* **10**, 979-994 (2015).

46. Horlbeck, M.A. et al. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *Elife* **5** (2016).

47. Maruyama, T. et al. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature biotechnology* **33**, 538-U260 (2015).

48. Madzak, C., Treton, B. & Blanchin-Roland, S. Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast Yarrowia lipolytica. *J Mol Microb Biotech* **2**, 207-216 (2000).

49. Blazeck, J., Liu, L.Q., Redden, H. & Alper, H. Tuning Gene Expression in Yarrowia lipolytica by a Hybrid Promoter Approach. *Appl Environ Microb* **77**, 7905-7914 (2011).

50. Magnan, C. et al. Sequence Assembly of Yarrowia lipolytica Strain W29/CLIB89 Shows Transposable Element Diversity. *Plos One* **11**, e0162363 (2016).

51. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research* **31**, 3406-3415 (2003).

52. Cherry, J.M. et al. Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic acids research* **40**, D700-D705 (2012).

SUPPLEMENTARY INFORMATION

Validating genome-wide CRISPR-Cas9 function in the non-conventional yeast *Yarrowia lipolytica*

Cory Schwartz[1], Jan-Fang Cheng[2], Robert Evans[2], Christopher A. Schwartz[3], James M. Wagner[4], Scott Anglin[5], Adam Beitz[5], Weihua Pan[6], Stefano Lonardi[6], Mark Blenner[5], Hal S. Alper[4,7], Yasuo Yoshikuni[2], and Ian Wheeldon[1]*

1. Chemical and Environmental Engineering, University of California Riverside, Riverside, CA 92521
2. DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, CA 94598
3. Department of Civil and Mechanical Engineering, United States Military Academy, West Point, NY 10996
4. McKetta Department of Chemical Engineering, The University of Texas at Austin, 200 E Dean Keeton St. Stop C0400, Austin, TX 78712
5. Chemical and Biomolecular Engineering, Clemson University, Clemson, SC 29634
6. Computer Science and Engineering, University of California Riverside, Riverside, CA 92521 USA
7. Institute for Cellular and Molecular Biology, The University of Texas at Austin, 2500 Speedway Avenue, Austin, TX 78712

*Correspondence: iwheeldon@engr.ucr.edu

Supplementary Figure 1. sgRNA design. Flowchart outlining sgRNA design strategy. On target efficiency calculated using previously constructed algorithm[1]. YALI1 is a recent assembly of the *Y. lipolytica* PO1f parent strain genome, W29[2].
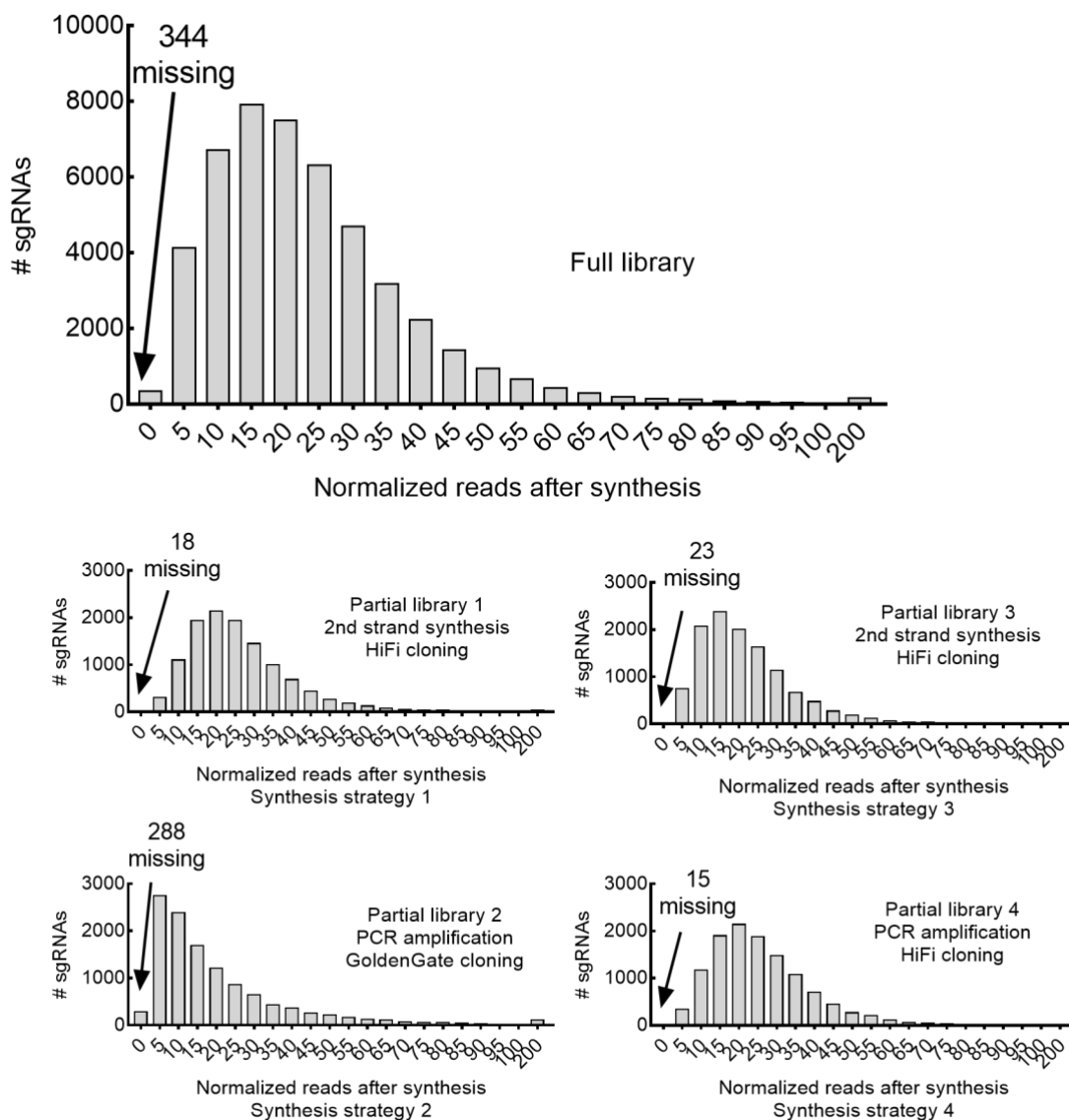
Supplementary Figure 2. Library cloning results. sgRNA abundance distribution in the full library and for each partial library, which were generated using 3 different cloning strategies
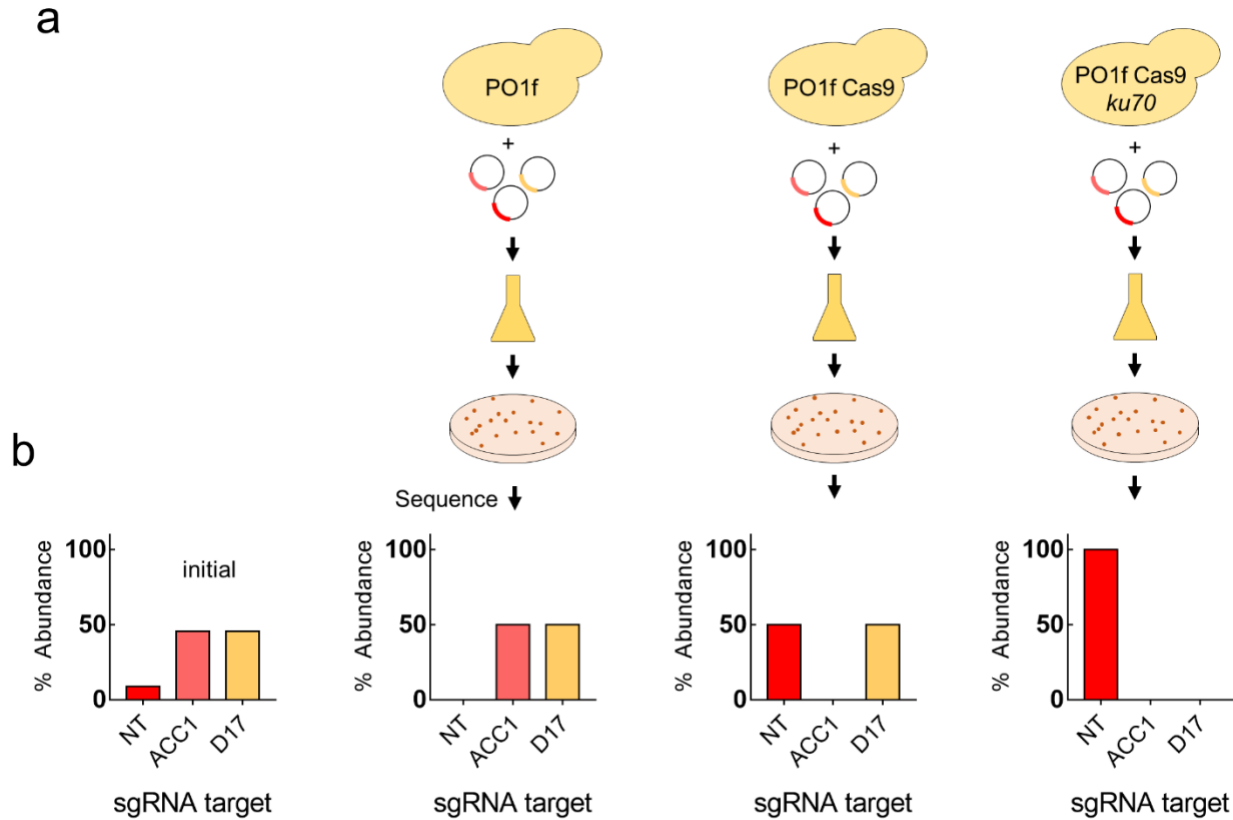
Supplementary Figure 3. Pilot scale cutting and fitness score experiments. (a) Small scale transformation of PO1f, PO1f Cas9, and PO1f Cas9 *ku70* with 3 different sgRNA expression plasmids. *ACC1* (light red) is an essential gene, *D17* (yellow) is a validated nonessential gene, and *NT* (red) represents a nontargeting sgRNA. (b) Percent abundance of sgRNAs before (initial) and after transformation into each strain.
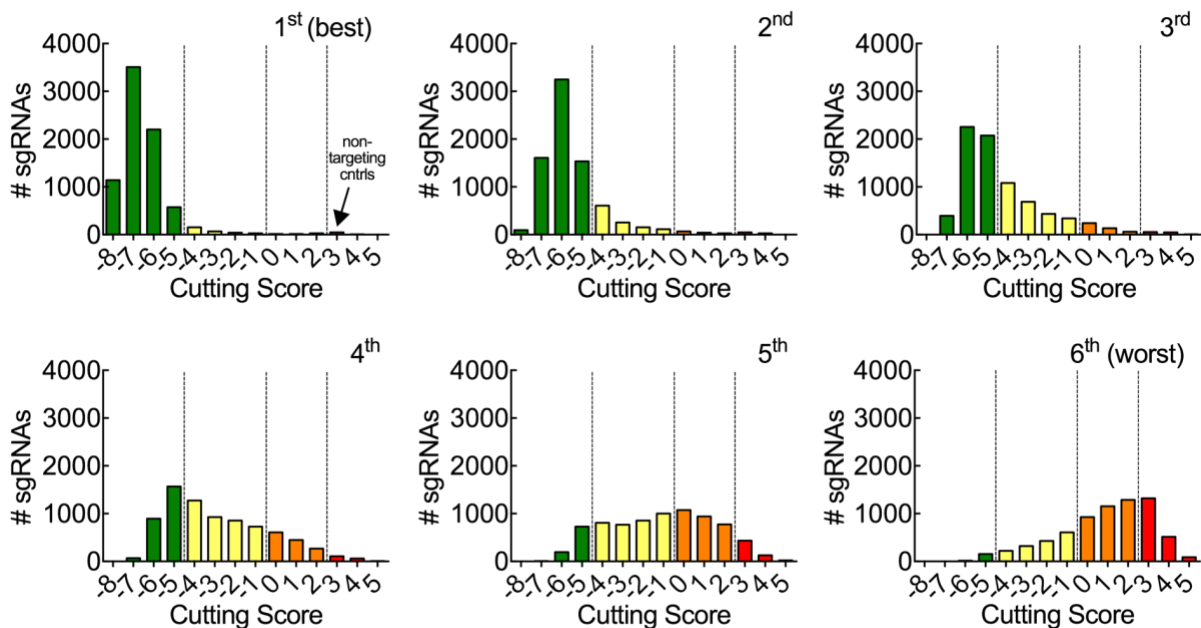
Supplementary Figure 4. Distribution of cutting scores across library. The sgRNAs for each gene were ranked by CS from 1st to 6th and separated into six subpopulations each representing 1-fold genome coverage and assigned to sgRNA classes (Excellent (CS<-5; green;), Moderate (-5<CS<-1; yellow), Poor (-1<CS< 2; orange), and Noncutter (>2; red)). Mean CS values were determined after two growth cycles (4 days) from triplicate biological samples.
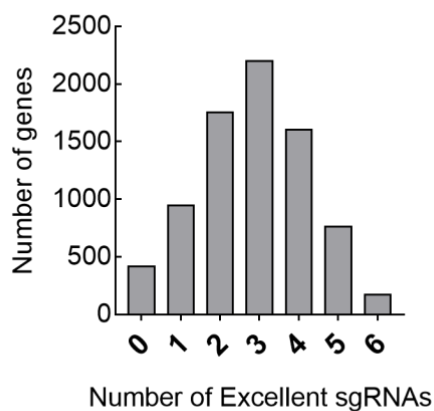

Supplementary Figure 5. Excellent sgRNAs per gene. Frequency of the number of excellent sgRNAs for each gene in the genome.
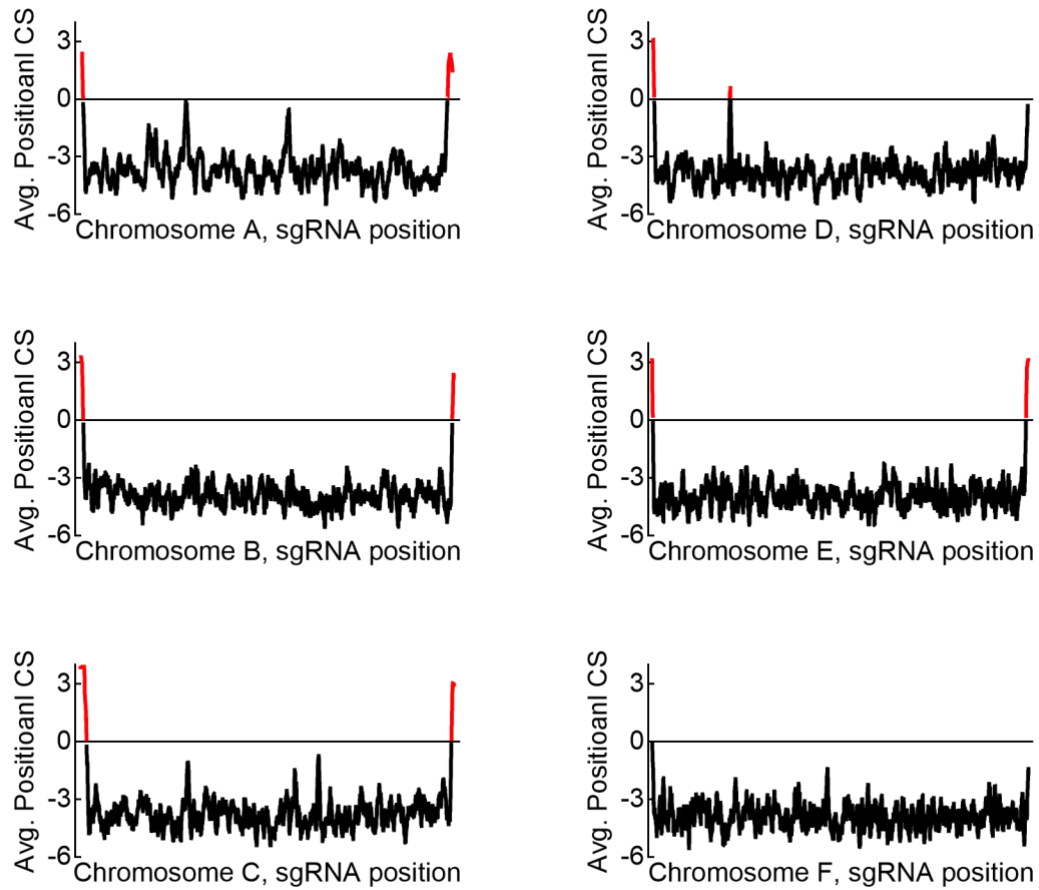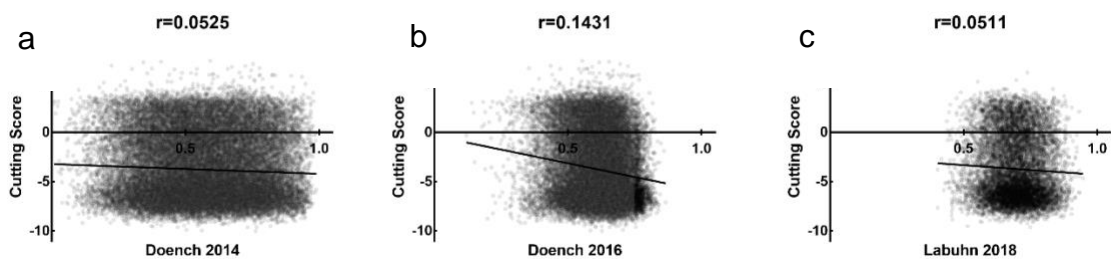
Supplementary Figure 6. Cutting score by chromosome position. Moving average of cutting score across chromosomes. Average CS is calculated by moving average of 50 sgRNAs, ordered by position in each chromosome.
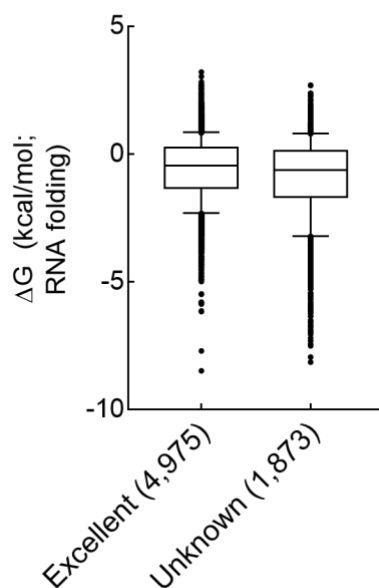
**Supplementary Figure 7. Secondary structure of excellent and noncutter sgRNAs**. Predicted secondary structure[3] for a representative population of excellent cutters and all noncutters that are not at chromosome ends or that contain a polyT motif.
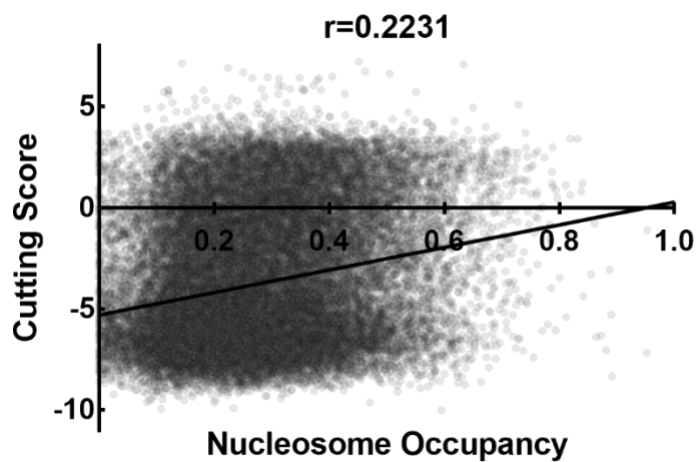
Supplementary Figure 8. Comparison of nucleosome occupancy to cutting score. Plot of cutting score for each sgRNA compared to average nucleosome occupancy.

Supplementary Figure 9. Comparison of sgRNA design algorithm score to cutting score. (a) sgRNA activity prediction from Doench et al. 2014[1] compared to cutting score. (b) sgRNA activity prediction from Doench et al. 2016[4] compared to cutting score. (c) sgRNA activity prediction from Labuhn et al. 2018[5] compared to cutting score.
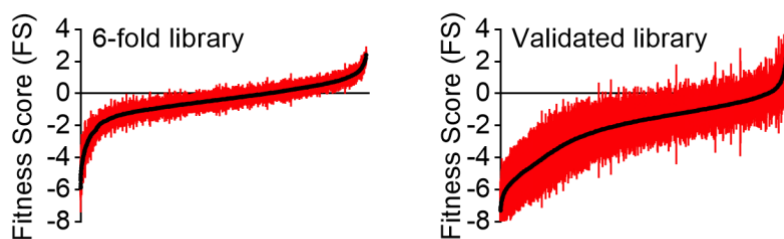
Supplementary Figure 10. Rank-ordered curves of fitness scores with error. The mean and standard deviation of three biological replicates is shown for the full 6-fold coverage library and for the validated library.
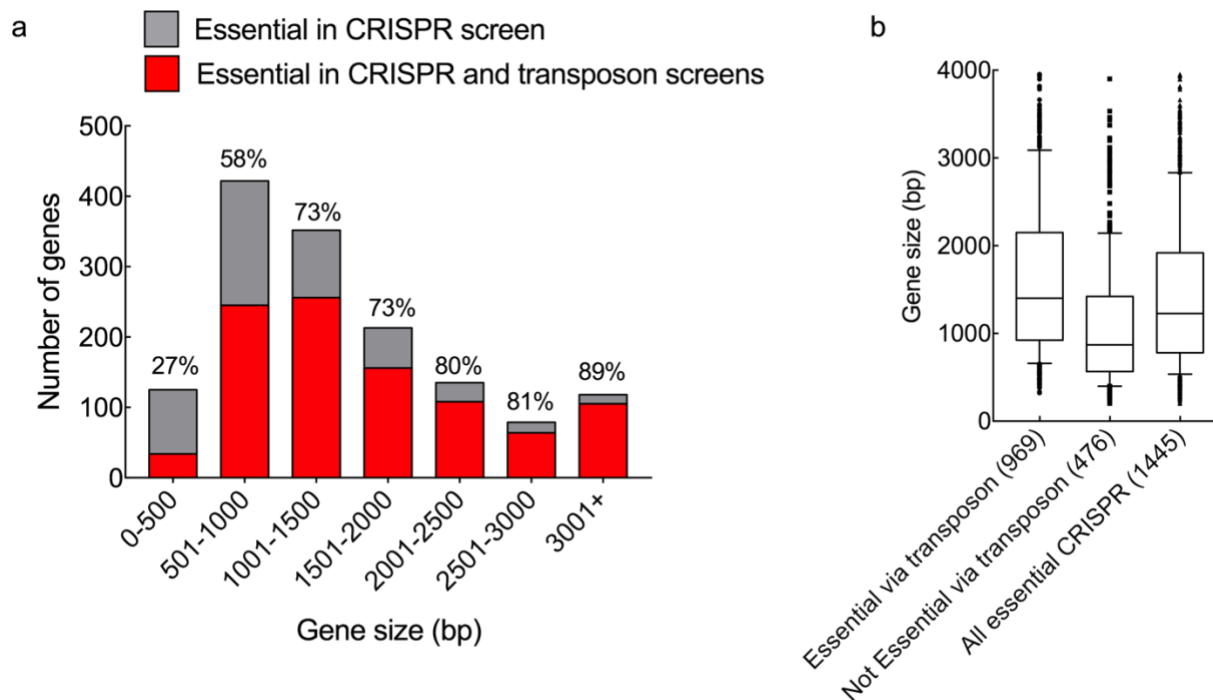
Supplementary Figure 11. Transposon and CRISPR essential gene comparison. (a) Comparison of essential genes identified in CRISPR and transposon screens. All CRISPR essential genes are shown, with the percentages representing the amount that were also identified as essential in the transposon screen. (b) Size distribution of CRISPR essential genes that were also identified as essential by the transposon screen and those not identified as essential in the transposon screen.
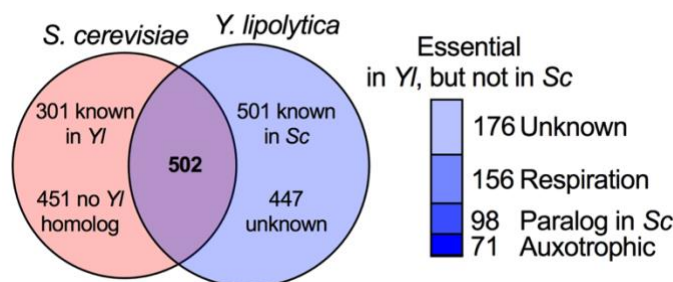
Supplementary Figure 12. Essential gene comparison to *Saccharomyces cerevisiae*. Comparison of essential genes identified in *Y. lipolytica* with known essential genes from *S. cerevisiae*. Respiration indicates genes annotated as being essential for respirative growth in *S. cerevisiae*, Paralog in Sc indicates genes that are duplicated in the *S. cerevisiae* genome, and Auxotroph indicates genes that are involved in nucleotide or amino acid biosynthesis (identified via manual inspection). Unknown or no *Yl* homolog indicate genes that did not have a clear homolog.
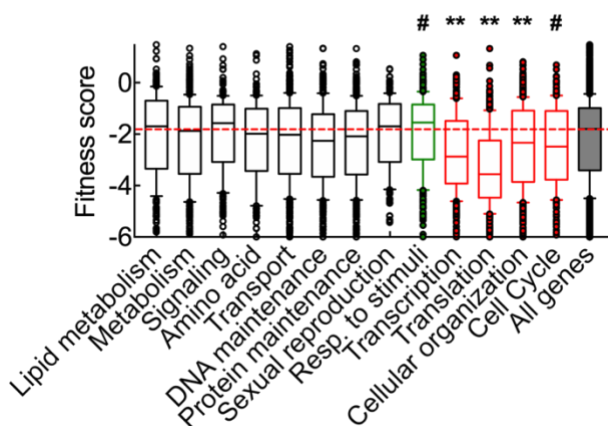
Supplementary Figure 13. GO term analysis and fitness score. Fitness score of groups of genes separated by gene ontology classification of genes by *S. cerevisiae* homolog. Significance was determined by 1-way ANOVA with post hoc analysis (# $p<0.05$; * $p<0.01$; **$p<0.0001$).

a
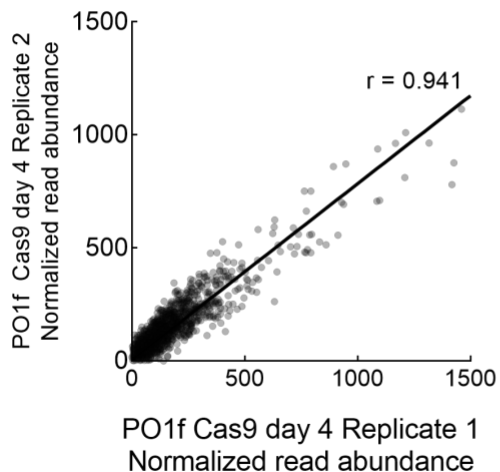
b

Strain	Time point	Comparison	Pearson
PO1f	Day 2	1 v. 2	0.988
		1 v. 3	0.982
		2 v. 3	0.980
	Day 4	1 v. 2	0.829
		1 v. 3	0.827
		2 v. 3	0.858
	Day 6	1 v. 2	0.818
		1 v. 3	0.829
		2 v. 3	0.855
PO1f Cas9	Day 2	1 v. 2	0.975
		1 v. 3	0.942
		2 v. 3	0.942
	Day 4	1 v. 2	0.941
		1 v. 3	0.958
		2 v. 3	0.964
	Day 6	1 v. 2	0.955
		1 v. 3	0.960
		2 v. 3	0.964
PO1f Cas9 ku70	Day 2	1 v. 2	0.972
		1 v. 3	0.976
		2 v. 3	0.972
	Day 4	1 v. 2	0.886
		1 v. 3	0.891
		2 v. 3	0.973
	Day 6	1 v. 2	0.877
		1 v. 3	0.875
		2 v. 3	0.968
Untransformed library		1 v. 2	0.976
		1 v. 3	0.977
		2 v. 3	0.985

Supplementary Figure 14. Pairwise comparison across biological replicates. (a) Pairwise comparison of PO1f Cas9 biological replicates 1 and 2. Linear regression analysis yielded a Pearson's coefficient of 0.941. (b) All possible pairwise comparisons of different replicates.

Supplementary Table 1: Selected essential and nonessential genes

		Putative Essential Genes	
Gene	Systematic name	Function	Ref.
ACC1	YALI1_C15991g	Acetyl-CoA carboxylase; lipid biosynthesis	unpublished
HMG1	YALI1_E05731g	HMG-CoA reductase; sterol biosynthesis	unpublished
ERG20	YALI1_E06759	Farnesyl pyrophosphate synthetase; sterol biosynthesis	[6]
FBA1	YALI1_E30886	Fructose 1,6-bisphosphate aldolase; glycolysis	[6]
KAR2	YALI1_E16731	ATPase; protein import to ER and protein folding mediator	[6]
DPB2	YALI1_F35907	Subunit B of DNA polymerase II	[6]
POL1	YALI1_E14644	DNA polymerase I catalytic subunit	[6]
PRO3	YALI1_B18955	Delta 1-pyrroline-5-carboxylate reductase; proline biosynthesis	[6]
FOL2	YALI1_A10254	GTP-cyclohydrolase I; folic acid biosynthesis	[6]
RPS5	YALI1_F11930	Protein component of 40S ribosomal subunit	[6]
MYO1	YALI1_F17836	Type II myosin heavy chain; cytokinesis and cell separation	[6]
ACT1	YALI1_D10600	Actin; structural protein and cytoskeletal functions	[6]

		Nonessential genes	Ref.
Gene name	Systematic name	Function	
POX2	YALI1_F14495	3-hydroxyacyl-CoA dehydrogenase and enoyl-CoA hydratase; peroxisomal beta-oxidation	[7]
GAL1	YALI1_C18258	Galactokinase; galactose catabolism	[8]
GAL7	YALI1_F31405	Galactose-1-phosphate uridyl transferase; galactose catabolism	[8]
POX5	YALI1_C32886	3-hydroxyacyl-CoA dehydrogenase and enoyl-CoA hydratase; peroxisomal beta-oxidation	[7]
BGL1	YALI1_F21504	beta-glucosidase; cellobiose catabolism	[8]
BGL2	YALI1_B18845	beta-glucosidase; cellobiose catabolism	[8]
KU70	YALI1_C11925	Subunit of Ku complex; nonhomologous end-joining	[9]
KU80	YALI1_E02574	Subunit of Ku complex; nonhomologous end-joining	[9]
POX3	YALI1_D32754	3-hydroxyacyl-CoA dehydrogenase and enoyl-CoA hydratase; peroxisomal beta-oxidation	[7]
POX4	YALI1_E32603	3-hydroxyacyl-CoA dehydrogenase and enoyl-CoA hydratase; peroxisomal beta-oxidation	[7]
ALK2	YALI1_F02132	n-alkane-inducible cytochrome P450; beta-oxidation	[10]
CAN1	YALI1_B25360	Arginine permease; knockout confers canavanine resistance	This work

33

Supplementary Table 2: Primers used in this study

Primer	Sequence	Use
Cr_1250	TATAAGAATCATTCAAAGGCGCGCATGGATAAGAAATACTCCATTGGCCTG	Cas9 integration cloning
Cr_1254	ATAACTAATTACATGAGGCTAGCTTACAGCATGTCCAGATCGAAATCG	Cas9 integration cloning
M13_F	GTAAAACGACGGCCAGT	RT-qPCR
M13_R	CACAGGAAACAGCTATGAC	RT-qPCR
Cr_1665	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTCCGGTTCGATTCCGGGTC	Forward primer Illumina
Cr_1666	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAGTCCGGTTCGATTCCGGGTC	Forward primer Illumina
Cr_1667	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTAGTCCGGTTCGATTCCGGGTC	Forward primer Illumina
Cr_1668	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTCCAGTAGTCCGGTTCGATTCCGGGTC	Forward primer Illumina
Cr_1669	CAAGCAGAAGACGGCATACGAGATTCGCCTTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Reverse primer Illumina
Cr_1670	CAAGCAGAAGACGGCATACGAGATATAGCGTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Reverse primer Illumina
Cr_1671	CAAGCAGAAGACGGCATACGAGATGAAGAAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Reverse primer Illumina
Cr_1672	CAAGCAGAAGACGGCATACGAGATATTCTAGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Reverse primer Illumina
Cr_1673	CAAGCAGAAGACGGCATACGAGATCGTTACCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Reverse primer Illumina
Cr_1709	CAAGCAGAAGACGGCATACGAGATGTCTGATGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Reverse primer Illumina
Cr_1710	CAAGCAGAAGACGGCATACGAGATTTACGCACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Reverse primer Illumina
Cr_1711	CAAGCAGAAGACGGCATACGAGATTTGAATAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Reverse primer Illumina
60mer_pool-F	GATTCCACCTGCCTGAACGT	Lib 2 cloning
spacer-AarI.rev	CGGCTCACCTGCAGTCAAAC	Lib 2 cloning
pLeu-mock-sgRNA.fwd	TAAAATCCATCGTTGCCATCGATGGGC	Lib 1, 3, 4 cloning
sgRNA-Rev2	CACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAA	Lib 1, 3, 4 cloning
Cr_1742	GCTCCGTCACGTGATATGTATTATC	Colony PCR and sequencing
Cr_1743	CTCGTGATACGCCTATTTTTATAGGTTAATG	Colony PCR and sequencing
Oligo pool 1	TAAAATCCATCGTTGCCATCGATGGGCCCCCGGTTCGATTCCGGGTCGGCGCAGGTTGACGTNNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG	Lib 1 sgRNAs
Oligo pool 2	GATTCCACCTGCCTGAACGTNNNNNNNNNNNNNNNNNNNNNGTTTGACTGCAGGTGAGCCG	Lib 2 sgRNAs
Oligo pool 3	TAAAATCCATCGTTGCCATCGATGGGCCCCCGGTTCGATTCCGGGTCGGCGCAGGTTGACGTNNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG	Lib 3 sgRNAs
Oligo pool 4	TAAAATCCATCGTTGCCATCGATGGGCCCCCGGTTCGATTCCGGGTCGGCGCAGGTTGACGTGACAGGCTCATCTCTAGAAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG	Lib 4 sgRNAs

References

1. Doench, J.G. et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature biotechnology* **32**, 1262-U1130 (2014).
2. Magnan, C. et al. Sequence Assembly of Yarrowia lipolytica Strain W29/CLIB89 Shows Transposable Element Diversity. *Plos One* **11**, e0162363 (2016).
3. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research* **31**, 3406-3415 (2003).
4. Doench, J.G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology* **34**, 184-+ (2016).
5. Labuhn, M. et al. Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications. *Nucleic acids research* **46**, 1375-1385 (2018).
6. Cherry, J.M. et al. Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic acids research* **40**, D700-D705 (2012).
7. Haddouche, R. et al. Roles of multiple acyl-CoA oxidases in the routing of carbon flow towards beta-oxidation and polyhydroxyalkanoate biosynthesis in Yarrowia lipolytica. *Fems Yeast Res* **10**, 917-927 (2010).
8. Schwartz, C., Curtis, N., Lobs, A.K. & Wheeldon, I. Multiplexed CRISPR Activation of Cryptic Sugar Metabolism Enables Yarrowia Lipolytica Growth on Cellobiose. *Biotechnol J*, e1700584 (2018).
9. Schwartz, C., Frogue, K., Ramesh, A., Misa, J. & Wheeldon, I. CRISPRi repression of nonhomologous end-joining for enhanced genome engineering via homologous recombination in Yarrowia lipolytica. *Biotechnol Bioeng* (2017).
10. Iida, T., Sumita, T., Ohta, A. & Takagi, M. The cytochrome P450ALK multigene family of an n-alkane-assimilating yeast, Yarrowia lipolytica: cloning and characterization of genes coding for new CYP52 family members. *Yeast* **16**, 1077-1087 (2000).