Title

Functional genomics for the oleaginous yeast Yarrowia lipolytica

Authors

Patterson, Kurt

Yu, James

Landberg, Jenny

et al.

Publication Date

2018-07-01

DOI

10.1016/j.ymben.2018.05.008

Metabolic Engineering 48 (2018) 184–196

Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/meteng

Functional genomics for the oleaginous yeast *Yarrowia lipolytica*

Kurt Patterson[a,1], James Yu[a,1], Jenny Landberg[a,2], Ivan Chang[a,3], Farbod Shavarebi[a], Virginia Bilanchone[a], Suzanne Sandmeyer[a,b,c,*]

[a] Department of Biological Chemistry, University of California, Irvine, Irvine, CA 92697-1700, USA
[b] Department of Chemical Engineering and Materials Science, University of California, Irvine, Irvine, CA 92697-1700, USA
[c] Institute for Genomics and Bioinformatics, University of California, Irvine, Irvine, CA 92697-1700, USA

ARTICLE INFO

Keywords:
Fitness
Functional genomics
Metabolism
Oleaginous yeast
Hermes
Transposon mutagenesis

ABSTRACT

Oleaginous yeasts are valuable systems for biosustainable production of hydrocarbon-based chemicals. *Yarrowia lipolytica* is one of the best characterized of these yeast with respect to genome annotation and flux analysis of metabolic processes. Nonetheless, progress is hampered by a dearth of genome-wide tools enabling functional genomics. In order to remedy this deficiency, we developed a library of *Y. lipolytica* insertion mutants via transposon mutagenesis. The Hermes DNA transposon was expressed to achieve saturation mutagenesis of the genome. Over 534,000 independent insertions were identified by next-generation sequencing. Poisson analysis of insertion density classified ~ 22% of genes as essential. As expected, most essential genes have homologs in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe,* and the majority of those are also essential. As an obligate aerobe, *Y. lipolytica* has significantly more respiration - related genes that are classified as essential than do *S. cerevisiae* and *S. pombe.* Contributions of non-essential genes to growth in glucose and glycerol carbon sources were assessed and used to evaluate two recent genome-scale models of *Y. lipolytica* metabolism. Fluorescence-activated cell sorting identified mutants in which lipid accumulation is increased. Our findings provide insights into biosynthetic pathways, compartmentalization of enzymes, and distinct functions of paralogs. This functional genomic analysis of the oleaginous yeast *Y. lipolytica* provides an important resource for modeling, bioengineering, and design of synthetic minimalized strains of respiratory yeasts.

1. Introduction

Oleaginous yeast provide a platform for biosustainable production of hydrocarbon-based compounds of industrial and biomedical interest (Beopoulos et al., 2011; Nicaud, 2012; Qiao et al., 2015; Yadav and Stephanopoulos, 2014; Zhu and Jackson, 2015). *Yarrowia lipolytica*, an obligate aerobic oleaginous yeast, thrives on diverse carbon sources ranging from plant oils to glycerol and acetate and can be engineered to accumulate up to 90% of its cell weight as lipid (Blazeck et al., 2014). Additional features of *Y. lipolytica* favoring biotechnology applications include Generally Regarded as Safe (GRAS) designation (Groenewald et al., 2014), reference genomes (Dujon et al., 2004; Magnan et al., 2016) and expression vectors. Nonetheless, despite metabolomic and transcriptomic studies, tools enabling functional genomics in this system are lacking. The impact of this deficiency is exacerbated by the fact that *Y. lipolytica* is phylogenetically distant from model respiro-fermentative yeasts, and so functions of many of its genes cannot be inferred from those of homologs (Dujon et al., 2004; Sherman et al., 2004).

Functional genomics of the model yeasts *Saccharomyces cerevisiae* (Giaever and Nislow, 2014) and *Schizosaccharomyces pombe* is largely based on high-throughput analysis of collections of targeted deletion knockout strains (Kim et al., 2010; Winzeler et al., 1999). However, despite their obvious value, construction of such collections is time and resource intensive. In addition, stringent genetic selection for specific knockout markers can promote enrichment for background suppressor mutations, resulting in a significant frequency of false non-essential classifications (Guo et al., 2013; Teng et al., 2013). These knockout collections also fail to survey unexpected features, such as intergenic transcripts and short open reading frames (ORFs) (Andrews and Rothnagel, 2014; Yoon and Brem, 2010). Transposon profiling provides an attractive alternative strategy for production of a mutant collection

* Corresponding author at: Department of Biological Chemistry, University of California, Irvine, Irvine, CA 92697-1700, USA
E-mail addresses: kurtp@uci.edu (K. Patterson), jamesy1@uci.edu (J. Yu), jenla@biosustain.dtu.dk (J. Landberg), ichang@jcvi.org (I. Chang), fshavare@uci.edu (F. Shavarebi), vwbilanc@uci.edu (V. Bilanchone), sbsandme@uci.edu (S. Sandmeyer).
[1] Equal contributors.
[2] Present address: Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens, Lyngby, Denmark.
[3] Present address: J. Craig Venter Institute, Capricorn Ln, La Jolla, CA 92037, USA.

https://doi.org/10.1016/j.ymben.2018.05.008
Received 17 March 2018; Received in revised form 16 May 2018; Accepted 16 May 2018
Available online 22 May 2018

appropriate for high-throughput and evolutionary analyses (Crook et al., 2016; Kumar and Snyder, 2001).

Approximately one quarter of *Y. lipolytica* genome features have no annotated function while others are assigned functions inferred from homology with genes in model yeasts and other fungi (Dujon et al., 2004; Magnan et al., 2016). Unlike *S. cerevisiae* and *S. pombe*, *Y. lipolytica* is oleaginous and an obligate aerobe. Even in cases in which *Y. lipolytica* genes have homologs in model yeasts, essential classification can help to differentiate paralogous gene functions. In this study, the Hermes Transposon (HTn), originally isolated from the housefly *Musca domestica* (Arensburger et al., 2011), was mobilized for saturation mutagenesis of the *Y. lipolytica* genome. A total of 534,589 independent Hermes insertion events were recovered from sequencing millions of isolated mutant colonies. The average number of independent events was calculated separately for both intergenic and intragenic regions based on sequencing data. Accounting for Hermes transposon insertion bias, an average of 1 hit every 24.37 bp was expected in intergenic regions, while an average of 1 hit every 97.5 bp was expected in intragenic regions. Insertional mutagenesis has previously been executed in *Y. lipolytica* to demonstrate its efficacy in rapid phenotype selection as well as introduction and excision of selection marker cassettes for facilitated strain engineering (Wagner et al., 2018). To further expand this tool for *Y. lipolytica*, we demonstrate the utility of this *Y. lipolytica* transposon library to classify essential and non-essential genes, provide serendipitous insights into gene function, identify non-coding upstream regulatory regions, evaluate genome-scale models, measure contributions of genes under different conditions, and enable reverse genetics by linking phenotypes to specific mutations. The library provides a means to carry out large-scale screens to select for desired phenotypes that can be linked to specific genes. Furthermore, this high-throughput screening approach is completely portable to other strains. Traditionally, metabolic engineering studies of lipogenesis in *Y. lipolytica* use genetic manipulation predicated upon combinations of known pathways and gene functions, alterations to growth conditions, as well as predictions based on genome scale models (Beopoulos et al., 2009). In a pilot experiment, we demonstrated the feasibility of using fluorescence-activated cell sorting (FACS) of the Hermes transposon library to isolate mutant strains with altered lipid metabolism on a genome-wide scale.

2. Materials and methods

2.1. Strains, plasmids, and growth conditions

Escherichia coli DH5α was grown at 37 °C in Lysogeny Broth (LB) medium supplemented with 100 μg/ml ampicillin to select for plasmid retention. *Y. lipolytica* wt strain CLIB89 was obtained from ATCC (20460). All yeast strains used were isogenic to CLIB89. To facilitate transposon mutagenesis and other investigations in this work, strains were modified and are listed in Table S1. Yeast were cultured at room temperature (22–25 °C) or at 28 °C as indicated. Cultures were grown in YPD [1% yeast extract (w/v), 2% peptone (w/v), and 2% dextrose (w/v)], YPG [1% yeast extract (w/v), 2% peptone (w/v), and 2% glycerol (w/v)], SD -Leu (nitrogen-enriched) medium [6.7 g/L yeast nitrogen base, 2% dextrose (w/v), inositol, adenine sulfate, and complete amino acids lacking leucine], or nitrogen-depleted medium [0.625 g/L $(NH_4)_2SO_4$, 5 g/L Na_2SO_4, 2% dextrose (w/v), 1.7 g/L yeast nitrogen base without amino acids and $(NH_4)_2SO_4$, and 35 mg/L uracil] (Guthrie and Fink, 1991).

Y. lipolytica strains, plasmids and primers used in this study are summarized in Table S1. A *URA3*-marked CEN/ARS18 donor plasmid (pJY3919) was constructed for Hermes transposon mutagenesis. This donor plasmid contains a selectable cassette that is mobilized by codon-optimized Hermes transposase expressed under the *TEF1* promoter (Muller et al., 1998; Tai and Stephanopoulos, 2013) from the same plasmid. The selectable donor cassette consists of *LEU2* (beta-isopropyl malate dehydrogenase) flanked by the Hermes terminal inverted repeats (TIRs). Two negative control plasmids were constructed based on the same vector. One contains TIR-flanked *LEU2* and lacks the transposase (pPS3911), while the other contains the transposase and lacks TIRs on either side of the *LEU2* marker (pMT3928) (Table S1).

Yeast transformations were performed using a modified lithium acetate transformation protocol (Chen et al., 1997; Gietz and Schiestl, 2007). Strains were inoculated into a 3 ml culture of YPD, grown overnight at 28 °C, diluted 1:50 into fresh YPD and harvested at 1.0–1.5 OD_{600}. Cell pellets were resuspended in 1× TE/LiOAc (10 mM Tris, 1 mM EDTA, 100 mM lithium acetate) at a ratio of 250 μL per 0.5 OD_{600}. Transformation reactions contained 300 μL of 1× TE/LiOAc/DMSO/PEG40 (TE/LiOAc plus 10% dimethyl sulfoxide, and 40% PEG 3350), 500–600 ng of plasmid DNA, and 50 μL of cell resuspension. Transformations were incubated at 30 °C with rotation for 3 h, heat shocked at 39 °C for 30 min, pelleted and resuspended in 200 μL water and plated onto the appropriate selection medium.

2.2. Hermes transposition in Y. lipolytica

For Hermes transpositions, uracil and leucine double auxotroph strains were transformed with pJY3919 or negative control plasmids pPS3911 and pMT3928. Test experiments indicated that 54.7% ± 6.7% of pJY3919 transformants had transposition events while pPS3911 or pMT3928 transformants had a negligible frequency of transposition. Analysis of *LEU2* copy number showed that only one in 28 tested strains had more than a single copy of *LEU2* (Fig. S1). For genome profiling, yJY1953 (Trial 1) and yJY2006 (Trial 2), were transformed with pJY3919. The transformation protocol was followed as described and cells were selected for pJY3919 on SD-Leu medium in 150 mm plates. After two days of growth, colonies were collected and then replated onto SD-Leu medium supplemented with 5-fluoroorotic acid (5FOA) (Boeke et al., 1987) in 150 mm plates and incubated at room temperature for another two days to allow colony formation. Room temperature was chosen as a relatively permissive condition, recognizing that growth at the temperature of maximum growth rate (28–30 °C) would likely cause more genes to be classified as essential. Approximately 2.2 million colonies were harvested and re-plated onto SD –Leu +5FOA medium to select for cells containing the *LEU2*-tagged transposon that no longer contained the *URA3*-tagged pJY3919 backbone. 5FOA-resistant colonies were pooled to represent the Gen0 population. For competitive growth experiments in YPD or YPG, 4 × 100 ml replicate cultures were inoculated with Gen0 cells with a starting OD_{600} = 0.05. For passaging (every 12 h of growth at 23 °C), cultures were re-inoculated to OD_{600} = 0.05 in fresh medium. At each passage the number of generations was estimated based on the OD_{600}, and at roughly 20 and 80 generations, 100 OD_{600} cell pellet from each replicate was harvested for genomic DNA isolation and processing. For simplicity these are referred to as Gen20 and Gen80.

2.3. Sequencing library preparation

Genomic DNA (gDNA) was isolated by the glass-bead method as described (Amberg et al., 2005). After extraction, gDNA was treated with DNase-free RNaseA (100 μg/ml) (Sigma-Aldrich) at 37 °C for 4 h, followed by Pronase (250 μg/ml) (Sigma-Aldrich) at 37 °C for 2 h. After both RNase and Pronase treatment, another phenol:chloroform extraction was performed. Approximately 30 μg of this treated DNA was sheared to less than 1 kb in a 100 μL volume using the Biorupter Standard (Diagenode). DNA size was confirmed by gel electrophoresis. Sheared DNA was concentrated using the Clean and Concentrator with 5× ChIP DNA Binding Buffer (Zymo Research). DNA ends were made blunt using the Fast DNA End Repair kit (Thermo Scientific) and 3′ adenylated using the Klenow Exo- enzyme with 0.2 mM dATP (Fisher Scientific). Custom sequencing adapters (0.5 μM) with Illumina multiplex barcodes (Table S1) were ligated to DNA in each sample using T4 ligase (Thermo Scientific) by incubating at 22 °C for 12 h followed by

65 °C for 10 min. Between enzyme reactions, samples were purified using the Agencourt AMPure XP purification beads (Beckman Coulter, Inc.). Hermes TIR–gDNA borders were enriched in PCR using the high-fidelity KOD polymerase (EMD Millipore) as follows [95 °C, 5 min; 5 cycles of 95 °C, 30 s; 65 °C, 30 s; and 70 °C, 30 s; followed by 20 cycles of 95 °C, 30 s; 61 °C, 30 s; and 70 °C, 30 s; and a final 70 °C incubation for 2 min]. Forward primer KP5113 anneals 26 bp from the downstream 3'-end of the TIR, and reverse primer KP5112 anneals to the adapter upstream of the Illumina multiplex barcode. The same PCR primers were used for all samples. To avoid PCR bias, each library sample was split into 16 technical replicates, amplified, and then recombined and size selected using the AMPure XP Beads. Each PCR reaction contained 25 ng of template DNA. After sizing and purification of PCR products, samples were submitted to the UCI Genomics High-throughput Facility for sequencing on the Illumina HiSeq. 2500.

2.4. Sequencing data analysis

Briefly, reads were first trimmed to remove the random 5-nt sequence, then sequences downstream of KP5113 primer binding site not matching the expected terminal 26 nts of the Hermes TIR were filtered out as they represent mis-primed reads. DNA sequences passing this filter were trimmed to remove the 26 nt TIR sequence and mapped to the CLIB89 YALI1 genome (Magnan et al., 2016) using Bowtie2 (Langmead and Salzberg, 2012). Each of the 534,589 unique reads mapped defined a Hermes insertion site; non-unique reads mapping to multiple positions, such as those representing insertions into tDNAs and transposons, were randomly assigned (less than 1.5% of reads). All raw Hermes insertion site positions and nearest annotated genes are provided in Table S2. Hermes insertion coordinates were mapped to gene annotations using BEDTools (Quinlan and Hall, 2010). Essential gene classification was done using custom Python v2.7 scripts, which are available upon request. Insertion site motif analysis was done using WebLogo (Crooks et al., 2004; Schneider and Stephens, 1990). Phenotype data from *S. cerevisiae* for comparison to *Y. lipolytica* was acquired from the *Saccharomyces* Genome Database (SGD) (Cherry, 2016). *Y. lipolytica* homologs to *S. cerevisiae* genes were acquired from the CLIB89 genome annotation (Magnan et al., 2016). For simulated transposition, custom scripts were used to scan the CLIB89 genome for all 5′nTnnnnAn3′ or 5′nAnnnnTn3′ sites, and these sites were mapped to the annotated CLIB89 genome using BEDTools. All graphs were prepared in RStudio.

2.5. Essential gene classification

Mutants were analyzed for genes with a statistically significant underrepresentation of Hermes insertions relative to an expected number of hits based on gene length (L). These genes were considered to be ones that were not recovered by sequencing due to detrimental effects of the Hermes insertion on growth. Similar to previous system-level analyses (Gerdes et al., 2003), our gene classification system (GCS) uses the Poisson distribution to test the probability of observing k insertions assuming a rate of 97.5 bp/hit for genes (λ) and a rate of 24.37 bp/hit for intergenic regions (λ_i). When the probability (P) of witnessing k insertions in a gene of length L was below 0.05, the gene was classified as essential.

$$P_k(L) = \frac{[\left(\frac{L}{\lambda}\right)*R]^k}{k!} * \quad e^{-[\left(\frac{L}{\lambda}\right)*R]}$$

Local features flanking a gene such as heterochromatin or gene density can impact the hit frequency; thus for the local hit density a 5′ and 3′ 1-kb flanking region (L_f) was used as a metric for determining the local correction ratio (R). R is defined as the ratio of observed local hit frequency over the expected. Observed local hits includes the sum gene hits (H_{gene}) and each 1-kb flank (H_{flank}); whereas expected hits includes the sum of expected gene hits $\left(\frac{L}{\lambda}\right)$ and flanking hits $\left(\frac{L_f}{\lambda_i}\right)$. Thus R is defined as

$$R = \frac{(H_{gene} + H_{flank})}{\left[\left(\frac{L}{\lambda}\right) + \left(\frac{L_f}{\lambda_i}\right)\right]}$$

When $R < 1$, it was multiplied by the expected gene hit value in the Poisson equation above; however, when $R \geq 1$ no correction was applied (i.e. $R = 1$).

Only unique insertions were included when calculating the GCS. Two classes of hits were excluded from the GCS calculation. In the first class, hits in several genes were restricted to introns, consistent with tolerance for intronic, but not exonic, disruption. Tolerance could be related to insertions simply being spliced out, as it is known that *Y. lipolytica* splicing can remove relatively large introns compared to other fungi (Mekouar et al., 2010). Alternatively, because fungal introns are at the extreme 5′ ends of genes, it is possible that effects of intronic insertions are alleviated by secondary downstream transcription or translation start sites. These interpretations are consistent with studies of DNA transposition of the related hAT family transposon piggyBac in *D. melanogaster* that showed that transposition into introns can modify, but not necessarily eliminate gene expression (Hacker et al., 2003). Intronic hits were excluded from the GCS Poisson analysis by subtracting them from H_{gene}, however they were included when calculating the genome-wide rate of gene hits (λ). In the second class, some genes were represented solely by hits in the extreme downstream end. In previous integration profiling, including Tn3 disruption profiling of *S. cerevisiae* (Ross-Macdonald et al., 1999) and Tf1 profiling of *S. pombe* (Guo et al., 2013), 3′-terminal hits were associated with expression of truncated proteins that retained significant activity. In order to avoid false negative findings based on lower phenotypic penetrance of 3′ hits, insertions restricted to the downstream 10% of the ORF were subtracted from H_{gene} when they accounted for ≥ 50% of the total H_{gene}. Genes that were sufficiently short such that expected number of hits was 0 or 1 and the insertion probability was greater than the threshold ($P < 0.05$) were classified as low-confidence essential (LC-essential). Finally, longer genes occasionally demonstrated an expected vs. observed hit margin sufficient to produce a $P < 0.05$, despite ample hits. Gen0 genes with hit density greater than 3 hits/kb/million reads (HKM) but with $P < 0.05$ were classified as non-essential.

Gen0, Gen20 or Gen80 samples grown and passaged in either YPD or YPG medium were analyzed as above except inter- and intra–genic hit frequencies (λ) were recalculated for each Gen80 population. After 80 generations, genes could remain non-essential, be reclassified as growth impaired (essential for survival in growth in glucose and glycerol), Gly-conditional (essential for survival in glycerol only) or Glu-conditional (essential for survival in glucose only). After the initial Poisson evaluation, logical refining of gene classifications was performed as described above for Gen0 to ensure accuracy. For example, low-flanking hit densities could allow a gene with 1 hit to have a P value > 0.05 and cause the gene to score non-essential. Refining takes into account additional parameters such as gene hit density (HKM) and the overall median of gene hits for essential and non-essential genes to flag possible miscalls. In some cases the density of gene hits at Gen20 were also considered.

2.6. Gene disruption

Y. lipolytica genes (*GUT2*, *KU70, MAE3, UTR1*) were disrupted by replacement of the ORF by homologous recombination (HR) using a LoxP-*URA3*-LoxP cassette flanked by sequences 5′ and 3′ of the ORF. Mutagenized transformants were colony purified and mutants were verified by PCR amplification and DNA sequencing. Genes (*ADE2*, *GPD1*, *HIS3*, *PEX10*, *SNF1*) were also disrupted by directed mutagenesis using the CRISPR-Cas9 protocol for *Y. lipolytica* (Schwartz et al., 2016).

For CRISPR-Cas9, primer sequences were incorporated into *Avr*II-restricted pCRISPRyL by Gibson assembly (Gibson, 2009) and plasmids used are listed (Table S1). Predicted essential genes *SNF1* and *HIS3* were disrupted using CRISPR-Cas9 as described above, except in the presence of a *URA3*-marked rescue plasmid containing a wild-type copy of the gene harboring a silent mutation that eliminated the CRISPR guide RNA PAM site.

2.7. Gene ontology analysis

GO information was obtained using the Blast2GO software (www.blast2go.com). Using the CLIB89 reference genome (Magnan et al., 2016), enriched GO terms were determined using the Fisher Exact test ($P < 0.05$) and false discovery rate (FDR) correction. For comparison to *Y. lipolytica*, *S. cerevisiae* essential genes (phenotype:inviable) were collected from the *Saccharomyces* Genome Database (SGD) (Cherry, 2016) and *S. pombe* essential genes (FYPO:0002061) were taken from PomBase (2016). GO enrichment for *S. cerevisiae* (sacCer3 reference genome) and *S. pombe* (ASM294v2 reference genome) was also done using Blast2GO. The Venn diagrams were prepared using venn.js software (https://github.com/benfred/venn.js).

2.8. Z-score calculations for mutant representation

Reads per insertion mutation ("hit") were used as a proxy for the representation of mutants in cultures. In this analysis, hits with fewer than 10 reads, within introns, or in the last 10% of the gene were filtered out unless they were the sole hit within a gene (raw reads are reported Table S2). Reads per hit in Gen0, Gen20Glu, Gen80Glu, Gen20Gly and Gen80Gly samples were determined, normalized by dividing by total reads per individual culture, log2 transformed, and used to determine the average normalized read count and standard deviation for each culture. These were used to determine Z-scores for each hit and these were averaged for the hits in each gene to give a final Z-score average per gene. These values are reported in Table S3.

2.9. Fluorescence-activated cell sorting

A frozen aliquot of Gen0 library cells was thawed, pelleted, resuspended at an $OD_{600} = \sim 0.1$ in SD -Leu (nitrogen-enriched) and grown in flasks with shaking at 200 rpm at 23 °C overnight. Cells were inoculated to $OD_{600} = 0.05$ and grown to a final OD = ~3.5. For isolation of a lipid-enriched population by cell sorting, cells were stained with the lipophilic fluorescent dye Bodipy [BODIPY® 493/503 (4,4-Difluoro-1,3,5,7,8-Pentamethyl-4-Bora-3a,4a-Diaza-*s*-Indacene, Molecular Probes]. Bodipy was dissolved in DMSO at a concentration of 5 mM and stored at − 80 °C until use. One OD_{600} of cells was stained in 1× PBS containing 8 μM Bodipy for 15 min in the dark at room temperature. This procedure was scaled up as needed. Stained cells were washed and resuspended in nitrogen-enriched medium at a concentration of 1×10^6/ml. Samples were kept on ice and in the dark. Cells were sorted using a JSAN cell sorter (Bay Biosciences, Japan) at a rate ~200 events/s. Bodipy was excited with a 488 nm laser, and fluorescent emission was captured using a 530/30 nm bandpass filter. Cells were fractionated into SD-Leu medium based on intensity of Bodipy staining (1% highly stained and 99% remaining). Non-stained cells were used as autofluorescence control. Collected cells from both the 1% and 99% stained-cell populations were plated onto YPD. For an initial screen, individual colonies from both fractions were analyzed by Bodipy fluorescence for lipid content. Cells were grown in 96-well plates to $OD_{600} = \sim 4–6$. Cells were fixed by addition of 1/9 volume of formaldehyde, incubated at room temperature with shaking for 20 min, washed with 1× PBS and stored at 4 °C until analyzed. Cells were stained with Bodipy as described above and fluorescence intensity was determined in 96-well plate format by flow cytometry using a NovoCyte flow cytometer (Acea Biosciences, Inc.). Fluorescence data were analyzed using FlowJo V10 software and expressed as the normalized geometric mean fluorescence intensity (MFI). Wilcoxon Rank Sum Test was used to determine statistical significance between MFI of collected fractions ($P < 0.05$).

2.10. Sequencing of insertion sites

Transposon insertion locations in mutants identified by Bodipy staining were determined from genomic DNA by inverse PCR (Sambrook and Russell, 2006). GDNAs were digested with *Bcu*1, *Nhe*1, *Xba*1 and *Xma*J1 or *Bam*HI, *Bgl*II and *Bcl*1 to generate compatible cohesive ends. Digested gDNAs were circularized by ligation and the transposon-genome junction sequence was amplified by PCR (Table S1). The PCR fragments were cloned into PCR4-Blunt TOPO vector (Life Technologies), the transposon-gDNA junctions were sequenced using primer VB5510, and sequences were mapped to the YALI1 (CLIB89) genome (Magnan et al., 2016).

2.11. Lipid determination

Lipid content of individual mutants isolated from the 1% lipid-rich fraction by cell sorting was determined and compared to the lipid content of the control strain (yJY1948) isogenic to yJY2006. To maximize lipid production, cell cultures were inoculated to an OD_{600} of 0.5 in 50 ml nitrogen-depleted medium in flasks and grown with shaking at 200 rpm at 28 °C for 120 h. A minimum of 4–5 biological replicates were analyzed for each strain. Cell pellets (~25 OD_{600} units) were collected, centrifuged and the supernatant removed and stored at − 80 °C. A duplicate sample was collected, desiccated to complete dryness and weighed to determine the cell dry weight. Fatty acids were extracted from ~25 OD_{600} frozen cell pellets to which 5 μL of internal standard was added (nonadecanoic acid; stock solution 0.015 g/ml) in the presence of 500 μL 10% barium hydroxide and 550 μL of 1,4-dioxane at 110 °C for 18–24 h. Samples were acidified with 6 M HCl to pH less than 4. Fatty acids were recovered in hexane and concentrated under a stream of N_2 to near dryness. Samples were derivatized by addition of 1 ml of 1 N HCl in methanol and heated at 80 °C for 30 min. Samples were cooled, and 1 ml 0.9% NaCl and 1 ml hexane was added, and vortexed for 5 min. Samples were centrifuged at $2000\times g$ for 10 min at 25 °C, the hexane layer removed and evaporated under N_2 stream to a final volume of ~1 ml. Analysis of derivatized fatty acids was performed by gas chromatography-flame ionization detector (GC-FID) on an Agilent GC 7890A (Agilent Technologies, Santa Clara, CA) equipped with a HP5MSI column (length 30 m, 0.250 mm ID, 0.25 μm film thickness). The oven temperature started at 80 °C, then 20 °C/min to 210 °C, hold 1 min, and 10 °C/min to 280 °C hold 1 min, 20 °C/min up to 320 °C hold for 3 min. The detector–FID was at 250 °C, H_2 flow 30 ml/min, Air 400 ml/min and helium as makeup gas at 25 ml/min. Peaks were identified using F.A.M.E. C8-C24 mix (Sigma-Aldrich). Wilcoxon Rank Sum Test was used to determine statistical significance between isolates.

2.12. ROC analysis of genome scale models

Two genome-scale models iNL895 (Loira et al., 2012) and iYali4 (Kerkhoven et al., 2016) were obtained from BioModels Database (2016) as MODEL11111900000 and MODEL1508190002 respectively. The models were loaded in Cobra Toolbox (Becker et al., 2007) for MATLAB, which was used to iteratively perform flux balance analysis (FBA) simulation on each model so that changes in predicted biomass production after a single gene deletion could be recorded and analyzed downstream. When predicted biomass productions are lower than a set threshold (50% of max biomass production), the corresponding deleted genes are suggested to be essential for fitness, while non-essential otherwise. The biomass predictions were then compared with the genes from Hermes transposon mutagenesis essential/non-essential

classification (899 genes in iNL895 and 901 genes in iYali4). For ROC analysis, Hermes-derived gene classifications were used as classifiers while biomass (arbitrary units) were used as predictors. Area under the curve (AUC) values with 99% confidence intervals were used to assess performance of each model. ROC analysis was done in RStudio using the pROC package (Robin et al., 2011).

3. Results and discussion

3.1. Activation of the Hermes transposon in Y. lipolytica

The Hermes transposon system was adapted to *Y. lipolytica* to enable high-throughput transposition and mutagenesis (Section 2). Our results demonstrated efficient mobilization of the Hermes *LEU2* marker in cells that did not incorporate the donor plasmid backbone (Fig. S1a) and that most mutant isolates tested had a single copy of *LEU2* (Fig. S1b). Analysis of transposon insertion target sites showed that 53% of 534,589 unique insertions conformed to the canonical insertion site motif 5′nTnnnnAn3′ (Evertts et al., 2007; Gangadharan et al., 2010; Guimond et al., 2003; Guo et al., 2013; Li et al., 2013) (Fig. S2a–c) and insertions into intergenic regions were 1.6-fold overrepresented (Fig. S2d). This observation contrasted with a relatively even distribution of consensus sequences in the genome (Fig. S2e). Similar results were observed for Hermes insertions in *S. cerevisiae* and *S. pombe* (Alexander et al., 2010; Wood et al., 2002). Thus, this system has the properties desired for high-throughput mutagenesis of *Y. lipolytica* and linking insertions to phenotypes and gene function. Our transposon profiling workflow diagram is shown in Fig. 1.

3.2. Classification of essential and non-essential genes by transposon profiling

In a relatively under-explored yeast such as *Y. lipolytica*, lists of essential/non-essential genes (Table S3) streamlines informed genome engineering of host strains, provides insights into metabolic differences with model yeasts, and defines the conditions under which genes are essential to survival. To score essentiality and incorporate adjustments for these additional inputs, we developed a Poisson-based Gene Classification Strategy (GCS) similar to that done for previous system-level analyses (Gerdes et al., 2003). This strategy calculates the probability (P) that Hermes insertions (“hits”) within a gene are significantly underrepresented (Section 2).

Overall, out of 8710 genomic features (Magnan et al., 2016), 1963 (22.5%) were classified as essential, 5907 non-essential (67.8%), and 813 (9.3%) low-confidence essential (LC essential) (Fig. 2a, Table S3). Hermes insertion profiles over specific genes and flanking sequence (Fig. 2b) illustrate patterns typical of essential, non-essential, and LC-essential genes. Essential genes have few to no hits and flanking regions populated with hits. Conversely, non-essential genes have hits in genes as well as flanking regions. LC- essential genes are those with relatively few or no insertions but which did not pass the threshold of significance because of gene size or regional deficit of insertions. Two short LC-essential genes flank *KU70* and illustrate the complications of classifying very short genes.

3.3. Validation of gene classification method

We evaluated the essential and non-essential classifications of *Y. lipolytica* gene features in several ways. In the first approach, we compared overall results of this classification to those for model yeasts based on systematic knockout collections. The fraction of essential genes (22.5%) in *Y. lipolytica,* was overall comparable to the 20% of ORFs for *S. cerevisiae* (Giaever et al., 2002) and 26% ORFs for *S. pombe* (Kim et al., 2010) called essential.

In a second approach, intragenic hit density or hits/kb/million sample hits (HKM) was plotted for all genomic features. This plot showed a bimodal distribution of intragenic HKM. Essential genes accounted for a significant fraction of very low insertion densities and non-essential genes accounted for a broad continuum of higher insertion densities (Fig. S3a, left panel). A display of intergenic hit density shows a generally unimodal distribution spanning essential and non-essential genes (Fig. S3a, right panel). A bimodal distribution of hit density for essential and non-essential genes was previously observed for Hermes saturation insertion profiling of the *S. pombe* genome, which was validated by comparison to a knockout collection (Guo et al., 2013; Kim et al., 2010). Scatter plots of expected versus observed hits in individual genes show the range of insertions within non-essential and essential gene groups (Fig. S3b).

In a third approach, our classification analysis was tested by engineering disruptions of genes we classified as non-essential and essential. These genes included five non-essential (*GUT2, KU70*, *MAE1*, *PEX10*, and *UTR1*) and three essential (*GPD1, HIS3*, and *SNF1*) genes. As expected, disruption of the non-essential gene set was efficiently accomplished under the conditions used for the initial classification and therefore validated our results (Table S3, Table S4). We classified *GPD1* as essential when cells were grown on glucose medium. *GPD1* encodes glycerol phosphate dehydrogenase. In *Y. lipolytica*, it is essential for growth in glucose as a carbon source, but non-essential on medium containing glycerol as a carbon source (Yuzbasheva et al., 2017). Cells disrupted for *GPD1* were viable when grown on glycerol but failed to grow on glucose.

Classified essential genes *HIS3* and *SNF1* were disrupted using CRISPR-Cas9 in the presence of a *URA3*-marked rescue plasmid containing a wild-type copy of the gene harboring a silent mutation that eliminated the CRISPR guide PAM site (Materials and Methods, Fig. S4a). Cells lacking the rescue plasmid failed to grow, confirming the essential classification under our experimental conditions (Fig. S4b,c). Further inspection of the data showed that all seven genes for histidine biosynthesis (*HIS1–7*) from phosphoribosyl pyrophosphate to L-histidine are essential. This result was surprising since *HIS3* was previously reported to be disrupted in *Y. lipolytica* (Kretzschmar et al., 2013).

SNF1 is a master regulator of lipid biosynthesis and disruption mutants are reported to display pleiotropic phenotypes, including elevated lipid production in *Y. lipolytica* (Kretzschmar et al., 2013; Seip et al., 2013). However, we found that mutants encoding a severely truncated gene did not survive in the absence of the *SNF1* rescue plasmid under our growth conditions on glucose medium (Fig. S4c). Others have also noted difficulties in reconciling lipid accumulation and flux analysis properties of different *snf1*Δ mutants (Kerkhoven et al., 2016). A likely explanation for the essential classification of *SNF1* and *HIS* biosynthetic pathway genes in comparison to other studies is a significant contribution to viability from unknown genetic factors in our strain background.

Taken together, our results validated the transposon profiling classification for a specific set of genes. In addition, validation enabled unexpected demonstration of nutritional requirements for *Y. lipolytica* strain CLIB89 and showed that *SNF1* and also *HIS1-7* genes of the histidine biosynthetic pathway are essential in this strain background under the growth conditions used in our experiments.

3.4. Essential genes in model yeasts are also essential in Y. lipolytica

Essential genes from *S. cerevisiae* (Dowell et al., 2010; Giaever et al., 2002; Kastenmayer et al., 2006) and *S. pombe* (Kim et al., 2010) that have *Y. lipolytica* homologs were assessed for essentiality in *Y. lipolytica.* Of the total number of homologous essential genes in *S. cerevisiae,* 73.4% were also essential in *Y. lipolytica,* while 9.8% and 16.9% were classified as LC essential and non-essential, respectively. Of the total number of homologous essential genes in *S. pombe,* 69.5% were also essential in *Y. lipolytica,* while 8.3% and 22.2% were classified as LC essential and non-essential, respectively (Fig. S5a). To further examine essential genes by product function, all of the essential genes from *Y.*

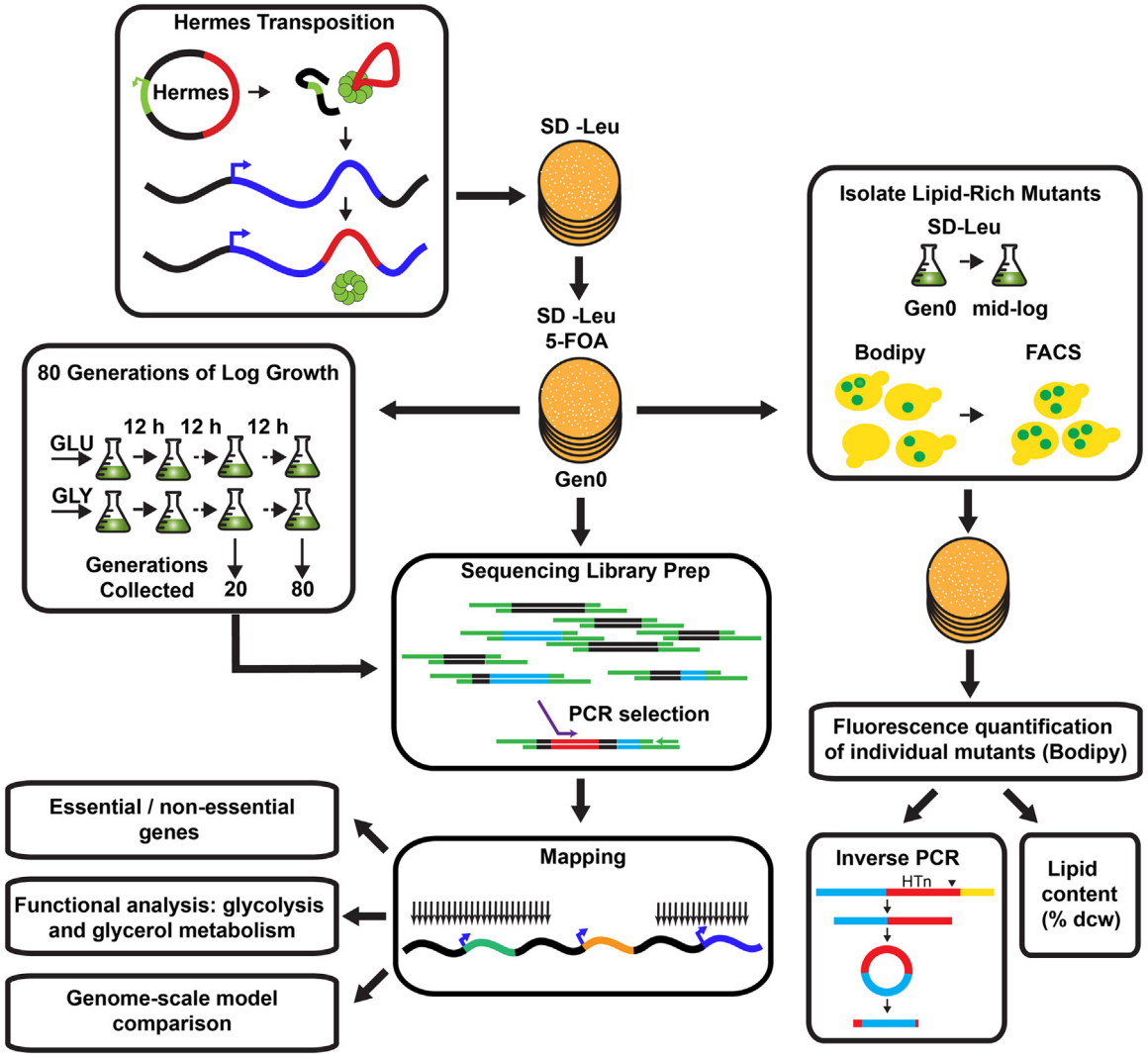

Fig. 1. Transposon profiling workflow. From top to bottom. Cells were transformed with plasmid (pJY3919) expressing Hermes transposase in *trans* to *LEU2*-tagged Hermes cassette and transformants were collected on SD–Leu medium; colonies were replated to select against donor plasmids and for transposed cells (Gen0). Cells were harvested and regrown ~ 80 generations in glucose (Glu) or glycerol (Gly). Cells were sampled at Gen0, Gen20, and Gen80 to identify the distribution of insertions, and features were classified according to whether they had expected levels of insertions (non-essential) or significantly fewer (essential). Other experiments included functional analysis of genes involved in glycolysis and glycerol metabolism, including contributions of genes to competitive growth and comparisons of genome scale models. To isolate lipid-rich mutants, the Bodipy-stained mutant library was sorted by fluorescence intensity as a measure of lipid content into the top 1% and bottom 99% fractions. Individual mutants were recovered and rescreened. Selected mutants were analyzed for lipid content and the location of the Hermes insertion was determined by inverse PCR (Section 2).

lipolytica, *S. cerevisiae* and *S. pombe* were also analyzed for GO term enrichment and compared (Table S6, Fig. S5b). Additionally, BLAST2GO analysis was performed for *Y. lipolytica* alone to identify GO categories enriched for essential genes (Fig. S5c, Table S5). For further details, refer to Fig. S5.

3.5. Application: Competitive growth and conditional essentiality

Classification of mutants isolated from colonies provided a relaxed context in which cells grew or died virtually irrespective of doubling time and the ability to compete for nutrients in medium containing glucose as the sole carbon source. In order to capture relative contributions of genes to long term batch culture, pooled mutants were passaged during logarithmic growth in rich medium with 2% glucose (YPD) and sampled at Gen20 and Gen80. To test for conditional essentiality based on carbon source, the experiment was also carried out in rich medium containing 2% glycerol (YPG). In addition to sequencing data collected for analysis of gene essentiality (Gen0), cultures were sampled and Hermes insertion sites were sequenced at Gen20 and Gen80 in order to profile the mutant population over time. Poisson-based GCS analysis was performed (Materials and Methods). Outgrowth of cultures in YPD or YPG reduced diversity of mutant populations (Fig. S6b, Table S7). Despite a significant reduction in population diversity for YPD and YPG cultures, inter- and intra- genic insertions were lost from cultures at an approximately equal rate. In the Gen0 population, 79.6% of hits were intergenic and after 80 generations 80.7% and 81.5% of remaining hits were intergenic for YPD and YPG, respectively. This suggests that hits in both inter- and intra-genic regions could reduce competitive fitness presumably by disrupting gene function. Essential genes lost more upstream insertions within 1-kb upstream of the ORF than did non-essential genes in both YPD and YPG. Mutant loss was more similar between these conditions for essential genes ($r = 0.84$) than for non-essential genes ($r = 0.69$) (Fig. S7c, d), suggesting that loss of mutations in promoters of essential genes was influenced to a greater degree by competitive growth.

Genes that were lost in both YPD and YPG outgrowth cultures by

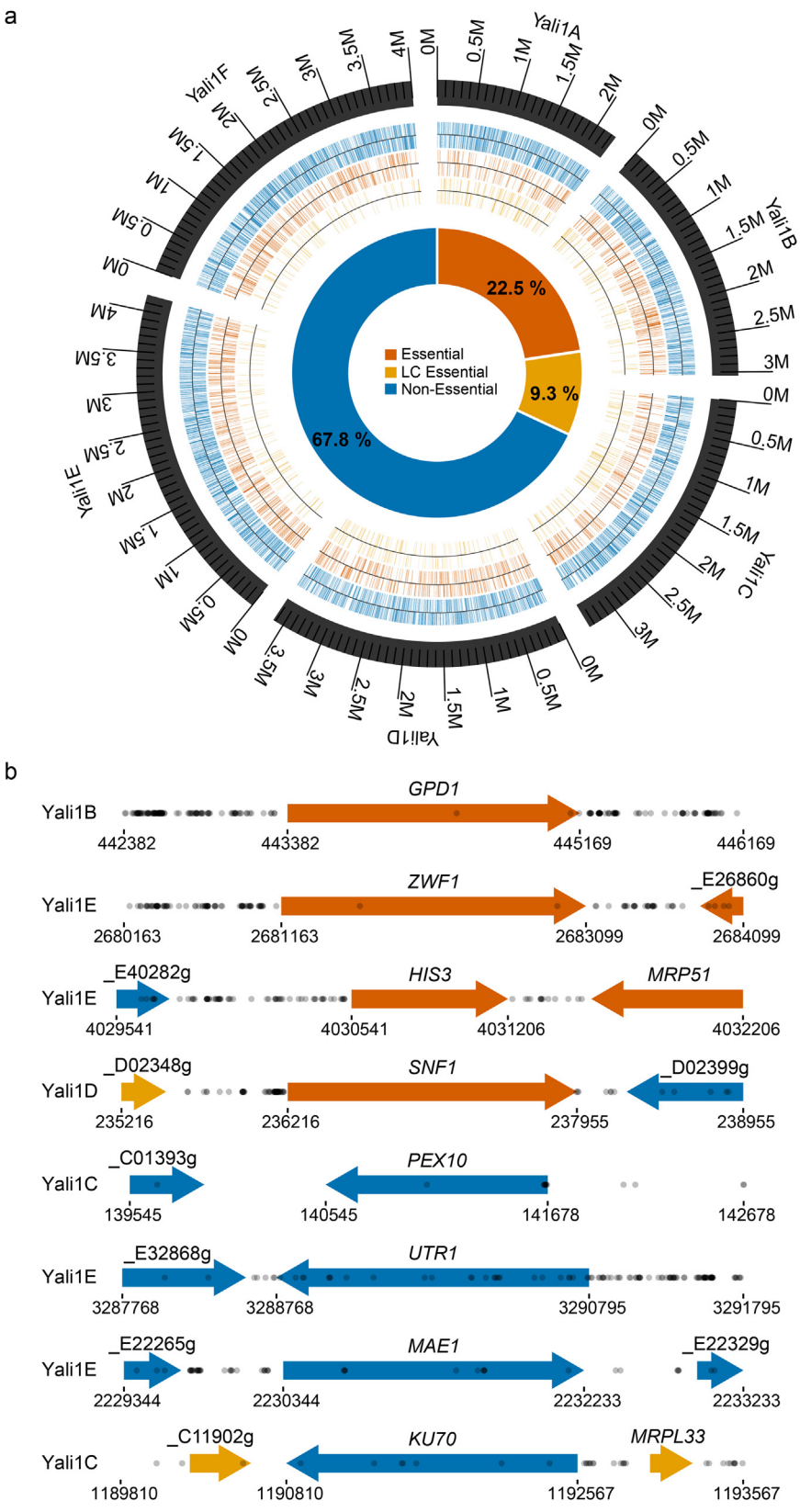

Fig. 2. Essential and non-essential genes classified by Hermes DNA transposon profiling. (a) Circos diagram showing positions of all non-essential genes (genes that met the threshold of expected numbers of insertions; blue), essential genes (genes with few or no insertions; orange) and low-confidence essential genes (LC essential; genes with relatively few or no insertions but which did not pass the threshold of significance because of size or regional deficit of insertions; gold). Chromosome size and name are indicated; gene strandedness is indicated by positive (outward) and negative (inward) lines. Central pie chart shows the distribution of classes. (b) Representative loci of essential, non-essential and LC-essential genes. Genes are shown with 1 kb of up- and down-stream flanking sequence. Hermes insertion site positions are indicated as grey dots, with regional hit density reflected in the darkness of overlapping dots. Gene names are represented as common names or YALI1 (CLIB89) gene codes positioned above each gene. Arrows indicate strandedness. Gene colors correspond to (a). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

Gen80 were classified as growth impaired (8.2%, 718 genes). GO analysis showed that 610 had associated GO terms and of these, 12 were enriched in Gen0 essential while 8 were unique to 17 growth-impaired categories. (Fig. S6a, Table S8). Compared to Gen0 essential GO-enriched genes, Gen80 growth-impaired GO-enriched genes were related to Pol III transcription, protein phosphorylation, nuclear transport, RNA modification, DNA recombination and repair, RNA modification, and peptidyl-amino acid modification.

The terms “Glu- and Gly-conditional” describe genes that were classified as non-essential in Gen0, but were classified as essential after

growth in either YPD or YPG, but not both. By Gen80, 17.5% of genes (1527) were classified as essential, 8.2% of genes were classified LC essential, and 1.4% of genes were classified as Glu-conditional (0.4%) or Gly-conditional (1%) (Fig. S6a). The greater number of genes in the Gly-conditional compared to the Glu-conditional class probably reflects the original selection in glucose as the carbon source, but may also indicate that additional genes are required for growth in glycerol.

3.6. Application: Contributions of genes to relative fitness

Applying GCS to liquid cultures over time revealed that the mutant population diversity was reduced more dramatically in YPG than in YPD (Fig. S6b), showing that glycerol conditions conferred a more competitive growth environment than glucose conditions. To further investigate the contributions of individual genes to fitness in competitive liquid culture, the change in the number of sequence reads per insertion mutation ("hit") was examined for each gene over time. A change in the number of reads per gene indicates a change in the contribution to fitness, and this change relative to all other genes was used to calculate relative fitness for each gene (Giaever et al., 2002; van Opijnen et al., 2009; Wiser and Lenski, 2015). We normalized the number of reads per hit in Gen0, Gen20, and Gen80 glucose and glycerol samples, generated a Z-score for each hit, and averaged all Z-scores within a gene to derive a final average Z-score reflecting the representation of that gene in each culture (Table S3, Materials and Methods). Analysis of the relative contribution of genes to fitness during competitive growth in YPD and YPG is consistent with the disappearance of mutants in the growth impaired and conditional classes of genes in Glu and Gly cultures, and also displays finer distinctions in representation (Fig. S8). For example, the *GUT1* gene scored as essential for growth on glycerol but not glucose media. In YPD, its fitness score increased from − 0.12 at Gen0 to − 0.08 by Gen80 but decreased to − 0.57 in YPG. The reduced fitness for *GUT1* mutants is consistent with scoring it as glycerol-conditional at Gen80. In YPD, 23 of the original 35 ORF insertions were still detected, while in YPG only 2 were identified at Gen80. Fitness scores corroborate the fact that glycerol conditions constitute a more competitive environment for mutants, but provide a more nuanced view of how reduced population diversity is accompanied by clonal expansion of competent subpopulations. This analysis and other inferences from individual mutations (Table S2, Table S3, Table S9) can be directly applied to strategies for bioengineering of this yeast for growth in liquid cultures.

3.7. Application: Comparative and functional analyses of homologs implicated in glycerol metabolism

Relatively few genes were identified as Glu- or Gly-conditional. At Gen80, 88 genes were classified as Gly-conditional and 34 as Glu-conditional (Table S10). Because *Y. lipolytica* is an obligate aerobe, it was expected that genes required for respiration would score as essential or growth impaired, but that there would be some additional genes specifically required for glycerol metabolism. This contrasts with *S. cerevisiae* for which respiration is not essential except for growth on 3-carbon or 2-carbon substrates that do not sustain fermentation. Therefore, it seemed likely that genes required for growth on glycerol in common between *S. cerevisiae* (Qian et al., 2012) and *Y. lipolytica* would include ones relatively specific for glycerol metabolism or gluconeogenesis. Comparison of these gene sets identified 9 homologs required for glycerol metabolism as well as genes related to a variety of other functions, and also 20 genes encoding proteins of unknown function (Table S10).

GUT1 (YALI1_F00654g) encoding glycerol kinase, converts glycerol to glycerol-3-phosphate, which is in turn converted to dihydroxyacetone phosphate (DHAP) by mitochondrial glycerol-3-phosphate dehydrogenase encoded by *GUT2* (YALI1_B18499g). DHAP can then be isomerized into glyceraldyde-3-phosphate to enter glycolysis (Fig. 3). *GUT1* was essential for growth in YPG (Gly-conditional) and *GUT2* was essential for growth in both YPD and YPG medium (growth impaired). The Gly-conditional phenotype of *GUT1* and growth-impaired phenotype of *GUT2* is consistent with behavior of independently verified gene disruptions in this work (Table S4), and elsewhere in *Y. lipolytica* (Beopoulos et al., 2008) and *S. cerevisiae* (Sprague and Cronan, 1977). This implies that the Gut1-Gut2 cytoplasmic-mitochondrial pathway operates for growth on glycerol in *Y. lipolytica*.

In many organisms, glycerol can also be metabolized in the cytoplasm by oxidation to dihydroxyacetone (DHA) by glycerol dehydrogenase encoded by a member of the broad aldo-keto reductase family of enzymes. In *S. cerevisiae*, this gene is designated *GCY1*. DHA is phosphorylated to DHAP by dihydroxyacetone kinases encoded by *DAK1* and *DAK2*. The Blastp in *Y. lipolytica* identified 12 candidate genes that encode an aldo-keto reductase, annotated in Fig. 3 as YPR (Yeast Possible Reductases). One of these YPR genes (YALI1_B20108g) was Gly-conditional, which suggests that it may be the aldo-keto reductase primarily associated with the oxidation of glycerol in this alternative pathway. In a similar manner, we identified three proteins with similarity to *DAK1* through BLASTp analysis (Magnan et al., 2016). However, only YALI1_F02508g was classified as essential, indicating it performs a non-redundant function. Since DHA accumulation is toxic in *S. cerevisiae* (Molin et al., 2003), the cytoplasmic pathway via *DAK1* might also be required despite the alternative Gut1-Gut2 pathway for glycerol metabolism.

3.8. Application: Variable penetrance of TCA cycle mutations

Because *Y. lipolytica* is an obligate aerobe, genes comprising the TCA cycle are expected to be essential. However, it was striking that mutants with insertions in genes that encoded components of the tricarboxylic acid cycle (TCA) cycle survived past Gen0 to be classified as growth impaired (Fig. 3, Table S11). Overall we identified 21 genes encoding functions associated with the TCA cycle. Of these, 16 are essential and 4 are growth impaired (Fig. 3). One possible explanation of the delay in the requirement for *MDH1, CIT1, LSC1, and LSC2* from Gen0 to Gen80 is that the mitochondrial proteins they encode are long lived and sufficient for Gen0 colony development (Guo et al., 2013).

3.9. Application: Identification of strains with altered lipid metabolism

The complexity of the Hermes library enables screening for desirable phenotypes over a wide range of genotypes. We carried out a preliminary study to investigate the feasibility of screening the transposon mutant library for genes involved in altered lipid metabolism. Flow cytometry screening indicated that the proportion of lipid-rich cells in the mutant library is ~5-fold greater than that of the control strain when grown in nitrogen-enriched medium, while no fold difference was observed under nitrogen-depleted conditions (Fig. S9). To identify mutants with altered lipid regulation in nitrogen-enriched growth conditions, Gen0 cells were sustained in logarithmic growth in nitrogen-enriched medium, stained with Bodipy and sorted by FACS. Cells were sorted by differential staining into two groups, the 1% most intensely stained cells and the remaining 99% fraction (Fig. 1, Materials and Methods). Retest of individual strains from the 1% and 99% fractions by cell cytometry under nitrogen-enriched conditions showed that the 1% fraction had significantly higher Bodipy fluorescence than the 99% fraction. This indicates the effectiveness of the method in isolating potential mutants of interest (Fig. 4a, Materials and Methods). Mutants with higher Bodipy staining in nitrogen-enriched conditions may have alterations in genes regulating pathways related to lipid biogenesis, and also have enhanced total lipid content when grown under nitrogen-depleted conditions. To test this hypothesis, several of the mutants that were previously selected by elevated Bodipy fluorescence were grown in flask cultures under conditions expected to cause maximal lipid content (nitrogen-depleted medium, 120 h at 28 °C) and the total lipid

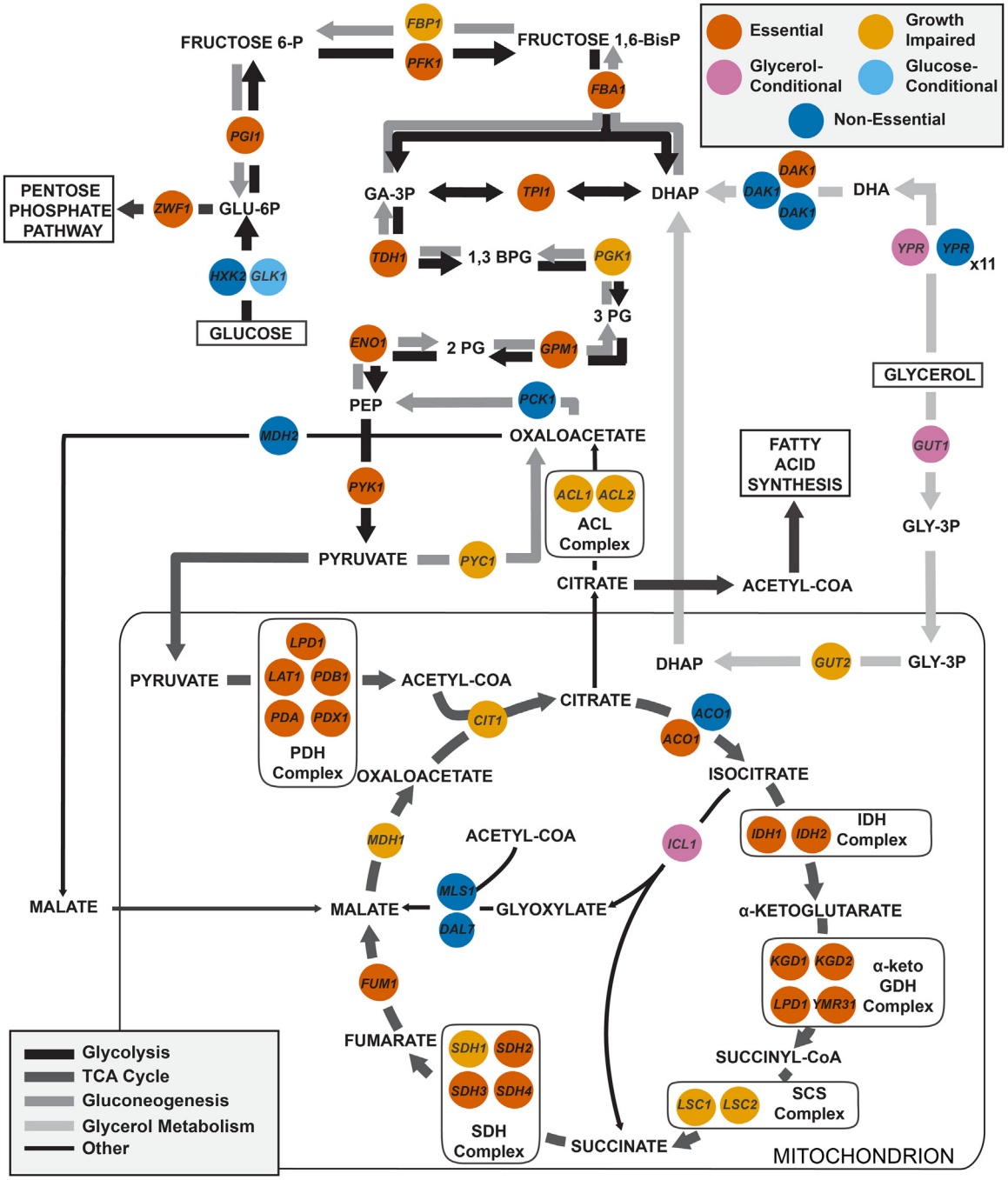

Fig. 3. Functional analysis of metabolic genes in glycolysis and glycerol metabolism. Each circle represents a gene and is color coded by its classification after 80 generations of log phase growth. Genes were classified as essential (orange), non-essential (blue), growth impaired (gold), Gly-conditional (reddish purple) or Glu-conditional (sky blue). Abbreviations are as follows: GLY-3p, glycerol-3-phosphate; DHA, dihydroxyacetone; DHAP, dihydroxyacetone phosphate; GLU-6P, glucose-6-phosphate; GA-3P, glyceraldehyde-3-phosphate; 1,3 BPG, 1,3-bisphosphoglycerate; 3PG, 3 phosphoglycerate; 2PG, 2-phosphoglycerate, PEP, phosphoenol pyruvate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

content was determined (Fig. 4b, Materials and Methods). Strain yFS2044 had significantly higher mean total lipid content [(26.3% lipid per dry cell weight (dcw))] and was 1.7-fold greater ($P = 0.03$) than the control (yJY1948) that is isogenic to CLIB89. The other strains tested (JY2042, yFS2043, yFS2045, yFS2046) appeared to have higher mean total lipid content, but values were not significant due to variability in the biological replicates under these experimental conditions. To investigate the nature of the genome disruption in selected strains, the location of Hermes transposon junctions with genomic DNA was identified and mapped to the CLIB89 YALI1 genome (Magnan et al., 2016) (Fig. 4c, Section 2). The transposon insertions in all strains analyzed were located in intergenic regions, consistent with positional bias previously observed (Fig. S3a). In all but one case, the insertion was between two divergently transcribed ORFs such that the expression of one or both genes could be affected. In strain yFS2044, insertion was 380-bp upstream of YALI1_F11261g, encoding a protein of unknown function. In strain yJY2042, the insertion was located 551-bp upstream of *PRB1*, a vacuolar protease. Interestingly, Prb1 is induced in nitrogen-limiting conditions (Zhu et al., 2012). In strains yFS2045 and yFS2046, the total lipid content was not significantly different from control. The results of this small-scale study demonstrate the potential of expanded investigation using the transposon mutant library in conjunction with

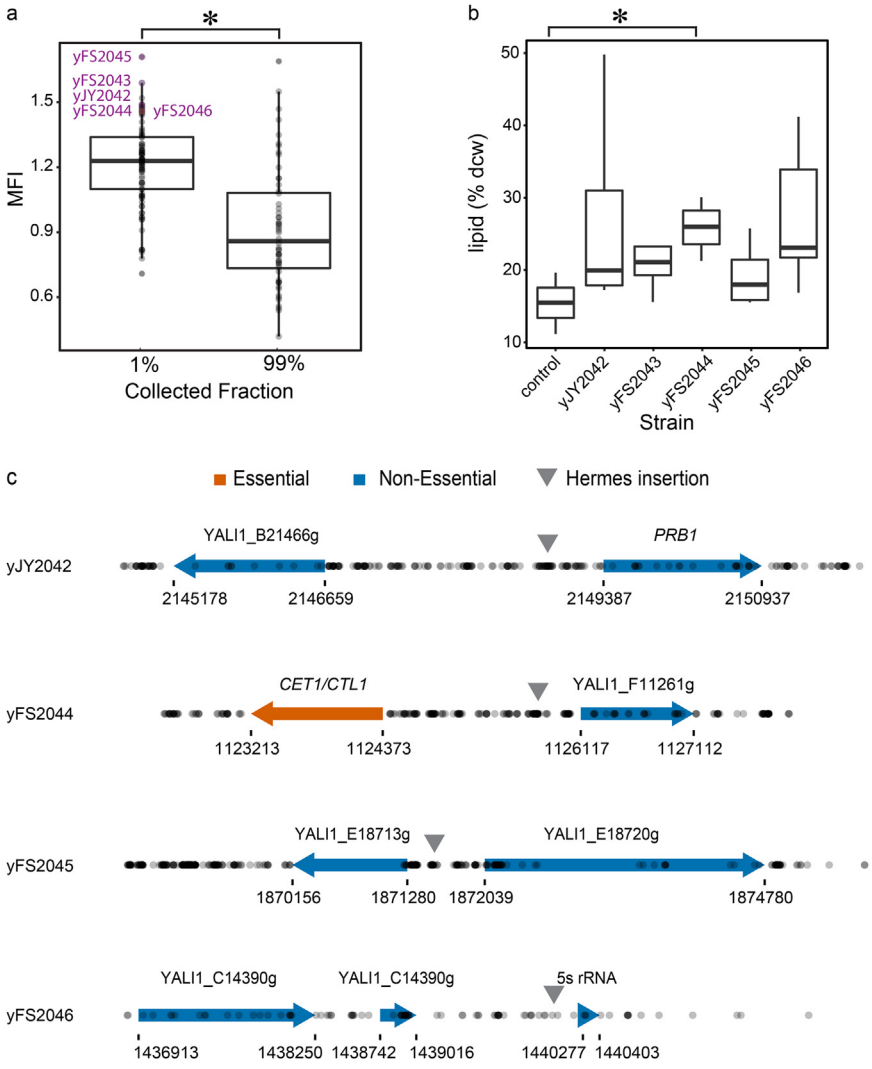

Fig. 4. Analysis of lipid-rich mutants isolated from Bodipy-stained mutant library using cell sorting. (a) Retesting of individual Bodipy-stained mutant cells from the original FACS-collected fractions (top 1% and remaining 99%) using NovoCyte cytometer. Boxplot of relative lipid content of individual Bodipy-stained mutants measured as normalized geometric mean fluorescence (MFI) in comparison to the control strain. Cells were grown in nitrogen-replete medium. Mutant strains that were subjected to further analysis are labeled. Asterisks (*) indicates significant difference in MFI, $P = 3.539 \times 10^{-10}$. (b) Lipid content [% lipid per dry cell weight (dcw)] of selected mutants grown in nitrogen-depleted medium for 120 h. Replicates n = 4 or 5. Asterisks (*) indicates significant difference, P = 0.03. (c) Position of Hermes insertion in selected mutant strains. Dots indicate the hit density from the Gen0 library. Grey triangles show insertion position in individual strains tested. Gene color: essential (orange), and non-essential (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Bodipy-staining and cell sorting to identify genes involved in different aspects of lipid metabolism.

3.10. Application: Functional genomics with in silico network modeling

Transposon profiling-based functional genomics is most useful if integrated with other genome-wide information for understanding biological processes. Genome-scale metabolic models (GEMs) are reconstructions of metabolic pathways at the systems level (Becker et al., 2007). Such models serve as frameworks for integrating multiple levels of heterogeneous data and so provide mechanistic insights into the relationship between genotype and phenotype. GEMs are constructed by linking genome annotations to reaction databases such as KEGG, thereby establishing a reaction network corresponding to known and hypothetical central metabolic genes of an organism (Kanehisa and Goto, 2000). GEMs can be combined with constraint-based linear/non-linear programming algorithms to predict the biomass growth rate of the organism given a particular growth condition. The standard way to measure the accuracy of and train GEMs is to compile a library of gene knockout phenotypes from experimental results, generate a "truth table" of true positives, false positives, true negatives, and false negatives, and compare with the GEMs' *in silico* predicted phenotypes. Hermes transposon profiling provided whole-genome information of knockout phenotypes such that all modeled genes and the reactions to which they map can be tested to evaluate and train the models (Fig. 5a, Table S12). We chose two GEMs for *Y. lipolytica*, iNL895 (Loira et al., 2012) and iYALI4 (Kerkhoven et al., 2016) and tested their classification predictions of a subset of genes (899 and 901 genes, respectively) against the classification of essential and non-essential genes determined by GCS. We sought to compare the overall performance of the two models. In addition, we tested for systematic weaknesses in the models by identifying GO groups that were enriched for genes where the models failed to correctly distinguish inviable and viable phenotypes. The iYALI4 and iNL895 models converged at about 80% true positives and 40% true negatives, meaning that they underestimated the fraction of essential genes. Specific comparison of the models to the transposon classification showed that the iNL895 model and the iYALI4 model similarly predicted a little less than half of the functional profiling prediction, with slight differences related to a small number of genes that differed between the two models (Fig. 5b). The receiver operating characteristic (ROC) curve showed no significant difference between the models (Fig. 5a). The models are in part based on *S. cerevisiae* GEM (Duarte et al., 2004; Nookaew et al., 2008). Examination of GO terms that were enriched for genes that were predominantly under-identified showed that mitochondrial genes including tRNA synthetases were among those processes (Table S12).

4. Conclusions

This work enabled functional genomics in *Y. lipolytica* and underscored the advantages of a system of random insertion mutagenesis. Over a half-million Hermes insertion mutants were derived, allowing a

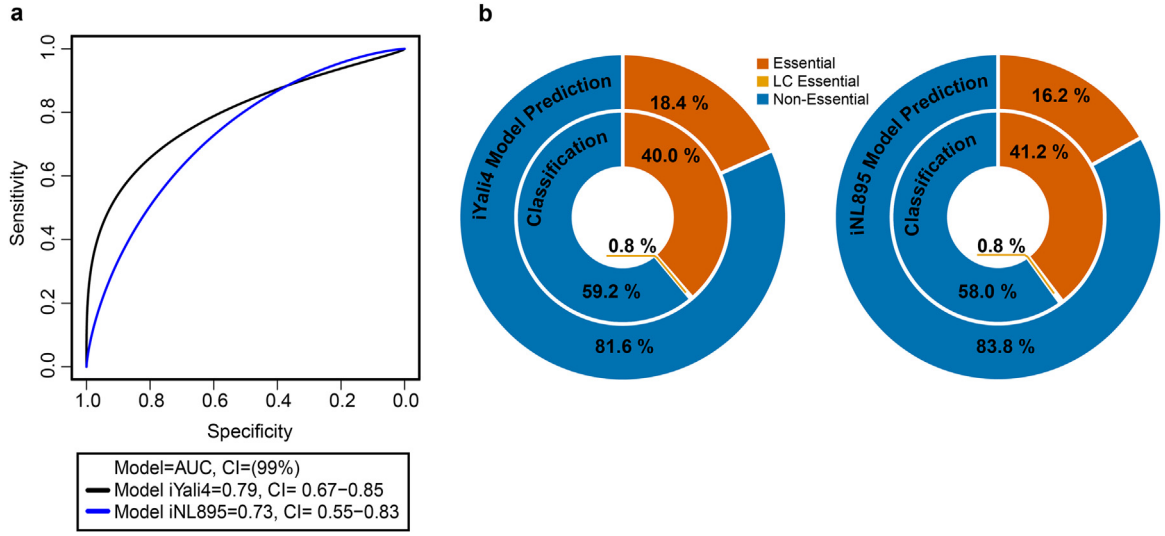

Fig. 5. Comparison of genome scale metabolic models for *Y. lipolytica*. The iNL895 (Loira et al., 2012) and iYALI4 models were evaluated using essential, non-essential and LC-essential classifications based on Hermes transposon profiling. (a) Models were allowed to execute overall genes at a thresholds of 0.5 and predict resulting biomass of single gene deletions. Receiver operating characteristics (ROC) curves show performance of model biomass predictions using the Hermes transposon profiling essential/non-essential classifications as binary classifiers. ROC curves were evaluated using the area under the curve (AUC) and 99% confidence intervals (CI). Both curves show better than random (slope = 1 and AUC ~ 0.5) performance, however the two curves are not significantly different from each other (bootstrap p-value = 0.4756). (b) Comparison of percentage of modeled genes classified by Hermes insertion profiling (inner circle) to performance of same genes classified by either iYali4 model (left) or iNL895 model (right). Gene color: essential (orange), LC essential (gold), and non-essential (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

high degree of granularity for assessing *Y. lipolytica* gene contributions to fitness. Essential genes were defined as those that had significantly less than the expected number of insertions after normalizing for gene length and flanking insertion frequency and filtering to remove intronic or 3′-bias. GCS was validated by comparisons to model yeasts, examination of conserved gene families, expectations based on known respiratory physiology of *Y. lipolytica*, and directed knockouts. Validation by these independent criteria showed that the method was sufficiently robust to draw useful inferences regarding the biology of *Y. lipolytica*.

The essential gene census derived here has multiple applications. First, on a practical level, random insertion patterns indicated that 22–30% of genes make critical contributions to survival and fitness. Attempts to attenuate function of these critical genes will now be better informed. Second, identification of unexpected non-essential genes highlighted redundancy of enzymes in the TCA cycle, and essentiality of glycerol catabolic pathways. Third, the list of essential genes will potentiate future systems biology of *Y. lipolytica* as illustrated through testing of GEMs and identification of classes of functions where the model does not perform well. Fourth, we highlighted the challenge for future investigations of the ~7% of genes with no known homolog or even recognizable domain that are nonetheless essential.

Gene contributions to fitness fall into a continuum. Some genes are absolutely essential for survival even within pure mutant colonies under the most permissive conditions, while a significant proportion of "non-essential" genes are required for cells to compete successfully in mixed cultures or under specific conditions. We highlight applications of quantifying the contributions of non-essential genes to growth. The calculation developed for relative mutant fitness differed from that described for essential gene determination in that it reflected representation of mutants in the cycling populations. This analysis identified mutants that were recovered in the Gen0 insertion library but were challenged in the mixed culture growth regime. The charts of fitness also serve as a guide to whether knockouts or specific mutants require careful interpretation, since in some instances the diversity of mutations within a single gene collapsed during growth from tens of insertions down to only a few. In contrast to deletion collections in which non-essential genes are represented by single mutants, the starting diversity of Hermes insertion mutations within genes can serve as a more complete guide for bioengineering to produce attenuated as well as knock-out phenotypes of both essential and non-essential genes.

One of the benefits of random insertion mutagenesis is recovery of phenotypes from insertions in noncoding regions, dubious ORFs or other unexpected features that can be linked to specific insertions. In a small-scale study, we demonstrated the feasibility of using FACS to identify mutants in lipid metabolism, thereby underscoring the potential of transposon profiling libraries to identify regulatory mutations.

The genomics resource described here will complement the expanding toolbox for development of *Y. lipolytica* as an industrial microbe. This Hermes transposon is portable to other *Y. lipolytica* strains and can be activated in other genetic backgrounds to identify synthetic phenotypes, dissect other biological processes, or, when coupled with fitness analysis, identify strains with robust survival. Genomewide profiling in *Y. lipolytica* will complement ongoing metabolomics, flux analysis and modeling studies to further enable the development of this unique single cell respiratory model system.

Acknowledgments

We thank Nancy Craig, Johns Hopkins University; and Henry Levin, National Institutes of Health Institute for Child Health and Human Development for helpful discussions and plasmid reagents; and Marquis Vawter, Xiaohui Xie, Jenny Wu, Diane Campbell, and Rishi Jajoo, UC Irvine, for helpful discussions of statistical analysis. We thank Cory Schwartz and Ian Wheeldon, UC Riverside, Riverside, CA, for helpful discussions and for making the CRISPR plasmid available prior to publication. We thank Tam Tran, Parth Sitlani, and Michael Aguilar for technical assistance. We thank the Staff of the Genomics High-Throughput Facility for technical support. We thank Michaela Hatch in the laboratory of C.W. Hughes, Dept. of Molecular Biology and Biochemistry, UCI, for assistance with cell sorting. We also acknowledge the support of Jennifer Atwood and the Chao Family Comprehensive Cancer Center Flow Cytometry Core Shared Resource, supported by the National Cancer Institute of the National Institutes of

Health under award number P30CA062203. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Funding

This research was supported by the National Science Foundation (EEC-0813570, B. Shanks, PI; subcontract SS) through the Engineering Research Center for Biorenewable Chemicals (CBiRC) Iowa State University, Ames, Iowa) and by a CBiRC Industrial Advisory Board Seed Grant (400-72-04-32-12F2) to SBS. This work was made possible, in part, through access to the Genomic High Throughput Facility Shared Resource of the National Cancer Institute Cancer Center Support Grant (CA-62203, R. Van Etten, PI) at the University of California, Irvine and NIH Shared Instrumentation Grants 1S10RR025496-01 and 1S10OD010794-01 to SS. KP was supported by the National Institute for Allergy and Infectious Diseases NIAID T32 AI07319 Molecular Biology of Eukaryotic Viruses (B. Semler, PI).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ymben.2018.05.008.

References

Alexander, R.P., Fang, G., Rozowsky, J., Snyder, M., Gerstein, M.B., 2010. Annotating non-coding regions of the genome. Nat. Rev. Genet. 11, 559–571.
Amberg, D.C., Burke, D.J., Strathern, J.N., 2005. Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Cold Spring Harbor, NY.
Andrews, S.J., Rothnagel, J.A., 2014. Emerging evidence for functional peptides encoded by short open reading frames. Nat. Rev. Genet. 15, 193–204.
Arensburger, P., Hice, R.H., Zhou, L., Smith, R.C., Tom, A.C., Wright, J.A., Knapp, J., O'Brochta, D.A., Craig, N.L., Atkinson, P.W., 2011. Phylogenetic and functional characterization of the hAT transposon superfamily. Genetics 188, 45–57.
Becker, S.A., Feist, A.M., Mo, M.L., Hannum, G., Palsson, B.O., Herrgard, M.J., 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat. Protoc. 2, 727–738.
Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J.L., Molina-Jouve, C., Nicaud, J.M., 2009. Yarrowia lipolytica as a model for bio-oil production. Prog. Lipid Res. 48, 375–387.
Beopoulos, A., Mrozova, Z., Thevenieau, F., Le Dall, M.T., Hapala, I., Papanikolaou, S., Chardot, T., Nicaud, J.M., 2008. Control of lipid accumulation in the yeast Yarrowia lipolytica. Appl. Environ. Microbiol. 74, 7779–7789.
Beopoulos, A., Nicaud, J.M., Gaillardin, C., 2011. An overview of lipid metabolism in yeasts and its impact on biotechnological processes. Appl. Microbiol. Biotechnol. 90, 1193–1206.
BioModels Database, 2016. EMBL-EBI.
Blazeck, J., Hill, A., Liu, L., Knight, R., Miller, J., Pan, A., Otoupal, P., Alper, H.S., 2014. Harnessing Yarrowia lipolytica lipogenesis to create a platform for lipid and biofuel production. Nat. Commun. 5, 3131.
Boeke, J.D., Trueheart, J., Natsoulis, G., Fink, G.R., 1987. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. Methods Enzymol. 154, 164–175.
Chen, D.C., Beckerich, J.M., Gaillardin, C., 1997. One-step transformation of the dimorphic yeast Yarrowia lipolytica. Appl. Microbiol. Biotechnol. 48, 232–235.
Cherry, M., 2016. Saccharomyces Genome Database.
Crook, N., Abatemarco, J., Sun, J., Wagner, J.M., Schmitz, A., Alper, H.S., 2016. In vivo continuous evolution of genes and pathways in yeast. Nat. Commun. 7, 13051.
Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190.
Dowell, R.D., Ryan, O., Jansen, A., Cheung, D., Agarwala, S., Danford, T., Bernstein, D.A., Rolfe, P.A., Heisler, L.E., Chin, B., Nislow, C., Giaever, G., Phillips, P.C., Fink, G.R., Gifford, D.K., Boone, C., 2010. Genotype to phenotype: a complex problem. Science 328, 469.
Duarte, N.C., Herrgard, M.J., Palsson, B.O., 2004. Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. Genome Res. 14, 1298–1309.
Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuveglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.M., Beyne, E., Bleykasten, C., Boisrame, A., Boyer, J., Cattolico, L., Confanioleri, F., De Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J.M., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, G.F., Straub, M.L., Suleau, A., Swennen, D., Tekaia, F., Wesolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P., Souciet, J.L., 2004. Genome evolution in yeasts. Nature 430, 35–44.
Evertts, A.G., Plymire, C., Craig, N.L., Levin, H.L., 2007. The hermes transposon of Musca domestica is an efficient tool for the mutagenesis of Schizosaccharomyces pombe. Genetics 177, 2519–2523.
Gangadharan, S., Mularoni, L., Fain-Thornton, J., Wheelan, S.J., Craig, N.L., 2010. DNA transposon Hermes inserts into DNA in nucleosome-free regions in vivo. Proc. Natl. Acad. Sci. USA 107, 21966–21972.
Gerdes, S.Y., Scholle, M.D., Campbell, J.W., Balazsi, G., Ravasz, E., Daugherty, M.D., Somera, A.L., Kyrpides, N.C., Anderson, I., Gelfand, M.S., Bhattacharya, A., Kapatral, V., D'Souza, M., Baev, M.V., Grechkin, Y., Mseeh, F., Fonstein, M.Y., Overbeek, R., Barabasi, A.L., Oltvai, Z.N., Osterman, A.L., 2003. Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. J. Bacteriol. 185, 5673–5684.
Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., Arkin, A.P., Astromoff, A., El-Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K.D., Flaherty, P., Foury, F., Garfinkel, D.J., Gerstein, M., Gotte, D., Guldener, U., Hegemann, J.H., Hempel, S., Herman, Z., Jaramillo, D.F., Kelly, D.E., Kelly, S.L., Kotter, P., LaBonte, D., Lamb, D.C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S.L., Revuelta, J.L., Roberts, C.J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D.D., Sookhai-Mahadeo, S., Storms, R.K., Strathern, J.N., Valle, G., Voet, M., Volckaert, G., Wang, C.Y., Ward, T.R., Wilhelmy, J., Winzeler, E.A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J.D., Snyder, M., Philippsen, P., Davis, R.W., Johnston, M., 2002. Functional profiling of the Saccharomyces cerevisiae genome. Nature 418, 387–391.
Giaever, G., Nislow, C., 2014. The yeast deletion collection: a decade of functional genomics. Genetics 197, 451–465.
Gibson, D.G., 2009. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. Nucleic Acids Res. 37, 6984–6990.
Gietz, R.D., Schiestl, R.H., 2007. Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. 2, 38–41.
Groenewald, M., Boekhout, T., Neuveglise, C., Gaillardin, C., van Dijck, P.W., Wyss, M., 2014. Yarrowia lipolytica: safety assessment of an oleaginous yeast with a great industrial potential. Crit. Rev. Microbiol. 40, 187–206.
Guimond, N., Bideshi, D.K., Pinkerton, A.C., Atkinson, P.W., O'Brochta, D.A., 2003. Patterns of Hermes transposition in Drosophila melanogaster. Mol. Genet. Genom. : MGG 268, 779–790.
Guo, Y., Park, J.M., Cui, B., Humes, E., Gangadharan, S., Hung, S., FitzGerald, P.C., Hoe, K.L., Grewal, S.I., Craig, N.L., Levin, H.L., 2013. Integration profiling of gene function with dense maps of transposon integration. Genetics 195, 599–609.
Guthrie, C., Fink, G.R., 1991. Guide to Yeast Genetics and Molecular Biology. 194 Academic Press, San Diego, CA.
Hacker, U., Nystedt, S., Barmchi, M.P., Horn, C., Wimmer, E.A., 2003. piggyBac-based insertional mutagenesis in the presence of stably integrated P elements in Drosophila. Proc. Natl. Acad. Sci. USA 100, 7720–7725.
Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30.
Kastenmayer, J.P., Ni, L., Chu, A., Kitchen, L.E., Au, W.C., Yang, H., Carter, C.D., Wheeler, D., Davis, R.W., Boeke, J.D., Snyder, M.A., Basrai, M.A., 2006. Functional genomics of genes with small open reading frames (sORFs) in S. cerevisiae. Genome Res. 16, 365–373.
Kerkhoven, E.J., Pomraning, K.R., Baker, S.E., Nielsen, J., 2016. Regulation of amino-acid metabolism controls flux to lipid accumulation in *Yarrowia lipolytica*. Syst. Biol. Appl. 2.
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, Y.S., Lee, M., Kim, L., Heo, K.S., Noh, E.J., Lee, A.R., Jang, Y.J., Chung, K.S., Choi, S.J., Park, J.Y., Park, Y., Kim, H.M., Park, S.K., Park, H.J., Kang, E.J., Kim, H.B., Kang, H.S., Park, H.M., Kim, K., Song, K., Song, K.B., Nurse, P., Hoe, K.L., 2010. Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. 28, 617–623.
Kretzschmar, A., Otto, C., Holz, M., Werner, S., Hubner, L., Barth, G., 2013. Increased homologous integration frequency in Yarrowia lipolytica strains defective in non-homologous end-joining. Curr. Genet. 59, 63–72.
Kumar, A., Snyder, M., 2001. Genome-wide transposon mutagenesis in yeast. Curr. Protoc. Mol. Biol (Chapter 13, Unit13 3).
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.
Li, M.A., Pettitt, S.J., Eckert, S., Ning, Z., Rice, S., Cadinanos, J., Yusa, K., Conte, N., Bradley, A., 2013. The piggyBac transposon displays local and distant reintegration preferences and can cause mutations at noncanonical integration sites. Mol. Cell Biol. 33, 1317–1330.
Loira, N., Dulermo, T., Nicaud, J.M., Sherman, D.J., 2012. A genome-scale metabolic model of the lipid-accumulating yeast Yarrowia lipolytica. BMC Syst. Biol. 6, 35.
Magnan, C., Yu, J., Chang, I., Jahn, E., Kanomata, Y., Wu, J., Zeller, M., Oakes, M., Baldi, P., Sandmeyer, S., 2016. Sequence assembly of Yarrowia lipolytica strain W29/CLIB89 shows transposable element diversity. PLoS One 11, e0162363.
Mekouar, M., Blanc-Lenfle, I., Ozanne, C., Da Silva, C., Cruaud, C., Wincker, P., Gaillardin, C., Neuveglise, C., 2010. Detection and analysis of alternative splicing in Yarrowia lipolytica reveal structural constraints facilitating nonsense-mediated decay of intron-retaining transcripts. Genome Biol. 11, R65.
Molin, M., Norbeck, J., Blomberg, A., 2003. Dihydroxyacetone kinases in Saccharomyces cerevisiae are involved in detoxification of dihydroxyacetone. J. Biol. Chem. 278, 1415–1423.

Muller, S., Sandal, T., Kamp-Hansen, P., Dalboge, H., 1998. Comparison of expression systems in the yeasts Saccharomyces cerevisiae, Hansenula polymorpha, Klyveromyces lactis, Schizosaccharomyces pombe and Yarrowia lipolytica. Cloning of two novel promoters from Yarrowia lipolytica. Yeast 14, 1267–1283.
Nicaud, J.M., 2012. Yarrowia lipolytica. Yeast 29, 409–418.
Nookaew, I., Jewett, M.C., Meechai, A., Thammarongtham, C., Laoteng, K., Cheevadhanarak, S., Nielsen, J., Bhumiratana, S., 2008. The genome-scale metabolic model iIN800 of Saccharomyces cerevisiae and its validation: a scaffold to query lipid metabolism. BMC Syst. Biol. 2, 71.
PomBase, 2016. Schizosaccharomyces Pombe Database. (pp).
Qian, W., Ma, D., Xiao, C., Wang, Z., Zhang, J., 2012. The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. Cell Rep. 2, 1399–1410.
Qiao, K., Imam Abidi, S.H., Liu, H., Zhang, H., Chakraborty, S., Watson, N., Kumaran Ajikumar, P., Stephanopoulos, G., 2015. Engineering lipid overproduction in the oleaginous yeast Yarrowia lipolytica. Metab. Eng. 29, 56–65.
Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.
Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., Muller, M., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinform. 12, 77.
Ross-Macdonald, P., Coelho, P.S., Roemer, T., Agarwal, S., Kumar, A., Jansen, R., Cheung, K.H., Sheehan, A., Symoniatis, D., Umansky, L., Heidtman, M., Nelson, F.K., Iwasaki, H., Hager, K., Gerstein, M., Miller, P., Roeder, G.S., Snyder, M., 1999. Large-scale analysis of the yeast genome by transposon tagging and gene disruption. Nature 402, 413–418.
Sambrook, J., Russell, D.W., 2006. The Condensed Protocols from Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Schneider, T.D., Stephens, R.M., 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18, 6097–6100.
Schwartz, C.M., Hussain, M.S., Blenner, M., Wheeldon, I., 2016. Synthetic RNA polymerase III promoters facilitate high-efficiency CRISPR-Cas9-mediated genome editing in Yarrowia lipolytica. ACS Synth. Biol.
Seip, J., Jackson, R., He, H., Zhu, Q., Hong, S.P., 2013. Snf1 is a regulator of lipid accumulation in Yarrowia lipolytica. Appl. Environ. Microbiol. 79, 7360–7370.
Sherman, D., Durrens, P., Beyne, E., Nikolski, M., Souciet, J.L., Genolevures, C., 2004. Genolevures: comparative genomics and molecular evolution of hemiascomycetous yeasts. Nucleic Acids Res. 32, D315–D318.
Sprague, G.F., Cronan, J.E., 1977. Isolation and characterization of Saccharomyces cerevisiae mutants defective in glycerol catabolism. J. Bacteriol. 129, 1335–1342.
Tai, M., Stephanopoulos, G., 2013. Engineering the push and pull of lipid biosynthesis in oleaginous yeast Yarrowia lipolytica for biofuel production. Metab. Eng. 15, 1–9.
Teng, X., Dayhoff-Brannigan, M., Cheng, W.C., Gilbert, C.E., Sing, C.N., Diny, N.L., Wheelan, S.J., Dunham, M.J., Boeke, J.D., Pineda, F.J., Hardwick, J.M., 2013. Genome-wide consequences of deleting any single gene. Mol. Cell. 52, 485–494.
van Opijnen, T., Bodi, K.L., Camilli, A., 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat. Methods 6, 767–772.
Wagner, J.M., Williams, E.V., Alper, H.S., 2018. Developing a piggyBac transposon system and compatible selection markers for insertional mutagenesis and genome engineering in Yarrowia lipolytica. Biotechnol. J. 13, 1800022.
Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., Chu, A.M., Connelly, C., Davis, K., Dietrich, F., Dow, S.W., El Bakkoury, M., Foury, F., Friend, S.H., Gentalen, E., Giaever, G., Hegemann, J.H., Jones, T., Laub, M., Liao, H., Liebundguth, N., Lockhart, D.J., Lucau-Danila, A., Lussier, M., M'Rabet, N., Menard, P., Mittmann, M., Pai, C., Rebischung, C., Revuelta, J.L., Riles, L., Roberts, C.J., Ross-MacDonald, P., Scherens, B., Snyder, M., Sookhai-Mahadeo, S., Storms, R.K., Veronneau, S., Voet, M., Volckaert, G., Ward, T.R., Wysocki, R., Yen, G.S., Yu, K., Zimmermann, K., Philippsen, P., Johnston, M., Davis, R.W., 1999. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science 285, 901–906.
Wiser, M.J., Lenski, R.E., 2015. A comparison of methods to measure fitness in Escherichia coli. PLoS One 10, e0126210.
Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P., Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D., Hidalgo, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E.J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M., Leather, S., McDonald, S., McLean, J., Mooney, P., Moule, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neil, S., Pearson, D., Quail, M.A., Rabbinowitsch, E., Rutherford, K., Rutter, S., Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R.G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J., Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I., Vanstreels, E., Rieger, M., Schafer, M., Muller-Auer, S., Gabel, C., Fuchs, M., Dusterhoft, A., Fritzc, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehrach, H., Reinhardt, R., Pohl, T.M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dreano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Aves, S.J., Xiang, Z., Hunt, C., Moore, K., Hurst, S.M., Lucas, M., Rochet, M., Gaillardin, C., Tallada, V.A., Garzon, A., Thode, G., Daga, R.R., Cruzado, L., Jimenez, J., Sanchez, M., del Rey, F., Benito, J., Dominguez, A., Revuelta, J.L., Moreno, S., Armstrong, J., Forsburg, S.L., Cerutti, L., Lowe, T., McCombie, W.R., Paulsen, I., Potashkin, J., Shpakovski, G.V., Ussery, D., Barrell, B.G., Nurse, P., 2002. The genome sequence of Schizosaccharomyces pombe. Nature 415, 871–880.
Yadav, V.G., Stephanopoulos, G., 2014. Metabolic engineering: the ultimate paradigm for continuous pharmaceutical manufacturing. ChemSusChem 7, 1847–1853.
Yoon, O.K., Brem, R.B., 2010. Noncanonical transcript forms in yeast and their regulation during environmental stress. RNA 16, 1256–1267.
Yuzbasheva, E.Y., Mostova, E.B., Andreeva, N.I., Yuzbashev, T.V., Fedorov, A.S., Konova, I.A., Sineoky, S.P., 2017. A metabolic engineering strategy for producing free fatty acids by the Yarrowia lipolytica yeast based on impairment of glycerol metabolism. Biotechnol. Bioeng.
Zhu, Q., Jackson, E.N., 2015. Metabolic engineering of Yarrowia lipolytica for industrial applications. Curr. Opin. Biotechnol. 36, 65–72.
Zhu, Z., Zhang, S., Liu, H., Shen, H., Lin, X., Yang, F., Zhou, Y.J., Jin, G., Ye, M., Zou, H., Zhao, Z.K., 2012. A multi-omic map of the lipid-producing yeast Rhodosporidium toruloides. Nat. Commun. 3, 1112.