nature biotechnology

BRIEF COMMUNICATIONS

Genome-scale engineering of *Saccharomyces cerevisiae* with single-nucleotide precision

Zehua Bao[1], Mohammad HamediRad[2], Pu Xue[2], Han Xiao[1,7], Ipek Tasan[3], Ran Chao[2], Jing Liang[4] & Huimin Zhao[1–6]

We developed a CRISPR–Cas9- and homology-directed-repair-assisted genome-scale engineering method named CHAnGE that can rapidly output tens of thousands of specific genetic variants in yeast. More than 98% of target sequences were efficiently edited with an average frequency of 82%. We validate the single-nucleotide resolution genome-editing capability of this technology by creating a genome-wide gene disruption collection and apply our method to improve tolerance to growth inhibitors.

Genome-scale engineering enables multiple hypotheses to be tested by producing genome-wide mutations in parallel. Existing approaches such as MAGE[1], TRMR[2], and CREATE[3] have mainly been applied in bacteria. Although CREATE was shown to work in yeast, in principle, efficient, high-throughput genome-wide engineering was not reported[3]. One problem with some existing genome-scale methods is that, because *Escherichia coli* cannot readily repair double-strand breaks, there is substantial selection pressure during mutagenesis for cells that have undergone homology-directed repair. The same is not true in yeast, and high-throughput approaches have not, thus far, been proven to work efficiently on a genome-wide scale.

Eukaryotic MAGE (eMAGE) enables genome engineering in yeast[4] but the editing efficiency of eMAGE relies on close proximity of target sequences to a replication origin and co-selection of a *URA3* marker. Although genome-wide engineering may be feasible using eMAGE, it was not demonstrated[4]. We report a CRISPR–Cas9- and homology-directed-repair (HDR)-assisted genome-scale engineering (CHAnGE) method that enables rapid engineering of *Saccharomyces cerevisiae* on a genome scale with precise and trackable edits.

To enable large-scale engineering using HDR, we synthesized the CRISPR guide sequence and the homologous recombination (HR) template in a single oligonucleotide (the CHAnGE cassette, **Fig. 1a**). In the CHAnGE cassette, we moved the long eukaryotic RNA promoter to the plasmid backbone to reduce oligonucleotide length, whereas the CREATE cassette includes a promoter. Cloning and delivering a pooled CHAnGE plasmid library into a yeast strain and subsequent editing will generate a yeast mutant library (**Fig. 1b**). The unique CHAnGE cassette in each plasmid serves as a genetic barcode for mutant tracking by next-generation sequencing (NGS).

We first applied CHAnGE to genome-wide gene disruption. To do this, previously described criteria[5–7] to maximize the efficacy and specificity of guide sequences were applied to design guides targeting each open reading frame (ORF) in the *S. cerevisiae* genome. Arbitrary weights were assigned to each criterion to derive a score for each guide (**Supplementary Table 1**). For each ORF, we selected four top-rank guides. For some ORFs, fewer guides were selected owing to short or repetitive ORF sequences. In total we used 24,765 unique guide sequences targeting 6,459 ORFs (~97.8% of ORFs annotated in the Saccharomyces Genome Database (SGD), **Supplementary Table 2**). We also included 100 non-editing guide sequences as controls. For each ORF-targeting guide, a 100-bp HR template with 50-bp homology arms and a centered 8-bp deletion was used. The deletion removes the protospacer-adjacent motif (PAM) sequence and causes a frameshift mutation for gene disruption (**Fig. 1a**). Adapters containing priming and BsaI sites were added to both ends of the oligonucleotide to facilitate cloning (**Supplementary Fig. 1**). CHAnGE cassettes are listed in **Supplementary Table 3**.

We measured editing efficiencies of CHAnGE cassettes with varying scores. In the designed library, 98.4% of the cassettes had a score of more than 60 (**Fig. 1c**). We tested 30 cassettes targeting *CAN1*, *ADE2*, and *LYP1* (**Supplementary Table 4**). Cassettes with a score > 60 had median and average editing efficiencies of 88% and 82%, respectively. Cassettes with a score <60 had median and average editing efficiencies of 81% and 61%, respectively (**Fig. 1d**). Considering that only 1.6% of cassettes in the library had a low score, these results suggest that CHAnGE cassettes enable efficient editing. Compared with eMAGE (from ~1.0% at a distance of 20 kb to >40% next to a replication origin), editing efficiency using CHAnGE was superior, independent of target site.

To generate a pooled plasmid library, we synthesized designed oligonucleotides on a chip and then assembled them into pCRCT[5] plasmids (**Fig. 1b**). Sequencing of 91 assembled plasmids revealed that 37.36% were correct (**Supplementary Fig. 2**), reflecting a 0.58% synthesis error rate per base. NGS of the plasmid library captured 95.5% of the designed guide sequences, which covered 99.5% of the targeted ORFs. *S. cerevisiae* was transformed by the plasmid library using heat shock to yield pooled single mutants, each containing an 8-nucleotide deletion in a single gene. 395-fold coverage was achieved

[1]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. [2]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. [3]Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. [4]Metabolic Engineering Research Laboratory, Science and Engineering Institutes, Agency for Science, Technology and Research, Singapore, Singapore. [5]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. [6]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. [7]Present address: State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences, and Laboratory of Molecular Biochemical Engineering, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. Correspondence should be addressed to H.Z. (zhao5@illinois.edu).

Received 1 May 2017; accepted 2 March 2018; published online 7 May 2018; doi:10.1038/nbt.4132

(**Supplementary Table 5**), ensuring the completeness of a collection of genome-wide gene deletions. The number of transformations can be scaled up to obtain efficiencies required for even larger library sizes. We screened our mutant library for *CAN1* mutants in the presence of L-(+)-(*S*)-canavanine and identified all four *CAN1*-targeting guides; non-edited controls were depleted since wild-type yeast cells are killed by canavanine (**Fig. 1e**). Some cassettes were not observed owing to the low NGS read depth (**Supplementary Table 5**). Reducing the synthesis error rate or assigning more reads to each sample could alleviate this problem.

We next used CHAnGE to engineer furfural tolerance. Selection with 5 mM furfural enriched *SIZ1* targeting guides (**Fig. 1f** and **Supplementary Fig. 3**), in line with previous findings[8]. Guide sequences targeting newly identified genes *SAP30* and *UBC*[4], were also enriched. All three disruption mutants grew faster in the presence of furfural compared with the wild-type parent (**Supplementary Fig. 4**). However, combining the individual gene disruptions into a single strain did not improve tolerance further (**Supplementary Fig. 5**), suggesting that these beneficial mutations are neither additive nor synergistic[9]. We selected SIZ1Δ1 (edited by CHAnGE cassette SIZ1_1) as the parental strain and repeated the CHAnGE workflow a second time. *LCB3* targeting guides were enriched in 10 mM furfural during the second round of evolution (**Fig. 1f**). Increased tolerance was confirmed by measuring the growth of wild-type, single, and double mutants in 10 mM furfural (**Fig. 1g**). Notably, the phenotype of the *LCB3* mutant was dependent on *SIZ1* disruption; *LCB3* targeting guides were not enriched in the first round of evolution, and the single *LCB3* disruption mutant LCB3Δ1 showed similar growth as wild type (**Fig. 1f**,**g**), showing epistasis. We also applied CHAnGE for directed evolution of acetic acid tolerance and achieved a 20-fold improvement (**Supplementary Note 1** and **Supplementary Figs. 6–8**).

Next, we applied CHAnGE to single-nucleotide resolution editing. Exogenous Siz1 mutations (F268A, D345A, I363A, S391D, F250A/F299A, FKSΔ) were previously shown to diminish SUMO conjugation to PCNA[10,11]. We designed seven CHAnGE cassettes to introduce these seven mutations and an insertion mutation (**Fig. 2a** and **Supplementary Figs. 9–12**). In each cassette, codon substitutions were placed between the homology arms. Unlike the CREATE cassette in which only one edit can be performed at a time, the CHAnGE cassette F250A F299A was designed to simultaneously introduce two distal codon substitutions (147 bp apart, **Supplementary Fig. 10**). Except for I363A, we observed all other designed Siz1 mutations with efficiencies of 80–100% (**Fig. 2b**). These results highlight the capability of CHAnGE to introduce mutations that are unlikely to occur spontaneously, such as those requiring two or three bases within a codon to be altered (e.g., F268A and S391D). F268A, D345A, S391D, FKSΔ, and AAA all showed improved furfural tolerance (**Fig. 2c**), suggesting that reducing PCNA sumoylation has a role in acquired furfural tolerance. An increased growth rate was not observed for F250A F299A, which may represent a difference between endogenously and episomally expressed mutants. We also designed eight CHAnGE cassettes targeting *CAN1* and *UBC4*, and achieved an average editing efficiency of 90% for 7/8 cassettes, which provides evidence that our method is generalizable to different loci (**Supplementary Note 2** and **Supplementary Figs. 13–17**).

Finally, we carried out tiling mutagenesis of the Siz1 SP-CTD domain. We first modified the CHAnGE cassette to reduce the length of homology arms to 40 bp, so that the sequence between the target codon and the PAM could be accommodated (**Fig. 2d**). We designed five CHAnGE cassettes with 40-bp homology arms targeting *UBC4*, and achieved an average editing efficiency of 86% (**Supplementary Fig. 17a**). To minimize the length of CHAnGE

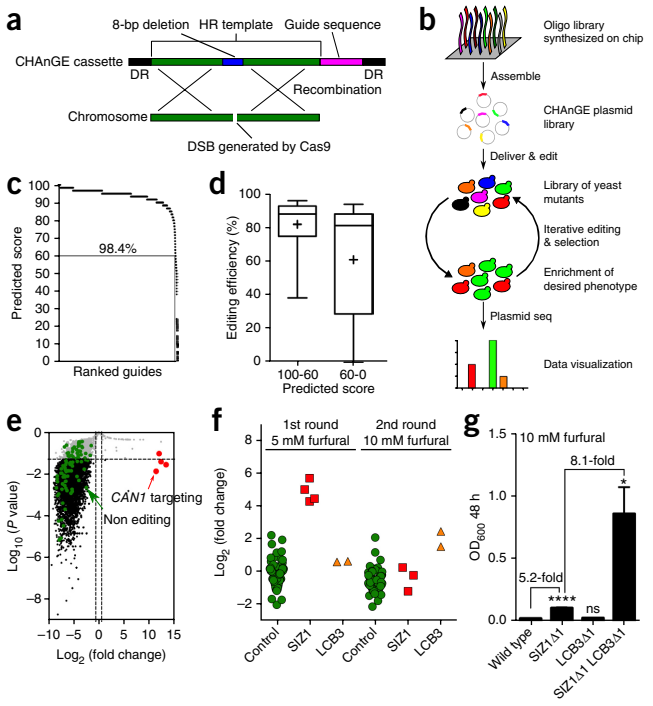

Figure 1 CHAnGE enables rapid generation of genome-wide yeast disruption mutants and directed evolution of complex phenotypes. (**a**) Design of the CHAnGE cassette. DR, direct repeat. (**b**) The CHAnGE workflow. (**c**) Distribution of guide sequences by predicted scores. (**d**) Editing efficiencies of CHAnGE cassettes with varying predicted scores. The box extends from the 25th to 75th percentiles. The line in the middle of the box is plotted at the median. The plus symbol denotes the mean. The whiskers go down to the smallest value and up to the largest. $n = 12$ for the group with scores over 60. $n = 18$ for the group with scores <60. (**e**) Genetic screening of *CAN1* disruption mutants in the presence of canavanine. Volcano plot is shown for canavanine-stressed libraries versus untreated libraries. The *x* axis represents enrichment levels of each guide sequence. The y axis represents $\log_{10}$-transformed *P* values. Significantly enriched guides ($P < 0.05$, fold change > 1.5) are denoted by black dots, all others by gray dots. Dotted lines indicate 1.5-fold ratio (*x* axis) and *P* value of 0.05 (y axis). $n = 2$ independent experiments. (**f**) Enrichment of guide sequences during the first round and second round directed evolution of furfural tolerance. Green circles, control guides; red squares, *SIZ1*-targeting guides; yellow triangles, *LCB3*-targeting guides. (**g**) Biomass accumulation of the wild-type and mutant strains in the presence of furfural. $n = 3$ independent experiments. Error bars, s.e.m. Two-tailed *t*-tests were performed to determine significance levels against the wild-type strain. $^*P < 0.05$. $^{****}P < 0.0001$. ns, not significant.

cassettes, we restricted the PAM–codon distance to 20 bp or less. Given that the density of NGG PAMs is one per 8 bp[12], there is a 93% chance of there being a PAM in this distance. We also used a genetic barcode within the donor to enable NGS tracking because 20-bp guides may not be unique (**Fig. 2d**). To evaluate editing efficiencies of CHAnGE cassettes with varying PAM–codon distances, we designed 30 CHAnGE cassettes to disrupt *CAN1*, *ADE2*, and *LYP1* (**Supplementary Table 4**). Cassettes with a PAM–codon distance up to 20 bp have 41% (median) and 47% (average) editing efficiencies, respectively. Cassettes with a PAM–codon distance of more than 20 bp have a less than 25% editing efficiency (**Fig. 2e**). We designed 580 CHAnGE cassettes (**Supplementary Table 6**) for saturation mutagenesis of the 29 amino acid residues of the SP-CTD domain, which consists of an α-helix and a β-strand. Amino acid residues from the C terminal of the α-helix and the entire β-strand interact extensively with SUMO (**Fig. 2f**). For example, E344 and D345 from the α-helix

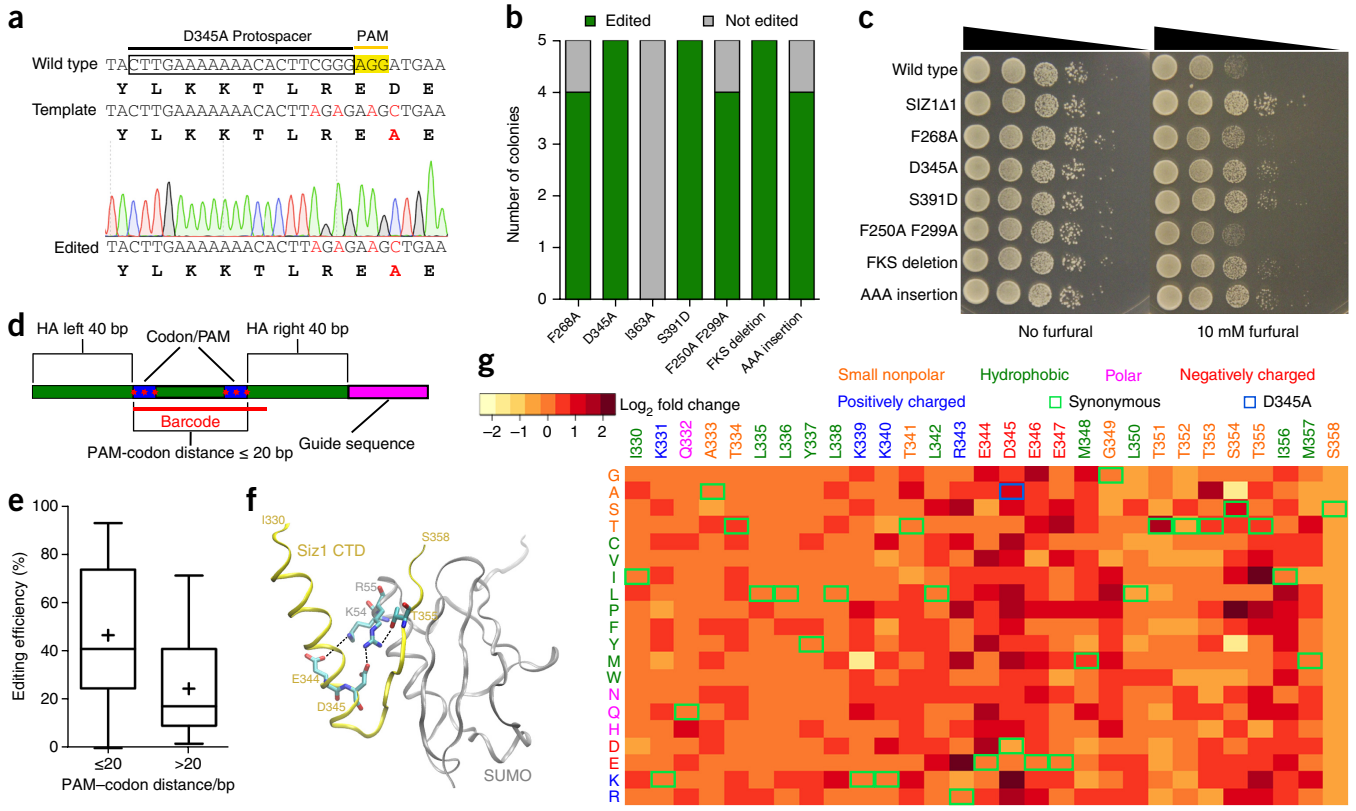

Figure 2 CHAnGE enables genome editing with a single-nucleotide resolution. (**a**) A representative figure showing the designed mutations in the Siz1 D345A CHAnGE cassette. The designed mutations in the HR template and the amino acid substitution are red. A Sanger sequencing trace file of a representative edited colony was shown at the bottom. (**b**) A summary of *SIZ1* precise editing efficiencies. For each mutagenesis, 5 randomly picked colonies were examined. (**c**) Spotting assay of *SIZ1* mutants in the presence of furfural. Black triangles denote serial dilutions. (**d**) Design of a modified CHAnGE cassette for single-nucleotide resolution editing. Blue rectangles denote the target codon and the PAM. Red stars denote mutations for codon substitution and PAM elimination. (**e**) Editing efficiencies of modified CHAnGE cassettes with varying PAM–codon distances. The box extends from the 25th to 75th percentiles. The line in the middle of the box is plotted at the median. The plus symbol denotes the mean. The whiskers go down to the smallest value and up to the largest. $n = 10$ for the group with distances less than 20 bp. $n = 20$ for the group with distances over 20 bp. (**f**) Crystal structure of Siz1 SP-CTD forming a complex with SUMO. Black dashed lines denote hydrogen bonds. PDB code 5JNE. (**g**) Heatmap showing the enrichment of 580 CHAnGE cassettes after selection with 5 mM furfural. Original and substitute amino acid residues are denoted on the top and at the left, respectively, and are colored according to the Lesk color scheme. Synonymous CHAnGE cassettes are denoted by green boxes. Cassette D345A is denoted by a blue box.

form hydrogen bonds with SUMO K54 and R55, respectively. T355 from the β-strand forms a hydrogen bond with SUMO R55 (ref. 10). When the yeast Siz1 mutant library was subject to furfural selection, we observed enrichment of the validated D345A, but no enrichment of most of the synonymous cassettes (**Fig. 2g** and **Supplementary Table 5**). Using this method we identified two enrichment hot spots centered around D345 and T355, consistent with molecular interactions between SP-CTD and SUMO.

CHAnGE is a trackable method to produce a genome-wide set of yeast mutants with single-nucleotide precision. Design of CHAnGE cassettes may be affected by the presence of BsaI sites and polyT sequences. Therefore, optimization using HR assembly and type II RNA promoters could expand the design space. Increasing the number of experimental replicates and design redundancy of CHAnGE cassettes should be considered to reduce false-positive rates. CHAnGE might be adopted for genome-scale engineering of higher eukaryotes, as preliminary experiments reveal precise editing of the human *EMX1* locus using a CHAnGE cassette (**Supplementary Fig. 18**), but improved efficiency of homology-directed repair in higher eukaryotes is needed first.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by the Carl R. Woese Institute for Genomic Biology at the University of Illinois at Urbana-Champaign and the US Department of Energy (DE-SC0018260). We thank A. Hernandez and C. Wright for assistance with next-generation sequencing, J. Zadeh for assistance with NGS data processing and analysis.

AUTHOR CONTRIBUTIONS

Z.B. and H.Z. conceived this project. Z.B., M.H., and H.X. designed the CHAnGE cassettes. R.C. and J.L. generated the ORF list and all possible guide sequences. M.H. sorted and selected the guide and homology arm sequences. Z.B., P.X., and I.T. performed the experiments. Z.B. analyzed the data. H.Z. supervised the research. Z.B. and H.Z. wrote the manuscript.

COMPETING INTERESTS

A patent application has been filed on this technology, on which H.Z. and Z.B. are authors.

1. Wang, H.H. *et al. Nature* **460**, 894–898 (2009).
2. Warner, J.R., Reeder, P.J., Karimpour-Fard, A., Woodruff, L.B. & Gill, R.T. *Nat. Biotechnol.* **28**, 856–862 (2010).
3. Garst, A.D. *et al. Nat. Biotechnol.* **35**, 48–55 (2017).
4. Barbieri, E.M., Muir, P., Akhuetie-Oni, B.O., Yellman, C.M. & Isaacs, F.J. *Cell* **171**, 1453–1467.e13 (2017).
5. Bao, Z. *et al. ACS Synth. Biol.* **4**, 585–594 (2015).
6. Cong, L. *et al. Science* **339**, 819–823 (2013).
7. Wang, T., Wei, J.J., Sabatini, D.M. & Lander, E.S. *Science* **343**, 80–84 (2014).
8. Xiao, H. & Zhao, H. *Biotechnol. Biofuels* **7**, 78 (2014).
9. Sandoval, N.R. *et al. Proc. Natl. Acad. Sci. USA* **109**, 10540–10545 (2012).
10. Streich, F.C. Jr. & Lima, C.D. *Nature* **536**, 304–308 (2016).
11. Yunus, A.A. & Lima, C.D. *Mol. Cell* **35**, 669–682 (2009).
12. Kim, H. & Kim, J.S. *Nat. Rev. Genet.* **15**, 321–334 (2014).

ONLINE METHODS

Plasmid construction. All plasmids for yeast genome editing were constructed by assembling a CHAnGE cassette with pCRCT using Golden Gate assembly[5]. For human *EMX1* editing, pX330A-1 × 3-EMX1 was similarly constructed using pX330A-1 × 3 (Addgene #58767). All CHAnGE cassettes were ordered as gBlock fragments (Integrated DNA Technologies, Coralville, Iowa) and the sequences are listed in **Supplementary Tables 3** and **4**.

CHAnGE library design and synthesis. All ORF sequences from *S. cerevisiae* strain S288c were downloaded from SGD and passed through CRISPRdirect[13] to generate all possible guide sequences. Only guide sequences with hit_20mer>0 were retained to exclude those targeting exon–intron junctions. A guide-specific 100-bp HR donor was assembled 5′ of each guide sequence. All assembled sequences were passed through four additional filters: no BsaI restriction site (to facilitate Golden Gate assembly), no homopolymer of more than four T's (to prevent early transcription termination), no homopolymer of more than five A's or more than five G's (to maximize oligonucleotide synthesis efficiency). Each guide sequence was then assigned an arbitrary score for assessing both genome editing efficiency and off-target effect (**Supplementary Table 1**). Specifically, artificial weights were assigned to each efficacy criterion so that higher scores will be given to guides with 35% to 75% GC content, with high purine content in the last four nucleotides, and targeting earlier regions of the ORF. To ensure targeting specificity, the score of a guide sequence decreases exponentially as the number of its off-target sites increases. An off-target site is defined as a site containing a matching 12-bp seed sequence[6] followed by a PAM. For each ORF, the top four guide sequences with the highest scores were selected for synthesis. For ORFs with less than four unique guide sequences available, less than four guide sequences were selected. The final library contains 24,765 unique guide sequences targeting 6,459 ORFs (**Supplementary Table 2**). For unknown reasons, there are five guide sequences for ORFs YOR343W-A and YBR089C-A, and six guide sequences for ORF YMR045C. An additional 100 non-targeting guide sequences with random homology arms were randomly generated and added to the library as non-editing control guide sequences. Adapters containing priming sites and BsaI sites were added to the 5′ and 3′ ends of each oligonucleotide for PCR amplification and Golden Gate assembly. The designed oligonucleotide library was synthesized on two 12,472 format chips and eluted into two separate pools (CustomArray, Bothell, WA).

Construction of a CHAnGE plasmid library. The two oligonucleotide pools were mixed at equal molar ratio. 10 ng of the mixed oligonucleotide pool was used as a template for PCR amplification with primers BsaI-LIB-for and BsaI-LIB-rev (**Supplementary Table 7**). The cycling conditions were 98 °C for 5 min, (98 °C for 45 s, 41 °C for 30 s, 72 °C for 6 s) × 24 cycles, 72 °C for 10 min, then held at 4 °C. 15 ng of the gel-purified PCR products were assembled with 50 ng pCRCT using Golden Gate assembly method followed by plasmid-safe nuclease treatment[5]. 25 parallel Golden Gate assembly reactions were performed and the resultant DNA was purified using a PCR purification kit (Qiagen, Valencia, CA). The purified DNA was transformed into NEB5α electrocompetent cells (New England Biolabs, Ipswich, MA) using Gene Pulser Xcell Electroporation System (Bio-Rad, Hercules, CA). 20 parallel transformations were conducted and pooled. The pooled culture was plated onto four 24.5 cm × 24.5 cm LB plates supplemented with 100 μg/mL carbenicillin (Corning, New York, NY). The plates were incubated at 37 °C overnight. The total number of colony-forming units was estimated to be between 1.2×10^7 and 4×10^7, which represents a 480- to 1,600-fold coverage of the CHAnGE plasmid library. Plasmids were extracted using a Qiagen Plasmid Maxi Kit.

Generation of yeast mutant libraries. Yeast strain BY4741 was transformed with 20 μg CHAnGE plasmid library per transformation using LiAc/SS carrier DNA/PEG method[14]. After heat shock, cells were washed with 1 mL double-distilled water once and resuspended in 2 mL synthetic complete minus uracil (SC-U) liquid media. 12 parallel transformations were conducted. 2 μL culture from each of three randomly selected transformations were mixed with 98 μL sterile water and plated onto SC-U plates for assessing transformation efficiency. The total number of colony-forming units was estimated to be 9.8×10^6, which represents a 395-fold coverage of the CHAnGE plasmid library. Using SIZ1Δ1 and BUL1Δ1 as parental strains, a 499- and 129-fold coverage was achieved, respectively. The rest of the cells were cultured in 12 15 mL falcon tubes at 30 °C, shaken at 250 r.p.m. Two days after transformation, 2 units of optical density at 600 nm (OD) of cells from each tube were transferred to a new tube containing 2 mL fresh SC-U liquid media. Four days after transformation, cultures from 12 tubes were pooled. 2 OD of pooled cells were transferred to each of 12 new tubes containing 2 mL fresh SC-U media. Six days after transformation, cultures from 12 tubes were pooled and stored as glycerol stocks in a –80 °C freezer.

Screening of yeast mutant libraries. A glycerol stock of pooled yeast mutants was thawed on ice. 3.125 OD of cells was inoculated into 50 mL of SC-U liquid media with or without growth inhibitor in a 250-mL baffled flask. Cells were grown at 30 °C, shaken at 250 r.p.m. and the optical density was measured periodically. 2 OD of cells from each of the untreated and stressed populations was collected when the optical density of the stressed population reached 2.

For canavanine resistance, 60 μg/mL L-(+)-(*S*)-canavanine (Sigma Aldrich, Saint Louis, MO) supplemented SC-UR media was used. For furfural tolerance, 5 mM and 10 mM furfural (Sigma Aldrich, Saint Louis, MO) supplemented SC-U media was used. For HAc tolerance, the pH of SC-U liquid media was adjusted to 4.5. Glacial acetyl acid was dissolved in double-distilled water, adjusted to pH 4.5, and then filtered to make 10% (v/v) HAc stock solution. Appropriate volumes of HAc stock solution were added to SC-U media (pH 4.5) to make 0.5% and 0.6% HAc supplemented SC-U media. The unstressed cells were grown in SC-U media (pH 5.6).

Next-generation sequencing. For each untreated or stressed library, 2 OD of cells was collected and plasmids were extracted using Zymoprep Yeast Plasmid Miniprep II kit (Zymo Research, Irvine, CA). To attach NGS adaptors, a first-step PCR was performed using 2 × KAPA HiFi HotStart Ready Mix (Kapa Biosystems, Wilmington, MA) with primers HiSeq-CHAnGE-for and HiSeq-CHAnGE-rev (**Supplementary Table 7**) and 10 ng extracted plasmid as template. The cycling condition is 95 °C for 3 min, (95 °C for 30 s, 46 °C for 30 s, 72 °C for 30 s) × 18 cycles, 72 °C for 5 min, and held at 4 °C. The PCR product was gel purified using a Qiagen Gel Purification kit. 10 ng PCR product from the first step was used in a second-step PCR to attach Nextera indexes using the Nextera Index kit (Illumina, San Diego, CA). The cycling condition is 95 °C for 3 min, (95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s) × 8 cycles, 72 °C for 5 min, and held at 4 °C. The second-step PCR products were gel purified using a Qiagen Gel Purification kit and quantitated with Qubit (ThermoFisher Scientific, Waltham, MA). 40 ng of each library was pooled. The pool was quantitated with Qubit. The average size was determined on a Fragment Analyzer (Advanced Analytical, Ankeny, IA) and further quantitated by qPCR on a CFX Connect Real-Time qPCR system (Biorad, Hercules, CA). The pool was spiked with 30% of a PhiX library (Illumina, San Diego, CA), and sequenced on one lane for 161 cycles from one end of the fragments on a HiSeq 2500 using a HiSeq SBS sequencing kit version 4 (Illumina, San Diego, CA).

NGS data processing and analysis. Fastq files were generated and demultiplexed with the bcl2fastq v2.17.1.14 conversion software (Illumina, San Diego, CA). 20-bp guide sequences were extracted from NGS reads using fastx_toolkit/0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). A bowtie index was prepared from the 24,865 designed guide sequences (**Supplementary Table 3**). Extracted guide sequences were mapped to the bowtie index using Map with Bowtie for Illumina (version 1.1.2) command in Galaxy (https://usegalaxy.org/) with commonly used settings. Unmapped reads were removed and reads mapped to each unique guide sequence were counted. The raw read counts per guide sequence were normalized to the total read counts of a library using the following equation: normalized read counts = (raw read counts × 1,000,000)/total read counts + 1. We used a threshold of two raw read counts in at least two of the four libraries (two biological replicates of untreated library and two biological replicates of stressed library) to keep a guide sequence. Genes with all observed guide sequences enriched (fold change > 1.5) were selected for further validation.

Construction of yeast single and double mutants. An aliquot of 5 mM furfural-stressed library (OD = 2) was plated onto a SC-U plate supplemented with 5 mM furfural. 24 random colonies were picked and genotyped by PCR and Sanger sequencing. One colony was confirmed to have a designed 8-bp deletion at *SIZ1* target site 1. This colony was stored as strain SIZ1Δ1. BY4741 strains SAP30Δ3, UBC4Δ3, and LCB3Δ1 were constructed using the HI-CRISPR method[5]. The gBlock sequences can be found in **Supplementary Table 3**. For constructing double mutants SIZ1Δ1 SAP30Δ3, SIZ1Δ1 UBC4Δ3, and SIZ1Δ1 LCB3Δ1, SIZ1Δ1 was used as the parental strain.

An aliquot of 0.5% HAc-stressed library (OD = 2) was plated onto a SC-U plate supplemented with 0.5% HAc. 32 random colonies were picked and genotyped by PCR and Sanger sequencing. Three colonies were confirmed to have a designed 8-bp deletion at *BUL1* target site 1. One of these colonies was kept and stored as a strain named BUL1Δ1. A BUL1Δ1 strain without HAc exposure and the SUR1Δ1 strain were constructed using the HI-CRISPR method[5]. For constructing double mutants BUL1Δ1 SUR1Δ1, BUL1Δ1 with HAc exposure was used as the parental strain.

All other yeast mutants with non-disruption mutations were constructed using the HI-CRISPR method. The gBlock sequences can be found in **Supplementary Table 4**. For each constructed mutant, pCRCT plasmids were cured as described elsewhere[15]. Briefly, a yeast colony with the desired gene disrupted was inoculated into 5 mL of YPAD liquid medium and cultured at 30 °C, shaken at 250 r.p.m. overnight. On the next morning, 200 µL of the culture was inoculated into 5 mL of fresh YPAD medium. In the evening, 50 µL of the culture was inoculated into 5 mL of fresh YPAD medium and cultured overnight. On the next day, 100–200 cells were plated onto an YPAD plate and incubated at 30 °C until colonies appeared. For each mutant, 20 colonies were streaked onto both YPAD and SC-U plates. Colonies that failed to grow on SC-U plates were selected.

Characterization of mutant strains for furfural or HAc tolerance. BY4741 wild-type or mutant strains were inoculated from glycerol stocks into 2 mL YPAD medium and cultured at 30 °C, shaken at 250 r.p.m. overnight, then streaked onto fresh YPAD plates. Three biological replicates of each strain were inoculated in 3 mL synthetic complete (SC) medium and cultured at 30 °C, shaken at 250 r.p.m. overnight. On the next morning, 50 µL culture was inoculated into 3 mL fresh SC medium and cultured at 30 °C, shaken at 250 r.p.m. overnight to synchronize the growth phase. After 24 h, 0.03 OD of cells were inoculated into 3 mL fresh SC medium (pH 5.6) supplemented with appropriate concentrations of furfural or 3 mL fresh SC medium (pH 4.5) supplemented with appropriate concentrations of HAc. Cell densities were measured at appropriate time points.

For spotting assays, each strain was inoculated in 3 mL SC medium and cultured at 30 °C, shaken at 250 r.p.m. overnight. On the next morning, 50 µL culture was inoculated into 3 mL fresh SC medium and cultured at 30 °C, shaken at 250 r.p.m. overnight to synchronize the growth phase. After 24 h, the OD was measured and the culture was diluted to OD 1 in sterile water; tenfold serial dilutions were performed for each strain. 7.5 µL of each dilution was spotted on appropriate plates. The spotted plates were incubated at 30 °C for 2 to 6 d.

Tiling mutagenesis of *SIZ1*. For the *SIZ1* tiling mutagenesis library, the length of homology arms was reduced to 40 bp to accommodate the sequence between the PAM and the targeted codon. The PAM–codon distance was limited to be no more than 20 bp to not exceed the length limit of high-throughput oligonucleotide synthesis. For each codon, 20 CHAnGE cassettes were designed for all possible amino acid residues. The *SIZ1* oligonucleotide library was synthesized on one 12,472 format chip (CustomArray, Bothell, WA). The *SIZ1* plasmid library was similarly constructed with downscaled numbers of Golden Gate assembly reactions and transformations. The total number of colony forming unit was estimated to be between 3.8×10^5 and 8×10^5, which represents a 655 to 1,379-fold coverage of the *SIZ1* plasmid library. The *SIZ1* yeast mutant library was similarly generated with four parallel transformations. The total number of colony-forming unit was estimated to be 1.9×10^6, which represents a 3,200-fold coverage. Screening of the library and NGS were performed using the same procedures as the genome-wide disruption library. For NGS data processing, we used mutation-containing regions in the CHAnGE cassettes as genetic barcodes (**Supplementary Table 6**) for mapping the reads. Zero mismatches were allowed for the mapping.

HEK293T culture, transfections, and genotyping. HEK293T cells were purchased from ATCC (CRL-3216) and maintained in DMEM with L-glutamine and 4.5 g/L glucose and without sodium pyruvate (Mediatech, Manassas, VA) supplemented with 10% FBS and 1% penicillin/streptomycin at 37 °C in a humidified CO_2 incubator. 2×10^5 cells were plated per well of a 24-well plate one day before transfection. Cells were transfected with Lipofectamine 2000 (ThermoFisher Scientific, Waltham, MA) using 800 ng pX330A-1 × 3-EMX1 and 2.5 µL of reagent per well. Cells were maintained for an additional 3 d before harvesting. Genomic DNA was extracted using QuickExtract DNA Extraction Solution (Epicentre, Madison, WI). 5 µg of genomic DNA was used as template for selective PCR using primers EMX1-selective-for and EMX1-selective-rev (**Supplementary Table 7**). PCR amplicons were gel purified and sequenced by Sanger sequencing.

Statistics. Data is shown as mean ± s.e.m., with *n* values indicated in the figure legends. All *P* values were generated from two-tailed *t*-tests using the GraphPad Prism software package (version 6.0c, GraphPad Software) or Microsoft Excel for Mac 2011 (version 14.7.3, Microsoft Corporation).

Code availability. All computational tools used for analyses of the NGS data are available from provided references in Methods. Custom batch scripts used for execution of these computational tools can be found in **Supplementary Code**.

Life Sciences Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The raw reads of the NGS data were deposited into the Sequence Read Archive (SRA) database (accession number: PRJNA419352) at the National Center for Biotechnology Information (NCBI).

13. Naito, Y., Hino, K., Bono, H. & Ui-Tei, K. *Bioinformatics* **31**, 1120–1123 (2015).
14. Gietz, R.D. & Schiestl, R.H. *Nat. Protoc.* **2**, 31–34 (2007).
15. Hegemann, J.H. & Heick, S.B. *Methods Mol. Biol.* **765**, 189–206 (2011).

doi:10.1038/nbt.4132

nature research

Corresponding author(s): Huimin Zhao

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

Sample sizes were chosen according to experience. Sample sizes are sufficient based on subjective judgments that saturation is reached.

2. Data exclusions

Describe any data exclusions.

All data are included for analysis.

3. Replication

Describe whether the experimental findings were reliably reproduced.

All attempts at replication were successful.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Randomization is not relevant because no human participants or animal subjects were involved in this study.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Blinding is not relevant because no group allocation was involved in this study.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Nature Biotechnology: doi:10.1038/nbt.4132

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Microsoft Excel for Mac 2011 version 14.7.3, GraphPad Prism version 6.0c, bcl2fastq v2.17.1.14, fastx_toolkit/0.0.13, public server at usegalaxy.org

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	All unique materials used are readily available from the authors or from standard commercial sources as specified in Methods.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	HEK293T cell line was purchased from ATCC.
b. Describe the method of cell line authentication used.	The cell line was not authenticated.
c. Report whether the cell lines were tested for mycoplasma contamination.	The cell line was not tested for mycoplasma contamination.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	The study did not involve human research participants.

Nature Biotechnology: doi:10.1038/nbt.4132