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 de e ed a CRISPR Ca 9- a d g -d ec ed- e a - a ed ge e- ca e e g ee g e d a ed CHA GE a ca a d e f a d f ec f c ge e c a a ea . M e a 98% f a ge e e ce e e eff c e ed ed a a e age f e e c f 82%. We a da e e g e- c e de e ge e-ed g ca ab f ec g b c ea g a ge e- de ge e d c ec a d a e d e e a ce g b .

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 *E* 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 *A3* 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 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 . 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 *CA 1*, *ADE2*, and *1* (**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. . 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

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

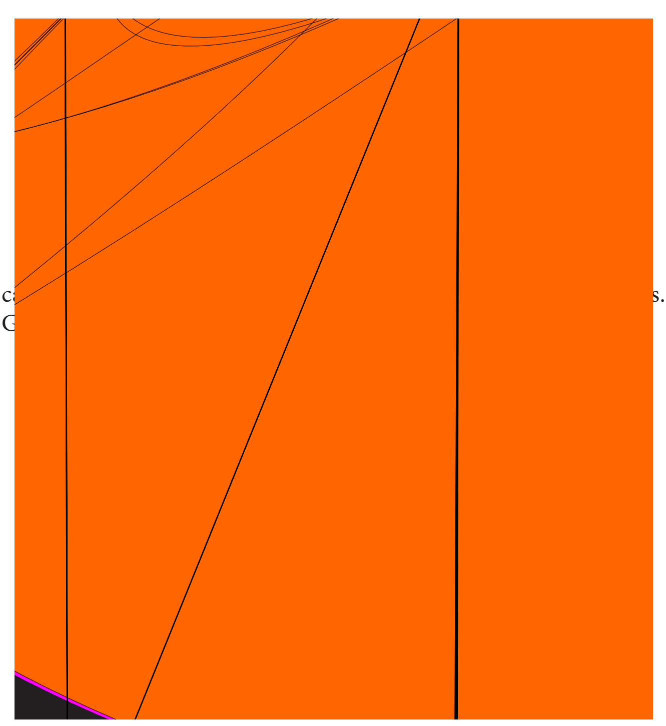
© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

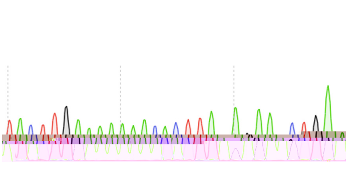
(**Supplementary Table 5**), ensuring the completeness of a collection of genome-wide gene deletions. The number of transformations can be scaled up to obtain efficiencies required for even larger library sizes. We screened our mutant library for *CA 1* mutants in the presence of L-(+)-( )-canavanine and identified all four *CA 1*-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.

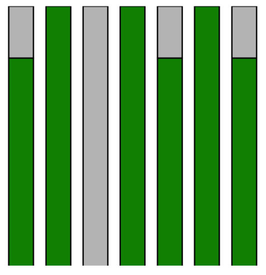
We next used CHAnGE to engineer furfural tolerance. Selection with 5 mM furfural enriched *I 1* targeting guides (**Fig. 1f** and **Supplementary Fig. 3**), in line with previous findings[8]. Guide sequences targeting newly identified genes *A 30* and *BC4*, were also enriched. All three disruption mutants 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. *CB3* 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 *CB3* mutant was dependent on *I 1* disruption; *CB3* targeting guides were not enriched in the first round of evolution, and the single *CB3* 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**).

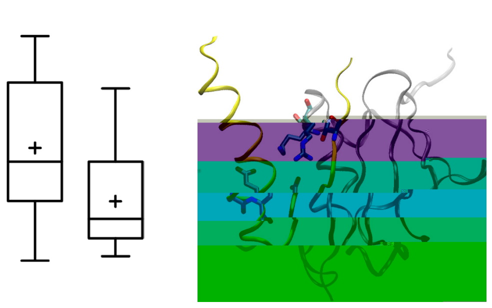
Next, we applied CHAnGE to single-nucleotide resolution editing. Exogenous Siz1 mutations (F268A, D345A, I363A, S391D, F250A/F299A, FKSΔ) were previously shown to diminish SUMO conjugation to PCNA[10,11]. We designed seven CHAnGE cassettes to introduce these seven mutations and an insertion mutation (**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 *CA 1* and *BC4*, 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**).

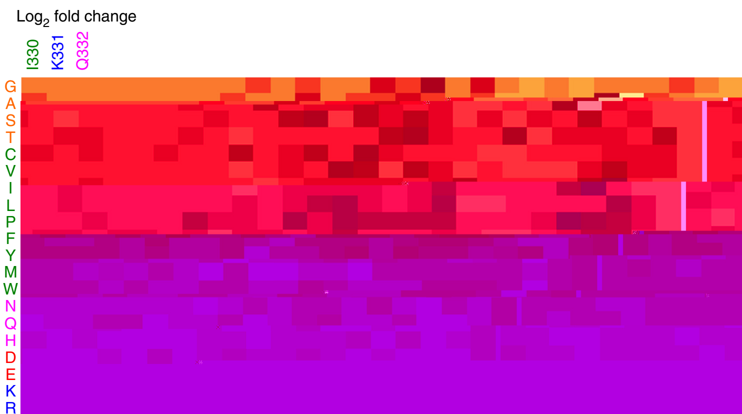
Finally, we carried out tiling mutagenesis of the Siz1 SP-CTD domain. We first modified the CHAnGE cassette to reduce the length of homology arms to 40 bp, so that the sequence between the target codon and the PAM could be accommodated (**Fig. 2d**). We designed five CHAnGE cassettes with 40-bp homology arms targeting *BC4*, and achieved an average editing efficiency of 86% (**Supplementary Fig. 17a**). To minimize the length of CHAnGE











© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

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 *E* *1* 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.

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

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

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 *E 1* 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 .
strain S288c were downloaded from SGD and passed through CRISPRdd

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

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 *I 1* 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 *B 1* 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

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

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.

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

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.

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

Nature Biotechnology: doi:10.1038/nbt.4132