MICROBIOLOGY

Base editor enables rational genome-scale functional screening for enhanced industrial phenotypes in *Corynebacterium glutamicum*

Ye Liu[1,2], Ruoyu Wang[1,2], Jiahui Liu[1,2,3], Hui Lu[1,2], Haoran Li[1,2], Yu Wang[1,2], Xiaomeng Ni[1,2], Junwei Li[1,2,3], Yanmei Guo[1,2], Hongwu Ma[1,2], Xiaoping Liao[1,2]*, Meng Wang[1,2]*

Genome-scale functional screening accelerates comprehensive assessment of gene function in cells. Here, we have established a genome-scale loss-of-function screening strategy that combined a cytosine base editor with approximately 12,000 parallel sgRNAs targeting 98.1% of total genes in *Corynebacterium glutamicum* ATCC 13032. Unlike previous data processing methods developed in yeast or mammalian cells, we developed a new data processing procedure to locate candidate genes by statistical sgRNA enrichment analysis. Known and novel functional genes related to 5-fluorouracil resistance, 5-fluoroorotate resistance, oxidative stress tolerance, or furfural tolerance have been identified. In particular, *purU* and *serA* were proven to be related to the furfural tolerance in *C. glutamicum*. A cloud platform named FSsgRNA-Analyzer was provided to accelerate sequencing data processing for CRISPR-based functional screening. Our method would be broadly useful to functional genomics study and strain engineering in other microorganisms.

INTRODUCTION

Genome-scale functional genomics screening is crucial for systematically defining genotype-phenotype associations, elucidating unknown gene function and discovering new genes for strain engineering (*1*–*3*). Canonical arrayed collections, with each mutant individually created, provide the gold standard in functional genomics profiling. However, the processes of construction and screening are laborious and costly. Traditionally, transposon-based library creation technique plays critical roles in functional genomics studies in a high-throughput manner, but because of different transposition efficiency and sequence preference, it usually causes nonuniform and incomplete gene disruptions in the genome (*2*). In addition, the transposon-based method cannot specifically target a subset of the genome, rendering it unprogrammable and less efficient. Recently, the powerful clustered regularly interspaced short palindromic repeats (CRISPR)–based functional genomics screenings have been used in prokaryotic and eukaryotic cells. The 20–base pair (bp) unique sequence (N20 sequence) of the single-guide RNA (sgRNA) can serve as a trackable barcode for identification of the corresponding target in a desired trait and can be quickly synthesized as a short oligonucleotide in parallel and quantified by next-generation sequencing (NGS). One representative strategy depends on the homology-directed repair of the double-strand breaking (DSB) generated by CRISPR-Cas system, which enables precise gene editing with DNA templates, such as CREATE and CHAnGE (*4*, *5*). However, the complex design of DNA template-guide pools, the cellular toxicity of DSB, high requirement for homologous recombination efficiency, and high homology arm length requirement make it still very challenging for application in many microorganisms. In addition, CRISPR interference (CRISPRi)–pooled screenings, which use the nuclease-deficient mutant of Cas protein for transcriptional repression of the target genes, have been used to identify gene essentiality and function (*1*, *6*, *7*). The simplicity in the sgRNA design and workflow makes it applicable in diverse microorganisms. However, CRISPRi is highly site dependent, and its moderate repression may not induce a notable phenotype. Moreover, for prokaryotic microorganisms with many operon structures, operon level repression will cause a collateral effect on the transcription of genes on both sides, which hinders precise phenotype mapping to the responsible target.

Base editor is a novel genome editing technology to introduce specific point mutations at target loci by combining the targeted specificity of CRISPR-Cas system and the catalytic activity of nucleobase deaminase (*8*). Up to now, base editors have been developed to not only induce initial base transition (C to T transition and A to G transition) (*9*–*11*) but also achieve base transversion (C to G transversion in mammalian cells and C to A transversion in bacterial cells) (*12*, *13*), enhancing their diversity and applicability. Because base editors do not generate DSBs, require DNA templates, or depend on homologous recombination, they are highly suitable for constructing genome-scale gene perturbation libraries. Recently, base editor–mediated large-scale library screenings have been reported in mammalian cells and yeast, demonstrating the excellence of this approach for functional genomics screenings in eukaryotic cells (*14*–*17*). To the best of our knowledge, although base editors have been applied in various prokaryotic microorganisms (*8*), functional genomics screenings with base editors in prokaryotes have not been exploited.

Corynebacterium glutamicum, as an important industrial workhorse, has been engineered to produce a variety of amino acids and value-added chemicals, such as organic acids, alcohols, aromatics, and proteins (*18*). *C. glutamicum* has many fundamental physiological properties that make it an outstanding model organism for industrial biomanufacturing, including generally regarded as safe, broad spectrum of carbon utilization, and flexible metabolism (*19*). In addition, more and more genetic manipulation tools have been developed in *C. glutamicum*, facilitating its strain engineering endeavors (*20*).

[1]Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, China. [2]National Technology Innovation Center of Synthetic Biology, Tianjin, China. [3]College of Biotechnology, Tianjin University of Science and Technology, Tianjin, China.
*Corresponding author. Email: liao_xp@tib.cas.cn (X.L.); wangmeng@tib.cas.cn (M.W.)

However, during the industrial fermentation, *C. glutamicum* encounters various environmental and metabolic stresses, including thermal stress, oxidative stress, end product toxicity, and chemical inhibitor, which are detrimental to the strain growth and productivity (*21*). The tolerance-related genes and mechanisms of *C. glutamicum* under different stresses are still not clear, presenting a challenge for further strain engineering.

Previously, we have established a multiplex automated *C. glutamicum* base editing method using CRISPR-Cas and activation-induced cytidine deaminase (AID) (*22*) and improved it with expanded targeting scope using Cas9 variants and adjustable editing window using sgRNAs with different lengths (*23*). Here, we have established a rational genome-scale loss-of-function screening strategy using a cytosine base editor–mediated gene inactivation library (Fig. 1A). In particular, we found that because of the fundamental difference between prokaryotes and eukaryotes and high background mutations observed in *C. glutamicum* under selective screening, previously reported data processing methods developed in *Saccharomyces cerevisiae* or mammalian cells for base editor–enabled genome-scale functional screening were not applicable in our case. In this study, we developed a new data processing procedure that used an edgeR-based statistical method to determine sgRNA enrichment, which can reduce the interference caused by background mutations and rigorously identify the sgRNAs related to the phenotypes. First, we created a genome-scale loss-of-function library, in which 3041 of the total 3099 genes (98.1%) in *C. glutamicum* ATCC 13032 genome were targeted with at least one sgRNA. Then, we proved that our method is effective by successfully identifying the genes with known

resistance functions toward 5-fluorouracil (5-FU). Novel genes related to 5-fluoroorotate (5-FOA) resistance in *C. glutamicum* were identified in our screening. In addition, to assess the performance of our method in the identification of genes related to stress tolerance, as a proof of concept, we performed screening under oxidative stress and the presence of furfural. Several related genes have been successfully identified. In particular, two genes, *purU* (*cgl0382*) and *serA* (*cgl1284*), were identified in the furfural tolerance screening and proven to be related to the furfural tolerance in *C. glutamicum*. Engineered strain with both *purU* deletion and *serA*^{Q184K} mutation has 1.93-fold higher cell biomass than the wild-type strain in the presence of furfural after 20 hours of cultivation. Last, an accompanying cloud platform (FSsgRNA-Analyzer) was provided for NGS data processing in CRISPR-based functional screening, which was built in a novel entirely cloud-based serverless architecture with high reliability, robustness, and scalability, and is capable of parallelly processing hundreds of tasks each analyzing thousands of targeting sgRNAs in minutes. Our base editor–enabled genome-scale functional screening method is specifically developed for prokaryotes and paves the way for its general application in precise and effective genome-scale (or a precise subset of) functional screening in many microorganisms.

RESULTS

Design and creation of the genome-scale gene inactivation library

The cytosine base editor can inactivate genes by precisely converting one of four specific codons (CGA, CAG, and CAA in coding strand
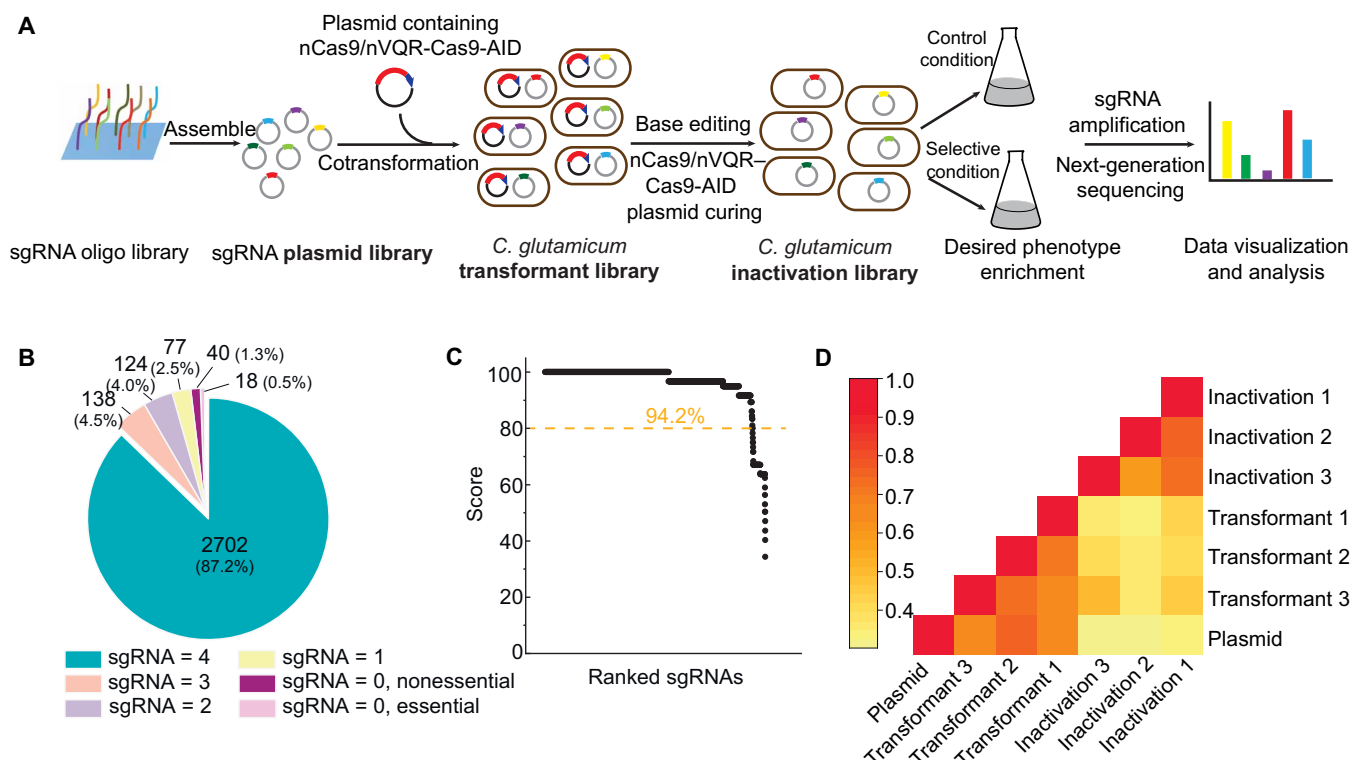

Fig. 1. Genome-scale functional screening of *C. glutamicum* with base editor–mediated inactivation library. (**A**) Workflow of this work. (**B**) Statistics of genes with different numbers of sgRNA in the sgRNA library. (**C**) Distribution of the scored sgRNAs. (**D**) Correlation analysis between different biological replicates during the construction of the inactivation libraries. Plasmid, sgRNA plasmid library; Transformant, *C. glutamicum* transformant library after cotransformation of base editor expression plasmid and sgRNA plasmid library; Inactivation, *C. glutamicum* inactivation library after base editing and base editor expression plasmid curing.

and CAA in noncoding strand) into premature stop codons (*24*). For the cytosine base editor combined with wild-type Cas9 [NGG protospacer-adjacent motif (PAM) requirement], 2856 of the total 3099 genes (88.0%) can be targeted for gene inactivation in *C. glutamicum* ATCC 13032, while other genes were inaccessible because of the limitation of NGG PAM requirement and ~5-bp editing window (*23*). To cover as many genes as possible, currently established different Cas9 variants with different PAM requirements were considered (*23*). Although xCas9 3.7 and Cas9-NG can broaden the PAM requirement to more relaxed NG PAM, base editors with VQR-Cas9 (NGA PAM requirement) and VRER-Cas9 (NGCG PAM requirement) exhibit higher editing efficiency and more stable performance in *C. glutamicum* (*23*). Therefore, two base editors were used in our study, and base editor with VQR-Cas9 was selected as the complement to base editor with wild-type Cas9, with additional 164 genes targeted, which covered more genes than the base editor with VRER-Cas9 with additional 55 genes targeted. Then, all the usable sgRNAs with the specific codons that can be edited to premature stop codons within the editing window were analyzed by our previously reported online tool called gBIG (http://gbig.ibiodesign.net). To select high-quality sgRNAs with maximized inactivation possibility and targeting specificity, criteria were designed to score each sgRNA by assessing the introduced position of the premature stop codon in the gene and the sgRNA off-target risk (table S1). In addition, sgRNAs with NGG PAMs were prioritized over those with NGA PAMs. Generally, the highest scores will be given to sgRNAs targeting the first half region of the genes with NGG PAMs and the lowest off-target risks. The top four sgRNAs with the highest scores were selected for each gene, and up to 87.2% of genes can be targeted with four sgRNAs (Fig. 1B). For genes with less than four sgRNAs available, we selected the truncated 18-nucleotide (nt) or extended 24-nt sgRNAs that can expand the target position to −15 or −21 (counting the PAM as positions 1 to 3) to inactive genes (*23*). Total 195 sgRNAs were supplemented in this way, which can additionally inactivate 16 genes inaccessible using canonical ~5-bp editing window (fig. S1D). We also selected the sgRNAs that target the cytosine of the reverse complementary sequence of ATG or TTG start codons (except for GTG start codon due to its possibility of being edited to ATG) to disturb the translation and ensured that the start codons were not followed by other in-frame ATG, GTG, or TTG. Total 82 sgRNAs were added in this way with another five genes perturbed (fig. S1D). An additional 100 nontargeting sgRNAs (NT sgRNAs) were included as controls. In total, 11,557 unique sgRNAs were designed, which target 3041 of the total 3099 genes (98.1%) in *C. glutamicum* ATCC 13032 (Fig. 1B). About 80.0% of the selected sgRNAs exhibits superior properties with a targeting region in the first half of genes and the lowest off-target risks (fig. S1). In addition, 94.2% of the scored sgRNAs had a score of more than 80 (Fig. 1C). To facilitate library creation, about 80 nt of oligonucleotide cassette for each sgRNA was designed, including priming sequences, restriction sites, and unique N20 sequence (fig. S2). All used oligonucleotide cassettes are listed in data S1.

The sgRNA plasmid library with NGG or NGA PAM requirement was constructed in *Escherichia coli* individually and cotransformed with base editor expression plasmid containing Cas9 or VQR-Cas9 into *C. glutamicum* separately. These two libraries were mixed by collecting enough transformants with 25-fold coverage of designed sgRNAs with NGG or NGA PAM requirements. Three biological replicates were generated independently by repeating *C. glutamicum*

transformation and transformant collection three times. After base editing, only sgRNA plasmids were preserved, while base editor expression plasmids were removed. The plasmid curing rate was up to 99.0 (fig. S3), which was sufficient to avoid the interference of continuous genome editing in the following screening. To track the variation of sgRNA composition from in silico design to library construction, we profiled each library via NGS (Table 1). For the final inactivation libraries after plasmid curation, above 91.0% mapping ratio of designed sgRNAs was achieved, covering more than 99.0% of the genes in the in silico library. There was a relatively good correlation between three biological replicates (Fig. 1D). In addition, to assess the editing efficiency of the final inactivation libraries, we randomly picked 10 colonies for each replicate, after scraping the cells of inactivation libraries on agar plates, and sequenced the sgRNAs and their targeting loci. As a result, we found that an average of 78.9% editing efficiency was obtained (fig. S4). In conclusion, these results demonstrated the high coverage and efficient editing of the inactivation libraries, sufficient for the following genome-scale functional screening.

Evaluation of sgRNA fitness in 5-FU resistance screening
To evaluate the performance of our genome-scale gene inactivation library in the functional screening, we first performed selectively lethal screenings to see whether target genes with known properties in *C. glutamicum* can be properly enriched. In *C. glutamicum*, *upp* (*cgl0684*) gene in pyrimidine metabolism is a commonly used selection marker, encoding a uracil phosphoribosyltransferase that is involved in the conversion of 5-FU to toxic metabolite named 5-fluoro–deoxyuridine monophosphate (dUMP) (Fig. 2A) (*22*). A mutant strain with an inactivated *upp* gene will survive in the presence of 5-FU. Therefore, for the genome-scale functional screening in the presence of 5-FU, the *upp* gene was expected to be enriched. The detailed screening procedure can be found in Materials and Methods. Briefly, the stored cells were first recovered overnight for seed culture and then transferred to the selective culture medium with 100 μM 5-FU and control culture medium without 5-FU, separately. The screening was terminated after about five to six doublings of the cells. After NGS analysis of sgRNA pools, we found that the correlations between biological replicates were poor, especially for the replicates in the selective culture condition (data S2), which was different from previous studies in *S. cerevisiae* or mammalian cells, in which replicates seem to be highly correlated (*14–17*). To investigate this phenomenon, we plated the cells of three replicates from the selective culture on agar plates, separately, and randomly picked about 50 single colonies from each replicate to determine their sgRNA sequence by Sanger sequencing. As expected, the designed sgRNA targeting *upp* gene was observed in all three replicates in the presence of 5-FU (Fig. 2B). Unexpectedly, sgRNAs targeting a variety of other genes were also observed from each replicate, but no common genes existed in three replicates, which was consistent with the poor NGS data correlations between replicates. We hypothesized that this phenomenon might be explained by two different reasons. First, new target genes with 5-FU resistance were discovered, although it seems unlikely because the number of new target genes enriched was too high. Second, background mutation, either from spontaneous mutation of *C. glutamicum* and/or elevated random mutation from overexpressing nucleobase deaminase, during the screening caused the inactivation of the *upp* gene. We sequenced the editing loci targeted by the sgRNA and the *upp* gene in the same single colonies,

Table 1. Statistics analysis of each constructed library via NGS.

Libraries	Total read counts	Mapped read counts	Read counts > 0		Read counts ≥ 20	
			sgRNA	Gene*	sgRNA	Gene†
			(ratio‡)	(ratio)	(ratio)	(ratio)
Plasmid (NGG)	10646744	9582402	7466	2837	7220	2812
			97.61%	97.76%	94.39%	96.90%
Plasmid (NGA)	7673608	6963096	3887	2024	3868	2020
			99.46%	98.54%	98.98%	98.34%
Transformant 1	10660732	10112918	11265	3036	10821	3033
			97.47%	99.84%	93.63%	99.74%
Transformant 2	11342596	10776907	11330	3035	10868	3033
			98.03%	99.80%	94.04%	99.74%
Transformant 3	9731896	9249060	11256	3039	10882	3035
			97.40%	99.93%	94.16%	99.80%
Inactivation 1	9111468	8606431	10931	3033	10528	3029
			94.58%	99.74%	91.10%	99.61%
Inactivation 2	9746060	9224673	11046	3034	10529	3029
			95.58%	99.77%	91.10%	99.61%
Inactivation 3	8551616	8088598	10929%	3033	10566	3027
			94.57%	99.74%	91.43%	99.54%

*Number of genes with at least one sgRNA that read count >0. in silico library. †Number of genes with at least one sgRNA that read count ≥20. ‡Mapping ratio to the
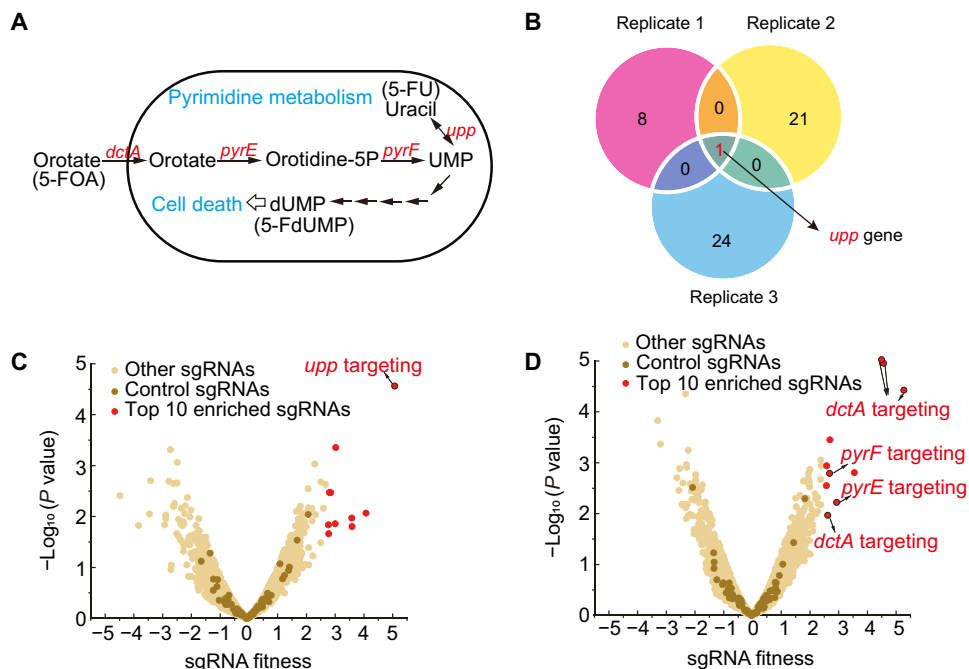

Fig. 2. Genome-scale functional screening in the presence of 5-FU or 5-FOA. (**A**) Pyrimidine metabolism in *C. glutamicum*. UMP, uridine monophosphate; 5-FdUMP, 5-fluorodeoxyuridine monophosphate. (**B**) Common enriched genes of three biological replicates determined by sgRNA Sanger sequencing of the picked colonies from the selective culture condition with 5-FU. (**C**) Volcano plot of sgRNA fitness relative to $-\log_{10}$ (P value) in the presence of 5-FU according to the edgeR-based analysis. (**D**) Volcano plot of sgRNA fitness relative to $-\log_{10}$ (P value) in the presence of 5-FOA.

and statistical results are listed in data S3. In all strains with sgRNAs not targeting the *upp* gene, mutations in the *upp* gene were always observed, explaining their resistance to 5-FU.

For NGS data processing, because of the data noise caused by background mutations and other factors, reported data processing method developed for eukaryotes is no longer suitable here, and more statistically rigorous testing methods need to be used to identify the sgRNAs that are actually relevant to the screened phenotype. Therefore, after obtaining read counts for all sgRNAs, edgeR, a widely used package for differential gene expression analysis (*25*), was used to calculate the sgRNA fitness and *P* value for each sgRNA between samples in selective and control culture conditions (details can be found in Materials and Methods). Enriched sgRNAs showing the concordant uptrends in all three comparisons between selective and control culture conditions will be reserved and ranked, reducing statistical noise caused by background mutations in the screening. For the screening results in the presence of 5-FU, the sgRNA targeting *upp* gene was successfully enriched with the highest sgRNA fitness value compared with the other sgRNAs (Fig. 2C), suggesting the reliability of our method. In an attempt to discover more genes related to 5-FU resistance, we picked the top 10 sgRNAs with the highest sgRNA fitness value (*P* < 0.05; hereafter, all ranked sgRNAs met this requirement) and reconstructed the mutant strains with corresponding genes inactivated individually. However, except for the *upp* inactivation strain, none of the other nine inactivation strains were insensitive to 5-FU (fig. S5A).

Genome-scale functional screening for 5-FOA resistance

For further evaluation of our new method, we performed the screening to discover functional genes in the presence of 5-FOA, an analog of orotate, which can be converted to the toxic 5-fluoro-dUMP by endogenous metabolite pathway (Fig. 2A). Using the similar screening and data processing workflow in the presence of 5-FU, the top 10 sgRNAs with the highest sgRNA fitness value were chosen (Fig. 2D). For the phenotype validation of the corresponding genes, inactivation strains were reconstructed individually. Strains with *dctA* (*cgl2595*), *pyrE* (*cgl2773*), or *pyrF* (*cgl1608*) inactivation had a well growth in the presence of 5-FOA, but other strains and the wild-type strain did not (fig. S5B), indicating that the functions

of *dctA*, *pyrE*, or *pyrF* were related to the 5-FOA resistance. In the pyrimidine de novo biosynthesis, *pyrE* and *pyrF* genes encode orotate phosphoribosyltransferase and orotidine-5′-phosphate decarboxylase, respectively. Inactivation of *pyrE* or *pyrF* gene renders cells resistant to 5-FOA by blocking the metabolic pathway from 5-FOA to 5-fluoro-dUMP. These genes have been developed as efficient selection markers in some microorganisms, such as *Thermus thermophilus* and Clostridia (*26–28*), but were rarely reported in *C. glutamicum*. *dctA* gene, encoding Na$^+$/H$^+$-dicarboxylate symporter in *C. glutamicum* (*29*), is speculated to be related to 5-FOA resistance by disturbing the transport of 5-FOA. Similar reports have confirmed the function of *dctA* in 5-FOA resistance in *Salmonella typhimurium* and *E. coli* (*30*), but it was first reported in *C. glutamicum* as far as we know. The functions of *dctA*, *pyrE*, or *pyrF* in the resistance to 5-FOA demonstrate their potential in serving as new selection markers in *C. glutamicum*. Hence, these results indicate that our genome-scale functional screening approach can identify targeted genes by functional evaluation of candidate genes targeted by top-enriched sgRNAs, and this approach is sensitive and reliable.

Genome-scale functional screening for higher oxidative stress tolerance

For identifying genes that may be related to oxidative stress, we performed screening in the presence of hydrogen peroxide (H$_2$O$_2$), one common type of reactive oxygen species, which is toxic to biological systems when excessively accumulated (*31*). The experiment and data processing workflow was similar to the screening in the presence of 5-FU or 5-FOA. After screening with 100 mM H$_2$O$_2$, two sgRNAs targeting *oxyR* (*cgl1925*) gene were enriched significantly with the highest sgRNA fitness value (Fig. 3A), which is in line with the previous report that its deletion leads to increased resistance to H$_2$O$_2$-induced oxidative stress in *C. glutamicum* (*32*). OxyR is a best-characterized LysR-type transcriptional regulator and plays as a transcriptional repressor of the H$_2$O$_2$-inducible antioxidant genes in *C. glutamicum*. In recent years, progress has been made for the regulatory mechanism in the oxidative stress responses in *C. glutamicum*, and more transcriptional regulators have been found involved in these responses. To the best of our knowledge, in
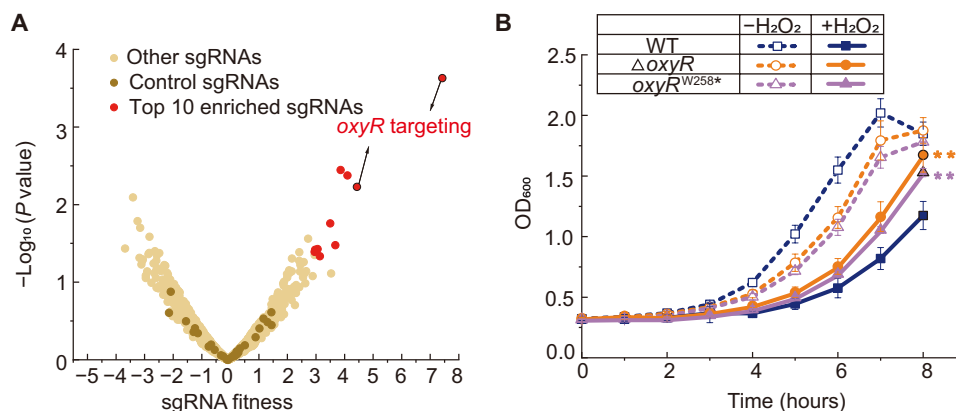

Fig. 3. Genome-scale functional screening under oxidative stress. (**A**) Volcano plot of sgRNA fitness relative to −log$_{10}$ (*P* value) under oxidative stress. (**B**) Effects of *oxyR* on oxidative stress tolerance. Δ*oxyR*, *oxyR* deletion by the suicide plasmid–mediated allele exchange; *oxyR*^{W258*}, *oxyR* inactivation for mutating tryptophan at position 258 to a stop codon by base editing. *N* = 3 independent experiments. Two-tailed *t* tests were performed to determine significance levels against the wild-type (WT) strain. **P* < 0.01.

addition to OxyR, deletion of several regulators, including multiple antibiotic resistance–type regulator CosR (33), xenobiotic response element–type regulator MsrR (34), and OsnR (35), has also been reported to increase resistance to H₂O₂ in *C. glutamicum*. Unexpectedly, the sgRNAs targeting these genes have not been enriched in our study. To investigate this potential contradiction toward literature, we first inactivated these four reported genes using base editor individually. Just to rule out any interference from base editing method itself, we also deleted these four genes individually using the suicide plasmid–mediated allele exchange method. In the control culture without H₂O₂, the mutants showed similar [for *osnR* (*cgl2920*) and *msrR* (*cgl2776*) genes] or reduced growth [for *cosR* (*cgl2711*) and *oxyR* genes] compared with the wild-type strain (fig. S6). However, for the selective culture in the presence of 100 mM H₂O₂, only *oxyR*-inactivated or *oxyR*-deleted mutants displayed enhanced growth (Fig. 3B). Mutants of the other three genes inactivated or deleted showed reduced growth in the selective culture, inconsistent with the previous reports. It is likely because our screening media and conditions are not exactly the same as reported in literature, and the used strains in previous reports are not *C. glutamicum* ATCC 13032. We also observed that there was little difference in growth of gene-inactivated and gene-deleted strains, proving the effectiveness of base editor in loss-of-function screening. The specific growth rates during exponential growth of the *oxyR*-deleted mutants were 1.28-fold higher than the wild-type strain. The cell biomass of the *oxyR*-deleted mutants after 8 hours of cultivation was 1.43-fold higher than the wild-type strain. These results supported the screening results using our genome-scale gene inactivation library, suggesting the reliability of our method. To discover more genes that may be involved in H₂O₂ tolerance, the genes corresponding to the top 10 enriched sgRNAs were inactivated using individually constructed base editor. However, except for *oxyR*-inactivated strain, none of the mutants showed improved growth compared with the wild-type strain (fig. S7).

Genome-scale functional screening for enhanced furfural tolerance

To demonstrate the applicability of our method in strain engineering, we first performed screening in the presence of furfural, one of the major toxic inhibitors on microbial growth on pretreated lignocellulose. *C. glutamicum* is considered as one of the most promising cell factories for producing fuels and chemicals from lignocellulose feedstock because of its capabilities of converting furfural into the less toxic compounds, furfuryl alcohol and furoic acid (36, 37). However, few genes have been reported to be involved in the detoxification of furfural in *C. glutamicum*, which renders strain engineering to further improve furfural tolerance challenging. In our study, after the selective screening in the presence of 20 mM furfural, the top 10 enriched sgRNAs were picked and the mutants were constructed individually with the corresponding genes inactivated (Fig. 4A). Under furfural stress, the growth of all mutants was hindered obviously. Compared with the wild-type strain, we observed that the mutant with *purU* gene inactivated showed improved furfural tolerance, which entered the exponential growth phase earlier than the wild-type strain (Fig. 4B). The mutant with *purU* gene deleted was also constructed, and there was no obvious growth difference compared with the mutant obtained by base editing (fig. S8A). After 20 hours of cultivation, the cell biomass of the inactivated mutant was 1.32-fold higher than the wild-type strain (Fig. 4B), which further confirmed that *purU* was related to furfural tolerance. *purU* gene encodes a formyltetrahydrofolate hydrolase in *C. glutamicum*, which reversibly catalyzes the 10-formyltetrahydrofolate to formate and tetrahydrofolate. As far as we know, it was the first report indicating that the function of *purU* gene relates to furfural tolerance, and the mechanism is still not known. Considering that formate and furfural both have an aldehyde group, we speculated that the enzyme encoded by *purU* may catalyze furfural and tetrahydrofolate to form an unknown toxic compound. The hypothesis may be supported by the result that mutant with overexpression of *purU* displayed a slight decrease in the growth compared with the wild-type strain in the presence of furfural (fig. S8C). Further investigation is needed to explain the mechanism in the future.

There is another sgRNA enriched significantly with the highest sgRNA fitness value, which targeted *serA* gene, encoding a D-3-phosphoglycerate dehydrogenase (PGDH). During the construction of the *serA*-inactivated strain by base editor, we observed that the target C in the specific codon CAG not only can be edited to T to form a premature stop codon but also can be edited to A to
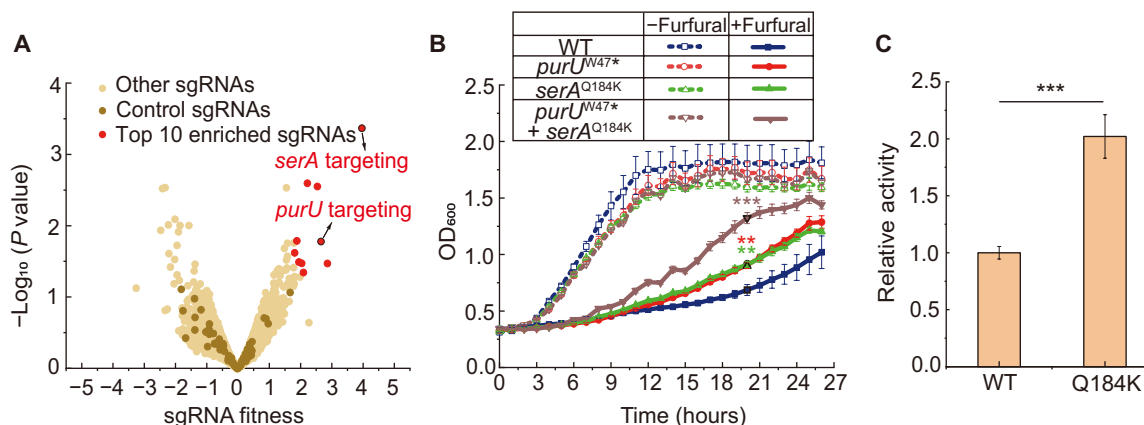

Fig. 4. Genome-scale functional screening in the presence of furfural. (**A**) Volcano plot of sgRNA fitness relative to −log₁₀ (*P* value) in the presence of furfural. (**B**) Effects of *purU* and *serA* on furfural tolerance. *purU*^W47*, *purU* inactivation for mutating tryptophan at position 47 to a stop codon by base editing; *serA*^Q184K, an amino acid transition Q184K in *serA* introduced by base editing. (**C**) Enzyme activity assay of PGDH using purified protein. WT, wild-type PGDH; Q184K, PGDH with amino acid transition Q184K. *N* = 3 independent experiments. Two-tailed *t* tests were performed to determine the significance levels against the wild-type strain. **P < 0.01 and ***P < 0.001.

introduce an amino acid transition Q184K (fig. S9). Compared with the wild-type strain, the strain with *serA*^{Q184K} mutation exhibited improved tolerance with a relatively short lag phase (Fig. 4B), while the mutant with *serA* inactivated showed decreased tolerance to furfural (fig. S8B). We also verified the above results by constructing the strain with *serA*^{Q184K} mutation and strain with *serA* deleted using the suicide plasmid–mediated allele exchange method, which showed almost the same cell growth as the mutant strains obtained with base editor (fig. S8B). After 20 hours of cultivation, the cell biomass of strain with *serA*^{Q184K} mutation was 1.33-fold higher than the wild-type strain. PGDH catalyzes the initial reaction in the L-serine biosynthetic pathway, converting D-3-phosphoglycerate (PGA) to phosphohydroxypyruvate concomitant with the formation of reduced form of nicotinamide adenine dinucleotide (NADH) (*38*, *39*). According to a previous report, both NADH and reduced form of nicotinamide adenine dinucleotide phosphate (NADPH) were required as cofactors during the detoxification process of furfural. Increased cellular NADH or NADPH pools might be beneficial for improving tolerance against furfural (*37*). Therefore, we hypothesized that the introduction of Q184K mutation in the *serA* gene may improve the catalytic activity of PGDH, which increased the formation of NADH. To verify our hypothesis, the wild-type PGDH and the PGDH harboring the Q184K mutation were heterogeneously expressed in *E. coli* and purified. As expected, a higher catalytic activity was observed with the PGDH^{Q184K} using PGA as substrate, which was 2.02-fold higher than the wild-type protein (Fig. 4C). In addition, the decreased furfural tolerance of the *serA*-inactivated strain supported the fact that PGDH plays a vital role in the detoxification process of furfural. However, the furfural tolerance of the strain was not enhanced by overexpression of *serA*, which needs further investigation (fig. S8C). For the other genes targeted by the top 10 enriched sgRNAs, strains were constructed individually with corresponding genes inactivated using the base editing method. However, no new target genes were discovered with improved furfural tolerance (fig. S10).

Then, *C. glutamicum* was engineered by combining the inactivation of *purU* gene and Q184K mutation of *serA* gene to further enhance its tolerance to furfural. In the absence of furfural, we observed that there was no obvious growth difference between different mutants (Fig. 4B). Under furfural stress, strains harboring the double mutations showed further improved tolerance compared with the strains with single mutations. The cell biomass of double mutant after 20 hours of cultivation was 1.93-fold higher than the wild-type strain, suggesting that the impact of these beneficial mutations was additive. Overall, these results demonstrate the great potential of our method in rapidly identifying new target genes for genetic engineering of microbial strains for enhanced performance.

FSsgRNA-Analyzer: A cloud platform for NGS data processing in CRISPR-based functional screening

For the sgRNA enrichment analysis, most of the available toolkits are programming language–based and require bioinformatics skills, which are not very convenient for experimental biologists (*40*). In addition, previous toolkits and cloud platforms such as PinAPL-Py (*41*) and CRISPRcloud (*42*) cannot be used directly in our study because of the different methods used for sgRNA statistical analysis. To facilitate the use of our new NGS data analysis workflow for experimental biologists, a cloud platform named FSsgRNA-Analyzer was developed (https://fssgrna.biodesign.ac.cn/), which is not only

applicable for our base editor–mediated functional screening but also valuable for other CRISPR-based functional screening. The cloud platform is based on a novel serverless architecture (Fig. 5A), enabling high reliability, robustness, and scalability. After users upload raw NGS data and target sgRNA data, FSsgRNA-Analyzer can execute the primary analysis for NGS data (read cleaning, mapping, and sgRNA counting) and then perform the secondary analysis (correlation or differential analysis) among samples under different culture/selection conditions (Fig. 5B). In general, the whole workflow can be completed in a few minutes. We use the Amazon DynamoDB to store job information, and users can track the past submission records and corresponding status information. Once the analysis is finished, the user can view or download the related results.

DISCUSSION

Here, we developed a cytosine base editor–enabled genome-scale loss-of-function screening strategy in *C. glutamicum* and demonstrate its powerful potential in the exploration of gene function and genetic engineering of strains. Our library can specifically target 98.1% of genes for inactivation in *C. glutamicum* ATCC 13032, with only 40 nonessential genes and 18 essential genes uncovered, which is superior to the previous reported single-gene disruption library constructed by random insertion of transposon into the *C. glutamicum* R genome, with only a coverage of nearly 80% of genes by analyzing 18,000 mutants (*43*). In addition, for nonessential genes without editable sites in *C. glutamicum* ATCC 13032 by base editor, we actually inactivated these genes using the suicide plasmid–mediated allele exchange method individually by introducing two successive premature stop codons (data S4) and verified the abovementioned phenotypes of these genes separately, ensuring the full coverage of nonessential genes in our study, and none of these genes are related to the abovementioned phenotypes (fig. S11). Although different DSB-mediated genome editing tools have been established in *C. glutamicum* (*20*), it is still challenging for their application in genome-scale library construction because of the toxicity of DSBs and inefficient homologous recombination. Unlike *S. cerevisiae*, the homology arm used in *C. glutamicum* is usually longer than 0.5 kb each, which is beyond current high-throughput DNA synthesis capability (*4*, *44*, *45*). On the contrary, for base editing, large-scale sgRNA pools can be quickly designed and synthesized with established sgRNA design tools, mature DNA synthesis technology, and low cost. Moreover, base editing technology has been established in many prokaryotic microorganisms, such as *E. coli*, *Bacillus subtilis*, *Streptomycetes*, and Clostridia (*8*, *46*, *47*). Given its superior simplicity and high editing efficiency, our method would generally be applied to other microorganisms.

In this study, several screening examples are presented to demonstrate the excellent capability of our method to enrich target genes in the screening with toxic chemical (5-FU or 5-FOA) tolerance, screening under environmental stress (oxidative stress), and screening in the presence of industrial inhibitors (furfural). For sgRNA analysis, we developed a new statistical data processing procedure, which can rigorously select the sgRNAs really related to the phenotypes and reduce the interference caused by background mutations. Notably, the top one selected sgRNA was always the one that we want to select in four cases, further showing the effectiveness of this procedure. We considered the genes with top-ranked sgRNA
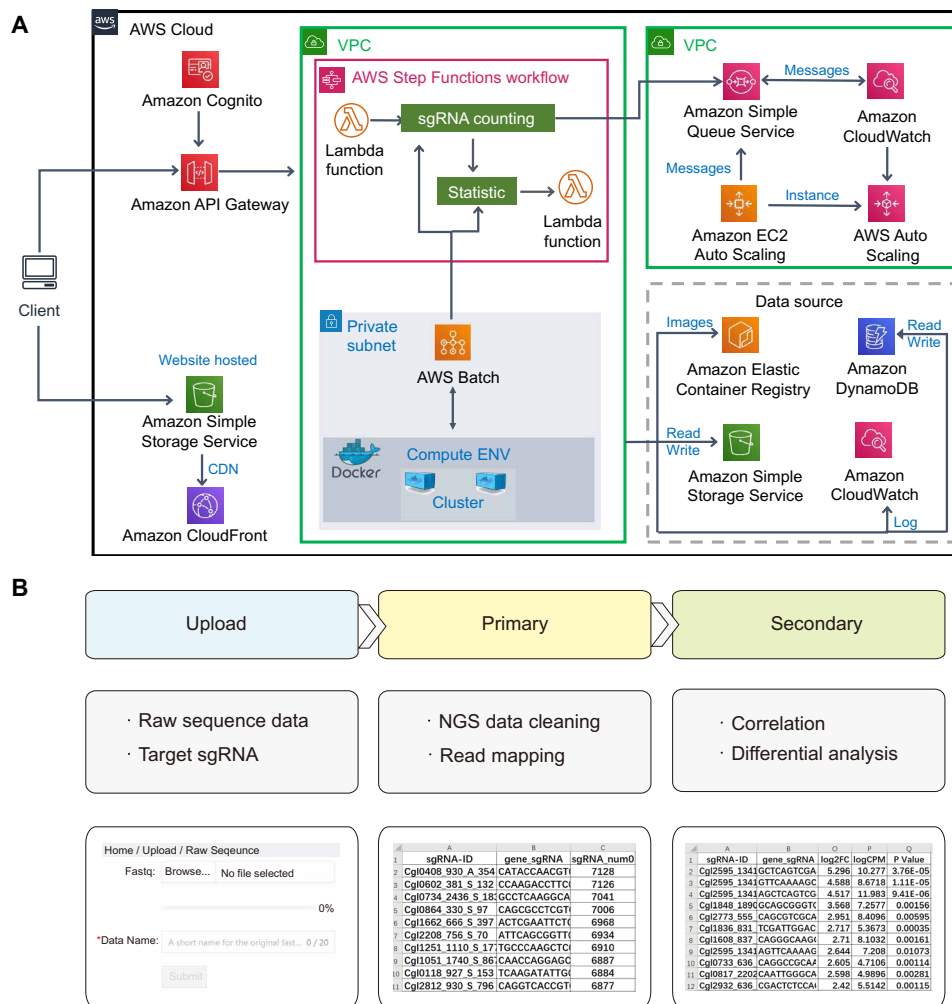
Fig. 5. Overview of FSsgRNA-Analyzer. (**A**) The architecture of FCsgRNA-Analyzer online service based on Amazon web services. (**B**) The workflow of FCsgRNA-Analyzer. CDN, content delivery network; VPC, virtual private cloud; ENV, environment.

targeted as candidates and exactly validated the function of each candidate by gene inactivation or deletion individually. We found that not all designed sgRNAs for each functional gene can be enriched in the screening. It may be explained by the following reasons. First, sgRNAs of samples in the control condition were filtered out because their read counts were below the minimum threshold, such as the other two sgRNAs targeting *oxyR* gene. Second, the stringent rule that a sgRNA needs to be enriched in all three replicates so that it would be considered valid prevents the sgRNAs with apparent enrichment in two replicates but slight depletion in the third replicate from being considered, such as the other sgRNA targeting *upp* gene. Third, some sgRNAs obviously lead to very low base editing efficiencies, which may be the reason why those sgRNAs cannot be enriched. On the other hand, we found that not all genes targeted by the top ranked sgRNAs showed the functions associated with the corresponding phenotypes. Combined with the data analysis results and the verification experiments, we observed that candidate genes with more than one sgRNA enriched (even not all sgRNA are ranked top 10) are more likely to be functionally relevant. In addition, sgRNAs with low abundance in the control condition tend to be false-positive interference

with large sgRNA fitness values. Deeper considerations are needed to reduce false-positive interference further while ensuring the enrichment of functional genes.

Recently, engineered base editors have been reported with improved performance, which may facilitate the development of base editor–mediated high-throughput functional genomics study. In particular, engineered base editors combined with new Cas variants with broad PAM compatibility, such as SpG (NGN PAMs) and SpRY (NRN>NYN PAMs) (*48*), will expand the targeting scope in the genome to increase the coverage of genes and reduce complexity in library creation. Established base editors with improved editing precision, including increased editing product purity, reduced off-target activity, and narrowed editing window, will be desirable for enhancing the predictability of the editing results and analysis of genotype-phenotype associations (*8, 49, 50*). In addition, our current functional screening phenotypes are all cell growth based. To expand the screenable phenotypes toward chemical or protein production, which is highly desired in the field of synthetic biology, fluorescence-activated cell sorting and fluorescence-activated droplet sorting can be readily combined with our base editor–mediated library. Moreover, genome-scale functional screening with multigene perturbation would

be within reach by combining base editor with other orthogonal genome perturbation tools, such as RNA interference and CRISPRi to investigate more complex phenotypes in the future.

MATERIALS AND METHODS
Bacterial strains and cultural conditions
Strains and plasmids used in this study are listed in Table 2. *E. coli* strain DH5α was used for molecular cloning and manipulation of plasmids. *E. coli* strain BL21 (DE3) was used as host for gene expression and protein purification. All *E. coli* strains were cultivated aerobically at 37°C in Luria-Bertani (LB) broth. Kanamycin (Km; 50 μg/ml) or chloramphenicol (Cm; 20 μg/ml) was added as required. *C. glutamicum* strain ATCC 13032 used as the initial host was cultivated aerobically at 30°C in LBHIS broth [tryptone (5 g/liter), NaCl (5 g/liter), yeast extract (2.5 g/liter), brain heart infusion (18.5 g/liter), and sorbitol (91 g/liter)] (51) or modified CGXII minimal medium

[$(NH_4)_2SO_4$ (20 g/liter), urea (5 g/liter), KH_2PO_4 (1 g/liter), K_2H-$PO_4 \cdot 3H_2O$ (1.3 g/liter), 3-morpholinopropanesulfonic acid (42 g/liter), $CaCl_2$ (0.01 g/liter), $MgSO_4 \cdot 7H_2O$ (0.25 g/liter), $FeSO_4 \cdot 7H_2O$ (0.01 g/liter), $MnSO_4 \cdot H_2O$ (0.01 g/liter), $ZnSO_4 \cdot 7H_2O$ (1 mg/liter), $CuSO_4$ (0.2 mg/liter), $NiCl_2 \cdot 6H_2O$ (0.02 mg/liter), protocatechuic acid (0.03 g/liter), biotin (0.2 mg/liter), thiamine (0.1 mg/liter), and glucose (5 g/liter)] (52) unless otherwise indicated. Agar was added at 18 g/liter for plates. Km (25 μg/ml), Cm (5 μg/ml), or isopropyl-β-D-thiogalactopyranoside (IPTG; 1 mM) was added as required for *C. glutamicum* derivatives.

Design and synthesis of the genome-scale sgRNA library
Open reading frame (ORF) annotations of 3099 genes were obtained from National Center for Biotechnology Information (GenBank, BA000036.3). The essential genes in *C. glutamicum* strain ATCC 13032 were determined by comparing the reported essential genes in *C. glutamicum* strain R (43). All possible sgRNAs with "NGG" or

Table 2. Strains and plasmids used in this study.

Strain or plasmid	Description*	Reference or source
Strain		
E. coli DH5α	General cloning host	TaKaRa
E. coli BL21 (DE3)	Gene overexpression host	Novagen
C. glutamicum ATCC 13032	Wild-type strain	ATCC
Plasmid†		
pK18*mobsacB*	Gene deletion/integration vector, *mob*, *sacB*, Km[R]	(54)
pET28a (+)	Expression vector of *E. coli*, IPTG-inducible promoter P_{T7}, Km[R]	Novagen
pnCas9(D10A)-AID[TS]	pXMJ19 carrying nCas9(D10A)-AID cassette driven by IPTG-inducible promoter P_{tac}, Cm[R], temperature-sensitive	(22)
pnVQR-Cas9(D10A)-AID[TS]	pnCas9(D10A)-AID[TS] derivative. D1135V, R1335Q, and T1337R mutations were introduced into nCas9(D10A).	(23)
pgRNA-*ccdB* (Bsa I)	pEC-XK99E derivative, carrying gRNA cassette driven by constitutive promoter P_{11F}, Km[R]. The 20-bp guide sequence was replaced with a Bsa I–*ccdB*–Bsa I cassette to facilitate easy assembly of guide sequence by Golden Gate assembly.	(22)
pgRNA-*ccdB* (Bbs I)	pgRNA-*ccdB* (Bsa I) derivative. The Bsa I–*ccdB*–Bsa I cassette was replaced with a Bbs I–*ccdB*–Bbs I cassette	This study
pnCas9(D10A)-AID-gRNA-*ccdB*[TS]	pnCas9(D10A)-AID[TS] derivative, carrying gRNA-*ccdB* cassette from pgRNA-*ccdB* (Bsa I)	(23)
pnVQR-Cas9(D10A)-AID-gRNA-*ccdB*[TS]	pnCas9(D10A)-AID-gRNA-*ccdB*[TS] derivative. D1135V, R1335Q, and T1337R mutations were introduced into nCas9(D10A).	(23)
pK18*mobsacB*-△*oxyR*	pK18*mobsacB* derivative carrying homologous arms for *oxyR* deletion	This study
pK18*mobsacB*-△*cosR*	pK18*mobsacB* derivative carrying homologous arms for *cosR* deletion	This study
pK18*mobsacB*-△*osnR*	pK18*mobsacB* derivative carrying homologous arms for *osnR* deletion	This study
pK18*mobsacB*-△*msrR*	pK18*mobsacB* derivative carrying homologous arms for *msrR* deletion	This study
pK18*mobsacB*-△*purU*	pK18*mobsacB* derivative carrying homologous arms for *purU* deletion	This study
pK18*mobsacB*-△*serA*	pK18*mobsacB* derivative carrying homologous arms for *serA* deletion	This study
pK18*mobsacB*-*serA*[Q184K]	pK18*mobsacB* derivative carrying homologous arms for introduction of Q184K mutation in *serA* gene	This study
pET28a-*serA*	pET28a(+) derivative, carrying *serA* gene driven by IPTG-inducible promoter P_{T7}, N-terminal His Tag, Km[R]	This study
pET28a-*serA*[Q184K]	pET28a-*serA* derivative, carrying a Q184K mutation in *serA* gene	This study

*Km[R] and Cm[R] represent resistance to kanamycin and chloramphenicol, respectively. †Plasmids with sgRNA expression cassette for changed guide sequences are not listed here. A full list of primers used for construction of sgRNA expression plasmids is shown in data S5.

"NGA" PAM requirement for genome-wide gene inactivation were analyzed using gBIG online tool (www.ibiodesign.net/gBIG). Each sgRNA was scored according to the criteria in table S1 by assessing the PAM position, PAM type, and sgRNA risk. To select all needed sgRNAs, first, the top four sgRNAs with the highest scores for each gene were collected. Second, for genes with less than four sgRNAs available, the truncated 18-nt or extended 24-nt sgRNAs that target position −15 or −21 to inactive genes were supplemented. Third, the sgRNAs that can disturb start codons were supplemented. For sgRNAs generated in steps 2 and 3, in-house Perl scripts and R codes were used for assistance. Fourth, the 100 NT sgRNAs were added, which were created using R2oDNA Designer (53). The final library contains 11,557 unique sgRNA sequences targeting 3041 genes, containing 7649 sgRNAs for NGG PAM requirement and 3808 sgRNAs for NGA PAM requirement. Priming sequences and restriction sites were added to the 5′ and 3′ ends of each sgRNA for polymerase chain reaction (PCR) amplification and Golden Gate assembly, which were designed differently to distinguish libraries with NGG or NGA PAM requirement (fig. S2). The final sgRNA oligonucleotide library was synthesized on a chip by Twist Bioscience (USA).

Construction of the sgRNA plasmid library
The sgRNA expression plasmid pgRNA-*ccdB* (Km resistance) with Bsa I sites reported in previous study was used for construction of sgRNA plasmid library with NGG PAM requirement (22). The Bsa I–*ccdB*–Bsa I cassette in the plasmid was replaced by a Bbs I–*ccdB*–Bbs I cassette, generating the new plasmid for construction of sgRNA plasmid library with NGA PAM requirement. The synthesized oligonucleotide library used as template was amplified separately with primer pairs specific to designed sequences for NGG or NGA PAM requirement. Then, the about 80-nt gel-purified PCR products were cloned into pgRNA-*ccdB* with Bsa I or Bbs I sites correspondingly using Golden Gate assembly. After transforming into *E. coli* strain DH5α competent cells, about 1×10^5 to 2×10^5 colonies were collected separately for sgRNA plasmid library with NGG or NGA PAM requirement. The colonies were washed from the agar plates by adding about 10 ml of sterile water. Then, the plasmids were extracted, and NGS was performed following the description to assess the plasmid library quality.

Generation of the genome-scale gene inactivation library in *C. glutamicum*
The temperature-sensitive nCas9-AID expression plasmid, pnCas9(D10A)-AID^TS (Cm resistance) (22), and the sgRNA plasmid library with NGG PAM requirement were cotransformed into *C. glutamicum* via electroporation. Cells were spread on LBHIS plates supplemented with Km and Cm after 2 to 3 hours of recovery. After 2 to 3 days of cultivation at 30°C, about 1.8×10^5 colonies for each biological replicate were obtained independently, which represents a 25-fold coverage of the designed sgRNAs library with NGG PAM requirement. Three biological replicates were generated. Similar to the above method, the temperature-sensitive nVQR-Cas9-AID expression plasmid, pnVQR-Cas9(D10A)-AID^TS (23), and the sgRNA plasmid library with NGA PAM requirement were cotransformed into *C. glutamicum*; then, about 9×10^4 colonies were obtained independently for each biological replicate, where the fold of the coverage was the same as the transformant library with NGG PAM requirement. Each transformant library biological replicate was generated by scraping colonies of each transformant library with NGG PAM requirement and library with NGA PAM requirement together according to the same proportion. After mixing, about 15 ml of sterile water was added to resuspend the cells. NGS of the each transformant library was performed following the description. For gene inactivation by base editing, the transformant cells were incubated into LBHIS medium supplemented with Km and Cm and cultivated overnight at 30°C. Then, the cells were transferred to fresh LBHIS supplemented with Km, Cm, and IPTG with an initial optical density at 600 nm (OD$_{600}$) of 0.5. After about 12 hours of cultivation with an OD$_{600}$ of 10 reached, the cells were collected and washed twice with sterilized 0.85% NaCl solution. To obtain the final inactivation library with nCas9-AID or nVQR-Cas9-AID expression plasmid curing, the washed cells after base editing were incubated into LBHIS medium only with Km with an initial OD$_{600}$ of 0.1 and cultivated at 37°C for 24 hours. The plasmid curing rate was calculated by counting colonies with Cm sensitivity after cell dilution and cultivation on LBHIS with Km and LBHIS with Km and Cm plates. The editing efficiency of the library was roughly evaluated by two-step Sanger sequencing. Ten colonies were picked from the agar plate for each biological replicate. First, the edited targets were determined by sgRNAs sequencing. Second, the editing results were obtained by target gene sequencing. NGS was also performed to validate the quality of the library. The final gene inactivation libraries were stored as glycerol stocks in a −80°C freezer.

Screening experiments
All screening experiments were performed in shake flask with 50 ml of liquid medium in a 250-ml baffled flask. For seed culture, the cells from glycerol stocks were recovered overnight in medium without growth inhibitor at 30°C. Then, the cells were transferred to fresh medium with or without growth inhibitor at an initial OD$_{600}$ of 0.0625. For each screening, we defined the culture with growth inhibitor as the selective culture and defined the culture without growth inhibitor as the control culture. After cultivation at 30°C, the cells were collected for further analysis when the OD$_{600}$ of culture reached 2. For 5-FU screening, CGXII medium supplemented with 100 μM 5-FU was used. For 5-FOA screening, CGXII medium supplemented with 5-FOA (400 μg/ml) and uracil (40 μg/ml) was used. For H$_2$O$_2$ tolerance, CGXII-TY [modified CGXII medium supplemented with tryptone (1 g/liter) and yeast extract (0.5 g/liter)] medium supplemented with 100 mM H$_2$O$_2$ was used. For furfural tolerance, LBHIS medium supplemented with 20 mM furfural was used.

NGS and data processing
For all samples, DNA fragments containing the sgRNA region (about 160 bp) were amplified by PCR using primers noted in data S5. For the sgRNA plasmid library, the extracted plasmids were used as the templates. For the transformant library and the final gene inactivation library, the extracted genomic DNAs of each biological replicate were used as the templates. For the screening experiments, the extracted genomic DNAs of each biological replicate in selective or control culture condition were used as the template. The NEBNext Ultra DNA Library Prep Kit was used to convert the amplicon into indexed libraries for NGS on the Illumina platform. Library construction and sequencing were performed by Novogene (Tianjin, China). Approximately 1 gigabyte of data were analyzed for each sample.

For NGS data processing, raw data (raw sequencing reads) in FASTQ format were first processed through in-house Perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adapter, reads containing N base, and low-quality reads from raw data. All the downstream analyses were based on the clean data with high quality. Clean data (trimmed reads) in FASTQ format were then transformed into FASTA format, and in-house Perl and R codes were used to match the reads with the target sgRNAs and compute read counts of target sgRNAs in each library. sgRNAs with <20 read counts in the initial library for each experiment were removed to increase statistical robustness. Statistical analysis of each sample during screening via NGS is listed in data S6. Then, for each experiment, correlations between biological replicates were calculated using in-house Python code.

Statistical analysis and sgRNA ranking

For each screening experiment, first, read counts of all selective and control culture conditions were combined. Only sgRNAs with positive counts in all samples were kept. In addition, we required the average read counts for all three replicates in the control condition to be larger than 50. All read counts were processed with the edgeR package using a generalized linear model to compute the $\log_2 FC$ (fold change) between samples in selective and control culture conditions according to Eq. 1

$$\log_2 FC = \log_2 \left(\frac{(\text{Read count})_{\text{selective}}}{(\text{Read count})_{\text{control}}} \right) \quad (1)$$

Second, the median of $\log_2 FC$ for all NT sgRNAs was calculated and applied to determine the sgRNA fitness in each sample according to Eq. 2

$$\text{sgRNA fitness} = \log_2 FC_{\text{sgRNA}} - \text{median}(\log_2 FC_{\text{NT sgRNA}}) \quad (2)$$

Third, to ensure the enrichment reliability of the analysis and reduce noise caused by spontaneous mutations in the screening, for sgRNAs with $\log_2 FC > 1$, only sgRNAs that show concordant uptrends in all three comparisons between selective and control samples were reserved. In this way, those sgRNAs showing high/low trend in only one or two comparisons were omitted. Last, those remaining sgRNAs were ranked according to the sgRNA fitness, and the top 10 significantly enriched sgRNAs were chosen for further experimental validation (data S7).

The architecture of cloud platform

For the sgRNA workflow, we used Amazon ECR to store Docker images, which packaged a set of bioinformatics software, such as in-house Perl, Python, and R scripts. We build a scalable, elastic, and easily maintainable batch engine using AWS Batch. This solution took care of our dynamically scaled compute resources in response to the number of runnable jobs in our job queue. Last, we used AWS Step Functions to coordinate the components of our applications, process messages passed from AWS API Gateway, and invoke the workflows asynchronously. We used AWS S3 to exchange data between jobs and store the result files. AWS API Gateway was used as the API server to handle the HTTP requests and route traffic to the correct backends. The static website was hosted by AWS S3 and sped up with AWS CloudFront.

Strain construction for phenotypic validation

The identified candidate genes in all screening experiments were individually inactivated by base editing or deleted by the suicide plasmid–mediated allele exchange method. Primers used here are listed in data S5. For base editing, the all-in-one tool plasmid pnCas9(D10A)-AID-gRNA-$ccdB^{\text{TS}}$ or pnVQR-Cas9(D10A)-AID-gRNA-$ccdB^{\text{TS}}$ was used (23). A pair of 24-bp primers containing the enriched sgRNA target sequence was designed and annealed to form a double-strand DNA (dsDNA) with cohesive ends. Then, the base editing plasmid was constructed by replacing the sgRNA-$ccdB$ cassette with the dsDNA using Golden Gate assembly and transformed into *C. glutamicum* via electroporation. The followed procedures of base editing and plasmid curing are consistent with the previous description in generation of the genome-scale gene inactivation library. For gene deletion, the suicide plasmid pK18$mobsacB$ was used, which contains the Km resistance gene (km^R) for the first round selection and the sucrose-lethal gene ($sacB$) for the second round selection (54). Two homologous arms (about 1000 bp) flanking at both sides of the target fragment were inserted into plasmid pK18$mobsacB$. The resultant plasmid was electroporated into *C. glutamicum*, and the cells were spread on LBHIS plates supplemented with Km for the first round selection. The cells with km^R integrated into the chromosome would survive. To further discard the false positives, colony PCR was performed using one primer located upstream of target sequence in chromosome and the other located downstream of homologous arms in the pK18$mobsacB$-derived plasmid. Only the colonies with correct PCR band were reserved and inoculated into liquid LBHIS medium for overnight culture. Next, the culture cells were spread on LAS agar plates [tryptone (10 g/liter), yeast extract (5 g/liter), sucrose (150 g/liter), and agar (18 g/liter)] for the second round selection. The mutant cells with the integrated plasmid eliminated would survive. To further discard the false positives caused by the spontaneous inactivation of $sacB$, colonies without Km resistance were selected by inoculating a single colony from LAS plate on LBHIS plate or LBHIS plate supplemented with Km, separately. Last, the correct mutants were validated by PCR and DNA sequencing of the target region.

For 5-FU or 5-FOA resistance verification, the mutants were incubated overnight in liquid medium at 30°C [CGXII medium for 5-FU resistance and CGXII medium supplemented with uracil (40 μg/ml) for 5-FOA resistance]; then, the cells were diluted and spread on agar plates supplemented with growth inhibitor [CGXII medium supplemented with 100 μM 5-FU for 5-FU resistance and CGXII medium supplemented with uracil (40 μg/ml) and 5-FOA (400 μg/ml) for 5-FOA resistance]. For H_2O_2 and furfural tolerance, the mutants were incubated at 48-well plates at 30°C for seed culture (CGXII-TY medium for H_2O_2 tolerance and LBHIS medium for furfural tolerance). Then, the cells were transferred to fresh medium with growth inhibitor (CGXII-TY medium supplemented with 100 mM H_2O_2 or LBHIS medium supplemented with 20 mM furfural) at 48-well plates at an appropriate OD_{600} (OD_{600} of 0.1 for H_2O_2 tolerance and OD_{600} of 0.2 for furfural tolerance) and cultivated at 30°C with shaking at 800 rpm using a MicroScreen HT shaker (Jieling, China). OD_{600} was automatically measured in 1-hour intervals.

For nonessential genes without editable sites by base editing, these genes were inactivated by introducing two successive premature stop codons in the first half of the ORFs using the suicide plasmid–mediated allele exchange method. The phenotypic validation of each mutant was the same as described above.

Protein expression and purification

serA and *serA*^Q184K were amplified using genomic DNAs of wild-type *C. glutamicum* strain and strain with *serA*^Q184K mutation as templates, respectively. The PCR product was inserted into the pET-28a (+) with an N-terminal His tag using the ClonExpress II One Step Cloning Kit (Vazyme Biotech, China). The resultant plasmid was transformed into *E. coli* BL21 (DE3) for heterogeneous expression. The recombinant strains were cultivated in LB medium at 37°C with shaking at 220 rpm. IPTG (0.5 mM) was used when the OD_{600} of the culture reached 0.6 to 0.8. After incubation at 20°C for 20 to 24 hours, cells were harvested and washed twice with 20 mM tris-HCl buffer and 500 mM NaCl (pH 7.5). The cells were resuspended in the same buffer and disrupted by sonication on ice. Then, the lysed cells were centrifuged at 10,000*g* for 30 min at 4°C, and the supernatant was reserved for further purification. The protein was purified using a His-Trap column (GE Healthcare, USA), desalted using an Amicon Ultra-4 centrifugal concentrator (30 kDa), and lastly stored in 20 mM tris-HCl buffer (pH 7.5) and 500 mM NaCl with 10% glycerol at −80°C. Protein concentration was determined with the Easy II Protein Quantitative Kit (BCA) (TransGen Biotech, China).

Enzyme activity assay of PGDH

The PGDH activity assay was performed according to the method described previously with slight modification (*39*). One milliliter of reaction mixture contained 100 mM tris-HCl (pH 7.5), 1 mM dithiothreitol, 5 mM EDTA-Na_2, NAD (2 mg/ml), and purified protein (300 µg). The reaction was started by adding 15 mM 3-phosphoglycerate, and the formation of NADH was measured spectrophotometrically at 340 nm. The relative activity is the ratio of the variant activity relative to the activity of wild-type PGDH.

Statistical analysis

Error bars indicate SDs from three parallel experiments. *P* values were generated from two-tailed *t* tests using the Microsoft Excel 2016 (Microsoft Corporation).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abq2157

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. H. Todor, M. R. Silvis, H. Osadnik, C. A. Gross, Bacterial CRISPR screens for gene function. *Curr. Opin. Microbiol.* **59**, 102–109 (2021).
2. H. Feng, Y. Yuan, Z. Yang, X. H. Xing, Z. Zhang, Genome-wide genotype-phenotype associations in microbes. *J. Biosci. Bioeng.* **132**, 1–8 (2021).
3. F. Rousset, D. Bikard, CRISPR screens in the era of microbiomes. *Curr. Opin. Microbiol.* **57**, 70–77 (2020).
4. Z. Bao, M. HamediRad, Genome-scale engineering of *Saccharomyces cerevisiae* with single-nucleotide precision. *Nat. Biotechnol.* **36**, 505–508 (2018).
5. A. D. Garst, M. C. Bassalo, G. Pines, S. A. Lynch, A. L. Halweg-Edwards, R. Liu, L. Liang, Z. Wang, R. Zeitoun, W. G. Alexander, R. T. Gill, Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat. Biotechnol.* **35**, 48–55 (2017).
6. L. Yao, K. Shabestary, S. M. Bjork, J. Asplund-Samuelsson, H. N. Joensson, M. Jahn, E. P. Hudson, Pooled CRISPRi screening of the *Cyanobacterium Synechocystis* sp PCC 6803 for enhanced industrial phenotypes. *Nat. Commun.* **11**, 1666 (2020).
7. T. Wang, C. Guan, J. Guo, B. Liu, Y. Wu, Z. Xie, C. Zhang, X. H. Xing, Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. *Nat. Commun.* **9**, 2475 (2018).
8. Y. Wang, Y. Liu, P. Zheng, J. Sun, M. Wang, Microbial base editing: A powerful emerging technology for microbial genome engineering. *Trends Biotechnol.* **39**, 165–180 (2020).
9. N. M. Gaudelli, A. C. Komor, H. A. Rees, M. S. Packer, A. H. Badran, D. I. Bryson, D. R. Liu, Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
10. K. Nishida, T. Arazoe, N. Yachie, S. Banno, M. Kakimoto, M. Tabata, M. Mochizuki, A. Miyabe, M. Araki, K. Y. Hara, Z. Shimatani, A. Kondo, Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, eaaf8729 (2016).
11. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
12. D. Zhao, J. Li, S. Li, X. Xin, M. Hu, M. A. Price, S. J. Rosser, C. Bi, X. Zhang, Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat. Biotechnol.* **39**, 35–40 (2020).
13. I. C. Kurt, R. Zhou, S. Iyer, S. P. Garcia, B. R. Miller, L. M. Langner, J. Grunewald, J. K. Joung, CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* **39**, 41–46 (2020).
14. P. Xu, Z. Liu, Y. Liu, H. Ma, Y. Xu, Y. Bao, S. Zhu, Z. Cao, Z. Wu, Z. Zhou, W. Wei, Genome-wide interrogation of gene functions through base editor screens empowered by barcoded sgRNAs. *Nat. Biotechnol.* **39**, 1403–1413 (2021).
15. R. E. Hanna, M. Hegde, C. R. Fagre, P. C. DeWeirdt, A. K. Sangree, Z. Szegletes, A. Griffith, M. N. Feeley, K. R. Sanson, Y. Baidi, L. W. Koblan, D. R. Liu, J. T. Neal, J. G. Doench, Massively parallel assessment of human variants with base editor screens. *Cell* **184**, 1064–1080.e20 (2021).
16. R. Cuella-Martin, S. B. Hayward, X. Fan, X. Chen, J. W. Huang, A. Taglialatela, G. Leuzzi, J. Zhao, R. Rabadan, C. Lu, Y. Shen, A. Ciccia, Functional interrogation of DNA damage response variants with base editing screens. *Cell* **184**, 1081–1097.e19 (2021).
17. P. C. Despres, A. K. Dube, M. Seki, N. Yachie, C. R. Landry, Perturbing proteomes at single residue resolution using base editing. *Nat. Commun.* **11**, 1871 (2020).
18. J. Becker, C. M. Rohles, C. Wittmann, Metabolically engineered *Corynebacterium glutamicum* for bio-based production of chemicals, fuels, materials, and healthcare products. *Metab. Eng.* **50**, 122–141 (2018).
19. J.-Y. Lee, Y.-A. Na, E. S. Kim, H.-S. Lee, P. Kim, The actinobacterium *Corynebacterium glutamicum*, an industrial workhorse. *J. Microbiol. Biotechnol.* **26**, 807–822 (2016).
20. Q. Z. Wang, J. Zhang, N. H. Al Makishah, X. M. Sun, Z. Q. Wen, Y. Jiang, S. Yang, Advances and perspectives for genome editing tools of *Corynebacterium glutamicum*. *Front. Microbiol.* **12**, 654058 (2021).
21. R. G. Stella, J. Wiechert, S. Noack, J. Frunzke, Evolutionary engineering of *Corynebacterium glutamicum*. *Biotechnol. J.* **14**, e1800444 (2019).
22. Y. Wang, Y. Liu, J. Liu, Y. Guo, L. Fan, X. Ni, X. Zheng, M. Wang, P. Zheng, J. Sun, Y. Ma, MACBETH: Multiplex automated *Corynebacterium glutamicum* base editing method. *Metab. Eng.* **47**, 200–210 (2018).
23. Y. Wang, Y. Liu, J. Li, Y. Yang, X. Ni, H. Cheng, T. Huang, Y. Guo, H. Ma, P. Zheng, M. Wang, J. Sun, Y. Ma, Expanding targeting scope, editing window, and base transition capability of base editing in *Corynebacterium glutamicum*. *Biotechnol. Bioeng.* **116**, 3016–3029 (2019).
24. C. Kuscu, M. Parlak, T. Tufan, J. Yang, K. Szlachta, X. Wei, R. Mammadov, M. Adli, CRISPR-STOP: Gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* **14**, 710–712 (2017).
25. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
26. M. Ehsaan, W. Kuit, Y. Zhang, S. T. Cartman, J. T. Heap, K. Winzer, N. P. Minton, Mutant generation by allelic exchange and genome resequencing of the biobutanol organism *Clostridium acetobutylicum* ATCC 824. *Biotechnol. Biofuels* **9**, 4 (2016).
27. S. A. Tripathi, D. G. Olson, D. A. Argyros, B. B. Miller, T. F. Barrett, D. M. Murphy, J. D. McCool, A. K. Warner, V. B. Rajgarhia, L. R. Lynd, D. A. Hogsett, N. C. Caiazza, Development of *pyrF*-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a pta mutant. *Appl. Environ. Microbiol.* **76**, 6591–6599 (2010).
28. A. Yamagishi, T. Tanimoto, T. Suzuki, T. Oshima, Pyrimidine biosynthesis genes (pyrE and pyrF) of an extreme thermophile, Thermus thermophilus. *Appl. Environ. Microbiol.* **62**, 2191–2194 (1996).
29. J. W. Youn, E. Jolkver, R. Kraemer, K. Marin, V. F. Wendisch, Characterization of the dicarboxylate transporter DctA in *Corynebacterium glutamicum*. *J. Bacteriol.* **191**, 5480–5488 (2009).
30. K. E. Baker, K. P. Ditullio, J. Neuhard, R. A. Klein, Utilization of orotate as a pyrimidine source by *Salmonella typhimurium* and *Escherichia coli* requires the dicarboxylate transport protein encoded by dctA. *J. Bacteriol.* **178**, 7099–7105 (1996).
31. J. Milse, K. Petri, C. Rueckert, J. Kalinowski, Transcriptional response of *Corynebacterium glutamicum* ATCC 13032 to hydrogen peroxide stress and characterization of the OxyR regulon. *J. Biotechnol.* **190**, 40–54 (2014).
32. H. Teramoto, M. Inui, H. Yukawa, OxyR acts as a transcriptional repressor of hydrogen peroxide-inducible antioxidant genes in *Corynebacterium glutamicum* R. *FEBS J.* **280**, 3298–3312 (2013).

33. M. Si, C. Chen, T. Su, C. Che, S. Yao, G. Liang, G. Li, G. Yang, CosR is an oxidative stress sensing a MarR-type transcriptional repressor in *Corynebacterium glutamicum*. *Biochem. J.* **475**, 3979–3995 (2018).

34. M. Si, C. Chen, J. Zhong, X. Li, Y. Liu, T. Su, G. Yang, MsrR is a thiol-based oxidation-sensing regulator of the XRE family that modulates *C. glutamicum* oxidative stress resistance. *Microb. Cell Fact.* **19**, 189 (2020).

35. H. Jeong, Y. Kim, H. S. Lee, OsnR is an autoregulatory negative transcription factor controlling redox-dependent stress responses in *Corynebacterium glutamicum*. *Microb. Cell Fact.* **20**, 203 (2021).

36. J. W. Choi, E. J. Jeon, K. J. Jeong, Recent advances in engineering *Corynebacterium glutamicum* for utilization of hemicellulosic biomass. *Curr. Opin. Biotechnol.* **57**, 17–24 (2019).

37. Y. Tsuge, Y. Hori, M. Kudou, J. Ishii, T. Hasunuma, A. Kondo, Detoxification of furfural in *Corynebacterium glutamicum* under aerobic and anaerobic conditions. *Appl. Microbiol. Biotechnol.* **98**, 8675–8683 (2014).

38. G. A. Grant, D-3-phosphoglycerate dehydrogenase. *Front. Mol. Biosci.* **5**, 110 (2018).

39. P. Peters-Wendisch, R. Netzer, L. Eggeling, H. Sahm, 3-Phosphoglycerate dehydrogenase from *Corynebacterium glutamicum*: The C-terminal domain is not essential for activity but is required for inhibition by L-serine. *Appl. Microbiol. Biotechnol.* **60**, 437–441 (2002).

40. A. Schuster, H. Erasimus, S. Fritah, P. V. Nazarov, E. van Dyck, S. P. Niclou, A. Golebiewska, RNAi/CRISPR screens: From a pool to a valid hit. *Trends Biotechnol.* **37**, 38–55 (2019).

41. P. N. Spahn, T. Bath, R. J. Weiss, J. Kim, J. D. Esko, N. E. Lewis, O. Harismendy, PinAPL-Py: A comprehensive web-application for the analysis of CRISPR/Cas9 screens. *Sci. Rep.* **7**, 15854 (2017).

42. H. H. Jeong, S. Y. Kim, M. W. C. Rousseaux, H. Y. Zoghbi, Z. Liu, CRISPRcloud: A secure cloud-based pipeline for CRISPR pooled screen deconvolution. *Bioinformatics* **33**, 2963–2965 (2017).

43. N. Suzuki, N. Okai, H. Nonaka, Y. Tsuge, M. Inui, H. Yukawa, High-throughput transposon mutagenesis of *Corynebacterium glutamicum* and construction of a single-gene disruptant mutant library. *Appl. Environ. Microbiol.* **72**, 3750–3755 (2006).

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

45. Y. B. Kim, A. C. Komor, J. M. Levy, M. S. Packer, K. T. Zhao, D. R. Liu, Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).

46. Y. Zhang, K. Yun, H. Huang, R. Tu, E. Hua, M. Wang, Antisense RNA interference-enhanced CRISPR/Cas9 base editing method for improving base editing efficiency in *Streptomyces lividans* 66. *ACS Synth. Biol.* **10**, 1053–1063 (2021).

47. S. Yu, M. A. Price, Y. Wang, Y. Liu, Y. Guo, X. Ni, S. J. Rosser, C. Bi, M. Wang, CRISPR-dCas9 mediated cytosine deaminase base editing in *Bacillus subtilis*. *ACS Synth. Biol.* **9**, 1781–1789 (2020).

48. R. T. Walton, K. A. Christie, M. N. Whittaker, B. P. Kleinstiver, Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).

49. B. Yang, L. Yang, J. Chen, Development and application of base editors. *CRISPR J.* **2**, 91–104 (2019).

50. K. A. Molla, Y. N. Yang, CRISPR/Cas-mediated base editing: Technical considerations and practical applications. *Trends Biotechnol.* **37**, 1121–1142 (2019).

51. Y. L. Ruan, L. J. Zhu, Q. Li, Improving the electro-transformation efficiency of *Corynebacterium glutamicum* by weakening its cell wall and increasing the cytoplasmic membrane fluidity. *Biotechnol. Lett.* **37**, 2445–2452 (2015).

52. C. Keilhauer, L. Eggeling, H. Sahm, Isoleucine synthesis in *Corynebacterium glutamicum*: Molecular analysis of the ilvB-ilvN-ilvC operon. *J. Bacteriol.* **175**, 5595–5603 (1993).

53. A. Casini, G. Christodoulou, P. S. Freemont, G. S. Baldwin, T. Ellis, J. T. MacDonald, R2oDNA designer: Computational design of biologically neutral synthetic DNA sequences. *ACS Synth. Biol.* **3**, 525–528 (2014).

54. A. Schafer, A. Tauch, W. Jager, J. Kalinowski, G. Thierbach, A. Puhler, Small mobilizable multipurpose cloning vectors derived from the *Escherichia coli* plasmids PK18 and PK19 selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene* **145**, 69–73 (1994).