- Genome-scale mapping of functional genes and loci in bacteria for industrial phenotypes

Yaomeng Yuan,[1#] Xihao Liao,[1#] Shuang Li,[1] Xin-hui Xing,[2] Chong Zhang[1]*

[1] MOE Key Laboratory for Industrial Biocatalysis, Institute of Biochemical Engineering, Department of Chemical Engineering, Tsinghua University, Beijing 100084, China; Center for Synthetic and Systems Biology, Tsinghua University, Beijing 100084, China.

[2] Institute of Biopharmaceutical and Health Engineering, Tsinghua Shenzhen International Graduate School, Shenzhen, Guangdong, 518055, China.

[#]These authors contributed equally.

Abstract

Genome-scale CRISPR/Cas9-based screening has been applied as a powerful strategy to improve phenotypes in strain engineering. Base editors, which offer an opportunity for the precise modification of genes without double-strand break and recombination, have been applied in screenings in eukaryotes. Here, we tested the great potential of base editors in high-throughput loss-of-function screenings by inactivating 89% genes in *Escherichia coli*. Furthermore, to fully exploit the ability of base editors to create mutations at single base resolution, we mutated around 31,566 individual sites across 462 stress response related genes in *E. coli* and screened for growth-promoting mutations under formate, isobutanol and furfural stress. As a result, we identified 170 gene candidates critical to growth, and scanned the function of tens of thousands of specific sites within target genes. This screening approach will accelerate large-scale interrogation of phenotype-genotype relationships in a broad range of hosts for metabolic engineering and synthetic biology.

Introduction

As the core of green bio-manufacturing and bio-economy, microbial cell factories are widely applied in the bio-production of fuels, medicines, chemicals and foods. To meet the requirements of industrial applications, engineering tools have been employed to optimize the target traits and develop better performed microbial cell factories. Numerous reported works mainly focus on the modification of the genes directly link to the target pathways. However, many complex phenotypes, such as resistance to toxic chemicals, protein secretion and chemical production, are controlled by a large number of genes and pathways(*1–3*). Due to the limited knowledge of biological system, it is difficult to identify and modify the related genetic loci to improve the target phenotypes at a system level.

Reverse engineering strategies, such as adaptive laboratory evolution and random mutagenesis, are broadly used to optimize the performance of microbial cell factories when there is a lack of prior knowledge(*4–6*). However, these methods suffer from a shortage that only a small number of mutations in the whole landscape can be detected, resulting in a loss of information(*6*). On the other hand, many engineered strains obtained by these strategies usually carry large amounts of mutations, making it laborious to identify the mutations which play a key role in the phenotype acquisition. Therefore, strategies that enable efficient genome editing are needed to map genotype-phenotype associations and guide the engineering of strains.

Recently, the CRISPR interference (CRISPRi) pooled screening have been applied in modifying gene expression levels to identify how each gene contributes to the target phenotype in a massively parallel manner(*3, 7, 8*). However, these approaches cannot reveal the role of specific positions within a locus. Moreover, mutations should be introduced to further modify the target genes identified by these tools for an optimized phenotype. Recent developed CRISPR-Cas9-associated gene-editing tools, such as CREATE, have allowed for the exploration of the effects of genomic loci at single base resolution(*9*). However, mutagenesis using these tools usually requires the formation of double-strand break (DSB) followed by homology-directed repair in the presence of donor DNA, which often needs a complex plasmid construction and strain background. An alternative is to use base editors, in which mutations are created by directly modifying the target bases without DSBs.

Base editors, where cytosine deaminase or adenine deaminase is fused to Cas9/Cas12 protein variants (e.g., dCas9, nCas9) to introduce specific mutations at target sites, have been shown to be effective in a variety of species(*10–13*). Recently, base editor associated loss-of-function screening has been successfully applied in eukaryotes(*14–16*). In this work, we demonstrated the great potential of base editor-mediated high-throughput screening of functional genes and loci in prokaryotes by using Target-AID(*12*), a previously reported cytosine base editor (CBE) with a narrow mutation window. A sgRNA library was designed to disrupt the function of 89% of the total genes in *Escherichia coli* by creating stop codons and applied to screen genes critical to growth. As a result, 339 sites across 170 genes were identified to have significant impacts to growth when nonsense mutations occurred. We also utilized this library to search for functional genes that can enhance the growth under formate stress, an important industrial phenotype. Harnessing the excellent ability of base editors to introduce mutations at single base resolution, we further created base substitutions at 31,566 individual sites across 462 genes which have been reported to be associated with stress response in *E. coli*. Three screening examples of mutations conferring improved formate, isobutanol and furfural resistance, were performed to demonstrate the great potential of base editors in functional loci mining. We successfully identified some known and novel genes and sites that are beneficial to the target phenotypes. Moreover, the obtained datasets provided a function profile of a large number of genetic sites, shedding light on the development of protein engineering, metabolic engineering and synthetic biology.

Results

Establishing a dual-induced Target-AID system in *E. coli* for pooled screening

In Target-AID, the *Petromyzon marinus* cytosine deaminase PmCDA119 (CDA), uracil DNA glycosylase inhibitor (UGI) from bacteriophage PBS2(*17*) and protein degradation tag (LVA tag)(*18*) were fused with the catalytically inactivated Cas9 variant (dCas9) from *Streptococcus pyogenes*. Guided by sgRNA, the dCas9-CDA-UGI-LVA protein can introduce a cytosine to thymine conversion at the -17 to -19 position from the PAM with a high frequency(*12*) (Fig. 1A). However, the dCas9 protein can also block the gene transcription initiation and elongation when binding to the target sites in *E. coli*, which may result a decreased gene expression level (known as CRISPRi effect). To distinguish the variants enriched during the screening are due to the contribution of base substitution rather than gene knockdown, we designed a dual-induced Target-AID system, where the expression of dCas9-CDA-UGI-LVA fusion protein was controlled by arabinose and the transcription of the corresponding sgRNA was regulated by rhamnose (Fig. S1). After

inducing mutagenesis, cells were washed and inoculated into the selection medium without inducers for screening.

To evaluate the CRISPRi effect of the system, we used sfGFP as a reporter gene and introduced 3 synonymous mutations in its ORF via Target-AID. To broaden the range of sgRNAs when designing the base editing library, a Cas9 protein variant, xCas9(*19*), with an expanded PAM of NG, was also used to construct the Target-AID_NG system. Almost no fluorescence repression was observed in both systems when no inducers were added, indicating that there was no significant CRISPRi effect or expression leakage. In addition, the Target-AID_NGG system showed a higher fluorescence repression when inducers were added, suggesting a stronger binding ability with target DNA (Fig. 1B). Singled colonies were further picked to confirm whether mutations were occurred by Sanger sequencing. As a result, the Target-AID_NGG showed a higher mutagenesis frequency (Fig. 1C) and was applied in the subsequent screenings.

Design and characterization of a loss-of-function library in genome scale

To explore the ability of Target-AID in searching functional genes via pooled screening, we designed a library that can introduce stop codons to disrupt genes in the whole genome, resulting a loss-of-function library. To demonstrate the capability of Target-AID in gene inactivation, we first designed a sgRNA targeting the chloramphenicol resistance (CmR) gene and used Target-AID to disrupt its function. The strain with edited CmR gene failed to grow on the medium containing chloramphenicol, indicating a successful loss-of-function (Fig. S2). For the library design, we have previously established pooled CRISPRi screening strategy and identified a library of sgRNA with high activity and specificity in *E. coli* (termed as GW_library)(*3*). There are four kinds of sgRNA that can generate stop codon: sgRNAs targeting the coding strand with a target codon of CGA/CAG/CAA in position -12 ~ -19, and targeting the noncoding strand with a target codon of CCA in position -12 ~ -20. We filtered sgRNAs that meet the four conditions in GW_lilbrary to obtain the STOP_library, which containing 19,086 targeting sgRNAs and 400 non-targeting sgRNAs. 3634 of the total 4098 genes (89%) of *E. coli* were targeted with at least one sgRNA. 41.2% of the 3634 genes were targeted by no more than 3 sgRNAs and the targeting positions of all sgRNAs in the open reading frame (ORF) were almost evenly distributed (Fig. S3). The STOP_library was constructed and transformed into *E. coli* BW25113 as described in Methods. The editing efficiency was estimated by randomly picking ~30 colonies and sequencing both sgRNA and the corresponding genomic sites. As a result, an 82% mutation rate was obtained (Table S1).

As a proof-of-concept, we first applied the library to identify essential genes of *E. coli* in Luria-Bertani broth. The Pearson coefficients of the two replicates indicated a good consistency (Fig. S4). When identifying the association between gene inactivation (genotype) and fitness (phenotype) at the gene level, we found that the data of most genes could not pass the statistical hypothesis test, mainly because: 1) most genes were targeted by no more than 3 sgRNAs, 2) some sgRNAs were not located in the front part of the gene ORF, resulting in incomplete loss-of-function. We thus analyzed the screening data at the sgRNA level. We found that there were 339 sgRNAs showed a significantly decreased fitness, targeting a total of 170 genes (Fig. 2B, 2C). Among them, 295 sgRNAs (87%) targeted 120 genes suggested as essential by Keio collection. GO analysis reveals that the remaining 50 genes were enriched in biological processes such as ATP biosynthetic

process, ATP metabolic process, ribose phosphate and purine nucleotide metabolic process, implying that these biological pathways are critical for bacterial growth (Fig. 2D).

Screening functional genes for formate tolerance using STOP_library

We then applied STOP_library to mining the critical genes conferring higher formate tolerance. Formate, together with other one-carbon compounds such as methanol, has great potential as a sustainable feedstock for bioproduction(*20*). Although synthetic *E. coli* strains that can assimilate formate were reported, the utilization of formate for biomass synthesis and chemical production is still limited due to its toxicity(*21*, *22*). On the other hand, formate is one of the major metabolites excreted by *E. coli* in fermentation. Production of formate is maximal in anaerobic conditions, but significant amounts of formate can also be excreted in oxygenated cultures, which is a key concern in large-scale fermentation(*23*). The screening was performed under 150 mM formate, which resulted a significantly decreased cell growth in *E. coli* BW25113 harboring non-targeting sgRNA (Fig. S5). The read counts of the two replicates showed a Pearson coefficient higher than 0.96 (Fig. S6). Several highly enriched mutations were reconstructed by Target-AID for growth verification (Fig. 3, Fig. S7). cvrA_Q451* variant showed a 1.31-fold higher growth under formate stress. *cvrA* encodes the K+/H+ antiporter, which is known to extrude K+ in exchange for external H+ to maintain a stable intracellular environment. To the best of our knowledge, this gene was identified as a candidate to improve the growth in formate for the first time. Further research is need to reveal the mechanism of the phenomenon. The disruption of the low affinity tryptophan permease (*tnaB*) involved in tryptophan transport also improved the formate tolerance. This gene, together with *tnaA*, tryptophan biosynthesis genes and indole transporter, was previously reported to improve the isobutanol and furfural tolerance when knocked down(*3*). However, we observed a decreased fitness when inactivating tnaA (STOP_sgRNA12079).

Designing a base editing library targeting stress response related genes

To bridge the gap between base mutagenesis and the corresponding phenotype, we further applied the Target-AID-mediated functional screening strategy to reveal beneficial mutations conferring higher growth in toxic chemicals, which was regarded as an important industrial phenotype. In detail, 462 genes related to stress response such as heat shock and organic solvent toxicity were chosen as targets(*3*, *9*, *24*). We selected the sgRNAs targeting these genes from GW_library, and further filtered those containing cytosine(s) in position -14 ~ -19. Together with 400 non-targeting sgRNAs, they constitute a directed mutagenesis library with a size of 31,966, which was named as AID_library. The library was transformed into *E. coli* BW25113 with a 400-fold coverage and ~82% editing efficiency (Table S1).

Screening beneficial mutations for improved growth in formate using AID_library

The AID_library was first applied to screen beneficial mutations improving growth in formate stress. The Pearson coefficients of the sgRNA read counts between two replicates were higher than 0.96, indicating a good consistency (Fig. S8). We noted that the positive mutations were highly enriched in *pcnB* and *crp* genes (Fig. 4). *pcnB* gene encodes poly(A) polymerase I, which plays a significant role in the global control of gene expression by influencing the rate of many transcripts degradation. Mutations in crp gene, which encodes cAMP-activated global transcriptional regulator, were reported to increase

tolerance of many toxic chemicals such as acetate and alcohols. We hypothesized that both *pcnB* and *crp* variants increased the phenotype by transcriptional regulating the expression level of other genes. Based on the screening results of STOP_library, we also paid attention to the fitness of sgRNAs targeting tnaA and tnaB. We found an increased fitness of AID_sgRNA29500 targeting tnaA and a decreased fitness of AID_sgRNA16627 targeting *tnaB* in the formate screening using AID_library. These results indicated that the modification of these two genes has a potential to further improve the tolerance to a variety of organic solvents including formate. We reconstructed several highly enriched variants (Fig. 5A) and their growth were shown in Fig. 5B. Among them, the oppD_P260L showed a 20% growth improvement compared to the non-targeting control.

Together with the positive variants obtained from STOP_library, we constructed strains harboring double mutations of oppD_P260L, cvrA_Q451* and tnaB_Q50* using Target-AID. Under formate stress, the variant carrying cvrA_Q451* and tnaB_Q50* mutations showed a slightly enhanced OD600 in stationary phase, implying that the beneficial effects of the two mutations are addictive (Fig. S9).

Screening beneficial mutations in isobutanol amd furfural using AID_library

Two more screening examples under 4 g/L isobutanol and 0.5 g/L furfural were performed, respectively (Fig. S5). Isobutanol, including some other higher alcohols, is considered to be the most promising substitute for gasoline(*25*). The bio-production of isobutanol has been achieved in industrial strains such as *E. coli* by applying metabolic engineering strategies (50 g/L at laboratory level)(*26*), but it is still necessary to increase the yield to lower the cost and meet the industrial demand. The toxicity of these biofuels to strains has been regarded as a key factor limiting the increase in their production(*27*). The two replicates of the screening under isobutanol showed a good consistency (Fig. S10). The screening results suggested that multiple positive mutations were enriched in gene *pcnB* and *crp* (Fig. 4), indicating that the two genes are potential targets capable of conferring multiple tolerance. We reconstructed some positive mutants and validated their growth (Fig. 6A, 6B).

Furfural is considered a major inhibitor of microbial fermentation since it is a toxic by-product in high concentration produced during the lignocellulose pretreatment(*28*). Furfural can consume the NAD(P)H, inhibit glycolysis and TCA cycle(*29*). It has been reported that the pyruvate dehydrogenase (PDH) complex can be inhibited by furfural(*30*). In our screening, multiple mutations in gene pdhR were highly enriched (Fig. 4, 6C). The pdhR is a pyruvate sensing bifunctional regulator for control of 16-23 targets, including PDH complex and genes for respiratory electron transport, fatty acid degradation and flagella formation. We reconstructed the highly enriched pdhR_V14V_E16K variant and find that its growth is slightly higher than the non-targeting control (Fig. 6D). This mutation is located in the HTH gntR-type conserved region of the protein, which may affect the inhibition of the PDH complex. Another positive variant in our screening, pdhR_R118C, has been demonstrated to derepress and constitutively activated the PDH complex synthesis(*31*), which may alleviate the inhibition of furfural. We also verified two more positive variants, but no improvement of growth was observed (Fig. 6C, S12). The slight fluctuation in the curve may be caused by the background noise of the instrument.

We also noted that the three screenings using AID_library shared highly similar negative sgRNAs, most of which targeted to essential genes such as *nusA*, *rpoB*, *rpoC*, *rho*. These

results indicated that these mutations disrupted the function of genes and adversely affected growth (Fig. S13). Taken together, the screening strategy mediated by Target-AID showed a great potential in the dissection of functional genes and loci.

Discussion

Large-scale functional screening mediated by the tools derived from CRISPR-Cas9 systems accelerated the mapping of genotype-phenotype association in a trackable and customized manner. Although CRISPRi pooled screening has been widely applied in the functional genomic studies in single gene level, the polar effect of polycistronic operons is one of the key concerns for false positives in prokaryotes. To solve this, we developed a genome-scale loss-of-function screening strategy in *E. coli* using base editors, where the genes were perturbed at translation instead of transcription level. We constructed a STOP_library targeting 89% of the total genes in the whole genome to disrupt the gene function and applied it to evaluate whether this strategy could identify essential genes. However, it was difficult to analyze the result at single gene level, since most genes were targeted by no more than 3 sgRNAs, and some sgRNAs were not located in the front part of the gene ORF, resulting in incomplete loss-of-function. In addition, some sgRNAs cannot be enriched or filtered out due to a low editing efficiency. Nevertheless, we observed that a large proportion of sgRNAs with significantly reduced fitness targeted essential genes, implying that our strategy can provide clues for the identification of functional loci or genes. Screening for formate tolerance was further performed as a proof-of-concept to demonstrate the potential of our strategy in industrial application. As a result, some gene candidates capable of enhancing the growth in the presence of formate were identified.

In prokaryotes, CREATE is a representative approach for the precise genome editing, where the base substitution is generated by homology-directed repair of DSBs in the existence of a donor DNA. Despite the great success, the toxicity of DSBs, the requirement of DNA templates and complex strain background hinder its application in a wider range. In contrast, base editors enable precise gene editing that is not limited by the aforementioned conditions. To fully harness the power of base editor in screening functional loci at base resolution, we then designed an AID_library which can introduce mutations at 31,566 individual sites across 462 genes. Three screenings were performed to demonstrate the capability of our strategy in identifying beneficial mutations that conferring higher growth in the presence of isobutanol, formate and furfural. We successfully identified some novel beneficial mutations, among which the oppD_P260L variant showed a 1.22-fold higher growth in the presence of formate.

When we were preparing this manuscript, a similar work carried out in *Corynebacterium glutamicum* ATCC 13032 was reported by Liu et al., in which 12,000 sgRNAs that can inactivate 98.1% of total genes were constructed and applied to identify genes related to the resistance of toxic chemicals(*32*). Their work demonstrated the great potential of base editor in identifying beneficial nonsense mutations via large-scale pooled screening in prokaryotes. Unlike their work, AID_library was applied here to further demonstrate the capability of base editor in identifying more types of mutations in *E. coli*. Our strategy not only enabled the identification of functional genes and sites in genome-scale, but also paves the way for mapping of the role of specific positions within target genes. We speculate some functional mutations cannot be enriched due to the low editing efficiency of some sgRNAs. Moreover, one sgRNA may create several kinds of mutation within the editing window. Different effects of these mutations on growth may interfere with sgRNA

fitness. For example, a mixture of neutral and beneficial mutations may lead to an underestimation of sgRNA fitness. On the other hand, although some sgRNAs showed increased fitness in the screening, the growth improvement of the corresponding variants may be too weak to detect in validation. In addition, based on the screening data analysis and growth verification, we found some enriched mutants with low initial read counts showed no significant growth improvement (Fig. S7, S12), which was also observed by Liu et al. For more accurate and efficient genotype-phenotype association, more researches should be performed to improve base editor performance and data analysis algorithms.

Recently, base editors with improved performance have been reported, paving the way for the development of functional loci identification mediated by base editors in a massively parallel manner. Developing base editors with broad PAM by using new Cas9 variants can widen the editable scope in genome, which can increase the number of target sites and perhaps endow the capability to analyze genotype-phenotype association at the single gene level(*33*). Establishing base editors with less by-standers, higher editing efficiency, reduced off-target activity and develop algorithms to predict editing outcomes will also facilitate the predictability and reduce the difficulty in library design and data analysis (*33–35*). Compared with CREATE, one of the major drawbacks of Target-AID is the limited base mutation type in *E. coli* (C to T). Recently, the C to A conversion was observed in *E. coli* MG1655 when using Ung-nCas9-AID fusion(*36*). Together with the development of adenine base editors (ABEs)(*11*), exploiting the potential of these new base editors may further expand their capability in functional screenings. Furthermore, by combining biosensors or chromogenic reactions with screening platforms such as fluorescence-activated cell sorting (FACS)(*37*), fluorescence-activated droplet sorting (FADS)(*38*), Droplet-based FACS(*39*), Gel-FACS(*40*), more complex phenotypes such as metabolite production and protein secretion can also be screened. Finally, a curable plasmid can be designed for the base editing system to allow the mutation combination via iterative screenings, which could potentially lead a better optimization of the desired phenotypes. Taken together, we are positive that the base editor-mediated functional screening will accelerate large-scale interrogation of phenotype-genotype relationships in a broad range of hosts for metabolic engineering and synthetic biology.

Materials and Methods

Strains and plasmid construction

All strains, plasmids and primers used in this work are listed in Table S2 and S3. *E. coli* BW25113 was obtained from the ATCC and *E. coli* DH5α was bought from Biomed. The dCas9-CDA-UGI-LVA expression cassette was amplified from pScI-dCas9-CDA-UL plasmid (Addgene #108551) and inserted into pBAD43 vector via Gibson assembly to obtain pdCas9-AID. The plasmid for sgRNA expression, termed as pTargetF, was derived from pTargetF_lac (3) by replacing the lac promoter with rhaBAD promoter. To facilitate library insertion into pTargetF, pTargetF_rhaBAD_RFP was further constructed by replacing the 20-bp target sequence with an RFP expression cassatte containing two BsmBI sites in opposite orientations at both ends. A pair of oligonucleotide DNAs for construction of the targeting sgRNA vector was designed as follows: 5’-tagc-(target sequence)-3’ and 5’-aaac- (reverse complement of the target sequence)-3’. Goldengate assembly was performed to insert the target sequences into

pTargetF_rhaBAD_RFP. All the constructed plasmids were confirmed by Sanger sequencing.

Mutagenesis assay

The BW25113 cells were chemically transformed with the appropriate plasmids and pre-cultured at 37℃ with 800 μL Luria–Bertani (LB) medium for 1 h. The cells were then transferred into 10 mL LB medium containing streptomycin (30 mg/L), ampicillin (100 mg/L), L-arabinose (100 mg/L) and L-rhamnose (300 mg/L), and grown overnight at 37℃ and 220 r.p.m. The cultures were then serially diluted onto LB agar plates supplemented with appropriate antibiotics and incubated overnight at 37℃ to form single colonies. Sanger sequencing was performed to verify the mutations.

Fluorescence characterization of the Target-AID system

For fluorescence characterization, *E. coli* s17-1 harboring *sfGFP* gene was used as the host. Three sgRNAs targeting the *sfGFP* gene, as well as the non-targeting sgRNA, were chemically transformed into s17-1 strain containing pdCas9-AID or pdCas9-AID-NG, respectively. The mutations were introduced according to the protocol described in Mutagenesis assay. Subsequently, overnight cultures in LB medium (with streptomycin, ampicillin) from single colonies of these strains were individually incubated in 10 mL fresh LB medium in 50-mL flasks (initial OD600 = 0.02) for 12h, and fluorescence was measured with an F-2500 Hitachi Fluorescence Reader (excitation, 488 nm; emission, 510 nm). Then, fluorescence was normalized to the culture OD600 value measured on an Amersham Bioscience spectrophotometer. The repression ratio was calculated by comparing the relative fluorescence with respect to the control strain expressing the non-targeting sgRNA.

Design and preparation of sgRNA library

In previous work, we have selected a set of sgRNAs with high activity and specificity in the whole genome of *E. coli*(*3*), termed as GW_library. For the design of AID_library, 462 target genes were chosen as editing targets, in which 35 genes related to the resistance of acetate(*9*), 97 genes related to the tolerance of furfural and isobutanol(*3*), and 330 genes related to thermo-tolerance(*41*). We selected the sgRNAs targeting the 462 genes from GW_library and further filtered those containing cytosine(s) in position -14 ~ -19. For the design of STOP_library, there are four kinds of sgRNA that can generate stop-codon: sgRNA targeting the coding strand with a target codon of CGA/CAG/CAA in position -12 ~ -19, and targeting the noncoding strand with a target codon of CCA in position -12 ~ -20. We filtered the sgRNAs meeting the four conditions in GW_lilbrary to obtain the STOP_library, which containing 19,086 targeting sgRNAs and 400 non-targeting sgRNAs. The plasmid libraries of AID_library and STOP_library was synthesized by Genewiz.

Screening experiments

The library plasmids were electroporated into *E. coli* BW25113 carrying the pdCas9-AID plasmid. For competent cell preparation, *E. coli* BW25113 containing the pdCas9-AID were grown in 100 mL LB medium without NaCl at 37 ℃ until the OD600 reached to 0.6. The cells were then collected by centrifugation at 4 ℃, washed for five times in ice-cold

deionized water and resuspended in 5 mL 15% glycerol. The prepared competent cells were mixed with 1μg library plasmid/mL competent cells, divided into 100μL and added into 25-well electroporation plates. For electroporation, BTX Harvard apparatus ECM 630 High Throughput Electroporation System was used and parameters were setting as 2.1 kV, 1 kΩ, 25 μF. The transformed cells were incubated in LB medium (1:4 v/v) for 1 h at 37 ℃, 220 r.p.m. 50 μL of the culture was diluted and streaked onto the LB agar plates with streptomycin (20 mg/L) and ampicillin (100 mg/L) to calculate the electroporation efficiency. The results suggested that the library coverage was ~ 400-fold. Subsequently, the cells were transferred into 100 mL LB medium containing streptomycin, ampicillin, L-arabinose (100 mg/L) and L-rhamnose (300 mg/L), and grown at 37 ℃ and 220 r.p.m. until an OD600 of ~1 was reached. We took 5 mL of the culture to extract plasmids as 'Before Library' for NGS. The remaining cells were further collected by centrifugation, washed twice and resuspended with M9 medium (12.8 g Na2PO4·7H2O, 3 g KH2PO4, 0.5 g NaCl, 1 g NH4Cl, 4 g glucose, 2 mM MgSO4, 0.1 mM CaCl2, for 1 L). 50 μL of the culture was diluted and streaked onto the LB agar plates, ~30 colonies were randomly picked for Sanger sequencing to calculate the mutagenesis efficiency. The remaining part of the cells were inoculated into 100 mL medium in a 500 mL flask (LB, M9+150 mM formate, M9+ 4 g/L isobutanol, M9+ 0.5 g/L furfural) with an initial OD600 of 0.03, and grown until an OD600 of 1 was reached. 5 mL of the culture was taken for plasmids extraction, termed as 'After Library'. Two biological replicates were performed for each screening.

Next generation sequencing (NGS) library sequencing and data analysis

The extracted plasmids were used as templates for PCR to amplify the N20 region of the sgRNAs. The PCR condition was set as follows: one reaction (50 μL) per library, 50 ng templates per reaction, primer NGS-seq-F/R, KAPA HiFi HotStart polymerase (KAPA Biosystems), 95 ℃for 3 min, 20 cycles (98 ℃, 20 s; 60℃, 15 s; 72 ℃, 30 s), 72 ℃ for 1 min). The amplified fragments were confirmed by 4% agarose gel electrophoresis, cleaned using a Gel Extraction Kit (Omega) and processed for NGS using standard Illumina preparation kits. Sequencing was carried out using a 2 × 150 paired-end configuration.

The raw NGS data processing and fitness calculation of sgRNA were performed as described previously(*3*). briefly, the raw NGS data was first de-multiplexed, filtered and the adaptor was removed to obtain clean data for each library. Then, we merged each of the two pairs by FLASH script and removed the reads without detected pairs. Customized python scripts were applied to calculate the read counts of each sgRNA. Due to the sequencing depth of each library is different, we used a normalization factor described in equation 1 to adjust the read counts.

$$\text{Normalization factor}_i = Read\ count_i / (\sum_i^n Read\ count_i / n) \quad (1)$$

Where i is the identity of sgRNA and n is the number of replicates. To increase statistical robustness, sgRNA with <20 normalized read counts in the plasmid library for each experiment were removed. Subsequently, the fitness for each sgRNA in two biological replicates was calculated using equation 2 and 3.

$$\text{Fitness'}_{sgRNA} = \text{Log}_2((\text{Read count})_{after}/(\text{Read count})_{before}) \quad (2)$$

$$\text{Fitness}_{\text{sgRNA}} = \text{Fitness'}_{\text{sgRNA}} - \text{median}(\text{Fitness'}_{\text{NCsgRNA}}) \quad (3)$$

In the screening of essential genes, the z score of each individual sgRNA was calculated by dividing FitnesssgRNA by the standard deviation σ (equation 4). The fitness of all NC sgRNAs was fitted to a t-distribution and calculated the σ value. The z score for each NC sgRNA was calculated by dividing the NC sgRNA fitness by the σ using equation 5.

$$\text{z score}_{\text{sgRNA}} = \text{Fitness}_{\text{sgRNA}}/\sigma_{\text{sgRNA}} \quad (4)$$

$$\text{z score}_{\text{NCsgRNA}} = \text{Fitness}_{\text{NCsgRNA}}/\sigma_{\text{t-distribution of sgRNA}} \quad (5)$$

We defined a False Discovery Rate (FDR) using equation 6, where False Positive Rate (FPR) was the cumulative frequency distribution of the z score of targeting sgRNA and FPR_NC was the cumulative distribution function of the t-distribution of z score of NC sgRNA (14). The sgRNAs with significantly decreased fitness were identified using the criteria with FDR less than 0.1.

$$\text{FDR} = \text{FPR}/\text{FPR_NC} \quad (6)$$

Mutant reconstruction and growth curve characterization

Mutations were introduced into *E. coli* BW25113 as described in 'mutagenesis assay'. The strains were stored at -80℃ after confirming mutations via Sanger sequencing. For growth curve verification, the stored strains were streaked onto the LB plates to isolate single colonies. The colonies were then inoculated in LB broth for 10 h shaking at 37 ℃, 220 r.p.m. and diluted 40-fold into M9 medium for 6 more hours' shaking. Subsequently, the cells were collected by centrifugation and inoculated into the medium containing corresponding organic solvents to characterize growth using Bioscreen C MBR (Oy Growth Curves Ab Ltd).

Statistical analysis

Error bars indicate SDs from 3 parallel experiments. GO analysis were performed at http://geneontology.org/. Plots were generated in GraphPad Prism 8 or Python 3.8 using the matplotlib (3.3.2) plotting libraries. Genome plots were generated in R using the OmicCircos package. All statistical analysis were performed using the pandas (1.1.3), SciPy (1.5.2) and NumPy (1.19.2) Python packages and the Microsoft Excel 2016 (Microsoft Corporation).

References

1. B. Özaydin, H. Burd, T. S. Lee, J. D. Keasling, Carotenoid-based phenotypic screen of the yeast deletion collection reveals new genes with roles in isoprenoid production. *Metab Eng*. **15**, 174–183 (2013).
2. Y. S. Jin, G. Stephanopoulos, Multi-dimensional gene target search for improving lycopene biosynthesis in Escherichia coli. *Metab Eng*. **9**, 337–347 (2007).
3. 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-net. *Nat Commun*. **9** (2018), doi:10.1038/s41467-018-04899-x.
4. A. L. Demain, "From natural products discovery to commercialization: A success story" in *Journal of Industrial Microbiology and Biotechnology* (2006), vol. 33, pp. 486–495.

5. S. S. Fong, A. R. Joyce, B. Palsson, Parallel adaptive evolution cultures of Escherichia coli lead to convergent growth phenotypes with different gene expression states. *Genome Res*. **15**, 1365–1372 (2005).
6. J. E. Barrick, R. E. Lenski, Genome dynamics during experimental evolution. *Nat Rev Genet*. **14** (2013), , doi:10.1038/nrg3564.
7. J. D. Smith, S. Suresh, U. Schlecht, M. Wu, O. Wagih, G. Peltz, R. W. Davis, L. M. Steinmetz, L. Parts, R. P. St.Onge, Quantitative CRISPR interference screens in yeast identify chemical-genetic interactions and new rules for guide RNA design. *Genome Biol*. **17** (2016), doi:10.1186/s13059-016-0900-9.
8. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*. **152** (2013).
9. 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).
10. H. A. Rees, D. R. Liu, Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet*. **19** (2018), , doi:10.1038/s41576-018-0059-1.
11. 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 T to G C in genomic DNA without DNA cleavage. *Nature*. **551** (2017), doi:10.1038/nature24644.
12. S. Banno, K. Nishida, T. Arazoe, H. Mitsunobu, A. Kondo, Deaminase-mediated multiplex genome editing in Escherichia coli. *Nat Microbiol*. **3**, 423–429 (2018).
13. 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 (1979)*. **353** (2016), doi:10.1126/science.aaf8729.
14. P. C. Després, A. K. Dubé, M. Seki, N. Yachie, C. R. Landry, Perturbing proteomes at single residue resolution using base editing. *Nat Commun*. **11** (2020), doi:10.1038/s41467-020-15796-7.
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** (2021), doi:10.1016/j.cell.2021.01.041.
17. W. Zhigang, D. G. Smith, D. W. Mosbaugh, Overproduction and characterization of the uracil-DNA glycosylase inhibitor of bacteriophage PBS2. *Gene*. **99** (1991), doi:10.1016/0378-1119(91)90030-F.
18. J. B. Andersen, C. Sternberg, L. K. Poulsen, S. P. Bjørn, M. Givskov, S. Molin, New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol*. **64** (1998), doi:10.1128/aem.64.6.2240-2246.1998.
19. J. H. Hu, S. M. Miller, M. H. Geurts, W. Tang, L. Chen, N. Sun, C. M. Zeina, X. Gao, H. A. Rees, Z. Lin, D. R. Liu, Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*. **556**, 57–63 (2018).
20. O. Yishai, S. N. Lindner, J. Gonzalez de la Cruz, H. Tenenboim, A. Bar-Even, The formate bio-economy. *Curr Opin Chem Biol*. **35** (2016), pp. 1–9.
21. S. Gleizer, R. Ben-Nissan, Y. M. Bar-On, N. Antonovsky, E. Noor, Y. Zohar, G. Jona, E. Krieger, M. Shamshoum, A. Bar-Even, R. Milo, Conversion of Escherichia coli to Generate All Biomass Carbon from CO2. *Cell*. **179**, 1255-1263.e12 (2019).
22. S. J. Kim, J. Yoon, D. K. Im, Y. H. Kim, M. K. Oh, Adaptively evolved Escherichia coli for improved ability of formate utilization as a carbon source in sugar-free conditions. *Biotechnol Biofuels*. **12** (2019), doi:10.1186/s13068-019-1547-z.
23. C. Kirkpatrick, L. M. Maurer, N. E. Oyelakin, Y. N. Yoncheva, R. Maurer, J. L. Slonczewski, Acetate and formate stress: Opposite responses in the proteome of Escherichia coli. *J Bacteriol*. **183**, 6466–6477 (2001).
24. O. Tenaillon, A. Rodríguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett, A. D. Long, B. S. Gaut, The molecular diversity of adaptive convergence. *Science (1979)*. **335** (2012), doi:10.1126/science.1212986.
25. V. F. Andersen, J. E. Anderson, T. J. Wallington, S. A. Mueller, O. J. Nielsen, "Vapor pressures of alcohol-gasoline blends" in *Energy and Fuels* (2010), vol. 24.
26. A. Baez, K. M. Cho, J. C. Liao, High-flux isobutanol production using engineered Escherichia coli: A bioreactor study with in situ product removal. *Appl Microbiol Biotechnol*. **90**, 1681–1690 (2011).
27. G. Stephanopoulos, Challenges in engineering microbes for biofuels production. *Science (1979)*. **315** (2007), pp. 801–804.
28. X. Wang, L. P. Yomano, J. Y. Lee, S. W. York, H. Zheng, M. T. Mullinnix, K. T. Shanmugam, L. O. Ingram, Engineering furfural tolerance in Escherichia coli improves the fermentation of lignocellulosic sugars into renewable chemicals. *Proc Natl Acad Sci U S A*. **110** (2013), doi:10.1073/pnas.1217958110.
29. H. S. Song, J. M. Jeon, H. J. Kim, S. K. Bhatia, G. Sathiyanarayanan, J. Kim, J. Won Hong, Y. Gi Hong, K. Young Choi, Y. G. Kim, W. Kim, Y. H. Yang, Increase in furfural tolerance by combinatorial overexpression

of NAD salvage pathway enzymes in engineered isobutanol-producing E. coli. *Bioresour Technol*. **245**, 1430–1435 (2017).

30. T. Modig, G. Lide; N, M. J. Taherzadeh, "Inhibition effects of furfural on alcohol dehydrogenase, aldehyde dehydrogenase and pyruvate dehydrogenase" (2002), (available at http://portlandpress.com/biochemj/article-pdf/363/3/769/737706/bj3630769.pdf).
31. FEBS Letters - December 20 1993 - Haydon - A mutation causing constitutive synthesis of the pyruvate dehydrogenase complex.
32. Y. Liu, R. Wang, J. Liu, H. Lu, H. Li, Y. Wang, X. Ni, J. Li, Y. Guo, H. Ma, X. Liao, M. Wang, "Base editor enables rational genome-scale functional screening for enhanced industrial phenotypes in Corynebacterium glutamicum" (2022), (available at https://www.science.org).
33. 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).
34. M. T. Nguyen Tran, M. K. N. Mohd Khalid, Q. Wang, J. K. R. Walker, G. E. Lidgerwood, K. L. Dilworth, L. Lisowski, A. Pébay, A. W. Hewitt, Engineering domain-inlaid SaCas9 adenine base editors with reduced RNA off-targets and increased on-target DNA editing. *Nat Commun*. **11** (2020), doi:10.1038/s41467-020-18715-y.
35. M. Arbab, M. W. Shen, B. Mok, C. Wilson, Ż. Matuszek, C. A. Cassa, D. R. Liu, Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. *Cell*. **182**, 463-480.e30 (2020).
36. D. Zhao, J. Li, S. Li, X. Xin, M. Hu, M. A. Price, S. J. Rosser, C. Bi, X. Zhang, New base editors change C to A in bacteria and C to G in mammalian cells. *Nat Biotechnol* (2020), doi:10.1038/s41587-020-0592-2.
37. M. Fang, T. Wang, C. Zhang, J. Bai, X. Zheng, X. Zhao, C. Lou, X. H. Xing, Intermediate-sensor assisted push-pull strategy and its application in heterologous deoxyviolacein production in Escherichia coli. *Metab Eng*. **33** (2016), doi:10.1016/j.ymben.2015.10.006.
38. S. Siedler, N. K. Khatri, A. Zsohár, I. Kjærbølling, M. Vogt, P. Hammar, C. F. Nielsen, J. Marienhagen, M. O. A. Sommer, H. N. Joensson, Development of a Bacterial Biosensor for Rapid Screening of Yeast p-Coumaric Acid Production. *ACS Synth Biol*. **6** (2017), doi:10.1021/acssynbio.7b00009.
39. J. M. Wagner, L. Liu, S. F. Yuan, M. v. Venkataraman, A. R. Abate, H. S. Alper, A comparative analysis of single cell and droplet-based FACS for improving production phenotypes: Riboflavin overproduction in Yarrowia lipolytica. *Metab Eng*. **47** (2018), doi:10.1016/j.ymben.2018.04.015.
40. C. Ma, Z. L. Tan, Y. Lin, S. Han, X. Xing, C. Zhang, Gel microdroplet–based high-throughput screening for directed evolution of xylanase-producing Pichia pastoris. *J Biosci Bioeng*. **128** (2019), doi:10.1016/j.jbiosc.2019.05.008.
41. O. Tenaillon, A. Rodríguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett, A. D. Long, B. S. Gaut, The molecular diversity of adaptive convergence. *Science (1979)*. **335** (2012), doi:10.1126/science.1212986.

Acknowledgments

The authors would like to thank associate Prof. Qiang Li in the department of Chemical Engineering, Tsinghua University for the support of Bioscreen C MBR.

Funding:
This work was supported by the National Key R&D Program of China [2018YFA0901500] and National Natural Science Foundation of China [U2032210].

Author contributions: C. Z., X. L. and Y. Y. conceived and designed this project. Y. Y., X. L. and S. L. performed the experiments. Y. Y. analyzed the data and wrote the initial manuscript draft. X. L., X. X. and C. Z. supervised the research and gave advice to the experiments and manuscript. All authors revised and approved the manuscript.

Competing interests: The authors declared no competing interests.

Data and materials availability: All data necessary for this work are available in the main text or the supplementary materials. The raw NGS data has been uploaded to NCBI (SRA:). Any other data or materials related to this work are available upon request from the corresponding author.

Fig.s and Tables

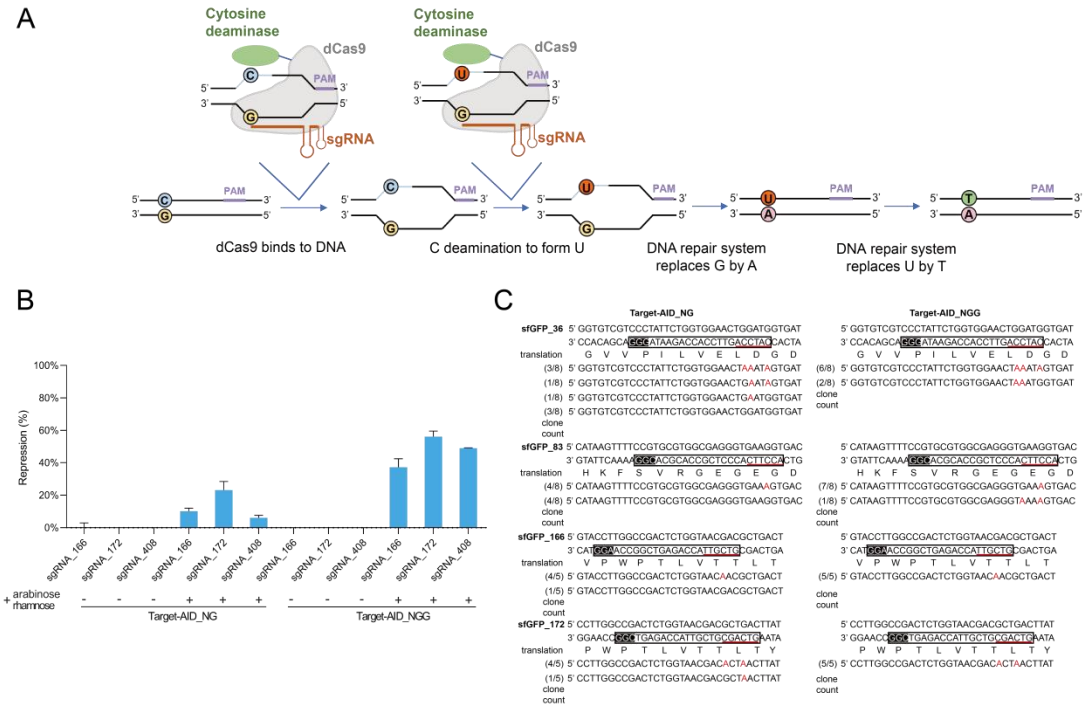

Fig. 1. Dual-induced Target-AID system in *E. coli*. (**A**) Target-AID mediated conversion of cytosine to thymine. (**B**) Characterization of fluorescence repression by Target-AID_NG and Target-AID_NGG. Target-AID_NG, base editors with dCas9 variant with NG PAM; Target-AID_NG, base editors with dCas9 with NGG PAM. (**C**) Editing efficiency confirming of Target-AID_NG and Target-AID_NGG by Sanger sequencing. sgRNA targeting region and the PAM were framed by white and black boxes, respectively. Editing window was underlined in red and the mutated bases were marked by red characters.

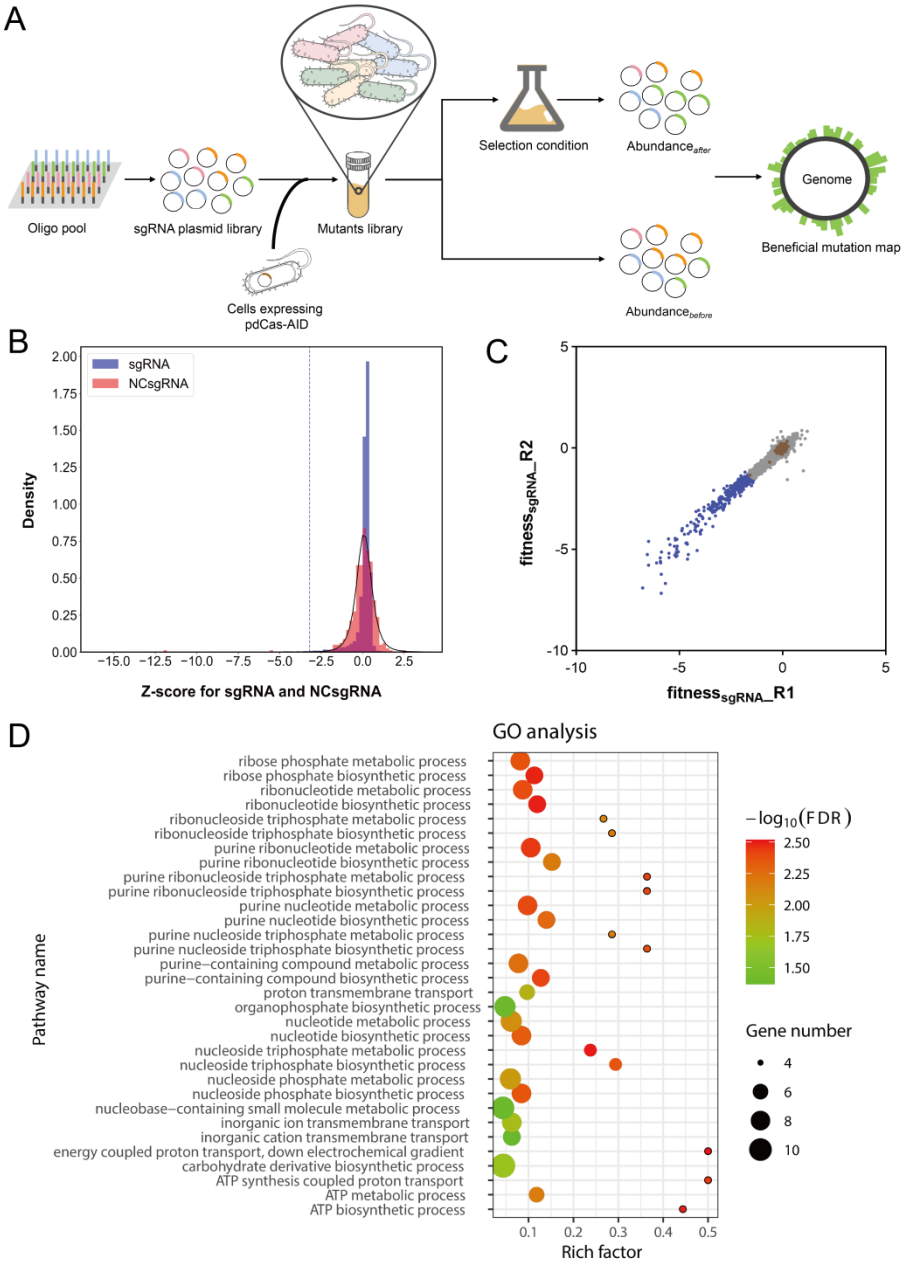

Fig. 2. High-throughput functional screening in *E. coli* using STOP_library. (**A**) Scheme of the Target-AID mediated pooled screening. (**B**) Distribution of sgRNA z-scores in the screening of essential genes. The sgRNAs to the left of the blue dotted line were considered to show a significantly decreased fitness. (**C**) sgRNA fitness of two replicates in essential gene screening. Brown and blue dots represent non-targeting (NC) sgRNAs and those with significantly decreased fitness. (**D**) GO analysis of the remaining 50 unessential genes.

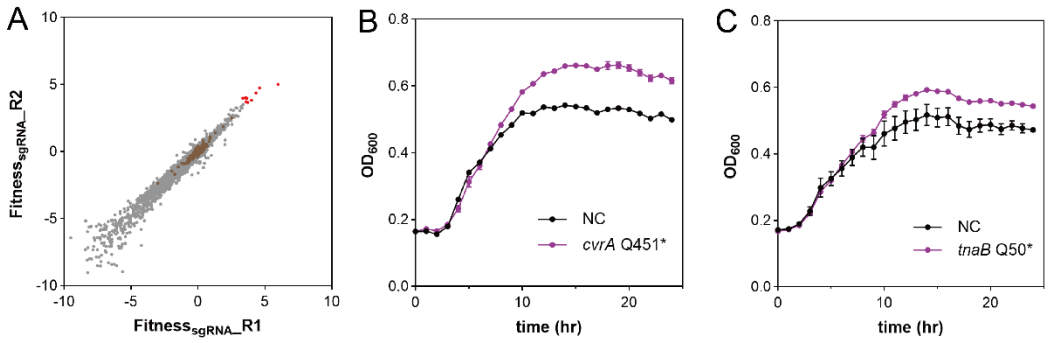

Fig. 3. Loss-of-function screening using STOP_library in the presence of formate. (**A**) sgRNA fitness of two replicates in the screening under formate stress. Brown and red dots represent NC sgRNAs and the top 10 enriched sgRNAs. (**B**) Growth verification of cvrA_Q451* variant under formate stress using STOP_library. (**C**) Growth verification of tnaB_Q50* variant under formate stress using STOP_library.

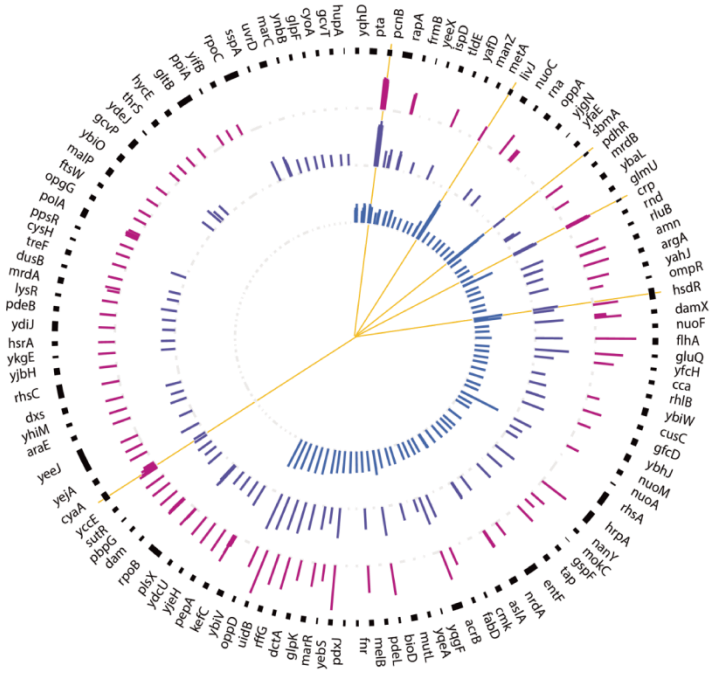

Fig. 4. Fitness of top 200 enriched sgRNA and the corresponding target genes under different conditions. From the outer circle to inner circle, the height of the bars represents the fitness of variants in the presence of formate, isobutanol and furfural. Yellow lines highlighted the genes with multiple enriched sgRNAs.

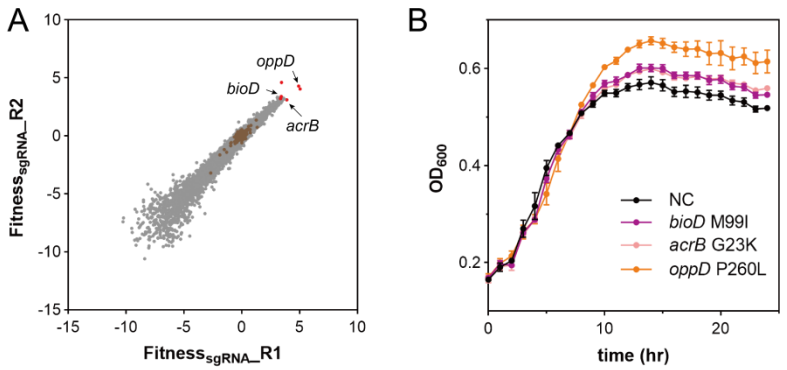

Fig. 5. Functional screening using AID_library in the presence of formate. (**A**) sgRNA fitness of two replicates in the screening under formate stress. Brown and red dots represent NC sgRNAs and the highly enriched sgRNAs. (**B**) Reconstruction and growth verification of enriched variants under formate stress.

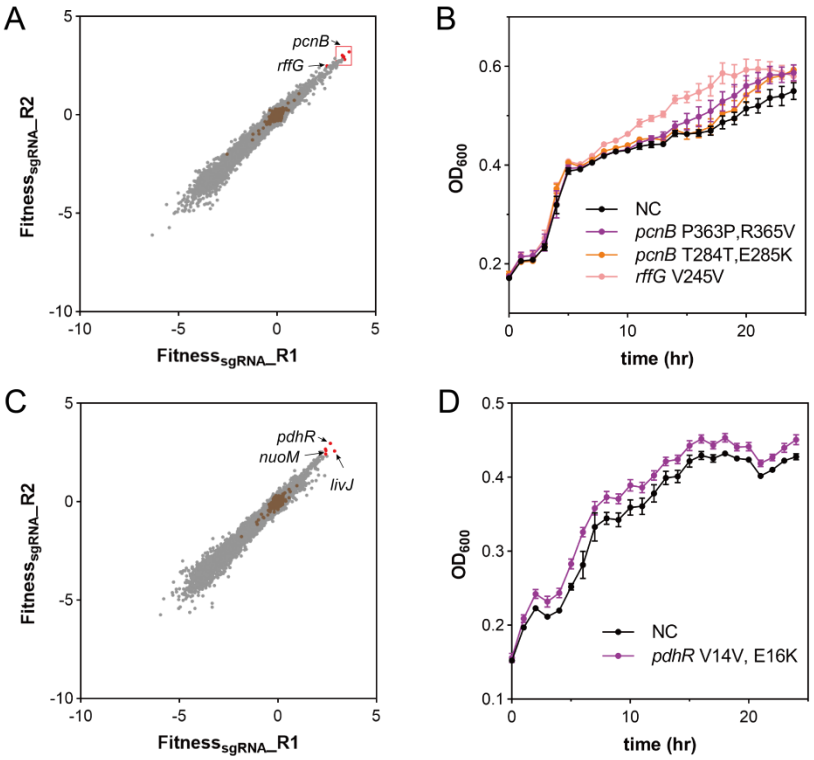

Fig. 6. Functional screening using AID_library in the presence of isobutanol or furfural. (**A**) sgRNA fitness of two replicates in the screening under isobutanol stress. Brown and red dots represent NC sgRNAs and the highly enriched sgRNAs. (**B**) Reconstructed variants with enhanced growth in the presence of isobutanol. (**C**) sgRNA fitness of two replicates in the screening under furfural stress. (**D**) Reconstructed variant with enhanced growth in the presence of furfural.