Retron-mediated multiplex genome editing and continuous evolution in *Escherichia coli*

Wenqian Liu[1,2], Siqi Zuo[1,2], Youran Shao[1,2], Ke Bi[1,2], Jiarun Zhao[1,2], Lei Huang[1,2], Zhinan Xu [1,2,*] and Jiazhang Lian [1,2,3,*]

[1]Key Laboratory of Biomass Chemical Engineering of Ministry of Education, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, China, [2]Institute of Bioengineering, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, China and [3]ZJU-Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou 311215, China

ABSTRACT

While there are several genome editing techniques available, few are suitable for dynamic and simultaneous mutagenesis of arbitrary targeted sequences in prokaryotes. Here, to address these limitations, we present a versatile and multiplex retron-mediated genome editing system (REGES). First, through systematic optimization of REGES, we achieve efficiency of ~100%, 85 ± 3%, 69 ± 14% and 25 ± 14% for single-, double-, triple- and quadruple-locus genome editing, respectively. In addition, we employ REGES to generate pooled and barcoded variant libraries with degenerate RBS sequences to fine-tune the expression level of endogenous and exogenous genes, such as transcriptional factors to improve ethanol tolerance and biotin biosynthesis. Finally, we demonstrate REGES-mediated continuous *in vivo* protein evolution, by combining retron, polymerase-mediated base editing and error-prone transcription. By these case studies, we demonstrate REGES as a powerful multiplex genome editing and continuous evolution tool with broad applications in synthetic biology and metabolic engineering.

GRAPHICAL ABSTRACT
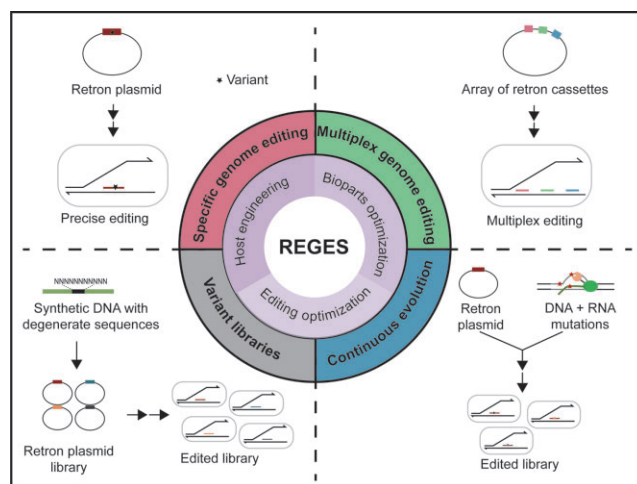
INTRODUCTION

With increasing interest in gene therapy, metabolic engineering and synthetic biology (1–3), there are expanding and continued demands for novel and efficient gene editing tools. Classic genetic engineering methods depend on homology-directed repair and often combine with phage-derived recombinases (e.g. RecET or Lambda Red) (4,5) to improve the recombination efficiency. Unfortunately, these methods often suffer from low efficiency and the most commonly used Cre/loxP and FLP/FRT systems will leave scars in the genome (6,7), hindering their broad applications in genetic engineering. To construct scarless mutant strains with high efficiency, genome-editing tools employing specific nucleases have been developed recently. These include zinc finger nucleases (ZFNs) (8), transcription activator-like effector nucleases (TALENs) (9) and clustered regularly interspaced short palindromic repeat-associated nucleases (CRISPR) (10–12). In contrast to ZFNs (13) and TALENs

---

*To whom correspondence should be addressed. Tel: +86 571 82993283; Fax: +86 571 82993283; Email: jzlian@zju.edu.cn
Correspondence may also be addressed to Zhinan Xu. Email: znxu@zju.edu.cn

(9), CRISPR-mediated editing only requires for changing the sequences of a short single guide RNA (sgRNA) for different targets. As a result, CRISPR has been rapidly becoming one of the most commonly used editing methods, with applications in almost all kingdoms of life.

Substitution mutations are indispensable and are ubiquitous in natural variation, which can alter the function of coding and regulatory sequences (14,15). In addition, most genetic mutation diseases in humans are caused by substitution mutations (16). However, current genome engineering tools still have limited abilities to introduce substitution mutations. The efficiency of oligonucleotide-mediated point mutations is <3%, even when combined with CRISPR (17). To address this challenge, base editors based on the CRISPR system have been developed (18). In this system, C to T mutations can be achieved with cytosine deaminase (CBE) (19,20), with the adenine base editor (ABE, A to G mutations) developed afterwards (21). However, base editors are restricted to some specific nucleotide substitutions in a short window. To enable all 12 types of substitutions, prime editor was developed by fusing a catalytically impaired Cas9 endonuclease to an engineered reverse transcriptase (22). Although well established in mammalian cells, prime editor suffers from poor performance in prokaryotes (e.g. *Escherichia coli*), with substitution, insertion and deletion reported with an editing efficiency of 6.8%, 12.2% and 26%, respectively (23). Multiplex automated genome engineering (MAGE), enabling nucleotide substitution with chemically synthesized single-stranded DNAs (ssDNAs), is one of the most efficient methods in introducing substitution mutation (24). Nevertheless, multiplex substitutions are introduced in multiple rounds of ssDNA transformation and automation devices are generally required.

Compared with the introduction of chemically synthesized ssDNAs *in vitro*, the generation of ssDNAs *in vivo* is expected to benefit recombineering-based genome editing, such as higher editing efficiency and easier manipulation. The retron cassette from *E. coli* BL21 is such a representative example that can generate multi-copy ssDNAs (msDNAs) *in vivo* (25). Retrons are found to mediate antiphage defense in bacteria (26) and the structure and defense mechanisms have been elucidated recently (27). Retron msDNAs can generate functional recombineering donors to introduce specific mutations (Supplementary Table S1), but the editing efficiency was reported to be low (28,29). Recently, by introducing a stronger ribosomal binding site (RBS) to overexpress the single-stranded annealing protein (SSAP) *bet* from Lambda phage (30), HiSCRIBE (high-efficiency Synthetic Cellular Recorders Integrating Biological Events) increased editing efficiency up to 60%. In addition, retron library recombineering (RLR) was reported to generate pooled and barcoded variant libraries with high efficiency (31). However, the retron system is still not systematically optimized, with an editing frequency of 3×10^{-7} for simultaneous two-locus editing (28), limiting their applications in modifying several single nucleotide polymorphisms at a time to treat a genetic disease (19), regulating many endogenous genes at once to trigger changes in cellular phenotype (32) or optimizing complex metabolic regulatory networks (24). Considering the continuous editing capability, the retron system is expected to enable multiplex genome editing.

In this work, we establish an efficient retron-mediated genome editing system (REGES) in *E. coli* (Figure 1). We first optimize the retron system in a modular and systematic manner, and the incorporation frequency of mutations was increased to ~100% for genomic substitutions, insertions, deletions as well as their combinations. In addition, by assembling an array of retron cassettes, we employ REGES for multiplex genome editing and can edit four genes with an efficiency of 25 ± 14%, representing the highest number of precise substitutions to be introduced in a single round of genome editing in *E. coli*. We then demonstrate REGES for the construction of pooled and barcoded variant libraries with degenerate RBS sequences to fine-tune the expression level of endogenous and exogenous genes, such as transcription factors to construct ethanol resistant strains and biotin-overproducing strains. Finally, we explored the capability of REGES for continuous *in vivo* protein evolution, via the combination of retron, orthogonal mutagenic RNA polymerase and cytidine deaminase to introduce random mutagenesis during the replication and transcription stages of *in vivo* ssDNA generation.

MATERIALS AND METHODS

Materials and cultivation conditions

Unless specifically mentioned, strains were cultivated at 37°C, and genome editing experiments were performed at 30°C. MacConkey agar medium (peptone 20 g l^{-1}, bile salts 1.5 g l^{-1}, sodium chloride 5 g l^{-1}, agar 13.5 g l^{-1}, neutral red 0.03 g l^{-1} and crystal violet 0.001 g l^{-1}) was used to determine the editing efficiency of *galK* mutations. Antibiotics, when needed, were added at the following concentrations: ampicillin, 100 mg l^{-1}; chloramphenicol, 34 mg l^{-1}; kanamycin, 50 mg l^{-1}; spectinomycin, 50 mg l^{-1}; streptomycin, 25 mg l^{-1} and sodium dodecyl sulfate, 0.1% v/v. PCR was performed with the high-fidelity enzyme PrimeSTAR® Max DNA Polymerase (Takara, Tokyo, Japan). Oligonucleotides were synthesized by TSINGKE Biological Technology (Hangzhou, China). All chemicals were purchased from Sigma (Sigma-Aldrich, St. Louis, MO).

Strain construction

E. coli strains used in this study were listed in Supplementary Table S2. DH5α (F$^-$ *end*A1 *sup*E44 *thi*-1 *recA*1 *relA*1 *gyr*A96 *deoR* phoA Φ80d*lacZ* Δ M15 Δ (*lac*ZYA-*arg*F) U169, *hsdR*17 (rK$^-$, mK$^+$), λ$^-$) was used for plasmid construction, and WT MG1655 (F$^-$ λ$^-$*ilvG$^-$ rfb*-50 *rph*-1) was used as the parental strain for REGES. All the basic strains in this study were constructed by CRISPR/Cas9 with the pTargetT and pCas system as previously described (11). To construct the Δ*recJ*, Δ*mutS*, Δ*xonA* and Δ*xseA* strains, DNA fragments including 500-bp upstream and 500-bp downstream were PCR amplified and assembled to generate donor DNAs. With the help of specific sgRNA expressed by pTargetT, *recJ*, *mutS*, *xonA* and *xseA* were individually knocked out in MG1655. MG06 (Δ*recJ* Δ*xonA*) was constructed by knocking out *recJ* in MG04 (Δ*xonA*). Donor DNAs used for the construction of MG01, MG07, MG08
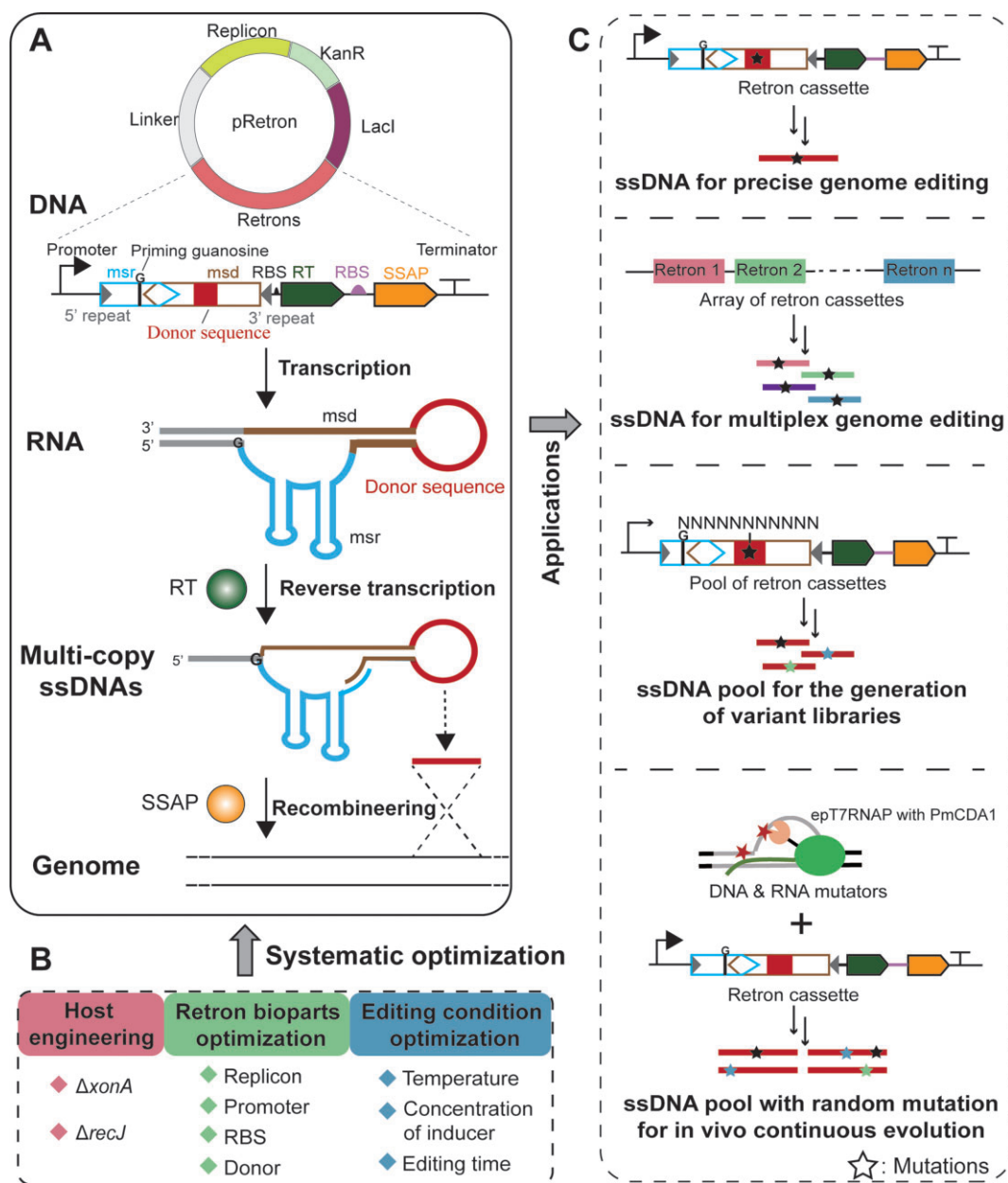
Figure 1. Design and establishment of REGES for genome editing and genome evolution applications. (**A**) Schematic representation of REGES. Retron is encoded as a polycistronic transcriptional cassette containing a promoter, self-complementary regions (grey), an *msr* (blue) with a priming guanosine residue (black), an *msd* (brown) with the desired donor sequence (red lines), an RBS to initiate the translation of reverse transcriptase (RT), an RT, an RBS (purple) to initiate the translation of SSAP, an SSAP and a terminator. Retron can generate ssDNAs (red lines) *in vivo*, which are subsequently incorporated into the targeted loci in the genome. (**B**) Systematic optimization of REGES parameters, including host engineering, retron bioparts optimization and editing condition optimization. (**C**) Through the design of *in vivo* generated ssDNAs, REGES enables versatile synthetic biology and metabolic engineering applications, such as precise genome editing (substitution, insertion and/or deletion), multiplex genome editing, ssDNA pool with degenerate RBS sequences for regulating the expression of endogenous and exogenous genes as well as ssDNA pool with random mutations for continuous *in vivo* protein evolution.

and MG11 were obtained by assembly of the mutant sequences of each gene (*galK*$_{off}$, *gfp*$_{off}$, *rfp*$_{off}$ and *cat*$_{off}$) with their upstream and downstream sequences. Transformants were verified by colony PCR and DNA sequencing.

Plasmid cloning

All constructs, template sequences and primers used in this study are listed in Supplementary Tables S3–S5. Plas-

mids were constructed by digestion/ligation, Gibson assembly or Golden-Gate assembly. The REGES editing plasmids expressing msr, msd, *RT* and *SSAP* was constructed as following: the WT retron cassette (Ec86) expressing *msr*, *msd* and *RT* was amplified from the genome of BL21(DE3). The *bet* gene was amplified from pKD46 (Novagen). The retron cassette and *bet* were assembled into a medium copy-number plasmid pET28a to construct pRE01. The template sequences for ssDNA production were inserted into

pRE01 between the *Eco*R I (New England Biolabs, Ipswich, MA) and *Hind* III sites. *CspRecT*, I-*Sce* I, promoter sequences, 5′UTR sequences and replicons were chemically synthesized by TSINGKE Biological Technology (Hangzhou, China). All plasmids harboring various targeting sequences were cloned into *Eco*R I and *Hind* III restriction sites. To construct multiplex editing plasmid, target plasmids (pRE13, pRE22) without RT and bet were first constructed. The replacement of promoter and terminator was performed by Gibson assembly. The reaction mixture was then purified and added to 100 μl DH5α competent cells, and electroporation was done in a 2-mm Gene Pulser cuvette (Bio-rad) under 2.5 kV, 25 μF and 200 Ω. Afterwards, 800 μl of SOC liquid medium was added to recover the cells at 37°C. After about 1 h, the transformants were plated onto LB solid medium containing the corresponding antibiotics.

Genome editing by REGES

Unless otherwise mentioned, the heat shock transformation method was used to transform editing plasmids to the corresponding *E. coli* strains. Competent cells were prepared according to the method as previously described (33). For editing in liquid medium, a single colony was inoculated to 50 ml of LB medium containing 50 mg l^{-1} kanamycin and 1 mM IPTG. After editing for ∼16 h, the mixed culture was streak out onto agar plate to isolate mono-clones. Alternatively, for editing in solid medium, *E. coli* strains were spread onto agar plates containing 50 mg l^{-1} kanamycin and 1 mM IPTG. After 16 h of growth, single colonies were isolated for genotyping analysis.

Editing efficiency assay

The *galK* reversion assay was performed as previously described (28). Briefly, strains after editing were plated onto MacConkey agar plates containing 50 mg l^{-1} kanamycin. The *galK* mutation efficiency was calculated by calculating the ratio of purple colonies to the total number of colonies. In the *gfp* and *rfp* reversion assay, strains after editing were plated on LB agar plates supplemented with kanamycin. The editing efficiency was the percentage of strains showing green fluorescence (or red fluorescence). The editing frequency of *rpsL* conversion was determined by plating cells on LB agar plates with or without streptomycin and calculating the ratio of streptomycin-resistant strains (plates with streptomycin) to the total number (plates without streptomycin). For *cat* reversion assay, cultures were spotted on LB agar plates with or without chloramphenicol, and the ratio of chloramphenicol-resistant strains to the total number was determined. The *tolC* editing efficiency was calculated by the ratio of SDS-sensitive strains to the total number. In addition to the phenotype-based validation, genome editing results were further verified by diagnostic PCR and Sanger sequencing.

Regulation of RFP expression and fluorescence intensity measurement

DNA containing degenerate RBS sequences (NNNNNNNNNNN) flanked by homologous regions on each side was synthesized and ligated into the *Eco*R I and *Hind* III sites of pRE15 to generate the pool of editing plasmid. After editing for 16 h, *E. coli* strains were spread into agar plates, and single colonies were picked, inoculated and cultured in 96-well plates overnight. To measure the fluorescence intensity, *E. coli* strains were washed twice, diluted in sterile water and measured at 555–584 nm. The fluorescence intensity (relative fluorescence units; RFU) was normalized to cell density (OD$_{600}$).

Regulation of IscR expression and isolation of *iscR* mutant strains

Barcoded plasmid library prepared as above was electroporated into BL21(DE3) △*recJ*△*xonA*/pAra-bioB-pMB1 and plated on LB plates with 1 mM IPTG. After 16 h of editing, the resultant colonies were scraped from the plates. Selection was performed by diluting the cell culture 100-fold into 5 ml LB with 100 mM arabinose to induce *bioB* overexpression to the cytotoxicity level. After isolating single colonies of the viable strains, they were picked for further characterization of biotin production capacity.

Quantification of biotin production

Bition quantification was performed as previously described with few modifications using *Lactobacillus plantarum* ATCC8014 as the biotin reporter strain (34). Sample supernatant was diluted ∼5000-fold and then mixed with the bioassay medium in a microtiter plate. Plates were incubated at dark at 37°C for 24 h. OD$_{600}$ was measured and biotin concentration was calculated using the standard curve with known concentration of biotin in the bioassay medium.

Plasmid elimination assay

Strains were inoculated into 5 ml of LB liquid medium containing arabinose. After culturing at 37°C for 1-9 h the diluted culture was spread onto an antibiotic-free LB solid medium and incubated overnight at 37°C. Colonies were then picked to kanamycin-containing and kanamycin-free plates to test their sensitivity to kanamycin. The elimination efficiency of the donor plasmids was determined based on the ratio of kanamycin-sensitive strains to the total number of *E. coli* strains. Plasmid elimination was further verified by diagnostic PCR with specific primers.

Construction of lycopene-producing strains and quantification of lycopene production

After simultaneous editing of multiple genomic loci, *E. coli* strains were spread onto LB agar plates containing kanamycin. The multiplex editing efficiency was determined by diagnostic PCR with specific primers and further verified by Sanger sequencing. After elimination of the editing plasmids, pAC-lyc containing the lycopene biosynthetic pathway was transformed to enable lycopene production (24). Lycopene extraction and quantification was performed as previously described with few modifications (24). 1 ml of cells were centrifuged at 12 000 rpm for 1 min and washed

twice with PBS buffer. The resultant cell pellets were re-suspended in 300 μl of acetone and incubated at 55°C for 15 min. After centrifuge at 12 000 rpm for 1 min, the absorbance of the supernatant was measured at 470 nm with a Varioskan Flash Multimode Reader (Thermo Scientific, Waltham, USA). Cell dry weight (DCW) was calculated based on a conversion coefficient of 0.36 for DCW/OD (35,36). The lycopene yield of all strains was normalized to DCW.

Evolution of sacB and viability assay

The mutator plasmid (Mutator-DNA or Mutator-RNA) and editing plasmid (pRsacB) were co-transformed into MG14. Cells harboring these plasmids were grown overnight in 5 ml of LB supplemented with chloramphenicol and kanamycin. The overnight cultures (50 μl) were mixed with 5 ml of LB supplemented with chloramphenicol, kanamycin, 1 mM IPTG and 100 mM arabinose and incubated at 30°C for 8 hours (Cycle#1). The 1/100 dilution-incubation cycle was repeated, and samples were taken at the end of each cycle. To determine the mutation rate of sacB, cells at each cycle were serial 10-fold diluted and then spotted onto selective (10 g l^{-1} sucrose) or nonselective (no sucrose) plates. The number of sucrose-resistant strains as a percentage of the total number of cells was calculated. Mutations in sacB were further determined by Sanger sequencing.

Statistics and reproducibility

All experimental data were at least in triplicate and expressed as mean ± standard error. All data analyses were performed by Excel or OriginPro.

RESULTS

### Design of REGES for multiplex genome editing and evolution

Retrons are encoded as a polycistronic transcriptional cassette, which contains a promoter, an *msr* (including a priming guanosine residue), an *msd*, self-complementary regions (5′ repeat and 3′ repeat), an RBS for translating reverse transcriptase (RT), an RT, an RBS for translating SSAP, an SSAP and a terminator (Figure 1A). After transcription, the msr-msd RNA folds into a secondary structure, which can be recognized by RT. Then RT uses the guanosine residue in msr as a priming site to reverse transcribe the msr sequence and produce a hybrid RNA–ssDNA molecule. We can insert templates (donor sequence) into msd to generate multi-copy ssDNAs that will subsequently be incorporated into the targeted genes (Figure 1A). To improve the recombineering efficiency in *E. coli*, we introduced *bet*, an SSAP from Lambda phage (37) to bind ssDNA fragments and anneal complementary sequences. Using standard synthetic biology parts, key parameters of REGES including host, retron bioparts and editing conditions can be optimized in a modular and systematic manner (Figure 1B). The high efficiency of REGES enables a series of synthetic biology and metabolic engineering applications. First, the production of ssDNAs with desired mutations enables precise genome editing, while ssDNA molecules targeted to

multiple loci can be employed for multiplex genome editing. Additionally, ssDNA pools can facilitate the generation of pooled and barcoded variant libraries, such as degenerate RBS sequences to fine-tune the expression level of endogenous and exogenous genes for strain engineering purposes. Finally, the incorporation of ssDNA pools with random mutations (through DNA-dependent transcription and RNA-dependent reverse transcription) can be utilized for gene-specific, continuous *in vivo* evolution (Figure 1C).

Systematic optimization of REGES for efficient genome editing

To facilitate the identification of the desired editing events, we constructed a reporter strain MG01 by introducing two continuous premature stop codons (L187*-L188*) in the galactokinase encoding gene (*galK*). On the MacConkey plate, strains with the function of *galK* recovered will turn purple (30,38). For the initially constructed lac-driven retron editing system, we could not detect any target edits on the MacConkey plate, indicating that the editing efficiency was less than 0.1% (Figure 2A and Supplementary Figure S1A). In previous studies, the efficiency of ssDNA-based recombination could be improved by removing nuclease genes (e.g. *xseA*, *recJ* and *xonA*) for the degradation of exogenous DNAs (39–42) as well as mismatch repair gene *mutS* (39). To explore these strategies in REGES, we constructed a series of strains with *xseA*, *recJ*, *xonA* and *mutS* being individually knocked out. The knockout of *xseA* (42), an ssDNA-specific exonuclease responsible for converting large ssDNA substrates into smaller oligonucleotides, no enhancement in recombinant efficiency was observed (Figure 2B). On the other hand, we found that the inactivation of *xonA* or *recJ*, encoding 3′ or 5′ ssDNA exonucleases (39), respectively, significantly improved the editing efficiency (Figure 2B), whose simultaneous deletion increased the editing efficiency by ∼130-fold over the original system (Figure 2B). We observed no significant differences in the *mutS* deficient strain (Figure 2B), indicating that MutS does not recognize sequences with insertions or mismatches larger than 4 nucleotides (43). Although the precise mechanism of retron is still unclear (44–46), our results showed ssDNAs were hybridized to the lagging strand of the replication fork, which is akin to the Lambda Red recombination mechanism (47). Accordingly, ssDNAs targeting to the lagging strand of *galK* achieved an editing frequency of 15 ± 3%, which was ∼300-fold higher than that to the leading strand (0.05 ± 0.03%, Figure 2B). Thus, according to the genome coordinate, we come to some basic rules to design the template sequences of REGES (Supplementary Figure S1B).

To further improve genome editing efficiency, we systematically optimized each module of REGES at replication (plasmid copy numbers), transcription (promoter activities) and translation (RBS strengths) levels. We achieved an editing efficiency (i.e. proportions of purple cells) of 29 ± 6% when p15A (∼5–10 copies) was used. In contrast, plasmids with pMB1 (∼20–40 copies) and pUC (∼600 copies) resulted in editing frequencies of 13 ± 4% and 7 ± 3%, respectively (Figure 2B). To further match the recombination capacity with the reverse transcription
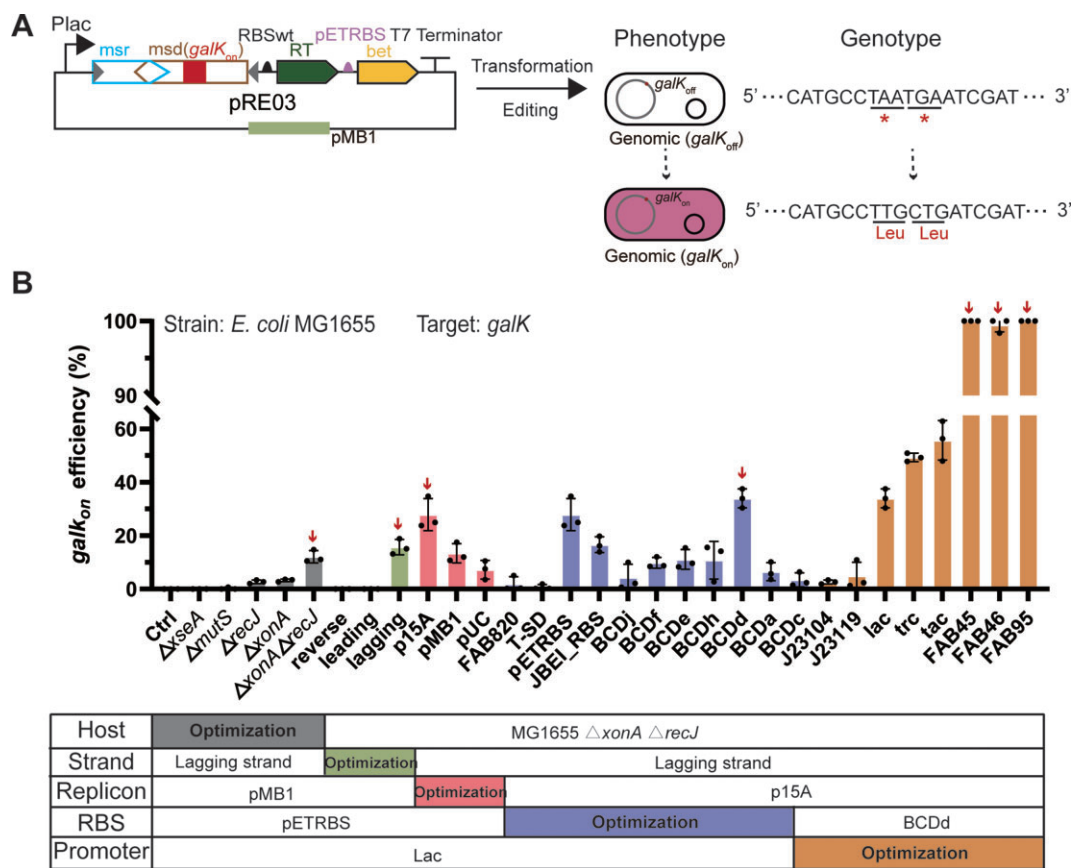
Figure 2. Systematic optimization of REGES for efficient genome editing. (**A**) A *galK* reversion assay was used to measure the efficiency of DNA editing within living cells. The *galK* stop codon sequences (TAATGA) were substituted to coding sequences (TTGCTG) using REGES. (**B**) Optimization of each element of REGES to improve genome editing efficiency, including host modification, targeting strand, copy number of retron vector, translational strength of SSAP and transcriptional activity of the retron cassette. pET-RBS: TAAGAAGGAGA. Ctrl represented wild-type MG1655. The best performance for each parameter was labelled with a red arrow.

efficiency, we evaluated 11 RBSs (48,49) with different translation initiation rate (Supplementary Figure S2A) for the expression of *bet* from Lambda phage. Interestingly, BCDd with medium strength resulted in the highest editing efficiency (34 ± 4%, Figure 2B), indicating that the expression level of *bet* should be carefully fine-tuned. Finally, we regulated the transcription of the retron cassette driven by eight promoters (49) of different strengths (Supplementary Figure S2B). As shown in Figure 2B, we could achieve ~100% editing when the strong IPTG-responsive FAB45, FAB46 and FAB95 were used, indicating that high expression of the retron system was beneficial for REGES.

Based on the optimized REGES, we set to evaluate the editing efficiency for a full set of possible engineering events, including substitution, insertion and deletion. While we achieved high efficiency for deletion and substitution, the insertion efficiency was much lower (Figure 3A), which has been reported for other ssDNA-mediated recombineering systems such as MAGE (24,41). Interestingly, when the editing length is <3 nucleotides, the genome editing efficiency was dramatically affected, probably due to the recognition and repair of the mutations by mutS of the MMR system (43).

We then employed REGES to edit more target genes related to important phenotypes in *E. coli*. For example, the mutant RpsL (K43R) conferring *E. coli* streptomycin resistance (50) was obtained with an efficiency of 91 ± 9% (Figure 3B). More importantly, *gfp* (encoding green fluorescent protein), *rfp* (encoding red fluorescent protein) and *tolC* (encoding outer membrane channel) were edited with an efficiency of 74 ± 7%, 94 ± 4% and 62 ± 4%, respectively (Figure 3B). We evaluated the ability of REGES for combinatorial mutations, such as substitution with deletion (tested for both *gfp* and *rfp*) and substitution with insertion (tested for *tolC*), which resulted in 62 ± 4% to 94 ± 4% editing efficiency (Figure 3B). These results demonstrated the power and versatility of REGES for genome editing in *E. coli*.

The last element for optimization was the donor of REGES (also known as the guide sequences). By placing mutations in different positions of the template sequences, we found positions from 20 to 55 in the ssDNAs to be optimal for genome editing (Figure 3C). Although increasing donor length was expected to provide more homology for annealing, we are surprised to find that larger donor impaired genome editing efficiency significantly, particularly when larger than 300 bp (Supplementary Figure S2C). In
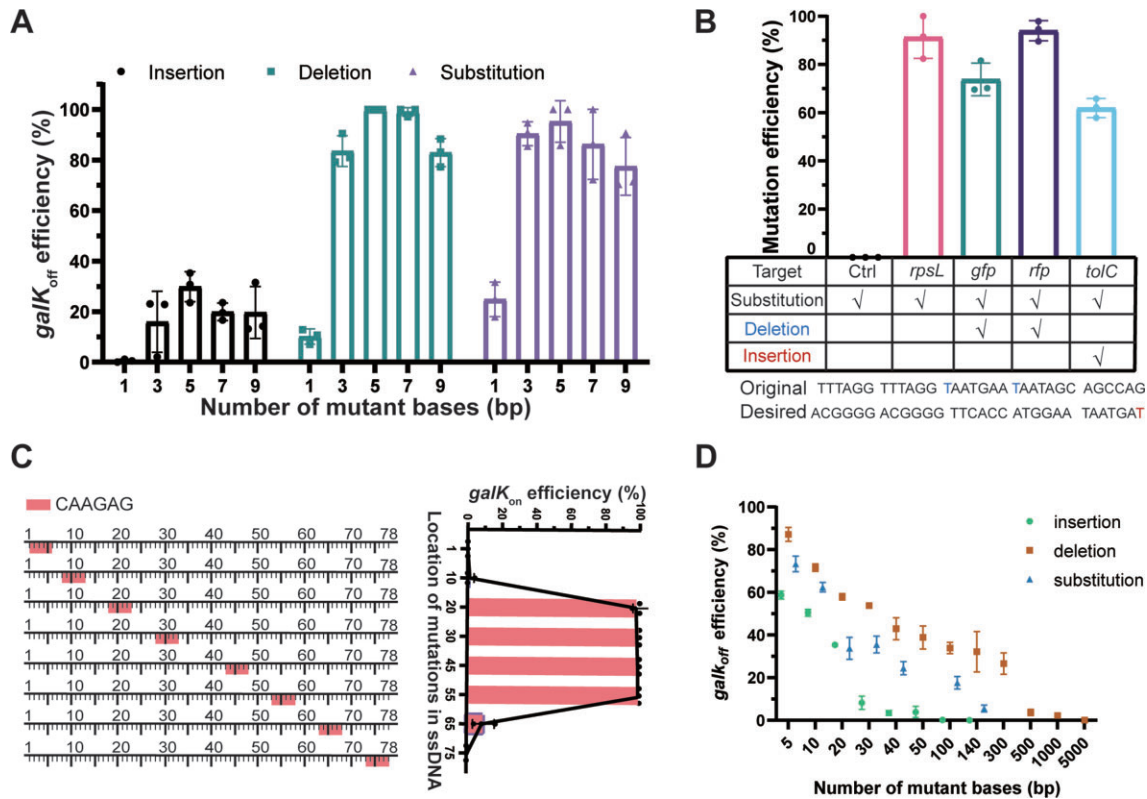
Figure 3. REGES for precise genome editing. (**A**) Editing efficiency of REGES for substitutions, deletions and insertions with different number of mutant bases. To ensure the consistency of the starting strains used for editing, we used a $galK_{on}$ to $galK_{off}$ mutation assay to measure the efficiency of DNA editing. (**B**) Applications of REGES in mutating essential gene and introducing combinatorial mutations (e.g. substitution with deletion as well as substitution with insertion). Ctrl represented REGES with a non-targeting plasmid (i.e. pRE15). (**C**) Effect of the location of mutations in ssDNA on the genome editing efficiency using REGES. A red box represented the mutated bases (CAAGAG) within the 78-bp ssDNA. (**D**) Characterization of editing efficiency as a function of the type and scale of genetic modifications. Error bars represented standard errors of three biological replicates.

other words, we should keep the guide sequences shorter than 300 bp to achieve efficient genome editing in *E. coli*. In addition, we optimized inducer concentration (Supplementary Figure S2D), editing time (Supplementary Figure S2E) and editing temperature (Supplementary Figure S2F) to maximize REGES efficiency. After systematic exploration, we summarized the optimal parameters for the best performance of REGES in Table 1. Briefly, retron cassette is transcribed by a strong promoter (e.g. FAB45, FAB46 and FAB95), with the translation of *bet* initiated by a medium-strength RBS (e.g. BCDd). The optimal window for REGES is less than 300 bp, with the mutation site ideally positioned in the middle of the donor. To induce ssDNA production and continuous editing, IPTG is added at a concentration of 1 mM and genome editing is performed at 30°C for 16 h.

After systematic optimization, we aimed to explore the maximum capacity of REGES. Generally, extended donor sequences allow for a broader spectrum of insertions or substitutions at specific positions within the target gene, but excessively long donor sequences were found to result in decreased editing efficiency of REGES (Supplementary Figure S2C). Considering that a donor of 200 bp achieved an editing efficiency of 87 ± 8% (Supplementary Figure S2C), we adopted this length for maximum capacity evaluation. As the homology regions required for Lambda Red

Table 1. Optimal parameters of REGES for maximal editing efficiency

Parameters	Optimal values
Genome modifications	$\Delta recJ$ & $\Delta xonA$
Promoter	FAB45/FAB46/FAB95
SSAP	bet
RBS of SSAP	BCDd
Replicon	P15A
Donor length[a]	<300 bp
Mutation position in the ssDNA	20 to 55
Concentration of inducer	1 mM IPTG
Temperature	30°C
Editing time	16 h

[a]Unless otherwise mentioned, the donor length used in this study was 78 bp.

recombination is ~30 bp (24,37), the maximum allowable length for insertions and substitutions tested here is ~140 bp. However, we failed to detect any desired modification when attempting to insert a 140 bp random fragment (from *sacB*) into the *galK* locus. Nevertheless, we could insert a 100 bp fragment with low efficiency (Figure 3D). For substitutions, we achieved a mutation efficiency of 6 ± 2% when trying to substitute a 140 bp random sequence (from *sacB*) with the same size in *galK* (Figure 3D). Notably, the efficiency of generating a substitution or deletion modification

is correlated with the size of the mutation. We achieved an efficiency of $27 \pm 5\%$ for the deletion of $\sim$300 bp. Although with extremely low editing efficiency, we could successfully delete up to 5000 bp sequences (Figure 3D).

REGES for multiplex genome editing

After systematic optimization, we then tested the performance of REGES for multiplex genome editing, which was highly desirable for synthetic biology and metabolic engineering applications. To minimize repetitive DNA sequences (32), we used two different REGES constructs, FAB45-retron and FAB46-retron, and T7 terminator of the first REGES cassette was replaced with L2U5H11 terminator (51). When multiple retron cassettes were arranged as tandem arrays, these two different constructs were used sequentially, leading to reduced homology between the adjacent expression cassettes. For two-locus editing, we chose *rpsL* and *gfp* as the reporter genes, which were simultaneously edited with an efficiency of $85 \pm 3\%$. By assembling another REGES cassette targeting *galK*, we achieved three-locus (*rpsL-gfp-galK*) editing with an efficiency of $69 \pm 7\%$. After the inclusion of a fourth REGES cassette targeting *cat*, we observed four-editing events with an efficiency of $25 \pm 14\%$ (Figure 4A), representing the maximal number of precise mutations to be introduced in a single round of genome editing in *E. coli*. The editing efficiency of each gene in the multiplex editing system was shown in Supplementary Figures S3A, S3B, S3C and S3D, respectively. To facilitate continuous or iterative genome editing, we incorporated an I-*Sce* I-mediated (52) plasmid curing system (Supplementary Figure S3E). Specifically, the expression of the homing endonuclease I-*Sce* I, which recognized and cleaved a specific recognition sequence (TAGGGATAACAGGGTAAT) to introduce double-strand DNA breaks (DSBs) into the retron plasmid, was tightly controlled by an arabinose-inducible promoter. The system enabled $\sim$100% plasmid elimination after 3 h of induction by arabinose (Supplementary Figure S3F).

For practical applications, we employed REGES and the plasmid elimination system for the construction of lycopene overproducing strains (Figure 4B). Based on previous studies, the combination of *idi*, *dxs* and *ytjC* engineering or the combination of *dxs*, *idi*, *dxr*, *rpoS* and *ytjC* engineering were determined to be the most effective in increasing lycopene production (24). We constructed and introduced pRE31 to knock down *ytjC*, and eight out of nine colonies were found to harbor the expected mutations (Figure 4C). After induction of the I-*Sce* I system to eliminate the donor plasmid, the resultant strain MG13 was further transformed with pRE-multi-2E or pRE-multi-4F, which were expected to perform *dxs-idi* double mutations or *idi-dxr-rpoS-dxs* quadruple mutations (Figure 4B). For the dual-editing event, REGES reached efficiencies up to $\sim$100% (Figure 4D); for the quadruple mutation assay, REGES reached efficiencies up to $18 \pm 9\%$ (Figure 4E). In addition, we found that all the selected colonies ($\sim$100%) were edited at least twice, and $60 \pm 14\%$ were edited in at least three loci (Figure 4E). DNA sequencing results showed that both strains were precisely edited at all the predeter-

mined loci (Supplementary Figure S4A and S4B). Compared with the control strain, lycopene production was increased by 3.9-fold and 4.5-fold in the two-locus editing strain and four-locus editing strain, respectively (Supplementary Figure S4C). Notably, it only took us two days to construct a high-yield lycopene producing strain harboring five-locus mutations, indicating the efficiency and time saving of REGES for the construction of microbial cell factories.

REGES for the generation of pooled and barcoded variant libraries

As REGES can efficiently introduce mutations in a relatively wide window, we propose REGES to incorporate a pooled and barcoded library with the desirable mutations into the *E. coli* genome. For example, we can incorporate a library of degenerate RBS sequences to regulate the expression level of endogenous and exogenous genes (Figure 5A). As a proof-of-concept, we constructed a reporter strain MG09 with the RFP expression cassette (harboring the original RBS sequences from pET-RBS, TAAGAAG-GAGA) integrated into the genome. We inserted degenerate RBS sequences (NNNNNNNNNNN) flanked by 34 bp upstream and 33 bp downstream homologous regions of *rfp* into *msd* to generate a pooled and barcoded retron library. After transforming the retron library and editing, we randomly picked colonies for fluorescence measurement and observed an expression range of $\sim$208-fold (Supplementary Figure S5A). Despite the strong translational capacity of pET-RBS, the expression level of RFP could still be enhanced by up to 1.6-fold via REGES-mediated RBS engineering (Figure 5B).

For practical applications, we chose to regulate the expression level of transcriptional factors. In previous studies, global transcription machinery engineering (gTME) has been found to enable the engineering of complex regulatory networks and used to increase protein expression (53), metabolite production (54) and stress resistance (55) via directed evolution. We proposed that modulating the expression of global transcription factors could also generate global diversity of transcription levels and be employed for strain engineering purposes. Here, we constructed retron libraries for fine-tuning the expression level of seven global transcription factors, including CRP (cyclic AMP receptor protein), FNR (fumarate and nitrate reduction regulatory protein), ArcA (aerobic respiration control regulator), IHF (integration host factor), HNS (histone-like nucleoid structuring protein), Fis (factor for inversion stimulation) and LRP (leucine responsive regulatory protein) (56,57), and chose ethanol tolerance as the target phenotype for engineering. Among these seven mutant libraries, we found the FNR mutant library enabled the grown of *E. coli* strains in high concentration of ethanol (50 g l^{-1}, Figure 5C and Supplementary Figure S5B), after which six mutant strains were picked for further characterization. To evaluate the strength of the RBS mutants, we fused the 5' 36-nt of FNR to the 5' end of eGFP (48) and found that all six constructs conferring ethanol resistance expressed lower levels of FNR than that of the wild-type (WT) strain (Figure 5C). Although there has been no report on improving ethanol tolerance by
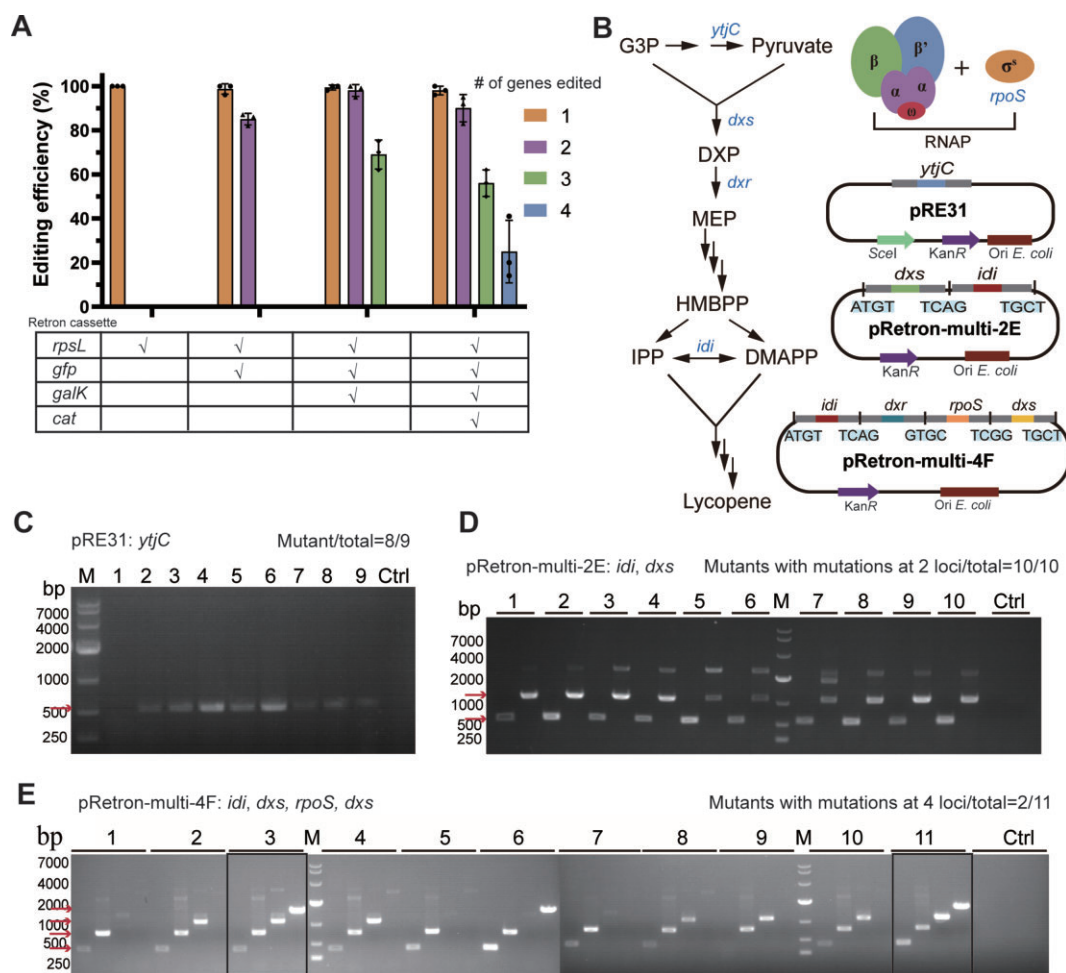
Figure 4. REGES for multiplex genome editing. (**A**) Efficiency distribution of clones in the population with different numbers of gene editing in each editing event, with *rpsL*, *gfp*, *galK* and *cat* chosen as the reporter genes. (**B**) Lycopene biosynthetic pathway and the edited targets. pRE31: A plasmid containing Retron system for editing of *ytjC* and elements for plasmid curing. pRetron-multi-2E: Retron system for simultaneous editing of two RBS (*dxs* and *idi*). pRetron-multi-4F: Retron system for simultaneous editing of four RBS (*idi*, *dxr*, *rpoS* and *dxs*). For the construction of lycopene-overproducing strains, *ytjC* was edited first with pRE31, followed by double editing (*dxs-idi*) with pRetron-multi-2E or quadruple editing (*idi-dxr-rpoS-dxs*) with pRetron-multi-4F. (**C**) PCR verification of *ytjC* editing. Genome editing of *ytjC* resulted in PCR amplicon of 520 bp. (**D**) PCR validation of genome editing by pRetron-multi-2E (*dxs-idi*). Genome editing of *dxs* and *idi* resulted in PCR amplicon of 540 and 1200 bp, respectively. (**E**) PCR validation of genome editing by pRetron-multi-4F (*idi-dxr-rpoS-dxs*). Genome editing of *idi*, *dxs*, *rpoS* and *dxs* resulted in PCR amplicon of 485, 750, 1200 and 1850 bp, respectively. Ctrl represented wild-type MG1655. M represented DNA Marker. Red arrows represented the size of the target bands.

engineering of FNR, our results clearly demonstrated that FNR repression enabled higher ethanol tolerance in *E. coli*. Additionally, the retron itself can serve as a barcode (unique RBS sequence), allowing for direct analysis of mutants from the screened libraries by sequencing the retron plasmids using universal primers. Our REGES-based global transcription engineering provides advantages in establishing genotype and phenotype relationships of the mutant strains.

After successfully demonstrating REGES in an *E. coli* K strain (MG1655), we extended the applications in the genome editing and global transcription engineering of an *E. coli* B strain, BL21(DE3). We aimed to increase the production of biotin, a high-value vitamin metabolite with broad applications. While the final step of biosynthesis pathway, encoded by the *bioB* gene, is the bottlenecke for biotin production (58), its overexpression dramatically inhibits cell growth, probably due to the depletion of FeS-

cluster (58). As IscR is a global regulator responsible for the assembly of FeS-cluster, we constructed a retron library to precisely regulate the expression level of IscR. As shown in Figure 5D, the overexpression of *bioB* completely inhibited cell growth when the inducer arabinose was higher than 50 mM. Luckily, cell growth could be rescued by the introduction of the retron library (Supplementary Figure S5C), indicating that the supply of FeS-cluster was enhanced in the IscR-engineered strains to minimize the cytotoxicity of *bioB* overexpression. Based on the growth-associated selection, we successfully obtained several biotin overproducing strains (Supplementary Figure S5D and Figure 5E). DNA sequencing results verified mutations in the RBS sequences (Supplementary Figure S5E), which were introduced by REGES. In addition, we found a correlation between improved IscR expression and increased biotin production, with maximum titers reaching almost 900 μg l^{-1} in M9 minimal media (Figure 5E).
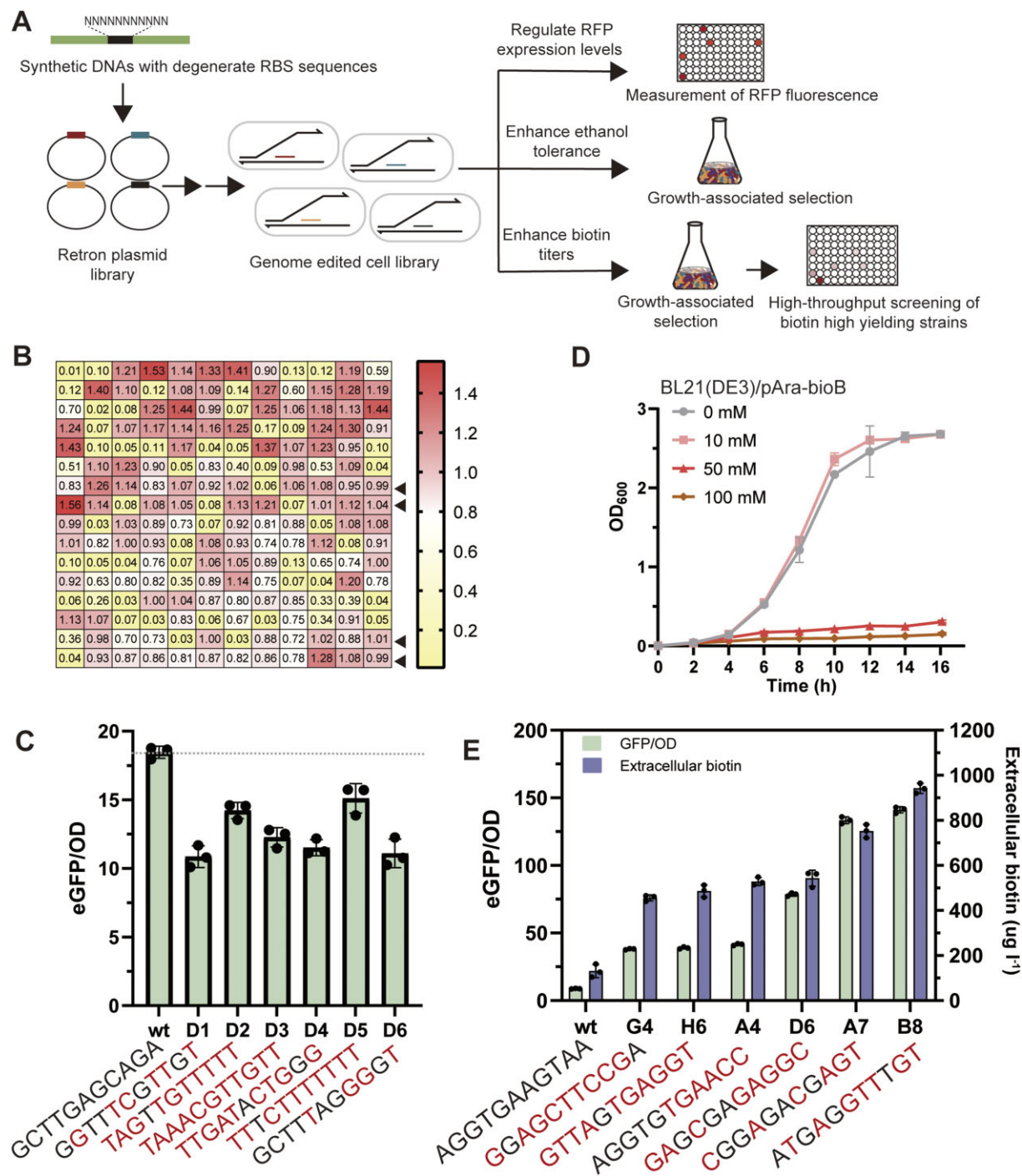
Figure 5. REGES for diversifying the expression level of endogenous and exogenous genes. (**A**) Schematic diagram of the regulation of gene expression levels through the incorporation of a library of degenerate RBS sequences. By mutating RBS sequences, it is possible to regulate the expression level of RFP, FNR to enhance ethanol resistance of *E. coli* and IscR to increase biotin production. (**B**) Fold changes in *rfp* expression level (RFP fluorescence) of the randomly selected strains. RFP fluorescence intensities were normalized to that with pET-RBS (triangle). (**C**) Expression level of eGFP when the translation was initiated by the mutant RBS (underneath the corresponding strain) of fumarate and nitrate reduction regulatory protein (FNR). In this assay, the original promotor of FNR was used. The dashed line represented the expression intensity of the wild-type RBS. (**D**) A growth-associated selection system based on the overexpression of *bioB*. The expression level of bioB was controlled by the concentration of arabinose. (**E**) Translational intensity of IscR and corresponding biotin production titers of the selected mutant strains. The mutant RBS sequence was present underneath the corresponding strain. WT represented the original RBS of *iscR*.

REGES for gene-specific continuous *in vivo* evolution

Finally, we employed REGES for continuous *in vivo* protein evolution. Unlike most of previous ssDNA-mediated genome editing and evolution systems (e.g. MAGE), where ssDNAs were synthesized and diversified *in vitro*, ssDNAs of REGES are generated through DNA-dependent transcription and RNA-dependent reverse transcription *in vivo*. This means that by introducing random mutations into the DNA templates and/or RNA templates used to generate ssDNAs, we can achieve continuous *in vivo* evolution of any selected proteins. T7 RNA polymerase (T7 RNAP) guided base editing (59) and eMutaT7 (60), which were constructed by fusing cytidine deaminase to T7 RNAP, have been demonstrated to enable *in vivo* random mutagenesis (Mutator-DNA). Additionally, error-prone T7RNAP (epT7RNAP) (61) has been reported to introduce random mutations to the transcribed RNAs (Mutator-RNA) with a frequency of $\sim2 \times 10^{-3}$. Therefore, we designed and established an orthogonal transcription system for REGES (Figure 6A), where the transcription of msr and msd were driven by an epT7RNAP fused with a cytidine deaminase from *Petromyzon marinus* (PmCDA1-epT7RNAP). This enabled the introduction of *in vivo* random mutations to the DNA template and RNA molecule, both of which could be incorporated into the generated ssDNAs for protein evolution applications. For convenient detection of mutagenesis and evaluation of mutation rate, we constructed a reporter strain by integrating a model counter-selection gene *sacB* into the genome. The resultant strain could only grow on sucrose when *sacB* was inactivated (62).

We initiated mutagenesis by adding arabinose and IPTG to the recombinant strains carrying the mutator plasmids and editing plasmid pRsacB. After four cycles of editing, we spread the mutated cells onto agar plates with or without sucrose. Compared with the control strains (Figure 6B), we found that strains expressing pRsacB and Mutator-DNA showed an increased mutation frequency, with an estimated mutation rate of $2.7 \pm 0.5 \times 10^{-5}$ (Figure 6C). Although functional, the Mutator-DNA based evolution efficiency was low, probably due to the relatively low mutation rate (60) and limited mutation space of PmCDA1 (19). On the other hand, Mutator-RNA generated sucrose-resistant clones at a frequency of $1.3 \pm 0.56 \times 10^{-4}$ (Figure 6C), which was lower than that reported for epT7RNAP (61). The mutation frequency was probably negatively affected by the decreased transcriptional activity of epT7RNAP, which has been reported in other systems as well (29). We further improved the mutagenesis efficiency via combining Mutator-DNA and Mutator-RNA (i.e. PmCDA1-epT7RNAP), resulting in an editing frequency of up to $2.4 \pm 0.076 \times 10^{-3}$ (Figure 6C). The resultant mutation rate was ~1000-fold higher than that of the control strain, indicating a synergistic effect between Mutator-DNA and Mutator-RNA for introducing random mutagenesis and continuous *in vivo* evolution.

We expected the variants in the sucrose medium to possess mutations decreasing the activity of sacB. After genotyping the *sacB* gene from ~50 colonies, we found Q230, Y237 and D247 to be mutated with high frequency (Figure 6B and Supplementary Figure S6A). As shown in

Figure 6D, the mutation at Q230 caused a premature stop codon to inactivate sacB; Y237 was recently reported to affect catalytic efficiency of this enzyme via site-directed mutagenesis studies (63); D247 was located at the active site of sacB and its mutation resulted in a 75- to 3500-fold decrease of K_{cat} (64). To assess any potential off-target mutagenesis effects on the genome, we determined the rifampicin resistance frequency of *E. coli* strains with or without mutators. We found no significant difference between the control and mutator strains (Supplementary Figure S6B), indicating that genome-wide off-target mutagenesis should not be a major concern for REGES mediated continuous *in vivo* mutagenesis.

DISCUSSION

Significant progress in bacterial genome manipulation has been made in recent years. In particular, the CRISPR system has revolutionized genome editing and served as the most successful tool for targeted mutagenesis in *E. coli*. Compared with the CRISPR-mediated genome editing (including base editor and prime editor) (23,65) and ssDNA-mediated recombination (such as MAGE) (24), REGES established in the present study demonstrates advantages in high editing efficiency, easy manipulation (*in vivo* generation of ssDNAs) and flexible mutagenesis ranges. We achieved $\sim100\%$, $85 \pm 3\%$, $69 \pm 14\%$ and $25 \pm 14\%$ efficiency for single-, two-, three and four-locus editing of *E. coli* genome, respectively. More importantly, the editing region of REGES is determined by the *in vivo* generated ssDNA donor, allowing for flexible editing of any selected range and a relatively broad editing window, particularly when compared with base editor and prime editor.

While retron-mediated genome editing systems have been established in recent years, the editing efficiency remains unsatisfactory (Supplementary Table S1). For example, the ability of retron to write multiple mutations into independent loci was explored in SCRIBE (Synthetic Cellular Recorders Integrating Biological Events), but only a limited number of colonies ($\sim3 \times 10^{-7}$) were found to possess two-locus writing events (28). As the WT retron cassette from *E. coli* BL21 is not efficient enough (29–31), we optimized REGES for maximal editing performance in a modular and systematic manner. After systematic optimization, including host modification (*recJ* and *xonA* knockout) to increase the stability and availability of intracellular ssDNAs, retron expression systems (replicons, promoters and RBSs), donor design (donor length and mutation positions) and editing conditions (Table 1), we achieved $\sim100\%$ genome editing efficiency for substitution, deletion and insertion. When compared with the recently published RLR, where CspRecT was used instead of bet to bind ssDNAs (31), the expression of CspRecT in *E. coli* resulted in cytotoxicity and cell growth retardation (Supplementary Figures S7A and S7B). Moreover, by using different promoters and terminators (i.e. FAB45-L2U5H11 and FAB46-T7terminator) for the expression of retron cassettes, we were able to simultaneously mutate up to four genes with an efficiency of $25 \pm 14\%$, representing the highest number genes to be precisely substituted in a single round of genome editing in *E. coli*. As *xonA* and *recJ* encode 3′ and 5′ exonucleases,
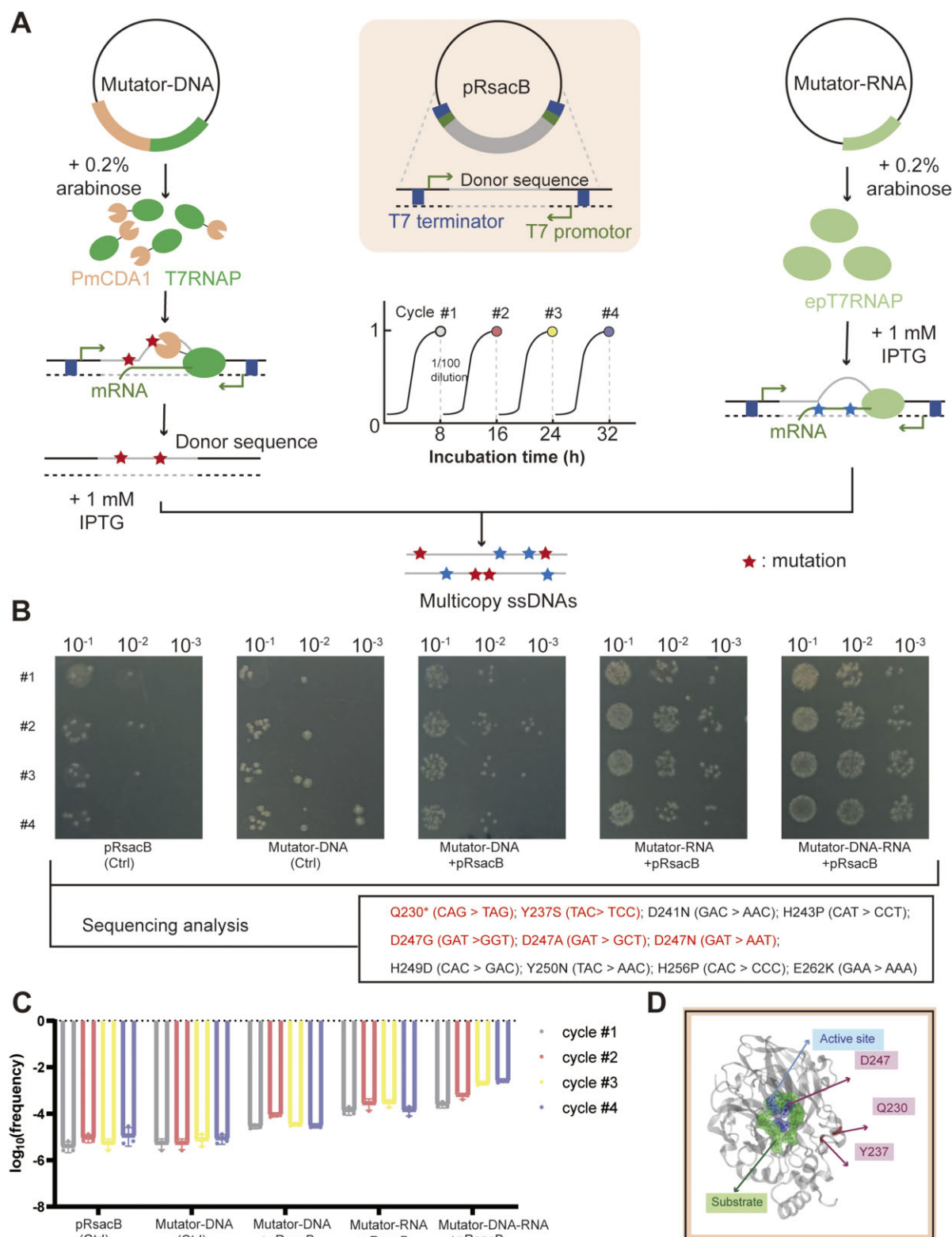
Figure 6. REGES-mediated continuous *in vivo* evolution. (**A**) Schematic illustration of the mutation system. Mutator-DNA and Mutator-RNA introduce random mutations from the DNA and RNA levels, respectively. The dilution-incubation cycle was repeated every 8 hours. T7RNAP: T7 RNA polymerase, PmCDA1: cytidine deaminase from *Petromyzon marinus*, epT7RNAP: error-prone T7 RNA polymerase. (**B**) Cell viability on sucrose medium. The same number of cells after each cycle of growth were serial-diluted and spotted onto sucrose agar plates. Mutations found in clones that survived on sucrose medium were listed and the most frequent mutations, Q230, Y237, D247, were shown in red. (**C**) Resistance frequency at each mutagenesis cycle for *E. coli* strains with different mutators. (**D**) Structure of sacB with the substrate. The active sites and mutation sites found in this study were highlighted.

which are involved in DNA repair and replication (66), disruption of these genes may lead to increased sensitivity to DNA damage in bacteria (67). Thanks to the high efficiency of REGES after optimization, we also achieved successful genome editing in WT genomic background, although with lower efficiency (Supplementary Figure S7C), indicating the potential of REGES for the construction of stable microbial cell factories for industrial applications. Nevertheless, considering the stability as well as the challenges in the assembly of repetitive DNA sequences, genome editing of more than five genes has not been attempted in the present study. In our future studies, machine learning can be employed to rationally design functional and highly nonrepetitive sequences, such as ELSA (32), enabling stable co-expression of many retron cassettes and simultaneous editing of more genomic loci.

Besides multiplex genome editing, we also demonstrate the use of REGES for continuous *in vivo* evolution of specific genes. While several methods for continuous *in vivo* evolution have been established, only a few can evolve the target protein specifically in *E. coli*. T7 RNAP guided base editing (59,60) have been recently demonstrated to enable *in vivo* random mutagenesis and one of the most powerful tools for continuous protein evolution. With the introduction of base editing mutations during T7RNAP transcription, the mutation region is started from T7 promoter and ended at T7 terminator. While gene-specific *in vivo* evolution can be achieved, the insertion of T7 promoter is expected to affect the expression level of the target gene. In other words, these systems are more suitable for the evolution of exogenous genes and have limited ability to evolve the endogenous genes. On the contrary, REGES-mediated evolution system employed epT7RNAP-PmCDA1 to drive the expression of retron cassette and random mutations introduced to the DNA and RNA templates were incorporated into the genome for protein evolution applications. In other words, no pre-editing of the target gene is needed, and its expression profile remains unchanged. Another advantage of the REGES-mediated evolution system is the flexibility in choosing the mutation region. As the insertion of T7 promoter and T7 terminator will disrupt the coding sequences, they are generally inserted in the upstream or downstream of the target gene, limiting the mutation region to the entire gene. On the other hand, the evolution region of REGES is determined by the *in vivo* generated ssDNA donor and in theory any selected region can be evolved (such as a specific domain or defined region within a coding sequences). Considering that the editing efficiency will be affected when the donor size exceeds 300 bp, REGES is more suitable for continuous *in vivo* evolution of specific structural domains or small proteins. The combination of random mutagenesis during replication (Mutator-DNA) and transcription (Mutator-RNA) stages for the generation of ssDNAs *in vivo* resulted in a mutation frequency of up to $2.4 \pm 0.076 \times 10^{-3}$, which is ~1000-fold higher than the control strain. Theoretically, we can also incorporate mutations at the reverse transcription stage, by the introduction or engineering of an error-prone reverse transcriptase (68,69). In this case, we can further improve REGES-mediated mutation rate for continuous *in vivo* protein evolution applications.

In summary, we established REGES for multiplex genome editing and continuous *in vivo* protein evolution. Systematic optimization of REGES parameters resulted in the editing of one, two, three and four genomic loci with efficiencies of ~100%, $85 \pm 3\%$, $69 \pm 14\%$ and $25 \pm 14\%$, respectively. In addition, REGES enabled the generation of pooled and barcoded variant libraries with degenerate RBS sequences to fine-tune the expression level of endogenous and exogenous genes. Finally, we employed REGES for gene-specific continuous *in vivo* evolution, by combining with T7RNAP-mediated base editing and error-prone transcription. REGES can be employed for rapid construction of microbial cell factories as well as other synthetic biology applications.

DATA AVAILABILITY

The authors declare that all data supporting the findings of this study are available in the manuscript and supporting information.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Baba,T., Ara,T., Hasegawa,M., Takai,Y., Okumura,Y., Baba,M., Datsenko,K.A., Tomita,M., Wanner,B.L. and Mori,H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, 2006.0008.
2. Zhang,H. and Wang,X. (2016) Modular co-culture engineering, a new approach for metabolic engineering. *Metab Eng.*, **37**, 114–121.
3. Weber,W. and Fussenegger,M. (2012) Emerging biomedical applications of synthetic biology. *Nat. Rev. Genet.*, **13**, 21–35.
4. Yu,D., Ellis,H.M., Lee,E.-C., Jenkins,N.A., Copeland,N.G. and Court,D.L. (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 5978–5983.
5. Muyrers,J.P.P., Zhang,Y., Testa,G. and Stewart,A.F. (1999) Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.*, **27**, 1555–1557.

6. Song,C.W. and Lee,S.Y. (2013) Rapid one-step inactivation of single or multiple genes in Escherichia coli. *Biotechnol J.*, **8**, 776–784.

7. Schweizer,H.P. (2003) Applications of the *Saccharomyces cerevisiae* Flp-FRT system in bacterial genetics. *J. Mol. Microbiol. Biotechnol.*, **5**, 67–77.

8. Urnov,F.D., Rebar,E.J., Holmes,M.C., Zhang,H.S. and Gregory,P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.*, **11**, 636–646.

9. Bogdanove,A.J. and Voytas,D.F. (2011) TAL effectors: customizable proteins for DNA targeting. *Science*, **333**, 1843–1846.

10. Jiang,W., Bikard,D., Cox,D., Zhang,F. and Marraffini,L.A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.*, **31**, 233–239.

11. Jiang,Y., Chen,B., Duan,C., Sun,B., Yang,J. and Yang,S. (2015) Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.*, **81**, 2506–2514.

12. Ran,F.A., Hsu,P.D., Wright,J., Agarwala,V., Scott,D.A. and Zhang,F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.

13. Segal,D.J., Beerli,R.R., Blancafort,P., Dreier,B., Effertz,K., Huber,A., Koksch,B., Lund,C.V., Magnenat,L., Valente,D. *et al.* (2003) Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry*, **42**, 2137–2148.

14. Lee,H., Popodi,E., Tang,H. and Foster,P.L. (2012) Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E2774–E2783.

15. Hein,R. and Tsien,R.Y. (1996) Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol.*, **6**, 178–182.

16. Abbasi,J. (2017) DNA base editing could reverse most disease-causing point mutations. *JAMA*, **318**, 2173–2173.

17. Lee,H.J., Kim,H.J. and Lee,S.J. (2020) CRISPR-Cas9-mediated pinpoint microbial genome editing aided by target-mismatched sgRNAs. *Genome Res.*, **30**, 768–775.

18. Zhang,X., Zhu,B., Chen,L., Xie,L., Yu,W., Wang,Y., Li,L., Yin,S., Yang,L., Hu,H. *et al.* (2020) Dual base editor catalyzes both cytosine and adenine base conversions in human cells. *Nat. Biotechnol.*, **38**, 856–860.

19. Komor,A.C., Kim,Y.B., Packer,M.S., Zuris,J.A. and Liu,D.R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, **533**, 420–424.

20. Jin,S., Fei,H., Zhu,Z., Luo,Y., Liu,J., Gao,S., Zhang,F., Chen,Y.H., Wang,Y. and Gao,C. (2020) Rationally designed APOBEC3B cytosine base editors with improved specificity. *Mol. Cell*, **79**, 728–740.

21. Gaudelli,N.M., Komor,A.C., Rees,H.A., Packer,M.S., Badran,A.H., Bryson,D.I. and Liu,D.R. (2017) Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, **551**, 464–471.

22. Anzalone,A.V., Randolph,P.B., Davis,J.R., Sousa,A.A., Koblan,L.W., Levy,J.M., Chen,P.J., Wilson,C., Newby,G.A., Raguram,A. *et al.* (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, **576**, 149–157.

23. Tong,Y., Jorgensen,T.S., Whitford,C.M., Weber,T. and Lee,S.Y. (2021) A versatile genetic engineering toolkit for *E. coli* based on CRISPR-prime editing. *Nat. Commun.*, **12**, 5206.

24. Wang,H.H., Isaacs,F.J., Carr,P.A., Sun,Z.Z., Xu,G., Forest,C.R. and Church,G.M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, **460**, 894–898.

25. Lim,D. and Maas,W.K. (1989) Reverse-transcriptase dependent synthesis of a covalently linked, branched DNA RNA compound in *Escherichia coli* B. *Cell*, **56**, 891–904.

26. Millman,A., Bernheim,A., Stokar-Avihail,A., Fedorenko,T., Voichek,M., Leavitt,A., Oppenheimer-Shaanan,Y. and Sorek,R. (2020) Bacterial retrons function in anti-phage defense. *Cell*, **183**, 1551–1561.

27. Wang,Y., Guan,Z., Wang,C., Nie,Y., Chen,Y., Qian,Z., Cui,Y., Xu,H., Wang,Q., Zhao,F. *et al.* (2022) Cryo-EM structures of *Escherichia coli* Ec86 retron complexes reveal architecture and defence mechanism. *Nat. Microbiol.*, **7**, 1480–1489.

28. Farzadfard,F. and Lu,T.K. (2014) Genomically encoded analog memory with precise *in vivo* DNA writing in living cell populations. *Science*, **346**, 825–833.

29. Simon,A.J., Morrow,B.R. and Ellington,A.D. (2018) Retroelement-based genome editing and evolution. *ACS Synth Biol*, **7**, 2600–2611.

30. Farzadfard,F., Gharaei,N., Citorik,R.J. and Lu,T.K. (2021) Efficient retroelement-mediated DNA writing in bacteria. *Cell Syst.*, **12**, 860–873.

31. Schubert,M.G., Goodman,D.B., Wannier,T.M., Kaur,D., Farzadfard,F., Lu,T.K., Shipman,S.L. and Church,G.M. (2021) High-throughput functional variant screens via *in vivo* production of single-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2018181118.

32. Reis,A.C., Halper,S.M., Vezeau,G.E., Cetnar,D.P., Hossain,A., Clauer,P.R. and Salis,H.M. (2019) Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays. *Nat. Biotechnol.*, **37**, 1294–1301.

33. Morrison,D.A. (1977) Transformation in *Escherichia coli*: cryogenic preservation of competent cells. *J. Bacteriol.*, **132**, 349–351.

34. Xiao,F., Wang,H., Shi,Z., Huang,Q., Huang,L., Lian,J., Cai,J. and Xu,Z. (2020) Multi-level metabolic engineering of *Pseudomonas mutabilis* ATCC31014 for efficient production of biotin. *Metab Eng.*, **61**, 406–415.

35. Ren,Q., Henes,B., Fairhead,M. and Thöny-Meyer,L. (2013) High level production of tyrosinase in recombinant *Escherichia coli*. *BMC Biotechnol.*, **13**, 18–28.

36. HamediRad,M., Chao,R., Weisberg,S., Lian,J., Sinha,S. and Zhao,H. (2019) Towards a fully automated algorithm driven platform for biosystems design. *Nat. Commun.*, **10**, 5150.

37. Datsenko,K.A. and Wanner,B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 6640–6645.

38. Ronda,C., Pedersen,L.E., Sommer,M.O. and Nielsen,A.T. (2016) CRMAGE: CRISPR optimized MAGE recombineering. *Sci. Rep.*, **6**, 19452.

39. Egbert,R.G., Rishi,H.S., Adler,B.A., McCormick,D.M., Toro,E., Gill,R.T. and Arkin,A.P. (2019) A versatile platform strain for high-fidelity multiplex genome editing. *Nucleic Acids Res.*, **47**, 3244–3256.

40. Dalia,T.N., Yoon,S.H., Galli,E., Barre,F.X., Waters,C.M. and Dalia,A.B. (2017) Enhancing multiplex genome editing by natural transformation (MuGENT) via inactivation of ssDNA exonucleases. *Nucleic Acids Res.*, **45**, 7527–7537.

41. Gallagher,R.R., Li,Z., Lewis,A.O. and Isaacs,F.J. (2014) Rapid editing and evolution of bacterial genomes using libraries of synthetic DNA. *Nat. Protoc.*, **9**, 2301–2316.

42. Mosberg,J.A., Gregg,C.J., Lajoie,M.J., Wang,H.H. and Church,G.M. (2012) Improving lambda red genome engineering in *Escherichia coli* via rational removal of endogenous nucleases. *Plos One*, **7**, e44638.

43. Sawitzke,J.A., Costantino,N., Li,X.T., Thomason,L.C., Bubunenko,M., Court,C. and Court,D.L. (2011) Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. *J. Mol. Biol.*, **407**, 45–59.

44. Lampson,B.C., Inouye,M. and Inouye,S. (2005) Retrons, msDNA, and the bacterial genome. *Cytogenet. Genome Res.*, **110**, 491–499.

45. Simon,A.J., Ellington,A.D. and Finkelstein,I.J. (2019) Retrons and their applications in genome engineering. *Nucleic Acids Res.*, **47**, 11007–11019.

46. Palka,C., Fishman,C.B., Bhattarai-Kline,S., Myers,S.A. and Shipman,S.L. (2022) Retron reverse transcriptase termination and phage defense are dependent on host RNase H1. *Nucleic Acids Res.*, **50**, 3490–3504.

47. Li,X.T., Costantino,N., Lu,L.Y., Liu,D.P., Watt,R.M., Cheah,K.S.E., Court,D.L. and Huang,J.D. (2003) Identification of factors influencing strand bias in oligonucleotide-mediated recombination in *Escherichia coli*. *Nucleic Acids Res.*, **31**, 6674–6687.

48. Mutalik,V.K., Guimaraes,J.C., Cambray,G., Lam,C., Christoffersen,M.J., Mai,Q.A., Tran,A.B., Paull,M., Keasling,J.D., Arkin,A.P. *et al.* (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods*, **10**, 354–360.

49. Kosuri,S., Goodman,D.B., Cambray,G., Mutalik,V.K., Gao,Y., Arkin,A.P., Endy,D. and Church,G.M. (2013) Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 14024–14029.

50. Wannier,T.M., Nyerges,A., Kuchwara,H.M., Czikkely,M., Balogh,D., Filsinger,G.T., Borders,N.C., Gregg,C.J., Lajoie,M.J., Rios,X. *et al.* (2020) Improved bacterial recombineering by parallelized protein discovery. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 13689–13698.

51. Moore,S.J., Lai,H.E., Kelwick,R.J., Chee,S.M., Bell,D.J., Polizzi,K.M. and Freemont,P.S. (2016) Ecoflex: a multifunctional moclo kit for *E. coli* synthetic biology. *ACS Synth. Biol.*, **5**, 1059–1069.

52. Tang,Q., Lou,C. and Liu,S.J. (2017) Construction of an easy-to-use CRISPR-Cas9 system by patching a newly designed EXIT circuit. *J. Biol. Eng.*, **11**, 32.

53. McKenna,R., Lombana,T.N., Yamada,M., Mukhyala,K. and Veeravalli,K. (2019) Engineered sigma factors increase full-length antibody expression in *Escherichia coli*. *Metab Eng.*, **52**, 315–323.

54. Chen,M., Liang,H., Han,C., Zhou,P., Xing,Z., Chen,Q., Liu,Y., Xie,G.A. and Xie,R. (2022) Engineering of global transcription factor FruR to redirect the carbon flow in *Escherichia coli* for enhancing L-phenylalanine biosynthesis. *Microb Cell Fact*, **21**, 222.

55. Zhang,H., Chong,H., Ching,C.B. and Jiang,R. (2012) Random mutagenesis of global transcription factor cAMP receptor protein for improved osmotolerance. *Biotechnol. Bioeng.*, **109**, 1165–1172.

56. Martinez-Antonio,A., Janga,S.C. and Thieffry,D. (2008) Functional organisation of *Escherichia coli* transcriptional regulatory network. *J. Mol. Biol.*, **381**, 238–247.

57. Martinez-Antonio,A. and Collado-Vides,J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol.*, **6**, 482–489.

58. Bali,A.P., Lennox-Hvenekilde,D., Myling-Petersen,N., Buerger,J., Salomonsen,B., Gronenberg,L.S., Sommer,M.O.A. and Genee,H.J. (2020) Improved biotin, thiamine, and lipoic acid biosynthesis by engineering the global regulator IscR. *Metab Eng.*, **60**, 97–109.

59. Cravens,A., Jamil,O.K., Kong,D., Sockolosky,J.T. and Smolke,C.D. (2021) Polymerase-guided base editing enables *in vivo* mutagenesis and rapid protein engineering. *Nat. Commun.*, **12**, 1579.

60. Park,H. and Kim,S. (2021) Gene-specific mutagenesis enables rapid continuous evolution of enzymes *in vivo*. *Nucleic Acids Res.*, **49**, e32.

61. Brakmann,S. and Grzeszik,S. (2001) An error-prone T7 RNA polymerase mutant generated by directed evolution. *Chem. Bio. Chem.*, **2**, 212–219.

62. Gay,P., Coq,D.L., Steinmetz,M., Berkelman,T. and Kado,C.I. (1985) Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J. Bacteriol.*, **164**, 918–921.

63. Raga-Carbajal,E., Diaz-Vilchis,A., Rojas-Trejo,S.P., Rudino-Pinera,E. and Olvera,C. (2021) The molecular basis of the nonprocessive elongation mechanism in levansucrases. *J. Biol. Chem.*, **296**, 10078.

64. Meng,G. and Futterer,K. (2003) Structural framework of fructosyl transfer in *Bacillus subtilis* levansucrase. *Nat. Struct. Biol.*, **10**, 935–941.

65. Banno,S., Nishida,K., Arazoe,T., Mitsunobu,H. and Kondo,A. (2018) Deaminase-mediated multiplex genome editing in *Escherichia coli*. *Nat. Microbiol.*, **3**, 423–429.

66. Cao,Z., Mueller,C.W. and Julin,D.A. (2010) Analysis of the *recJ* gene and protein from *Deinococcus radiodurans*. *DNA Repair*, **9**, 66–75.

67. Prada Medina,C.A., Aristizabal Tessmer,E.T., Quintero Ruiz,N., Serment-Guerrero,J. and Luis Fuentes,J. (2016) Survival and SOS response induction in ultraviolet B irradiated *Escherichia coli* cells with defective repair mechanisms. *Int. J. Radiat Biol.*, **92**, 321–328.

68. Halvas,E.K., Svarovskaia,E.S. and Pathak,V.K. (2000) Role of murine leukemia virus reverse transcriptase deoxyribonucleoside triphosphate-binding site in retroviral replication and *in vivo* fidelity. *J. Virol.*, **74**, 10349–10358.

69. Roberts,J.D., Bebenek,K. and Kunkel,T.A. (1988) The accuracy of reverse-transcriptase from HIV-1. *Science*, **242**, 1171–1173.