CRISPRi-FGP: a web-based genome-scale CRISPRi sgRNA design tool for prokaryotes

Tongjun Xiang[1, †], Huibao Feng[1, †], Xin-hui Xing[1] and 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.

[*] Chong Zhang, Email: chongzhang@tsinghua.edu.cn

[†] Joint Authors

ABSTRACT

CRISPRi screening is emerging as a powerful strategy for functional genomic study. However, off-target effects that result from mismatch tolerance between sgRNAs and targets are regarded as the major problem in CRISPRi applications. To avoid off-target binding, here we propose a web-based tool named CRISPRi-Functional Genomics for Prokaryotes (CRISPRi-FGP) for genome-scale sgRNA library design for prokaryotic organisms (http://www.thu-big.net/sgRNA_design/). For this tool, we established a powerful quality control principle that identifies off-target hits with high accuracy. Moreover, we established a built-in database containing 1,300 common prokaryotes as a design source. In addition, CRISPRi-FGP also supports sgRNA design for new organisms using uploaded reference files. Finally, we found that compared to existing sgRNA design tools, CRISPRi-FGP could identify off-target binding hits with higher precision.

INTRODUCTION

With the boosting accumulation of bacterial genomes, genotype-phenotype mapping is drawing more attention as a powerful strategy for functional genomic study, which provides insights into the microbiology and engineering of microorganisms (1–3). To investigate the function of a specific gene, the straightforward approach is to perturb the gene and assay the consequent phenotypic changes. Recently, the CRISPRi screening method was established for associating transcriptional perturbations with corresponding phenotypic changes in a massively parallel manner (4). A typical CRISPRi screening starts with a pooled single-guide RNA (sgRNA) library, in which each sgRNA is designed to

guide a nuclease-dead Cas9 (dCas9) protein to a specific genetic locus flanked by a 3'-NGG protospacer adjacent motif (PAM) via 20-bp Watson-Crick base pairing (5), resulting in transcriptional inhibition of the target gene. After transformation of the library, cells with different sgRNAs perform differently in subsequent high-throughput screening, such as fluorescence-activated cell sorting (6, 7) or fitness screening (8, 9), and can be further identified through next generation sequencing (NGS). However, false positive results sometimes occur owing to unexpected binding. This phenomenon, known as "off-targeting", is because the binding process can tolerate several mismatches (10). Hence, a highly accurate and robust sgRNA library design tool is needed to avoid off-target binding in CRISPRi screening.

To date, a variety of tools have been established for sgRNA design (11–16). However, these applications are all designed for the CRISPR/Cas system, rather than the CRISPRi system. Note that the occurrence of cleavage must undergo conformational gating after adequate base pairs are formed between the sgRNA and the target DNA (17). Therefore, off-target binding is more likely to happen compared to off-target cleavage, suggesting that existing tools might have problems in accurately identifying off-target binding hits. On the other hand, quality control criteria used by existing sgRNA design tools are developed using experience from eukaryotes. Considering the fundamental differences between prokaryotic and eukaryotic genomes, such as the accessibility differences of the genome in a less organized chromosome structure (18, 19) and the variety in intracellular conditions, the applicability of these tools on prokaryotes might be problematic.

To address the above issues, we have developed CRISPRi-Functional Genomics for Prokaryotes (CRISPRi-FGP), a web-based tool that allows users to easily and rapidly design genome-scale sgRNA libraries for CRISPRi screening in prokaryotes. CRISPRi-FGP employs a rigorous off-target identification algorithm, which is specifically optimized for prokaryotes, to predict the binding specificity of each potential target site in the genome (11, 20–22). This tool has a built-in database of 1,300 common microorganisms and accepts uploaded reference files for new organisms. In addition, we showed that several sgRNAs designed by existing tools bound to off-target hits identified by CRISPRi-FGP, suggesting that CRISPRi-FGP has superior performance in designing sgRNA libraries for CRISPRi screening. Therefore, through rigorous design and solid experiments, we have demonstrated that CRISPRi-FGP can serve as a promising tool to support functional genomic studies in prokaryotes.

MATERIALS AND METHODS

Reference files and data pre-processing

To design the sgRNA library of a specific microorganism, two reference files are required: the sequence and annotation files. The sequence file is a single contig genome file (FASTA format). The annotation file contains the features of all genes of the organism, including genomic name, coding type, location, strand, phase, and functional attributes, etc. CRISPRi-FGP accepts two types of annotation files: (i) a generic feature format file (GFF) and (ii) a protein table file (PTT) or RNA table file (RNT). The files can be obtained from a variety of publicly available databases (such as NCBI, EMBL, and DDBJ, etc.), or generated by a genome annotation pipeline (like PGAP (23)). Once the two files are uploaded, a coding list is constructed for correlating gene features with their DNA sequences, which contains information on all genes with unique identifiers.

sgRNA library design workflow

The design procedure of CRISPRi-FGP contains three major steps, and the workflow of functional blocks is shown in Figure 2. The first step is gene clustering (23,24). In this step, genes with high similarity are considered as the same function (24, 25) and are grouped into a cluster. To this end, CRISPRi-FGP employs BLASTN (26) with default parameters for sequence alignment. It uses a strict threshold of: < 0.001 evalue, > 95% identity, > 95% hit coverage, and > 95% query coverage. In the end, each cluster might contain one or more genes. The identifier of each cluster (clusterID) is defined as the smallest identifier of the genes it contains. For example, all clusters of multi-copy coding genes of *E. coli* MG1655 are listed in Supplementary Table S1. After grouping all clusters, each nucleobase belonging to only one cluster is labeled with the clusterID. The second step is the search for candidate sgRNAs by regular expressions. Two types of guides, N20NGG and N20NAG, corresponding to canonical and non-canonical PAM sequences (27), are identified from both chains of the genome. The candidate sgRNAs are categorized into three groups: (i) sgRNAs with an NGG PAM targeting the non-template strand; (ii) sgRNAs with an NGG PAM targeting the template strand; and (iii) sgRNAs with an NAG PAM. Three FASTA files (FASTA1, FASTA2, and FASTA3) are generated corresponding to the group (i), (ii), and (iii), respectively. In each file, the PAM-proximal 12-bp regions (seed regions) of all candidate sgRNAs are saved for their essential roles in binding (9). A toy example is shown in

Supplementary Figure S1. The third step is the quality control of sgRNAs. Specifically, three alignments are performed according to the user defined target strand (template or non-template) using Seqmap (21). For example, if the non-template strand is selected, FASTA1 would be aligned with FASTA1, FASTA2, and FASTA3. In each alignment, sequence pairs with two or fewer mismatches are picked for subsequent scoring evaluations given the fact that two or more mismatches in the seed region would significantly weaken the binding affinity (11, 28). The metrics of the scoring evaluation are defined according to the distance of mismatches to the PAM, as mismatches are generally better tolerated at the 5′ end of the sgRNA than at the 3′ end (14). Regions from the 3' end to the 5' end are as follows: region I (7 nt), II (5 nt), and III (the rest of the sgRNA) (10, 22). If the PAM is NGG, the mismatch penalties of region I, II, and III are 8, 4.5, and 2.5, respectively, and the penalties for the NAG PAM are 10, 7, and 3, respectively (see Supplementary Figure S2). One off-target site is considered significant when the penalty score is less than the user-defined threshold (4). We recommend a threshold of 18–21 for genome-scale library design. Finally, sgRNAs within the upper and lower limits of GC-content defined by the user are reserved.

Design of negative control sgRNAs

Negative control sgRNAs are sgRNAs with no significant target across the genome. These can serve as negative controls in screening experiments to evaluate the impact of experimental procedures, such as the effects of transforming plasmids on the cell phenotype. CRISPRi-FGP designs negative control sgRNAs by generating random N20 sequences. The penalty scores for both NGG and NAG PAMs are 25, and the GC content limits are the same as the sgRNA library. The presence of five or more consecutive identical bases is avoided. The number of total negative control sgRNAs is defined by the user.

Visualization of results

The final designed sgRNA library is arranged in a table, in which sgRNAs are classified with their targeted genes. For each gene, sgRNAs are ranked by their distance to the start codon. This is because sgRNAs targeting closer to transcription start sites generally have higher knock-down activities (20). This table is temporarily saved on the server and can be easily downloaded from the website.

The sgRNA library can also be inspected on a graphical interactive page, which is built with the D3GB genome browser (29). The whole genome and all sgRNAs are displayed on a zoomable page so that users can zoom in to inspect the region of interest or search for a specific locus. Genes and designed sgRNAs are displayed at their appropriate locations in the genome.

DNA manipulation and reagents

Plasmid extraction and DNA purification were performed using kits from Vazyme. PCR reactions were carried out using KOD One™ PCR Master Mix from TOYBO Life Science. All PCR primers were ordered from Azenta (Supplementary Table S2). Plasmids were constructed using a Gibson Assembly mix containing 10 U/μL T5 exonuclease, 2 U/μL Phusion High-Fidelity DNA polymerase, and 40 U/μL Taq DNA ligase, which were all from New England Biolabs. Antibiotic concentrations for kanamycin and ampicillin were 50 and 100 mg/L, respectively. All strains and plasmids used in this work are listed in Supplementary Table S3.

Characterization of potential off-target sgRNAs from existing tools

The experiment was conducted in *Escherichia coli* MCm, which was constructed in previous work by integrating a chloramphenicol expression cassette cloned from pKM154 (Addgene plasmid #13036) into the *smf* locus of wild-type *E. coli* K12 MG1655 (4). We first designed sgRNA libraries for *Corynebacterium glutamicum* ATCC 13032 using CHOPCHOP (30) and Synthego (https://design.synthego.com/). The sgRNAs designed by these tools that were highly ranked and were considered to have potential off-target hits by CRISPRi-FGP were chosen as candidates. The candidate sgRNAs were next compared with the genome of *E. coli* MCm using CRISPRi-FGP to ensure no possible hits. During both runs of CRISPRi-FGP, the off-target penalty score was set to 18.

Plasmid construction

The reporting system was constructed in two plasmids, one that carried dCas9, and another that contained sgRNA and *mcherry*. The dCas9-expressing plasmid pdCas9-J23111 was constructed in previous work (4). The plasmid for sgRNA and *mcherry* expression was modified from pN20test-114mCherry-r0-m1 (11). In this plasmid, the expression of sgRNA was controlled by a J23119 promoter. The N20 sequence was inserted upstream of the −35 region of the J23114 promoter, which

controlled the expression of *mcherry*. Finally, 21 different pN20test-114mCherry plasmids were constructed (Supplementary Figure S3). Plasmids were constructed using the Gibson Assembly method for PCR products (primers are listed in Supplementary Table S2; the template was the original pN20test-114mCherry-r0-m1 plasmid). All constructed plasmids were confirmed by Sanger sequencing.

Cell growth conditions and strain construction

Cells were grown in a 48-well deep-well plate with 1 mL of LB medium in each well at 37°C and 220 rpm. All 21 experimental groups were grown with kanamycin and ampicillin, and an empty control group for measuring autofluorescence (MCm-PdCas9-J23111) was grown with kanamycin.

Strains were first grown overnight, and a two-round 5‰ dilution and 10 h growth were carried out to ensure the stability of mCherry and reduce cell agglomeration. Cultures were next diluted in fresh LB medium with antibiotics to OD_{600} = 0.02 and then grown for 4 h to the logarithmic phase. After cultivation, 5 μL of culture medium was diluted into 200 μL of phosphate-buffered saline (PBS) for cytometry assay. Three independent biological replicates were prepared for each strain.

Flow cytometry assay and data processing

The flow cytometry assay was performed on an LSRFortessa flow cytometer (BD Biosciences) using a 96-well plate. Gating based on the FSC area and SSC area was carried out to exclude non-cell particles. The autofluorescence was measured using the MCm/PdCas9-J23111 strain and was subtracted in the following analysis. The mCherry fluorescence intensity for each group was calculated from the resulting positive area in the region corresponding to the PE-Texas Red area.

For each cytometry assay, the log10-transformed fluorescence intensity distribution was matched with a two-component Gaussian mixture model (30) with parameters $(\lambda, \mu_1, \mu_2, \sigma_1, \sigma_2)$ through the expectation–maximization algorithm, where λ and $1-\lambda$ refer to the mixing coefficients of two Gaussian components; μ_1 and σ_1 are the mean and standard deviation of the first Gaussian component, respectively; and μ_2 and σ_2 are the mean and standard deviation of the second Gaussian component, respectively (Equation 1).

$$f(x) = \lambda N(\mu_1, \sigma_1^2) + (1-\lambda) N(\mu_2, \sigma_2^2) \quad (1)$$

The mean expression strength was calculated with Equation 2.

$$Mean = Mean_1 \times Mean_2 = exp(m_1 + V_1/2) \times exp(m_2 + V_2/2) \quad (2)$$

where $m_1 = \lambda_i \mu_{1i}\, log(10)$, $V_1 = (\lambda_i \sigma_{1i}\, log(10))^2$, $m_2 = (1 - \lambda_i)\mu_{2i}\, log(10)$ and $V_2 = ((1 - \lambda_i)\sigma_{2i}\, log(10))^2$.

The repression rate (rr) of each group was calculated using Equation 3.

$$rr = \frac{Mean_{Ri} - Mean_{Ri-PC}}{Mean_{Ri-NC} - Mean_{Ri-PC}} \quad (3)$$

where *i* = 1–4.

RESULTS

Implementation of the CRISPRi-FGP web tool

We have developed CRISPRi-FGP, a web-based tool for genome-scale sgRNA library design for the CRISPRi system in prokaryotes. It is available at http://www.thu-big.net/sgRNA_design/. It employs a powerful algorithm for identifying off-target hits and outputs the results on an interactive page (Figure 1).

Input and configuration

CRISPRi-FGP has a built-in database containing over 1,300 prokaryotes and accepts uploaded reference files (see **Materials and Methods**). Five parameters are required for a design, which are described as follows: (i) design target, which can be chosen either from the coding sequence (CDS) or RNA coding genes (RNA); (ii) off-target threshold, which is the penalty score used in the sgRNA quality control (see **Materials and Methods**); (iii) GC limits, which are the upper and lower bounds of the GC content of each sgRNA; (iv) spacer length, which is the length of the designed sgRNA (≥ 12 bp); and (v) target strand, which can be selected either from the template or non-template, and refers to the targeted strand of the sgRNAs. Note that in the case of common CRISPRi systems, the non-template is preferred for this parameter to ensure effective gene silencing (27). We provide the other choice for applications such as genome editing, base editing (31, 32), or dynamic imaging (33), in which there is no strand preference.

Process and output

After configuration, a request is sent to the server to start the workflow (Figure 2). A progress bar is displayed on the page to show the running status. For a typical genome-scale library design, we expect the total process time to be 5 to 20 min, according to different genome sizes. The resulting sgRNA library is displayed in a table, in which sgRNAs are classified by their targeted genes and labeled with their start positions (see **Materials and Methods**). An interactive page is also generated by the D3GB genome browser (29) to show the sgRNA library, in which all genes and sgRNAs are mapped in their natural positions in the genome.

CRISPRi-FGP is superior to existing tools in avoiding off-target binding

We compared CRISPRi-FGP with two widely used sgRNA design tools, CHOPCHOP and Synthego, both of which can design sgRNAs for prokaryotes. We discovered that some high-ranked sgRNAs designed by these tools had potential off-target hits by CRISPRi-FGP (Table 1). We constructed a fluorescence reporter system based on CRISPRi to test the binding affinity of these sgRNAs with their off-target hits identified by CRISPRi-FGP. In this system, the N20 sequence (corresponding to one potential off-target hit) was inserted upstream of the promoter of *mcherry*. Once the dCas9-sgRNA complex bound to the N20 sequence, the expression of *mcherry* was repressed, resulting in a different fluorescence intensity. For each sgRNA (EXP) and its potential off-target hit, two more sgRNAs were designed as positive (PC) and negative control (NC). The PC was the same as the N20 sequence and was used to represent full repression; meanwhile, the NC had no binding affinity with the N20 sequence and was used to represent the fluorescence intensity without repression (Figure 3A). As sgRNAs bind with the endogenous genome, the fluorescence intensity may be influenced; therefore, the design and experiment were conducted on different organisms. Specifically, all sgRNAs were first designed in *Corynebacterium glutamicum* ATCC 13032, from which the off-target hits were identified (see **Materials and Methods**). Candidate sgRNAs were next compared with the genome of *Escherichia coli* MCm to exclude sgRNAs with a possible hit in the *E. coli* genome. Finally, four sgRNAs (R1–R4) with their off-target hits were selected (Table 1). To ensure these sgRNAs alone had no impact on fluorescence intensity, we further designed a blank control N20 sequence (blank) with no binding affinity to any sgRNA (Table 2). These groups showed similar mCherry fluorescence

intensities in flow cytometry assays (Figure 3B) and no apparent differences were found, demonstrating that the repression rates were primarily contributed by off-target effects. The results showed that all four sgRNAs designed by CHOPCHOP and Synthego exhibited binding affinities with potential off-target hits predicted by CRISPRi-FGP, leading to reduced fluorescence intensity compared to the negative control groups (Figure 3C, $P < 1.36 \times 10^{-5}$ for all groups, two-tailed *t* test; Supplementary Figure S4). The average repression rate was 69.47% (see **Materials and Methods**), ranging from 51.12% (R1) to 90.70% (R4).

DISCUSSION

In this study, we developed CRISPRi-FGP, a web-based tool that was demonstrated to be efficient for sgRNA library design for CRISPRi screening. This tool accepts standard annotation files, which can be acquired from publicly available databases. In the case of organisms that lack annotations, such information can be generated by automatic genome annotation programs, such as the Prokaryotic Genome Annotation Pipeline (PGAP) (23). For example, based on PGAP, we designed an sgRNA library for CRISPRi screening for *Corynebacterium glutamicum* CGMCC 1.15647, which successfully identified genes conducive to nanobody VHH secretion (34). This universality of CRISPRi-FGP dramatically extends its applicability in functional genomic studies for a variety of prokaryotes.

One potential limitation of CRISPRi-FGP is in quality control. While it is commonly accepted that the PAM-proximal region, to a large extent, determines the binding affinity of an sgRNA, the scoring principle is completely based on experience and lacks a rational explanation. Furthermore, the contributions of different mismatched nucleobase pairs are considered equal. However, it has been shown that the stability of different mismatches varies (35, 36). Some mismatches have stability close to Watson-Crick pairs, including rU/dG, rG/dT, and rG/dG (37). This variety of stabilities was ignored in the current underlying system principles. To address this issue, the thermodynamics of the entire binding process should be considered. For example, we proposed a calculation model that predicts off-target binding affinities based on a Markov chain using nucleic acid thermodynamic parameters reported in the literature (11). However, the parameters of mismatches are still insufficient. Therefore, a comprehensive measurement of nearest-neighbor parameters might be required to help predict the binding affinities of sgRNAs.

DATA AVAILABILITY

Raw data of the flow cytometry assays and plasmid maps related to this work can be accessed via Github (https://github.com/falconxtj/CRISPRi-FGP).

SUPPLEMENTARY DATA

ACKNOWLEDGEMENT

We thank Dr. T. Wang (Tsinghua University) for building the original demo of the sgRNA design tool. We also thank members from the C. Zhang and H. Xing laboratories for critical discussions of this work.

FUNDING

This work was supported by the National Key R&D Program of China [2018YFA0901500 to X.X.]; the National Natural Science Foundation of China [21938004 to X.X.]; the National Natural Science Foundation of China [U2032210 to C.Z.]; and the Open Project Funding of State Key Laboratory of Biocatalysis and Enzyme Engineering of Hubei University [SKLBEE2020001].

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

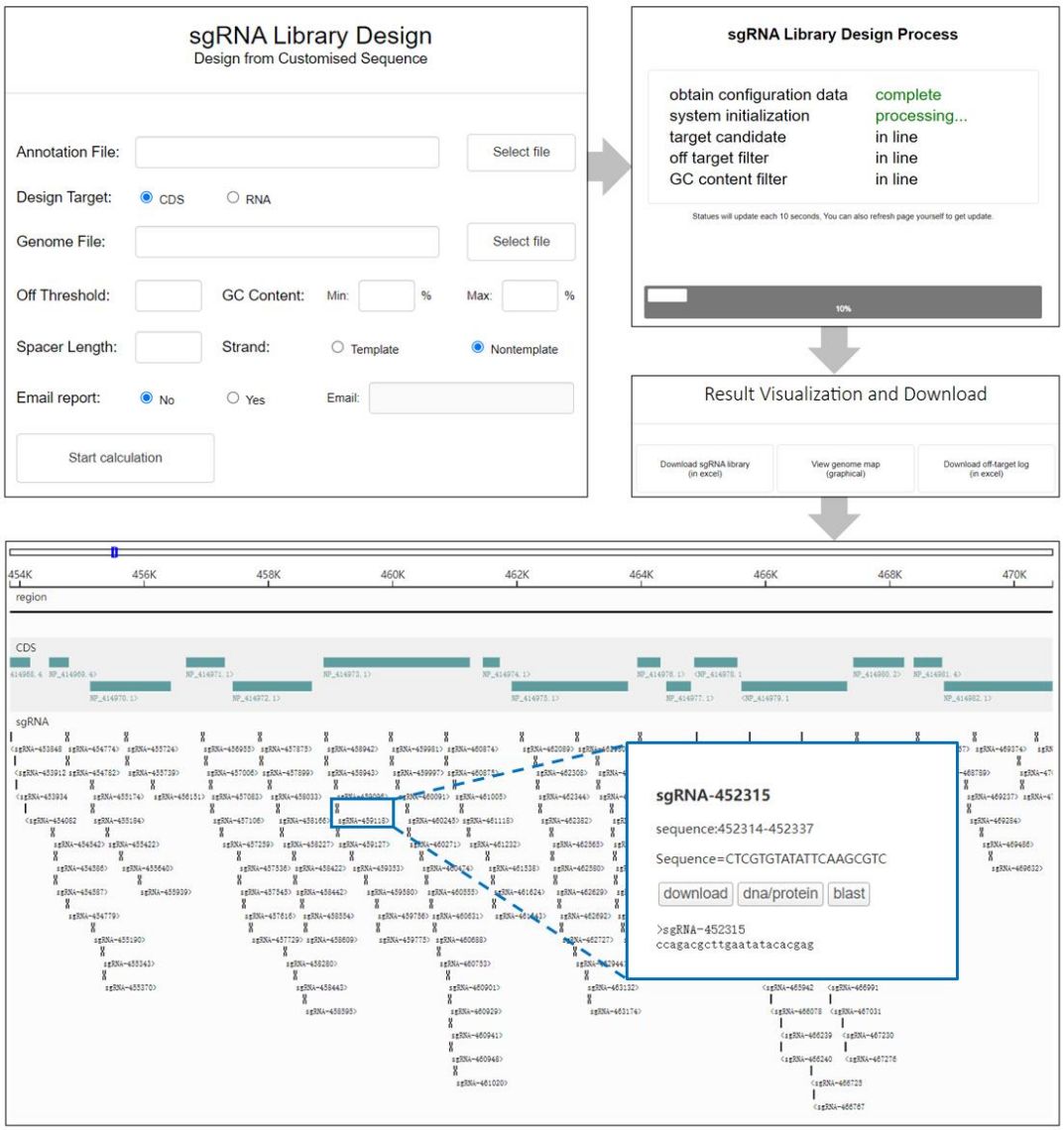
1. van Opijnen,T., Bodi,K.L. and Camilli,A. (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods*, **6**, 767–772.
2. Garst,A.D., Bassalo,M.C., Pines,G., Lynch,S.A., Halweg-Edwards,A.L., Liu,R., Liang,L., Wang,Z., Zeitoun,R., Alexander,W.G., *et al.* (2017) Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat Biotechnol*, **35**, 48–55.
3. Freed,E.F., Winkler,J.D., Weiss,S.J., Garst,A.D., Mutalik,V.K., Arkin,A.P., Knight,R. and Gill,R.T. (2015) Genome-Wide Tuning of Protein Expression Levels to Rapidly Engineer Microbial Traits. *ACS Synth Biol*, **4**, 1244–1253.
4. Wang,T., Guan,C., Guo,J., Liu,B., Wu,Y., Xie,Z., Zhang,C. and Xing,X.-H. (2018) Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. *Nat Commun*, **9**, 2475.
5. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A

programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (1979)*, **337**, 816–821.
6. Hawkins,J.S., Silvis,M.R., Koo,B.M., Peters,J.M., Osadnik,H., Jost,M., Hearne,C.C., Weissman,J.S., Todor,H. and Gross,C.A. (2020) Mismatch-CRISPRi Reveals the Co-varying Expression-Fitness Relationships of Essential Genes in Escherichia coli and Bacillus subtilis. *Cell Syst*, **11**, 523-535.e9.
7. Lian,J., Schultz,C., Cao,M., HamediRad,M. and Zhao,H. (2019) Multi-functional genome-wide CRISPR system for high throughput genotype–phenotype mapping. *Nature Communications 2019 10:1*, **10**, 1–10.
8. Chen,P., Michel,A.H. and Zhang,J. (2022) Transposon insertional mutagenesis of diverse yeast strains suggests coordinated gene essentiality polymorphisms. *Nature Communications 2022 13:1*, **13**, 1–15.
9. de Bakker,V., Liu,X., Bravo,A.M. and Veening,J.W. (2022) CRISPRi-seq for genome-wide fitness quantification in bacteria. *Nature Protocols 2021 17:2*, **17**, 252–281.
10. Hsu,P.D., Scott,D.A., Weinstein,J.A., Ran,F.A., Konermann,S., Agarwala,V., Li,Y., Fine,E.J., Wu,X., Shalem,O., *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*, **31**, 827–832.
11. Feng,H., Guo,J., Wang,T., Zhang,C. and Xing,X. (2021) Guide-target mismatch effects on dCas9–sgRNA binding activity in living bacterial cells. *Nucleic Acids Res*, **49**, 1263–1277.
12. Shalem,O., Sanjana,N.E. and Zhang,F. (2015) High-throughput functional genomics using CRISPR–Cas9. *Nat Rev Genet*, **16**, 299–311.
13. Heigwer,F., Kerr,G. and Boutros,M. (2014) E-CRISP: fast CRISPR target site identification. *Nat Methods*, **11**, 122–123.
14. Sander,J.D. and Joung,J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*, **32**, 347–355.
15. Montague,T.G., Cruz,J.M., Gagnon,J.A., Church,G.M. and Valen,E. (2014) CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res*, **42**, W401–W407.
16. Concordet,J.-P. and Haeussler,M. (2018) CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res*, **46**, W242–W245.
17. Sternberg,S.H., LaFrance,B., Kaplan,M. and Doudna,J.A. (2015) Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature*, **527**, 110–113.
18. Struhl,K. (1999) Fundamentally Different Logic of Gene Regulation in Eukaryotes and Prokaryotes. *Cell*, **98**, 1–4.
19. Kuzminov,A. (2014) The Precarious Prokaryotic Chromosome. *J Bacteriol*, **196**, 1793–1806.
20. Guo,J., Wang,T., Guan,C., Liu,B., Luo,C., Xie,Z., Zhang,C. and Xing,X.-H. (2018) Improved sgRNA design in bacteria via genome-wide activity profiling. *Nucleic Acids Res*, **46**, 7052–7069.
21. Jiang,H. and Wong,W.H. (2008) SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics*, **24**, 2395–2396.
22. Gilbert,L.A., Horlbeck,M.A., Adamson,B., Villalta,J.E., Chen,Y., Whitehead,E.H., Guimaraes,C., Panning,B., Ploegh,H.L., Bassik,M.C., *et al.* (2014) Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, **159**, 647–661.
23. Li,W., O'Neill,K.R., Haft,D.H., Dicuccio,M., Chetvernin,V., Badretdin,A., Coulouris,G., Chitsaz,F., Derbyshire,M.K., Durkin,A.S., *et al.* (2021) RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucleic Acids Res*, **49**, D1020–

D1028.
24. McTavish,H., LaQuier,F., Arciero,D., Logan,M., Mundfrom,G., Fuchs,J.A. and Hooper,A.B. (1993) Multiple copies of genes coding for electron transport proteins in the bacterium Nitrosomonas europaea. *J Bacteriol*, **175**, 2445–2447.
25. Schrider,D.R. and Hahn,M.W. (2010) Gene copy-number polymorphism in nature. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 3213–3221.
26. Johnson,M., Zaretskaya,I., Raytselis,Y., Merezhuk,Y., McGinnis,S. and Madden,T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res*, **36**, W5–W9.
27. Qi,L.S., Larson,M.H., Gilbert,L.A., Doudna,J.A., Weissman,J.S., Arkin,A.P. and Lim,W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–1183.
28. Boyle,E.A., Andreasson,J.O.L., Chircus,L.M., Sternberg,S.H., Wu,M.J., Guegler,C.K., Doudna,J.A. and Greenleaf,W.J. (2017) High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proceedings of the National Academy of Sciences*, **114**, 5461–5466.
29. Barrios,D. and Prieto,C. (2017) D3GB: An Interactive Genome Browser for R, Python, and WordPress. *J Comput Biol*, **24**, 447–449.
30. Feng,H., Li,F., Wang,T., Xing,X., Zeng,A. and Zhang,C. (2022) Deep-learning-assisted Sort-Seq enables high-throughput profiling of gene expression characteristics with high precision. *bioRxiv*, 10.1101/2022.11.23.517705.
31. 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.
32. 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 2015 533:7603*, **533**, 420–424.
33. Chen,B., Gilbert,L.A., Cimini,B.A., Schnitzbauer,J., Zhang,W., Li,G.W., Park,J., Blackburn,E.H., Weissman,J.S., Qi,L.S., *et al.* (2013) Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell*, **155**, 1479–1491.
34. Yu,X., Li,S., Feng,H., Liao,X., Xing,X.H., Bai,Z., Liu,X. and Zhang,C. (2023) CRISPRi-microfluidics screening enables genome-scale target identification for high-titer protein production and secretion. *Metab Eng*, **75**, 192–204.
35. Sugimoto,N., Nakano,S., Katoh,M., Matsumura,A., Nakamuta,H., Ohmichi,T., Yoneyama,M. and Sasaki,M. (1995) Thermodynamic Parameters To Predict Stability of RNA/DNA Hybrid Duplexes. *Biochemistry*, **34**, 11211–11216.
36. Watkins,N.E., Kennelly,W.J., Tsay,M.J., Tuin,A., Swenson,L., Lee,H.-R., Morosyuk,S., Hicks,D.A. and SantaLucia,J. (2011) Thermodynamic contributions of single internal rA·dA, rC·dC, rG·dG and rU·dT mismatches in RNA/DNA duplexes. *Nucleic Acids Res*, **39**, 1894–1902.
37. Sugimoto,N., Nakano,M. and Nakano,S. (2000) Thermodynamics−Structure Relationship of Single Mismatches in RNA/DNA Duplexes. *Biochemistry*, **39**, 11270–11281.

TABLE AND FIGURES LEGENDS

A



B

Location	Strand	Length	Gene	Guide sequence	Guide position	Guide number
20815 ..21078	-	87	*rpsT*	AAGTACGCATCATAGAGCGA	20988	1
190..255	+	21	*thrL*	CGCACCGTTACCTGTGGTAA GGTGGTGCTAATGCGTTTCA	249 210	2

C

Guide sequence	Off-target sequence	penalty Score	Off-target number
TTAAAACCAGTCAGTTGCTC	TGAAGGCGTGCCAGTTGCTC TAAAGAACAGTCAGATGCTC	19.0 15.5	2
ATCTCGGCATTCAATTTCTT	ATAGAGGAAGTGAATTTCTT CAAACGCCAGTCAATTTCTT	19.0 17.0	2

Figure 1. Workflow and results of a CRISPRi-FGP query. **(A)** The main design process starts when a "start calculation" order is received, and the website automatically redirects to the progress page that displays the real-time progress of the calculation. After the calculation is complete, CRISPRi-FGP redirects again to the results page. Users are free to download the result tables and view the sgRNA library on the interactive page. **(B)** The designed sgRNA library is grouped by its targeted gene. **(C)**

The off-target log lists sgRNAs predicted to have potential off-target hits.

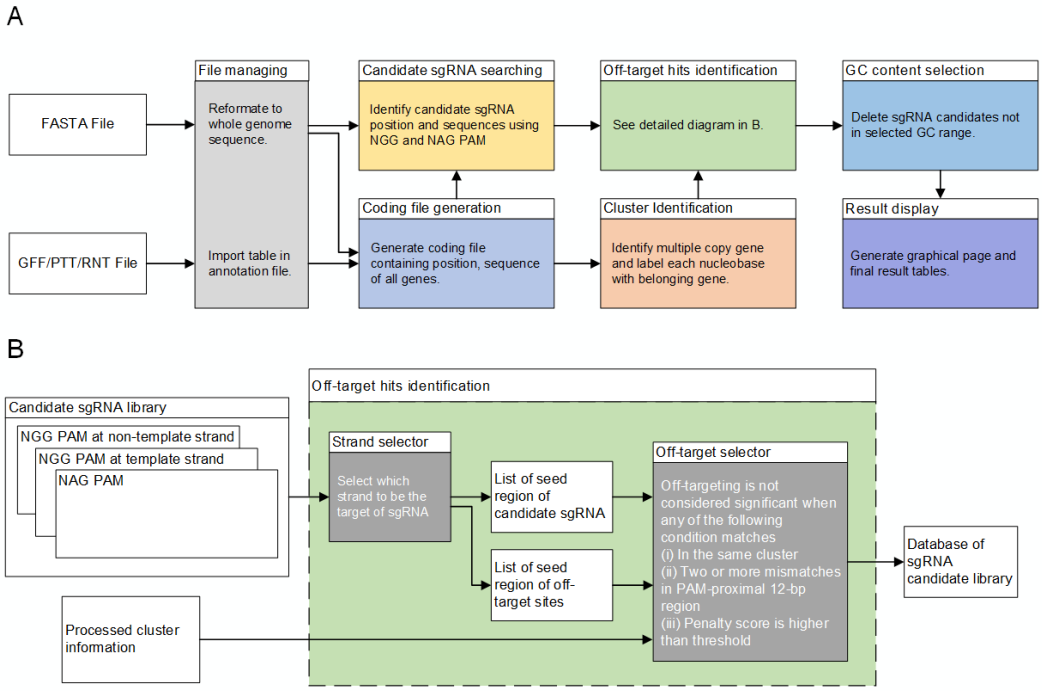


Figure 2. Functional blocks of CRISPRi-FGP. **(A)** The flowchart of CRISPRi-FGP. Two input data files (FASTA and either GFF or PTT/RNT) are processed by multiple modules. A coding file is generated based on these two files. The subsequent modules perform the candidate sgRNA search, quality control (off-target hits identification, and GC content selection), and results display. **(B)** The off-target hits identification module generates three FASTA files containing 12 bp PAM-proximal sequences of sgRNA candidates. Alignments are performed among these three files to identify potential off-target hits.

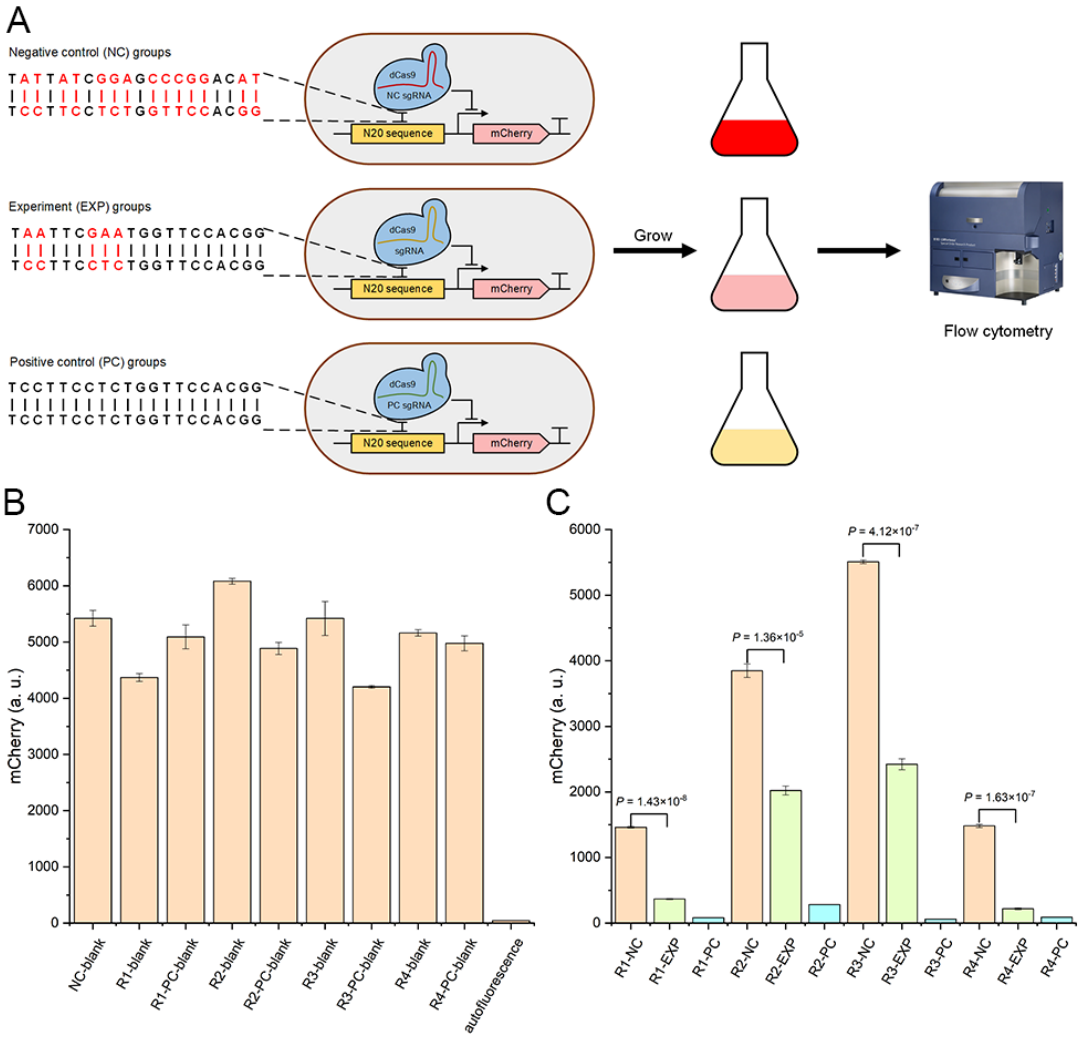


Figure 3. Flow cytometry assay of off-target binding affinities. **(A)** The binding affinity of each mismatched sgRNA was compared with positive controls and negative controls. Fluorescence intensities of the three groups were measured by flow cytometry. **(B)** The fluorescence intensities of the blank control groups demonstrated that the expression of sgRNAs only had a moderate influence on *mcherry* expression. Therefore, the fluorescence intensity variations between EXP, PC, and NC groups originated from the differences in binding affinities of sgRNAs. **(C)** All sgRNAs from the EXP groups exhibited significant off-target binding affinities. In both **(B)** and **(C)**, error bars represent the standard deviations of biological replicates (n = 3).

Table 1. sgRNAs designed by other tools and potential off-target hits discovered by CRISPRi-FGP

No.	Design tool	Rank*	Target gene	sgRNA sequence	Off-target gene	Potential off-target hits discovered by CRISPRi-FGP
R1	chopchop	1	*Cgl0025*	T***AA***TTC***GAA***TGGTTCCACGG	*Cgl2364*	T***CC***TTC***CTC***TGGTTCCACGG
R2	chopchop	3	*Cgl0044*	***GCACGAT***GCCCACCACACCG	*Cgl1968*	***CGGATGG***GCCCACCACACCG
R3	Synthego	3	*Cgl0005*	***TTCT***G***CGC***ATGGCTTCGGCG	*Cgl0369*	***GCGC***G***GCG***ATGGCTTCGGCG
R4	Synthego	2	*Cgl0059*	***GAG***CT***CAG***ATCGCTCAACAT	*Cgl1680*	***TCT***CT***TGA***ATCGCTCAACAT

*For the rank of sgRNA in other tools, the ranking method differed for each tool

Table 2. sgRNAs and N20 sequences of each test group

Test group	sgRNA sequence	N20 sequence
R1-NC	TATTATCGGAGCCCGGACAT	
R1-EXP	TAATTCGAATGGTTCCACGG	TCCTTCCTCTGGTTCCACGG
R1-PC	TCCTTCCTCTGGTTCCACGG	
R2-NC	TATTATCGGAGCCCGGACAT	
R2-EXP	GCACGATGCCCACCACACCG	CGGATGGGCCCACCACACCG
R2-PC	CGGATGGGCCCACCACACCG	
R3-NC	TATTATCGGAGCCCGGACAT	
R3-EXP	TTCTGCGCATGGCTTCGGCG	GCGCGGCGATGGCTTCGGCG
R3-PC	GCGCGGCGATGGCTTCGGCG	
R4-NC	TATTATCGGAGCCCGGACAT	
R4-EXP	GAGCTCAGATCGCTCAACAT	TCTCTTGAATCGCTCAACAT
R4-PC	TCTCTTGAATCGCTCAACAT	
NC-blank	TATTATCGGAGCCCGGACAT	
R1-blank	TAATTCGAATGGTTCCACGG	
R1-PC-blank	TCCTTCCTCTGGTTCCACGG	
R2-blank	GCACGATGCCCACCACACCG	
R2-PC-blank	CGGATGGGCCCACCACACCG	CATGTATTATACACGAAGTT
R3-blank	TTCTGCGCATGGCTTCGGCG	
R3-PC-blank	GCGCGGCGATGGCTTCGGCG	
R4-blank	GAGCTCAGATCGCTCAACAT	
R4-PC-blank	TCTCTTGAATCGCTCAACAT	