Briefings in Bioinformatics, 2020, pp. 1–13

doi: DOI HERE
Advance Access Publication Date: Day Month Year
Paper

PAPER

Multi-gene editing: current approaches and beyond

Hui Peng,[1,2] Yi Zheng,[1] Zhixun Zhao[1] and Jinyan Li[1,*]

[1]Data Science Institute, University of Technology Sydney, PO Box 123, Ultimo, NSW 2007, Australia and [2]School of Computing, National University of Singapore, 13 Computing Drive, 117417, Singapore

[*]Corresponding author. jinyan.li@uts.edu.au

FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

Abstract

CRISPR/Cas9 multi-gene editing is an active and widely studied topic in the fields of biomedicine and biology. It involves a simultaneous participation of multiple single-guide RNAs (sgRNAs) to edit multiple target genes in a way that each gene is edited by one of these sgRNAs. There are possibly numerous sgRNA candidates capable of on-target editing on each of these genes with various efficiencies. Meanwhile, each of these sgRNA candidates may cause unwanted off-target editing at many other genes. Therefore, selection optimization of these multiple sgRNAs is demanded so as to minimize the number of sgRNAs and thus reduce the collective negative effects caused by the off-target editing. This survey reviews wet-lab approaches to the implementation of multi-gene editing and their needs of computational tools for better design. We found that though off-target editing is unavoidable during the gene editing, those disfavored cuttings by some target genes' sgRNAs can potentially become on-target editing sites for some other genes of interests. This off-to-on role conversion is beneficial to optimize the sgRNA selection in multi-gene editing. We present a preference cutting score to assess those beneficial off-target cutting sites which have a few mismatches with their host genes' on-target editing sites. These potential sgRNAs can be prioritized for recommendation via ranking their on-target average cutting efficiency, the total off-target site number and their average preference cutting score. We also present case studies on cancer-associated genes to demonstrate tremendous usefulness of the new method.

Key words: CRISPR/Cas9, multi-gene editing, off-target editing, preference cutting score

Introduction

The CRISPR/Cas9 system (the clustered, regularly interspaced, short palindromic repeats/CRISPR-associated protein 9 system) is one of the most popular gene editing tools to modify an organism's DNA [33, 22, 12]. Recently, multi-gene editing has been of strong interests in a wide range of applications where the genes are functioning dependently and cooperatively. Multi-gene editing involves a simultaneous participation of multiple sgRNAs to edit multiple target genes in a way that each gene is edited by one of these sgRNAs. The applications include studies on complex diseases such as cancers which are caused by the dysfunction of many genes [80, 10, 70], investigations on genes forming regulation networks [58, 43], and gene groups that are often involved in the same pathway to execute special biological functions [11, 27]. Multi-gene editing has been also intensively studied for crop quality enhancement [64, 76, 88, 89, 30, 52, 34], for microbial production improvement [2, 60], and for model organism construction [37, 90]. In particular, Otoupal et al. [60] studied the disruption of both CAR2 and URA3 genes in a single transformation to improve the ability of generating bioproducts by Rhodosporidium toruloides; Xie et al. [96] designed synthetic genes with a tandemly arrayed tRNA-gRNA architecture to boost multiple gene editing capability of the CRISPR-Cas9 system; and Zhang et al. [105] reported a gRNA-tRNA array for CRISPR-Cas9 (GTR-CRISPR) for multiplexed engineering of Saccharomyces cerevisiae. These studies demonstrated the possibility of multi-gene editing with CRISPR/Cas9 to solve real-life problems, where as many as six genes were edited at one time [105].

On the other hand, off-target editing effect generated by the CRISPR/Cas9 system largely prevents its usage in critical fields such as gene therapy [56] and crops' genetic modification [29]. New gene editing systems [40, 42] have even been proposed to overcome this huge challenge, or adjustments have been proposed to change the interior system [31, 41], or to select optimal on-target cutting sites to avoid off-target editing

© The Author 2020. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

events [18, 47]. To understand the benefits and concerns of the current approaches in the implementation of multi-gene editing, we compare their key steps and make our critical review.

We observed that computational tools are essentially important to assist the design of an optimal multi-gene editing system, and we found that optimal CRISPR/Cas9 systems should consider a panel of mutually beneficial sgRNAs. Although some existing computational methods have been proposed to predict the on-target cutting efficiency of sgRNAs [18, 65, 61] and to detect their possible off-target editing sites [47, 62, 13], all of these methods are just capable of predicting sgRNAs to target single genes. So far, there is nearly no good computational method to solve the multi-gene editing design problem.

Beyond these methods and ideas, taking as a perspective, we propose a novel method to recommend an optimal panel of sgRNAs for multi-gene editing. Our main idea is to select as less sgRNAs as possible to edit multiple genes simultaneously with a higher average cutting efficiency but lower total off-target effects. This goal is achieved by prioritizing those beneficial off-target cutting sites of the selected sgRNA for targeting one gene of interests and treating them as the on-target cutting sites for editing the other genes of interests. We define a preference cutting score to measure the efficiency differences of a cutting site (on-target cutting or off-target cutting) comparing to the other editing sites. Then, one or more selected sgRNAs are ranked for recommendation to edit single or multi-genes by considering the cutting efficiency, off-target effect and preference cutting score. As case studies, we applied our ranking method to 717 cancer genes from the Cancer Gene Census (CGC) [75]. We present an off-line tool for helping multi-gene editing tool design. The tool and its source codes are available from Github: https://github.com/PennHui2016/MultiGeneEditing.

Multi-gene editing: current applications and methods

The key definition of CRISPR/Cas9 multi-gene editing is the simultaneous participation of multiple sgRNAs to edit the multiple target genes in a way that each gene is edited by one of the sgRNAs. Multi-gene editing has been adopted to edit genomes of various species such as human [15, 57, 109, 9, 7]; other animals, e.g., hamster [23, 73], mouse [104, 54, 107], Zebrafish [36, 67, 5], rabbit [99, 49]; plants, e.g., wheat [89, 6], maize [97, 64], Arabidopsis [93, 106, 25, 51], tomato [46, 30, 34], rice [96, 91, 92], Cabbage [52], rapeseed [76]; microorganisms, e.g., yeast [72, 32, 87, 1], Candida albicans [83, 35], Bacillus subtilis [94, 108, 48, 95], Staphylococcus aureus [21], Pichia pastoris [50, 100], Aspergillus niger [20], and Myceliophthora thermophila [45].

The implementation of these CRISPR/Cas systems to edit the candidate genes consists of two phases of work: system design and the formation of the complete CRISPR/Cas complex in vivo.

Multi-gene editing system design

The phase of editing system design needs to choose the type of the system and then to elaborate on its structure. The CRISPR/Cas9 system [44] and the CRISPR/Cpf1 (or CRISPR/Cas12a) system [104] are the two most widely used gene-editing systems for implementing multi-gene editing. Both of them are composed of a Cas protein for playing the role of endonuclease and a RNA guide part for positioning. Detailed structures of these two systems have been described and compared in extensive literature [39, 38, 98, 19, 77, 79, 28]. Most parts of these two systems are fixed, only the spacer sequence in the CrRNA is changed accompanying with the candidate target gene. Thus, after the system type of CRISPR/Cas is decided, only the spacer sequences are required to consider and design.

The CRISPR/Cas systems of multi-gene editing can be classified into Single-CRISPR/Cas system Multi-Gene Editing (SCMGE) where only one CRISPR/Cas system is expected to edit more than one gene [82, 102, 49], and Multi-CRISPR/Cas system Multi-Gene Editing (MCMGE) where each candidate gene is targeted by its specificity-oriented CRISPR/Cas system [15, 104, 69]. SCMGE is useful to edit those homology genes whose sequences are similar to each other, on the other hand for those non-homology genes, MCMGE is suitable.

The strategy for SCMGE spacer design is to find an optimal conserved locus across all the homology genes. While for MCMGE, current issues are handled by separately selecting one optimal spacer for each candidate gene via existing computational tools [17, 53].

For single gene editing, various computational tools have been proposed to help select the spacer sequence (see reviews [14, 16, 4]). Only a few of the tools supports the design of spacers for multi-gene editing. Prykhozhij et al. [63] proposed the tool CRISPR MultiTargeter which can select common and unique spacers for similar genes. For the selection of spacers for the design of MCMGE, we have not yet known any computational tools. Though, for different genes, one can run an existing computational tool several times to match the selection to each candidate gene, there exist some drawbacks. Firstly, it is not convenient to run the tool a lot of times if many candidate genes are required to edit. Secondly, the total off-target effect are hard to be summarised. Thirdly, these designs have not considered the optimized situation where a smaller number of spacers can be grouped to edit a larger number of genes through those beneficial off-target sites. This is our motivation to propose a computational optimization tool for assisting the practical applications of multi-gene editing.

In vivo CRISPR/Cas system assembly

After the design of a CRISPR/Cas system for multi-gene editing, the next phase is to implement an in vivo assembly of the CRISPR/Cas system. The two primary parts of a CRISPR/Cas system, i.e., the Cas protein and the sgRNA, are always expressed from RNA ploy-II and RNA ploy-III respectively. For SCMGE, the in vivo assembly of the editing system is the same as the single gene editing process. However, for MCMGE, the main difficulty is to express multiple sgRNAs and assemble different editing systems for different candidate genes.

Currently, several strategies have been adopted to express multi-sgRNAs in vivo. For example, several sgRNAs can be co-expressed through multiple sgRNA expression cassettes with their own promoters [97, 87]. Or different sgRNAs can be combined as a tandem array and expressed from a single promoter. In the tandem array, two sgRNAs are separated by a cutting feature such as the sequences that can be recognised by the RNA endonuclease Csy4 (Csy4 cleavage) [26], the hammerhead ribozyme and HDV ribozyme flanked 5' and 3' of the sgRNA (self-cleavage) [101] and the tRNA which can be cut by endogenous RNase P and RNase Z (tRNA

Processing) [81, 105]. Another strategy is to express the crRNA array containing several spacers separated by a direct repeat with a promoter [3, 71]. Three assembly methods have been used to synthesis sgRNA arrays such as Oligo-based method, combining PCRs and Golden Gate or Gibson assembly and direct synthesis and ligation method [53].

Different carriers are often used to help express the multi-gene editing systems. For prokaryotes, helper plasmids (single or multiple plasmids) are always needed [2]. For eukaryotes, helper plasmid can be used as well, or the expression components of Cas protein and the gRNAs can be integrated into the target genome DNA [97, 2]. For some cases, researchers have co-injected the Cas protein and the sgRNAs into the zygotes [44, 59]. More detailed reviews about the in vivo CRISPR/Cas system assembly are referred to [66, 55, 2, 1, 17, 53]. Our own collected literature list from Google scholar (https://scholar.google.com/) and pubmed (https://pubmed.ncbi.nlm.nih.gov/) using the key words of "CRISPR" and "multi-gene" is attached at **Supplementary file 1**.

Optimization of CRISPR/Cas9 multi-gene editing through prioritizing beneficial off-target editing sites: the idea beyond the current approaches

An off-target cutting is an event happened at a genome site located away from the objective cutting site of an sgRNA designed for the target gene. If we also need to edit this off-target cutting site's host gene, then we can use the same sgRNA to target the two genes simultaneously. This is our perspective beyond the current approaches to optimise the selection on a panel of sgRNAs for multi-gene editing. Endo et. al. [24] also adopted this strategy to implement multi-gene knockout in rice. Next, we present a definition to describe a preference cutting score to measure the efficiency differences of a cutting site (on-target cutting or off-target cutting) comparing to the other editing sites. Then, one or more selected sgRNAs are ranked for recommendation to edit single or multi-genes by considering the cutting efficiency, off-target effect and preference cutting score. We also compare different recommendation methods to understand the differences in the sgRNA panels for editing cancer associated genes.

Preference cutting scores for a list of editing sites

The CRISPR/Cas9 complexes of an sgRNA may bind to different genes. We define a numerical cutting score to measure the editing preference of this sgRNA at these editing sites based on the difference of the editing efficiency at a given site compared with all the other editing sites. It is named a preference cutting score at an editing site for the sgRNA. An sgRNA with a higher cutting preference at an on-target editing site than its off-target sites is selected to achieve a higher specificity. Let $C = \{c_1, c_2, c_i, \ldots, c_n\}$ be the set of editing sites (including the on-target and off-target sites) of an sgRNA sg, the preference cutting score (PCS) at a site c_i is defined by equation (1):

$$PCS(c_i, sg) = \frac{1}{n-1} \sum_{j=1, j\neq i}^{n} (f(c_i, sg) - f(cj, sg)) \times \frac{|f(*, sg) < f(c_i, sg)|}{n-1} \quad (1)$$

where $f(c_k, sg)$ is the editing efficiency of sg at site c_k, $|f(*, sg) < f(c_i, sg)|$ means the number of sites having a cutting efficiency lower than c_i in C. PCS includes two parts: the left part is the average difference between the on-target cutting efficiency and all the off-target cutting efficiencies; the right part is a penalty factor to compute the percentage of the off-target cutting sites having higher cutting efficiencies than the on-target cutting site. The average difference helps identify an optimal on-target site that has a higher cutting efficiency than all of the off-target cutting sites. In particular, the multiplication of the penalty factor $\frac{f(*,, sg) < f(c_i, sg)|}{n-1}$ can give a penalty to the situations when there are many editing sites having bigger cutting efficiencies than our expected one. This can help avoid selecting those sgRNAs whose off-target cutting sites have an equal or even higher cutting efficiency comparing to their on-target sites. Though the off-target number is small, the off-target effect can be still significant.

The on-target cutting efficiency of an sgRNA for a given gene is predicted by our recently published TSAM method [61], which can recommend a list of potential spacer sequences, their target site information and can predict the corresponding on-target cutting efficiencies as well.

Off-target editing efficiency estimation

For an sgRNA, its genome wide possible off-target sites are predicted via our published SVM ensemble method (ESC), which has been proved to exceed the performance of the state-of-the-art methods [62]. The output of this off-target detection tool is a list of potential off-target sites with detailed genome location information, mismatch numbers and predicted probabilities. However, it does not predict the exact off-target editing efficiency. There already exist some methods that can perform this prediction such as Elevation [47]. However, the predicted efficiencies may be not comparable to the on-target efficiencies predicted by TSAM as they are not in the same scale.

To address this problem, we make estimations on the off-target cutting efficiency based on its on-target editing efficiency and the distribution of the mismatches between the off-target site sequence and the on-target site sequence. Doench et al. [18] studied the impact of mismatch positions and types on the cutting efficiency by introducing various types of mutations to different positions of the sgRNAs and measuring the decreasing rates of the cutting efficiencies. For an sgRNA's spacer sequence with the length of 20, there are 12 types of mismatches at each position. Doench et al. provided a percent-activity profile of these 240 kinds of mismatches $PA = \{pa_1^1, \cdots, pa_j^i, pa_{20}^{12}\}$, where $i \in [1, 12]$ is the mismatch type number, $j \in [1, 20]$ is the position number and $pa_j^i \in [0, 1]$ (See **Supplementary file 2**). Given an off-target site sequence $otSeq$ and its corresponding sgRNA spacer sequence $sgSeq$, we calculate its n mismatches $mis(otSeq, sgSeq) = \{mis_1, mis_2, \cdots, mis_k, \cdots, mis_n\}$ and their percent-activity values $pa(otSeq, sgSeq) = \{p_1, p_2, p_k, \cdots, p_n\}$. Let the on-target cutting efficiency of the sgRNA with $sgSeq$ be $on(sgSet)$, then the cutting efficiency at $otSeq$ is computed using equation (2):

$$Ofc(otSeq) = \prod_{p_k \in pa(otSeq, sgSeq)} on(sgSet) \cdot p_k \quad (2)$$

Algorithm 1 Optimal sgRNA selection for single gene editing.

Require: gene sequence g_seq of candidate gene A;
1: Adopt the tool TSAM to obtain the potential on-target cutting sites of A where $CS(g_seq) = \{cs_1, cs_2, \cdots, cs_i, \cdots, cs_n\}$ and compute their cutting efficiencies $CE(g_seq) = \{ce_1, ce_2, \cdots, ce_i, \cdots, ce_n\}$;
2: **for** cs_i **in** $CS(g_seq)$ **do**
3: Predict the off-target cutting sites with ESC: $OT(cs_i) = \{ot_1, ot_2, \cdots, ot_j, \cdots, ot_m\}$;
4: Estimate the off-target site cutting efficiency at ot_j as $otce_j$ via above formula (2)
5: Compute the PCS of cs_i:
6: $PCS(cs_i) = \frac{1}{m}\sum_{j=1}^{m}(ce_i - otce_j) \cdot \frac{|otce_j < ce_i|}{m}$
7: **end for**
8: Rank cs_i:
9: $rank(cs_i) = (rank(ce_i) + rank(|OT(cs_i)|) + rank(PCS(cs_i)))/3$
10: **Output** The rank list of A's on-target cutting sites $CS(g_seq)'$

Selection of an optimal sgRNA for single-gene editing

Given a gene of interests, there are potentially many sgRNAs for the on-target editing of this gene. An optimal sgRNA to edit this gene is determined primarily by considering the sgRNA's on-target editing efficiency, in addition to the number of its off-target sites and its preference cutting score. The resulted CRISPR/Cas9 system should edit the gene with a high efficiency meanwhile having a low off-target editing effect. The special exploit of the preference cutting score is to avoid any high-efficiency off-target editing event. The following pseudo codes in **Algorithm 1** describe the steps for selecting an optimal sgRNA in single-gene editing.

The key step of **Algorithm 1** is to sort the on-target sgRNAs into a descending order according to their average rank over the on-target cutting efficiency, the off-target number and their PCS. $|OT(cs_i)|$ stands for the number of off-target sites in terms of the on-target editing site cs_i. In fact, one can change the weights of these three types of ranks instead of using the same weights. However, we don't have datasets for optimizing them, thus averaging strategy is adopted currently. **Fig 1(a)** shows a simple schematic diagram of single gene editing design.

Most potential cutting partner (MPCP) of a gene and the gene's most beneficial off-target cutting site (MBOCS)

Suppose A is a gene. Denote $sg(A) = \{sg_1^A, sg_2^A, \ldots, sg_t^A, \ldots, sg_n^A\}$ as n number of sgRNAs which are capable of editing A at different sites. Each of these sgRNAs, for example sg_t^A, has a number of off-target sites. We denote these off-target sites as a set $OFF(sg_t^A)$, and denote the on-target editing site of sg_1^A as $on(sg_1^A)$. We calculate the PCS and cutting efficiency of sg_1^A at every site in $OFF(sg_1^A)$. These sites in $OFF(sg_1^A)$ are then ranked in terms of their averages over the cutting efficiencies (ce) and PCSs: $2 * ce * PCS/(ce + PCS)$, where we used the harmonic mean instead of arithmetic mean to avoid the large difference between the two terms.

The top ranked off-target site of sg_1^A is the most beneficial off-target cutting site (MBOCS) of sg_1^A (namely a local MBOCS). Meanwhile, the host gene of this MBOCS is the local most potential cutting partner (MPCP) of A. We repeat the process for every sg_t^A to produce n local MBOCSs and their corresponding local MPCPs.

Then we rank all the local MBOCSs to find the best local MBOCS as the final global MBOCS, thus the MPCP of gene A (namely the global MPCP) is determined. This ranking process includes three steps: (a) combine the local MBOCS (lm_t) and its related sgRNA's on target site $on(sg_t^A)$ as a site-pair $< lm_t, on(sg_t^A) >$; (b) compute the site-pair's average cutting efficiency $ace_t = 2 * ce(lm_t) * ce(on(sg_t^A))/(ce(lm_t) + ce(on(sg_t^A)))$, the average PCS as $aPCS_t$ (in the same way as ace) and the off-target site number $numOT_t = |OFF(sg_t^A)| - 1$ (here, the MBOCS is not regarded as an off-target site); (c) rank $ace, aPCS, numOT$ separately and calculate the average rank for each pair as its final rank. We denote the final MPCP as AP, where the corresponding sgRNA is sg_p^A and its MBOCS is $OFF(sg_p^A)[t]$.

The purpose of defining the MPCP of gene A is two-fold: firstly, we can identify the most possible unwanted mutation that the CRISPR/Cas9 system of sg_p^A would introduce when editing A; secondly, we may select sg_p^A to implement a paired gene editing (i.e., editing on A and AP) with an optimal overall performance.

We note that if gene B is the most beneficial off-target cutting site of gene A, it does not necessarily mean gene A is the most beneficial off-target cutting site of gene B. But it is highly possible that gene A is a top-ranked beneficial off-target cutting site of gene B. An ideal case is that A and B can be mutually the most beneficial off-target cutting sites.

The following **Algorithm 2** with detailed steps can be used to obtain the MPCP and MBOCS of given gene A. **Fig 1(b)** shows an example of detecting MPCP and MBOCS.

Algorithm 2 considers only 20% top-ranked potential on-target cutting sites to determine the MPCP and MBOCS of a gene. There exist two-tier loops in the algorithm. In the inner loop, we find the local MBOCS and local MPCP for each considered on-target cutting site cs_i' while the outside loop returns the global MBOCS and MPCP of A.

Recommendation of an sgRNAs panel for multi-gene editing

For the simultaneous editing of multiple genes, three approaches can be used to design the CRISPR/Cas9 system:

- The first approach is to separately select an optimal sgRNA for each of the genes and then construct a co-expressing vector or tandemly arrayed tRNA-gRNA unit to implement the multi-gene editing; we name this approach one-to-one (OTO) (see an example in **Fig 2(a)**). For this approach, we can apply our **Algorithm 1** to find the best sgRNAs for each of the candidate genes.
- The second approach is to select a common sgRNA that can target all of the genes in the group. The approach is named one-to-all (OTA) (see an example in **Fig 2(c)**). This approach is practically applicable only when these genes are homology genes, having the same cutting site in each sequence. We find the common sgRNAs by intersecting the lists of on-target cutting sites for each of the candidate genes derived from **Algorithm 1**.
- The third approach is to detect beneficial off-target editing sites within the group of genes. This approach may result in less number of sgRNAs to target all of the genes. This approach is named less-to-all (LTA) (see an example in **Fig 2(b)**). There are two situations for this approach: s1–one sgRNA LTA (os-LTA), where we can choose one sgRNA from the candidate list of a gene and it can off-target all of the remaining genes in the group; s2–multi-sgRNA LTA,

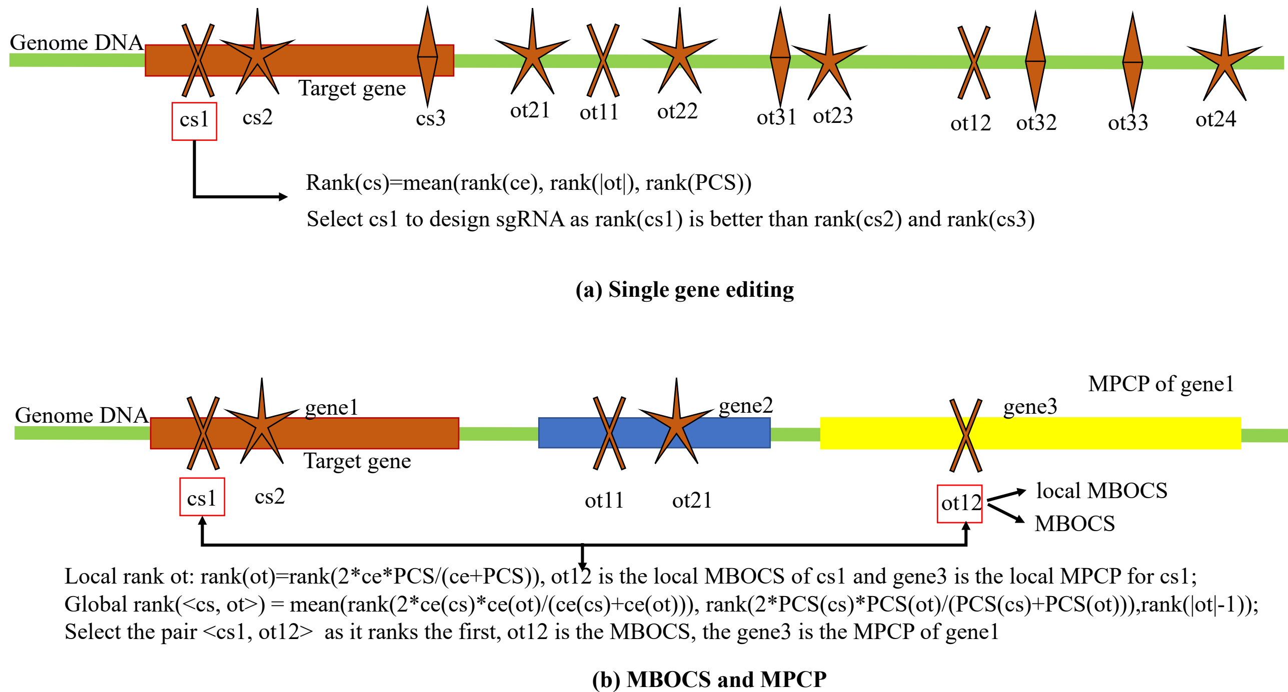

Fig. 1. Schematic diagram for single gene editing, and for searching MPCP and MBOCS. Here cs stands for the on-target cutting site, ot represents the off-target site, ce is the cutting efficiency, PCS is the PCS score and $|ot|$ gives the number of off-target sites; ot11 and ot12 are the off-target sites for the sgRNA designed with cs1.

where for a subgroup of the genes, our above OTA or os-LTA can be adopted and the remaining genes require OTO. For example in the case of 6 genes to be edited, OTA can be used for 2 of them, os-LTA can be used for another 2 and the remaining two suit for the OTO, thus totally 4 sgRNAs are required. For the os-LTA, we implement it via successively selecting one gene from the group (or subgroup in s2) to get its candidate sgRNAs together with each sgRNA's off-target sites, then search whether the remaining genes contain the off-target sites. If found, we turn the off-target cutting to an on-target cutting of the corresponding gene.

The OTO approach has nearly no application limits while OTA and LTA can just work for those genes contain identical or highly similar cutting sites. OTA and LTA may have less off-target effect and the resulting CRISRP/Cas9 system is simpler than that by OTO. For a given group of candidate genes to be edited, all these three approaches are applicable. The potential sgRNA/sgRNAs are sorted according to their average ranks of the mean cutting efficiencies, total off-target site numbers and mean PCS scores (in the way similar to previous single gene editing).

One important restriction in the recommendation of sgRNAs for multi-gene editing is that we do not allow more than one sgRNAs to edit one gene. This is to avoid generating unintended mutations which may result in elimination of a long sequence in the gene when the two cutting sites are close to each other [68]. Suppose three genes A, B, C are required to be edited. When OTO is used, three sgRNAs sg_A, sg_B, sg_C can be designed for targeting the three genes separately. If there exists an sgRNA sg_{common} that targets all of the three genes, then we require the OTA strategy. For the LTA, we require that one or two sgRNAs could be selected. For the one-sgRNA based LTA, the sgRNA sg_{lta1} will generate on-target cutting on one of the three genes, e.g. A and off-target editing on both of the two remaining genes, i.e. B and C. For the two-sgRNA based LTA, one sgRNA sg_{lta1} will on-target one gene A and off-target the gene B while another sgRNA sg_{lta2} will on-target the remaining gene C. The situation of sg_{lta1} on-target cutting A and off-target cutting B while sg_{lta2} on-target cutting B and off-target cutting C is not considered. Our method recommends the best sgRNA/sgRNAs derived by these three approaches for further wet-lab experimental validation.

Case study on cancer associated genes

Cancer associated genes were collected from the Cancer Gene Census (https://cancer.sanger.ac.uk/cosmic/census?tier=all) on April 24, 2019. Related data of 723 cancer genes were exported from the two tiers. Six genes (IGH, IGK, IGL, TRA, TRB and TRD) were not labeled with ensembl ids [103], and they were excluded from our analysis. The annotation information of the remaining 717 genes were obtained from the ensembl database (version GRCh38.92) [103] (See **Supplementary file 3**).

Pairwise-editing of cancer genes and their MPCPs: safety analysis and effectiveness

We detected MPCPs for all of the 717 cancer genes. We also conducted analysis on the safety and effectiveness of the pairwise-editing by our suggested sgRNA at each cancer gene and its MPCP (see **Supplementary file 4** for more details).

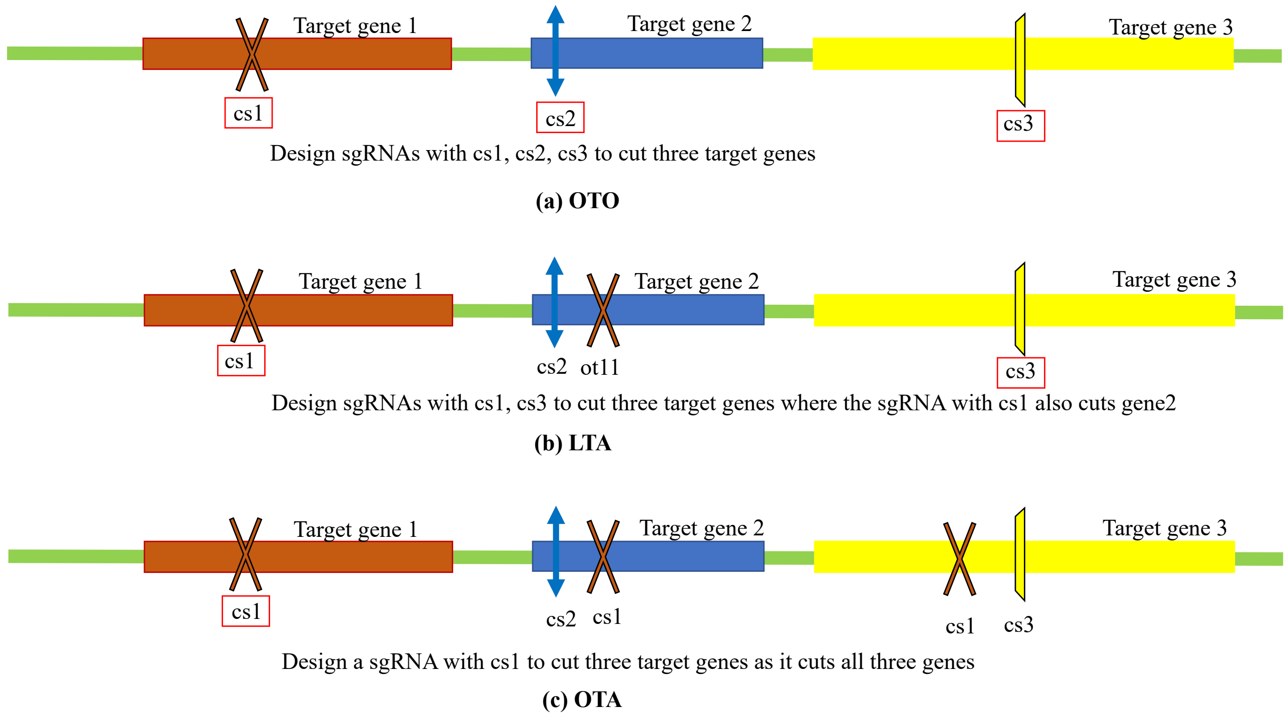

Fig. 2. Illustrations of the three strategies OTO, LTA and OTA. Here cs stands for the on-target cutting site, ot represents the off-target site. ot11 and ot12 are the off-target sites for the sgRNA designed with cs1.

Pairwise editing safety analysis.
It was reported that small distance between two nicking sites may introduce indels [68]. Such fragment deletion should be avoided when a cancer gene and its MPCP are edited by one spacer sg_t. We examined the distance between every cancer gene and its MPCP. We found that most of the cancer genes' MPCP are located at different chromosomes. Only about 7.5% (54 out of 717) of the cancer genes and their MPCPs are located at the same chromosome, and the average distance between these cutting sites at the cancer genes and those at their MPCPs is 4.86E+7bp.

The nearest one is between gene KLK2 (ensembl id: ENSG00000167751) and its MPCP KLK3 (ensembl id: ENSG00000142515), the distance is about 18487bp. It is understood that these cancer genes' MPCPs are always far from themselves. This suggests that it is safe to implement these pairwise-editings on a cancer gene and its MPCP via the recommended sgRNA.

Effectiveness.
We found that the cutting efficiencies at these editing sites for the cancer genes and their MPCPs are high. The average cutting efficiency (in the range [0, 1]) of the editing sites at these cancer genes is about 0.78, the average cutting efficiency at their MPCPS is about 0.58, and the average cutting efficiency of each pair is much higher than 0.5. The preference cutting scores (PCS) of the cancer genes and their MPCPs are also high: the average PCS of the sites for the 717 cancer genes is 0.66 and the average PCS for their MPCPs is 0.46. By our definition of the PCS, a bigger PCS implies that the sgRNA's on-target editing efficiency is higher than that at its off-target cutting sites. This suggests that our method is capable of selecting an optimal sgRNA sg_t to edit a cancer gene and its MPCP simultaneously with a higher on-target cutting efficiency and specificity than using any other sgRNA. A boxplot in **Fig 3** shows a detailed distribution of these cutting efficiency values.

Recommendation of sgRNAs for pairwise editing on cancer-associated genes

Of the 717 cancer genes, there are 256686 unique pairs. Given a pair of these genes, we attempted to recommend one sgRNA for each of the two genes for editing (namely, by the OTO strategy) or recommend one single sgRNA to edit the two genes simultaneously (namely, by the OTA/LTA strategy). If more than one sgRNAs are qualified by the OTA/LTA strategy, only the top ranked one was recommended.

We found that about 57.3% of these pairs can be edited by single sgRNAs with high cutting efficiencies (namely, 147162 out of 256686 pairs, including 4614 pairs determined by the OTA strategy and 145692 determined by the LTA strategy, where 3144 pairs were determined by both OTA and LTA). In addition, there are 42 pairs of cancer genes where one gene is the other gene's MPCP. Comparing with sgRNAs selected by the OTO strategy, the sgRNAs selected by OTA/LTA have a higher average on-target cutting efficiencies to edit 24 pairs of genes; and the sgRNAs selected by OTA/LTA have less number of off-target sites to edit 22023 pairs of genes. In terms of the PCS scores, the OTA/LTA strategies yield better scores on 37 pairs. Furthermore, three pairs of cancer genes can be edited by single sgRNAs with higher average on-target cutting efficiency (AOCE) and less off-target sites (OTS) than taking the OTO approach (see **Table 1**).

For example, three cancer associated genes SSX1, SSX2 and SSX4 can be edited by a single sgRNA containing the spacer sequence of 'ACTACGCATGAAAGGTGGGA'. It cuts

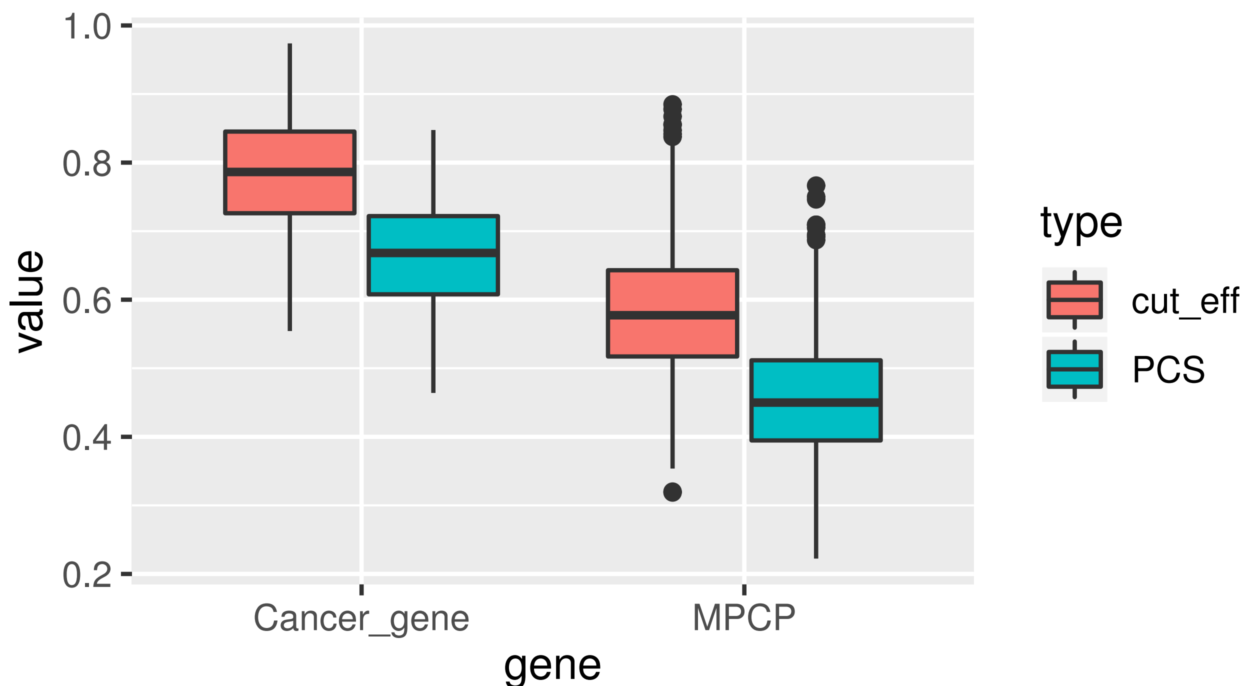

Fig. 3. A boxplot showing the cutting efficiencies and the PCS scores of the cutting sites for the 717 cancer genes and their MPCPs. "cut_eff" means the cutting efficiency.

Table 1. Three cancer gene pairs that can be edited by single sgRNAs with higher on-target cutting efficiency and less off-target sites.

Strategy	Cancer gene pair	sgRNA	AOCE	OTS	PCS
OTO	SSX1-SSX2	AGCCTGCCGAAAGTCATCTG, GCTATGCACCTGATGACGAG	0.8050	151	0.6756
OTA	SSX1-SSX2	AGCCTGCCGAAAGTCATCTG	0.8099	81	0.6688
OTO	SSX1-SSX4	AGCCTGCCGAAAGTCATCTG, GCTATGCACCTGATGACGAG	0.8050	151	0.6755
LTA	SSX1-SSX4	ACTACGCATGAAAGGTGGGA	0.8521	100	0.5989
OTO	SSX2-SSX4	GCTATGCACCTGATGACGAG, GCTATGCACCTGATGACGAG	0.8308	140	0.7095
LTA	SSX2-SSX4	ACTACGCATGAAAGGTGGGA	0.8521	100	0.5989

the three genes with on-target cutting efficiencies of 0.8947 (on SSX4), 0.8134 (on SSX1) and 0.8134 (on SSX2), while introducing 98 off-target sites. If we use the OTO strategy, three sgRNAs 'AGCCTGCCGAAAGTCATCTG' (SSX1), 'GCTATGCACCTGATGACGAG' (SSX2) and 'GCTATGCAC CTGATGACGAG' (SSX4) are required that have on-target editing efficiencies of 0.7807, 0.8308 and 0.8308 respectively, and there would be 221 off-target sites. Literature work has reported that human synovial sarcoma is caused by a chromosome translocation, which generates the SS18-SSX fusion protein by the fusion of one of the three SSX genes (SSX1, SSX2 and SSX4) with the SS18 gene [78]. Simultaneously editing the three SSX genes could help investigate the detailed mechanism of the human synovial sarcoma, and the sgRNA with the spacer sequence 'ACTACGCATGAAAGGTGGGA' is recommended useful for the investigations.

Editing five cancer associated genes TP53, BRCA2, BRCA1, PTEN, NF1 to generate the murine models of ovarian high-grade serous carcinoma

Five cancer associated genes TP53 (ensembl id: ENSG00000141510), BRCA2 (ensembl id: ENSG00000139618), BRCA1 (ensembl id: ENSG00000012048), PTEN (ensembl id: ENSG00000171862), and NF1 (ensembl id: ENSG00000196712) have been all found to be related to the ovarian high-grade serous carcinoma [8, 74]. Previous studies have knocked out some or all of the five genes to generate the murine models of ovarian high-grade serous carcinoma [84, 86, 85]. To test the performance of our sgRNA recommendation method, our tool is applied to select the sgRNAs for an optimal design of the CRISPR/Cas9 system editing the five genes simultaneously. Our method recommended three groups of spacers (see **Table 2**).

In the first group, three spacers are recommended to construct the simplest CRISPR/Cas9 system. The spacer 'GAGCACAGTAGAACTAAGGG' is predicted to target BRCA2 and NF1 with an on-target editing efficiencies 0.7351

Table 2. Our recommended spacers for editing the five cancer-associated genes.

spacer number	strategy	spacer	genes	OCS	OTS	PCS
3	LTA	GAGCACAGTAGAACTAAGGG; GGCTGGGAACGCCGGAGAGT; CCATTGTTCAATATCGTCCG.	BRCA2;NF1;PTEN; BRCA1;TP53	0.7351; 0.1519;0.5427; 0.1960;0.7919	690	0.1759
4	LTA	CTTACAGCAGTAGTATCATG; CATGACACCCCACCGGAAGT; GGCTGGGAACGCCGGAGAGT; CCATTGTTCAATATCGTCCG.	BRCA2;NF1;PTEN; BRCA1;TP53	0.7990;0.8371;0.5427; 0.1960;0.7919	322	0.3693
5	OTO	TCAGGATGAAGGCCTGATGT; CTTACAGCAGTAGTATCATG; CATGACACCCCACCGGAAGT; ACAGACTGATGTGTATACGT; CCATTGTTCAATATCGTCCG.	BRCA1;BRCA2;NF1; PTEN;TP53	0.8249;0.7990;0.8371; 0.8262;0.7919	256	0.6930

spacer means the selected spacer sequence; **OCS** is the on-target cutting efficiencies for the five genes (the same order as the genes)

Algorithm 2 Determine MPCP and MBOCS.

Require: gene sequence g_seq of candidate gene A;
1: Obtain the rank list of A's on-target cutting sites $CS(g_seq)'$ according to **Algorithm 1**;
2: $CS(g_seq)' = \{cs'_1, cs'_2, \cdots, cs'_i, \cdots, cs'_n\}$;
3: $CE(g_seq)' = \{ce'_1, ce'_2, \cdots, ce'_i, \cdots, ce'_n\}$;
4: $topN = int(n * 0.2)$
5: **for** t **in** 1 *to* $topN$ **do**
6: $OT(cs'_t) = \{ot_1, ot_2, \cdots, ot_k, \cdots, ot_K\}$;
7: **for** ot_k **in** $OT(cs'_t)$ **do**
8: Estimate the off-target site cutting efficiency at ot_k as $otce_k$;
9: Compute its PCS score:
10: $PCS(ot_k) = \frac{1}{K-1}\sum_{j=1\&j\neq k}^{K}(otce_k - otce_j) \cdot \frac{|otce_j < otce_k|}{K-1}$
11: Compute the average value of ot_k's cutting efficiency and PCS score as $avg_otce_pcs(ot_k) = 2 * otce_k * PCS(ot_k)/(otce_k + PCS(ot_k))$
12: **end for**
13: Obtain $ot_{k'}$, where it has the highest $avg_otce_pcs(ot_{k'})$ and we regard $ot_{k'}$ as the MBOCS of cs'_t
14: Compute the average cutting efficiency of cutting site pair $\langle cs'_t, ot_{k'}\rangle$ as $avgOn(\langle cs'_t, ot_{k'}\rangle) = 2*ce'_t*otce_{k'}/(ce'_t + otce_{k'})$
15: Compute the total number of off-target sites for cutting site pair $\langle cs'_t, ot_{k'}\rangle$ as $numOT(\langle cs'_t, ot_{k'}\rangle) = K - 1$
16: Compute the average PCS score for cutting site pair $\langle cs'_t, ot_{k'}\rangle$ as $avgPCS(\langle cs'_t, ot_{k'}\rangle) = 2 * PCS(cs'_t) * PCS(ot_{k'})/(PCS(cs'_t) + PCS(ot_{k'}))$
17: **end for**
18: Rank the cutting site pair $\langle cs'_t, ot_{k'}\rangle$:
19: $rank(\langle cs'_t, ot_{k'}\rangle) = (rank(avgOn(\langle cs'_t, ot_{k'}\rangle)) + rank(numOT(\langle cs'_t, ot_{k'}\rangle)) + rank(avgPCS(\langle cs'_t, ot_{k'}\rangle)))/3$
20: Annotate the $ot_{k'}$, obtain its host gene
21: **Output** The rank list of cutting cite pairs $\langle cs''_i, ot'_{k'}\rangle$, the rank first ot'_1 is the MBOCS and its host gene $G_{ot'_1}$ is the MPCP of A

and 0.1519 respectively. The spacer 'GGCTGGGAACGC CGGAGAGT' targets PTEN (0.5427) and BRCA1 (0.1960) and the third spacer 'CCATTGTTCAATATCGTCCG' (0.7919) binds to TP53. There are 690 potential off-target editing sites by these three sgRNAs and the resulting PCS score is 0.1759. In the second group having four spacers, the spacer 'GGCTGGGAACGCCGGAGAGT' is predicted to edit both PTEN and BRCA1. Each of the remaining three sgRNAs is to edit a different gene of the remaining three. With this sgRNA panel, the on-target cutting efficiency on NF1 is improved, the total number of off-target sites is decreased (to 322), and the PCS score is also improved (to 0.3693). Using the spacers recommended by the OTO strategy, each of the five genes is targeted by its own optimal spacer. The cutting efficiency and specificity are both improved. But the resulted CRISPR/Cas9 system becomes much more complex. These recommendations have their own advantages and shortcomings, end users are suggested to choose their favourable spacer group to carry out the multi-gene editing experiments.

Conclusion

We conducted a brief survey on the applications of multi-gene editing and commented on their methods. We found that bioinformatics to the system design of multi-gene editing is important to optimize the editing performance. So far, identification of beneficial off-target editing site has not been studied. As a novel perspective, we constructed a novel computational tool to help design CRISPR/Cas9 systems for multi-gene editing by considering beneficial off-target editing sites.

We adopted the idea of turning the harmful off-target editing into beneficial editing sites of additional genes of our interests. We searched those spacers that are designed for a given gene (on-target) but also targeting other genes (off-target), and evaluated their editing performance. The performance evaluation included the comparison of the on-target cutting efficiencies for all the genes, the total number of off-target sites and the preference cutting scores. The cutting efficiencies of the sgRNA cutting the additional genes are approximated by multiplying the on-target cutting efficiency at the given genes' editing site with the percent-activity values.

Three strategies (OTO, LTA and OTA) were applied for the selection of spacers for the multi-gene editing. The LTA and OTA strategies can design simpler CRISPR/Cas9 systems and sometimes can reduce the off-target site numbers. The OTO strategy is to select an optimal spacer to each of the target genes. This type of strategy can always achieve the highest on-target cutting efficiencies but sometimes introduce

more off-target sites; and the resulted CRISPR/Cas9 system is complex. These selected spacers are ranked according to three indices such as the average cutting efficiency, the total off-target site number and the PCS score. The top-ranked spacers are recommended for further wet lab validation. The developed tool has great potential to make contributions to the multi-gene editing studies and the related CRISPR/Cas9 system research.

Key Points

- Survey on the current wide-range applications of multi-gene editing in the fields of biomedicine and biology;
- Perspectives and discussions on a novel off-to-on role conversion idea to turn off-target editing sites into on-target editing sites such that the number of sgRNAs is minimized in the multi-gene editing, meanwhile the collective negative effects caused by the off-target editing are reduced;
- Preference cutting scores to identify and rank beneficial off-target editing sites for the recommendation of an optimised panel of sgRNAs in the multi-gene editing;
- Comparative case studies using three approaches to editing five cancer associated genes TP53, BRCA2, BRCA1, PTEN, NF1 to generate the murine models of ovarian high-grade serous carcinoma, and large scale evaluation on cancer-associated genes; and
- The source code, software tools and the bench mark data sets are public available.

Competing interests

There is NO Competing Interest.

Author contributions statement

H.P. and J.L. conceived the experiment(s), H.P. conducted the experiment(s), H.P., Y.Z., Z.Z. and J.L. analysed the results. H.P. and J.L. wrote the paper and all reviewed the manuscript.

Acknowledgments

This work is supported in part by funds from the Australian Research Council (ARC DP180100120).

References

1. Adames, Neil R and Gallegos, Jenna E and Peccoud, Jean. Yeast genetic interaction screens in the age of CRISPR/Cas. *Current Genetics*, 65(2):307–327, 2019.
2. Belén Adiego-Pérez, Paola Randazzo, Jean Marc Daran, René Verwaal, Johannes A Roubos, Pascale Daran-Lapujade, and John van der Oost. Multiplex genome editing of microorganisms using CRISPR-Cas. *FEMS Microbiology Letters*, 366(8):fnz086, 2019.
3. Bao, Zehua and Xiao, Han and Liang, Jing and Zhang, Lu and Xiong, Xiong and Sun, Ning and Si, Tong and Zhao, Huimin. Homology-integrated CRISPR–Cas (HI-CRISPR) system for one-step multigene disruption in Saccharomyces cerevisiae. *ACS Synthetic Biology*, 4(5):585–594, 2015.
4. Bradford, Jacob and Perrin, Dimitri. A benchmark of computational CRISPR-Cas9 guide design methods. *PLoS Computational Biology*, 15(8):e1007274, 2019.
5. Breunig, Christopher T and Durovic, Tamara and Neuner, Andrea M and Baumann, Valentin and Wiesbeck, Maximilian F and Köferle, Anna and Götz, Magdalena and Ninkovic, Jovica and Stricker, Stefan H. One step generation of customizable gRNA vectors for multiplex CRISPR approaches through string assembly gRNA cloning (STAgR). *PLoS One*, 13(4):e0196015, 2018.
6. Camerlengo, Francesco and Frittelli, Arianna and Sparks, Caroline and Doherty, Angela and Martignago, Damiano and Larré, Colette and Lupi, Roberta and Sestili, Francesco and Masci, Stefania. CRISPR-Cas9 multiplex editing of the α-amylase/trypsin inhibitor genes to reduce allergen proteins in durum wheat. *Frontiers in Sustainable Food Systems*, 4:104, 2020.
7. Campa, Carlo C and Weisbach, Niels R and Santinha, António J and Incarnato, Danny and Platt, Randall J. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. *Nature Methods*, 16(9):887–893, 2019.
8. Cancer Genome Atlas Research Network and others. Integrated genomic analyses of ovarian carcinoma. *Nature*, 474(7353):609–615, 2011.
9. Cao, Jian and Wu, Lizhen and Zhang, Shang-Min and Lu, Min and Cheung, William KC and Cai, Wesley and Gale, Molly and Xu, Qi and Yan, Qin. An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. *Nucleic Acids Research*, 44(19):e149–e149, 2016.
10. Zachary R Chalmers, Caitlin F Connelly, David Fabrizio, Laurie Gay, Siraj M Ali, Riley Ennis, Alexa Schrock, Brittany Campbell, Adam Shlien, Juliann Chmielecki, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Medicine*, 9(1):34, 2017.
11. Lynne Chantranupong, Sonia M Scaria, Robert A Saxton, Melanie P Gygi, Kuang Shen, Gregory A Wyant, Tim Wang, J Wade Harper, Steven P Gygi, and David M Sabatini. The CASTOR proteins are arginine sensors for the mTORC1 pathway. *Cell*, 165(1):153–164, 2016.
12. Van Trung Chu, Timm Weber, Benedikt Wefers, Wolfgang Wurst, Sandrine Sander, Klaus Rajewsky, and Ralf Kühn. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology*, 33(5):543–548, 2015.
13. Guohui Chuai, Hanhui Ma, Jifang Yan, Ming Chen, Nanfang Hong, Dongyu Xue, Chi Zhou, Chenyu Zhu, Ke Chen, Bin Duan, et al. DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome Biology*, 19(1):80, 2018.
14. Chuai, Guohui and Wang, Qilong and Liu, Qi. In silico meets in vivo: towards computational CRISPR-based sgRNA design. *Trends in biotechnology*, 35(1):12–21, 2017.
15. Cong, Le and Ran, F Ann and Cox, David and Lin, Shuailiang and Barretto, Robert and Habib, Naomi and Hsu, Patrick D and Wu, Xuebing and Jiang, Wenyan and Marraffini, Luciano A and others. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121):819–823, 2013.
16. Cui, Yingbo and Xu, Jiaming and Cheng, Minxia and Liao, Xiangke and Peng, Shaoliang. Review of CRISPR/Cas9 sgRNA design tools. *Interdisciplinary Sciences: Computational Life Sciences*, 10(2):455–465, 2018.
17. Ding, Wentao and Zhang, Yang and Shi, Shuobo. Development and Application of CRISPR/Cas in

Microbial Biotechnology. *Frontiers in Bioengineering and Biotechnology*, 8:711, 2020.

18. John G Doench, Nicolo Fusi, Meagan Sullender, Mudra Hegde, Emma W Vaimberg, Katherine F Donovan, Ian Smith, Zuzana Tothova, Craig Wilen, Robert Orchard, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, 34(2):184–191, 2016.
19. Dong, De and Ren, Kuan and Qiu, Xiaolin and Zheng, Jianlin and Guo, Minghui and Guan, Xiaoyu and Liu, Hongnan and Li, Ningning and Zhang, Bailing and Yang, Daijun and others. The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature*, 532(7600):522–526, 2016.
20. Dong, Hongzhi and Zheng, Junwei and Yu, Dou and Wang, Bin and Pan, Li. Efficient genome editing in Aspergillus niger with an improved recyclable CRISPR-HDR toolbox and its application in introducing multiple copies of heterologous genes. *Journal of Microbiological Methods*, 163:105655, 2019.
21. Dong, Xiaoyun and Jin, Yingli and Ming, Di and Li, Bangbang and Dong, Haisi and Wang, Lin and Wang, Tiedong and Wang, Dacheng. CRISPR/dCas9-mediated inhibition of gene expression in Staphylococcus aureus. *Journal of Microbiological Methods*, 139:79–86, 2017.
22. Jennifer A Doudna and Emmanuelle Charpentier. The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213):1258096, 2014.
23. Eisenhut, Peter and Klanert, Gerald and Weinguny, Marcus and Baier, Laurenz and Jadhav, Vaibhav and Ivansson, Daniel and Borth, Nicole. A CRISPR/Cas9 based engineering strategy for overexpression of multiple genes in Chinese hamster ovary cells. *Metabolic Engineering*, 48:72–81, 2018.
24. Masaki Endo, Masafumi Mikami, and Seiichi Toki. Multigene knockout utilizing off-target mutations of the CRISPR/Cas9 system in rice. *Plant and Cell Physiology*, 56(1):41–47, 2015.
25. Feng, Zhengyan and Zhang, Zhengjing and Hua, Kai and Gao, Xifeng and Mao, Yanfei and Botella, Jose Ramon and Zhu, Jian-Kang. A highly efficient cell division-specific CRISPR/Cas9 system generates homozygous mutants for multiple genes in Arabidopsis. *International Journal of Molecular Sciences*, 19(12):3925, 2018.
26. Ferreira, Raphael and Skrekas, Christos and Nielsen, Jens and David, Florian. Multiplexed CRISPR/Cas9 genome editing and gene regulation using Csy4 in Saccharomyces cerevisiae. *ACS Synthetic Biology*, 7(1):10–15, 2018.
27. David A Fruman, Honyin Chiu, Benjamin D Hopkins, Shubha Bagrodia, Lewis C Cantley, and Robert T Abraham. The PI3K pathway in human disease. *Cell*, 170(4):605–635, 2017.
28. Fu, Becky Xu Hua and Smith, Justin D and Fuchs, Ryan T and Mabuchi, Megumu and Curcuru, Jennifer and Robb, G Brett and Fire, Andrew Z. Target-dependent nickase activities of the CRISPR–Cas nucleases Cpf1 and Cas9. *Nature Microbiology*, 4(5):888–897, 2019.
29. Caixia Gao. The future of CRISPR technologies in agriculture. *Nature Reviews Molecular Cell Biology*, 19(5):275–276, 2018.
30. Ryosuke Hashimoto, Risa Ueta, Chihiro Abe, Yuriko Osakabe, and Keishi Osakabe. Efficient multiplex genome editing induces precise, and self-ligated type mutations in tomato plants. *Frontiers in Plant Science*, 9:916, 2018.
31. Ayal Hendel, Rasmus O Bak, Joseph T Clark, Andrew B Kennedy, Daniel E Ryan, Subhadeep Roy, Israel Steinfeld, Benjamin D Lunstad, Robert J Kaiser, Alec B Wilkens, et al. Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nature Biotechnology*, 33(9):985–989, 2015.
32. Horwitz, Andrew A and Walter, Jessica M and Schubert, Max G and Kung, Stephanie H and Hawkins, Kristy and Platt, Darren M and Hernday, Aaron D and Mahatdejkul-Meadows, Tina and Szeto, Wayne and Chandran, Sunil S and others. Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas. *Cell Systems*, 1(1):88–96, 2015.
33. Hsu, Patrick D and Lander, Eric S and Zhang, Feng. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157(6):1262–1278, 2014.
34. Nan Hu, Zhiqiang Xian, Ning Li, Yudong Liu, Wei Huang, Fang Yan, Deding Su, Jingxuan Chen, and Zhengguo Li. Rapid and user-friendly open-source CRISPR/Cas9 system for single-or multi-site editing of tomato genome. *Horticulture Research*, 6(1):7, 2019.
35. Huang, Manning Y and Woolford, Carol A and Mitchell, Aaron P. Rapid Gene Concatenation for Genetic Rescue of Multigene Mutants in Candida albicans. *Msphere*, 3(2), 2018.
36. Jao, Li-En and Wente, Susan R and Chen, Wenbiao. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proceedings of the National Academy of Sciences*, 110(34):13904–13909, 2013.
37. Yu Jiang, Biao Chen, Chunlan Duan, Bingbing Sun, Junjie Yang, and Sheng Yang. Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. *Applied and Environmental Microbiology*, 81(7):2506–2514, 2015.
38. Jiang, Fuguo and Doudna, Jennifer A. CRISPR–Cas9 structures and mechanisms. *Annual Review of Biophysics*, 46:505–529, 2017.
39. Jiang, Fuguo and Taylor, David W and Chen, Janice S and Kornfeld, Jack E and Zhou, Kaihong and Thompson, Aubri J and Nogales, Eva and Doudna, Jennifer A. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*, 351(6275):867–871, 2016.
40. Yongsub Kim, Seung-A Cheong, Jong Geol Lee, Sang-Wook Lee, Myeong Sup Lee, In-Jeoung Baek, and Young Hoon Sung. Generation of knockout mice by Cpf1-mediated gene targeting. *Nature Biotechnology*, 34(8):808–810, 2016.
41. Benjamin P Kleinstiver, Michelle S Prew, Shengdar Q Tsai, Nhu T Nguyen, Ved V Topkar, Zongli Zheng, and J Keith Joung. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nature Biotechnology*, 33(12):1293–1298, 2015.
42. Benjamin P Kleinstiver, Shengdar Q Tsai, Michelle S Prew, Nhu T Nguyen, Moira M Welch, Jose M Lopez, Zachary R McCaw, Martin J Aryee, and J Keith Joung. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nature Biotechnology*, 34(8):869–874, 2016.
43. Mo Li and Juan Carlos Izpisua Belmonte. Deconstructing the pluripotency gene regulatory network. *Nature Cell Biology*, 20(4):382–392, 2018.
44. Li, Wei and Teng, Fei and Li, Tianda and Zhou, Qi. Simultaneous generation and germline transmission

of multiple gene mutations in rat using CRISPR-Cas systems. *Nature Biotechnology*, 31(8):684–686, 2013.

45. Li, Xiaolin and Liu, Qian and Sun, Wenliang and He, Qun and Tian, Chaoguang. Improving cellulases production by Myceliophthora thermophila through disruption of protease genes. *Biotechnology Letters*, 42(2):219–229, 2020.
46. Li, Xindi and Wang, Yanning and Chen, Sha and Tian, Huiqin and Fu, Daqi and Zhu, Benzhong and Luo, Yunbo and Zhu, Hongliang. Lycopene is enriched in tomato fruit by CRISPR/Cas9-mediated multiplex genome editing. *Frontiers in Plant Science*, 9:559, 2018.
47. Jennifer Listgarten, Michael Weinstein, Benjamin P Kleinstiver, Alexander A Sousa, J Keith Joung, Jake Crawford, Kevin Gao, Luong Hoang, Melih Elibol, John G Doench, et al. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nature Biomedical Engineering*, 2(1):38–47, 2018.
48. Liu, Dingyu and Huang, Can and Guo, Jiaxin and Zhang, Peiji and Chen, Tao and Wang, Zhiwen and Zhao, Xueming. Development and characterization of a CRISPR/Cas9n-based multiplex genome editing system for Bacillus subtilis. *Biotechnology for Biofuels*, 12(1):197, 2019.
49. Liu, Huan and Sui, Tingting and Liu, Di and Liu, Tingjun and Chen, Mao and Deng, Jichao and Xu, Yuanyuan and Li, Zhanjun. Multiple homologous genes knockout (KO) by CRISPR/Cas9 system in rabbit. *Gene*, 647:261–267, 2018.
50. Liu, Qi and Shi, Xiaona and Song, Lili and Liu, Haifeng and Zhou, Xiangshan and Wang, Qiyao and Zhang, Yuanxing and Cai, Menghao. CRISPR–Cas9-mediated genomic multiloci integration in Pichia pastoris. *Microbial Cell Factories*, 18(1):1–11, 2019.
51. Lowder, Levi G and Zhou, Jianping and Zhang, Yingxiao and Malzahn, Aimee and Zhong, Zhaohui and Hsieh, Tzung-Fu and Voytas, Daniel F and Zhang, Yong and Qi, Yiping. Robust transcriptional activation in plants using multiplexed CRISPR-Act2. 0 and mTALE-Act systems. *Molecular Plant*, 11(2):245–256, 2018.
52. Cunfa Ma, Chenzeng Zhu, Min Zheng, Mengci Liu, Dejun Zhang, Baoli Liu, Qinfei Li, Jun Si, Xuesong Ren, and Hongyuan Song. CRISPR/Cas9-mediated multiple gene editing in Brassica oleracea var. capitata using the endogenous tRNA-processing system. *Horticulture Research*, 6(1):20, 2019.
53. McCarty, Nicholas S and Graham, Alicia E and Studená, Lucie and Ledesma-Amaro, Rodrigo. Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nature Communications*, 11(1):1–13, 2020.
54. Merenda, Alessandra and Andersson-Rolf, Amanda and Mustata, Roxana C and Li, Taibo and Kim, Hyunki and Koo, Bon-Kyoung. A protocol for multiple gene knockout in mouse small intestinal organoids using a CRISPR-concatemer. *Journal of Visualized Experiments*, (125):e55916, 2017.
55. Minkenberg, Bastian and Wheatley, Matthew and Yang, Yinong. CRISPR/Cas9-enabled multiplex genome editing and its application. In *Progress in Molecular Biology and Translational Science*, volume 149, pages 111–132. Elsevier, 2017.
56. Ana M Moreno, Xin Fu, Jie Zhu, Dhruva Katrekar, Yu-Ru V Shih, John Marlett, Jessica Cabotaje, Jasmine Tat, John Naughton, Leszek Lisowski, et al. In situ gene therapy via AAV-CRISPR-Cas9-mediated targeted gene regulation. *Molecular Therapy*, 26(7):1818–1827, 2018.
57. Nissim, Lior and Perli, Samuel D and Fridkin, Alexandra and Perez-Pinera, Pablo and Lu, Timothy K. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Molecular Cell*, 54(4):698–710, 2014.
58. Joana Osório. Gene regulation: landscape and mechanisms of transcription factor cooperativity. *Nature Reviews. Genetics*, 17(1):5, 2016.
59. Ota, Satoshi and Hisano, Yu and Ikawa, Yoshiya and Kawahara, Atsuo. Multiple genome modifications by the CRISPR/Cas9 system in zebrafish. *Genes to Cells*, 19(7):555–564, 2014.
60. Peter B Otoupal, Masakazu Ito, Adam P Arkin, Jon K Magnuson, John M Gladden, and Jeffrey M Skerker. Multiplexed CRISPR-Cas9-Based Genome Editing of Rhodosporidium toruloides. *Msphere*, 4(2):e00099–19, 2019.
61. Hui Peng, Yi Zheng, Michael Blumenstein, Dacheng Tao, and Jinyan Li. CRISPR/Cas9 cleavage efficiency regression through boosting algorithms and Markov sequence profiling. *Bioinformatics*, 34(18):3069–3077, 2018.
62. Hui Peng, Yi Zheng, Zhixun Zhao, Tao Liu, and Jinyan Li. Recognition of CRISPR/Cas9 off-target sites through ensemble learning of uneven mismatch distributions. *Bioinformatics*, 34(17):i757–i765, 2018.
63. Prykhozhij, Sergey V and Rajan, Vinothkumar and Gaston, Daniel and Berman, Jason N. CRISPR multitargeter: a web tool to find common and unique CRISPR single guide RNA targets in a set of similar sequences. *PLoS One*, 10(3):e0119372, 2015.
64. Weiwei Qi, Tong Zhu, Zhongrui Tian, Chaobin Li, Wei Zhang, and Rentao Song. High-efficiency CRISPR/Cas9 multiplex gene editing using the glycine tRNA-processing system-based strategy in maize. *BMC Biotechnology*, 16(1):58, 2016.
65. Md Khaledur Rahman and M Sohel Rahman. CRISPRpred: A flexible and efficient tool for sgRNAs on-target activity prediction in CRISPR/Cas9 systems. *PLoS One*, 12(8):e0181943, 2017.
66. Raitskin, Oleg and Patron, Nicola J. Multi-gene engineering in plants with RNA-guided Cas9 nuclease. *Current Opinion in Biotechnology*, 37:69–75, 2016.
67. Ramanagoudr-Bhojappa, Ramanagouda and Carrington, Blake and Ramaswami, Mukundhan and Bishop, Kevin and Robbins, Gabrielle M and Jones, MaryPat and Harper, Ursula and Frederickson, Stephen C and Kimble, Danielle C and Sood, Raman and others. Multiplexed CRISPR/Cas9-mediated knockout of 19 Fanconi anemia pathway genes in zebrafish revealed their roles in growth, sexual development and fertility. *PLoS Genetics*, 14(12):e1007821, 2018.
68. Ran, F Ann and Hsu, Patrick D and Lin, Chie-Yu and Gootenberg, Jonathan S and Konermann, Silvana and Trevino, Alexandro E and Scott, David A and Inoue, Azusa and Matoba, Shogo and Zhang, Yi and others. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154(6):1380–1389, 2013.
69. Reis, Alexander C and Halper, Sean M and Vezeau, Grace E and Cetnar, Daniel P and Hossain, Ayaan and Clauer, Phillip R and Salis, Howard M. Simultaneous repression

of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays. *Nature Biotechnology*, 37(11):1294–1301, 2019.

70. Jiangtao Ren, Xiaojun Liu, Chongyun Fang, Shuguang Jiang, Carl H June, and Yangbing Zhao. Multiplex genome editing to generate universal CAR T cells resistant to PD1 inhibition. *Clinical Cancer Research*, 23(9):2255–2266, 2017.
71. Roux, Indra and Woodcraft, Clara and Hu, Jinyu and Wolters, Rebecca and Gilchrist, Cameron LM and Chooi, Yit-Heng. CRISPR-mediated activation of biosynthetic gene clusters for bioactive molecule discovery in filamentous fungi. *ACS Synthetic Biology*, 2020.
72. Ryan, Owen W and Skerker, Jeffrey M and Maurer, Matthew J and Li, Xin and Tsai, Jordan C and Poddar, Snigdha and Lee, Michael E and DeLoache, Will and Dueber, John E and Arkin, Adam P and others. Selection of chromosomal DNA libraries using a multiplex CRISPR system. *Elife*, 3:e03703, 2014.
73. Schmieder, Valerie and Bydlinski, Nina and Strasser, Richard and Baumann, Martina and Kildegaard, Helene Faustrup and Jadhav, Vaibhav and Borth, Nicole. Enhanced genome editing tools for multi-gene deletion knock-out approaches using paired CRISPR sgRNAs in CHO cells. *Biotechnology Journal*, 13(3):1700211, 2018.
74. Roland F Schwarz, Charlotte KY Ng, Susanna L Cooke, Scott Newman, Jillian Temple, Anna M Piskorz, Davina Gale, Karen Sayal, Muhammed Murtaza, Peter J Baldwin, et al. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. *PLoS Medicine*, 12(2):e1001789, 2015.
75. Zbyslaw Sondka, Sally Bamford, Charlotte G Cole, Sari A Ward, Ian Dunham, and Simon A Forbes. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nature Reviews Cancer*, 18(11):696–705, 2018.
76. Qinfu Sun, Li Lin, Dongxiao Liu, Dewei Wu, Yujie Fang, Jian Wu, and Youping Wang. CRISPR/Cas9-Mediated Multiplex Genome Editing of the BnWRKY11 and BnWRKY70 Genes in Brassica napus L. *International Journal of Molecular Sciences*, 19(9):2716, 2018.
77. Sundaresan, Ramya and Parameshwaran, Hari Priya and Yogesha, SD and Keilbarth, Mark Walter and Rajan, Rakhi. RNA-independent DNA cleavage activities of Cas9 and Cas12a. *Cell Reports*, 21(13):3728–3739, 2017.
78. Svejstrup, Jesper Q. Synovial sarcoma mechanisms: a series of unfortunate events. *Cell*, 153(1):11–12, 2013.
79. Swarts, Daan C and Jinek, Martin. Cas9 versus Cas12a/Cpf1: Structure–function comparisons and implications for genome editing. *Wiley Interdisciplinary Reviews: RNA*, 9(5):e1481, 2018.
80. Collin J Tokheim, Nickolas Papadopoulos, Kenneth W Kinzler, Bert Vogelstein, and Rachel Karchin. Evaluating the evaluation of cancer driver genes. *Proceedings of the National Academy of Sciences*, 113(50):14330–14335, 2016.
81. Vazquez-Vilar, Marta and Bernabé-Orts, Joan Miquel and Fernandez-del-Carmen, Asun and Ziarsolo, Pello and Blanca, Jose and Granell, Antonio and Orzaez, Diego. A modular toolbox for gRNA–Cas9 genome engineering in plants based on the GoldenBraid standard. *Plant Methods*, 12(1):1–12, 2016.
82. Voloshanenko, Oksana and Gmach, Philipp and Winter, Jan and Kranz, Dominique and Boutros, Michael. Mapping of Wnt-Frizzled interactions by multiplex CRISPR targeting of receptor gene families. *The FASEB Journal*, 31(11):4832–4844, 2017.
83. Vyas, Valmik K and Barrasa, M Inmaculada and Fink, Gerald R. A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families. *Science Advances*, 1(3):e1500248, 2015.
84. Walton, Josephine and Blagih, Julianna and Ennis, Darren and Leung, Elaine and Dowson, Suzanne and Farquharson, Malcolm and Tookman, Laura A and Orange, Clare and Athineos, Dimitris and Mason, Susan and others. CRISPR/Cas9-mediated Trp53 and Brca2 knockout to generate improved murine models of ovarian high-grade serous carcinoma. *Cancer Research*, 76(20):6118–6129, 2016.
85. Walton, Josephine and Farquharson, Malcolm and Mason, Susan and Blagih, Julianna and Ennis, Darren and Leung, Elaine and Dowson, Suzanne and Athineos, Dimitris and Rami, Damiano and Stevenson, David and others. Abstract PR14: CRISPR/Cas9-mediated Trp53, Brca1, Brca2, Pten, and Nf1 knockout to generate improved murine models of ovarian high-grade serous carcinoma. In *CLINICAL CANCER RESEARCH*, volume 24, pages 36–37. AMER ASSOC CANCER RESEARCH 615 CHESTNUT ST, 17TH FLOOR, PHILADELPHIA, PA . . . , 2018.
86. Walton, Josephine B and Farquharson, Malcolm and Mason, Susan and Port, Jennifer and Kruspig, Bjorn and Dowson, Suzanne and Stevenson, David and Murphy, Daniel and Matzuk, Martin and Kim, Jaeyeon and others. CRISPR/Cas9-derived models of ovarian high grade serous carcinoma targeting Brca1, Pten and Nf1, and correlation with platinum sensitivity. *Scientific Reports*, 7(1):16827, 2017.
87. Laiyou Wang, Aihua Deng, Yun Zhang, Shuwen Liu, Yong Liang, Hua Bai, Di Cui, Qidi Qiu, Xiuling Shang, Zhao Yang, et al. Efficient CRISPR–Cas9 mediated multiplex genome editing in yeasts. *Biotechnology for Biofuels*, 11(1):277, 2018.
88. Mugui Wang, Yanfei Mao, Yuming Lu, Zhidan Wang, Xiaoping Tao, and Jian-Kang Zhu. Multiplex gene editing in rice with simplified CRISPR-Cpf1 and CRISPR-Cas9 systems. *Journal of Integrative Plant Biology*, 60(8):626–631, 2018.
89. Wei Wang, Qianli Pan, Fei He, Alina Akhunova, Shiaoman Chao, Harold Trick, and Eduard Akhunov. Transgenerational CRISPR-Cas9 activity facilitates multiplex gene editing in allopolyploid wheat. *The CRISPR Journal*, 1(1):65–74, 2018.
90. Xiaolong Wang, Jing Liu, Yiyuan Niu, Yan Li, Shiwei Zhou, Chao Li, Baohua Ma, Qifang Kou, Bjoern Petersen, Tad Sonstegard, et al. Low incidence of SNVs and indels in trio genomes of Cas9-mediated multiplex edited sheep. *BMC Genomics*, 19(1):397, 2018.
91. Wang, Fujun and Wang, Chunlian and Liu, Piqing and Lei, Cailin and Hao, Wei and Gao, Ying and Liu, Yao-Guang and Zhao, Kaijun. Enhanced rice blast resistance by CRISPR/Cas9-targeted mutagenesis of the ERF transcription factor gene OsERF922. *PLoS One*, 11(4):e0154027, 2016.
92. Wang, Mugui and Mao, Yanfei and Lu, Yuming and Tao, Xiaoping and Zhu, Jian-kang. Multiplex gene editing in rice using the CRISPR-Cpf1 system. *Molecular Plant*, 10(7):1011–1013, 2017.

93. Wang, Zhi-Ping and Xing, Hui-Li and Dong, Li and Zhang, Hai-Yan and Han, Chun-Yan and Wang, Xue-Chen and Chen, Qi-Jun. Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biology*, 16(1):144, 2015.
94. Westbrook, Adam W and Moo-Young, Murray and Chou, C Perry. Development of a CRISPR-Cas9 tool kit for comprehensive engineering of Bacillus subtilis. *Applied and Environmental Microbiology*, 82(16):4876–4895, 2016.
95. Wu, Yaokang and Liu, Yanfeng and Lv, Xueqin and Li, Jianghua and Du, Guocheng and Liu, Long. CAMERS-B: CRISPR/Cpf1 assisted multiple-genes editing and regulation system for Bacillus subtilis. *Biotechnology and Bioengineering*, 117(6):1817–1825, 2020.
96. Kabin Xie, Bastian Minkenberg, and Yinong Yang. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences*, 112(11):3570–3575, 2015.
97. Xing, Hui-Li and Dong, Li and Wang, Zhi-Ping and Zhang, Hai-Yan and Han, Chun-Yan and Liu, Bing and Wang, Xue-Chen and Chen, Qi-Jun. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biology*, 14(1):327, 2014.
98. Yamano, Takashi and Zetsche, Bernd and Ishitani, Ryuichiro and Zhang, Feng and Nishimasu, Hiroshi and Nureki, Osamu. Structural basis for the canonical and non-canonical PAM recognition by CRISPR-Cpf1. *Molecular Cell*, 67(4):633–645, 2017.
99. Yan, Quanmei and Zhang, Quanjun and Yang, Huaqiang and Zou, Qingjian and Tang, Chengcheng and Fan, Nana and Lai, Liangxue. Generation of multi-gene knockout rabbits using the Cas9/gRNA system. *Cell Regeneration*, 3(1):12, 2014.
100. Yang, Yankun and Liu, Guoqiang and Chen, Xiao and Liu, Meng and Zhan, Chunjun and Liu, Xiuxia and Bai, Zhonghu. High efficiency CRISPR/Cas9 genome editing system with an eliminable episomal sgRNA plasmid in Pichia pastoris. *Enzyme and Microbial Technology*, page 109556, 2020.
101. Yoshioka, Shin and Fujii, Wataru and Ogawa, Tetsuhiro and Sugiura, Koji and Naito, Kunihiko. Development of a mono-promoter-driven CRISPR/Cas9 system in mammalian cells. *Scientific Reports*, 5(1):1–8, 2015.
102. Zhiming Yu, Qiyuan Chen, Weiwei Chen, Xian Zhang, Fengling Mei, Pengcheng Zhang, Mei Zhao, Xiaohong Wang, Nongnong Shi, Stephen Jackson, et al. Multigene editing via CRISPR/Cas9 guided by a single-sgRNA seed in Arabidopsis. *Journal of Integrative Plant Biology*, 60(5):376–381, 2018.
103. Daniel R Zerbino, Premanand Achuthan, Wasiu Akanni, M Ridwan Amode, Daniel Barrell, Jyothish Bhai, Konstantinos Billis, Carla Cummins, Astrid Gall, Carlos García Girón, et al. Ensembl 2018. *Nucleic Acids Research*, 46(D1):D754–D761, 2018.
104. Zetsche, Bernd and Heidenreich, Matthias and Mohanraju, Prarthana and Fedorova, Iana and Kneppers, Jeroen and DeGennaro, Ellen M and Winblad, Nerges and Choudhury, Sourav R and Abudayyeh, Omar O and Gootenberg, Jonathan S and others. Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nature Biotechnology*, 35(1):31–34, 2017.
105. Yueping Zhang, Juan Wang, Zibai Wang, Yiming Zhang, Shuobo Shi, Jens Nielsen, and Zihe Liu. A gRNA-tRNA array for CRISPR-Cas9 based rapid multiplexed genome editing in Saccharomyces cerevisiae. *Nature Communications*, 10(1):1–10, 2019.
106. Zhengjing Zhang, Yanfei Mao, Si Ha, Wenshan Liu, Jose Ramon Botella, and Jian-Kang Zhu. A multiplex CRISPR/Cas9 platform for fast and efficient editing of multiple genes in Arabidopsis. *Plant Cell Reports*, 35(7):1519–1533, 2016.
107. Zhang, He and Pan, Hong and Zhou, Changyang and Wei, Yu and Ying, Wenqin and Li, Shuting and Wang, Guangqin and Li, Chao and Ren, Yifei and Li, Gen and others. Simultaneous zygotic inactivation of multiple genes in mouse through CRISPR/Cas9-mediated base editing. *Development*, 145(20), 2018.
108. Zhang, Kang and Duan, Xuguo and Wu, Jing. Multigene disruption in undomesticated Bacillus subtilis ATCC 6051a using the CRISPR/Cas9 system. *Scientific Reports*, 6(1):1–11, 2016.
109. Zhou, Yuexin and Zhang, Hongmin and Wei, Wensheng. Simultaneous generation of multi-gene knockouts in human cells. *FEBS Letters*, 590(23):4343–4353, 2016.

Hui Peng. Hui is currently working as a research fellow at National University of Singapore. His research interests include bioinformatics and data analysis.

Yi Zheng. Yi is an engineer in the company AllDimensions, Chengdu, China. His research interests include data mining and medical informatics.

Zhixun Zhao. Zhixun is currently pursuing the Ph.D. degree in data analytics and bioinformatics with the Advanced Analytics Institute, University of Technology Sydney, Australia. His research interests include data mining, machine learning, and their applications in bioinformatics.

Jinyan Li. Jinyan is a Professor of Data Science in the Faculty of Engineering and IT in the University of Technology Sydney. His research is on bioinformatics and data mining algorithms.