CHINESE JOURNAL OF ANALYTICAL CHEMISTRY
Volume 48, Issue 1, January 2020
Online English edition of the Chinese language journal

Available online at www.sciencedirect.com
ScienceDirect

Cite this article as: Chinese J. Anal. Chem., 2020, 48(1): 1–12

REVIEW

Enrichment Reporter System of Genome Editing Positive Cells

ZHANG Li-Xian[1], WANG Qi-Fei[1], CAO Yang[1], YE Cheng-Yu[1], CHEN Zhi-Yang[1], XU Qi-Jie[1], ZOU Bing-Jie[2], SONG Qin-Xin[1,2,*], ZHOU Guo-Hua[1,2]

[1] Key Laboratory of Drug Quality Control and Pharmacovigilance of Ministry of Education, China Pharmaceutical University, Nanjing 210009, China
[2] Department of Pharmacology, Jinling Hospital, Medical School of Nanjing University, Nanjing 210002, China

Abstract: Genome editing has become a vital tool in medical biology research. The critical mission of facilitating the progress of genome editing is to enrich different genes into positive edited cells quickly and effectively and carry out the targeted research. In recent years, researchers have established various reporter systems for selection and enrichment of editing-induced positive cells, which are based on genome repair mechanisms, such as non-homologous end joining, homology-directed repair, single strand annealing and inversion, and the principle of the expression of fluorescent protein or resistance tag after genome repair. The T7E1 assay or sequencing method can analyze the mutation of enriched cells with the results of lower background signals and higher mutation ratio. Therefore, these reporter systems can profit the characterization of genome editing effectiveness. Besides, positive cells can be cultured continuously, so this technology possesses a promising prospect in mutated cell line construction and the research of mutated cell functions. This article summarized the design principles and applications of these reporter systems and would provide a reference to construct a more perfecte evaluating system for genome editing.

Key Words: Genome editing; Repair mechanism; Cell enrichment; Reporter system; Review

1 Introduction

Genetic engineering started when monkey virus SV40 and *E. coli* bacteriophage λDNA spliced into a loop for the first time[1]. Subsequently, the discovery and research of genome repair mechanisms such as homologous recombination (HR), non-homologous end joining (NHEJ), and endonucleases provided a theoretical basis for modern genome editing technology[2,3]. In recent years, researchers have found some nucleases based on long sequence recognition, such as meganuclease[4–7] or homing endonuclease, zinc finger nucleases (ZFNs)[8–12], transcription activator-like effector nucleases (TALENs)[13–17], structure-guided nucleases (SGN)[18] and clustered regularly interspaced short palindromic repeats associated protein Cas9 (CRISPR/Cas9)[19–22]. These nucleases can accurately recognize the target sequence and induce double-strand break (DSB), and then through homology-directed repair (HDR)[23] or NHEJ[24] and other mechanisms to edit the target sequence. Genome editing technology has been applied in various fields, such as the establishment of disease models[26–28], gene therapy[29], and gene screening[30] through different transfection techniques, such as liposomes, electroporation, viral packaging, and transmembrane peptides[25]. However, the off-target effects inherent in genome editing tools[31,32] may induce side effects such as non-target editing, erroneous phenotypes, or lethality in practical applications. The T7E1 digestion test and the amplicon sequencing analysis of the target fragment can only reflect local off-target effects, so deep sequencing[33] for the whole genome can be able to evaluate off-target effects

Received 6 June 2019; accepted 8 November 2019
*Corresponding author. Email: songqinxin@cpu.edu.cn
This work was supported by the National Natural Science Foundation of China (No. 81673390), the Jiangsu Provincial Key Research and Development Program of China (No. BE2016745), the Jiangsu Provincial Science Fund for Distinguished Young Scholars of China (No. BK20180005), the Open Project Program of MOE Key Laboratory of Drug Quality Control and Pharmacovigilance of China (No. DQCP2017MS01), the College Students Innovation Project for the R&D of Novel Drugs, China (No. J1310032), and the Jiangsu Provincial Qing Lan Project of China.

DOI: 10.1016/S1872-2040(19)61206-5

reasonably. In terms of reducing off-target effects, LNA-replaced[34], chemically synthesized base-substituted sgRNAs such as 2'-F, 2'-*O*-methyl, 2'-*O*-methyl-3'-phosphorothioate and 2'-*O*-methyl-3'-thiophosphonoacetate[35,36] significantly reduced the off-target effect of Cas9 and improved editing efficiency; the amino acid sequence in Cas9 protein was replaced to generate a mutant, called Cas9x, with more specificity or wider recognition range[37].

Different cell lines, loci, and genome editing tools will significantly affect editing efficiency, resulting in fewer genome editing positive cells, and a large number of wild-type cells will interfere in mutation analysis. Although high-throughput deep sequencing can resolve mutant sequences, the cost is high, data analysis is relatively difficult, and subsequent cell culture studies cannot be performed. In general, the T7E1 digestion test can be used as an early evaluation scheme for DNA level editing efficiency to screen optimal reaction system. However, to continue studying living cells with mutated genomes, a method that can enrich positive cells intuitively, quickly, and efficiently is needed. Currently, genome-editing tools are used to target exogenous target genes integrated with fluorescent proteins (such as mRFP, mCherry, DsRed, AsRED and eGFP) or resistance proteins (such as anti-puromycin and anti-hygromycin). Relevant methods to characterize genome editing positive results by using functional mechanisms such as HDR, NHEJ, single-strand annealing (SSA)[38], and inversion rearrangement to induce reading frame shifts for achieving functional protein expression were successively reported. Subsequent efficient enrichment of genome editing-positive cells was accomplished employing fluorescence-activated cell sorting (FACS) or resistance screening, which has the advantages of strong versatility, rapid detection, and simple analysis. These systems have been applied to validation of various genome-editing tools (Table 1). This article reviews the reporter systems based on the four-genome repair mechanisms (NHEJ, SSA, HDR and inversion rearrangement) and their applications in genome editing positive cell sorting and enrichment.

Table 1 Comparison of reporter systems for enrichment

Name	Name	Mechanism	Site	Tool	Fold/Ratio	References
pRTG	Fluorescence-fluorescence	NHEJ*	TP53	ZFN	5%[a], 13[b], 20[c]	[39]
pRTGG	Fluorescence-fluorescence	NHEJ*	BMP1	Cas9	9.4%[a], 18[b], 7.4[c]	[40]
			F13A1	Cas9	8.2%[a], 7.8[b], 10[c]	
			PPP1R1B	Cas9	7.8%[a], 4.6[b], 4.3[c]	
pRTKG	Fluorescence-magnetism	NHEJ*	CCR5	ZFN	12[b], 21[c]	[41]
			TP53	ZFN	17[b]	
			CD81	ZFN	17[b]	
			BRCA1	TALEN	17[b]	
pRTHG	Fluorescence-resistance	NHEJ*	CCR5	ZFN	16[b], 8.5[c]	[41]
			BRCA1	TALEN	15[b]	
LD	Resistance integration	NHEJ*	HSPA1A	Cas9	5.5[b]	[42]
			HSPA1B	Cas9	6.1[b]	
			HSPA1L	Cas9	3.4[b]	
			HSPA6	Cas9	6.6[b]	
pRGZG	Fluorescence-fluorescence	SSA*	MSTN	ZFN	18.33%[a]	[43]
C-Check	Fluorescence-fluorescence	SSA*	IAPP	TALEN	80%[a], 2.99%[b]	[44]
			MAPT	Cas9	10%–20%[a]	
			SORL1	Cas9	>15%[a]	
			IGF1R	Cas9	97.9%[d]	
			CBX5	Cas9	86.9%[d]	
pSCR	Fluorescence-fluorescence	SSA*	DAZL	Cas9	24.79%[a], 4[b]	[45]
			PLZF	Cas9	54.32%[b]	
			ACR	Cas9	40.49%[b]	
pSSA-RPG	Dual-fluorescence-resistance	SSA*	AAVS1	Cas9	34.8[b], 21.1[c]	[46]
			CCR2	Cas9	13.18[b], 34.6[c]	
			CCR5a	Cas9	12.3–13.2[b], 11.2–18.1[c]	
			CCR5b	Cas9	6.3[b], 11.8[c]	
			CCR5	ZFN	27.7–34.2[b], 14.9–18.8[c]	
Rep/Don	Fluorescence-resistance	SSA+HDR*	CCR5	Cas9	34.09%[d]	[47]
			Lnc-sscg3623	Cas9	18.18%[d]	
pdT	Fluorescence	NHEJ#	AAVS1	Cas9	26.3%–56.1%[a]	[48]
PTV	Fluorescence-resistance	HDR#	NFE2L2	Cas9	8%–15%[a], 2–4[b]	[49]
			A1CF	Cas9	2.98[b]	
			RBM47	Cas9	1.4[b]	
			NFE2L2	Cas9	1.49[b]	
iGFP	Fluorescence	Inversion#	GFP	Cas9	19.5%–27.7%[a]	[50]

Note: * means Exogenous; # means Endogenous; a means FACS; b means T7E1 assay; c means Sanger sequencing; d means amplicon analysis.

2 Genome repair mechanisms

ZFN or TALEN causes 5' sticky ends on the double strands of the target gene through Fok I, while Cas9 relies on its two internal cleavage domains, RuvC and HNH, to generate a blunt end between the third and fourth bases upstream of PAM[51]. Both forms of DSB trigger genomic repair. NHEJ usually results in base deletion, insertion or substitution (Fig.1A) at the break site. NHEJ is an error-prone repair method and is typically used for gene knockouts. NHEJ restores gene expression by causing two DSBs at both ends of a pathological insertion mutation[52]. Both HDR and SSA belong to homologous recombination repair mechanisms depending on homologous arm. When a donor containing DSB upstream and downstream homologous sequences is present, the break site can be accurately repaired based on the donor sequence, which can reverse pathological mutations or insert functional genes (Fig.1B). The triggering condition of SSA is that DSB occurs in the space between a pair of long direct repeat sequences. The cut is digested from the 5' end to the 3' end, and the complementary sequence at the 3' protruding end is annealed to complete the repair (Fig.1C). Chromosome inversion (Fig.1D) is associated with a variety of genetic diseases and tumorigenesis [50], which mostly occurs at the same time when both ends of a long sequence are broken, and then the sequence direction is reversed by 180°.

3 Intracellular free exogenous reporter system

The target sequence can be integrated into an exogenous reporter plasmid based on different repair mechanisms. Then the effect of genome editing can be characterized by the expression of fluorescent or resistance proteins after the co-transfection of both reporter and genome editing plasmids.

3.1 NHEJ based dual-fluorescence reporter system

In 2011, Kim *et al.*[39] used exogenous reporter plasmids encoding fluorescent proteins to indicate ZFNs and TALENs-induced genome editing positive cells. The functional structure of the dual-fluorescence reporter plasmid (Fig.2) is mRFP-Target-eGFP (pRTG), where mRFP can be used to reflect the evaluation of transfection efficiency. There is a stop codon between the target and eGFP, which makes eGFP unable to express normally. If ZFNs or TALENs induce DSBs on the target without homologous donors, there is a high probability that NHEJ repair will result in base deletion. If the number of deletions is $3n + 1$ ($n \geq 0$ and $n \in$ N), then frameshift will occur downstream of the nick position, so that eGFP is normally expressed. pRTG was transfected into HEK293 cells with a ZFNs expression plasmid targeting TP53 gene. FACS obtained 5% mRFP$^+$ eGFP$^+$ cells. T7E1 digestion test verified that the proportion of TP53 mutations in mRFP$^+$

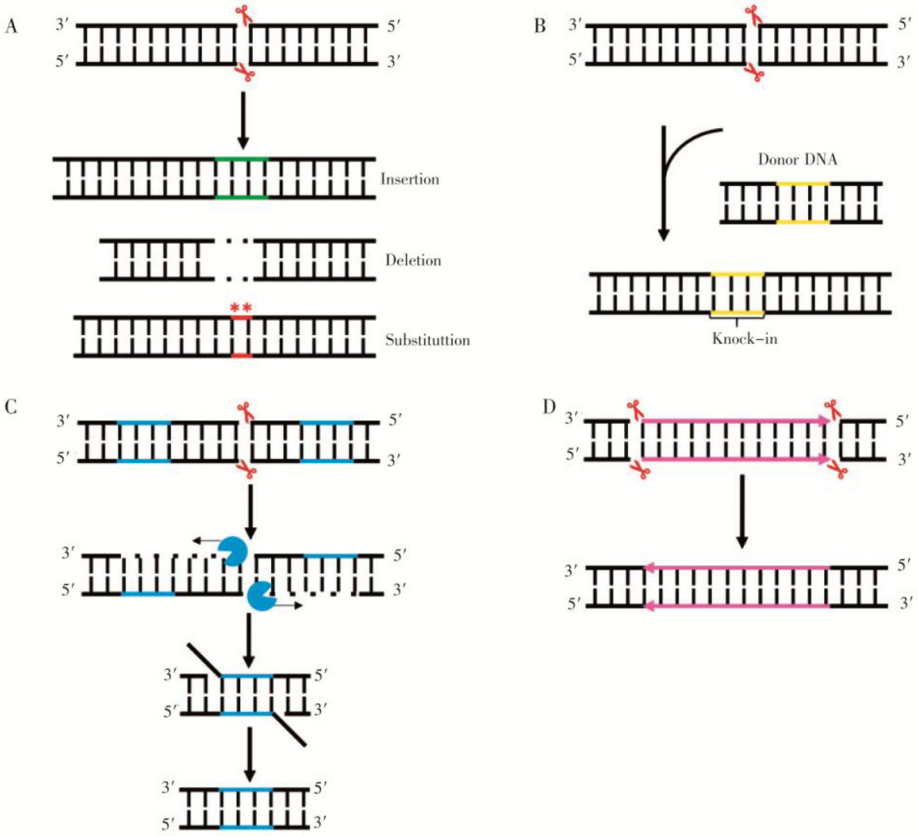

Fig.1 Four types of genome repair mechanism: A: Non-homologous end joining; B: Homology directed repair; C: Single strand annealing; D: Inversion

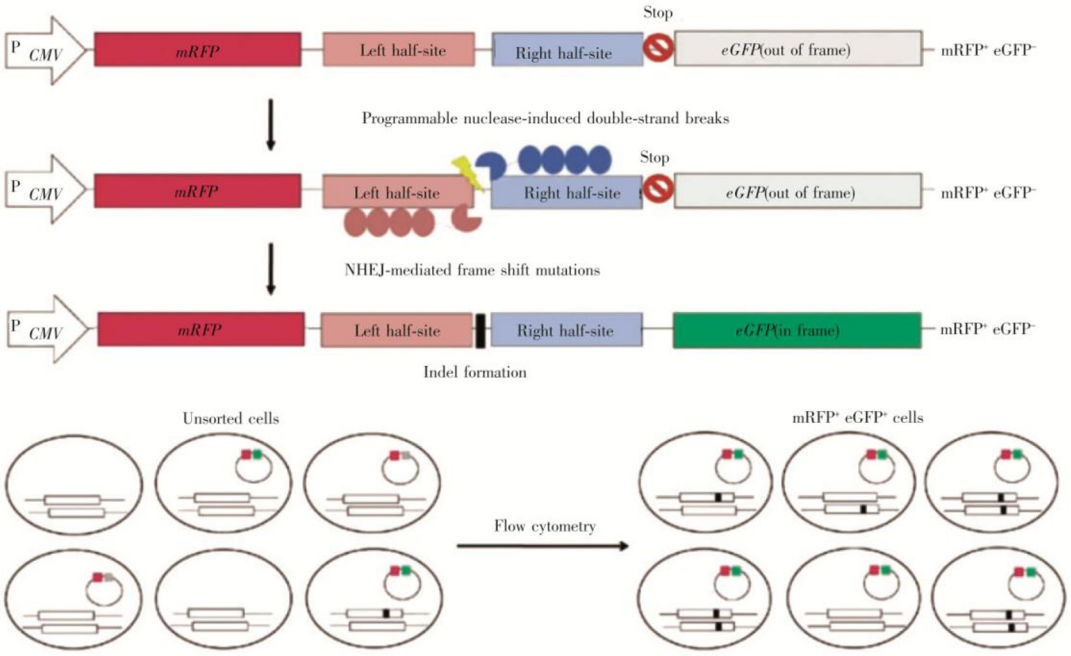

Fig.2 Principle of pRTG reporter system and cell selection[39]

eGFP$^+$ cells was about 13 times higher than that of mRFP$^+$ eGFP$^-$ and mRFP$^-$ eGFP$^-$ cells. Sequencing confirmed that the mutation ratio of mRFP$^+$ eGFP$^+$ cells was 20%. Besides, the mutation ratios of human CCR5 and mouse Thumbd3 genes were 5.8–11 times and 92 times of that of unsorted cells, respectively. This shows that pRTG system is feasible in indicating and enriching genome editing positive cells.

In 2014, Kim *et al*.[40] used the pRTG plasmid as the basic backbone, inserted an eGFP sequence after the original eGFP sequence, and placed the newly inserted sequence outside the open reading frame (ORF), thereby constructing a new dual-fluorescence reporter plasmid (mRFP-Target-eGFP-eGFP, pRTGG). Compared to a single eGFP sequence, a double eGFP sequence can enter the ORF and be expressed with greater probability. pRTGG was co-transfected with CRISPR/Cas9 expression plasmid targeting BMP1, F13A1, and PPP1R1B into HEK293T cells, and were enriched to 9.4%, 8.2% and 7.8% of mRFP$^+$ eGFP$^+$ cells by FACS; T7E1 test showed mutation frequency of mRFP$^+$ eGFP$^+$ cells was 18, 7.8 and 4.6 times higher than that of negative cells. Sequencing results showed that the mutation times were 7.4, 10, and 4.3 times, respectively. In 2018, they compared the gene interference efficiency of Cas9n (D10A) and Cas9 by using pRTGG[53] and found that the targeted cleavage efficiency of Cas9n that relied on a pair of nickase cleavage enzymes was often higher than that of wild-type Cas9. This discovery expanded the application of pRTG-related reporter system.

Since the concept of an exogenous dual-fluorescence reporter plasmid was proposed in 2011, Chinese scholars have also done many researches on it. Wang *et al*.[54] used pRTG to improve the enrichment efficiency of CRISPR/Cas 9-induced pig BMP15 gene mutant cells. Gao *et al*.[55] inserted puromycin resistance (Puro^R) gene downstream of eGFP based on pRTG, and Cas9 was used to target TMEM215 gene of HEK293T cells. Flow cytometry and puromycin screening were used. These two methods could achieve a 2.75-fold increase in mutation rate.

Although the reporter plasmids mentioned before can be efficiently enriched for nuclease-induced mutation cells, there are also some shortcomings in FACS, such as the dependence of flow cytometry equipment and the interference of fluorescent background between double fluorescent proteins.

3.2 NHEJ based fluorescence-magnetism and fluorescence-resistance reporter system

To overcome the shortcomings of the dual fluorescent reporter system, Kim *et al*.[41] designed two reporter plasmids for magnetic screening and resistance screening based on pRTG. The former introduces a truncated MHC class I molecule, H-2K^k, which is expressed only in rare mouse germlines such as AKR/J or CBA/J and is not expressed in human cells, so that it can be used as a highly specific screening marker. The functional domain is mRFP-Target-H-2K^k-eGFP (pRTKG). The fusion protein of H-2K^k and eGFP will turn to be two independent proteins by T2A self-cleaving. Using magnetic beads modified with H-2K^k antibody to separate transfected cells magnetically, positive cells can be enriched with significantly increased mutation ratio. For the resistance screening reporter plasmid (mRFP-Target-Hygro-eGFP and pRTHG), the H-2K^k gene

was replaced with a hygromycin resistance gene (HygroR). The positive rate of induced mutations on CCR5, TP53, CD81, BRCA1 and other gene loci reached 10-20 times that of the unsorted group.

Although the fluorescence-magnetism and fluorescence-resistance reporter systems do not rely on FACS, the former requires H-2K^k antibody-modified magnetic beads, which are expensive and have not been commonly commercialized. Cells may exist acquired hygromycin resistance so that false-positive results may occur.

3.3 NHEJ based linear donor reporter systems

Homologous recombination only occurs in G2 and S phases of the cells, so the integration of exogenous double-stranded DNA into the cell genome through the HDR mechanism is very unlikely. Studies on ZFNs, TALENs, and CRISPR/Cas9 have shown that when a foreign plasmid or donor generates DSB under the action of a nuclease, it will integrate into the genome in a homologous recombination-independent manner[56]. Zhou *et al*.[42] proposed an NHEJ-based linear donor reporter system in 2016 (Fig.3). The linear donor (LD) consists of 4 parts, with one or a pair of sgRNA recognition sequences on 5' end or 5' and 3' ends. The central part is the CMV-Puro^R sequence, which is in the middle of two extension sequences containing stop codons that block 6 ORFs. Also, there is 20 bp at both ends of the outer side of this linear donor as protection sequences. When Cas9 induces double-strand breaks on the genome and donor targets, the donor will be integrated into the cell genome and express puromycin resistance protein, which can survive drug screening to achieve enrichment.

LD simultaneously enriched HeLaOC cells that were knocked out by four genes (HSPA1A, HSPA1B, HSPA1L and HSPA6). T7E1 detected the mutation ratios of 64 loci in HSPA1A, HSPA1B, HSPA1L and HSPA genes were increased by 5.5, 6.1, 3.4 and 6.6 times, respectively. Different from the previously reported linear donor-dependent HR integration strategies, this donor itself carries the CMV promoter, so there is no need to integrate within the ORF of the gene of interest or restore the ORF after integration. However, dual-site sgRNAs may exacerbate the off-target effect, and the mechanism cannot control the insertion direction of the donor, so the application range of this linear reporter system is limited.

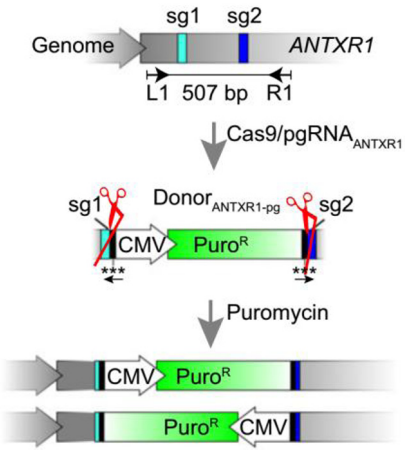

Fig.3 Principle of linear donors[42]

In summary, the advantage of the NHEJ based reporter systems lies in the extremely simple design of the reporter vectors. The open reading frame shifts by base deletion or insertion after repair to achieve the reporter function. However, the biggest limitation is that the number of deleted or inserted bases must meet $3n + 1$ or $3n + 2$ ($n \geq 0$ and $n \in$ N), so other forms of mutations cannot be accurately reflected.

3.4 SSA based dual-fluorescence reporting system

In 2014, Zhang *et al*.[43] designed a dual fluorescence reporter system (mRFP-GFP$_{left}$-target-ZFN-target-GFP$_{right}$, pRGZG) coupled with a self-degrading ZFN expression cassette. The downstream eGFP sequence was interrupted into two segments of GFP$_{left}$ and GFP$_{right}$, and the 3' and 5' ends of the two shared a 200 bp direct repeat sequence. The expression sequences of the left and right arms of the ZFN were inserted into it, and the other two pairs of targets of the left and right arms of the ZFN (Target) were placed upstream and downstream of the ZFN expression cassette. When ZFN expresses, it not only cuts the genomic target but also recognizes and cuts the target on the reporter plasmid and triggers SSA repair, recombining into a complete eGFP expression sequence. The ZFN expression cassette will be digested so that ZFN expression is terminated, and the effect of blocking the continuous expression of ZFN and reducing cytotoxicity is achieved. pRGZG enriched ZFN-induced MSTN-mutated sheep fetal fibroblasts to 18.33% of mRFP$^+$ eGFP$^+$ cells, and compared with the control group, the cell viability decreased by only about 20%, while that of normal transfected ZFN cells decreased by about 40 %. It can be seen that the pRGZG system can efficiently enrich genome editing positive cells and reduce the cytotoxicity of ZFN to a certain extent.

Although the assembly platform of ZFN is relatively mature, each zinc finger enzyme relies on the internal three amino acid residues to recognize three bases, so it depends on the design of upstream and downstream sequences, and its cytotoxicity is still a issue that cannot be ignored[57]. In contrast, CRISPR/Cas9 technology only needs designing of sgRNA sequences that are complementary to the target and can cooperate with Cas9 protein to achieve the manipulation of the target gene[51], which greatly simplifies the design steps before experiment. In 2016, Zhou *et al*.[44] designed the C-Check reporter plasmid based on the SSA repair mechanism (Fig.4). C-Check consists of two expression cassettes, a pair of truncated eGFP expression sequences used to trigger SSA repair, and an AsRED expression sequence used to determine transfection

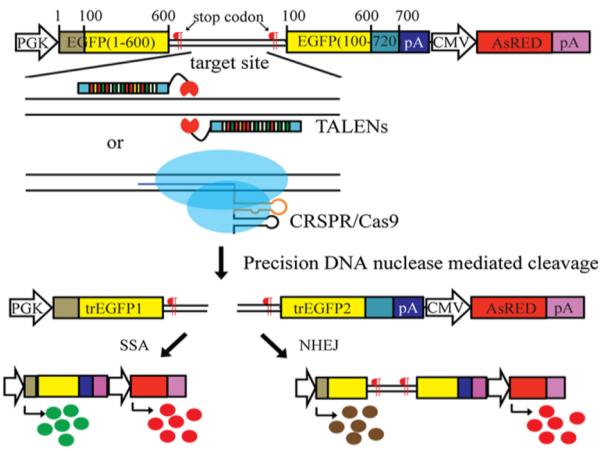

Fig.4 Principle of C-Check reporter system[44]

efficiency and establish standardization. eGFP was split into two fragments of 1–600 bp and 100–720 bp, which completely interfered with the fluorescence property. At the same time, the two parts shared a 500 bp homology arm in order to trigger SSA repair.

To verify the function of C-Check, TALEN targeted the IAPP gene in the genome of HEK 293T cells. After FACS sorting, AsRED$^+$ EGFP$^+$ cells accounted for about 80%, and the control group was less than 10%. Among the more difficult transfecting primary pig fibroblasts, a mutation rate of 2.99% was detected by T7E1, and Sanger sequencing results showed 3.125% of the mutant target gene. Cas9/sgRNA targeted the MAPT and SORL1 genes in the genome of HEK 293T cells, and FACS enriched to 11.7%–18.1% of the double-fluorescent positive cells, while the control group was only 2.37% and 2.91%. So far, the C-Check reporter system could enrich double-positive cells that were clearly distinguished from the control group, proving its high reporter performance. The researchers co-transfected C-Check and IGF1R-targeted Cas9/sgRNA expression plasmid into HEK293T cells. FACS sorted four groups of double-positive cells from weak to strong based on the EGFP fluorescence signal. Amplicon analysis showed that the frequency of indels increased significantly with the increase in EGFP fluorescence intensity (8.7%–97.9%), indicating that IGF1R mutations induced by CRISPR/Cas9 were efficiently enriched. Similarly, the C-Check reporting system also enriched 86.9% of CBX5 mutant cells in the cancer cell line MCF-7, with reduced mRNA levels and CBX protein expression, further confirming that C-Check is a highly efficient reporter system for editing positive cells.

Most of the above exogenous reporters plasmids need co-transfection with nuclease-expressing plasmids, but the proportion of the two components entering the cells is difficult to assess. The CRISPR/Cas9 genome editing system originally required co-transfection of Cas9 expression plasmid and tracr/crRNA complex. Doudna and Charpentier chimerically expressed crRNA:tracrRNA[58] as single-guided RNA (sgRNA), enabling Cas9/sgRNA to express within a single plasmid, thereby increasing transfection efficiency. Therefore, if the reporter system and Cas9/sgRNA expression plasmid are co-encapsulated into a single expression system, the enrichment efficiency of mutant cells may be improved. Liu *et al.*[45] interrupted mCherry into two parts, each containing about 300 bp of repeat sequences as SSA repair arms. sgRNA targets were inserted in the middle, and then the mCherry expression cassette was integrated into the Cas9/sgRNA expression plasmid (sgRNA-mCherry-CRISPR, pSCR), in which the copGFP was used to characterize transfection efficiency. T7E1 detected mutation percentages of the three target sites of the DAZL gene in copGFP$^+$ mCherry$^+$ cells were 24.79%, 12.26% and 17.18%, respectively, and the mutation percentages increased by 1.87, 4.99 and 4.00 times, respectively. Besides, the mutation enrichment rates of double-copy sgRNA targeting PLZF and ACR genes of murine melanoma B16 cells were up to 54.32% and 40.49%.

The advantage of the dual-fluorescence reporter system is that one fluorescent protein is responsible for indicating transfection efficiency, and another fluorescent protein is indicative of genome editing efficiency. Therefore, enrichment analysis of cells that are positive for dual fluorescence signals can simplify and expedite the study of mutation types. However, the emission wavelengths of the two fluorescent proteins overlap to a certain degree, and the resulting non-specific background may interfere with FACS sorting, leading to the collection of false double-positive cells, so appropriate thresholds need to be drawn to distinguish.

3.5 SSA based fluorescence-resistance-fluorescence reporter system

Ren *et al.*[46] designed a new fluorescence-resistance-fluorescence reporter plasmid pSSA-RPG (DsRed-Puro^R-eGFP). The Puro^R gene is interrupted into two parts: upstream (Puro^{R}L) and downstream (Puro^{R}R). The sgRNA recognition site and PAM sequence are inserted in the middle. The 200 bp direct repeat sequence shared between the 5 ‘end of Puro^{R}L, and the 3’ end of Puro^{R}R is used as the SSA repair arm. Downstream of Puro^{R}R, a T2A cleavage peptide sequence is connected to the eGFP expression sequence, and the eGFP sequence is outside the normal ORF because the puromycin resistance gene sequence is interrupted. At this time, the cell will display DsRed$^+$ eGFP$^-$ Puro^{R-}. When DSB is produced on the target, SSA repair is triggered, and Puro^{R}L and Puro^{R}R are recombined into a complete puromycin resistance gene. Then the ORF is reset for eGFP normally expressing, and the cells show DsRed$^+$ eGFP$^+$ Puro^{R+}. T7E1 detected the four sites of AAVS1, CCR2, CCR5a, and CCR5b in the genome of HEK 293T cells. The proportion of mutations at four sites in the drug-screening group was 43.9%, 57.8%, 27.1% and 24.7%,

respectively. The mutation times were 34.8, 13.8, 12.3 and 6.3 times, respectively. The sequencing results showed that the mutation ratios were 86.6%, 72.7%, 30.8% and 45.0%, respectively, and the mutation times were 21.1, 34.6, 18.1 and 11.8 times, respectively. After FACS enriched CCR5a mutant cells with DsRed$^+$ eGFP$^+$, T7E1 detected 31.4% of the cells as mutation-positive, with mutation times of 13.2. Sequencing showed that the mutation ratio and mutation times were 23.5% and 11.2 times, respectively.

In order to study whether the reporter system based on the SSA was superior to the aforementioned NHEJ based reporter system, the research team also designed the corresponding reporter plasmid (pNHEJ-RPG), inserting a stop codon before the complete Puro^R and eGFP sequences. The principle is consistent with the dual-fluorescence reporter plasmids designed by Kim *et al*.[39,40]. The enrichment efficiency of pSSA-RPG at the AAVS1 site was 1.15–1.34 times higher than that of pNHEJ-RPG, showing a small efficiency improvement. However, analysis of the pSSA-RPG sequence revealed that Puro^{R}R had several ATG start codons, so the leakage of downstream eGFP may occur during intracellular transcription, leading to false-positive results. This may be the main disadvantage of both double markers with Puro^R and eGFP.

Recent studies have found that the probability of a single allele being edited in the cell genome is significantly higher than that of double alleles[59–61]. However, in the field of disease cell model construction, transgenic animal model construction, and gene therapy, it is urgent to quickly screen biallelic mutations. Therefore, Wu *et al*.[47] transformed pSSA-RPG to construct two reporter-integrated donor plasmids (Rep/Don), as shown in Fig.5. One inserted the homologous arms (Rep/DonPG, Fig.5A) upstream and downstream of the target in the genome in the upstream and downstream of the Puro^R-eGFP expression cassette, and another one replaced the reporting sequence inside the homology arms on both sides with Zeo^R-mRFP expression cassette (Rep/DonZR, Fig.5B). When Cas9 cuts the genome and the target of Rep/Don, SSA repair occurs within Rep/Don to restore the expression of resistance and fluorescent proteins. The outer homology arm mediates HDR, integrating Puro^R-eGFP or Zeo^R-mRFP into the cell genome. Functional fluorescent protein can be used as a primary screen to reflect Cas9 cleavage activity, and then puro and zeo dual drug screening can be used to enrich biallelic-modified cells efficiently. The proportion of biallelic mutations in CCR5 locus of human HEK293T cells and Lnc-sscg3623 locus of pig PK15 cells reached 34.09% and 18.18%, respectively, and the biallelic mutation rate of Rep/DonPG or Rep/DonZR single transfection control group was only 2.33%–6.81%, so Rep/Don could be used as a highly efficient reporter system for enrichment of biallelic mutations.

More and more researchers favor the exogenous reporter system based on the SSA repair mechanism because of high repair efficiency and not having to consider base number changes. This type of reporting system has a common defect with the NHEJ-based reporting system, which is the background signal caused by leakage expression of fluorescent proteins or partial overlap of the excitation light bands of two fluorescent proteins[61]. In order to overcome this problem, many studies have combined resistance screening with fluorescence sorting, and standardized fluorescence signal processing with fluorescent indicator proteins, thereby further specificity and enrichment efficiency are improved. Besides, according to Li *et al*.[56], increasing the expression level of sgRNA could directly enhance the cleavage efficiency of Cas9, and then promote the enrichment rate.

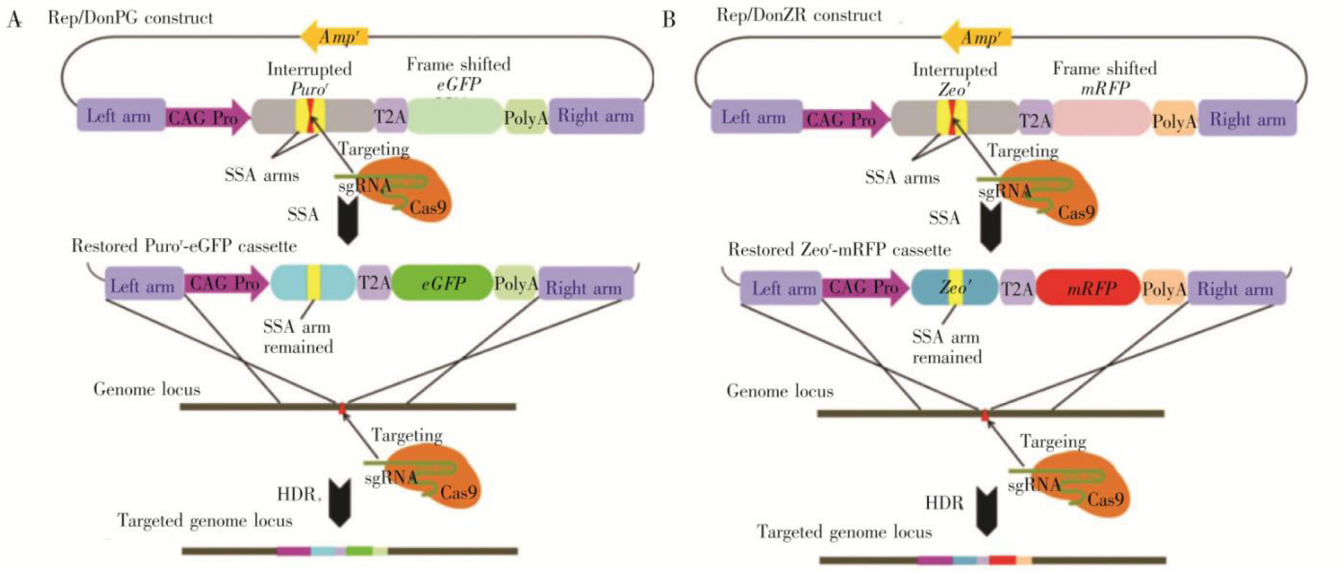

Fig.5 Principle of Rep/Don donor reporter system[47]: (A) Rep/DonPG construct; (B) Rep/DonZR construct

4 Cell genome integrated endogenous reporter system

Compared with the exogenous reporter system, integrating a reporter system based on the corresponding repair mechanism into the cell genome can build an endogenous reporter system, so there is no need to consider the proportion of multiple plasmids co-transfection or the transfection efficiency. Moreover, the integrated reporter system is consistent with the copy number of the target gene, which can more accurately reflect the mutation caused by the nucleases to the target and can also increase the proportion of mutations enrichment. However, building stable expressing cell lines is often a time-consuming and labor-intensive process, so there are fewer studies on integrated endogenous reporter systems.

4.1 NHEJ based reporter system

D'Astolfo *et al.*[48] designed an endogenous dTomato reporter system (pdT) based on the NHEJ repair mechanism, as shown in Fig.6. The pdT structure is EF1a-ATG-AAVS1-dTomato. AAVS1 as the target sequence also interrupts the dTomato reading frame, so it cannot be expressed normally. When Cas9 cuts AAVS1, the downstream ORF shifts, resulting in dTomato positive cells. Lentiviral vectors constructed KBM7 and H1 human embryonic stem cells. After the first round of Cas9/sgRNA transfection, the dTomato intensity in KBM7 cells and H1 human embryonic stem cells reached 33.8% and 10.2%, respectively. After the second round of transfection, the dTomato intensity increased to 56.1% and 26.3%, respectively. At the same time, off-target experiments showed that dTomato positive signal was about 0.2%, which was much lower than the fluorescence intensity of the experimental group. The above results indicated that the dTomato reporter system could efficiently enrich mutation-positive cells, but the research group did not analyze the mutation of the target gene, and further sequencing results are needed to prove the stability of the reporter system.

4.2 HDR based reporter system

Piggybac transposon (PB) can efficiently integrate large fragments of DNA, thereby creating a cell line with stable expression of foreign genes. Wen *et al.*[49] used PB to construct a donor-dependent fluorescence-resistance reporter system (Piggybac target vector, PTV), as shown in Fig.7. The highly active PB transposase (hyPBase) integrated the CMV-Puro^R-pA-H2B-GFP-pA expression element into the cell genome, and pA (polyadenylation signal) terminated GFP expression downstream of the Puro^R gene. At the same time, a pair of Cas9D10A recognition sequences was between Puro^R and GFP. The plasmid donor contained Puro^R-T2A-H2B-GFP sequence, where Puro^R and H2B-GFP were PTV homology

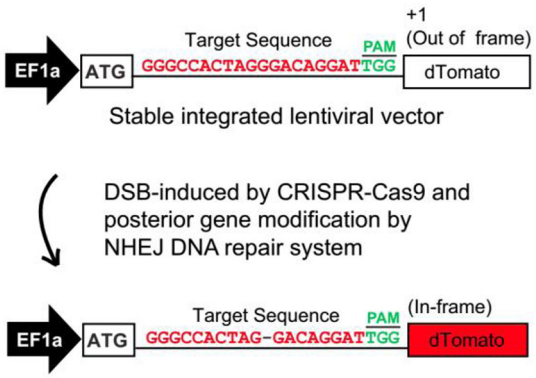

Fig.6 Principle of endogeneous dTomato reporter system induced by NHEJ[48]

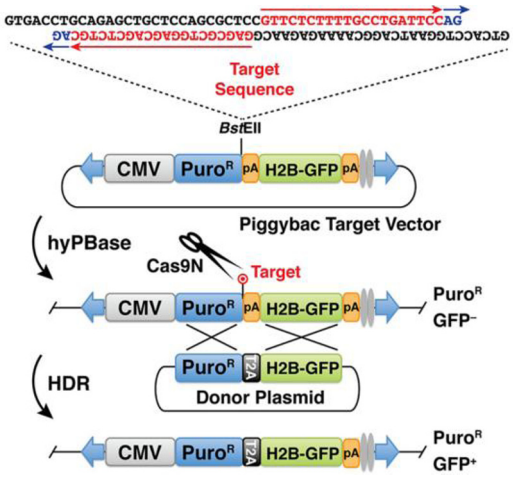

Fig.7 Principle of endogenous PuroR-GFP reporter system induced by HDR[49]

arms, and the target sequence in PTV generated DSB.

After FACS sorted Hela and DLD1 cell lines, GFP$^+$ cells accounted for about 8% and 15%, respectively. T7E1 digestion verified that the NFE2L2 site mutation rates reached 33.1% and 9.5%, respectively, which was 2–4 times higher than that of the unsorted group. Then the GFP sequence was replaced with Hygro^R, and the drug screening enriched to about 60, 120 and 100 live DLD1 cells. The mutation rate of three gene loci (A1CF, RBM47, and NFE2L2) by T7E1 test was 19.7%, 51.5%, and 29.5%, respectively. The non-specific gRNA control group was relatively reduced by 1.4-3 times. Although the target mutation times of the cell genome does not increase a lot after enrichment, the system does make a greater contribution to the study of targeted insertion of the target gene under the premise of a lower incidence of HDR.

4.3 Inversion based reporter system

Some regions of the chromosome have a strong tendency to rearrange, leading to the occurrence of hereditary diseases or cancer. Li *et al.*[50] constructed a reporter system that directly reflected CRISPR/Cas9-induced sequence rearrangements (Fig.8).

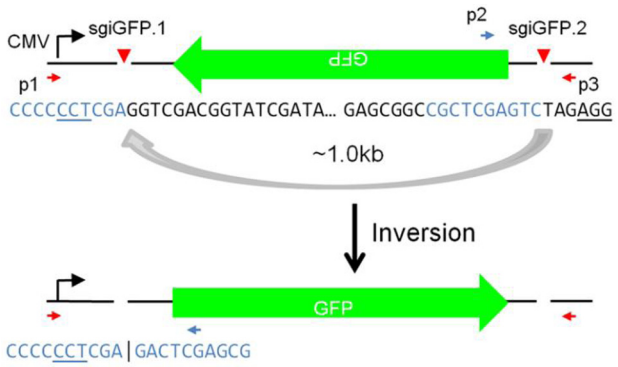

Fig.8 Principle of endogenous GFP reporter system induced by rearrangement[50]

The inverted GFP (Inverted GFP, iGFP) sequence was inserted downstream of the CMV promoter, and GFP was not expressed at this time. When Cas9 created double-stranded breaks at positions about 1.0 kb upstream and downstream of the GFP expression cassette, the direction of iGFP might be reversed, so that GFP might express normally. A retrovirus was used to integrate iGFP into the chromosomes of murine 3T3 cells, so a single-copy stable expression cell line of iGFP was constructed to quantify iGFP rearrangements. FACS analysis revealed 23.6% ± 4.1% of GFP^+ cells, and obvious bands were found by PCR. The iGFP rearrangement reporter system has certain application prospects for studying induced chromosome rearrangements.

In the existing endogenous reporter system, NHEJ, as a simpler repair mechanism, has a higher incidence than that for HDR or chromosomal inversion, and higher efficiency for the enrichment of positive cells. Constructing a cell line with stable expression of reporter system is a time-consuming and labor-intensive process, but it has the advantages that an exogenous reporter system cannot match, such as no need for co-transfection, and the same copy number of the reporter system target and genomic target. However, because the integrated target sequence is difficult to change, each constructed reporter cell line can only enrich the corresponding target gene, which has poor versatility and narrow application range.

5 Outlooks

At present, although the free exogenous reporter system has issues such as degradation and uneven distribution, it possesses advantages of flexible design and convenient transfection. It can quickly use different genome editing tools for research in different cell lines. Although the endogenous integrated reporter system can more stably characterize the effect of genome editing, the longer time cost in cell line construction cannot be ignored. Therefore, the types of exogenous reporter systems are far more than those of endogenous reporter systems. At the same time, fluorescence-fluorescence, fluorescence-resistance, fluorescence-magnetism, and other reporter systems are derived based on the common and efficient NHEJ and SSA mechanisms and inefficient repair mechanisms such as HDR and inversion rearrangement. The first two reporter systems have higher economic availability and design simplicity, and rapid positive enrichment by using mature FACS technology or drug screening method. The significance of the fluorescence-magnetism reporter system lies in its innovation, while the expensive antibody modification and cumbersome elution steps reduce the ease of system. Although the above reporter system implements a rapid and efficient enrichment strategy, objective internal and external factors such as heterogeneity between cell lines, differences in the identification and targeting efficiency of different gene loci, and cleavage activity of different genome editing tools, as well as the sensitivity differences in FACS, resistance screening, and magnetic separation, may lead to low enrichment ratios of positive cells. Therefore, in order to minimize or exclude negative and false positive cells, and to achieve a higher or complete proportion of positive cell enrichment, the joint application of two or more reporter systems or an integrated reporter system incorporating multiple enrichment strategies may be required. In addition, these reporter systems also require cytotoxicity tests to verify their safety and stability, which will help ensure high reproducibility and downstream culture of enriched mutant cell lines.

In the future, the genome editing positive cell enrichment reporter system will continue to improve and focus on two efficient repair mechanisms, NHEJ and SSA. At the same time, it also needs to follow up with the development of genome editing tools in real-time. For example, redesigning the sequence of the reporter vector in applying to non-cutting CRISPR genome editing systems that rely on transposons, or exploring its application in RNA editing or single base editing to assist and accelerate the development of new genome editing tools. Besides, efforts should be made to study reporter systems that can reflect clinically significant non-error prone gene repair mechanisms, such as HDR, in order to promote the development and application of precise genome editing.

References

[1] Jackson D A, Symons R H, Berg P. *P. Natl. Acad. Sci. USA*, **1972**, 69(10): 2904–2909

[2] Capecchi MR. *Science*, **1989**, 244(4910): 1288–1292

[3] Moore J K, Haber J E. *J. Mol. Cell Biol.*, **1996**, 16(5): 2164–2173

[4] Thierry A, Dujon B. *Nucleic Acids Res*., **1992**, 20(21): 5625–5631

[5] Thierry A, Perrin A, Boyer J, Fairhead C, Dujon B, Frey B, Schmitz G. *Nucleic Acids Res*., **1991**, 19(1): 189–190

[6] Smith J, Grizot S, Arnould S, Duclert A, Epinat J C, Chames P, Prieto J, Redondo P, Blanco F J, Bravo J, Montoya G, Paques F,

Duchateau P. *Nucleic Acids Res*., **2006**, 34(22): e149

[7] Boissel S, Jarjour J, Astrakhan A, Adey A, Gouble A, Duchateau P, Shendure J, Stoddard B L, Certo M T, Baker D, Scharenberg A M. *Nucleic Acids Res*., **2014**, 42(4): 2591–2601

[8] Kim Y G, Cha J, Chandrasegaran S. *P. Natl. Acad. Sci. USA*, **1996**, 93(3): 1156–1160

[9] Wolfe S A, Nekludova L, Pabo C O. *Annu. Rev. Biophys. Biomol. Struct*., **2000**, 29: 183–212

[10] Bibikova M, Beumer K, Trautman J K, Carroll D. *Science*, **2003**, 300(5620): 764

[11] Urnov F D, Miller J C, Lee Y L, Beausejour C M, Rock J M, Augustus S, Jamieson A C, Porteus M H, Gregory P D, Holmes M C. *Nature*, **2005**, 435(7042): 646–651

[12] Miller J C, Holmes M C, Wang J, Guschin D Y, Lee Y L, Rupniewski I, Beausejour C M, Waite A J, Wang N S, Kim K A, Gregory P D, Pabo C O, Rebar E J. *Nat. Biotechnol*., **2007**, 25(7): 778–785

[13] Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. *Science*, **2009**, 326(5959): 1509–1512

[14] Moscou M J, Bogdanove A J. *Science*, **2009**, 326(5959): 1501

[15] Mahfouz M M, Li L, Shamimuzzaman M, Wibowo A, Fang X, Zhu J K. *P. Natl. Acad. Sci. USA*, **2011**, 108(6): 2623–2628

[16] Christian M, Cermak T, Doyle E L, Schmidt C, Zhang F, Hummel A, Bogdanove A J, Voytas D F. *Genetics*, **2010**, 186(2): 757–761

[17] Li T, Huang S, Jiang W Z, Wright D, Spalding M H, Weeks D P, Yang B. *Nucleic Acids Res*., **2011**, 39(1): 359–372

[18] Xu S, Cao S, Zou B, Yue Y, Gu C, Chen X, Wang P, Dong X, Xiang Z, Li K, Zhu M, Zhao Q, Zhou G. *Genome Biol*., **2016**, 17(1): 186

[19] Bolotin A, Quinquis B, Sorokin A, Ehrlich S D. *Microbiology*, **2005**, 151(8): 2551–2561

[20] Garneau J E, Dupuis M E, Villion M, Romero D A, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan A H, Moineau S. *Nature*, **2010**, 468(7320): 67–71

[21] Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero D A, Horvath P. *Science*, **2007**, 315(5819): 1709–1712

[22] Mali P, Yang L, Esvelt K M, Aach J, Guell M, DiCarlo J E, Norville J E, Church G M. *Science*, **2013**, 339(6121): 823–826

[23] San F J, Sung P, Klein H. *Annu. Rev. Biochem*., **2008**, 77: 229–257

[24] Lieber M R. *Annu, Rev. Biochem*., **2010**, 79: 181–211

[25] Aubrey B J, Kelly G L, Kueh A J, Brennan M S, O'Connor L, Milla L, Wilcox S, Tai L, Strasser A, Herold M J. *Cell Rep*., **2015**, 10(8): 1422–1432

[26] El Refaey M, Xu L, Gao Y, Canan B D, Adesanya T M A, Warner S C, Akagi K, Symer D E, Mohler P J, Ma J, Janssen P M L, Han R. *Circ. Res*., **2017**, 121(8): 923–929

[27] Ren J, Liu X, Fang C, Jiang S, June CH, Zhao Y. *Clin. Cancer Res*., **2017**, 23(9): 2255–2266

[28] Kurochkin I V, Guarnera E, Berezovsky I N. *Trends Pharmacol. Sci*., **2018**, 39(1): 49–58

[29] Doench J G. *Nat. Rev. Genet*., **2018**, 19(2): 67–80

[30] Yue H H, Bao Y J, Zhou X M. *Acta Laser Biol. Sin.*, **2017**, 26(6): 490–494

[31] Fu Y, Foden J A, Khayter C, Maeder M L, Reyon D, Joung J K, Sander J D. *Nat. Biotechnol*., **2013**, 31(9): 822–826

[32] 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, Cradick T J, Marraffini L A, Bao G, Zhang F. *Nat. Biotechnol*., **2013**, 31(9): 827–832

[33] Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna J A, Liu D R. *Nat. Biotechnol*., **2013**, 31(9): 839–843

[34] Cromwell C R, Sung K, Park J, Krysler A R, Jovel J, Kim S K, Hubbard B P. *Nat. Commun*., **2018**, 9(1): 1448

[35] Hendel A, Bak R O, Clark J T, Kennedy A B, Ryan D E, Roy S, Steinfeld I, Lunstad B D, Kaiser R J, Wilkens A B, Bacchetta R, Tsalenko A, Dellinger D, Bruhn L, Porteus M H. *Nat. Biotechnol*., **2015**, 33(9): 985–989

[36] Rahdar M, Mcmahon M A, Prakash T P, Swayze E E, Bennett C F, Cleveland D W. *P. Natl. Acad. Sci. USA*, **2015**, 112(51): 7110–7117

[37] Hu J H, Miller S M, Geurts M H, Tang W, Chen L, Sun N, Zeina C M, Gao X, Rees H A, Lin Z, Liu D R. *Nature*, **2018**, 556(7699): 57–63

[38] Bhargava R, Onyango D O, Stark J M. *Trends Genet*., **2016**, 32(9): 566–575

[39] Kim H, Um E, Cho S R, Jung C, Kim H, Kim J S. *Nat. Methods*, **2011**, 8(11): 941–943

[40] Ramakrishna S, Cho S W, Kim S, Song M, Gopalappa R, Kim J S, Kim H. *Nat. Commun*., **2014**, 5: 3378

[41] Kim H, Kim M S, Wee G, Lee C I, Kim H, Kim J S. *PLoS One*, **2013**, 8(2): e56476

[42] Zhou Y, Zhang H, Wei W. *FEBS Lett*., **2016**, 590(23): 4343–4353

[43] Zhang C, Xu K, Hu L, Wang L, Zhang T, Ren C, Zhang Z. *Biotechnol. Lett*., **2015**, 37(2): 299–305

[44] Zhou Y, Liu Y, Hussmann D, Brogger P, Al-Saaidi R A, Tan S, Lin L, Petersen T S, Zhou G Q, Bross P, Aagaard L, Klein T, Ronn S G, Pedersen H D, Bolund L, Nielsen A L, Sorensen C B, Luo Y. *Cell. Mol. Life Sci*., **2016**, 73(13): 2543–2563

[45] Liu W, Li S, Zhang Y, Li J, Li Y. *FEBS Open Bio*., **2018**, 8(7): 1167–1175

[46] Ren C, Xu K, Liu Z, Shen J, Han F, Chen Z, Zhang Z. *Cell. Mol. Life Sci*., **2015**, 72(14): 2763–2772

[47] Wu Y, Xu K, Ren C, Li X, Lv H, Han F, Wei Z, Wang X, Zhang Z. *FEBS Lett*., **2017**, 591(6): 903–913

[48] D'astolfo D S, Pagliero R J, Pras A, Karthaus W R, Clevers H, Prasad V, Lebbink R J, Rehmann H, Geijsen N. *Cell*, **2015**, 161(3): 674–690

[49] Wen Y, Liao G, Pritchard T, Zhao T T, Connelly J PPruett-Miller S M, Blanc V, Davidson N O, Madison B B. *J. Biol. Chem*., **2017**, 292(15): 6148–6162

[50] Li Y, Park A I, Mou H, Colpan C, Bizhanova A, Akama-Garren

E, Joshi N, Hendrickson E A, Feldser D, Yin H, Anderson D G, Jacks T, Weng Z, Xue W. *Genome Biol*., **2015**, 16: 111

[51] Jiang F, Doudna J A. *Annu. Rev. Biophys*., **2017**, 46: 505–529

[52] Cox D B, Platt R J, Zhang F. *Nat. Med*., **2015**, 21(2): 121–131

[53] Gopalappa R, Suresh B, Ramakrishna S, Kim H H. *Nucleic Acids Res*., **2018**, 46(12): e71

[54] Wang M, Shi H, Huang X, Liu X F, Qin Y Fg, Liu X H, Chen Y S, He Z Y. *Hereditas (Beijing)*, **2017**, 39(1): 48–55

[55] Gao X T, Fu W, Chen J J, Hu B, Li D H, Huang S Y, Li G H, Zhang Y P, Liu L, Liang Y M. *Prog. Mod. Biomed*., **2016**, 16(4): 652–656

[56] Lackner D H, Carre A, Guzzardo P M, Banning C, Mangena R, Henley T, Oberndorfer S, Gapp B V, Nijman S M, Brummelkamp T R, Burckstummer T. *Nat. Commun*., **2015**, 6: 10237

[57] Li H L, Nakano T, Hotta A. *Dev. Growth Differ*., **2014**, 56(1): 63–77

[58] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna J A, Charpentier E. *Science*, **2012**, 337(6096): 816–821

[59] Wang Y, Li J, Xiang J, Wen B, Mu H, Zhang W, Han J. *Protein Cell,* **2016**, 7(2): 152–156

[60] He X, Tan C, Wang F, Wang Y, Zhou R, Cui D, You W, Zhao H, Ren J, Feng B. *Nucleic Acids Res*., **2016**, 44(9): e85

[61] Ye L, Wang J, Beyer A I, Teque F, Cradick T J, Qi Z, Chang J C, Bao G, Muench M O, Yu J, Levy J A, Kan Y W. *P. Natl. Acad. Sci. USA,* **2014**, 111(26): 9591–9596