A simple method based on Sanger sequencing and MS Word wildcard searching to identify Cas9-induced frameshift mutations

Hui Jie[#], Zhuoling Li[#], Ping Wang, Linjie Zhao, Qian Zhang, Xiaomin Yao, Xiangrong Song, Yinglan Zhao and Shaohua Yao

Recent advances in targeted genome editing have enabled sequence-specific modifications in eukaryotic genomes. As it can be easily reprogrammed, the clustered regularly interspaced short palindromic repeat (CRISPR)-Cas9 nuclease system has been studied extensively and is now a widely used genome editing tool. Generally, Cas9 nucleases are designed to target the coding regions in exons of protein-coding genes, which are expected to cause frameshift indel mutations and interrupt protein expression. In such cases, it is often necessary to separate single clones that harbor double frameshift mutant alleles from clones that harbor the wild-type allele or an in-frame mutant allele. We developed a simple and efficient method to identify frameshift mutations in diploid genomes based on Sanger sequencing and MS Word wildcard searching (SWS). As indel mutations induced by Cas9 are varied, Sanger sequencing of PCR products from a single mutant genome will generate double peaks that begin at the indel sites. By positioning the putative sequences deduced from the double peak regions in the sequencing graph onto the wild-type sequence by MS Word wildcard searching, it is possible to predict exactly how many nucleotides were deleted or inserted in each allele of the genome. The SWS strategy greatly facilitates the process of identifying single clones with biallelic frameshift mutations from pooled cells or model organisms.

Recent, rapid advances in engineered nuclease-mediated genome editing have provided efficient tools for gene function studies and forward genetic screening. Three major nuclease families, zinc-finger nucleases (ZFNs),[1] transcription activator-like effector nucleases (TALENs)[2] and the CRISPR/Cas9 nucleases,[3] have been studied extensively and are widely used to induce site-specific genome modifications. Upon binding to their target DNAs, these nucleases induce double-strand breaks (DSBs), which, in most cases, are repaired by the non-homologous end-joining repair system resulting in random insertions and/or deletions (indels).[4–6] ZFNs and TALENS are chimera nucleases composed of a sequence-specific DNA-binding module and a non-sequence-specific DNA cleavage domain from the Fok1 endonuclease.[1,7] CRISPR/Cas9 is an RNA guided system composed of a Cas9 protein and a single guide RNA (sgRNA), both of which are responsible for target recognition.[3,8–10] SgRNA recognizes complementary DNA sequences and the Cas9 protein recognizes the protospacer recognition motif (PAM) of target

DNAs, which is a unique sequence adjacent to RNA-binding sites. Upon binding to its target, Cas9 catalyzes DSBs at several base pairs upstream of PAM.[11,12] Reprogramming the CRISPR/Cas9 system for new targets is unprecedentedly easy, and only requires changing the first approximately 20 bases in the sgRNA. Besides, compared with the ZFNs and TALENs, CRISPR/Cas9 requires only a PAM in the target DNA and tolerates DNA methylation, and thus has a higher number of targetable sites.[3,13,14] These advantages of the CRISPR/Cas9 system have greatly improved the ability to make desired mutations or modifications in genomes.[14–18]

In most applications, Cas9 nucleases are designed to induce indel mutations in coding regions in exons of protein-coding genes. In gene function studies, it is usually required to isolate single clones harboring biallelic frameshift mutations from clones harboring wild-type or in-frame mutant alleles. Although the presence of indel mutations in specific target regions can be determined easily by a variety of methods, including limited endonuclease assay,[19] T7E1 and

State Key Laboratory of Biotherapy, Cancer Center, West China Hospital, Sichuan University, 17# 3rd Section of South Renmin Road, Chengdu 610041, Sichuan, China
Correspondence: Associate Professor S Yao, PhD or Professor Y Zhao, MD, State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University, 17# 3rd Section of South Renmin Road, Chengdu, Sichuan 610041, China.
E-mail: shaohuayao@scu.edu.cn or zhaoyinglan@scu.edu.cn
#These authors contributed equally to this work.

SURVEYOR assays,[4,20] high-resolution melting curve assay,[21] and heteroduplex mobility assay,[22] these methods cannot predict whether the indel mutations are frameshift mutations. Determining the exact mutation status of each allele can be achieved by PCR cloning and subsequent sequencing of single alleles.[23,24] Western blotting also can be used to check the absence of proteins of interest to predict the mutation status, because biallelic frameshift mutations result in protein truncations and sometimes reduced expression levels as a result of non-sense mediated mRNA decay. However, both these methods are time-consuming and laboratory intensive.

Here, we developed a simple but efficient method based on Sanger sequencing and the wildcard search function of MS Word, which we called the SWS strategy. By Sanger sequencing of PCR products of single-cell clones and subsequent analysis of the sequencing trace with wildcard searching, we could accurately and reliably predict if indels existed in each allele of a diploid genome and their exact numbers. This method can easily identify single-cell clones harboring biallelic frameshift indels without the need for plasmid cloning and sequencing of individual bacterial colonies to identify exact indels within each allele. We also compared SWS with the recently developed CRISP-ID online tool, which serves a similar purpose as SWS.[25]

MATERIALS AND METHODS
SgRNAs Design and Plasmid Construction
SgRNAs were designed to target the fourth exon of the phosphoserine aminotransferase 1 (*PSAT1*) gene with the CRISPR gRNA design tool (https://www.dna20.com/eCommerce/cas9/input). Three target sites were chosen to test for efficiency. The target sites and the primers used to amplify the target sites are listed in Table 1. SgRNA oligos were cloned into plasmid pX459 (Addgene, Cambridge, MA, USA) following an online protocol from the Feng Zhang laboratory (https://www.addgene.org/static/cms/files/Zhang_lab_Lenti CRISPR_library_protocol.pdf).[26] Briefly, single-stranded oligo nucleotides were phosphorylated and annealed into double strands with a PCR machine and ligated to *Bbs*I-digested pX459 plasmids.

Table 1 Oligos used for Cas9 vector construction

Oligo name	Primer sequence (5′ to 3′)
PSAT1 sgRNA1 F	CACCGTAGCTGACCAAGCTCCTGTC
PSAT1 sgRNA1 R	AAACGACAGGAGCTTGGTCAGCTAC
PSAT1 sgRNA2 F	CACCGAAACTTCTTGGCTTCTTCTG
PSAT1 sgRNA2 R	AAACCAGAAGAAGCCAAGAAGTTTC
PSAT1 sgRNA3 F	CACCGAAAATCACCTTATAGTTGTC
PSAT1 sgRNA3 R	AAACGACAACTATAAGGTGATTTTC

Cell Culture, Transfection and Isolation of Single-Cell Clones
Eca-109 esophageal squamous cell carcinoma cells were maintained in RPMI-1640 medium (HyClone, Logan, UT, USA) with 10% FBS, at 37 °C with 5% CO_2. Then, 2.5 μg pX459-sgRNA plasmids were transfected into the eca-109 cells using Lipofectamine 3000 reagent (Thermo Fisher, Shanghai, China). At 24 h post-transfection, 1 μg/ml puromycin was added to the medium to select positively transfected cells. Three days post-puromycin selection, the cells were seeded in 96-well plates with serial dilutions.

Genomic DNA Extraction and PCR Protocol
The eca-109 cells were harvested at about 90% confluent and lysed with lysis buffer (100 μg/ml proteinase K, 10 mM Tris at pH = 8.0, and 2 mM EDTA) at 55 °C for 2 h. The lysates were boiled for 5 min, then centrifuged at $16\,200 \times g$ for 5 min. The supernatants were transferred to new tubes for PCR analysis. Prime STAR Max premix (TAKARA, Dalian, China) was used to amplify the genomic DNA as follows: 98 °C for 5 min; followed by 35 cycles of 98 °C for 10 s, 52 °C for 15 s, 72 °C for 5 s; then 72 °C for 10 min; and 16 °C for 10 min. The primers used to amplify the *PSAT1* DNA fragment containing sgRNA targets were: *PSAT1*-F: 5′-GCACTTGGTAGAGCATCACTTG-3′ and *PSAT1*-R: 5′-AGCTAGTACAGCCACCCAATCT-3′.

Sanger Sequencing
The PCR products were sequenced by the Sanger method using specific forward primer (Tsingke, Chengdu, China). For sequencing single allele, 2 μl PCR products were ligated to a blunt-end cloning vector (Vazyme, Nanjing, China) to generate single-allele-derived colonies. Then, 3–5 single colonies from each ligation were sequenced with the *PSAT1* forward primer to ensure that both alleles were covered.

Western Blot Analysis
Cells (1×10^6) were lysed on ice in RIPA buffer (Beyotime, Nantong, China) for 30 min in the presence of protease inhibitors (Selleck, Houston, TX, USA). The whole-cell lysates were centrifuged for 15 min at $16\,200 \times g$. The supernatants were transferred into new tubes and protein concentrations were determined by Coomassie brilliant blue G250 (Bio-Rad, CA, USA). The lysates were mixed with protein loading buffer and boiled for 5 min. Then, the samples were separated by SDS-PAGE gel, and blotted onto polyvinylidene fluoride membranes. Blots were probed with anti-PSAT1 (Protein-Tech, Chicago, USA) and anti-beta-actin (Boster, Wuhan, China) antibodies. Blots were developed with secondary goat anti-mouse IgG HRP (Boster, Wuhan, China) and bands were visualized using a chemiluminescent detector.

RESULTS
SWS Strategy Design
Indels generated by the non-homologous end-joining repair mechanism generally occur around DSBs and are often
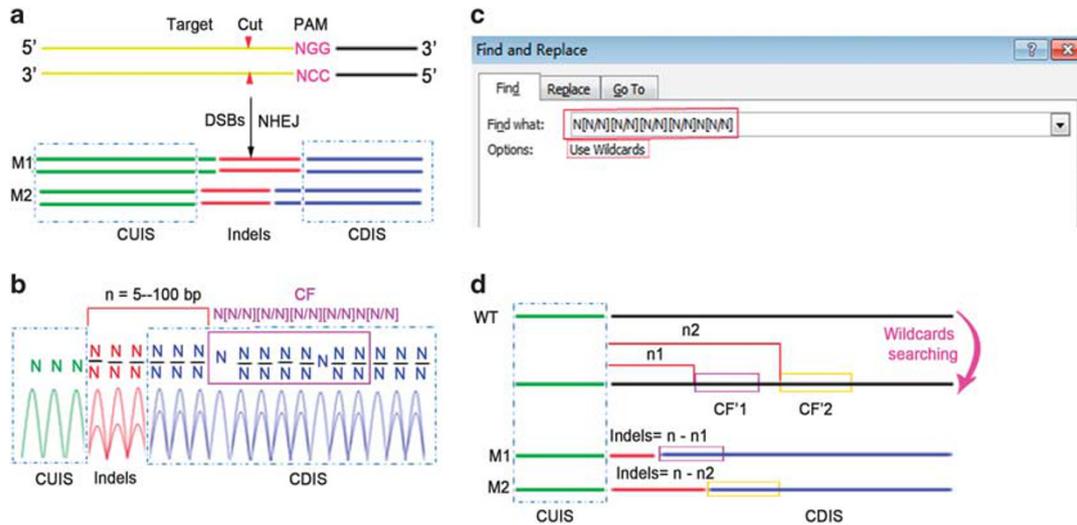
Figure 1 Overview of SWS strategy design. (**a**) A schematic graph showing the patterns of indels induced by CRISPR/Cas9 system. sgRNAs target site (yellow); PAM (5'NGG3', pink); cleavage site cut (red triangle); indels (red bars). CUIS stands for the common DNA sequences upstream of indels, which are identical to corresponding wild-type genomic sequences and shared by both alleles (green bars). CDIS is the common DNA sequences downstream of indels, which are identical to corresponding wild-type genomic sequences and shared by both alleles (blue bars). (**b**) A graph showing the analysis of Sanger sequencing graph. The Sanger sequencing graph was made by MS drawing software. Green single peaks indicate CUIS; red double peaks indicate putative indels; blue double peaks indicate CDIS. A 7- to 10-bp length CF (N[N/N][N/N][N/N][N/N]N[N/N]) 5- to 100-bp (n) downstream of CUIS is selected for wildcard searching. (**c**) A screenshot showing MS Word wildcards searching of CF (N[N/N][N/N][N/N][N/N]N[N/N]) in wild-type sequence. (**d**) A schematic graph showing the identification of corresponding CF's (CF'1 and CF'2) in wild-type sequence and the calculation of the length of indels.

several to tens of base pairs long.[4,16] For the *Streptococcus pyogenes* Cas9 (spCas9) nuclease, DSBs occur about 3-bp (base pair) upstream of PAM.[11] The common DNA sequences flanking each side of indels, which are identical to corresponding wild-type genomic sequences and shared by both alleles, are referred to as CUIS or CDIS, respectively, (Common Upstream/Downstream Identical Sequence) (Figure 1a). In a single Sanger sequencing reaction, a template from a single diploid genome containing indels in specific loci will generate double peaks in the sequencing graph because indels are random (Figure 1b). In most cases, the double peaks begin at the end of a CUIS. In every position, these double peaks represent two defined nucleotides in each allele. We hypothesized that if we positioned a certain fragment (referred to as CF) within a CDIS in the sequencing graph back to the wild-type sequence (Figure 1c), then the exact length of indels can be deduced. The corresponding fragment in the wild-type sequence is referred to as CF' (Figure 1d). As the CF in the sequencing graph was derived from two different alleles, two corresponding CF's (CF'1 and CF'2) were identified in the wild-type sequence (Figure 1d). Considering the limited length of the PCR products used for sequencing, determination of even a small CF is enough to predict the position of CF's. This can be achieved using the wildcard search function in MS Word that allows searches for seven wildcards plus defined letters in a given sequence (Figure 1c). As shown in Figure 1d, by comparing the distance between a CF and CUIS in the sequencing graph (referred

to as n, Figure 1b) and the distance between each CF' and CUIS in the wild-type sequence (referred to as n1 or n2, Figure 1d), we can accurately deduce the lengths of indels in each allele.

Verification of the SWS Strategy with Defined Sequences
As a first step to determine if the SWS strategy was practicable, we tested it with two DNA sequences containing known indels (Figure 2a). We chose two DNA fragments derived from the coding region of the green fluorescent protein (*GFP*) gene that had been interrupted by Cas9 and repaired by non-homologous end joining (one with a 1-bp deletion; the other with a 29-bp deletion). These two fragments were mixed together in equal amounts to imitate a diploid genome that contained indel mutations in both alleles (Figure 2b). Sanger sequencing of this mixer generated double peaks beginning at the end of the CUIS, as expected (Figure 2b). By searching for a 7-bp *GFP*-CF (G[C/T][A/G][G/C][A/T][A/G]C) that was 14-bp downstream from the CUIS in the sequencing graph with the wildcard search function, we successfully positioned the CF into the *GFP* wild-type sequence and found two corresponding *GFP*-CF's (*GFP*-CF'1 GCAGAAC; *GFP*-CF'2 GTGCTGC) (Figure 2c). As shown in Figures 2b and c, the distance between the *GFP*-CF and CUIS in the sequencing graph was 14 bp ($n = 14$ bp) and the distances between each *GFP*-CF' and the CUIS in the wild-type sequence were 15 bp and 43 bp ($n1 = 15$ bp; $n2 = 43$ bp). Therefore, the deduced
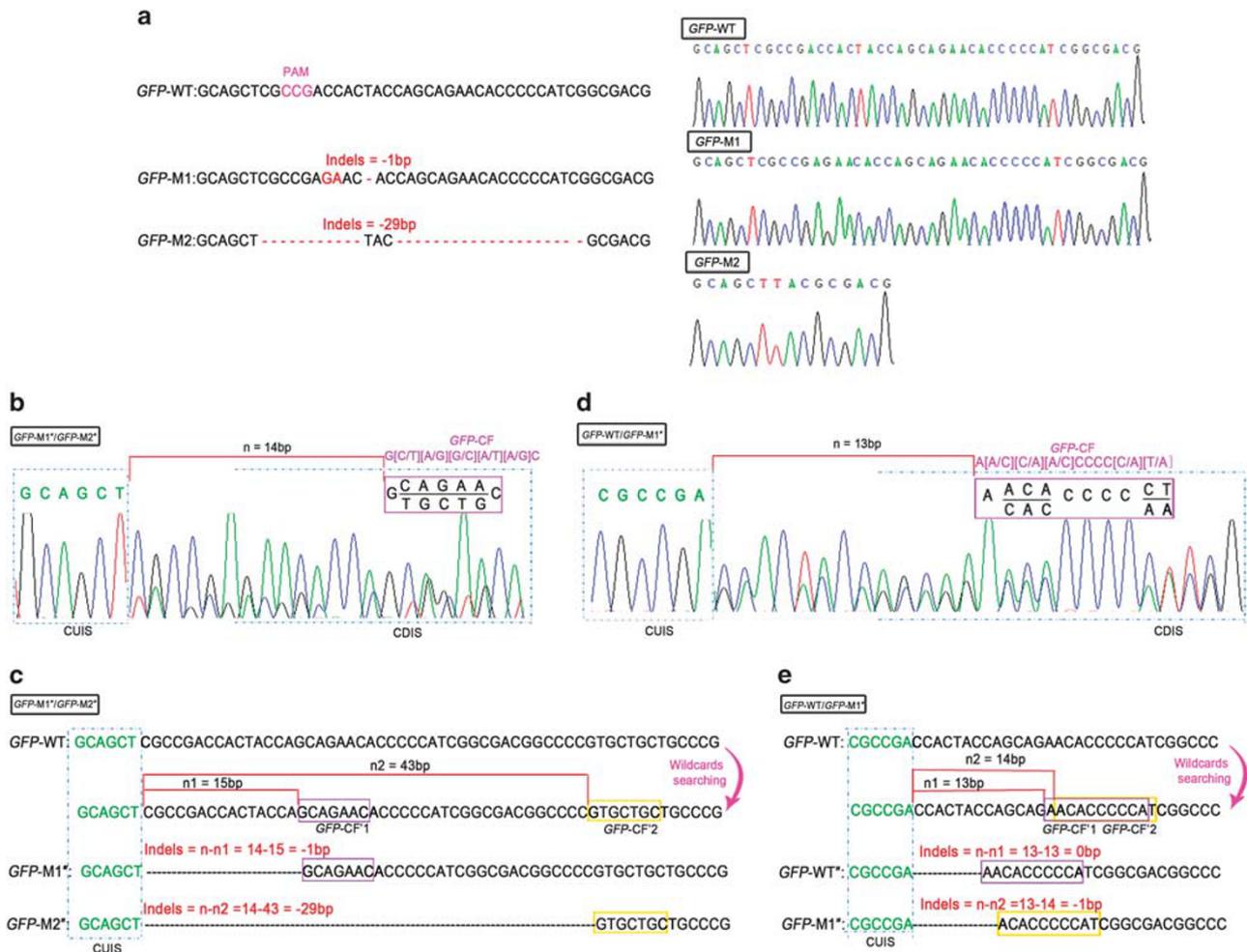
Figure 2 Verification of SWS strategy using sequences with defined indels. (**a**) Sequences of DNA fragments from wild-type or mutant *GFP*. *GFP*-M1 contained 1- bp deletion (−1 bp); *GFP*-M2 contained 29- bp deletion (−29 bp). (**b**, **c**) Analysis of sequencing graph of *GFP*-M1 and M2 mixture. In the sequencing graph, a *GFP*-CF (G[C/T][A/G][G/C][A/T][A/G]C) 14-bp downstream of CUIS were used for searching (**b**). Two corresponding CF's (*GFP*-CF'1 GCAGAAC; *GFP*-CF'2 GTGCTGC) were found in *GFP* wild-type sequence and the distances between each CF' and CUIS were 15 bp (n1) and 43 bp (n2) (**c**). The deduced numbers of indels were n–n1 = 14–15 = − 1 bp (*GFP*-M1) and n–n2 = 14–43 = − 29 bp (*GFP*-M2). (**d**, **e**) Analysis of sequencing graph of *GFP*-WT and M1 mixture. In the sequencing graph, a *GFP*-CF (A[A/C][C/A][A/C]CCCC[C/A][T/A]) 13-bp downstream of CUIS were used for searching (**d**). Two corresponding *GFP*-CF' (*GFP*-CF'1 AACACCCCCA; *GFP*-CF'2 ACACCCCCAT) were found in *GFP* wild-type sequence and the distance between each CF' and CUIS were 13 bp (n1) and 14 bp (n2) (**e**). The deduced numbers of indels were n–n1 = 13–13 = 0 bp (*GFP*-WT) and n–n2 = 13–14 = − 1 bp (*GFP*-M1).

numbers of indels within each allele were n–n1 = 14–15 = − 1 bp and n–n2 = 14–43 = − 29 bp, which are identical to the actual indel numbers. When we chose another *GFP*-CF (C[G/A][A/C][C/T][G/A][G/C]C[C/T]) that was 32-bp downstream of the CUIS, we got the same prediction (Supplementary Figures 1A and B). Next, we mixed one mutant *GFP* sequence with one wild-type sequence to imitate a heterozygous mutation. One allele was predicted to contain a 1-bp deletion (or 29-bp deletion) and one allele was predicted to contain no indels, as shown in Figures 2d and e and Supplementary Figures 1C and D. Taken together, these data demonstrated the accuracy and reliability of the SWS strategy.

Screening Single-Cell Clones Harboring Homozygous Frameshift Mutant from Cas9-Treated Cells by SWS Strategy

Encouraged by the verification results, we determined if the SWS strategy could predict unknown indels in a diploid genome from single-cell clones. We used the human esophageal cancer cell line eca-109 as a model, and designed three non-overlapping sgRNAs targeting human *PSAT1* and cloned them into the pX459 vector (Figure 3a). The pX459 vector expressed the sgRNA, humanized *Streptococcus pyogenes* (hSpCas9), and a 2A linked puromycin-*N*-acetyltransferase, which enabled puromycin selection of positively transfected cells. To test which sgRNA could produce targeted
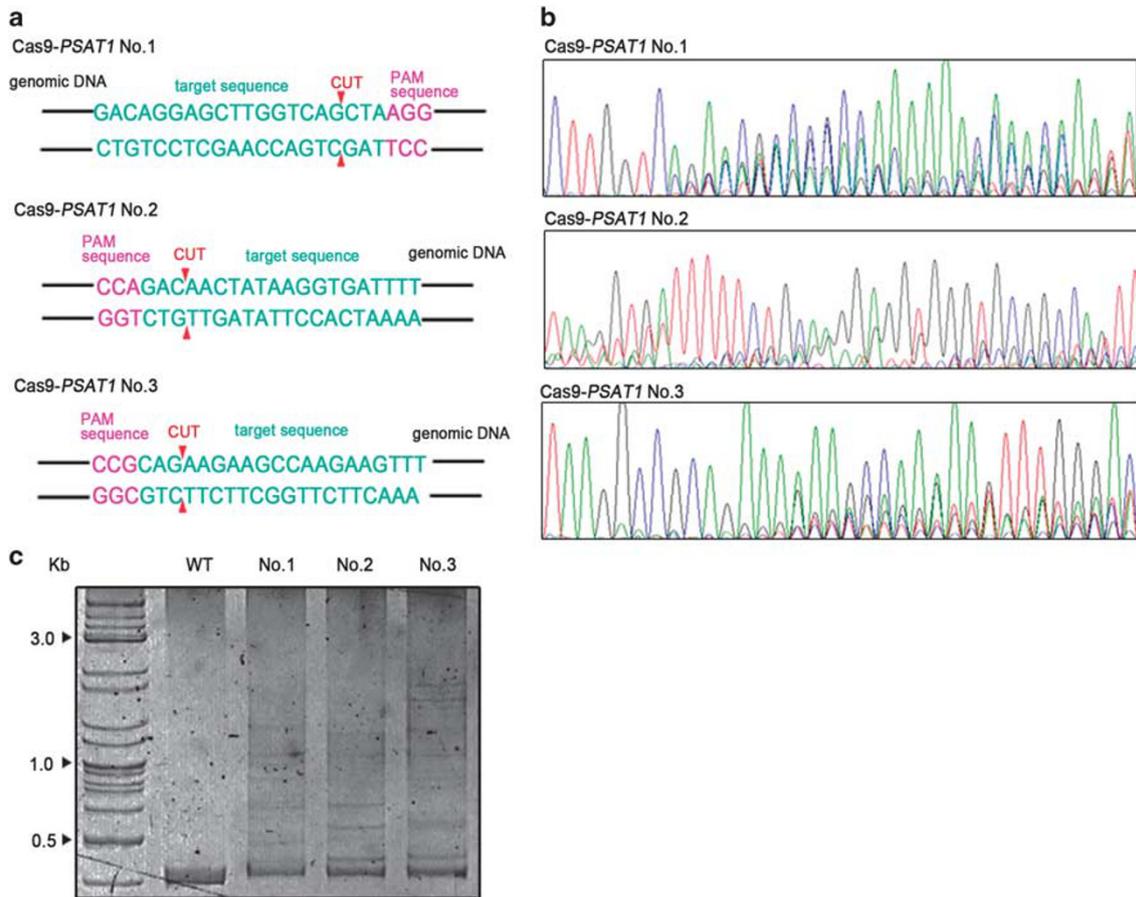
Figure 3 Screening of effective sgRNAs against human *PSAT1* gene. (**a**) A schematic graph showing the design of sgRNAs targeting human *PSAT1* gene. (**b, c**) Efficiency of *PSAT1* sgRNAs analyzed with heteroduplex mobility assay (**b**) and DNA Sanger sequencing (**c**).

cleavage and induce indel mutations, we transfected the three plasmids separately into eca-109 cells. Then, we extracted genomic DNA from a portion of the transfected cells in each clone for PCR amplification with *PSAT1*-specific primers. The PCR products were analyzed by heteroduplex mobility assay and sequenced. All three sgRNAs (sgRNA1, 2, 3) were found to be effective, as evidenced by the presence of bands with limited mobility in the heteroduplex mobility assay and composite sequence traces in the sequencing graphs (Figures 3b and c).

We chose the sgRNA1- and sgRNA3-treated cells for further single-cell clone analysis. We seeded single cells into 96-well plates by serial dilution. After 4 weeks, these clones were harvested for genomic DNA extraction. The genomic DNA was amplified by PCR and analyzed with Sanger sequencing. Among the 32 clones analyzed, we identified 12 that were single clones and harbored indels in at least one allele, and five of them contained biallelic frameshift mutations (Table 2).

Figure 4a shows the sequencing graph for one of the sgRNA1-derived single-cell clones (clones 1–9). We chose an

8-bp *PSAT1*-CF ([A/G]G[A/C][C/A]G[A/C]AG) that was 6-bp downstream of the CUIS and mapped it back to the wild-type sequence with wildcard searching, and two corresponding *PSAT1*-CF's (*PSAT1*-CF′1 GGCCGCAG; *PSAT1*-CF′2 AGAA-GAAG) were identified (Figure 4b). By comparing the distance (n) between the *PSAT1*-CF and CUIS in the sequencing graph and the distances (n1 and n2) between each *PSAT1*-CF′ and CUIS in the wild-type sequence, we deduced that both alleles of this clone contained frameshift indels; one allele contained a 2-bp insertion (indel = n−n1 = 6−4 = 2) and the other contained a 4-bp deletion (indel = n−n2 = 6−10 = −4) (Figure 4b). Another *PSAT1*-CF ([A/G]A[C/A]TAT[C/A][A/G][A/T]T) 33-bp downstream of the CUIS gave the same prediction (Supplementary Figures 2A and B). The Sanger sequencing and analysis for one of the sgRNA3-derived single-cell clones (clones 3–4) predicted that this clone contained biallelic frameshift indels (Figures 4d and e). The presence of biallelic frameshift mutations in these two clones was further confirmed by western blot analysis (Supplementary Figure 2C).

To demonstrate the accuracy of these predictions, we cloned the PCR products into a blunt-end ligation vector

Table 2 Indels verification in *PSAT1*

Code	Predicted by SWS (bp)	Confirmed single-allele sequence by Sanger sequencing of individual bacterial colony	Gene type	Predicted by CRISP-ID (bp)
1	WT	GGTGACAGGAGCTTGGTCAGCTAAGGCCGCAGAAGA		
1–4	+1	GGTGACAGGAGCTTGGTCAGGCTAAGGCCGCAGAAGA	Frameshift/in-frame	Identical
	− 3	GGTGACAGGAGCTTGGTC---TAAGGCCGCAGAAGA		Identical
1–8	− 3	GGTGACAGGAGCTTGGTC---TAAGGCCGCAGAAGA	In-frame/frameshift	Failed to make prediction
	− 13	GGTGACAGGAGCTTGGTC---T---------AAGA		
1–9	+2	GGTGACAGGAGCTTGGTCAGAGCTAAGGCCGCAGAAGA	Frameshift/frameshift	Identical
	− 4	GGTGACAGGAGCTTGGTCAG----GGCCGCAGAAGA		Identical
1–13	− 9	GGTGACAGG--------AGCTAAGGCCGCAGAAGA	In-frame/frameshift	Identical
	− 10	GGTGACAGGA----------CTAAGGCCGCAGAAGA		Identical
1–14	+2	GGTGACAGGAGCTTGGTCAGAGCTAAGGCCGCAGAAGA	Frameshift/frameshift	Identical
	− 2	GGTGACAGGAGCTTGGTCAG–AAGGCCGCAGAAGA		Identical
1–18	− 17	GGTGACAGGAGCTT----------------GAAGA	Frameshift/in-frame	Identical
	− 24	GGTGA---------------------CAGAAGA		Identical
3	WT	CAGCTAAGGCCGCAGAAGAAGCCAAGAAGTTTGGG		
3–4	+1	CAGCTAAGGCCGCAGAAAGAAGCCAAGAAGTTTGGG	Frameshift/frameshift	Identical
	− 14	CAGCTAAGGCCGCAG--------------TTTGGG		Identical
3–7	+3	CAGCTAAGGCCGCAGAACAAGAAGCCAAGAAGTTTGGG	In-frame/ in-frame	Identical
	− 3	CAGCTAAGGCCGCAG---AAGCCAA GAAGTTTGGG		Identical
3–9	+2	CAGCTAAGGCCGCAGAAAAGAA GCCAAGAAGTTTGGG	Frameshift/frameshift	Identical
	− 14	CAGCTAAGGCCGCAG-------------TTTGGG		Identical
3–11	+1	CAGCTAAGGCCGCAGAAAGAAGC AAGAAGTTTGGG	Frameshift/frameshift	Identical
	− 1	CAGCTAAGGCCGCAGA-GAAGCCAAGAAGTTTGGG		Identical
3–12	0	CAGCTAAGGCCGCAGAAGAAGCCAAGAAGTTTGGG	WT/frameshift	Identical
	− 26	CAG-----------------------TTTGGG		Identical
3–19	− 1	CAGCTAAGGCCGCA-AAGAAGCCAAGAAGTTTGGG	Frameshift/in-frame	Identical
	−9	CAGCT-TGGTC---------AGCCAAGAAGTTTGGG		Identical

Abbreviations: +: insertion; −: deletion; Blue: sgRNA; Green: PAM; Red: indels; WT: wild type.

(pZero-blunt) and sequenced the resulting bacterial colonies (cloning and sequencing strategy). The sequencing results were compared with wild-type sequences to obtain actual status of the indels. In all cases, the actual numbers of indels determined by cloning and sequencing strategy were identical to the ones predicted by the SWS strategy (Figures 4c and f and Table 2).

To compare the SWS strategy with the recently published CRISP-ID online tool, we analyzed the same sequencing graphs using CRISP-ID. We found that CRISP-ID produced the same predictions as SWS in all cases, except for the clones 1–8 (Table 2 and Supplementary Figures 3A–E). CRISP-ID failed to make a prediction to this clone. Several non-overlapping double peaks were present in the sequencing graph of clones 1–8, indicating that CRISP-ID may require high-quality sequencing results.

DISCUSSION

CRRIPR/Cas9 system provides a powerful approach for the rapid generation of gene knock-out cell lines or model organisms. However, current methods to characterize indels within the resultant clones are time-consuming and laboratory intensive. In this work, we developed a simple but efficient strategy to identify single-cell clones harboring biallelic frameshift mutations. Our strategy requires only two simple steps: Sanger sequencing and MS Word wildcard searching. We showed that by analyzing double peaks in the sequencing graph with the wildcard search function of MS Word, we could predict exactly how many nucleotides were deleted or inserted in each allele of diploid genomes. Compared with CRISP-ID, the SWS strategy is a little inconvenient and cannot be used to predict indels in triploid genomes. However, it is not
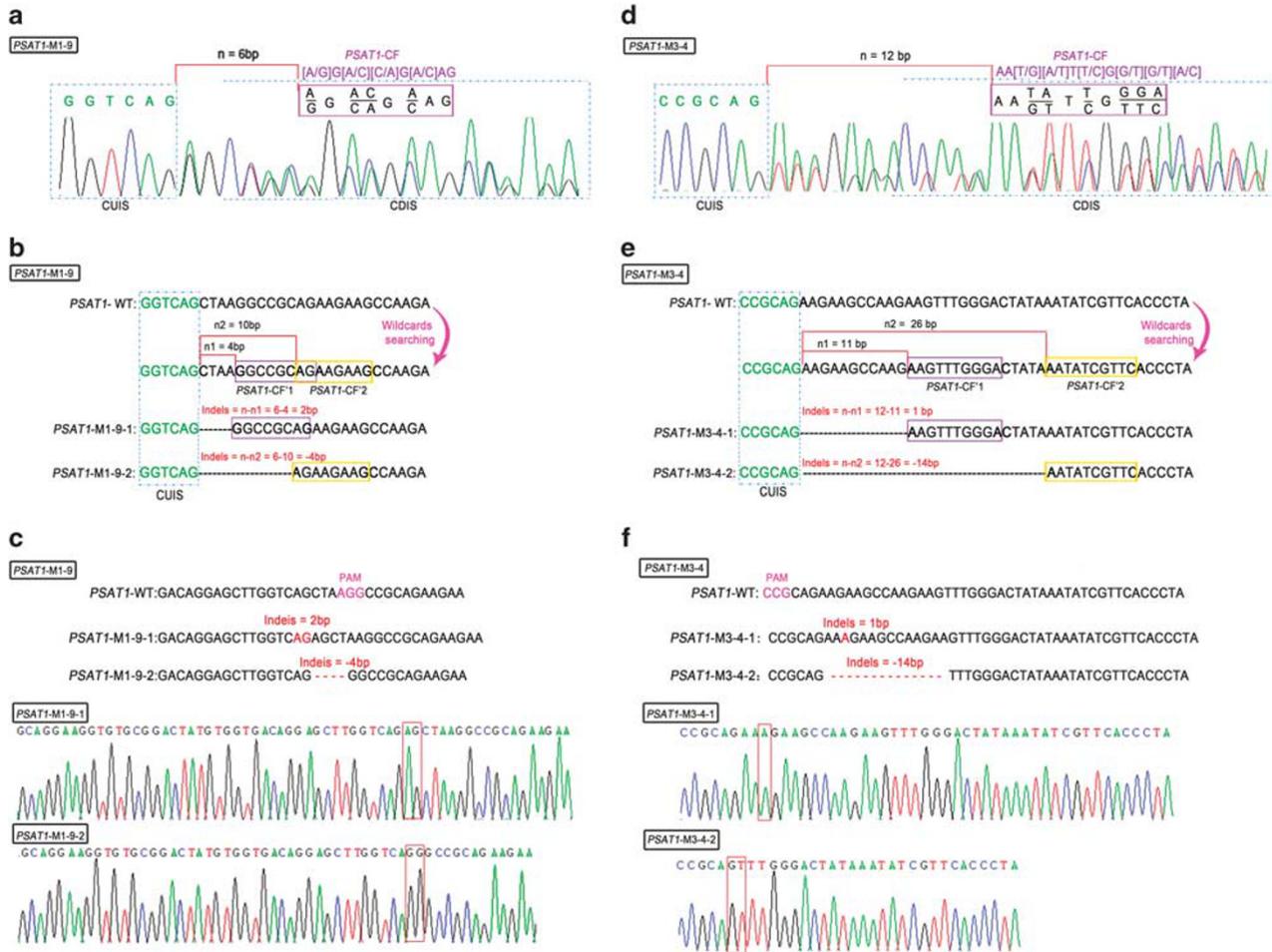
Figure 4 Screening single-cell clones harboring homozygous frameshift mutant from pooled mammalian cells with SWS. (**a**, **b**) Analysis of sequencing graph of a single clone from sgRNA1- treated cells (clones 1–9). In the sequencing graph, a *PSAT1*-CF ([A/G]G[A/C][C/A]G[A/C]AG) 6-bp downstream of CUIS were used for searching (**a**). Two corresponding *PSAT1*-CF′ (*PSAT1*-CF′1 GGCCGCAG; *PSAT1*-CF′2 AGAAGAAG) were identified in *PSAT1* wild-type sequence and the distance between each *PSAT1*-CF′ and CUIS were 4 bp (n1) and 10 bp (n2) (**b**). The deduced number of indels in one allele was n–n1 = 6–4 = 2 bp; the one in the other allele was was n–n2 = 6–10 = − 4 bp. (**c**) The actual indels of each allele in clones 1–9 determined by cloning and sequencing strategy. PCR products from clones 1–9 were cloned into a blunt-end ligation vector (pZero-blunt) and the resulting single colonies were sequenced. Compared with the wild-type allele, one mutant allele had 2-bp insertion and the other had 4-bp deletion. (**d**, **e**) Analysis of sequencing graph of a single clone from sgRNA3-treated cells (clones 3–4). In the sequencing graph, a *PSAT1*-CF (AA[T/G][A/T]T[T/C]G[G/T][G/T][A/C]) 12- bp downstream of CUIS were used for searching (**c**). Two corresponding *PSAT1*-CF′ (*PSAT1*-CF′1 AAGTTTGGGA; *PSAT1*-CF′2 AATATCGTTC) were found in *PSAT1* wild-type sequence and the distances between each *PSAT1*-CF′ and CUIS were 11 bp (n1) and 26 bp (n2) (**d**). The deduced number of indels in one allele was n–n1 = 12–11 = 1 bp; the one in the other allele was n–n2 = 12–26 = − 14 bp. (**f**) The actual indels of each allele in clones 3–4 determined by cloning and sequencing strategy. PCR products from clones 3–4 were cloned into pZero-blunt and the resulting single colonies were sequenced. Compared with the wild-type allele, one mutant allele had 1-bp insertion and the other one had 14-bp deletion.

fastidious about the quality of the sequencing graph, and thus may be more reliable. We believe that the SWS strategy will be a useful tool for genotyping of cell lines and model organisms that contain diploid genomes.

DISCLOSURE/CONFLICT OF INTEREST
The authors declare no conflict of interest.

1. Urnov FD, Rebar EJ, Holmes MC, *et al*. Genome editing with engineered zinc finger nucleases. Nat Rev Genet 2010;11:636–646.
2. Bogdanove AJ, Voytas DF. TAL effectors: customizable proteins for DNA targeting. Science 2011;333:1843–1846.
3. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell 2014;157:1262–1278.
4. Le Cong FAR, Lin* DCS, Barretto R, *et al*. Multiplex genome engineering using CRISPR/Cas systems. Science 2013;339:819–823.
5. Chu VT, Weber T, Wefers B, *et al*. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol 2015;33:543–548.

6. Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem 2010;79:181–211.

7. Li T, Yang B. TAL effector nuclease (TALEN) engineering. Methods Mol Biol 2013;978:63–72.

8. Nishimasu H, Cong L, Yan WX, *et al*. Crystal structure of *Staphylococcus aureus* Cas9. Cell 2015;162:1113–1126.

9. Nishimasu H, Ran FA, Hsu Patrick D, *et al*. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 2014;156:935–949.

10. Ma Y, Zhang L, Huang X. Genome modification by CRISPR/Cas9. FEBS J 2014;281:5186–5193.

11. Jinek M, Chylinski K, Fonfara I, *et al*. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012;337:816–821.

12. Ran FA, Cong L, Yan WX, *et al*. *In vivo* genome editing using *Staphylococcus aureus* Cas9. Nature 2015;520:186–191.

13. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 2014;32:347–355.

14. Cox DB, Platt RJ, Zhang F. Therapeutic genome editing: prospects and challenges. Nat Med 2015;21:121–131.

15. Kang Y, Zheng B, Shen B, *et al*. CRISPR/Cas9-mediated Dax1 knockout in the monkey recapitulates human AHC-HH. Hum Mol Genet 2015;24:7255–7264.

16. Xue W, Chen S, Yin H, *et al*. CRISPR-mediated direct mutation of cancer genes in the mouse liver. Nature 2014;514:380–384.

17. Wu Y, Liang D, Wang Y, *et al*. Correction of a genetic disease in mouse via use of CRISPR-Cas9. Cell Stem Cell 2013;13:659–662.

18. Liang P, Xu Y, Zhang X, *et al*. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. Protein Cell 2015;6:363–372.

19. Chuan Yu YZ, Yao S, Wei Y. A PCR based protocol for detecting indel mutations induced by TALENs and CRISPR/Cas9 in Zebrafish. PLoS ONE 2014;9:e98282.

20. Cho SW, Kim S, Kim JM, *et al*. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 2013;31:230–232.

21. Bassett Andrew R, Tibbit C, Ponting Chris P, *et al*. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. Cell Rep 2014;6:1178–1179.

22. Zhu X, Xu Y, Yu S, *et al*. An efficient genotyping method for genome-modified animals and human cells generated with CRISPR/Cas9 system. Sci Rep 2014;4:6420.

23. Liao J, Karnik R, Gu H, *et al*. Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nat Genet 2015;47:469–478.

24. Bohne F, Langer D, Martiné U, *et al*. Kaiso mediates human ICR1 methylation maintenance and H19 transcriptional fine regulation. Clin Epigenet 2016;8:47.

25. Dehairs J, Talebi A, Cherifi Y, *et al*. CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing. Sci Rep 2016;6:28973.

26. Masago K, Fujita S, Muraki M, *et al*. Next-generation sequencing of tyrosine kinase inhibitor-resistant non-small-cell lung cancers in patients harboring epidermal growth factor-activating mutations. BMC Cancer 2015;15:908.