Using Sanger sequencing to facilitate CRISPR- and TALEN-mediated genome editing workflows

In this application note, we show:

- Sanger sequencing by capillary electrophoresis can be used to determine the efficiency of genome editing in primary transformed cultures
- Sanger sequencing is an efficient method to confirm successful genome edits in transformed cultures, as well as screen secondary clones for successful editing events
- Applied Biosystems™ Minor Variant Finder Software can be used to determine the frequency of SNP changes in clones isolated from secondary cultures

Introduction

Ever since the double-helical structure of DNA was elucidated, researchers have developed techniques for manipulating DNA sequences. However, directing precise sequence changes at user-defined sites has remained a difficult and tedious challenge. Limited successes have been achieved with oligonucleotides, small molecules, or self-splicing introns, but the development of site-directed zinc finger nucleases (ZFNs) and TAL effector nucleases (TALENs) have facilitated sequence-specific manipulations. Nevertheless, difficulties of protein design, synthesis, and validation have slowed adoption of these engineered nucleases for routine use. The most recent gene editing technology, the CRISPR-Cas9 system, largely overcomes these difficulties [1]. In fact, the CRISPR-Cas9 system has proven to be so easy and inexpensive that one investigator has stated that it has brought about the “democratization of gene targeting” [2]. Thus, the CRISPR-Cas9 system is poised to transform genome editing.

CRISPR-Cas9 technology is derived from a bacterial adaptive immune system. It is a two-component system that depends on an enzyme (Cas9) to cleave double-stranded DNA, and a guide RNA (gRNA) that directs the enzyme to the correct location in the genome. If a repair template is not provided, the break produced by the enzyme is repaired by an error-prone nonhomologous end joining (NHEJ) process. This results in a heterogeneous population of cells with different insertions or deletions (indels) around the user-defined break. This process can be exploited to generate cell lines with a specific gene knocked out. Alternatively, if a repair sequence is provided, the sequence around the break can be repaired using the repair sequence as a template. By selecting the sequence for the repair template, precise sequence changes can be introduced at a user-defined locus within the genome.

However, to obtain a clonal population with a homologous genome edit, several clones from the primary transformed pool of cells need to be screened. This necessitates two rounds of screening. First, a primary screen must be performed to determine the relative fraction of cells containing an edit. Knowing the efficiency of the edit will determine the number of single-cell clones that will need to be isolated for expansion. Next, a secondary screen must be performed to identify the clones derived from a single cell that have the desired edit. Sanger sequencing by capillary electrophoresis (CE) can provide information at both screening stages. Sanger sequencing has been the gold standard for sequence determination for several years due to its simple, cost-effective workflow and uncomplicated data analysis. The data produced by CE allows unambiguous identification of sequence changes as well as detection of mixed single-nucleotide polymorphisms (SNPs) in a population. For these reasons, Sanger sequencing by CE can be a valuable part of any genome editing workflow.

Workflow overview

Thermo Fisher Scientific has integrated all the tools necessary for genome editing and downstream analysis (Figure 1). The Invitrogen™ GeneArt™ design tool facilitates the design and ordering of target-specific gRNAs for CRISPR-mediated genome editing or TALs for TALEN-mediated genome editing. Invitrogen™ transfection reagents offer several options for delivery of genome editing tools into eukaryotic cells. In addition, Invitrogen™ TOPO™ TA cloning vectors and competent cells facilitate the sequence analysis of primary transformants. Gibco™ media is available for growing the primary transformants and secondary cultures following clonal expansion. Finally, Applied Biosystems™ sequencing instruments and reagents enable the determination of specific genomic editing events. In this application note, we demonstrate how this workflow comes together to generate and identify mutations in the human hypoxanthine phosphoribosyl transferase (*HPRT*) gene.

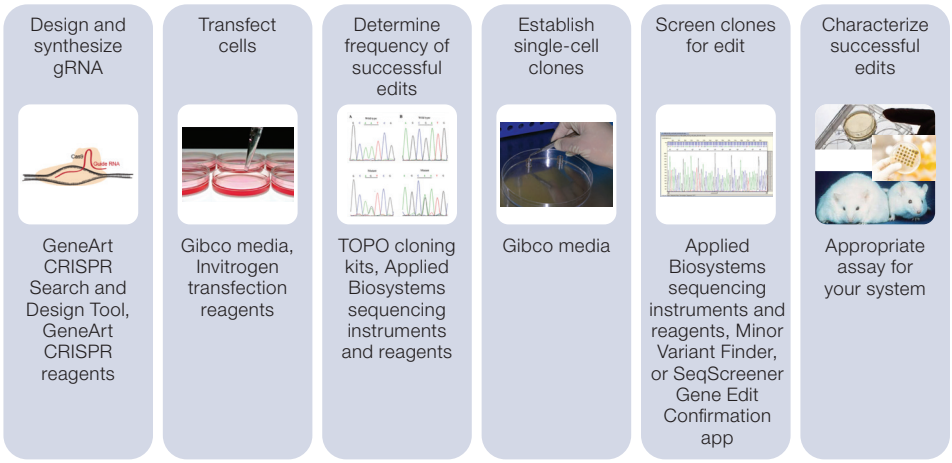

Figure 1. Overall workflow for CRISPR genome editing. Thermo Fisher Scientific provides the tools, reagents, and competence required for success at each step of the workflow.

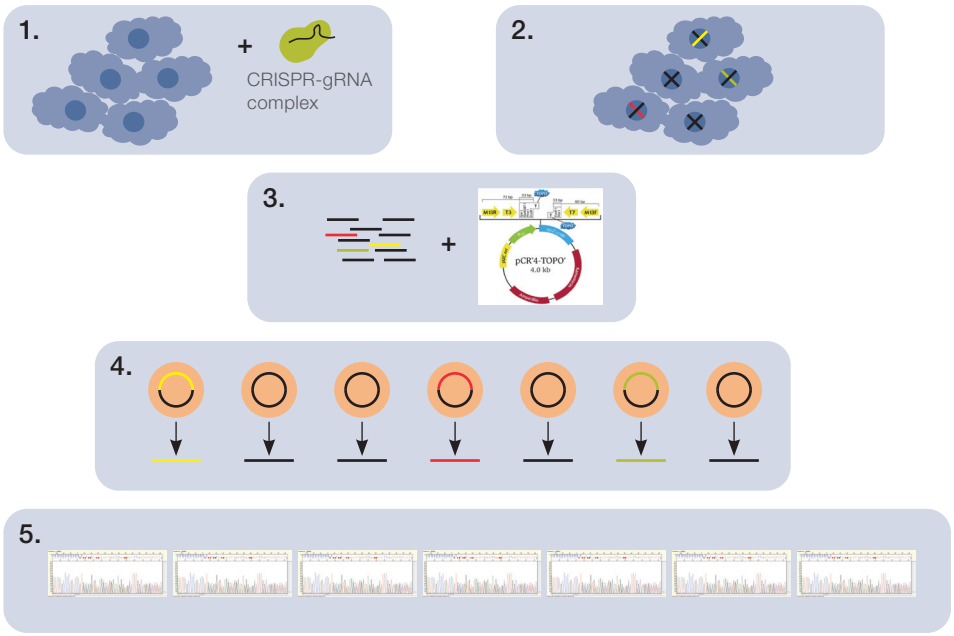

Figure 2. Steps for determining the efficiency of an edit using TOPO cloning and Sanger sequencing by CE. 1. Transfect gRNA and Cas9 mRNA into cells. **2.** Incubate cells to allow processing of genomic change. **3.** Purify genomic DNA from the cell culture, PCR-amplify the engineered locus from the heterogeneous culture, and clone PCR fragments into TOPO vector. **4.** Isolate plasmids from single colonies and PCR-amplify the insert. **5.** Sequence the insert. The efficiency of the edit is the ratio of the number of inserts with an engineered change to the total number of inserts sequenced. Higher efficiency will likely result in fewer secondary clones that need to be screened to identify specific cells with the change.

A brief overview of the steps used to generate and analyze a primary culture with *HPRT* mutations is shown in Figure 2. The target-specific CRISPR RNA (crRNA) sequence within the gRNA was designed to a *HPRT*-specific locus. The gRNA was synthesized via *in vitro* transcription using the Invitrogen™ GeneArt™ Precision gRNA Synthesis Kit. Following synthesis and purification, gRNA was cotransfected with Cas9 mRNA into 293FT cells using Invitrogen™ Lipofectamine™ MessengerMAX™ Transfection Reagent. The cells were harvested 78 hours posttransfection. The cell lysates were then used along with primers flanking the *HPRT* target to generate PCR amplicons no greater than 600 bp in length. The PCR products were then subcloned using the Invitrogen™ Zero Blunt™ TOPO™ PCR Cloning Kit and transformed into Invitrogen™ TOP10 *E. coli* cells. Ninety-six bacterial colonies were picked per transformed pool of gene-edited cells and processed for DNA isolation using the Invitrogen™ PureLink™ 96 Plasmid Purification System and subjected to Sanger sequencing. The resulting sequencing data was then analyzed to measure the percent of PCR products containing accurately edited sequence and to select which clonal isolates to maintain. Alternatively, although it was not performed for this study, the PCR product could be sequenced directly, without subcloning into TOPO cells.

Demonstration of sequencing results in the primary screen

In any genome editing experiment, the nuclease cleavage and repair process is not completely efficient or accurate. Therefore, before moving on to clonal isolation of engineered cells, the fraction of cells containing an edit should be determined. One way to do this is to PCR-amplify the region edited from primary transformant cultures and subclone into a plasmid. By sequencing a large number of plasmids, the fraction containing an edit can be determined. This also gives a first glance into the overall gene knockout or editing efficiencies and type of indel changes that might have occurred.

After transfecting 293FT cells with gRNA and Cas9 mRNA, we subcloned the locus from primary transformants and sequenced 96 clones. Of those 96 clones, 84 aligned with the target sequence. Only 12 clones had no editing event in the amplified region. Seventy-two clones had at least one sequence deviation from the wild-type sequence, for an overall efficiency of about 86%. This gave us a good idea of how many secondary single-cell clones needed to be screened to find a desired pure knockout clone. Interestingly, some of the TOPO™ clones had a mixed sequence (Figure 3). This could be due to either the bacterial colony having two distinct plasmids, or the DNA not being derived from a single colony. Nevertheless, it is clear that an editing event is present. Note that sequence is uniform up to the red arrow. After that, each position consists of two peaks, indicating two different sequences are present. It is not easy to separate the sequences at this level; however, it is clear that CE sequencing can show that an edit occurs at the correct location even if downstream sequences can't be read accurately.

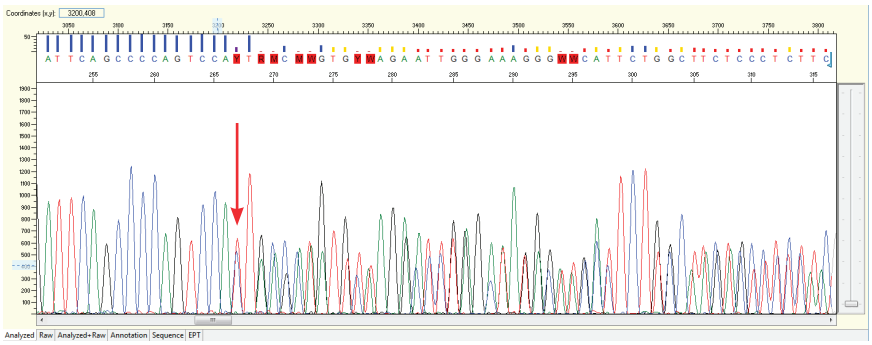

Figure 3. Example CE trace of a mixed clone, containing two different edited sequences. Notice that the sequence is uniform up to the red arrow, after which there are two different sequences present in approximately equal amounts. Because one sequence contains a deletion, it is out of register with the other sequence and can't be easily read.

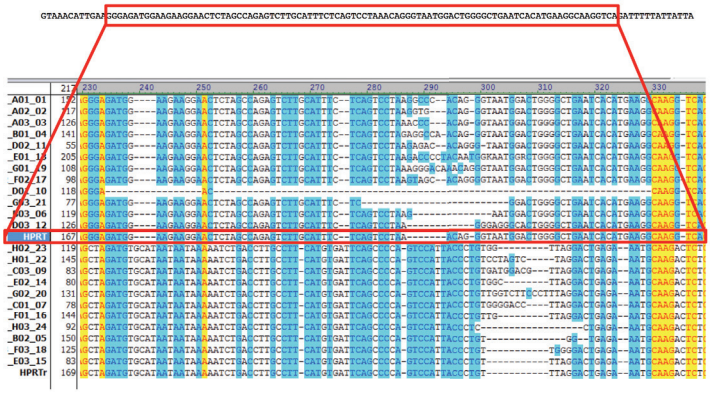

Figure 4. Sequences within the *HPRT* locus produced by NHEJ of DNA following cleavage by the CRISPR-Cas9 system. Genomic editing events can produce a variety of sequence changes, especially in the absence of a repair template. Each line of sequence shown is derived from a different TOPO clone and aligned to show differences. The entire guide RNA sequence used is shown at the top; the boxed sequence emphasizes the regions shown below. The normal *HPRT* locus is labeled in dark blue on the left. Yellow boxes with red font are nucleotides that are identical to wild-type *HPRT* (unchanged); blue and white boxes illustrate nucleotide differences.

Analysis of sequences present in many different colonies revealed the spectrum of changes introduced by the editing complex (Figure 4). Each sequence shown was from a different TOPO clone and represented a different molecule in the primary transformant culture. Deletions and insertions are apparent around the gRNA site and are not confined to a specific base. Since the editing complex can introduce a wide variety of changes, a collection of clones derived from the primary transformant culture should be sequenced to profile and predict what edits may be expected in the secondary screen.

Demonstration of sequencing results in the secondary screen

After generating a primary transformed culture, and while that culture was being characterized, single cells from the heterogeneous primary culture were obtained by limiting dilution. Clones were grown for 14 days, and lysed as described in the Invitrogen™ GeneArt™ Genomic Cleavage Detection Kit manual. Sequences of the locus around the putative edit were PCR-amplified using target-specific primers (forward: 5′-GTGTTAATTTCAAACATCAGCAGC-3′, reverse: 5′-GTCTTCTTGTTTATGGCCTCC-3′). The resulting PCR products were subjected to Sanger sequencing by CE and analyzed using Applied Biosystems™ Sequence Scanner Software v2.

Several cultures derived from a single eukaryotic cell were established and Sanger sequencing was used to analyze the *HPRT* locus in samples. In one case (Figure 5), the edit resulted in an insertion of a single T, which changed the reading frame and presumably resulted in a loss-of-function allele. In this case, because there is a single peak, the cell line is either homozygous for the engineered change, or hemizygous (the homologous locus on the other chromosome was deleted). Two different nucleotide peak traces could be observed if the cell line was heterozygous.

Overlapping traces can also be observed when the culture is not clonally pure. We have developed Applied Biosystems™ Minor Variant Finder Software for quantifying allelic frequencies from CE-generated sequencing traces. This software is useful for determining the fraction of cells in a culture containing a modified allele. If the edit results in an SNP, Minor Variant Finder Software can be used to determine the percentage of cells in culture with the SNP. For example, a culture may have an edit that changed an A to a G, but a fraction of wild-type cells are still present. Minor Variant Finder Software can determine the fraction of loci in the culture containing the change by comparing the sequence of the culture with a homogeneous control sequence (Figure 6). Furthermore, Minor Variant Finder Software is sensitive enough to detect a contaminating sequence among diploid cells to around 10%; that is, 1:10 cells are variant (this would show up as an allele frequency of 5%). However, Minor Variant Finder Software is at this time not suitable for determining the frequency of indels or other genomic rearrangements, since

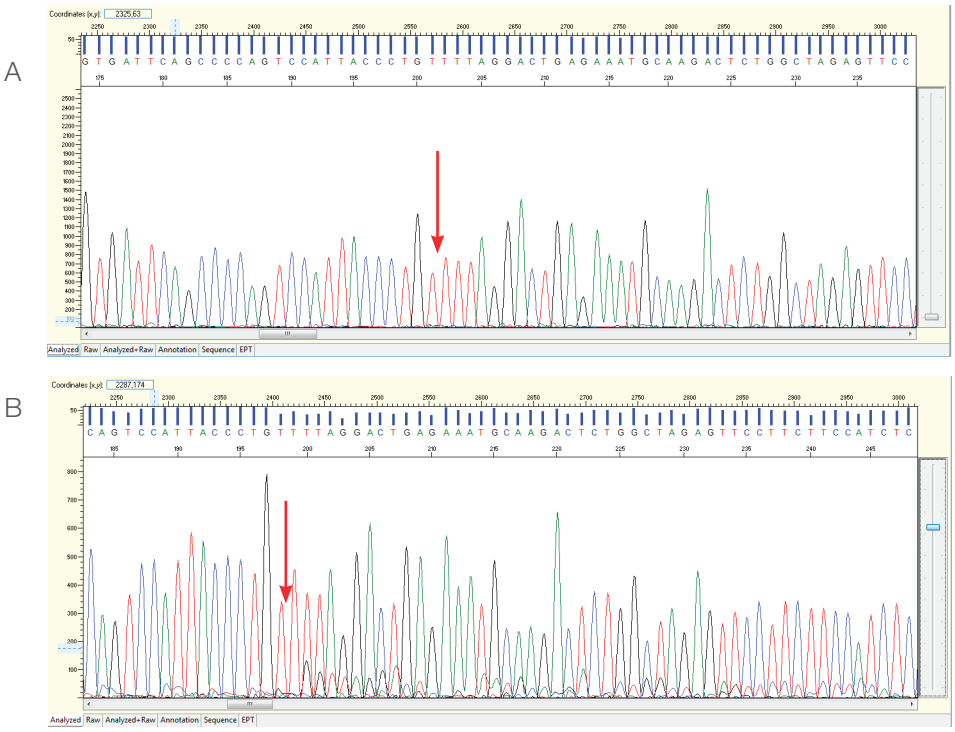

Figure 5. Examples of Sanger sequencing traces from a secondary screen. (A) Sequence is homogeneous and monoclonal. **(B)** Sequence is heterogeneous at the arrow and thus is not derived from a single clone. Note the distinct peaks at each position downstream of the red arrow.

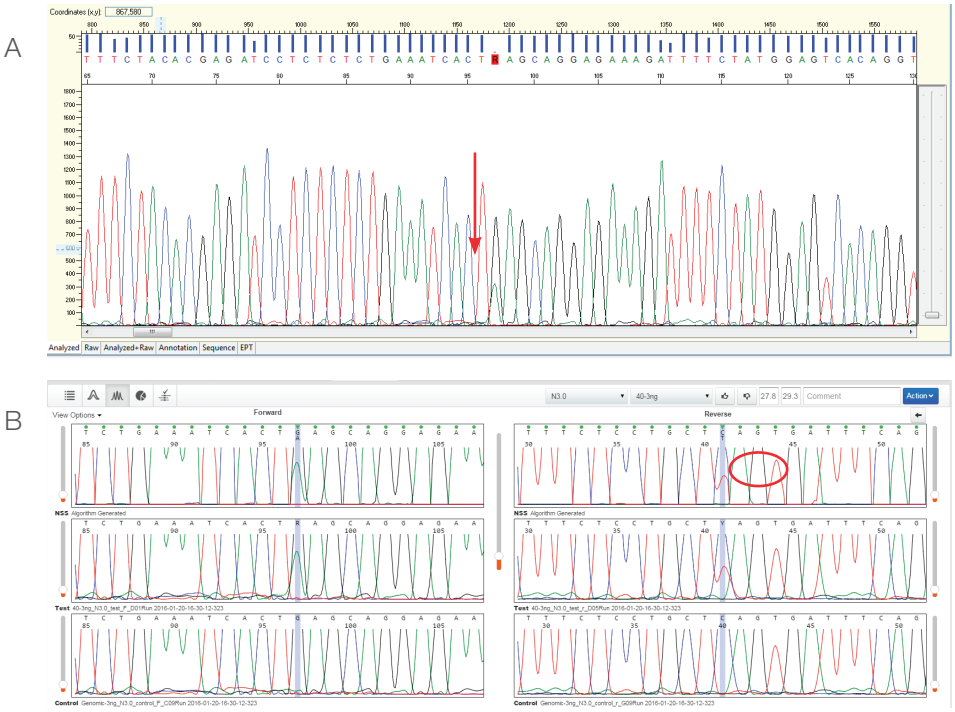

Figure 6. SNP detection and analysis in a secondary clone. (A) Sanger sequencing trace of a heterogeneous clone with an SNP. **(B)** Minor Variant Finder Software analysis and quantification of the frequency of the SNP. In this culture, the SNP is present at a frequency of around 28%.

the deconvolution of the out-of-register or mixed sequences is not yet possible. Nevertheless, the Applied Biosystems™ SeqScreener Gene Edit Confirmation software was designed to facilitate analysis of indels, substitutions, and other rearrangements that might arise from genomic editing workflows. For more information, see reference 3.

applied
biosystems

Conclusions

In this application note, we have shown how Sanger sequencing by capillary electrophoresis and Minor Variant Finder Software can be used in a genome editing workflow. We show the results of a CRISPR-mediated edit, but the principles applied here can also be used for ZFN- or TALEN-mediated editing workflows. The simplicity, cost-effectiveness of the workflow, and uncomplicated data analysis make Sanger sequencing by capillary electrophoresis a valuable part of any genome editing workflow.

References

1. Doudna JA and Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096.
2. Travis J (2015) Making the cut: CRISPR genome editing technology shows its power. *Science* 350:1,456–1,457.
3. https://test.apps.thermofisher.com/apps/help/MAN0019454/GUID-13CF6DAF-343B-4969-9459-E910C7B01B43.html

Find out more at **thermofisher.com/seqscreener**

ThermoFisher
SCIENTIFIC