Genetic analysis tools for genome editing workflows

In this paper we show:

- Why genome editing efficiency is important
- Various genetic analysis methods, each with particular strengths, that can be used to evaluate genome editing efficiency
- Thermo Fisher Scientific's complete workflow solutions for genome editing experiments

Introduction

Genome editing is poised to revolutionize our understanding and control of diverse biological systems. Many different systems have been developed to engineer precise changes to the sequence of a locus, including CRISPR-Cas9, TALEN, and zinc-finger nucleases. The CRISPR-Cas9 system, the easiest, most precise, and most widely adopted genome editing technology, is a system based on a bacterial adaptive immune response. Its two components are Cas9 enzyme to cleave double-stranded DNA and a guide RNA (gRNA) that directs the enzyme to a specific location within the genome. Repair of a double-strand break can occur via multiple mechanisms. In the absence of a repair template, the break produced by the Cas9 enzyme is repaired by an error-prone, nonhomologous end-joining process that results in a heterogeneous population of cells with different insertions or deletions (indels) around the gRNA-defined break. This process can be exploited to generate cell lines with random deletions around a specific sequence, producing a functional knockout. Alternatively, if a repair template is provided, a user-defined sequence change can be introduced at a specific locus within the genome via error-free homologous recombination. This process can be used to overexpress a novel gene, create disease-relevant cell models, or tag endogenous genes with reportable moieties.

In any genome editing experiment, the repair process is not completely efficient or accurate. Therefore, to obtain a clonal population homogeneous for a desired genome edit, independent clones from the primary transformed pool of cells need to be screened. This necessitates two rounds of analysis. 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. For some experimental paradigms, if the efficiency is high enough, a secondary screen of clonally pure cells might not be necessary. For example, if the study involves analyzing the response of cells with a loss of function for a particular gene, having a homogeneous population of a single knockout might not be required. Analysis of a mixed pool where a vast majority of cells have some kind of loss-of-function knockout may provide the desired experimental model.

Therefore, before moving on to clonal isolation of engineered cells, the percent editing efficiency in the transfected cell pool should be determined. Depending on the model system, the efficiency will dictate whether the primary pool can be used as a model, or will dictate the approximate number of single cells that need to be examined to find a desired edit in a secondary screen. In any case, calculating the efficiency of editing is a critical step in any genome editing workflow.

Description of workflows and tools from Thermo Fisher Scientific

We offer complete workflow solutions for genome editing experiments, including target-specific editing tool design, synthesis and delivery reagents, and a wide selection of analytical tools for verifying and validating gene modifications (Figure 1).

Choosing the gRNA sequence

The first step of a genome editing experiment is designing the gRNA. The Invitrogen™ GeneArt™ CRISPR Search and Design Tool allows you to search our database of more than 600,000 predesigned CRISPR gRNAs targeting human and mouse genes. The tool also accepts a FASTA sequence of interest that is queried for *de novo* gRNA designs using our proprietary algorithms. Up to 25 gRNA sequences per gene are provided with recommendations based on potential off-target effects for each CRISPR sequence. Similarly, the Invitrogen™ GeneArt™ TAL Effector Search and Design Tool can be used to find the optimal sequence for TALEN-mediated genome edits. Oligonucleotide templates pre-tagged with T7 promoter sequences, needed for *in vitro* gRNA transcription, can be ordered directly using the design tool. Alternatively, ready-to-use gRNAs can be ordered through the Invitrogen™ TrueGuide™ Synthetic gRNA website. From here, custom and predesigned gRNAs for mouse and human hosts can be ordered. In most cases, it may be beneficial to choose the top 2 or 3 recommended designs and test each gRNA individually for best editing performance. Invitrogen™ TrueGuide™ Synthetic gRNA along with TrueCut™ Cas9 Protein v2 are highly recommended to maintain high cell viability while working with challenging cell types.

Synthesizing gRNA

Once the gRNA sequences are chosen, the gRNAs themselves have to be obtained. For synthesis of gRNAs, we recommend using the Invitrogen™ GeneArt™ Precision gRNA Synthesis Kit. This kit provides a complete system for rapid synthesis of gRNA, starting with the assembly of template DNA using target-specific oligos tagged with the T7 promoter, through an *in vitro* transcription (IVT) reaction, and ending with a rapid purification step that yields transfection-ready gRNA. If you are deploying ready-to-use TrueGuide Synthetic gRNAs, these synthesis steps can be bypassed. The gRNA is then complexed with Cas9 nuclease and introduced into target cells using one of a variety of methods. For example, the nuclease can be supplied in DNA format using Invitrogen™ GeneArt™ CRISPR Nuclease Vector, mRNA format using

Genome editing workflow and our products

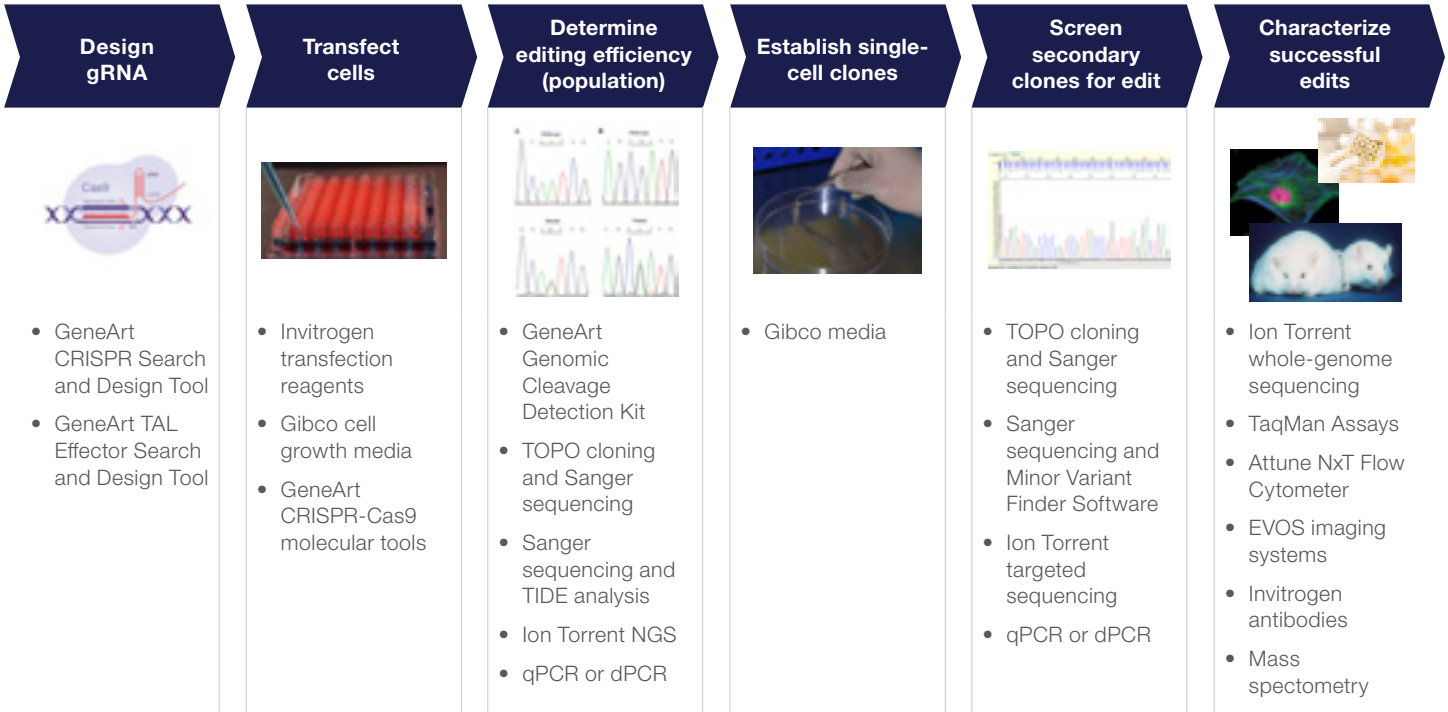


Figure 1. Thermo Fisher Scientific portfolio for genome editing. From gRNA design and cell transfection reagents through genetic analysis tools and phenotype analysis solutions, we have a complete set of tools comprising optimized, validated systems that have been designed to work together to facilitate genome editing projects.

Invitrogen™ GeneArt™ CRISPR Nuclease mRNA, or ready-to-use protein format. For all transient transfections we recommend using our purified recombinant Invitrogen™ TrueCut™ Cas9 Protein v2. Finally, arrays of predesigned gRNAs are available as lentivirus stocks for functional genomic screening. For a complete description of available CRISPR-Cas9 formats and recommendations based on cell type, please refer to **thermofisher.com/genomeedit**

Introducing gRNA and Cas9 into cells

The Cas9–gRNA complexes can be introduced into cell lines by lipid-mediated transfection using Invitrogen™ Lipofectamine™ CRISPRMAX™ Cas9 Transfection Reagent. Alternatively, the Invitrogen™ Neon™ Transfection System can be used to introduce the complexes into cells by electroporation. The Neon system is highly recommended for maximizing editing efficiency in hard-to-transfect cells. Once transfected, the cells are allowed to recover. A pool of primary transformants, representing a variety of cleavage events centered around the gRNA, is then used to determine the editing efficiency.

Determining fraction of cells containing an edit

Before advancing to labor-intensive and expensive experiments, the fraction of cells in the pool of primary transformants with a successful editing event needs to be determined. Some of the most common methods using genetic analysis are described in the following sections. While all of the methods can be used to give quantitative information for editing efficiency, some methods have advantages over others in terms of the nature of the desired edit, throughput, and workflow. Other phenotypic methods that take advantage of the specific gene being edited, such as enzyme activity, can also be used but are not considered here. Once the fraction containing an edit is determined, the next steps can be considered. For example, it might be necessary to obtain genetically homogeneous clones of cells, where the exact sequence of the edit is known, for further studies. Alternatively, if the fraction of cells containing the edit in the primary pool is sufficiently high, studies can be performed on the mixed pool without clonal isolation. On the other hand, if the editing efficiency is low, it might be advantageous to start over with a different gRNA or transfection method.

Isolating homogeneous clones of edited cells

The number of single clones needing to be screened to obtain a clonal cell line with the desired knockout can be estimated based on editing efficiency and expected viability. For example, if a homozygous knockout with mutations in both copies of a gene is desired, and the resulting cleavage detection efficiency is 50%, then the probability of having both alleles knocked out in any single cell is 25% (0.5 x 0.5). Since the probability of an indel leading to a frameshift is 2/3, the chance of having a homozygous knockout of that indel is 10–11% per cell (0.5 x 0.5) x (0.66 x 0.66). For isolating single cells by limiting dilution, 0.8 cells per well of a 96-well plate is recommended as a target. If it is assumed that only 20% of cells are expected to survive, and at least ten 96-well plates are plated in this manner, only 192 cells are expected to survive. Therefore, based on the example of editing efficiency provided, the number of homozygous knockout clones in the surviving cells is expected to be 19–21 cells (192 x 10–11%). Please note that certain cell types will require specific media conditions for maintaining clonal cell viability.

Validating secondary clones for edited sequence

Since secondary cultures of cells grown from a single isolated cell are expected to be genetically homogeneous, the method for determining if the edit was successful can be more varied. In these cases, anything from Applied Biosystems™ TaqMan® SNP Genotyping Assays to phenotypic analysis can be used, again depending on factors such as the nature of the edit, workflow and throughput needs, and cost. Some considerations for performing secondary screens are described in the final section of this paper.

Determining editing efficiency in pools of primary transformants

Genomic cleavage detection on agarose gels

A commonly used method for determining editing efficiency makes use of the Invitrogen™ GeneArt™ Genomic Cleavage Detection (GCD) Kit. In this method, lysates of primary pools of cells are amplified using locus-specific primers. The resulting PCR products are then denatured and reannealed so that mismatches are generated when strands with an indel anneal to strands with no indel or a different indel sequence. The mismatches are subsequently detected and cleaved with an enzyme. The resulting fragments are analyzed by gel electrophoresis and band densitometry. For sequences designed for gRNAs, both the GeneArt CRISPR and TALEN design tools have the option of suggesting primers that can be used for GCD assays for each specific gRNA.

Examples of agarose gel results obtained with GCD assays are shown in Figure 2. Several different loci are shown, both untreated and treated with the detection enzyme. The presence of two smaller cleavage products indicates that the transformed pool contains successful edits.

The sizes of the expected cleaved products can be determined where the gRNA is centered within the amplified region. Comparing the signal intensity of the cleaved fragments by densitometry or fluorescence to the total amount of DNA (intact wild-type amplicon + cleavage fragments) results in an estimate of the editing efficiency. In the examples shown, the efficiency ranged from about 10% to 50%.

Summary of GCD analysis on agarose gels:

- GCD on agarose gels provides a simple mechanism for examining editing efficiencies.
- The size resolution of agarose gels is relatively low, so the sizes of the cleavage products can only be estimated.
- Accurate quantification requires that the scanning system not be saturated.
- Loading and analyzing a large number of samples on agarose gels can be tedious.
- Unlike sequencing-based methods, GCD does not identify the types of indels that may be expected from the editing process.

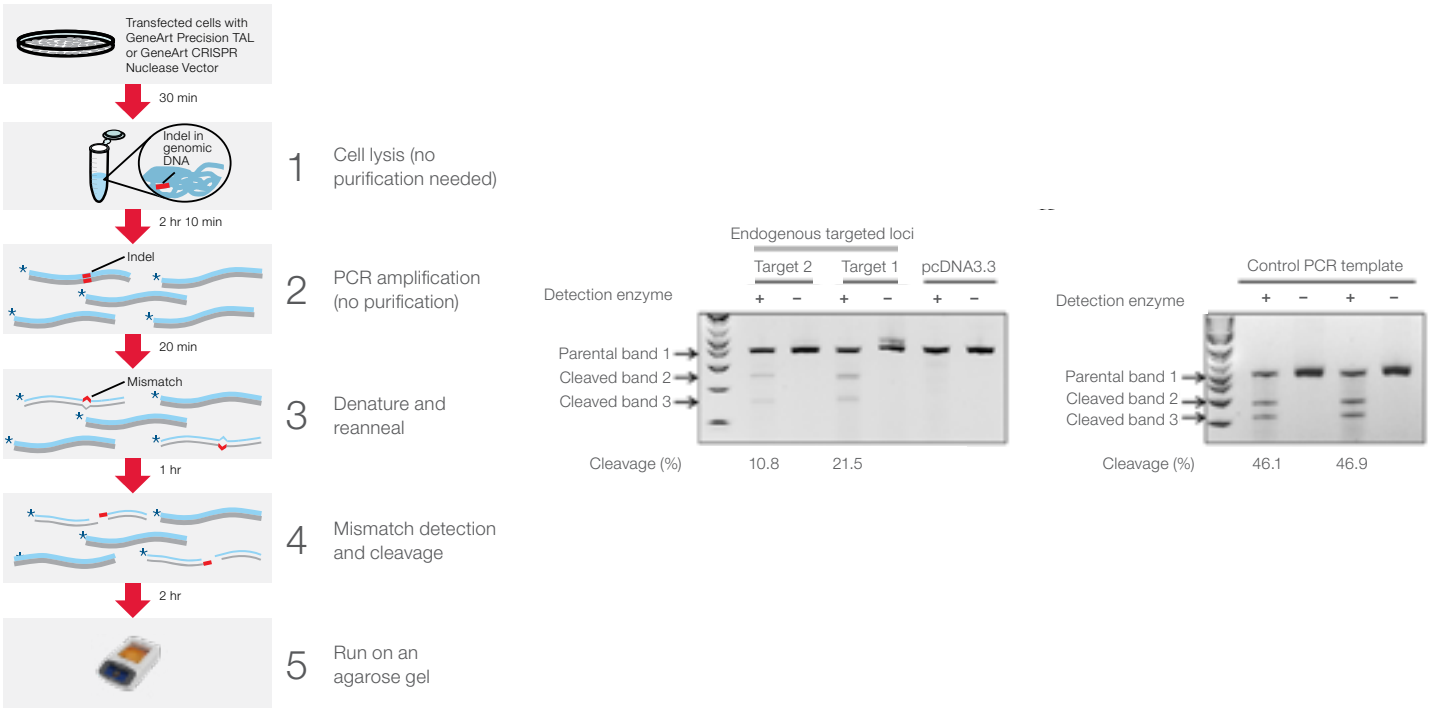


Figure 2. Determining editing efficiency using GCD and agarose gels. Analysis of editing efficiency on agarose gels is simple, but quantification and fragment sizing can be inexact.

GCD with fragment analysis

The GCD assay described above can also be analyzed using capillary electrophoresis (CE). For this method, one of the amplification primers for the locus is labeled with a fluorescent dye. The workflow for PCR amplification from a lysate of primary transformants, followed by treatment with the detection enzyme, is the same as previously described for agarose gels. The cleavage fragments are then resuspended in Applied Biosystems™ Hi-Di™ Formamide with size standards, separated by CE, and analyzed using standard fragment analysis software.

Editing efficiency using GCD followed by CE was tested on lysates from edits of five different loci. Extracts from control normal cells (not treated with gRNA) showed single peaks of the predicted size for the amplicon (Figure 3, example using *HPRT* amplicon). The size was not affected by denaturation followed by treatment with the detection enzyme. In contrast, cells treated with target-specifc gRNA, which is expected to create random deletions in the *HPRT* locus, display a peak corresponding to an intact amplicon, as well as peaks up to 15 nt smaller, when no detection enzyme is added. These smaller peaks correspond to deletions in the amplicon, presumably as a result of the edit. When treated with the single-stranded nuclease, new peaks in the range of 229–251 nt are visible, indicative of the cleavage products of the mismatch. By calculating the total height of the cleavage peaks relative to the total height of all peaks, an editing efficiency of 68% can be calculated. Similar results were obtained using extracts of pools edited at four other loci (data not shown).

Figure 3. Determining editing efficiency using GCD and fragment analysis by CE. Using the same general method used for agarose gels, editing efficiency can be analyzed by CE. Here, one of the PCR primers is labeled with a fluorescent tag. When control cells lacking gRNA were analyzed, a single peak corresponding to the *HPRT* parental locus was observed both with and without the detection enzyme. A mixed pool of primary transformants treated with target-specific gRNA showed the parental peak (471 nt). However, in the presence of the detection enzyme a cluster of peaks in the 229–251 nt range, indicative of cleavage products, was also detected. In addition, a cluster of peaks around 460 nt was also visible, indicative of the reannealed edits of deleted sequences without mismatches. Editing efficiency was calculated by determining the fraction of the total of all peaks (cleavage plus parental) in the 229–251 nt peaks. Orange peaks are size markers labeled with Applied Biosystems™ LIZ™ dye.

Analysis of fragments by CE is typically performed using Applied Biosystems™ GeneMapper™ Software. However, we have released a cloud-based fragment analysis tool, Applied Biosystems™ Peak Scanner™ Software. This software provides the peak analysis functionality of GeneMapper Software, but with a cleaner, easier-to-use interface. Since Peak Scanner Software is a cloud-based application, the addition of new functional modules is facilitated. In addition, cloud connectivity makes it easier for investigators to share data. Results of GCD analyses using Peak Scanner Software are equivalent to those obtained using GeneMapper Software (Figure 4).

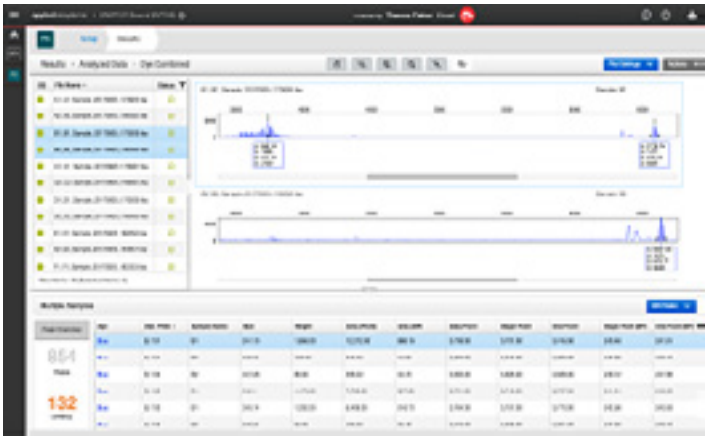


Figure 4. Cloud-based Peak Scanner Software facilitates fragment analysis. GCD fragments can be analyzed using free Peak Scanner Software. The functionality of GeneMapper Software is preserved in Peak Scanner Software, but with cloud-based storage and a streamlined user interface. GCD analysis of a *RELA* amplicon obtained from a primary pool of *RELA* gRNA–treated cells is shown; the top trace shows fragments produced when treated with the detection enzyme, and the bottom trace is in the absence of the enzyme.

Summary of GCD analysis by CE:

- GCD followed by CE provides finer size resolution than agarose gels. Differences down to a single base pair can be discerned.
- Summation of peak heights can provide more accurate quantification of the amount of DNA in a particular fragment. Calculations based on peak size (or area) can be more precise than calculations based on cleaved fragments from agarose gels.
- Throughput can be increased by using different dye colors for primer labeling, combining GCD assays from different loci in a single capillary. Although dye-labeled primers need to be chosen for each locus queried, a system has been described that performs similar analyses using universal dye-labeled primers [1].
- Analysis of GCD fragments is simplified using Peak Scanner software.

Sanger sequencing with Tracking Indels by Decomposition (TIDE) software

Genome editing efficiencies can also be determined using Sanger sequencing. Primers flanking the edited site can be designed and used to amplify the locus from the primary pool of edited cells. The GCD primers suggested by the GeneArt CRISPR and TALEN design tools can be used to generate amplicons that will cover the edited region and can be sequenced. Alternatively, PCR primers flanking the gRNA target sequence can be designed using rules similar to those for standard PCR primers. The resulting amplicons are then sequenced using Applied Biosystems™ BigDye™ Terminator or BigDye™ Direct chemistries. The traces from a sequencing reaction derived from a mixed pool containing a successful edit will contain a stretch of well-defined sequence, followed by a stretch of mixed-base sequence downstream of where the edit occurred (Figure 5). Visual inspection of these cues will often indicate whether an edit was successful, but determining the efficiency and nature of sequences present in the mixed pool is rarely possible from a simple inspection.

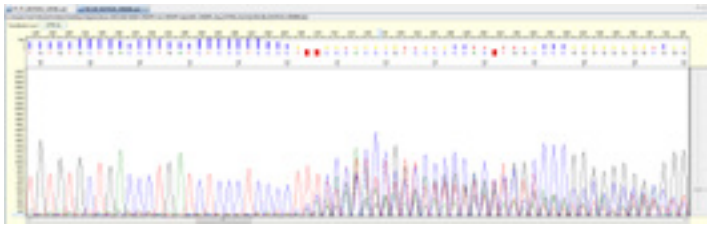


Figure 5. Sanger sequencing of a mixed pool of edited cells. An example trace of an amplicon from a pool of primary transformants. Notice that the sequence is uniform up to the red arrow, after which there are different sequences present. The mixture contains various insertions and deletions, and although it is clear that an edit has taken place, sequences and efficiency cannot be determined.

To facilitate this analysis, Brinkman et al. [2] devised methods for deconvoluting mixed-base sequencing traces resulting from genome editing experiments. They packaged these methods into a software tool called TIDE (https://tide.nki.nl). Starting with the gRNA sequence and sequencing traces from the amplicons of unmodified and edited DNA, TIDE software determines the likelihood that traces of different indel lengths contribute to the mixed trace. A histogram showing the various contributions of insertions and deletions (e.g., +2, +1, 0, –1, –2) is returned, as well as an overall calculation of the efficiency from the entire population (Figure 6). Although the exact sequences of the molecules in the edited population are not determined, establishing a measurement of the range of insertions and deletions that are present can help determine whether to proceed to the next steps of the experiment.

Extracts from primary pools of cells with edits at different loci were analyzed for editing efficiency using TIDE software (Figures 6 and 7). Note that the frequency calculated and spectrum of possible mutations is similar in the forward and reverse sequencing directions. For *HPRT*-edited cells (Figure 6), the overall efficiency was about 80%, with the major species appearing to be a +1 insertion. In the *RELA*-edited cells (Figure 7), the efficiency was lower at about 20%. TIDE software also provides an R^2 value as a measure of goodness-of-fit and highlights the statistical significance for each indel. Although the R^2 value is high for the *HPRT* locus, it is significantly lower for the *RELA* sequences, indicating lower confidence in those values.

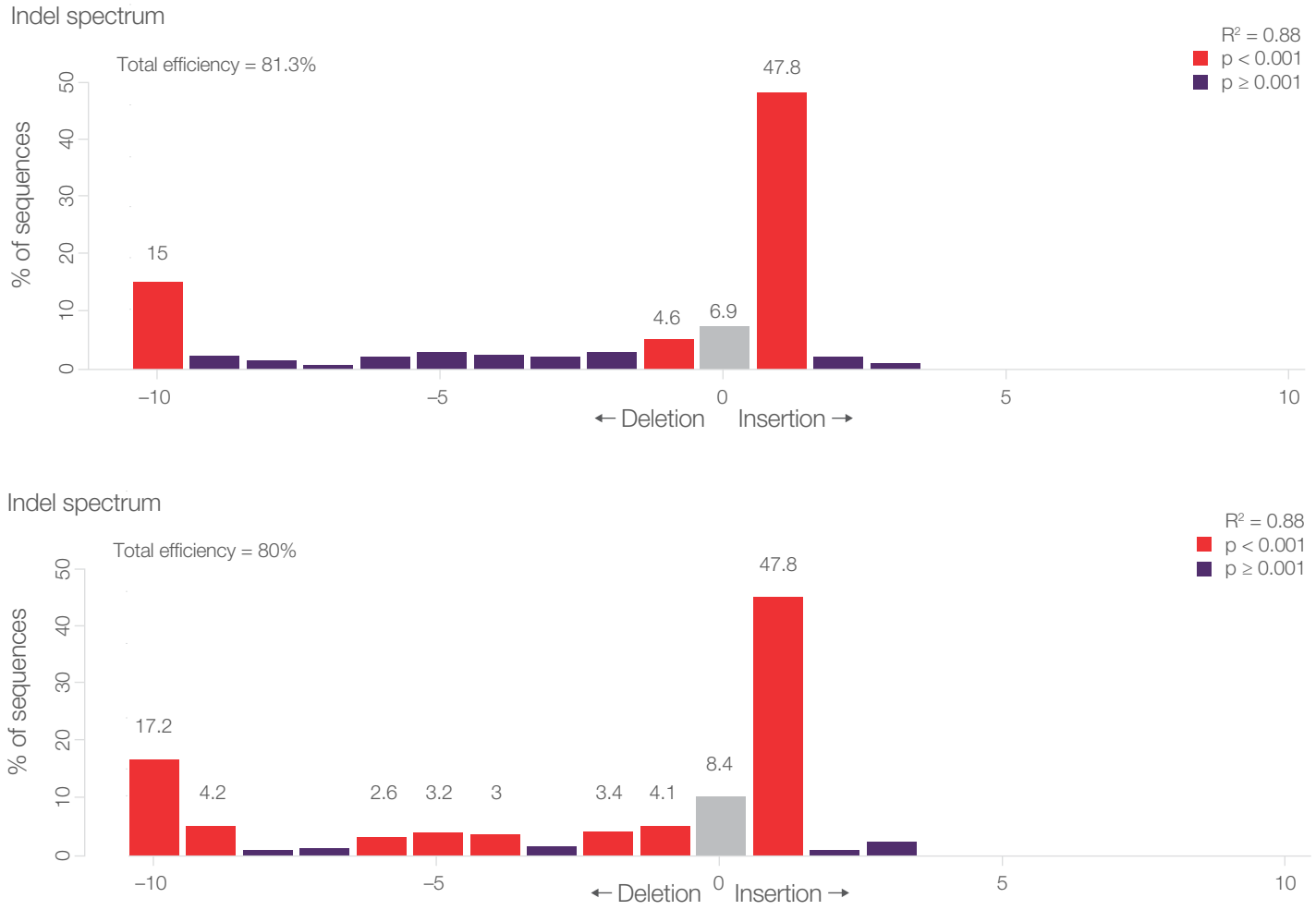


Figure 6. TIDE software can be used to evaluate editing efficiency. The software is able to take a mixed sequencing trace (e.g., Figure 5) and predict the percentages that can be contributed by different indel lengths. Red bars indicate the fractions calculated with high confidence; purple bars are indicative of lower confidence. Note that the spectrum of mutations and frequency of edits calculated match between the two strands.

Importantly, settings in TIDE software allow the user to change the window where the deconvolution occurs. Changing the window from the default value of 10 nt to 25 nt reveals that many of the edits in the *RELA* locus were deletions greater than 15 nt and were therefore missed using the default settings (Figure 7). The R^2 value is also higher with the increased window size, and the resulting calculated efficiency (68%) is more in line with that determined by other methods (see below).

Summary of analysis by Sanger sequencing:

- Sanger sequencing followed by analysis with TIDE software offers a useful way to get information from primary pools of edited cells.
- Well-established sequencing workflows provide a straightforward method for determining the efficiency and types of deletions to expect in clonally expanded cells.
- When using TIDE software, it might be necessary to change the query window from the default 10 nt to determine if larger deletions are present.
- Since homology-directed repair happens less frequently than double-strand break repair, any SNPs or other directed edits may be lost among the random deletions present in the primary pool. However, an update was introduced (TIDER software) that can calculate the efficiency from template-directed edits (e.g., SNPs) [3].
- Use of TIDE software is free of charge for single-locus queries. Batch queries for high-throughput studies are possible but might require purchasing a license. For the most up-to-date details, please visit the TIDE website (https://tide.nki.nl).

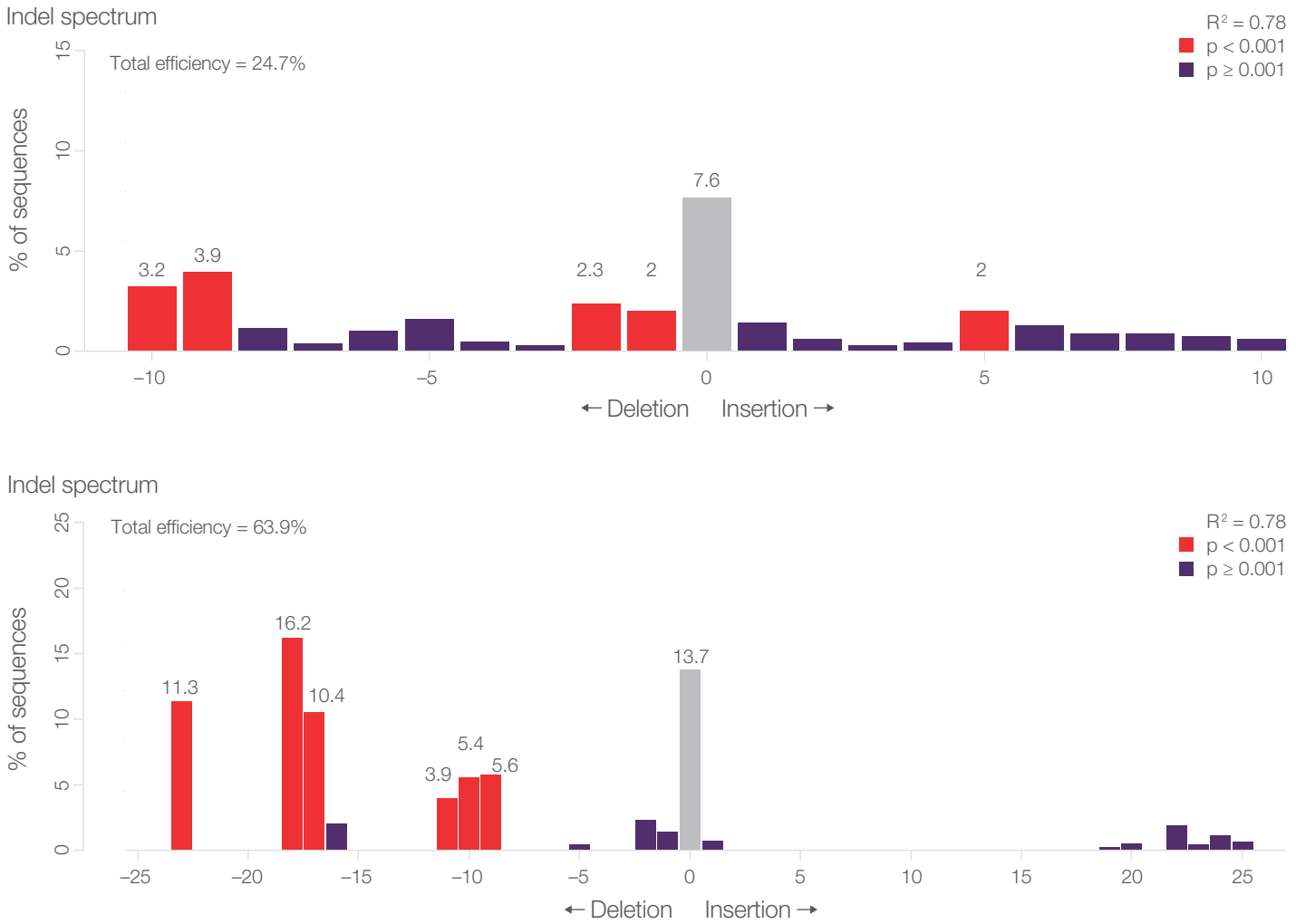


Figure 7. Changing analysis parameters in TIDE software can have a large effect on the efficiency calculation. By default, TIDE software operates in a –10 to +10 window around the gRNA start site. However, changing the size of the window can increase the accuracy of the calculation. Here, the overall efficiency in a *RELA*-edited amplicon is ~25% with a default setting, while expanding the window to –25 to +25 produces a result (~64% efficiency) closer to the real value (see Figure 11).

Sanger sequencing with TOPO cloning

The efficiency of a genome editing experiment can also be determined by subcloning the amplicons into plasmids and sequencing individual plasmids (Figure 8). This method gives precise information about the types of mutations that are generated, since individual molecules are sequenced. To illustrate this method, we amplified DNA from cell lysates using target-flanking primers to generate PCR amplicons that were no greater than 600 base pairs in length. The PCR products were then subcloned using the Invitrogen™ Zero Blunt™ TOPO™ PCR Cloning Kit and transformed into TOP10 *E. coli* cells. Ninety-six bacterial colonies were picked per transformed pool of edited cells and processed for DNA isolation using the Invitrogen™ PureLink™ 96 HQ Mini Plasmid DNA Purification Kit and then subjected to Sanger sequencing. The resulting sequencing data were then analyzed to measure the percentage of PCR products containing accurately edited sequence and to select which clonal isolates to maintain. By sequencing a large number of plasmids, the fraction containing an edit can be determined. This method also gives a first glance at the overall gene knockout or editing efficiencies and types of indel changes that have occurred.

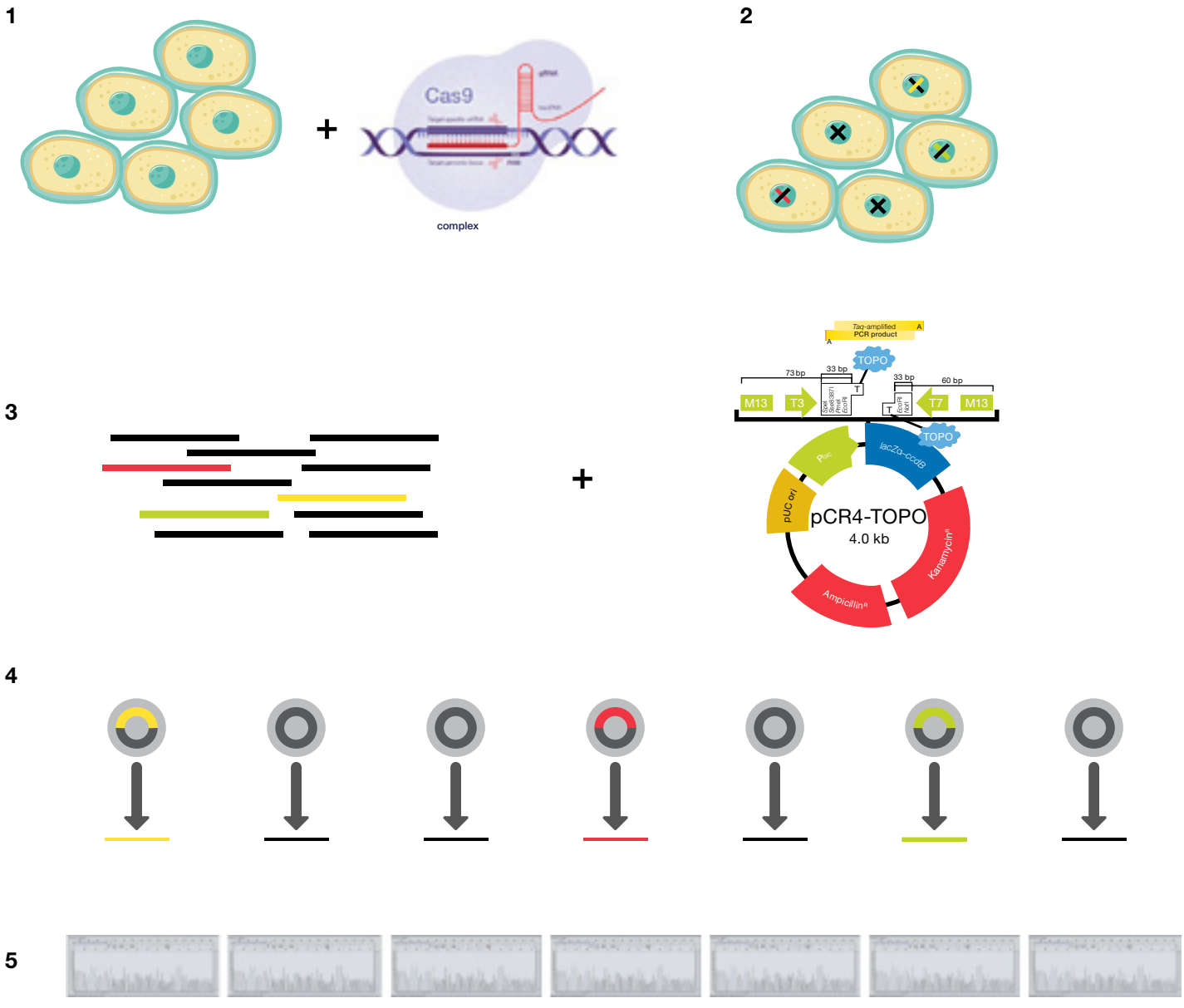


Figure 8. Calculating efficiency by plasmid subcloning followed by sequencing. Step 1: Transfect cells with gRNA and Cas9 mRNA. Step 2: Incubate cells to allow processing of genomic change. Step 3: Purify genomic DNA from the cell culture, PCR-amplify the engineered locus from the heterogeneous culture, and clone PCR fragments into the TOPO vector. Step 4: Isolate plasmids from single colonies and PCR-amplify the inserts. Step 5: Sequence the inserts. The efficiency of the edit is the percentage of the total number of inserts sequenced that have an engineered change.

After transfecting our cells with gRNA and Cas9 mRNA, we used TOPO cloning to subclone the locus from primary transformants, and Sanger sequencing to analyze 96 clones. Of the 96 clones analyzed, 84 gave alignments with the target sequence. Only 12 clones had no editing event in the amplified region; 72 clones had at least one sequence deviation from the wild-type sequence, for an overall efficiency of about 86%. For further details, please see reference 4.

Summary of analysis by TOPO cloning:

- Subcloning followed by Sanger sequencing can provide an estimate of the exact nature of the changes present in the primary pool.
- Sequencing more plasmids will provide greater depth of knowledge, as rarer species present in the pool are more likely to be revealed as more colonies are sequenced.
- Editing efficiency becomes more accurate as more plasmids are sequenced.
- The workflow of subcloning into bacteria followed by sequencing can be time-consuming and tedious.

dPCR

Digital PCR (dPCR) provides a mechanism for precisely determining the fraction containing an edit, by partitioning a PCR reaction into thousands of submicroliter reaction chambers. By counting the number of chambers that are positive for queried sequences, the number of molecules containing those sequences in the original sample can be calculated. This approach can be adapted for experiments to determine the efficiency of genome editing. In this strategy, custom Applied Biosystems™ TaqMan® probes are designed so that one probe is placed over the desired edited sequences, and a second probe is placed in a region of the amplicon that is not likely to be affected by the edits (Figure 9A). When PCR is performed on the partitioned sample, most of the partitions should be positive for the probe targeting the unedited region, while partitions that have not undergone an edit will also be positive for the probe targeting the edited region. Counting the fraction of wells that are positive for only one probe will give an indication of the efficiency of the edit.

To show the utility of this method, we obtained lysates from cells containing random edits of various human genes. One microliter of lysate was used with custom-designed Applied Biosystems™ TaqMan® Assays and the QuantStudio™ 3D Digital PCR System, following the standard protocol

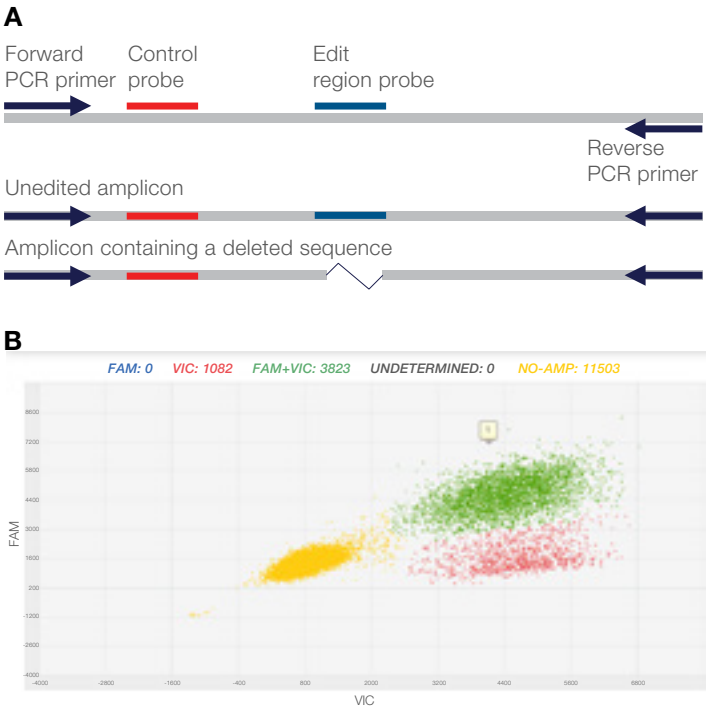


Figure 9. Calculating editing efficiency by dPCR. (A) By designing a custom TaqMan SNP assay, editing efficiency can be estimated by dPCR. To do this, one probe is chosen that is not likely to be affected by the cleavage and repair events (probe labeled with VIC™ dye, here shown in red), and one probe is chosen that covers the original unaltered sequence in the region targeted for editing (probe labeled with FAM™ dye, blue). **(B)** Wells that are positive only for the VIC probe have had an edit occur (red points); wells that are positive for both FAM and VIC probes (green points) have had no edit occur. The fraction of the total (wells positive for VIC probe plus FAM + VIC probes) that is positive only for the VIC probe gives the editing efficiency. Note that some optimization of probe placement might be required.

and thermal cycling program for dPCR. Data were analyzed using Applied Biosystems™ QuantStudio™ 3D AnalysisSuite™ Cloud Software (Figure 9B). As an example, a lysate from cells edited at the *CDK4* locus resulted in 3,827 wells that were positive for both probes and 1,082 wells that were positive for the control probe only. This translated to a 28% editing efficiency. Similar determinations were made with lysates from cells edited at four other genes (Figure 11).

Summary of analysis by dPCR:

- Editing efficiency can be determined by dPCR.
- The digital nature of data acquisition means that a precise value for the efficiency of the edit can be obtained.
- dPCR methods can be modified to determine the efficiency of more targeted edits, such as SNPs or translocations. For example, the edit-specific probe can be targeted to the desired single-nucleotide change, and the fraction that is positive for this probe calculated.

- Successful use of dPCR in genome editing experiments might require optimization of assay design, including PCR conditions and placement of primers and probes.

Next-generation sequencing

The editing efficiencies can also be determined using Ion Torrent™ next-generation sequencing (NGS) approaches. In general, amplicons derived from targeted loci in the primary pool of edited cells are sequenced. Cells from each pool are collected and lysed in buffer provided with the GeneArt Genomic Cleavage Detection Kit. The targeted region is then amplified using primers flanking the edited site. It is important to choose primers that will generate amplicons compatible with the expected read length. We also recommend using a high-fidelity PCR enzyme to avoid any sequence errors that could compromise data quality. Barcoded libraries can be constructed from the PCR samples using, for example, the Ion Chef™ System, Ion Xpress™ Barcode Adapters 1-96 Kit, Ion Xpress™ Plus Fragment Library Kit, and Ion PGM™ Hi-Q™ Chef Kit. Once constructed, libraries can be sequenced on Ion PGM™ or Ion S5™ instruments. Detailed protocols for preparing and running Ion Torrent™ libraries can be found in the respective product user manuals.

NGS provides a robust method for analyzing the pool of primary transformants (Figure 10). The efficiency of editing can be calculated by comparing the number of reads for edited species to the total number of reads for the locus. In addition, the nature of the edits at the targeted locus can be determined by analyzing the sequences that are returned. Sequencing depth and the ability to get multiple reads per individual amplicon drives the accuracy and reliability of this method. Having up to 384 barcodes available also makes this method amenable to high-throughput workflows. However, one of the drawbacks to analysis by NGS is the scarcity of automated data analysis tools. Customized data analysis pipelines may be able to assist with automated data analysis but would have to be developed. Moreover, data sets as large as those generated by NGS might not be needed to address the efficiency of an editing experiment. The complexity of the chemistry and data analysis can affect the time to obtain an answer, and unless sequencing runs are multiplexed, NGS can also be relatively expensive on a per-sample basis.

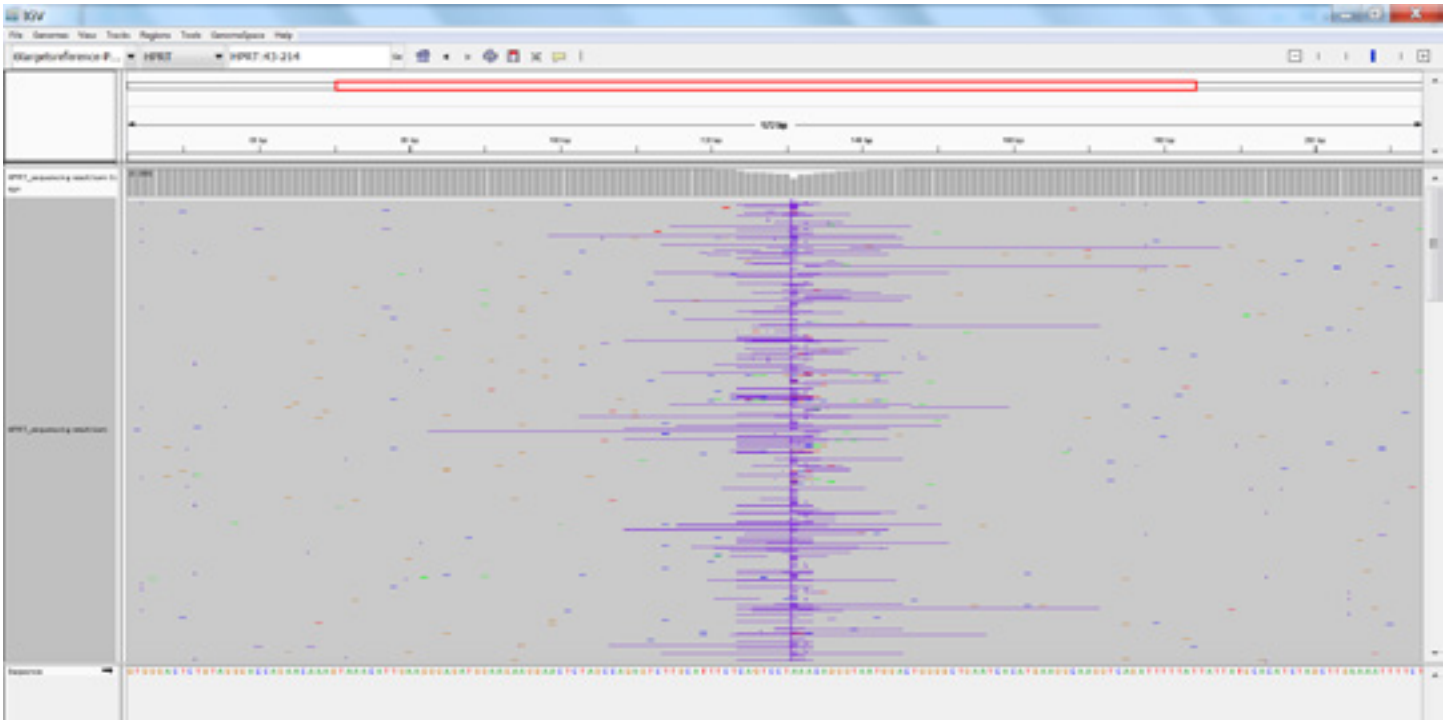


Figure 10. NGS analysis of mixed pools of cells edited at the *HPRT* locus. Reads from an Ion Torrent NGS analysis were imported into the Integrative Genomics Viewer (IGV) and aligned over the gRNA sequence. Each row represents a separate read; purple lines in the row are the sequences deleted. Other modified bases are shown by short bars of different colors. The gray bars above the reads panel show the relative number of unmodifed bases at that position; note that most of the changes cluster around the gRNA site. By analyzing data in IGV, an indication of the efficiency and types of edits present in a mixed pool is readily obtained.

Summary of analysis by NGS:

- Genome edits in a primary pool can be analyzed by next-generation sequencing.
- Increasing read depth provides increasing detail of edits present in the pool and accuracy of the efficiency determination.
- NGS solutions can be amenable to high-throughput workflows.
- The complexity of the data generated may require custom data analysis tools.
- NGS can be relatively expensive unless samples are multiplexed.

Method comparison

We have shown here that there are many different ways to determine the efficiency of primary edits, and there are others not mentioned herein. To compare the values obtained by the methods previously discussed, we determined the editing efficiency in lysates from pools of cells edited for five different loci (Figure 11). To ensure direct comparisons, aliquots of the same lysate were used for each determination. Overall, the frequencies observed were similar, with a couple of outliers. For example, the *HPRT* editing frequency obtained by GCD with agarose analysis was higher than those obtained by GCD with fragment analysis or Sanger sequencing; this could be due to the imprecise nature of quantifying intensities on agarose gels. Another example was that the *CDK4* editing frequency obtained by dPCR was lower than those of other methods; this could be due to suboptimal assay design. Finally, the TIDE frequency determination using *CHUK* lysate was lower than those of the other methods; this was due to the lower quality of the sequencing traces (data not shown).

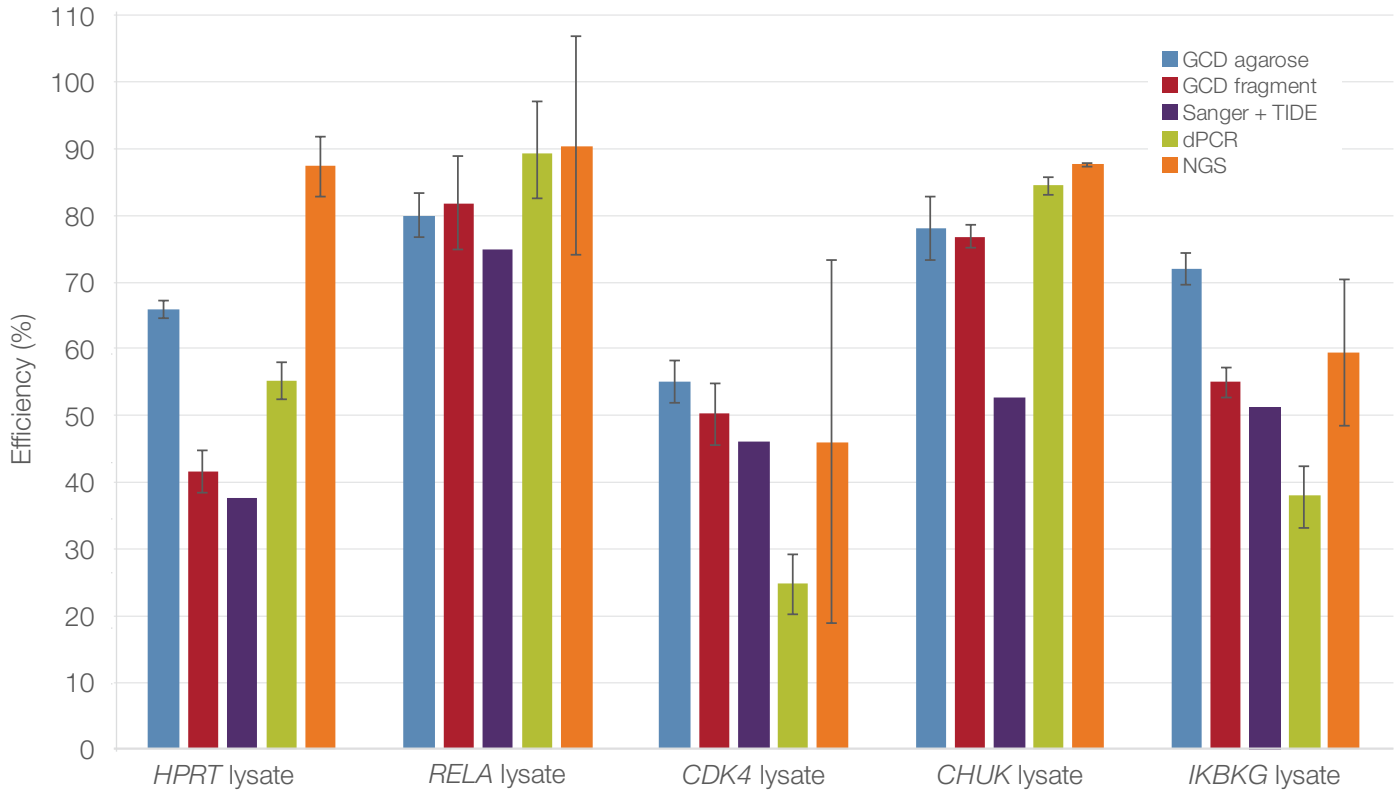


Figure 11. Comparison of editing efficiencies obtained by different techniques. Each lysate of primary pools of knockout edits of the indicated genes in A549 cells was analyzed using the techniques shown. Determination of efficiency of editing is mostly similar by electrophoretic methods (GCD and Sanger sequencing) and more divergent by dPCR and NGS.

	Primary analysis (efficiency of edits)	Primary analysis (sequencing information gained)	Secondary analysis (confirmation of edit)	Highest relative throughput	Lowest relative cost
Knockout (random deletion)	• GCD • Sanger sequencing and TIDE software	• NGS • TOPO cloning and Sanger sequencing	• Sanger sequencing	• GCD • NGS	• GCD
SNP	• dPCR • Sanger sequencing and TIDE software	• NGS • TOPO cloning and Sanger sequencing	• TaqMan Genotyping Assay • Sanger sequencing	• TaqMan Genotyping Assay • NGS	• TaqMan Genotyping Assay
Insertion	• dPCR • TOPO cloning and Sanger sequencing	• NGS • TOPO cloning and Sanger sequencing	• TaqMan Genotyping Assay • Sanger sequencing	• TaqMan Genotyping Assay • NGS	• TaqMan Genotyping Assay
Defined deletion	• TOPO cloning and Sanger sequencing • Sanger sequencing and TIDE software	• NGS • TOPO cloning and Sanger sequencing	• Sanger sequencing	• GCD • NGS	• Sanger sequencing

Figure 12. Methods for determining efficiency of cleavage. Choosing a method for determining the efficiency of an edit depends on the type of edit desired, throughput needed (ability to analyze large numbers of samples), and complexity of the workflow. The methods recommended for various edits and experimental variables are shown. All of the methods can provide quantitative information, but some have relative advantages over others.

Each method has advantages and disadvantages for analyzing primary transformants. While each method can return information on editing efficiency, many variables will dictate which method is most appropriate for a given experimental system. Some of the variables that should be considered include the type of edit desired, whether exact sequence information is needed, and the experimental workflow. As a starting point, suggested methods for some of the more common and important experimental considerations are shown in Figure 12. For a given experiment, it might not be necessary to obtain a precise determination of editing efficiency. Rather, it might be sufficient to see that the edit worked "well enough" to choose a best-performing gRNA before moving to the next phase of the experiment.

Considerations for secondary validation of homogeneous cell populations

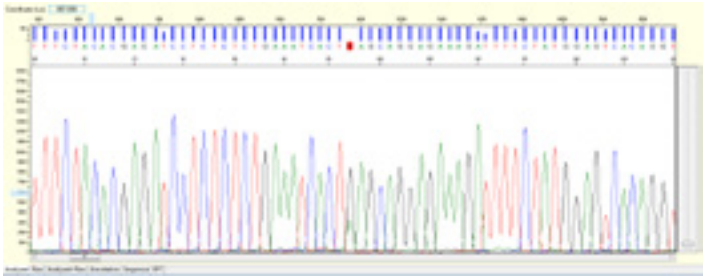
Analyzing a primary pool is only one part of a genome editing experiment. Some investigators may choose to perform a secondary screen of isolated colonies in order to obtain a genetically homogeneous culture of cells. In such secondary screens, determining the editing efficiency is not required. Nevertheless, the clonality and purity of the derived clone will need to be determined. In addition, on- and off-target effects will need to be examined. Many of the genomic analysis tools described here will be helpful in these screens.

For example, Sanger sequencing is a valuable tool for validating secondary clones. In a relatively straightforward workflow, lysates from the culture can be amplified and sequenced as previously described. Here, the expectation is that a homogeneous culture will produce a clean trace without contaminating peaks. Thus, in addition to confirming the monoclonality of the culture, the exact sequence of the edit at the locus can be obtained. If the sequencing trace shows heterogeneity, another round of single-cell selection might be required. In the case of engineered SNPs, analysis using Applied Biosystems™ Minor Variant Finder Software might be useful (Figure 13). This freely available software tool was developed to detect a variant SNP at frequencies as low as 5% of the total from Sanger sequencing traces [5]. Although Minor Variant Finder Software will only determine variant frequencies of SNPs, TIDER software [3] could be used to analyze the frequency of other sequences in a mixed population.

In addition, genotyping approaches based on TaqMan Assays can be extremely useful in the validation process. In a primary pool, there is a heterogeneous population of sequences present, only a small fraction of which might have the desired edit. This could complicate the reaction efficiency and data interpretation in a primary screen. In contrast, in a secondary analysis of monoclonal cultures, screening for the desired edit can be done much more efficiently. Since TaqMan genotyping assays are designed to query defined sequences, they can be used to identify cultures that have a specific desired edit.

Finally, NGS can be useful for characterizing a secondary clone of edited cells. For example, sequencing the edited locus in a monoclonal cell culture can provide information about the exact nucleotides changed, as well as a quantification of other species present if the culture is not pure. It might also be useful to sequence loci related to the targeted gene, or even the whole genome, to determine if any off-target effects occurred during the editing process.

A



B

Figure 13. Detecting SNP edits in mixed-population sequencing traces. Minor Variant Finder Software is a user-friendly desktop software package specifically designed for the detection and reporting of single-nucleotide variants (SNVs) in Sanger sequencing traces. **(A)** An example of a sequencing trace with a minor SNP. **(B)** Minor Variant Finder Software operates by (1) reducing background, (2) identifying variants, and (3) determining allele frequency in forward and reverse strands. Variants present at frequencies as low as 5% can be detected using this software. In the example shown, the variant frequency is 12.9% in the forward direction and 10.8% in the reverse direction.

Conclusions

Thermo Fisher Scientific has developed tools and solutions for every step in the genome editing workflow. Our portfolio, including Applied Biosystems™ genetic analysis tools, is built on 30 years of innovation and can grow with your research needs. Our set of tools comprises optimized, validated systems that have been designed to work together to help answer your important scientific questions faster and with less effort. Since every investigation is unique, we offer a range of genome editing solutions that cater to the needs of your laboratory and research.

References

1. Yang Z, Steentoft C, Hauge C et al. (2015) Fast and sensitive detection of indels induced by precise gene targeting. *Nucleic Acids Res* 43:e59.
2. Brinkman EK, Chen T, Amendola M et al. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* 42:e168.
3. Brinkman EK, Kousholt AN, Harmsen T et al. (2017) Easy quantification of template-directed CRISPR-Cas9 editing. Preprint from bioRxiv: **http://dx.doi.org/10.1101/218156**
4. Thermo Fisher Scientific Inc. (2016) Using Sanger sequencing to facilitate CRISPR- and TALEN-mediated genome editing workflows. Application note COL31142.
5. Thermo Fisher Scientific Inc. (2016) Low-level somatic variant detection in tumor FFPE samples by Sanger sequencing. Application note COL31176.

Ordering information

Product	Quantity	Cat. No.
GeneArt Genomic Cleavage Detection Kit	20 rxns	A24372
Zero Blunt TOPO PCR Cloning Kit for Sequencing	25 rxns	450031
SeqStudio Genetic Analyzer System with SmartStart Orientation	1 system	A35644
SeqStudio Cartridge v1	500 reactions	A33671
SeqStudio Cartridge v2	1,000 reactions	A41331
3500 Genetic Analyzer	1 system	4440466
3500xL Genetic Analyzer	1 system	4440467
GeneMapper Software 6	1 license	A38892
Peak Scanner Software v1.0	1 license	4381867
BigDye Terminator v3.1 Cycle Sequencing Kit	100 rxns	4337455
BigDye Direct Cycle Sequencing Kit	100 rxns	4458687
GeneScan 500 LIZ Dye Size Standard	800 rxns	4322682
QuantStudio 7 Flex Real-Time PCR System, 384-well, desktop	1 system	4485701
Custom TaqMan SNP Genotyping Assays	1,500 rxns	4331349
QuantStudio 3D Digital PCR Instrument with Power Cord	1 instrument	4489084
QuantStudio 3D Digital PCR 20K Chip Kit v2 and Master Mix	8 x 12 chips	A26317
Ion S5 System	1 system	A27212

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

ThermoFisher SCIENTIFIC

For Research Use Only. Not for use in diagnostic procedures. © 2019 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. TaqMan is a registered trademark of Roche Molecular Systems, Inc., used under permission and license. **COL08859 0319**