

Sep 02, 2021

Rapid Quantitative Evaluation of CRISPR Genome Editing by TIDE and TIDER

Book Chapter

Eva Karina Brinkman¹, Bas van Steensel¹¹Division of Gene Regulation and Oncode Institute, Netherlands Cancer Institute, Amsterdam, The Netherlands

1 Works for me

Share

dx.doi.org/10.17504/protocols.io.bqzmmx46

Springer Nature Books

satyavati Kharde

ABSTRACT

Current genome editing tools enable targeted mutagenesis of selected DNA sequences in many species. However, the efficiency and the type of introduced mutations by the genome editing method are largely dependent on the target site. As a consequence, the outcome of the editing operation is difficult to predict. Therefore, a quick assay to quantify the frequency of mutations is vital for a proper assessment of genome editing actions. We developed two methods that are rapid, cost-effective, and readily applicable: (1) TIDE, which can accurately identify and quantify insertions and deletions (indels) that arise after introduction of double strand breaks (DSBs); (2) TIDER, which is suited for template-mediated editing events including point mutations. Both methods only require a set of PCR reactions and standard Sanger sequencing runs. The sequence traces are analyzed by the TIDE or TIDER algorithm (available at <https://tide.nki.nl> or <https://deskgen.com>). The routine is easy, fast, and provides much more detailed information than current enzyme-based assays. TIDE and TIDER accelerate testing and designing of DSB-based genome editing strategies.

DOI

dx.doi.org/10.17504/protocols.io.bqzmmx46

EXTERNAL LINK

https://link.springer.com/protocol/10.1007/978-1-4939-9170-9_3

PROTOCOL CITATION

Eva Karina Brinkman, Bas van Steensel 2021. Rapid Quantitative Evaluation of CRISPR Genome Editing by TIDE and TIDER. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.bqzmmx46>

MANUSCRIPT CITATION please remember to cite the following publication along with this protocol

Brinkman E.K., van Steensel B. (2019) Rapid Quantitative Evaluation of CRISPR Genome Editing by TIDE and TIDER. In: Luo Y. (eds) CRISPR Gene Editing. Methods in Molecular Biology, vol 1961. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-9170-9_3

KEYWORDS

CRISPR-Cas, Genome editing, Indel mutation, Mutagenesis, DNA mutational analysis/methods, Web tool, Algorithm, Sanger sequencing

LICENSE

————— This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited


CREATED


Dec 23, 2020


LAST MODIFIED


Sep 02, 2021

OWNERSHIP HISTORY

Dec 23, 2020  Megan Freund

May 07, 2021  Emma Ganley protocols.io

Aug 24, 2021  Satyavati Kharde

Aug 26, 2021  satyavati Kharde

PROTOCOL INTEGER ID

45837

GUIDELINES

1. Introduction:

CRISPR-based systems are popular and widely used for genome editing in the field of molecular biology. CRISPR endonuclease Cas9 introduces a DSB into the genomic DNA with high precision. Due to the error-prone repair mechanisms of the cell, this often results in insertions or deletions at the targeted site [1]. This is exploited to make functional knock-outs of specific genes and regulatory elements [2, 3, 4]. Alternatively, to gain more control over the nature of the mutations, strategies have been developed that introduce small nucleotide changes around a precisely targeted site by using a donor template [5, 6]. In the latter approach the genomic DNA around the DSB break is replaced by the DNA of the donor template through homology-directed repair (HDR), resulting in the introduction of a designed mutation with high accuracy [7, 8]. This precise editing creates the possibility to generate and study specific disease-causing nucleotide variants [6, 9]. Typically, one starts with a homogeneous cell line and ends up with a pool of cells with a complex mix of indels and/or designer mutations [10, 11, 12]. To study a mutation of interest, clonal mutant lines need to be isolated from the cell pool. Because this is a very labor-intensive process it is important to know *a priori* the efficiency in which the desired mutation(s) have been introduced. However, a complicating factor is that the efficacy of the programmable nucleases can vary dramatically depending on the sequence that is targeted. In addition, different cell types have a varying performance in transfection capability. These factors make the efficacy of CRISPR experiment difficult to predict. For this reason it is usually necessary to test several guide RNAs (gRNAs) that lead the endonuclease to the site of interest. This is even more critical when a template-directed strategy is applied, which often has a low efficiency because HDR repair pathways are generally less active than error-prone non-templated repair [10, 12]. Hence, a quick and easy assay to estimate the frequencies of the diverse introduced mutations in the cell pool is of key importance.

We developed two methods that can accurately quantify the efficiency of either indels or template-directed mutations in a pool of cells. Both methods are rapid and cost-effective. The method TIDE (Tracking of Indels by DEcomposition) identifies and quantifies indels. It requires only a pair of standard Sanger sequence traces of two PCR products [13]. The sequence traces are then analyzed using an easy-to-use web tool. Note that TIDE can only detect overall indel frequencies, but not nucleotide substitutions or specifically designed indels. For the latter purpose we developed TIDER (Tracking of Insertions, DEletions, and Recombination events) [14]. This method can estimate the incorporation frequency of template-directed mutations, including point mutations, and distinguish them from a background of additional indels that originate from competing erroneous repair pathways. Although TIDER can also quantify indels alone, TIDE is slightly simpler to implement and therefore more suited for the assessment of non-templated editing experiments. The corresponding web tools for both TIDE and TIDER are freely accessible at <http://tide.nki.nl>.

4. Notes:

1. *Primer design recommendations for control and experimental sample.* Primers a, b need to cover the CRISPR target site. The length of the PCR product can vary, but there should be at least >50 bp up- and downstream of

- the break site for the alignment (see **Notes 6** and **7**) and decomposition windows respectively (see **Note 9**).
2. *Primer design recommendations for reference sample.* Primer c, d should carry the designed mutation(s) as present in the donor template (see [Section 3.2](#), **Note 3**). It is advised to include at least 10 complementary nucleotides on the 3' side of the mutation(s).
 3. *Donor plasmid contamination in isolated genomic DNA.* Potentially, a donor template that was transfected into the cells could co-purify with genomic DNA and be co-amplified in the PCR if it contains the primer sequences. This could result in an overestimation of the HDR events. This is generally not a problem with short ssODN donors, but with plasmid templates with long homology arms the primers a, b should be chosen outside of these homology arms. Alternatively, the donor plasmid may be cleared from the cells by a few passages of culturing.
 4. *Nuclease type.* TIDE(R) is currently designed for regular Cas9. But it can be used to analyze data from another nuclease, by entering in the web tool the DNA sequence around the expected cut site. The TIDE(R) web tool assumes that the DSB is induced between nucleotides 17 and 18 of the guide RNA sequence string (Fig. 3f). Note that if the exact breakpoint is unknown, TIDE will estimate the amount of the indels correctly, but the nucleotide composition of the +1 insertion will not be reliable. TIDER will only work when the exact cutting position is known and when the nuclease is a blunt cutter.
 5. *No guide RNA match.* Sometimes a mismatch occurs in the control sequence at the location of the sgRNA. This will stop the TIDE(R) analysis. In this case, edit the base annotation in the chromatogram file into IUPAC nucleotides of the expected control sequence (Fig. 3g). The peak signals in the chromatogram should not be altered. Viewing and editing of chromatogram files can be performed with Snapgene or Chromas software.
 6. *Alignment cannot be performed.* By default, the alignment window begins at nucleotide number 100, because the first part of the sequence read tends to be of low quality. The end of the alignment window is set automatically at 15 bp upstream of the break site. When this window is too small or when the break site is located upstream of nucleotide 100, the alignment cannot be performed correctly. Then the start of the alignment window should be set manually closer to nucleotide number 1 (Fig. 3c).
 7. *Incorrect alignment.* When the beginning of the sequence trace is of poor quality, the alignment function can make a mistake. This results in a quality plot with a high aberrant sequence signal along the whole length of the sequence trace (Fig. 3d). The aberrant sequence signal should only increase around the expected cut site (blue dashed line). In case of poor alignment, the start of the alignment window needs to be adjusted until a proper alignment is achieved (default of 100).
 8. *Quality plot recommendations.* In the experimental sample, around the break site a consistently elevated signal is expected, which is due to indels introduced at the break site. The starting position of this elevated signal may be used to verify that breaks were induced at the expected location. The control trace should have a low and equally distributed aberrant sequence signal along the whole trace. The reference trace in the case of TIDER should only have high scores at the positions of the altered nucleotides. Fluctuations in the control and reference signal reflect local variation in the quality of the sequence trace. Near the end of the sequencing traces the aberrant signal is often high, typically due to the lower quality of the trace toward the end (Fig. 3a). When a sequence stretch of poor local quality is present in the decomposition window the calculations of TIDE(R) are compromised. The boundaries of the decomposition window can be manually adjusted to remove the region that is of low quality; this will improve the estimations. Another area to avoid in the decomposition window is a stretch of repetitive sequences. These regions can be recognized in the quality plot as a sudden stretch without aberrant nucleotides (Fig. 3b). Such region might confound the decomposition of the sequence trace.
 9. *Decomposition window recommendations.* For TIDE, the default decomposition window spans the entire sequence trace from the break site until the end of the sequence minus the size of the maximum indel. When the boundaries of the decomposition window cannot fulfill this constraint, the software will report that the boundaries are not acceptable. For example, this can occur when the break site is too close to the end of sequence trace. To address this, the decomposition window boundaries should be set further apart or a smaller indel size should be chosen. Alternatively, new primers have to be designed according to **Note 1**. For TIDER the decomposition window is by default 20 bp upstream of the break to 80 bp downstream from the break. This smaller window compared to TIDE has more discriminatory power for subtle designed base pair changes.
 10. *Goodness of fit.* R^2 is a measure for the reliability of the estimated values. For example, if the R^2 value is 0.95, it means that 95% of the variance can be explained by the model; the remainder 5% consists of random noise, very large indels, non-templated point mutations, and possibly more complex mutations. Decomposition results with a low R^2 must be interpreted with caution. A low R^2 can be caused when the settings are not optimal or when the sequence quality is not good (see **Note 15**). A low R^2 value can also arise when a sequence stretch with a poor local quality is present in the decomposition window (see **Note 8**). Furthermore, the presence of indels larger than the maximum indel size that is considered can affect the R^2 (default of 10). By default these

are not modeled, which may result in a low R^2 score. The size range of indels that are modeled can be manually changed to larger number to test if this improves the fit (Fig. 3e).

11. *Allele-specific indels.* The different bars in the plot represent the insertions, deletions, and/or template-directed mutations in the cell population. These mutations are not specific of an allele. To determine allele-specific information a cell clone needs to be isolated and analyzed again by TIDE(R). A diploid cell gives a percentage of ~50% per mutation.
12. *Overall efficiency.* The overall efficiency refers to the estimated total fraction of DNA with mutations around the break site. It is calculated as $R^2 \times 100\%$ wild type.
13. *Distal designed mutations.* It has been reported that the incorporation of donor template sequence is less efficient when the designed point mutations are further away from the break site [19]. This often leads to a variation in incorporation frequently of the distal and proximal designed mutations as can be observed in the quality plots. Such a situation may confound TIDER estimates. The decomposition window can be restricted to either the proximal or the distal mutations to resolve the individual efficiencies.
14. *Natural versus designed mutations.* In general, TIDER is able to discriminate “naturally” occurring deletions and insertions from templated “designed” indels. Only in the presence of a small designed deletion (-1, -2) near the expected break site the designed mutation may be underestimated [14]. In case the designed mutation consists of an insertion larger than +1, TIDER does not consider natural insertions of the same size, because the decomposition becomes less robust. This is generally acceptable, because natural insertions larger than +1 are rarely observed [13, 17].
15. *Poor sequence quality.* When the sequence has poor quality overall, TIDE(R) will yield poor results with a low R^2 value (see **Note 10**) since too much noise is present in the data. The quality plot will show an overall high aberrant sequence signal in the control (the reference) and the experimental sample, before and after the break site (see **Note 8**). It is recommended to check the chromatograms of the samples (Fig. 3h) for poor sequencing quality. If so, these samples cannot be analyzed reliably by TIDE(R). Note that sometimes the peak signals in the chromatogram appear normal, but the file can contain wrongly unannotated or additional annotated nucleotides (Fig. 3i). TIDE(R) gives a warning when the spacing between the nucleotides in the chromatogram of the sequence trace is not consistent, which is often an indication for wrongly unannotated or additional annotated nucleotides. In case of this warning, the chromatograms should be carefully investigated (use Snapgene or Chromas software).

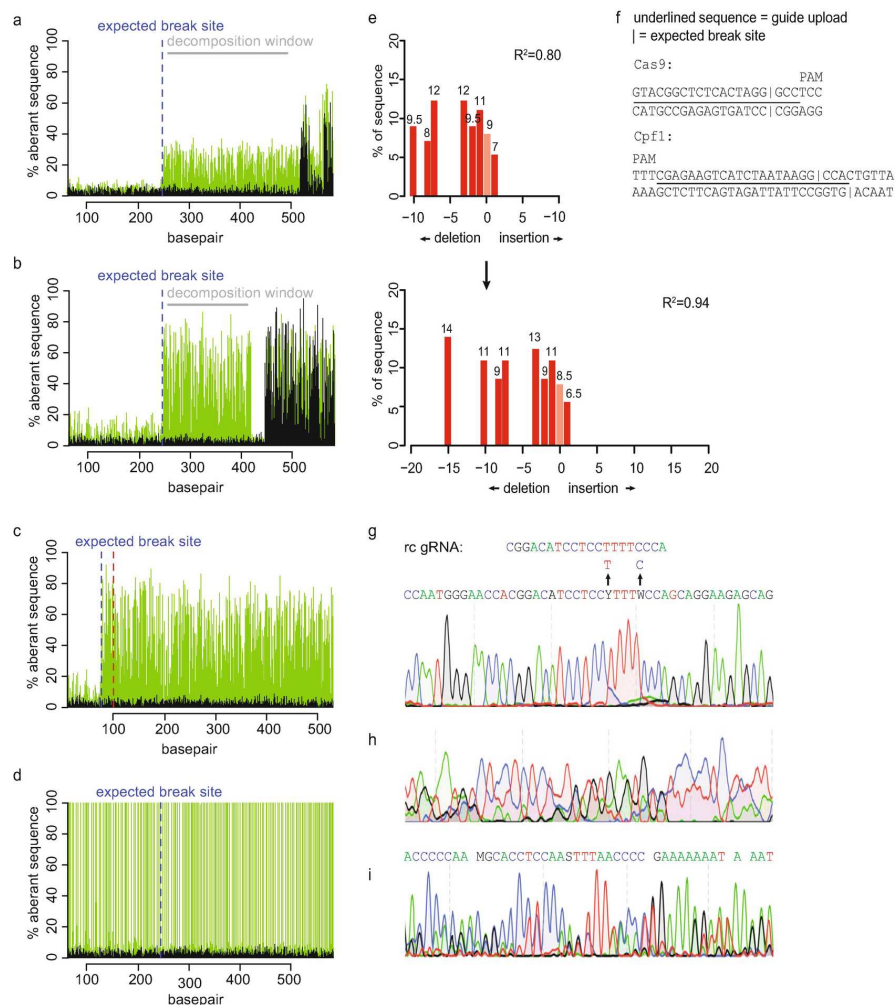


Fig. 3

Troubleshooting with TIDE and TIDER. All parameters in TIDE(R) have default settings but can be adjusted if necessary.

Different settings are often a remedy to solve error messages. (a–i) Examples of most common error messages with the recommended setting changes. (a, b) Avoid the decomposition window to overlap with high aberrant signal in the control. This occurs often near the ends of the sequence traces (a) or in a stretch of repetitive sequences (b). (c, d) Alignment problems can occur when the alignment window is too small (default is from 100 until 15 bp upstream break site) (c) or when the wrong nucleotides in the files are aligned (d). The alignment window can be changed closer or further to 1 in the sequence trace. (e)

The presence of indels larger than the default of 10 is not included in decomposition and can result in low R^2 score. Indel size can be changed. (f) The use of other nucleases than Cas9 in TIDE(R) works when the guide RNA string is mimicked to the 20 nt Cas9 guide RNA that cuts between nucleotide 17 and 18. (g) A mismatch in the IUPAC nucleotide annotation that prevents the recognition of guide RNA in the control sequence can be solved by editing the chromatogram file to the expected nucleotides. (h, i) Poor sequence quality will not give reliable results in TIDE(R)

Acknowledgements:

We thank Marcel de Haas, Stefano Manzo, and Ruben Schep for critical reading of the manuscript. This work was supported by the Netherlands Organization for Scientific Research ZonMW-TOP grant 91211061, and European Research Council Advanced Grant 694466.

Competing Interests: EKB and BvS declare competing financial interests: As inventors of TIDE and TIDER software, they receive licensing payments under their employer's rewards-to-inventors scheme.

References:

- Brinkman EK, Chen T, de Haas M, Holland HA, Akhtar W, van Steensel B (2018) Kinetics and fidelity of the repair of Cas9-induced double-strand DNA breaks. *Mol Cell* 70:801–813 [CrossRef](#) [Google Scholar](#)
- Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA et al (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819–823 [CrossRef](#) [Google Scholar](#)
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided

- DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821 [CrossRef](#) [Google Scholar](#)
4. Mali P, Esvelt KM, Church GM (2013) Cas9 as a versatile tool for engineering biology. *Nat Methods* 10:957–963 [CrossRef](#) [Google Scholar](#)
 5. Yoshimi K, Kunihiro Y, Kaneko T, Nagahora H, Voigt B, Mashimo T (2016) ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. *Nat Commun* 7:10431 [CrossRef](#) [Google Scholar](#)
 6. Inui M, Miyado M, Igarashi M, Tamano M, Kubo A, Yamashita S, Asahara H, Fukami M, Takada S (2014) Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system. *Sci Rep* 4:5396 [CrossRef](#) [Google Scholar](#)
 7. Storici F, Snipe JR, Chan GK, Gordenin DA, Resnick MA (2006) Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing. *Mol Cell Biol* 26:7645–7657 [CrossRef](#) [Google Scholar](#)
 8. Wang Z, Zhou ZJ, Liu DP, Huang JD (2008) Double-stranded break can be repaired by single-stranded oligonucleotides via the ATM/ATR pathway in mammalian cells. *Oligonucleotides* 18:21–32 [CrossRef](#) [Google Scholar](#)
 9. Ma H, Marti-Gutierrez N, Park SW, Wu J, Lee Y, Suzuki K, Koski A, Ji D, Hayama T, Ahmed R et al (2017) Correction of a pathogenic gene mutation in human embryos. *Nature* 548(7668):413–419 [CrossRef](#) [Google Scholar](#)
 10. Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE (2016) Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* 34:339–344 [CrossRef](#) [Google Scholar](#)
 11. Horlbeck MA, Witkowsky LB, Guglielmi B, Replogle JM, Gilbert LA, Villalta JE, Torigoe SE, Tjian R, Weissman JS (2016) Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *Elife* 5:e12677 [CrossRef](#) [Google Scholar](#)
 12. Yang L, Guell M, Byrne S, Yang JL, De Los Angeles A, Mali P, Aach J, Kim-Kiselak C, Briggs AW, Rios X et al (2013) Optimization of scarless human stem cell genome editing. *Nucleic Acids Res* 41:9049–9061 [CrossRef](#) [Google Scholar](#)
 13. Brinkman EK, Chen T, Amendola M, van Steensel B (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* 42:e168 [CrossRef](#) [Google Scholar](#)
 14. Brinkman EK, Kousholt AN, Harmsen T, Leemans C, Chen T, Jonkers J, van Steensel B (2018) Easy quantification of template-directed CRISPR/Cas9 editing. *Nucleic Acids Res* 46(10):e58 [CrossRef](#) [Google Scholar](#)
 15. Green MR, Sambrook J (2012) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY [Google Scholar](#)
 16. Kunkel TA (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci U S A* 82:488–492 [CrossRef](#) [Google Scholar](#)
 17. van Overbeek M, Capurso D, Carter MM, Thompson MS, Frias E, Russ C, Reece-Hoyes JS, Nye C, Gradia S, Vidal B et al (2016) DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Mol Cell* 63:633–646 [CrossRef](#) [Google Scholar](#)
 18. Dorsett Y, Zhou Y, Tubbs AT, Chen BR, Purman C, Lee BS, George R, Bredemeyer AL, Zhao JY, Soderger E et al (2014) HCoDES reveals chromosomal DNA end structures with single-nucleotide resolution. *Mol Cell* 56:808–818 [CrossRef](#) [Google Scholar](#)
 19. Beumer KJ, Trautman JK, Mukherjee K, Carroll D (2013) Donor DNA utilization during gene targeting with zinc-finger nucleases. *G3* 3:657–664 [CrossRef](#) [Google Scholar](#)

MATERIALS TEXT

2. Materials

2.1 Guide RNA Design

TIDE and TIDER are suitable for any species in which genomic editing experiments can be performed. CRISPR guide RNAs can be designed using various online design web tools (e.g., <http://crispr.mit.edu/>, <https://chopchop.rc.fas.harvard.edu/>, <https://www.deskgen.com/>).

2.2 DNA Purification Buffers and Solutions

Usually, 1–3 days after transfection genomic DNA is isolated. Genomic DNA of a minimum of 1000 cells should be isolated

to get a comprehensive sampling of the complexity of the mutations that are introduced by the repair of the CRISPR-Cas9 double strand break. A standard genomic DNA isolation Kit (e.g., BioLine ISOLATE II Genomic DNA Kit) can be used according to the manufacturer's protocol. DNA can also be isolated with the protocol for isolation of high-molecular-weight DNA from mammalian cells using proteinase K and phenol/chloroform extraction [15].

2.3 PCR Amplification of Control and Experimental Sample DNA

PCR reactions are carried out with primers surrounding the expected break site. We advise to amplify and sequence a stretch of DNA 500–1500 bp enclosing the designed editing site. The projected break site should be located preferably ~200 bp downstream from the sequencing start site.

1. Genomic DNA.
2. PCR primers (*see* **Note 1**).
3. PCR master mix (example makes 50 µL):

A	B
21-x uL	H2O
2 uL	Primer a (10 µM stock)
2 uL	Primer b (10 µM stock)
x uL	Genomic DNA (~50 ng)
25 uL	2× pre-mix of buffer, Taq polymerase, and dNTPs (e.g., BioLine MyTaq)

4. PCR program:

A	B	C	D
Step	Temperature	Time (min:s)	Number of cycles
Initial denaturation	95 °C	1:00	1
Denaturation	95 °C	0:15	25-30x
Annealing	55–58 °C	0:15	
Extension	72 °C	0:10	
Final extension	72 °C	1:00	
	4 °C	Hold	

5. Check an aliquot of the PCR product on 1-2% agarose gel. A sharp single band should be visible.
6. Purify the PCR product using a kit according to manufacturer's instructions (e.g., BioLine ISOLATE II PCR and Gel Kit).

2.4 Two-Step PCR Amplification of Reference DNA (TIDER Only)

1. Genomic DNA .
2. PCR primers (*see* **Notes 2 and 3**).
3. PCR master mix (example makes 50 µL):

A	B	C
	PCR mix1	PCR mix2
21-x µL	H2O	H2O
2 µL	Primer a (10 µM stock)	Primer d (10 µM stock)
2 µL	Primer c (10 µM stock)	Primer b (10 µM stock)
x µL	Genomic DNA (~50 ng)	Genomic DNA (~50 ng)
25 µL	2× pre-mix of buffer, Taq polymerase, and dNTPs (e.g., BioLine MyTaq)	2× pre-mix of buffer

4. PCR program:

A	B	C	D
Step	Temperature	Time (min:s)	Number of cycles
Initial denaturation	95 °C	1:00	1
Denaturation	95 °C	0:15	25x
Annealing	55–58 °C	0:15	
Extension	72 °C	0:10	
Final extension	72 °C	1:00	
	4 °C	Hold	

5. Purify PCR product using kit and manufacture instructions (e.g., BioLine ISOLATE II PCR and Gel Kit).

6. Anneal the following two PCR products for 1 min 95 °C, cool down to 20 °C (0.1 degrees/s).

A	B
48 µL	Annealing buffer (=10 mM Tris, 50 mM NaCl, 1 mM EDTA)
1 µL	PCR mix1
1 µL	PCR mix2

7. Extend the annealed products and amplify the joined product.

A	B
18 µL	H2O
2 µL	Primer a (10 µM stock)
2 µL	Primer b (10 µM stock)
3 µL	Annealed oligo mix
25 µL	2× pre-mix of buffer, Taq polymerase, and dNTPs (e.g., BioLine MyTaq)

8. PCR program:

A	B	C	D
Step	Temperature	Time (min:s)	Number of cycles
Initial denaturation	95 °C	1:00	1
Denaturation	95 °C	0:15	25x
Annealing	55–58 °C	0:15	
Extension	72 °C	0:10	
Final extension	72 °C	1:00	
	4 °C	Hold	

9. Check the PCR product on 1–2% agarose gel. A sharp single band should be visible.

10. Purify the PCR product using a kit and manufacturer's instructions (e.g., BioLine ISOLATE II PCR and Gel Kit).

2.5 Sanger Sequencing

We strongly recommend that all PCR products (control, experimental sample(s), and for TIDER also the reference) are sequenced in parallel. Purified PCR samples are prepared for Sanger sequencing with the following protocol or can be send for commercial Sanger sequencing.

1. Purified PCR samples (100 ng).

2. PCR primers. Similar primers as in Section 2.3 above can be used (*see* **Notes 1 and 3**).

3. PCR master mix (example makes 20 µL):

A	B
15.5-x µL	H2O
0.5 µL	Primer a or primer b (10 µM stock)
x µL	Purified PCR samples (100 ng)
4 µL	BigDye (e.g., BigDye® terminator v3.1 of Applied Biosystems)

PCR program:

A	B	C	D
Step	Temperature	Time (min:s)	Number of cycles
Initial denaturation	96 °C	1:00	1
Denaturation	96 °C	0:30	30x
Annealing	50 °C	0:15	
Extension	60 °C	4:00	
	4 °C	Hold	

4. Samples are analyzed by a Sanger sequence instrument (e.g., Applied Biosystems 3730 × I DNA Analyzer). Sequence trace files must be saved in .ab1 or .scf format.

2.6 Equipment

1. Cell counter
2. Microcentrifuge
3. PCR cycler
4. Nanodrop

2.7 Software

The TIDE and TIDER web tools are both available at <https://tide.nki.nl> or <https://deskgen.com>.

SAFETY WARNINGS

Personal protective equipment (PPE) should be worn at all times (gloves, lab coat, and eye protection).

For hazard information and safety warnings, please refer to the SDS (Safety Data Sheet).

3.1 Control and Experimental Sample Generation

- 1 For both methods genomic DNA is isolated from the cell pool that was transfected with the nuclease or guide RNA alone (control) and from cells exposed to both Cas9 and guide RNA (experimental sample). For TIDER the experimental sample is also co-transfected with the donor template.
- 2 Then a region of about 500–1500 base pairs around the target site is amplified by PCR from DNA of the control and experimental sample (Fig. 1a, b).

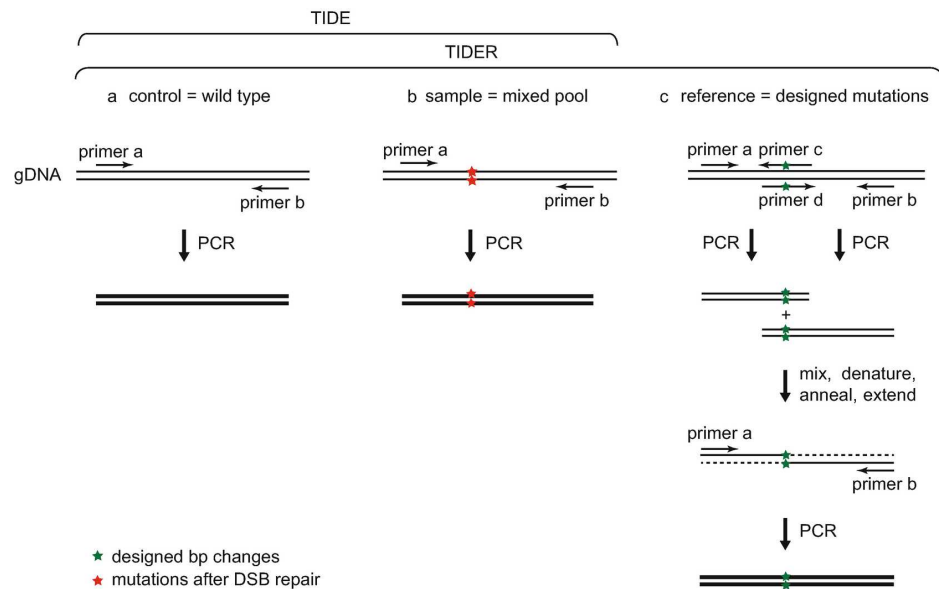


Fig. 1
Method to generate the required input samples for TIDE and TIDER. Control and test samples can be obtained by PCR using primers spanning the CRISPR target site (primers a, b). The reference sequence (TIDER only) can be created in a similar way as site-directed mutagenesis [16] (see Section 3.2 for detailed explanation)

- Next, the PCR amplicons are subjected to conventional Sanger sequencing. In the PCR product of the experimental sample, the sequence trace may consist of a combination of multiple sequences derived from unmodified DNA and DNA that has acquired a mutation (Fig. 2a).

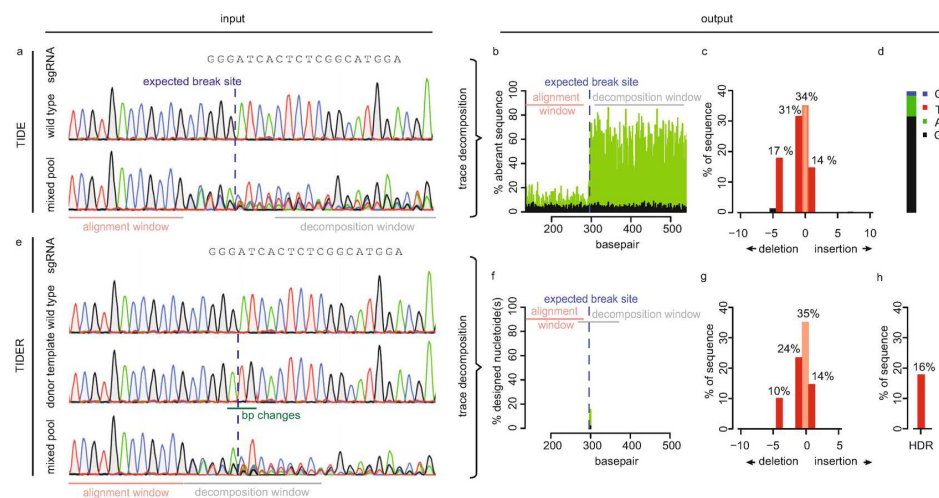


Fig. 2
Overview of TIDE and TIDER algorithm. Due to imperfect repair (and repair by homology-directed repair with a donor template) after cutting by a targeted nuclease, the DNA in the cell pool consists of a mixture of indels (and designed mutations). The various introduced mutations in the pool are disentangled by TIDE or TIDER. (a) TIDE requires as input a guide RNA sequence string and two sequences are required: (1) wild-type control, (2) composite test sample. (b) For quality control the aberrant sequence signal is visualized in control (black) and treated sample (green), the expected break site (vertical dotted line), region used for alignment (pink bar), and the region used for decomposition (gray bar). A constant composite sequence signal is yielded after the break site. (c) Trace decomposition yields the spectrum of indels with their frequencies. (d) In presence of +1 insertions, the base composition is estimated. (e) Input files for TIDER are identical to TIDE and one additional sequence file with designed mutations in the used donor template. (f) Quality plot showing only the proportion of desired mutated nucleotide(s) as designed in donor template that is/are present in the control (black) and treated sample (green). The region for alignment (pink bar) and decomposition (gray bar) as used in TIDER are represented. (g and h) Decomposition gives the spectrum of indels (g) and the HDR events (h) with their frequencies

3.2 Reference Sample Generation (TIDER Only)

- 4 TIDER is required for genome editing experiments in the presence of a donor template. In addition to the control and experimental sample trace (*see* Section 3.1), TIDER requires one extra Sanger sequencing trace called “reference.” The reference is similar to the control sequence, except that it carries the desired base pair changes as designed in the donor template (Fig. 2e). There are two paths to obtain the reference sequence as described below:

- 4.1 The reference sequence can be easily created in a 2-step PCR protocol based on site-directed mutagenesis [16].

Here, two additional primers are required that overlap and carry the desired mutation(s) (mutated primers c, d, which are reverse complement of each other) (Fig. 1c). These primers are used in combination with the primers used for the amplification of the control and experimental sample (control primers a, b). The control forward primer a is combined with the reverse mutated primer c and the forward mutated primer d with the control reverse primer b, resulting in two PCR amplicons that incorporate the designed mutations.

Then the two amplicons are denatured and hybridized at the complementary ends in an annealing reaction. The second PCR uses the annealing mixture as a template and the control forward and reverse (primers a and b) as primers. This PCR starts with an extension step followed by exponential amplification. This results in a PCR product carrying the designed mutations (*see* **Notes 2** and **3**).


- 4.2 Alternatively, the reference DNA can be ordered as synthesized DNA. The design should include a similar DNA code as the PCR product of the control sample, except that it should carry the designed mutation(s) as in the donor template. The annealing sequences for the forward and reverse primers (a, b) should also be present in the synthesized fragment. Similar to the control and test sample, the reference can be amplified with primer a, b (*see* **Note 3**).

3.3 Web Tool

- 5 

Process the PCR products of the control, optional reference, and experimental sample by conventional Sanger sequencing.

- 6 The resulting sequence trace files (.ab1 or .scf format) are then uploaded into the TIDE or TIDER web tool (both available at <http://tide.nki.nl> and <https://deskgen.com>). In addition, a character string representing the guide RNA sequence (20 nt) is required as input (*see* **Notes 4** and **5**).

- 7 

Then, the software will perform several calculations.

First, the guide RNA sequence is aligned to the control sequence in order to determine the position of the expected Cas9 break site.

Next, in all Sanger sequence traces an alignment window is automatically selected that runs from 100 to 15 bp upstream of the break site. The sequence segment in this window of the experimental sample (and the optional reference) is aligned to that of the control in order to determine any offset between the sequence reads. Users may change the default settings for these calculations, which is necessary when alignment problems occur with these settings (*see* **Notes 6** and **7**).

Subsequently, two output plots are generated: one plot that can help with quality control and one that displays the indel/HDR spectrum.

3.4 Quality Control

For generation of the quality control plot the signals of all nucleotides: A, G, T, C at each position in the sequence file are

8 used. In general, each position in the sequence trace is represented by one predominant nucleotide signal indicative of the actual nucleotide. The minor signals from the other three nucleotides are normally considered as background. In TIDE(R) the percentage of these aberrant nucleotides is plotted along the sequence trace of the control and the experimental sample. Thus, a value of 0% at a position indicates that the detected nucleotide does not differ from the control sequence while a value of 100% indicates that the expected nucleotide was not detected at all (and instead only one or more of the other three nucleotides) (Fig. 2b). The percentages of aberrant nucleotides in the control should be low along the whole sequence trace. However, the experimental sample consists of a mixture of multiple sequences due to the presence of indels and possible point mutations. Around the break site the sequences start to deviate from the control, which is visible with consistently elevated signal of the aberrant sequence signal. Note that there is a 25% chance that an identical nucleotide in a mutated sequence is found as is present in the control sequence at the same position, because there are only 4 different nucleotides available. This plot allows the user to visually inspect the sequence deviation caused by the targeted nuclease and enables to verify the alignments and quality of the data. It is important to confirm that:

- (1) the break site is located as expected,
- (2) the aberrant signal is only increasing around the break site and
- (3) remains elevated downstream of the break site.

The sequence trace downstream of the break site is decomposed into its individual sequence components. The region used for this purpose is marked as the decomposition window. All parameters in TIDE(R) have default settings but can be adjusted if necessary. The user can interactively change the alignment and decomposition windows. Choosing a different decomposition window is often a remedy to circumvent locally poor sequence traces, which should be avoided (see **Notes 8-10**).

9 For TIDER two additional quality plots are generated.

In one, the aberrant signal of the reference trace compared to the control trace is plotted. This can be used to verify whether the designed mutation(s) is/are present at the expected location.

In the second one, the percentage of the designed mutation(s) present in the experimental sample is plotted, representing the relative incorporation of the donor template (Fig. 2f).

3.5 Mutation Detection by Decomposition

10 For the detection of individual mutations with the corresponding frequencies, the TIDE and TIDER software perform a decomposition of the mixed sequence signal in the experimental sample. This composite sequence trace is a linear combination of the wild type (control) and the mutated sequences. For TIDE, the decomposition is performed on a sequence segment downstream of the break site. As a rule of thumb, the larger the decomposition window is chosen, the more robust the estimation of mutations is (see **Note 9**).

To perform the decomposition, generate a set of sequence trace models that contain all possible indels of size {0..n} (n is by default set to 10).

The models are derived from the control trace and contain all nucleotide peak signals of the decomposition window shifted by the appropriate number of positions to the left or right.

A wild-type trace (shift 0) is also added as a model.

Then, using non-negative linear modeling the combination of trace models that can best explain the composite sequence trace in the experimental sample is determined (Fig. 2c) (see **Note 11**). An R^2 value is calculated as a measure of the goodness of fit (see **Notes 10 and 12**), and the statistical significance of the detection of each indel is calculated.

11 For TIDER the mutation detection is more complex. It is mandatory that the decomposition window in TIDER covers the location of the designed mutation(s) in the donor template (see **Notes 9 and 13**). In contrast to TIDE, the decomposition window of TIDER spans by default only 100 bp. In case only few base pair changes are introduced, the sequence with the designed mutation will be very similar to the wild-type sequence. The smaller decomposition window of TIDER emphasizes the difference between the control and reference better.

Simulations of all possible insertions and deletions are generated from the control file and placed in a decomposition

matrix together with the control and reference. Subsequently, decomposition of the experimental sample is performed thereby choosing the best combination of the models in the decomposition matrix. This results in an estimation of the incorporation frequency of template-directed mutation(s) and distinguishes these from the background of indels that are introduced by error-prone repair (*see* **Note 14**).

- 12 The reliability of TIDE and TIDER depends on the quality of the input samples (*see* **Note 15**). For an accurate TIDE(R) estimation it is recommended that (1) $R^2 > 0.9$ and (2) aberrant signals upstream of the break site are below 10% in the quality plot. This applies to all files: control, reference, and experimental sample. To verify the results the samples can be sequenced from the opposite strand (*see* **Note 13**).

3.6 Sequence Determination of the +1 Insertion (TIDE Only)

- 13 During repair of CRISPR-Cas9 a single base pair is frequently inserted at one of the DNA ends of the break [13, 17, 18]. TIDE provides an estimate of the base composition of this insertion. This may be of interest if one wishes to obtain a particular sequence variant (Fig. 2d). For longer insertions this base calling is computationally complicated and currently not implemented.