A detection method for the capture of genomic signatures: From disease diagnosis to genome editing

Orléna Benamozig[a,b,c], Lou Baudrier[a,b,c], and Pierre Billon[a,b,c,*]

[a]The University of Calgary, Cumming School of Medicine, Department of Biochemistry and Molecular Biology, Calgary, AB, Canada
[b]Robson DNA Science Centre, Calgary, AB, Canada
[c]Arnie Charbonneau Cancer Institute, Calgary, AB, Canada
[*]Corresponding author: e-mail address: pierre.billon@ucalgary.ca

Contents

Methods in Enzymology
ISSN 0076-6879
https://doi.org/10.1016/bs.mie.2021.08.012

Abstract

Variations in the genetic information originate from errors during DNA replication, error-prone repair of DNA damages, or genome editing. The most common approach to detect changes in DNA sequences employs sequencing technologies. However, they remain expensive and time-consuming, limiting their utility for routine laboratory experiments. We recently developed DinucleoTidE Signature CapTure (DTECT). DTECT is a marker-free and versatile detection method that captures targeted dinucleotide signatures resulting from the digestion of genomic amplicons by the type IIS restriction enzyme AcuI. Here, we describe the DTECT protocol to identify mutations introduced by CRISPR-based precision genome editing technologies or resulting from genetic variation. DTECT enables accurate detection of mutations using basic laboratory equipment and off-the-shelf reagents with qualitative or quantitative capture of signatures.

1. Introduction

Nucleic acids encode critical information for human physiology and medicine (Miga et al., 2020). A spectrum of variations contribute to the evolution and adaptation of species, underlies, or predisposes to disorders, and impacts disease burden (Ellegren & Galtier, 2016; Genomes Project et al., 2015; The International HapMap Consortium, 2015; Karczewski et al., 2020; Landrum et al., 2020; Lek et al., 2016). Mutations originate from error-prone repair of DNA damage induced by cellular processes or environmental mutagens and from errors introduced during replication but unresolved by the mismatch repair pathway (Alexandrov et al., 2013; Helleday, Eshtad, & Nik-Zainal, 2014; Zou et al., 2021). Although decoding genetic variants with a role in human disease is crucial, the large fraction of documented variants is uncharacterized. Connecting variants to phenotypes requires installing desired mutations with genome editing technologies, and methods to detect nucleic acid changes.

CRISPR-based precision genome editing technologies, such as CRISPR-mediated homology-directed repair, base editing, or prime editing, introduce desired modifications by directing the activity of nucleases and mutagenic enzymes at desired genomic locations (Anzalone, Koblan, & Liu, 2020; Billon et al., 2017). For instance, CRISPR-prime editing utilizes an engineered reverse transcriptase that copies the information encoded in an

RNA template directly in the genome, thereby enabling the insertion of any desired genomic changes. These technologies introduce site-specific DNA lesions, such as double-strand breaks (DSBs), resolved by various cellular DNA repair mechanisms (Nambiar et al., 2019; Yeh, Richardson, & Corn, 2019). Precision genome editing enables manipulating genetic information in cellular and animal models for functional interrogation of coding sequences and non-coding regulatory elements (Pickar-Oliver & Gersbach, 2019). Our ability to rapidly and accurately model or correct disease-associated variants in model systems is crucial to accelerate basic and translational research.

Typical approaches used to detect nucleic acid sequences are mostly limited to sequencing technologies, primer- or probe-specific mutations, or approaches that require specific instrumentation (Brinkman, Chen, Amendola, & van Steensel, 2014; Clement et al., 2019; Findlay, Vincent, Berman, & Postovit, 2016; Guell, Yang, & Church, 2014; Lindsay et al., 2016; Qiu et al., 2004). However, these methods require sophisticated detection strategies with weak sensitivity, precision, and accuracy, limiting their broad adoption (Germini et al., 2018). Also, despite recent spectacular advances, sequencing technologies remain expensive and time-consuming for basic laboratory experiments. Therefore, there is a need for rapid, versatile, and easy to implement laboratory detection assays.

Restriction endonucleases and ligases have played a pivotal role in the DNA recombinant technology revolution by enabling the assembly of selected DNA molecules (Khan et al., 2016). Restriction endonucleases are members of restriction-modification systems, a primitive bacterial innate immune system that protects hosts by destroying mobile genetic elements (Labrie, Samson, & Moineau, 2010). Endonucleases have revolutionized the DNA repair field by developing unique tools that simplify the efficient insertion of site-specific DSBs (Berkovich, Monnat, & Kastan, 2007; Chailleux et al., 2014; Rouet, Smih, & Jasin, 1994; Shanbhag, Rafalska-Metcalf, Balane-Bolivar, Janicki, & Greenberg, 2010). DNA ligases exhibit critical enzymatic activities during DNA replication and repair (Pascal, 2008). Ligases seal two DNA ends by forming phosphodiester bonds between 3′-hydroxyl and 5′-phosphoryl termini (Green & Sambrook, 2019). Endonucleases and ligases are ubiquitous tools in modern molecular biology for laboratory DNA manipulation and genetic studies. In particular, restriction enzymes from the type IIS family ("S" for "shifted," because DNA cleavage is shifted to one side of the recognition motif) have been utilized in molecular biology for their properties to cleave unknown sequences (Drmanac et al., 2010; Klompe, Vo, Halpin-Healy, & Sternberg, 2019; Pingoud, Wilson, & Wende, 2014; Ran et al., 2013; Shendure et al., 2005).

Here, we describe a rapid-to-execute and easy-to-implement protocol that utilizes AcuI, a type IIS restriction endonuclease discovered in *Acinetobacter calcoaceticus*, and a DNA ligase to identify nucleic acid signatures (Billon et al., 2020). AcuI is programmed to expose desired nucleic acid signatures that are subsequently captured by the ligation of standard DNA adaptors. This protocol allows the quantitative and qualitative identification of genetic variation induced by precision genome editing or natural variation. We illustrate the protocol by detecting two mutations introduced with CRISPR technologies in the *DPY30* gene, which encodes for a core subunit of the SET1/MLL histone methyltransferase complex (Xue et al., 2019), and the *SAGE1* (sarcoma antigen 1) gene, which encodes for a cancer biomarker and target for immunotherapy (Djureinovic et al., 2016). We introduce these transversion mutations in human cells using CRISPR-mediated prime editing (Anzalone et al., 2019) and use DTECT to determine editing efficiency (Billon et al., 2020). We also describe how positive and negative controls from non-edited samples are used to generate high confidence results. Finally, we discuss the strengths and limitations of DTECT compared to other detection methods.

2. Overview of the procedure

DTECT is a capture-based detection method that requires the digestion of a particular amplicon and ligation of specific adaptors (Billon et al., 2020). The experimental workflow of DTECT comprises six steps (Fig. 1). Briefly, DTECT requires the selection of the desired dinucleotide that includes the variant base(s). The selection of a "*dinucleotide signature*" of interest enables the design of a particular primer named "AcuI-*tagging primer*" (Fig. 1, Step I). This primer is adjacent to the targeted dinucleotide (Fig. 1, Step I dinucleotide in red) and amplifies the DNA sequence by PCR (Fig. 1, Step II). The AcuI-tagging PCR allows the insertion of an AcuI motif (5′-CTGAAG-3′) 14 bp upstream of the dinucleotide of interest (Fig. 1, Step II in green) and the addition of a 5′ handle used to detect the ligated product in step VI (Fig. 1, Step II in blue). The resulting amplicon is then digested with AcuI to generate two products of digestion, a small fragment (60 bp) and a long DNA fragment (>100 bp) (Fig. 1, Step III). Next, the small DNA fragment (60 bp), which contains the 3′ overhang dinucleotide signature, is isolated using solid-phase reversible immobilization (SPRI) magnetic beads (Fig. 1, Step IV). Then, the dinucleotide signature is captured by ligating complementary or non-complementary DNA adaptors (Fig. 1, step V). The adaptors are selected from a library of 16 standard adaptors that covers all the possible dinucleotides (Fig. 2A).

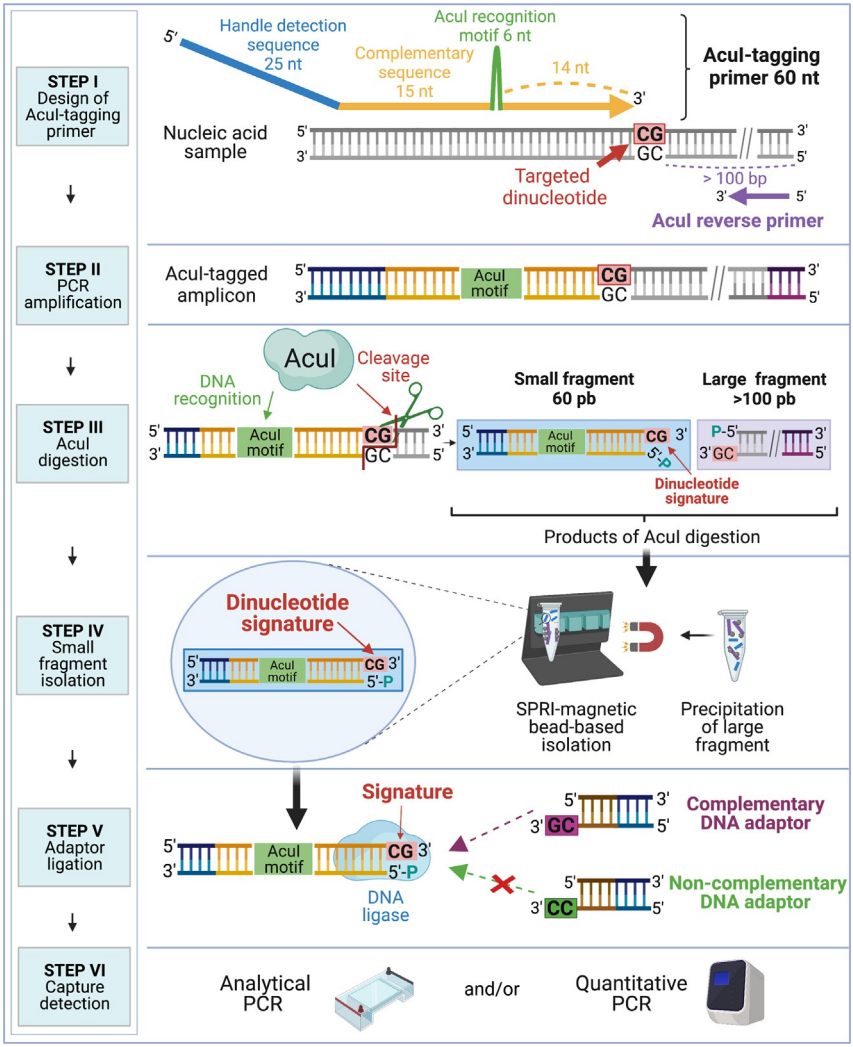

Fig. 1 Schematic representation of DTECT. DTECT is composed of six steps (Step I–VI). Step I: The targeted genomic locus contains a dinucleotide of interest (CG, in red). The AcuI-tagging oligonucleotide contains a detection handle (blue), two complementary sequences to the genomic locus (yellow) and an AcuI hairpin (green). The AcuI-tagging oligonucleotide is adjacent to the targeted dinucleotide. The reverse primer is positioned at >100 bp from the targeted dinucleotide (purple). Step II: The selected dinucleotide signature is amplified. Step III: AcuI digestion of the AcuI-tagged amplicon generates a 3′ dinucleotide overhang at the dinucleotide of interest (CG) and 5′ phosphates at the cleavage site. Two products of digestion are generated, a small fragment of 60 bp and a large fragment of >100 bp. Step IV: The small fragment is isolated using SPRI-magnetic beads. Step V: The signature is captured using complementary (CG) and non-complementary (CC) adaptors. Step VI: The captured material is amplified and revealed by analytical gel or quantitative PCR.

The ligated product serves as a template for detection by analytical or quantitative PCR (Fig. 1, Step VI) using unique primers that anneal to the 5′ end of the AcuI-tagging oligo and 3′ end of the adaptors (Fig. 1, Step V in blue). These PCRs quantify and reveal the presence of the genomic signatures in the original nucleic acid samples.

3. Materials and equipment

Here is the list of materials, reagents, and equipment required for DTECT. All reagents are commercially available and can be purchased from different sources. Moreover, the materials required for DTECT constitute standard equipment for a basic biology laboratory.

3.1 Key resources table

Reagent or resource	Source	Identifier
Chemicals, peptides, and recombinant proteins		
AcuI	NEB	R0641
T4 DNA Ligase	Invitrogen	15224025
Q5 High-Fidelity DNA Polymerase	NEB	M0491
dNTPs	NEB	N0447
Loading Dye	NEB	B7024
1 kb Plus DNA Ladder	NEB	N3200
Critical commercial assays		
Agencourt AMPure XP magnetic beads	Beckman Coulter	A63881
Power SYBR Green PCR Master Mix	Applied Biosystems	4367659
SYBR Gold Nucleic Acid Gel Stain	Invitrogen	S11494
Zymoclean gel DNA recovery Kit	Zymo Research	D4008
Quick Extract DNA Extraction Solution	Lucigen	QE09050
Deposited data		
DTECT—Plasmid for standard curve	Addgene	139333
Experimental models: Cell lines		
HEK293T cells	ATCC	CRL-11268

Oligonucleotides		
Primers for PCR	This paper	Table 1
Software and algorithms		
Excel		Microsoft
Other		
QuantStudio 6 Flex Real-Time PCR System	Applied Biosystems	4484642
ProFlex 3 × 32-well PCR System	Applied Biosystems	4484073
ChemiDoc Touch Gel Imaging System	Bio-Rad	1708370
12-Tube Magnetic Separation Rack	NEB	S1509

3.2 Reagents

- Agarose (Fisher BioReagents, #BP160-500)
- Tris Base (Fisher BioReagents, #BP152-5)
- Ethylenediaminetetraacetic Acid (Fisher BioReagents, #BP120-1)
- Sodium Hydroxide (Fisher BioReagents, #BP359-212)
- Acetic Acid, Glacial (Fisher BioReagents, #BP2401-212)
- Sodium Chloride (Fisher BioReagents, #BP358-10)
- Dimethyl Sulfoxide (Fisher BioReagents, #BP231-1)
- Phosphate buffered saline (Sigma-Aldrich, #P4417-100TAB**)**
- Molecular Biology Grade Water (Fisher BioReagents, #BP281920)

3.3 Equipment

- NanoDrop 2000c Spectrophotometer (Thermo Scientific, #ND-2000C)
- Mini Centrifuge (DLAB Scientific, #9031001012)
- Vortex Mixer (DLAB Scientific, #8031102000)
- Centrifuge 5425 (Eppendorf, #5405000042)
- Centrifuge 5810R (Eppendorf, #022628258)
- Dry Block Heaters (VWR, #12621-104)
- PowerPac Basic Power Supply (Bio-Rad, #1645050)
- Blue Light Transilluminator (VWR, #76151-834)
- Horizontal Mini S Gel Electrophoresis System (VWR, #76314-718)

3.4 Consumables

- MicroAmp Optical 384-Well Reaction Plate (Applied Biosystems, #4309849)
- MicroAmp Optical Adhesive Film (Applied Biosystems, #4311971)
- PCR tubes (Sarstedt, #72.991.002)
- Microtubes 1.5 mL (Sarstedt, #72.690.301)
- Tips, 10 μL, Eppendorf/Gilson type (Sarstedt, #70.1116.210)
- Filter Pipette tip 1–20 μL (Globe Scientific, #150910)
- Filter Pipette tip 1–200 μL (Globe Scientific, #150920)
- Filter Pipette tip 1–1000 μL (Globe Scientific, #150935)
- Razor blades (VWR, #55411-050)

3.5 Reagent setup and stock solutions

Timing: 1 h

– Preparation of the 500 mM EDTA stock solution: 186 g EDTA is dissolved into 800 mL of distilled H_2O. The pH is adjusted to 8.0 with NaOH. Finally, the final volume is adjusted to 1 L with distilled H_2O.
– Preparation of the 1 M Tris HCl pH 8.0 stock solution: 121.1 g of tris base is dissolved into 800 mL of distilled H_2O. HCl is slowly added until the pH reaches 8.0. Then, the final volume of the solution is adjusted to 1 L with distilled H_2O.
– Preparation of TAE 50 ×: The 50 × concentrated stock solution is prepared by dissolving 242 g of tris base into ~700 mL distilled H_2O. Then, 57.1 mL of glacial acetic acid and 100 mL of 500 mM EDTA (pH 8.0) are added. Finally, the final volume is adjusted to 1 L with distilled H_2O.
– The TE (10 mM Tris HCl pH 8.0, 1 mM EDTA pH 8.0) solution is prepared from the 1 M Tris HCl pH 8.0 and 500 mM EDTA pH 8.0 stock solutions. 1 mL of 1 M Tris HCl and 200 μL of 500 mM EDTA are mixed into 98.8 mL distilled H_2O.
– Dilution of the SYBR Gold solution: 10 μL of the SYBR Gold stock solution is diluted in 990 μL DMSO.

3.6 Software for analysis

– Spreadsheet software such as Microsoft Excel, Apple Numbers, or Google Sheets is needed for statistical analysis and quantification.

3.7 Preparation of the adaptor library

Timing: ~30 min

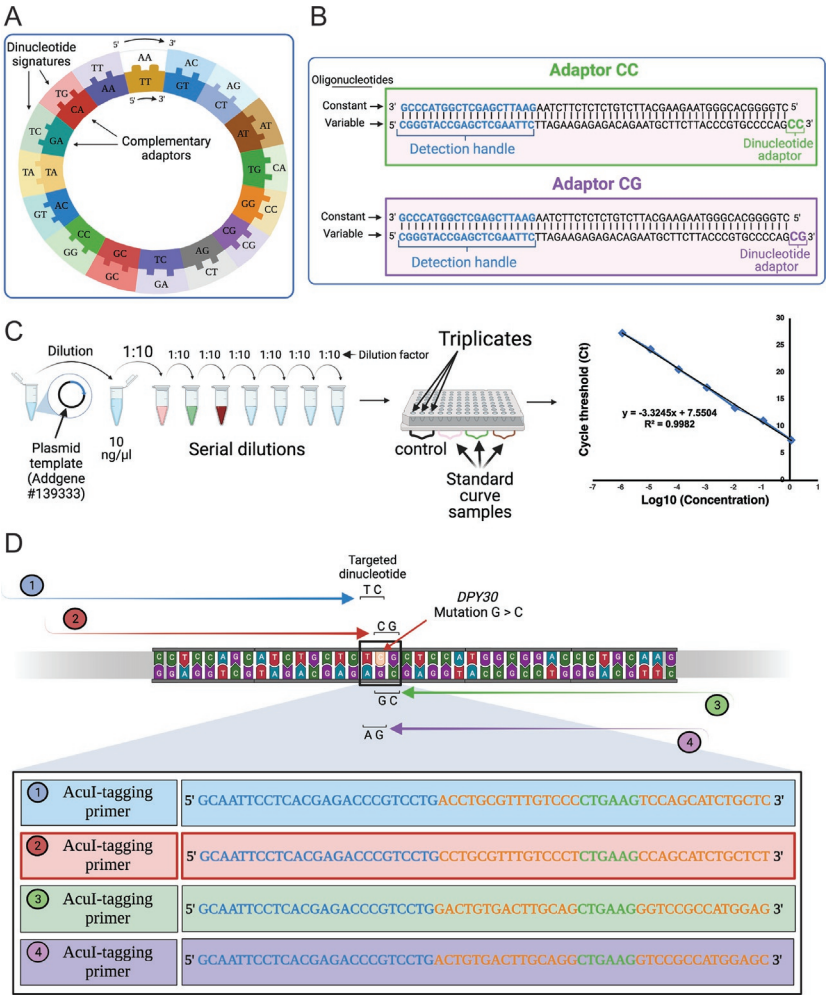

Fig. 2 Preparation of the library of adaptors and standard curve, and illustration of the flexibility of DTECT. (A) Illustration of the library of 16 adaptors and their respective complementary dinucleotide signatures. (B) Sequences of the complementary constant and variable oligonucleotides to generate the dsDNA adaptors. The common detection handle (blue) and the dinucleotides of the CC and CG adaptors (green and purple) are represented. (C) Steps required for the generation of the standard curve. Serial dilutions of a plasmid template are prepared and tested in triplicate to generate a unique standard curve common to all DTECT experiments. (D) Schematic representation of the targeted *DPY30* sequence surrounding the edited base (C > G). Four independent dinucleotides from the two strands can be generated to detect the base of interest. The sequences for the four respective AcuI-tagging primers used to capture the respective signatures are indicated.

A unique library of 16 double-stranded DNA adaptors is used to capture each possible dinucleotide signature. Fig. 2A illustrates the signature/adaptor combinations. This library is prepared from 17 individual oligonucleotides, consisting of 16 variable oligonucleotides of 60 nt and one constant oligonucleotide of 58 nt (oligonucleotide sequences are in Table 1). Each variable oligonucleotide contains a 58 nt sequence complementary to the constant oligonucleotide and has one of the 16 dinucleotides at the 3′ end (Fig. 2B, green and purple). In addition, adaptors contain a handle sequence (Fig. 2B, in blue) to analyze the captured products (detailed in part 5).

The double-stranded DNA library is prepared by combining the constant oligonucleotide with each of the 16 variable oligonucleotides as described below.

a. Adaptor oligonucleotides are regular oligos resuspended at a concentration of 100 μM in TE buffer.

b. Oligonucleotides are assembled in a 20 μL reaction. 2.5 μL of the constant oligonucleotide (100 μM) and 2.5 μL of each variable oligonucleotide (100 μM) are mixed with 4 μL of 1 × ligase buffer (5 ×) and 11 μL H_2O.

c. The mixture containing the constant oligonucleotide and one of the variable oligonucleotides is annealed by incubating for 5 min at 95 °C, followed by a temperature decrease from 95 to 15 °C at a ramp of 0.5 °C/s.

d. After annealing, 100 μL of molecular biology grade H_2O is added to generate a stock of double-stranded DNA adaptors at 5 μM.

e. Adaptors are stored at −20 °C. Each adaptor preparation captures up to 240 reactions.

f. (*Optional*) The adaptors can be visualized on agarose or polyacrylamide gels to confirm successful annealing.

3.8 Generation of the standard curve

Timing: 120 min

The detection of the ligated product requires a single pair of primers that anneal to the 5′ handle of the AcuI-tagging oligos and 3′ end of the adaptors (Fig. 1, step V in blue). Consequently, a unique standard curve is necessary to detect the captured material in all experiments, irrespective of the mutation type or genomic locus.

Here is the protocol to generate the unique standard curve (Fig. 2C).

a. The plasmid containing a product of DTECT ligation (Addgene #139333) is diluted at a concentration of approximately 10 ng/μL.

Table 1 Oligonucleotides used in this study.

Name	Sequence (5′ → 3′)	Description
OB1	ctggggcacgggtaagaagcattctgtctctcttctaagaattcgagctcggtacccg	Constant adaptor
OB2	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**aa**	AA-Variable adaptor
OB3	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**ac**	AC-Variable adaptor
OB4	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**ag**	AG-Variable adaptor
OB5	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**at**	AT-Variable adaptor
OB6	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**ca**	CA-Variable adaptor
OB7	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**cc**	CC-Variable adaptor
OB8	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**cg**	CG-Variable adaptor
OB9	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**ct**	CT-Variable adaptor
OB10	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**ga**	GA-Variable adaptor
OB11	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**gc**	GC-Variable adaptor
OB12	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**gg**	GG-Variable adaptor
OB13	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**gt**	GT-Variable adaptor
OB14	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**ta**	TA-Variable adaptor
OB15	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**tc**	TC-Variable adaptor
OB16	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**tg**	TG-Variable adaptor

Continued

Table 1 Oligonucleotides used in this study.—cont'd

Name	Sequence (5′ → 3′)	Description
OB17	cgggtaccgagctcgaattcttagaagagagacagaatgcttcttacccgtgccccag**tt**	TT-Variable adaptor
OB18	gcaattcctcacgagacccgtcctg	DTECT qPCR for
OB19	cgggtaccgagctcgaattcttagaag	DTECT qPCR rev
OB210	gcaattcctcacgagacccgtcctgttgtgaagagaggatctgaagaaataacggccaac	AcuI tagging PCR *SAGE1*
OB211	taagttgaggtcacgaaagataattca	PCR *SAGE1* rev
OB212	gcaattcctcacgagacccgtcctgcctgcgtttgtccctctgaagccagcatctgctct	AcuI tagging PCR *DPY30*
OB213	gatgaaggaagattgtagctcgtag	PCR *DPY30* rev

List of oligonucleotide sequences used to generate the library of adaptors, detect the captured material, and detect *SAGE1* and *DPY30* mutations. OB1-OB17 are used to construct the library of adaptors (3.7). OB1 corresponds to the constant oligonucleotide. OB2-OB17 correspond to the variable oligonucleotides. The dinucleotide specific to each adaptor is underlined. OB18 and OB19 are primers used to detect the captured material by analytical or quantitative PCR (3.7, 5.1, and 5.2). OB210-OB213 are DTECT oligos to capture *SAGE1* and *DPY30* mutations.

The exact concentration of the dilution must be accurately determined with a spectrophotometer (e.g., nanodrop). In our example, the measured concentration is 11.4 ng/μL.

b. Several 10-fold serial dilutions of the plasmid are prepared by adding 1 μL of the plasmid in 9 μL of H_2O. This dilution corresponds to "sample 1" at approximatively 1 ng/μL (1.14 ng/μL in our example).

c. 1 μL of "sample 1" is added into 9 μL of H_2O to obtain "sample 2" at approximatively ~0.1 ng/μL (0.114 ng/μL in our example).

d. The successive 10-fold serial dilutions are repeated to obtain a total of 7 samples.

e. The qPCR master mix containing the DTECT primers, named OB18 and OB19 (sequences listed in Table 1), is prepared. Each sample is tested in technical triplicates. In addition, a "negative control" sample that consists of no DNA template is also tested.

Master mix component	Final concentration	Volume (μL)—10 μL/reaction	Volume (μL)—Master mix for 26 reactions
H_2O	–	3.8	98.8
SYBR Green (2×)	1×	5	130
OB18, 100 μM	1 μM	0.1	2.6
OB19, 100 μM	1 μM	0.1	2.6
Diluted samples	–	1	–

f. 9 μL of the master mix is loaded into 24 qPCR wells.

g. 1 μL of the respective sample dilutions is added to the wells in triplicate.

h. 1 μL of H_2O is added in the last three qPCR wells as negative controls.

i. The qPCR machine is set up with the following program:

Temperature (°C)	Time (s)	PCR phase	Number of cycles
50	120	Hold stage	1
95	600		
95	10	PCR stage	40
60	30		
95	15	Melting	1
60	60		
95	15		

j. The cycle threshold (Ct) values are retrieved for each triplicate sample (columns 1–3 in the table below).

k. The average Ct values obtained from the triplicates are calculated using the formula "=AVERAGE(A2:C2)" (column 4 in the table below). Next, the concentrations of the 10-fold dilutions are determined using the formula "=LOG10(E2)" (columns 5 and 6).

Triplicate 1 (Ct)	Triplicate 2 (Ct)	Triplicate 3 (Ct)	Average Ct	Concentration (ng)	Log10 (concentration)
7.173	7.690	7.527	7.464	1.14	0.057
11.195	11.025	10.958	11.059	0.114	−0.943
13.671	13.412	13.285	13.456	0.0114	−1.943
17.009	17.143	17.482	17.211	0.00114	−2.943
20.427	20.540	20.742	20.570	0.000114	−3.943
24.120	24.351	24.333	24.268	0.0000114	−4.943
27.567	27.171	27.208	27.315	0.00000114	−5.943
30.759	30.506	31.126	30.797	H_2O (control)	

l. The threshold cycle as a function of the dilution factor is plotted in a graph representing the linear regression of the "Average Ct" as a function of "log10 (concentration)."

Each experimental point of the standard curve should be aligned to confirm the inverse correlation between the log of the quantity of DNA and the Ct value. In addition, the efficiency of OB18 and OB19 primers can be calculated from the slope using the formula: $\text{Efficiency} = 10^{(-1/\text{slope})}$. Primer efficiency should be close to 100%, and the coefficient correlation (R^2) should be close to 1. In our example, the following standard curve formula determines the amount of captured material: $y = -3.3245 \times +7.5504$. Part 5.2 describes how to measure the relative frequency of genomic signatures using the standard curve.

4. Step-by-step dinucleotide signature capture

The DTECT protocol contains five steps for the signature capture (4.1–4.5) and a detection analysis step (5.1 and/or 5.2). The signature capture consists of (4.1) the design of two primers for the AcuI-tagging PCR,

(4.2) AcuI-tagging PCR to tag the desired signature of interest with the AcuI motif and insert the detection handle sequence, (4.3) digestion of the amplicon with AcuI, (4.4) isolation of the small digestion fragment and (4.5) DNA adaptor ligation to capture desired signatures. The DNA ligated product is subsequently analyzed by quantitative (5.1) and/or qualitative PCR (5.2).

4.1 Design of primers for the AcuI-tagging PCR

a. Selection of the dinucleotide of interest

The first objective is to select a dinucleotide signature of interest that must include the base (or bases) probed for modification (Fig. 1, step I in red). Notably, a selected genomic base can be included in four independent dinucleotides (Fig. 2D). Precisely, two dinucleotides can be designed using the upstream or downstream nucleotide of the targeted base from each DNA strand, as illustrated in Fig. 2D for the *DPY30* target. The selection of the dinucleotide provides substantial flexibility in the design of the AcuI-tagging primer. The generation of diverse dinucleotide signatures to capture the same genomic base is a striking advantage of DTECT (advantages are described in Section 6.2), which helps bypass certain limitations of the method (limitations are described in Section 6.1).

b. General design guidelines for AcuI-tagging primers

Timing: ~5 min

The AcuI-tagging primer has a total length of 60 nt and includes three essential parts. First, the 5′ end contains a 25 nt handle detection sequence (5′-GCAATTCCTCACGAGACCCGTCCTG-3′) to detect the captured material in step 5. Second, it includes 29 nt of the targeted nucleic acid sequence adjacent to the targeted genomic sequence. Third, an AcuI recognition motif (5′-CTGAAG-3′) is inserted strictly at 14 nt from the 3′ end (Fig. 1, step I in green). Consequently, the targeted sequence complementary to the nucleic acid sequence is interspaced by the AcuI motif in two independent parts of 15 nt and 14 nt, respectively (Fig. 1, step I in yellow).

c. Design of PCR reverse primers

Timing: ~5 min

In addition to the AcuI-tagging primer, a reverse primer is designed to amplify the targeted dinucleotide during the AcuI-tagging PCR (Fig. 1, step I). The reverse primer is a locus-specific primer located at a distance greater than 100 bp from the selected dinucleotide. The selection of a primer sequence at a distance > 100 bp is important for the rapid and efficient

separation of the two DNA fragments generated during AcuI digestion, as detailed in Section 4.4 and illustrated in Fig. 1, step IV.

For the design of the reverse primer, we recommend using "*Primer 3*" (https://bioinfo.ut.ee/primer3/) (Koressaar et al., 2018) with the following parameters: The primer length is selected between 25 and 30 nt, and the optimal melting temperature (Tm) should be between 57 and 63 °C. We also recommend selecting the "Mispriming library" with the species of interest to design a highly specific primer. For example, to amplify human genomic DNA, "HUMAN" should be selected at "Mispriming library =."

d. Primer design to capture the DPY30 and SAGE1 mutations introduced with prime editing

To illustrate this protocol, we introduce two transversion mutations into *DPY30* and *SAGE1* genes using prime editing (Fig. 3). We design AcuI-tagging primers and reverse primers to amplify the signatures and distinguish the edits from the reference signatures (Fig. 3).

Design of the AcuI-tagging oligonucleotide for the detection of the *DPY30* mutation (G > C):

First, we collect the 29 nt genomic sequence upstream of the nucleotide targeted for modification: "5′-CCTGCGTTTGTCCCTCCAGCATC TGCTCT-3.′" Next, we insert the AcuI motif sequence "5′-CTGAAG-3′" 14 nt upstream of the 3′ end from the targeted nucleotide "5′-CCTGCGTTTGTCCCT-*CTGAAG*-CCAGCATCTGCTCT-3′." Finally, we introduce the handle detection sequence (5′-GCAATT CCTCACGAGACCCGTCCTG-3′) at the 5′ end of the oligonucleotide: "5′-*GCAATTCCTCACGAGACCCGTCCTG*CCTGCGTTTGTCCC TCTGAAGCCAGCATCT
GCTCT-3′" (Oligonucleotide named OB212). Oligonucleotide sequences are available in Table 1.

Design of the AcuI-tagging oligonucleotide for the detection of the *SAGE1* mutation (C > G):

First, we collect the 29 nt genomic sequence that is upstream of the nucleotide targeted for modification: "5′-TTGTGAAGAGAGGATAAA TAACGGCCAAC-3′." Next, we insert the AcuI motif sequence "5′-CTGAAG-3′" 14 nt upstream of the 3′ end from the targeted nucleotide "5′-TTGTGAAGAGAGGAT-*CTGAAG*-AAATAACGGCCAAC-3′". Finally, we introduce the handle detection sequence (5′-GCAATTCC TCACGAGACCCGTCCTG-3′) at the 5′ end of the oligonucleotide: "5′-*GCAATTCCTCACGAGACCCGTCCTG*TTGTGAAGAGAGGA TCTGAAGAAATAACGGCCAAC-3′" (Oligonucleotide named OB210). Oligonucleotide sequences are available in Table 1.

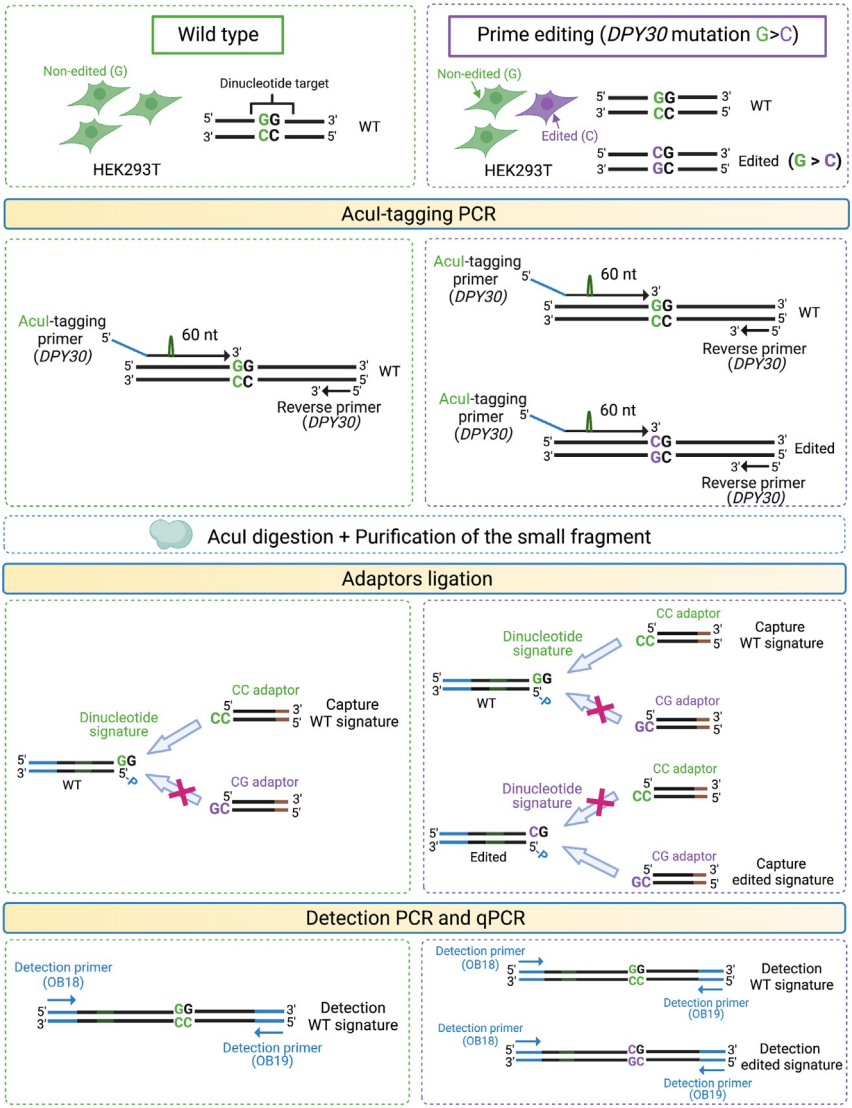

Fig. 3 DTECT capture of newly created signatures by CRISPR-mediated prime editing at the *DPY30* and *SAGE1* loci. Illustration of DTECT to capture the WT and edited signatures on the control (left) and edited samples (right) at the *DPY30* locus. An AcuI-tagging PCR is performed to amplify the signatures of interest in the respective samples. Adaptors complementary to the WT and edited signatures are utilized to capture signatures and are detected by PCR and qPCR using oligos OB18 and OB19.

The reverse primers are designed computationally using "*primer 3*," as described in 4.1.c (sequences available in Table 1). Reverse oligonucleotide sequences are named OB211 and OB213 for *SAGE1* and *DPY30*, respectively. Oligonucleotide sequences are available in Table 1.

4.2 Amplification of the targeted dinucleotide using the AcuI-tagging primer and PCR purification

Timing: ~60–90 min

The second step of the protocol consists of amplifying the locus of interest with the desired targeted dinucleotide by PCR using primers designed in Section 4.1. This protocol focuses on DTECT, so we refer the readers to excellent protocols for detailed experimental design for genome editing (Anzalone et al., 2019; Bak, Dever, & Porteus, 2018; Huang, Newby, & Liu, 2021; Ran et al., 2013; Santos, Kiskinis, Eggan, & Merkle, 2016). A critical negative control consists of the transfection of a non-targeting guide RNA to obtain an unedited control genomic sample.

a. Cells edited at the *SAGE1* and *DPY30* loci and unedited (control) cells are harvested by centrifugation at 300 × *g* for 3 min. Next, the medium is removed, and the cell pellet is washed with PBS. Then, cells are centrifuged at 300 × *g* for 3 min, and the PBS is removed.

b. The genomic DNA (gDNA) is extracted with the Quick Extract solution. First, the cell pellet is resuspended in 50 μL Quick Extract solution. Next, the solution is mixed by vortexing for 15 s and incubated at 65 °C for 6 min. Once the incubation is completed, the solution is mixed by vortexing for 15 s.

c. The tube is transferred to a heat block for 2 min at 98 °C.

d. H_2O is added to dilute the gDNA at approximatively 200 ng/μL as determined by nanodrop quantification. The gDNA preparation can be stored in 50–100 μL aliquots at −80 °C or immediately used for the AcuI-tagging PCR. If the genomic DNA samples are stored, the sample should be heated for 2 min at 98 °C before adding samples in the PCR reactions.

Note: The DTECT protocol can start here if DNA has been extracted with different approaches.

e. The AcuI-tagging PCR is conducted in a 25 μL reaction. First, a PCR master mix is prepared to amplify the loci of interest (*SAGE1* and *DPY30*) in the edited and unedited samples.

Master mix component	Final concentration	Volume (μL)—25 μL/reaction	Volume (μL)—Master mix for 2.5 reactions
H_2O	–	18.05	45.125
Q5 buffer, 5×	1×	5	12.5
dNTP, 10 mM	0.1 μM	0.2	0.5
Q5 polymerase, 2 units/μL	0.5 units	0.25	0.625
AcuI-tagging primer, 100 μM	1 μM	0.25	0.625
Reverse primer, 100 μM	1 μM	0.25	0.625
gDNA, 200 ng/μL	–	1	–

f. One AcuI-tagging PCR is performed per edited and unedited genomic DNA samples. 24 μL of the master mix is loaded in two PCR tubes, and 1 μL (~200 ng/μL) of the respective genomic DNA is added. The AcuI-tagging of the *SAGE1* targeted dinucleotide is conducted with primers OB210 and OB211. The AcuI-tagging of the *DPY30* targeted dinucleotide is conducted with primers OB212 and OB213.

g. The PCR machine is set up with the following conditions.

Temperature	Time	PCR phase	Cycle number
95 °C	30 s	Denaturation	1
95 °C	10 s	Denaturation	
58 °C	10 s	Annealing	40
72 °C	45 s	Extension	
72 °C	60 s	Extension	1

h. A 2% (wt/vol) agarose gel is prepared by combining 1 g of agarose in 50 mL 1× TAE. The mixture is melted by heating until agarose has dissolved (approximatively 1–2 min). The solution is swirled carefully to ensure complete solubilization of the agarose and prevent excess evaporation of the buffer.

Note: Appropriate gloves must be worn to protect from burns.

i. The agarose is carefully poured into a gel cassette, bubbles are removed with a pipette tip, and a gel comb is added. The gel is left to cool down at room temperature.

j. Once the agarose gel has solidified, the molecular weight marker is prepared by mixing 5 μL DNA ladder, 1 μL loading dye, 0.5 μL diluted SYBR gold. Then, the samples are prepared by adding 1 μL diluted SYBR gold and 5 μL loading dye to the PCR tubes.

k. The AcuI-tagging PCRs are loaded on the gel to verify the successful amplification of the loci and extract the amplicons.

Expected result: A successful PCR should be a single band at the size of the expected amplicon. In the *DPY30* and *SAGE1* examples, the amplicons have an expected length of 582 and 646 bp, respectively.

l. The four AcuI-tagged amplicons are excised from the gel with a clean razor blade and purified using a Zymoclean Gel DNA Recovery Kit. To avoid contamination between samples, the razor blade is washed with ethanol and H_2O between samples. Alternatively, the AcuI-tagging PCR can be purified on a column.

m. The purified AcuI tagged amplicons are quantified using a nanodrop and diluted at ~20 ng/μL in the elution buffer.

4.3 Digestion of the AcuI-tagged genomic amplicon with AcuI

Timing: 80 min

The third step of the DTECT protocol requires digestion of the AcuI-tagged amplicons with AcuI (Fig. 3). AcuI cleaves at a fixed and predictable position 14/16 bp downstream of its recognition motif, generating a 3′ dinucleotide signature and leaving 5′ phosphates (Fig. 3).

a. The AcuI digestion master mix is prepared as indicated below.

Master mix component	Final concentration	Volume (μL)—20 μL/reaction	Volume (μL)—Master mix for 4.5 reactions
H_2O	–	15.75	70.875
CutSmart Buffer, 10×	1×	2	9
AcuI, 5000 units/mL	1.25 units	0.25	1.125
AcuI-tagged PCR, 20 ng/μL	40 ng	2	–

b. 18 μL of the master mix is loaded into tubes, and 2 μL of the purified AcuI-tagged amplicon (40 ng/μL) is added.
c. The reactions are incubated at 37 °C for 60 min and heated at 65 °C for 20 min to inactivate AcuI.

Pause point: The AcuI-digested amplicons can be stored at −20 °C or immediately used for the isolation step (step 4.4).

4.4 Isolation of the small AcuI-digested fragments by solid-phase reversible immobilization (SPRI)

Timing: ~10 min

DTECT requires the isolation of the small fragments (60 bp) to facilitate the ligation of DNA adaptors. The isolation of the small fragment is achieved by the precipitation of the large fragment by SPRI using magnetic beads. The stock of beads must be resuspended by mixing the stock bottle. Once well resuspended, a small aliquot (~800 μL) is placed in a 1.5 mL microcentrifuge tube and stored at 4 °C. The aliquot of beads should be placed at room temperature approximately 1 h before step 4.4.

a. 10 μL of the digestion reactions are transferred in new microcentrifuge tubes, and 18 μL of AMPure XP magnetic beads are added (volume ratio of DNA:beads = 1:1.8). The solutions are mixed by pipetting up and down 10 times and are incubated at room temperature for 5 min.
b. After incubation, the tubes are placed on a magnetic rack until the beads are pulled to the side of the tube by the magnetic field, which takes approximately 3–5 min. The small fragment is solubilized in the supernatant, and the large fragment is bound to the magnetic beads.
c. 20 μL of the supernatant is transferred into a new 1.5 mL microcentrifuge tube, and 40 μL H_2O is added to dilute the DNA.

Pause point: The isolated fragments can be stored at −20 °C or immediately used for the ligation.

4.5 The capture of dinucleotide signatures

Timing: 70 min

The final step of the capture is the ligation of DNA adaptors that are complementary or non-complementary to the reference signatures and edited signatures (Fig. 3). In the examples presented here, the selected dinucleotide signature is 5′-GG-3′ for *DPY30* and 5′-CA-3′ for *SAGE1*; and the expected edited signatures is 5′-CG-3′ for *DPY30* and 5′-GA-3′ for *SAGE1*. For simplification, only the *DPY30* target is illustrated in Fig. 3.

a. Adaptors are selected for their complementarity to the reference and edited dinucleotide signatures. We, therefore, use the complementary adaptors 5′-CC-3′ and 5′-TG-3′ to capture the reference dinucleotide signatures; and the complementary adaptors 5′-CG-3′ and 5′-TC-3′ for the mutations inserted with prime editing into *DPY30* and *SAGE1*, respectively.

b. Independent ligation master mixes for each adaptor are prepared.

Master mix component	Final concentration	Volume (μL)—10 μL/reaction	Volume (μL)—Master mix for 2.5 reactions
H_2O	–	6.5	16.25
T4 ligase buffer, 5×	1×	2	5
T4 ligase, 1 unit	0.5 units/μL	0.5	1.25
Adaptor, 20×	1×	0.5	1.25
Isolated material (Step IV)	–	0.5	–

c. 9.5 μL of the master mix is loaded into PCR tubes, and 0.5 μL of the isolated material is added to the ligation reactions.

d. The reaction is incubated at 25 °C for 60 min followed by 65 °C for 10 min to heat inactivate the DNA ligase.

Pause point: The ligation tubes can be stored at −20 °C or immediately used to analyze the captured material (steps 5.1 or/and 5.2).

5. Capture detection and quantification

The ligation reactions can either be analyzed by analytical or quantitative PCR (Fig. 3). While analytical PCRs determine the presence of a particular signature, quantitative PCRs determine the relative quantity of the respective signatures in the nucleic acid samples.

5.1 Analytical detection

Timing: 45 min

The analytical PCR detects the ligation product (Fig. 4A–B). The detection of the ligated product reveals the presence of the particular signature in the original nucleic acid sample.

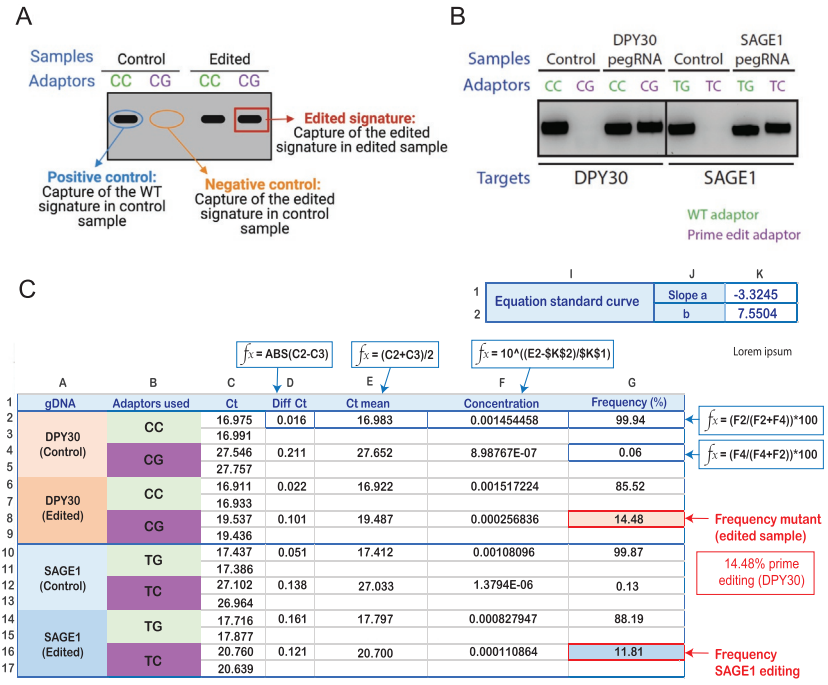

Fig. 4 Example of analytical and quantitative analysis of *DPY30* and *SAGE1* edited sites with prime editing. (A) Illustration of expected results from an analytical gel. WT signature (positive control) is specifically detected in the control sample but not edited signature (negative control). The detection of a captured signature in the edited sample indicates successful editing. (B) Agarose gel of DTECT for the capture of prime editing at the *DPY30* and *SAGE1* loci. (C) Mathematical formulas and quantifications of the newly created signatures with prime editing using qPCR.

a. A PCR master mix is prepared to amplify the captured material

Master mix component	Final concentration	Volume (µL)—12.5 µL/reaction	Volume (µL)—Master mix for 4.5 reactions
H_2O	–	9.075	40.84
Q5 buffer, 5×	1×	2.5	11.25
dNTP, 10 mM	0.05 mM	0.2	0.9
OB18, 100 µM	0.5 µM	0.0625	0.281
OB19, 100 µM	0.5 µM	0.0625	0.281
Q5 polymerase, 2 units/µL	1 unit	0.1	0.45
Ligation products (Step V)	–	0.5	–

b. 12 μL of the PCR master mix is loaded into PCR tubes, and 0.5 μL of the respective ligations is added.

c. The analytical PCR is conducted under the following conditions:

Temperature (°C)	Time (s)	PCR phase	Cycle number
95	60	Denaturation	1
95	10	Denaturation	
65	5	Annealing	15–25 (see below)
72	7	Extension	
72	60	Extension	

The number of PCR cycles is adjusted based on the expected efficiency. For example, a low abundance of genomic variants (≤1% frequency) is detected with 23–25 PCR cycles. On the other hand, the detection of greater amounts of edited products can be achieved between 15 and 22 PCR cycles. For instance, given that we did not know the expected editing level of *DPY30* and *SAGE1*, we conducted 22 PCR cycles (Fig. 4B).

d. Before starting the PCR, a 2% agarose gel is prepared, as described in 4.2f–i.

e. 0.5 μL of diluted SYBR gold and 2.5 μL loading dye are added to the reaction, and the PCR and DNA ladder is loaded on the agarose gel. Then, the gel is run in 1 × TAE until the DNA has resolved.

f. Gels are developed using a gel imaging system (e.g., ChemiDoc or LI-COR Odyssey).

Expected result: If the experiment is successful, it is expected that the capture of the reference signature in the non-edited sample (positive control) generates an unambiguous amplicon that migrates at 120 bp (Fig. 4A, in blue control sample). As expected, this amplicon corresponds to the detection handle, the genomic DNA sequence interspaced with the AcuI motif, the dinucleotide signature, and the adaptor (Billon et al., 2020). In addition, the capture using the variant adaptor in the non-edited sample (negative control) generates no amplification (or background) (Fig. 4A, in orange control sample). A signal in the edited sample captured with the adaptor corresponding to the variant confirms the successful incorporation of the mutation (Fig. 4A, edited signature in red edited sample). The development of the analytical gels for the *SAGE1* and *DPY30* prime editing experiment is shown in Fig. 4B. As expected, a strong capture was observed in the positive control and no

detection in the negative control. However, by capturing the edited signature in the samples that were edited with prime editing, the newly created signature was unambiguously detected with DTECT (Fig. 4B, CG adaptor for *DPY30* and TC adaptor for *SAGE1* in pegRNA samples).

5.2 Quantitative detection

Timing: 90 min

Quantitative detection enables the determination of the relative amount of captured material by the respective adaptors (Fig. 4C).

a. The amount of captured material is quantified by qPCR in technical duplicates.

b. A qPCR master mix is prepared as indicated below.

Master mix component	Final concentration	Volume (μL)—10 μL/reaction	Volume (μL)—Master mix for 18 reactions
H_2O	–	3.8	68.4
SYBR green, 2×	1×	5	90
OB18, 100 μM	1 μM	0.1	1.8
OB19, 100 μM	1 μM	0.1	1.8
Ligated products	–	1	–

c. 9 μL of the master mix is loaded into a qPCR plate, and 1 μL of the respective ligation reactions is added in duplicates.

d. The qPCR plate is sealed with a transparent adhesive film to prevent sample evaporation during PCR heating and loss during transportation. The plate is centrifuged to ensure the samples are at the bottom of the plate.

e. The qPCR machine is set with the following program.

Temperature (°C)	Time (s)	PCR phase	Number of cycles
50	120	Hold stage	1
95	600		
95	10	PCR stage	40
60	30		
95	15	Melting	1
60	60		
95	15		

5.2.1 Quantification and statistical analysis

Downstream analysis of the experiment is shown in Fig. 4C. The standard curve equation (determined in Section 3.8) and experimental Ct scores (Fig. 4C, column C) determine the relative frequency of each genomic signature captured by the respective adaptors (Fig. 4C, column G).

First, the difference in Ct between technical duplicates is calculated (Fig. 4C, column D). Usually, when the difference between technical duplicates is >0.5 Ct, the qPCR samples are repeated. Then, the average cycle threshold (Ct mean) is calculated from the technical duplicates of each sample (Fig. 4C, column E) using the formula "=(C2 + C3)/2." Next, the standard curve generated in step 3.7 is used to determine the relative quantity of each signature captured using the formula: "Concentration = 10 ^ ((Mean Ct − 7.5504)/−3.3245)" (Fig. 4C, column F).

The relative abundance between the wild type (WT) and the variant is determined as follows: "Frequency Mutant = 10 ^ (Concentration Mutant/(Concentration Mutant + Concentration WT)) x 100" and "Frequency WT = 10 ^ (Concentration WT/(Concentration Mutant + Concentration WT)) x 100" (Fig. 4C, column G).

Expected result: If successful, the relative abundance associated with the negative control should be close to 0%, representing the background capture. Here, the DPY30 background capture is at 0.06% (Fig. 4C, G4) and the SAGE1 capture background is at 0.13% (Fig. 4C, G12). Interestingly, the prime editing frequency at these loci is 14.48% (Fig. 4C, G8) and 11.81% (Fig. 4C, G16) for *DPY30* and *SAGE1*, respectively.

6. Limitations and advantages of DTECT

6.1 Limitations of DTECT

DTECT has three limitations caused by the sequences surrounding the targeted dinucleotide, the specificity of AcuI, and the error-prone repair of lesions introduced by genome editing technologies.

The first limitation occurs when one or multiple AcuI recognition motifs (5′-CTGAAG-3′) are present in the targeted nucleic acid sequence near the dinucleotide of interest. If genomic AcuI motif(s) are included in the AcuI-tagged amplicon, the AcuI digestion generates additional undesired signature(s), thereby potentially affecting targeted capture. However, given that AcuI-tagging primers can be designed from both strands (Fig. 2D), only when two AcuI motifs are positioned on both sides of the targeted dinucleotide DTECT is affected. Consequently, genomic AcuI motifs present near

the target site are problematic only if two AcuI motifs are close (<30 nt) to the targeted dinucleotide both upstream and downstream. Nevertheless, a 6-nucleotide motif is approximately present only every 4 kb, limiting the probability of two interfering AcuI motifs.

Although type IIS enzymes cleave far from their recognition motif, their helpful activity can come at the expense of their precision. The slippage is an unspecific activity that occurs when cleavage is shifted by one nucleotide. Slippage activity of certain type IIS enzymes is more frequent than the canonical cleavage (Lundin et al., 2015). However, AcuI is not prone to frequent slippage, estimated at only 1.1% (Lundin et al., 2015), which explains the high precision of DTECT. Moreover, the dinucleotide signatures induced by AcuI slippage are predictable based on the sequence flanking the targeted dinucleotide. Therefore, the dinucleotide induced by the slippage can be avoided by the appropriate selection of the adaptors (Billon et al., 2020).

Finally, DTECT can only capture the identity of a two-nucleotide window. Certain byproducts of the repair of lesions introduced by genome editing technologies can occur in a larger window. For instance, DSB repair by end-joining leads to the generation of predictable insertion-deletion (indel) mutations. Consequently, the indels generated near the targeted dinucleotide signature are susceptible to disrupt the proximal sequence of the targeted signature, thereby inducing erroneous AcuI-tagging PCR amplification. This limitation results in an overestimation of the frequency of precision genome editing by DTECT, as we revealed previously (Billon et al., 2020). However, we also demonstrated that it is possible to position the targeted dinucleotide at a location where no indels are induced (generally ~10 bp from the DSB). However, indels are rare byproducts of CRISPR-mediated base editing, prime editing, and naturally occurring genetic variants, suggesting that DTECT is less affected in these contexts.

6.2 Strengths and advantages of DTECT

DTECT is a simple detection method that relies on standard molecular biology techniques (PCR, digestion, and ligation), off-the-shelf reagents, and enzymes (AcuI and T4 ligase) and requires minimal equipment (thermocycler and qPCR). In addition, DTECT has a short turnaround time (~4–5 h) with limited hands-on time.

Notably, one of the most significant strengths of DTECT is its versatility and flexibility. A unique library of 16 standard adaptors is sufficient to identify all dinucleotide signature types (Fig. 2A), and each genomic base can be

detected using 4 distinct signatures (Fig. 2D). Importantly, DTECT is sensitive as it can distinguish signatures present at low frequency (<1%) (Fig. 4C) in complex samples, including precious patient specimens (Billon et al., 2020). The high sensitivity of DTECT is mediated by the complementarity between the signature and the specific adaptor. In addition, the ligation creates a stable ligation product by covalently linking the adaptors and the signature. Finally, the detection is compatible with analytical and quantitative detection modalities, which provide complementary advantages. Indeed, it has the dual benefit of quantifying the relative frequency of signatures for precision genome editing experiments (Fig. 4C) and detecting the presence of a specific signature for the rapid genotyping of cellular and animal models, like we previously demonstrated (Billon et al., 2020).

Strikingly, DTECT has limited technical variabilities across distinct experimental conditions, samples, and mutation types and is conducted in a highly controlled experimental setup. Indeed, a unique pair of detection primers are used in all experiments to quantify the amount of captured material, regardless of the mutation or sample type. This is a significant advantage because a unique standard curve is sufficient (see part 3.7) to quantify the capture of all signatures. In contrast, methods using mutation-specific primers or probes require separate evaluation of the primer efficiencies and do not provide positive controls to ensure experiment quality. This is a problem because, unlike with DTECT, primers/probes-specific mutations have variable and unpredictable efficiencies due to the competition between reference and variant alleles. Another remarkable advantage of DTECT is the inherent controls to determine the efficiency and specificity of signature capture, as shown in Fig. 4 for the prime editing experiments. For instance, negative controls provide the experimental background as measured by capturing the variant signature in the control sample (Fig. 4C). Positive controls ensure that DTECT assays have been successfully conducted by capturing the WT signature in reference nucleic acid samples (Fig. 4C). Positive controls are, therefore, crucial to help new users to implement DTECT. The ease of implementing DTECT and the presence of controls are substantial advantages for its adoption.

DTECT is a robust assay with several advantages over alternative detection methods for the rapid and accurate identification of nucleic acid sequences. We previously compared the performance of DTECT to sequencing technologies to detect precision genome editing and genetic variants (Billon et al., 2020). Strikingly, DTECT shows comparable

performance to Next-Generation Sequencing (NGS) for its precision to detect desired signatures (Billon et al., 2020). However, NGS has a larger window of detection and provides highly precise quantification of the mixture of modified alleles. Unlike DTECT, which takes a few hours to perform, sequencing technologies are time-consuming (several days to weeks) and expensive. DTECT and NGS are complementary as DTECT can validate the presence of genetic signatures before a deeper analysis of low-frequency mutations/editing byproducts by NGS. We also compared DTECT with Sanger sequencing-based detection methods (Billon et al., 2020). We observed that a significant fraction of reactions does not generate the high-quality profiles required for accurate quantification, making DTECT more reliable than Sanger sequencing.

DTECT can be applied to identify changes introduced by modern precision genome editing technologies and natural variants in virtually any biological system. DTECT requires a small amount of nucleic acid as starting material, only sufficient to conduct the AcuI-tagging PCR. This is advantageous for clinical applications in which samples are limited. For example, we previously demonstrated the utility of DTECT to identify oncogenic signatures in the bone marrow specimens from pediatric cancer patients (Billon et al., 2020). Consequently, DTECT will be helpful for the identification of pathogenic variants in cellular and animal models, human biopsies, and fluid sampling.

6.3 Optimizations and troubleshooting

With the presence of positive and negative controls and the use of commercial reagents, no optimizations are required to implement DTECT. Nevertheless, in rare cases (<5%), we observed that the AcuI-tagging PCR is inefficient at hard-to-amplify loci. One solution is to amplify the targeted genomic locus using standard primers and conduct the AcuI-tagging PCR using the locus amplicon as template DNA.

7. Concluding remarks

This step-by-step protocol describes how to identify nucleic acid signatures and associated genetic variation introduced by modern precision genome editing technologies, including CRISPR-mediated base editing and prime editing or naturally occurring mutations. The DTECT protocol is easy to conduct. It only requires oligo design, PCR amplification, restriction

digestion, DNA ligation, and analytical/quantitative PCR. Furthermore, DTECT only requires commercially available reagents, is rapid to execute (~4–5 h), providing same-day quantitative determination of editing frequency. The development of novel detection methods for genetic signatures will improve the characterization of variations of uncertain significance and accelerate clinical diagnostics.

References

Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393.

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., et al. (2013). Signatures of mutational processes in human cancer. *Nature*, *500*(7463), 415–421. https://doi.org/10.1038/nature12477.

Anzalone, A. V., Koblan, L. W., & Liu, D. R. (2020). Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology*, *38*(7), 824–844. https://doi.org/10.1038/s41587-020-0561-9.

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576*(7785), 149–157. https://doi.org/10.1038/s41586-019-1711-4.

Bak, R. O., Dever, D. P., & Porteus, M. H. (2018). CRISPR/Cas9 genome editing in human hematopoietic stem cells. *Nature Protocols*, *13*(2), 358–376. https://doi.org/10.1038/nprot.2017.143.

Berkovich, E., Monnat, R. J., Jr., & Kastan, M. B. (2007). Roles of ATM and NBS1 in chromatin structure modulation and DNA double-strand break repair. *Nature Cell Biology*, *9*(6), 683–690. https://doi.org/10.1038/ncb1599.

Billon, P., Bryant, E. E., Joseph, S. A., Nambiar, T. S., Hayward, S. B., Rothstein, R., et al. (2017). CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons. *Molecular Cell*, *67*(6), e1064. https://doi.org/10.1016/j.molcel.2017.08.008. 1068-1079.

Billon, P., Nambiar, T. S., Hayward, S. B., Zafra, M. P., Schatoff, E. M., Oshima, K., et al. (2020). Detection of marker-free precision genome editing and genetic variation through the capture of genomic signatures. *Cell Reports*, *30*(10), e3286. https://doi.org/10.1016/j.celrep.2020.02.068. 3280–3295.

Brinkman, E. K., Chen, T., Amendola, M., & van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research*, *42*(22), e168. https://doi.org/10.1093/nar/gku936.

Chailleux, C., Aymard, F., Caron, P., Daburon, V., Courilleau, C., Canitrot, Y., et al. (2014). Quantifying DNA double-strand breaks induced by site-specific endonucleases in living cells by ligation-mediated purification. *Nature Protocols*, *9*(3), 517–528. https://doi.org/10.1038/nprot.2014.031.

Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., et al. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature Biotechnology*, *37*(3), 224–226. https://doi.org/10.1038/s41587-019-0032-3.

Djureinovic, D., Hallstrom, B. M., Horie, M., Mattsson, J. S. M., La Fleur, L., Fagerberg, L., et al. (2016). Profiling cancer testis antigens in non-small-cell lung cancer. *JCI Insight*, *1*(10), e86837. https://doi.org/10.1172/jci.insight.86837.

Drmanac, R., Sparks, A. B., Callow, M. J., Halpern, A. L., Burns, N. L., Kermani, B. G., et al. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science*, *327*(5961), 78–81. https://doi.org/10.1126/science.1181498.

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews. Genetics*, *17*(7), 422–433. https://doi.org/10.1038/nrg.2016.58.

Findlay, S. D., Vincent, K. M., Berman, J. R., & Postovit, L. M. (2016). A digital PCR-based method for efficient and highly specific screening of genome edited cells. *PLoS One*, *11*(4), e0153901. https://doi.org/10.1371/journal.pone.0153901.

Germini, D., Tsfasman, T., Zakharova, V. V., Sjakste, N., Lipinski, M., & Vassetzky, Y. (2018). A comparison of techniques to evaluate the effectiveness of genome editing. *Trends in Biotechnology*, *36*(2), 147–159. https://doi.org/10.1016/j.tibtech.2017.10.008.

Green, M. R., & Sambrook, J. (2019). Ligation and ligases. *Cold Spring Harbor Protocols*, *2019*-(8). https://doi.org/10.1101/pdb.top101352.

Guell, M., Yang, L., & Church, G. M. (2014). Genome editing assessment using CRISPR Genome Analyzer (CRISPR-GA). *Bioinformatics*, *30*(20), 2968–2970. https://doi.org/10.1093/bioinformatics/btu427.

Helleday, T., Eshtad, S., & Nik-Zainal, S. (2014). Mechanisms underlying mutational signatures in human cancers. *Nature Reviews. Genetics*, *15*(9), 585–598. https://doi.org/10.1038/nrg3729.

Huang, T. P., Newby, G. A., & Liu, D. R. (2021). Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nature Protocols*, *16*(2), 1089–1128. https://doi.org/10.1038/s41596-020-00450-9.

The International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature*, *437*(7063), 1299–1320. https://doi.org/10.1038/nature04226.

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), 434–443. https://doi.org/10.1038/s41586-020-2308-7.

Khan, S., Ullah, M. W., Siddique, R., Nabi, G., Manan, S., Yousaf, M., et al. (2016). Role of recombinant DNA technology to improve life. *International Journal of Genomics*, *2016*, 2405954. https://doi.org/10.1155/2016/2405954.

Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., & Sternberg, S. H. (2019). Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature*, *571*(7764), 219–225. https://doi.org/10.1038/s41586-019-1323-z.

Koressaar, T., Lepamets, M., Kaplinski, L., Raime, K., Andreson, R., & Remm, M. (2018). Primer3_masker: Integrating masking of template sequence with primer design software. *Bioinformatics*, *34*(11), 1937–1938. https://doi.org/10.1093/bioinformatics/bty036.

Labrie, S. J., Samson, J. E., & Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nature Reviews. Microbiology*, *8*(5), 317–327. https://doi.org/10.1038/nrmicro2315.

Landrum, M. J., Chitipiralla, S., Brown, G. R., Chen, C., Gu, B., Hart, J., et al. (2020). ClinVar: Improvements to accessing data. *Nucleic Acids Research*, *48*(D1), D835–D844. https://doi.org/10.1093/nar/gkz972.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, *536*(7616), 285–291. https://doi.org/10.1038/nature19057.

Lindsay, H., Burger, A., Biyong, B., Felker, A., Hess, C., Zaugg, J., et al. (2016). CrispRVariants charts the mutation spectrum of genome engineering experiments. *Nature Biotechnology*, *34*(7), 701–702. https://doi.org/10.1038/nbt.3628.

Lundin, S., Jemt, A., Terje-Hegge, F., Foam, N., Pettersson, E., Kaller, M., et al. (2015). Endonuclease specificity and sequence dependence of type IIS restriction enzymes. *PLoS One*, *10*(1), e0117059. https://doi.org/10.1371/journal.pone.0117059.

Miga, K. H., Koren, S., Rhie, A., Vollger, M. R., Gershman, A., Bzikadze, A., et al. (2020). Telomere-to-telomere assembly of a complete human X chromosome. *Nature*, *585*(7823), 79–84. https://doi.org/10.1038/s41586-020-2547-7.

Nambiar, T. S., Billon, P., Diedenhofen, G., Hayward, S. B., Taglialatela, A., Cai, K., et al. (2019). Stimulation of CRISPR-mediated homology-directed repair by an engineered RAD18 variant. *Nature Communications*, *10*(1), 3395. https://doi.org/10.1038/s41467-019-11105-z.

Pascal, J. M. (2008). DNA and RNA ligases: Structural variations and shared mechanisms. *Current Opinion in Structural Biology*, *18*(1), 96–105. https://doi.org/10.1016/j.sbi.2007.12.008.

Pickar-Oliver, A., & Gersbach, C. A. (2019). The next generation of CRISPR-Cas technologies and applications. *Nature Reviews. Molecular Cell Biology*, *20*(8), 490–507. https://doi.org/10.1038/s41580-019-0131-5.

Pingoud, A., Wilson, G. G., & Wende, W. (2014). Type II restriction endonucleases- -a historical perspective and more. *Nucleic Acids Research*, *42*(12), 7489–7527. https://doi.org/10.1093/nar/gku447.

Qiu, P., Shandilya, H., D'Alessio, J. M., O'Connor, K., Durocher, J., & Gerard, G. F. (2004). Mutation detection using surveyor nuclease. *BioTechniques*, *36*(4), 702–707. https://doi.org/10.2144/04364PF01.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, *8*(11), 2281–2308. https://doi.org/10.1038/nprot.2013.143.

Rouet, P., Smih, F., & Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Molecular and Cellular Biology*, *14*(12), 8096–8106. https://doi.org/10.1128/mcb.14.12.8096-8106.1994.

Santos, D. P., Kiskinis, E., Eggan, K., & Merkle, F. T. (2016). Comprehensive protocols for CRISPR/Cas9-based gene editing in human pluripotent stem cells. *Current Protocols in Stem Cell Biology*, *38*, 5B 6 1–5B 6 60. https://doi.org/10.1002/cpsc.15.

Shanbhag, N. M., Rafalska-Metcalf, I. U., Balane-Bolivar, C., Janicki, S. M., & Greenberg, R. A. (2010). ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks. *Cell*, *141*(6), 970–981. https://doi.org/10.1016/j.cell.2010.04.038.

Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M., et al. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, *309*(5741), 1728–1732. https://doi.org/10.1126/science.1117389.

Xue, H., Yao, T., Cao, M., Zhu, G., Li, Y., Yuan, G., et al. (2019). Structural basis of nucleosome recognition and modification by MLL methyltransferases. *Nature*, *573*(7774), 445–449. https://doi.org/10.1038/s41586-019-1528-1.

Yeh, C. D., Richardson, C. D., & Corn, J. E. (2019). Advances in genome editing through control of DNA repair pathways. *Nature Cell Biology*, *21*(12), 1468–1478. https://doi.org/10.1038/s41556-019-0425-z.

Zou, X., Koh, G. C. C., Nanda, A. S., Degasperi, A., Urgo, K., Roumeliotis, T. I., et al. (2021). A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nature Cancer*, *2*(6), 643–657. https://doi.org/10.1038/s43018-021-00200-0.