

Ride the Tide:
observing CRISPR/Cas9 genome editing by the numbers

Eva Karina Brinkman

This research was conducted under the auspices of the Oncology Graduate School
at the Netherlands Cancer Institute, Amsterdam

Printed by: Gildeprint - the Netherlands

Cover designed by: Stan Brinkman



**Ride the Tide:
observing CRISPR/Cas9 genome editing by the numbers**

Ga met de stroom mee: beschouw CRISPR/Cas9 genoombewerking op basis van getallen

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. R.C.M.E. Engels

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 17 april om 11.30 uur

door

Eva Karina Brinkman
geboren te Vlaardingen

Promotiecommissie

Promotoren: Prof.dr. B. van Steensel
Prof.dr. T.K. Sixma

Overige leden: Prof.dr. F. G. Grosveld
Prof.dr. R. Kanaar
Prof.dr. H. van Attikum

TABLE OF CONTENT

Chapter 1	Introduction	6
Chapter 2	Easy Quantitative Assessment of Genome Editing by Sequence Trace Decomposition	40
Chapter 3	Easy Quantification of Template-Directed CRISPR/Cas9 Editing	58
Chapter 4	Rapid Quantitative Evaluation of CRISPR Genome Editing by TIDE and TIDER	84
Chapter 5	Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks	100
Chapter 6	Genome-Wide Monitoring of Chromatin Effects on Cas9-Induced Double-Strand Break Repair	142
Chapter 7	General Discussion	172
Addendum	Summary	192
	Nederlandse Samenvatting	195
	Abbreviations	298
	Curriculum Vitae	200
	List of Publications	201
	PhD Portfolio	202
	Acknowledgement	204

Chapter 1

INTRODUCTION

ABSTRACT

Targeted genome editing has become a powerful genetic tool for modification of DNA sequences in their natural chromosomal context. CRISPR RNA-guided nucleases have recently emerged as an efficient targeted editing tool for multiple organisms. Hereby a double strand break is introduced at a targeted DNA site. During DNA repair genomic alterations are introduced which can change the function of the DNA code. However, our understanding of how CRISPR works is incomplete and it is still hard to predict the CRISPR activity at the precise target sites. The highly ordered structure of the eukaryotic genome may play a role in this. The organization of the genome is controlled by dynamic changes of DNA methylation, histone modification, histone variant incorporation and nucleosome remodelling. The influence of nuclear organization and chromatin structure on transcription is reasonably well known, but we are just beginning to understand its effect on genome editing by CRISPR.

PART 1: General Introduction

GENOME EDITING

Genome editing technologies make it possible to make precise changes in a DNA sequence, regardless of cell type or organisms. This gives an almost unlimited number of potential applications in the field of life sciences, for example to design model organisms with specific genotypes or to develop gene therapy strategies for use in health care or to improve crops and livestock for agriculture. A particular active area of genetic editing is that of patient-derived stem cells to create models for diseases including polycystic kidney disease (PKD) (1) or long QT syndrome (2). In the latter, patient-derived pluripotent stem cells were isolated to create isogenic cell lines. These cells can be differentiated to any cell type of interest to study or correct the disease. With these cell lines, it is possible to investigate the effect of gene mutations to a disease phenotype. Although genome editing strategies for disease therapies or plant breeding are making great progress, many hurdles still need to be overcome. Also, legislation and social acceptance are under active debate.

For the effective modification or regulation of genomic information a molecular machine is required with a DNA binding domain linked to an effector domain. The DNA binding domain is designed to bind specifically to a DNA sequence of a target gene. Several approaches for genome editing have been developed using targeted nucleases. The nuclease is directed to specific sequences in the genome where DNA modification is desired and introduces a DNA double-stranded break (DSB). Subsequently, the break activates the endogenous repair machinery of the host to restore the genome. In this process errors can be introduced that modify the targeted sequence (3-5).

Established targeted nucleases are meganucleases, zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR) with CRISPR associated nucleases (Cas9). Meganucleases are generated by engineering existing restriction enzymes, typically enzymes with a long DNA recognition sequence (e.g., 14-40 bp) (3, 6). The large recognition site provides the specificity to the target site to occur only once in the genome (FIGURE 1A). ZFNs and TALENs are recombinant proteins constructed of a customized DNA binding domain fused to the nuclease domain of the FokI restriction enzyme (FIGURE 1B-C). The DNA binding domain consists of a series of repeats that are only variable at a few residues. Each repeat region has specificity for a specific DNA motif. The various repeat moieties can be connected to each other into an array that binds at a dedicated DNA sequence (7-11). These platforms have made it possible to make significant progress, but each has its own

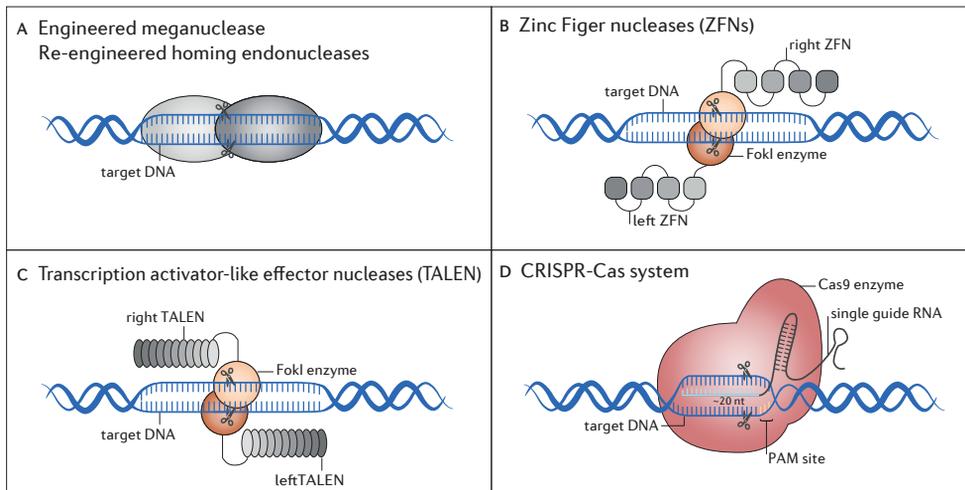


Figure 1: Targeted nucleases. Adapted from Thermo Fisher Scientific Inc. Schematics summarizing various approaches of genome engineering. (A) The meganuclease bound to its DNA target. The catalytic domain is shown in grey, which determines DNA sequence specificity and contains nuclease activity. (B) Zinc-finger nucleases (ZFNs) recognize DNA using three base pair recognition motifs. Three-four motifs are fused in tandem recognizing adjacent sequences to give unique specificity to a particular genomic locus. The motifs are linked to FokI nuclease that digests the DNA as a dimer. (C) Transcription activator-like effector nucleases (TALENs) recognize DNA through modules that include repeat-variable di-residues. As with ZFNs, two TALENs are used that cut DNA using the FokI nuclease dimer. (D) CRISPR/Cas9 system recognizes specific DNA using a guide RNA that brings Cas9 to its complementary DNA around a protospacer adjacent motif. Two domains of Cas9 are responsible for DNA cleavage on either stand of double strand DNA: the HNH domain cleaves the complementary DNA strand, whereas the RuvC-like domain cleaves the non-complementary DNA strand.

drawbacks. The generation of a meganuclease/ZFN/TALEN is a demanding and/or time-consuming process, making these approaches less suitable for multiplexing many targeted nucleases in a single cell.

More recently, a platform based on the bacterial CRISPR/Cas9 nuclease has been widely adopted by the scientific community for genome editing, largely because of the ease with which the target specificity can be generated (FIGURE 1D). This enables the performance of large scale, high throughput studies. In contrast to most known DNA-binding proteins, Cas9 is an RNA-guided nuclease and is targeted to a specific location in the DNA where its guide RNA base pairs with complementary DNA. Cas9 can be reprogrammed to target new sites by changing the sequence composition of the guide RNA. To serve as a genome editing tool, the natural endonuclease activity of Cas9 has been codon optimized for sequence-specific editing of the DNA in a wide range of organisms, including bacteria (12), fungi (13), plants (14) and animals (15-20).

The success of targeted nuclease genome editing tools is dependent on two processes. First, the specificity and efficiency of the generation of a DSB at a desired location in the genome by the targeted nuclease. Second, the efficacy and fidelity of the endogenous DNA

repair mechanism in the cell. Here, we will outline the influence of nuclear organization and chromatin structure on both these processes necessary for CRISPR/Cas9 genome editing.

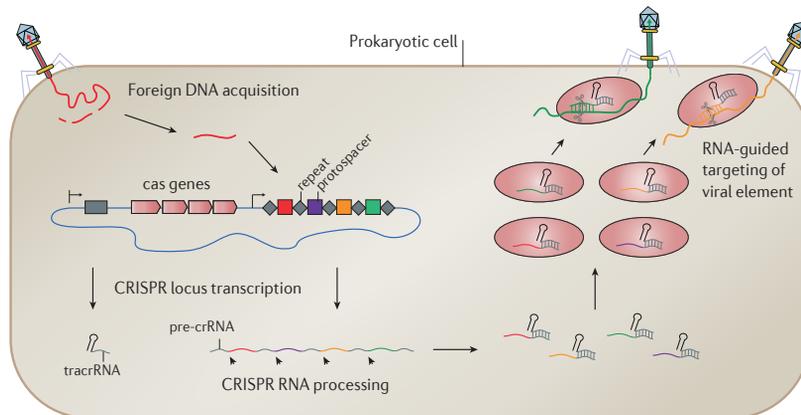


Figure 2: CRISPR/Cas9 bacterial immune system. Adapted from the Doudna lab. Bacteria and archaea possess adaptive immunity against foreign genetic elements using CRISPR–Cas systems. Upon infection, new foreign DNA sequences are captured and integrated into the host CRISPR locus as new spacers. The CRISPR locus is transcribed and processed to generate mature CRISPR RNAs (crRNAs), each encoding a unique spacer sequence. Each crRNA associates with Cas effector proteins that use crRNAs as guides to silence foreign genetic elements that match the crRNA sequence.

PART 2: CRISPR/Cas9

CRISPR

The functions of CRISPR and CRISPR-associated (Cas) genes (nucleases) are essential in adaptive immunity of many bacteria and in the majority of characterized Archaea to protect them against viruses and plasmids (21-24) (FIGURE 2). As the name suggests the CRISPR system incorporates sequences from foreign DNA in its own genome in arrays of repeat sequences. These repeat arrays are transcribed into CRISPR RNAs (crRNAs). Each crRNA consist of a constant part of the repeat and the specific incorporated foreign DNA, which is known as the ‘protospacer’ sequence. The crRNAs associate with a second RNA, the transactivating CRISPR RNA (tracrRNA) (25). crRNA-tracrRNA hybrids together from a guide RNA that recruits Cas enzymes to bind and cleave incoming pathogenic DNA carrying the complementary sequence of the protospacer (26-31). To prevent cleavage of the protospacer sequences which are incorporated in its own bacterial or archaeal DNA, the Cas nuclease needs to have a direct interaction with an additional short motif. This motif is positioned in the target DNA right next to the protospacer-

encoded sequence called protospacer adjacent motif (PAM) (26-28, 32). These PAMs are absent in the CRISPR repeat arrays of the bacterial or archaeal genome (FIGURE 1D).

Exploring the Cas9 nuclease from *Streptococcus pyogenes* led to a system that is now worldwide employed as a genome editing tool (26). The DNA target sequence is specified by a 20-nt target recognition segment in the guide RNA. Cas9 is guided to the site and upon binding; Cas9 uses its HNH and RuvC catalytic domains to create a precise DSB three nucleotides before the end of the target sequence (26, 32). This break occurs only when the target site is located adjacent to a PAM sequence that matches 5'-NGG. To simplify the system, the crRNA–tracrRNA duplex was fused into a chimeric single guide RNA (sgRNA) (26). Thus, any desired DNA sequence of the form N₂₀-NGG can recruit the Cas9 nuclease by simply customizing the first 20 nucleotides of the guide RNA.

A large variety of Cas9 proteins exists in different bacteria and they efficiently induce genome editing (32-34). In addition to *Streptococcus pyogenes* (SpCas9), these include *Neisseria meningitidis* (NmCas9), *Staphylococcus aureus* (SaCas9) and *Streptococcus thermophiles* (StCas9). The various orthologues increase the usability of the CRISPR systems, because these Cas9 enzymes recognize alternative PAM sequences and use distinct crRNAs and tracrRNAs. Another interesting outlook is the combinatorial use of orthologous Cas9. Targeted gene knockout and targeted transcription activation become possible in a single cell. Recently, the first combinatorial CRISPR screen was demonstrated as a proof-of-principle (35). However, with such an approach, the number of gRNA combinations is increasing exponentially when adding more target genes. Thus, it is limited to a preselected group of gRNAs to keep the strategy feasible.

APPLICATIONS OF CRISPR

CRISPR as a targeted nuclease

The CRISPR system has made it possible to knock-out target genes in various cell types and organisms more quickly and more efficiently. The main advantage of CRISPR technologies is the ease with which only ~80-nt sgRNAs need to be synthesized to direct Cas9 to unique target sequences and cut the DNA effectively. This gives the possibility to use the Cas9 platform for large-scale genome-wide knockout screens in a search of genes that contribute to a biological process of interest (FIGURE 3A). With the previously available techniques, this was not feasible (36-39). Researchers were only able to perform large-scale with RNA interference (RNAi)-based screens with pooled-RNA libraries (40). The RNAi molecules inhibit gene expression by pairing to complementary mRNA molecules (36, 41-44).

Analogous to RNAi screens, sgRNA libraries have been generated for CRISPR/Cas9 targeting gene coding regions. This approach generates mutations at the targeted loci that may cause complete loss of gene function. Sequencing of the sgRNAs in the library treated cell pool show gains or losses of particular sgRNAs that identify genes of interest (FIGURE 3A) (45). A CRISPR/Cas9 screen usually results in a knockout and a more pronounced phenotype due to a complete loss of function instead of a knock-down seen in a RNAi screen. CRISPR-based screens have already successfully identified essential genes (46, 47) and drug targets (48, 49). In addition to targeting the coding DNA, CRISPR-based screening is also used to characterize enhancer elements and regulatory sequences (50, 51). This type of analysis is important to clarify the role of the non-coding genome.

CRISPR as a targeted modifier

Beyond targeted genome editing, the CRISPR/Cas9 system can also be employed to monitor specific chromosomal loci or to regulate endogenous gene expression in living cells. For that purpose a nuclease-deactivated Cas9 (dCas9) variant was engineered. It carries D10A and H840A mutations that disrupt the HNH and RuvC cleavage domains. dCas9 has been fused to effector domains such as GFP, transcriptional activators, repressors, and epigenetic modifiers (41, 52-55) that can subsequently be targeted by sgRNAs to specific sites in the genome (26, 32). For example, an eGFP-dCas9 fusion has been used to visualize DNA with repetitive sequences, such as telomeres, using a single sgRNA. For a locus without repetitive sequences 26 to 36 tiled sgRNAs across a 5-kb stretch of DNA were required to visualize a locus *in vivo* (56) (FIGURE 3C). This imaging strategy provides a new possibility to study the conformation and dynamics of chromosomes *in vivo*. Furthermore, it has been demonstrated that dCas9 fused to the transcriptional activation domain VP64 or the transcriptional repression domain KRAB (the Krüppel-associated box domain) can respectively upregulate and downregulate the expression of targeted genes in human (52, 57-59) and mouse cells (60) (FIGURE 3D-E). The use of dCas9 fusions is exciting because it offers the opportunity to regulate multiple genes in multiple ways (i.e. using activation and repression) in a single cell without with overexpression constructs. This bring us closer to the possibility to reprogram cells by tuning defined sets of genes with high precision and thereby controlling cell behaviour and identity.

CHALLENGES IN CRISPR-MEDIATED GENOME EDITING

Success of this technique is dependent on two processes; targeting the nuclease to the correct place and mutagenesis by imperfect repair of the DSB. Although Cas9 has great potential for both research and therapeutics, improvements can still be made. In contrast to RNAi, ZFN or TALEN which in principle can target any sequence, the target sites for

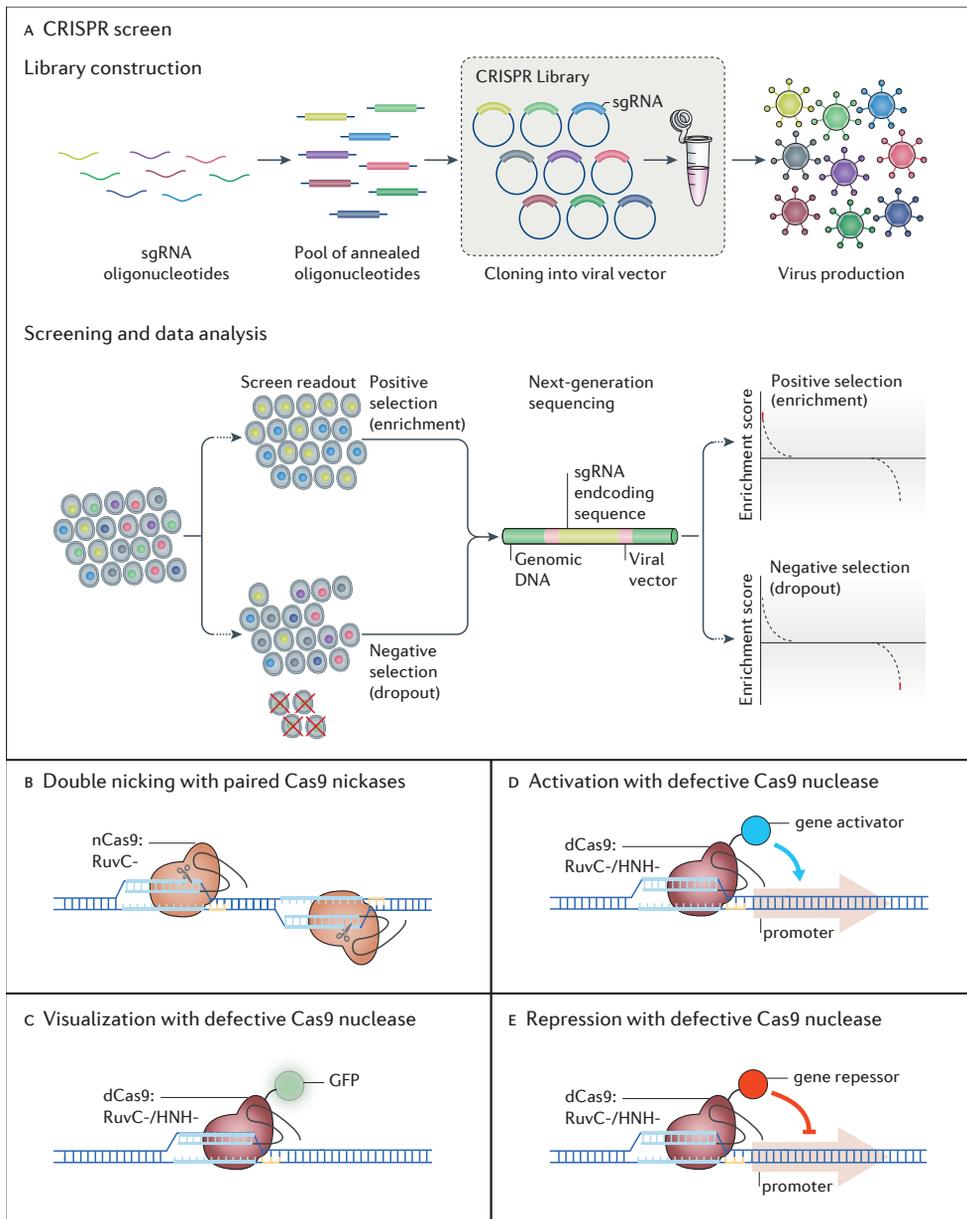


Figure 3: CRISPR/Cas9 System Applications. (A) Adapted from Lopes et al. (190). The sgRNA oligonucleotides targeting suitable Cas9 cleavage sites are synthesized, annealed with 3' and 5' cloning primers, pooled and cloned into viral constructs to produce a sgRNA expression library, from which viruses are produced to confer stable expression of sgRNAs in cells. Virus transduction of cells should ideally be performed such that each cell expresses only one sgRNA, but that all the sgRNAs are expressed in the transduced cell population, to maintain the complexity of the library. The transduced cells are subjected to a proliferation-based screening selection to identify sgRNAs that confer cell growth advantage or disadvantage according to the designed assay, and next-generation sequencing is used to assess which sgRNAs were enriched or depleted (shown in red) in the selected cell population. (B) A pair of Cas9 nickases (Cas9n). A mutation in one of the cleave domains of Cas9 results in a site specific single-strand nick. A pair of Cas9n/sgRNA complexes can nick both strands simultaneously and introduce a staggered double-stranded break.

(legend continued on next page)

CRISPR/Cas9 are limited to DNA stretches adjacent to a PAM sequence (NGG). With the development of Cas9 orthologues, a much broader spectrum of target sites in the genome became available already. Additionally, it should be noted that for a complete gene knockout it is mandatory that all copies of a particular gene are mutated by CRISPR/Cas9. This makes a knock out screen more challenging as normal cells usually have two alleles. More than two alleles are often found in cancer cells (61).

A more significant challenge lies in the specificity and efficacy of the method, especially for use in clinical applications. A big concern is potential off-target cleavage activity where a designed Cas9/sgRNA also induces a DSB elsewhere in the genome other than the intended target site, resulting in unwanted mutations (62-64). Reducing the concentration of either the Cas9 or sgRNA in the cells or minimalizing the duration of exposure of the CRISPR complex using inducible systems will diminish this problem (62, 65). Yet, these strategies often come at the cost of efficiency, which is important to successfully obtain a desired model system without having to screen hundreds of cells. In screens only efficient gRNAs are detected with a clear phenotype above the background. For clinical applications, one should keep in mind that an enormous number of affected cells have to be effectively mutated to influence the disease in a patient with for example Duchenne Muscular Dystrophy or Retinitis Pigmentosa. This emphasizes the importance of an efficient system.

Specificity

To assess CRISPR specificity, guide RNA variants containing one to four mismatches in the protospacer region have been generated and tested for their capability to guide Cas9 nuclease to a reporter-gene (63) or endogenous gene target sites (15, 62). Mismatches at the 5' end often appeared to be harmless for the recognition of the intended target. In contrast substitutions in the 3' end are less tolerated. Therefore, it was concluded that the stretch of 8-12 base pairs at the 3' end (seed area) is very important for target recognition (12, 15, 26, 66, 67). However, this rule does not apply to all single or double mismatches; it has been reported that some mismatches in the 5' end decreased the specificity, while other mismatches in the 3' end did not have a marked effect (63).

A complementary approach to study the specificity was to investigate the activity of Cas9 to target potential off-target sites, (i.e. loci that have few nucleotide mismatches compared to the designed gRNA target sequence). Algorithms were developed to find possible off-target sites in the human genome that differ 1–6 nucleotides with the on-target site (62, 63, 68-71). Sequence analysis revealed that off-target sites that differ

The D10A mutation (RuvC-mutant) renders cleavage of only the strand complementary to the sgRNA and generated a 5' overhang. The H840A/N863A mutation (HNH-mutant) cuts only the strand similar to the sgRNA and leaves a break with a 3' overhang. (C-E) Nuclease-deficient Cas9 with mutations in both cleavages domains can be fused with various effector domains allowing specific localization. For example with fluorescent proteins (C) transcriptional activators (D) and repressors (E).

by as many as five positions in the protospacer region, can still be edited by the CRISPR system (63). In addition the alternative PAM sequence, NAG, appeared to be effective for targeting by the sgRNA-Cas9 complex (62). Surprisingly, some research groups observed that the resulting insertion and deletion (indel) mutations at these off-target sites have sometimes comparable frequencies as those for the on-target site (63, 70, 71). Another whole-exome sequence study of three CRISPR treated K562 cell lines did not find evidence for Cas9-induced off-target mutations (72). Overall, these results suggest that the contribution of off-target editing is variable for diverse guide RNAs and that it is possible to target locations in the genome with high specificity. Based on published studies it is still difficult to predict the precision of a particular guide RNA.

Efficiency

Like the variability in cleaved off-target sites, there is also a high variability of on-target efficacy. Several research groups have published web-based software for the identification of CRISPR target sites and potential off-target sites in the organism of interest (e.g., CHOPCHOP (<http://chopchop.cbu.uib.no/>) (73) and the CRISPR Design Tool (<http://crispr.mit.edu/>) (62)). Nevertheless there is a lack of knowledge regarding the underlying rules that determine whether the CRISPR/Cas9 system will effectively target a given region of interest. To obtain an effective guide that introduces indels normally a few gRNAs should be tested per target site. For the incorporation of designed mutations by donor template-mediated HDR various guide and template combinations should be tested to find an efficient one.

To improve gRNA design for nuclease Cas9 to maximize the predictability of efficiency, several groups extracted information from large data sets to find correlations for specific sequence compositions (74, 75). Doench *et al.* constructed a library of sgRNAs targeting all possible sites across a handful of genes and tested their ability to make full gene knockout using antibody staining and flow cytometry readout (39). Sequence features were revealed that make sgRNAs most effective in various contexts. For example, a sgRNA expressed from a U6 promoter in mammalian cells should not contain a stretch of four or more uracils (U's) in a row otherwise RNA polymerase III will prematurely terminate the transcript (76). Also a stretch of U's near the 3' end of the guide sequence is unfavourable for Cas9–sgRNA binding (36). In general long stretches of the same nucleotide greatly decrease sgRNA activity (41). The results are mixed concerning the effect of GC content. A paper by Wang *et al.* suggests that sgRNAs with a very high or low GC content are less effective when combined with nuclease Cas9 (36). Another study reported that variations in GC content did not significantly change the effectiveness of dCas9 fused to effectors (41). Although progress has been made in predicting more effective gRNAs, it is obvious that more factors than gRNA sequence alone affect CRISPR/Cas9 efficacy.

IMPROVED CRISPR MOLECULES

To improve the specificity and efficiency of the CRISPR system a closer look has been taken at the *SpCas9* protein structure, guide RNA secondary structure, spacer sequence composition and length.

Optimization at the Cas9 level

To reduce the off-target effect of Cas9, Cas9 nickase mutants were developed with mutations of the catalytic residues (D10A in RuvC or H840A in HNH) (15, 17, 32) (FIGURE 3B). In contrast to wild-type Cas9, the nickase variants introduce gRNA-targeted single-strand breaks in DNA instead of the double-strand breaks. By using two Cas9 nicking enzymes directed by a pair of gRNAs targeting opposite strands of a locus results in a DSB while minimizing off-target activity (15, 69, 77, 78). Alternatively, protein engineering of *SpCas9* produced a high fidelity variant with reduced non-specific DNA contact while retaining on-target activity (79, 80). Therefore, substitutions were introduced into the Cas9 domain that interacts with the gRNA and the target DNA resulting in the variants with high specificity called eCas9(1.1) (79) and Cas9-HF1 (80). It is thought that the substitutions diminished the stability of the Cas9–gRNA interaction that introduces conformation changes necessary for active cleavage, thereby favouring the on-target cleavage (79, 80). An alternative explanation is that the engineered Cas9 molecules are unable to undergo conformational change to activate the NHN nuclease domain when bound to mismatched targets (81). Based on this hypothesis, new Cas9 variants were developed with high specificity: HypaCas9 (81) and evoCas9 (82). Both carried mutations in the REC3 domain. This domain binds to the RNA–DNA duplex and is believed to be important for linking the active part of the NHN domain (81).

Optimization at guide RNA level

At the RNA level, it was shown that off-target effects were minimized by decreasing the length of the gRNA–DNA pairing to 17–18 bp. Longer constructs can compensate for mismatches and still retain robust binding, while shorter gRNAs have less complementary RNA to bind the DNA and as a result are more sensitive to mismatches. gRNAs with decreased pairing length generally functioned efficiently at the intended target site and provides a simple flexible approach to minimize the off-target effects. The use of shorter gRNA does not impair the targeting range because a site of 17 or 18 nt of complementarity is equally unique in the human genome as those target site of 20 nt (70).

gRNAs were studied that comply with the above mentioned rules to improve efficiency, but it resulted in poor cleavage activity *in vitro* and *in vivo*. These impaired gRNA sequences revealed that potentially hairpin structures could be formed in the protospacer region of the gRNA. Substitutions that disrupt these predicted hairpins improved cleavage whereas

control substitutions in different areas of the gRNA were neutral (83). A model was proposed where the constant scaffold of the gRNA binds strongly to the Cas9 protein while the protospacer sequence has more conformational freedom. This allows the protospacer sequence to invade the target DNA strand, but also to form potentially harmful secondary RNA structures, which explains why secondary structure of the protospacer sequence affects Cas9 activity (84). Considering that the RNA structure has a key role for the binding of the complex to the target DNA, it should be noted that the chimeric sgRNA is 10 bp shorter than the native crRNA-tracrRNA duplex (26). It was shown that this does not reduce functionally *in vitro*, while conflicting results are found *in vivo*. Dang *et al.* showed enhanced efficiency when extending the sgRNA, while Hsu *et al.* do not find an effect of extension (62, 85). Probably, the enhanced effect is dependent on the target site, but the extension of sgRNA has no reported negative effect. The Cas9/sgRNA complex seems optimal when it is stable enough to form a complex, but flexible enough to engage at target site to form a subtle ‘weak-ish’ interaction that is sensitive for mismatches.

PART 3: DNA Repair

DNA CLEAVAGE FOLLOWED BY DNA REPAIR

In order to perform successful targeted genome editing, a DSB is first introduced at the to-be-modified genomic location as explained above. Second, repair of the break has to occur whereby the DNA is edited. The break can be repaired by intrinsic cellular mechanisms, such as non-homologous end joining (NHEJ) or homology directed repair (HDR). In addition to these, other (back up) DNA repair pathways have been described and are grouped under the name of alternative end joining (A-EJ) (86). The roles of the latter are less well understood. Moreover, the choice of pathway is also not completely clear. The relative usage of the various pathways may depend on species, cell type, phase of the cell cycle and chromatin state in which the DNA damage is encountered.

DNA damage response

Cells have evolved mechanisms that act upon damage of the DNA, collectively referred to as DNA damage response (DDR). Sensors detect DNA lesions after which a series of signal transductions is initiated by protein kinases (87). The initial activated kinases are the ataxia telangiectasia-mutated protein (ATM), the ATM and Rad3-related kinase (ATR) and the DNA-dependent protein kinase (DNA-PK). These kinases have the ability to phosphorylate

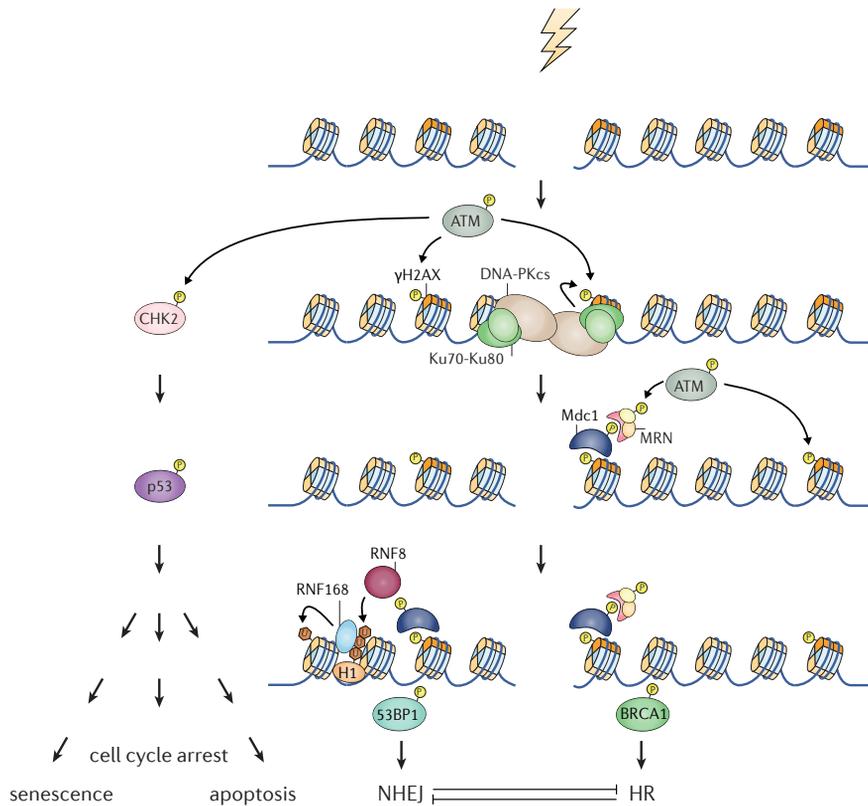


Figure 4: DNA damage response. Figure elements adapted from (191-193). Schematic representation of DNA damage response signalling pathway. ATM responds to DNA double-strand breaks and is activated, this is followed by the phosphorylation of H2AX and localization of MDC1 to the break. Then several repair proteins are recruited to the site of damage e.g. via ubiquitination of H2AX by RNF8 and RNF168 E3 ubiquitin ligases precedes recruitment of repair proteins such as BRCA1 and 53BP1. ATM also regulates cell-cycle checkpoints through the activation of CHK2 and p53. See text for further details.

residue 139 of histone variant H2AX (γH2AX) at the chromatin flanking the breakage site (88, 89). Proteins involved in repair and checkpoint activation are then recruited to the DSB site, visible as foci in immunofluorescence (FIGURE 4). ATM, MDC1, the MRN complex and the RING finger E3 ubiquitin ligases RNF8 and RNF168 are among the earliest factors found in DNA damage foci (90). 53BP1 and BRCA1 appear later and their recruitment depends on the aforementioned upstream factors (91). Phosphorylation of 53BP1 or BRCA1 by ATM plays a role in the selection of the repair pathway to control resection of the DNA ends. It has become clear that the degree of 5' to 3' resection has a major impact on the choice of repair pathway. DNA-ends with long 3' overhanging tails are destined for HR repair (92, 93). 53BP1 is a negative regulator of resection (94) while BRCA1 promotes the removal of 53BP1 to enable resection (95), but how cells switch from a preferred NHEJ to a resection-dependent pathway is unclear.

Canonical nonhomologous end-joining

The canonical form of NHEJ ensures that the broken DNA ends are joined together. In this process, the first protein that responds to double-strand breakage is Ku70/Ku80, which is present in very high concentrations in cells (FIGURE 5A). Once bound, the Ku-heterodimer serves as a scaffold to recruit other NHEJ factors to the damage site including DNA-PKcs (96), X-ray cross-complementing protein 4 (XRCC4) (97-99), DNA ligase 4 (98), XRCC4-like factor (XLF) (100) and end-processing enzymes (polymerases μ and λ , and the Artemis nuclease). Upon binding of DNA-PKcs to the DNA-Ku complex, the Ku-heterodimer translocates further on the DNA strand (101, 102). At the DNA ends, the DNA-PKcs molecule forms a specific structure that holds the two sites close together (103-105). This complex of Ku and DNA-PKcs prevents access to nucleases and ligases to process the DNA termini (105, 106). Subsequently, DNA-PKcs is activated, which in turn mediates autophosphorylation as well as phosphorylation of other NHEJ factors (107). Autophosphorylated DNA-PKcs causes a large conformational change that is thought to promote its dissociation from the DNA ends. Access of end-processing enzymes to the termini of the double strand break is then allowed (108-110). End-processing includes the removal of mismatched nucleotides by nucleases and/or resynthesis by DNA polymerases to create ends that are compatible for ligation. Different end-processing enzymes are active, depending on the status of the DNA termini. Ligase 4 and its co-factor XRCC4 anneal the DNA ends. Although the exact role of XLF is unknown, it interacts with the XRCC4/DNA ligase 4 complex and is therefore thought to participate in the ligation step (111).

Homology directed repair

Higher eukaryotes are also capable of repairing DSBs by using the sister chromatid as a homologous template (FIGURE 5B). As consequence this homologous recombination pathway (HR) is limited to the S and G2 phases of the cell cycle, in which the sister chromatid copy is generated by DNA replication. The initial step in HR is DNA nucleolytic end resection at the break site by the MRN complex (comprising Mre11, Rad50, Nbs1). The MRN complex is an important sensor of DNA DSBs and promotes long-distance resection by the endo/exonucleases Exo1 and Dna2 together with additional proteins such as BLM helicase, CtIP and the tumour suppressor protein BRCA1 (112-117). During resection, nucleotides are removed from the 5' ends leaving long 3' single-stranded DNA (ssDNA) overhangs on both sides of the fracture. These 3' ssDNA tails are coated and stabilized by the replication protein A (RPA) complex. This complex is then displaced by Rad51 recombinase, forming Rad51 nucleoprotein filament. BRCA1 promotes the recruitment of BRCA2 (118, 119) which assists loading of Rad51 (120). The Rad51 recombinase then performs strand invasion by pairing with the complementary strand of the sister chromatid, thereby forming a D-loop. The invading strand is extended by DNA polymerase using the sister chromatid as a template until it reaches the area homologous to other side of the fracture. The lagging

strand also has a 3' overhang and can recover by either forming another junction with the homologous chromatid followed by gap filling or by extension along the receiving DNA duplex. DNA ligation links the DNA ends and newly synthesized sequences together.

Alternative and microhomology-mediated end joining

Studies have shown that in addition to C-NHEJ and HR a different pathway of DSB processing is operational. It is based on the simple end joining principles, but slower than C-NHEJ (half-lives from 30 minutes to 20 hours) (121-124). This repair route, commonly named alternative end joining (A-EJ) is Ku or ligase 4 independent (125). Proteins involved in A-EJ are PARP1, the MRN complex and CtIP that perform DNA end processing. PARP1 accumulates factors to promote ligation including the ligation complex XRCC1/Lig3 (86) (FIGURE 5C). Occasionally, microhomologies are utilized in this pathway to process the DSB, although the use of microhomologies is not an exclusive feature of A-EJ. Therefore, this subset of A-EJ is also termed microhomology-mediated end joining (MMEJ) (125-127). It is thought that A-EJ will engage at DSBs when either C-NHEJ or HR have attempted to process the DSB but somehow failed. Thus, at these DNA ends, factors of either C-NHEJ or HR can be present when A-EJ takes over the DSB processing.

DNA REPAIR IN CHROMATIN CONTEXT

There is increasing evidence that the chromatin micro-environment and specific histone marks around DSBs are crucial for the efficiency and fidelity of DNA repair pathways. Monitoring of chromosomally integrated fluorescent reporter substrates demonstrate that C-NHEJ and HR are strongly influenced by chromosomal location (128). This variation of pathway usage may be explained that direct repair of DSBs in compact structures like heterochromatin is a challenge that cells need to overcome to preserve genome integrity (129-132). It has been proposed that after damage in compacted DNA, the chromatin needs to decondense first before repair proteins have access to the lesions (129, 130). ATM kinase seems to play a role in this process. It has been reported that in ATM null cells the majority of DSBs (~85%) is repaired with normal kinetics while the remaining breaks stay unrepaired for longer times after damage (133). However, inhibition of ATM in parallel with knockdown of the heterochromatin proteins KRAB-associated protein 1 (KAP-1) or heterochromatin protein 1 (HP-1) rescues these persistent DSBs (129). This finding supports the idea that phosphorylation of KAP-1 at residue Ser824 by ATM drives the relaxation of heterochromatin (129-131). In addition, decreasing the chromatin compaction by histone deacetylases (HDACs) or by reducing the levels of linker histone H1, enhances DDR signalling (134, 135). In euchromatin, ATM inhibition had no major effect on the repair of DSBs (129).

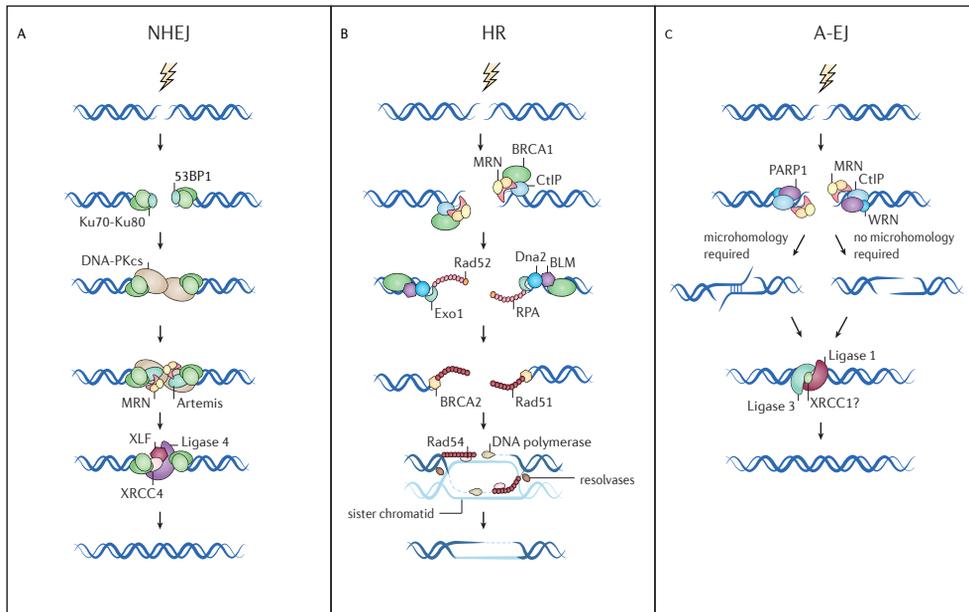


Figure 5: DNA repair pathways. Adapted from Iliakis *et al.* (194). (A) Canonical non-homologous end joining pathway. (B) Homologous recombination repair. (C) Alternative end joining. Two models of DSB repair by A-EJ are shown with or without using microhomologies. See main text for explanation.

Repair kinetics

Chromatin complexity and the need for relaxation upon damage has been suggested to delay repair kinetics in an effort to concentrate effector proteins to the damage site (130). DSB repair within heterochromatin was found to be roughly 2-fold slower than repair within regions of euchromatin (129). Inhibition of ATM lowered the rate of heterochromatic DSB repair further, while having little effect on repair of euchromatin (129). In contrast to this observation, Janssen *et al.* noticed similar kinetics for DSB repair in euchromatin and heterochromatin of *Drosophila* (136). Live cell imaging assays revealed that cells can use either C-NHEJ or HR to repair DSBs in heterochromatic and euchromatic regions of the genome with similar kinetics. A difference in the spatial displacement of the majority of the heterochromatin breaks is observed, but the movement is absent in the euchromatin breaks (136, 137). The relocation of foci has been shown to require the presence of resection proteins (137). It seems that breaks and their repair in heterochromatin behave differently when compared to euchromatin breaks and that the kinetics of chromatin decompaction may differ in various model systems.

DNA mobility

It has been suggested that there is a connection between the compaction state of the chromatin and the ability of a damaged locus to relocate (136, 137). In the absence of

damage, movements of the DNA are constrained by multiple cellular and physical properties leading to the retention of chromosomes within defined regions of the nucleus called 'chromatin territories' (138). In yeast, there is clear evidence that DSBs induce chromatin mobility (139). In higher eukaryotes, the issue of DSB is controversial due to conflicting results (140-144). As mentioned above, in *Drosophila*, single and global DNA damage leads to expansion of pericentromeric heterochromatin and relocation of heterochromatic foci to the periphery of the heterochromatin domains (136, 137). A similar relocation was observed upon single ion micro-irradiation of mouse chromocenters that represent constitutive heterochromatin (140). Conversely, UV or γ -rays induced DSBs were found to have only limited mobility, but did lead to a localized decondensation of chromatin (141, 142). Moreover, induction of multiply damaged sites (including DSBs, single strand breaks and base damages) did not cause relocation nor did nuclease induced DSBs in a heterochromatic transgene locus carrying >100 repeats (143, 144).

In yeast, resection of DNA ends was found to be key in regulating the mobility of breaks in the process of homology search (145). In mammalian cells, the constraint on mobility was shown to be dependent on Ku80, a component of C-NHEJ repair pathway. It appeared that the C-NHEJ machinery is tethered to the DSB ends for rapid repair, thereby limiting mobility (143). From these findings it was proposed that DSB relocation in heterochromatin depends on resection and DNA repair pathway choice (137). But how this choice is made in heterochromatin remains unclear. It is clear that a simple model where euchromatin or heterochromatin determines the type of damage repair pathway is an oversimplification.

Organizing DNA repair in the nucleus

The classic definition of transcriptionally active, open euchromatin, and compacted, silent heterochromatin, understates the high diversity of chromatin states. For example heterochromatin can have various different chromatin make-ups (146). The most compacted form of heterochromatin is typically rich in deacetylated histones and histone H3 trimethylation on lysine 9 (H3K9me3). This mark can be bound by repressive proteins such as KAP-1 and HP-1 (147-149). Another form of heterochromatin is more flexible and its level of compaction can change, for example during differentiation. This state is represented by histone H3 trimethylation on lysine 27 (H3K27me3) and polycomb-repressive complexes (146). Moreover, chromatin that is often found in genomic regions associated to nuclear lamina (lamina associated domains) is abundant in H3K9me2 and the boundaries are enriched with H3K27me3 (150). The majority of genes in these regions are silenced (146).

Several studies investigating DSB repair show that kinetics and choice of repair pathway may vary between the different heterochromatic compartments. It was found that DSBs induced by I-SceI in a mammalian locus that was experimentally tethered to the nuclear lamina could not recruit the HR associated factors BRCA1 and Rad51. Instead the breaks

were mainly repaired by C-NHEJ. Also the A-EJ pathway appeared to be active at these DSBs, possibly repairing breaks in which resection already had taken place (151). Interestingly, these breaks induced near the nuclear membrane did not relocate to areas that were more permissive for HR and were rather repaired by alternative end joining (151).

In contrast, DSBs introduced near nuclear pores, where the chromatin microenvironment is more open when compared to that at the lamina, use both C-NHEJ and HR pathway for repair (151). Also breaks in centromeric and pericentric heterochromatin Ku80 was able to recruit both the C-NHEJ protein Ku80 as well as the HR protein Rad51. Recruitment of Ku80 to the break occurs throughout the cell cycle and leaves locus positionally stable. Recruitment of Rad51 seems to be domain specific, centromeric lesions tether Rad51 protein at all stages of the cell cycle and relocate the foci toward euchromatin. DSBs in pericentric heterochromatin however recruit Rad51 exclusively at post-replicative chromatin at the periphery of the heterochromatin domain. The recruitment of Rad51 throughout the cell cycle in centromeric breaks is surprising since HR normally requires a sister chromatid for DNA repair, though the Rad51 recruitment is enhanced in G2. One might speculate that this could account for HR being licensed throughout the cell cycle and perhaps uses its own repeats in cis as a template or persist until it passes through S-phase (152, 153).

Although both types of heterochromatic domains are condensed, they are unique in chromatin modifications, DNA sequence and histone variant composition. Pericentric heterochromatin is enriched in H3K9me3 and HP-1s, while in the centromere core domain no H3K9me3 could be detected. The centromere core domain consists of nucleosomes carrying H3 and the H3 variant, CENP-A (154). The H3 nucleosomes comprise marks for active chromatin including H3K4me2, H3K36 methylation and H3 acetylation, (154). It was shown that H3K36me3 promote DNA ends resection and HR (155, 156). Possibly only the marks present at the centromere make the chromatin permissive for resection in G1.

In pericentric heterochromatin, DSBs are positionally stable in G1 and can recruit C-NHEJ factors. In S/G2, resection takes place and the DSBs are relocated to the periphery of the heterochromatin, where they are retained by Rad51. It has been proposed that the spatial movement of the break site prevents the activation of mutagenic pathways and illegitimate recombination between repetitive sequences in trans. As centromeres from different chromosomes are spatially separated within the nucleus and do not cluster together, the risk of chromosomal translocations is minimal in the presence of active HR.

Along the same line of thought two studies were performed in which DSBs were generated within repeats of nucleoli of mammalian cells using endonucleases and showed that the choice of repair pathways regulates the spatial movement of the break (153, 157). In both cases the DSBs and the rDNA chromatin itself were detected at the periphery of nucleoli indicating that relocation had occurred. The relocation was associated with transcriptional

silencing. Inhibition of ATM blocked the transcriptional silencing and prevented the reorganization of nucleoli and the rDNA. In addition, blocking of C-NHEJ resulted in enhanced nucleolar reorganization and transcriptional silencing, which was not observed by blocking HR. Repair of rDNA by HR was also found to generate a loss of rDNA repeats; this effect was increased by loss of C-NHEJ (158). These complementary studies suggest that C-NHEJ occurs rapidly within nucleoli to maintain rDNA transcription. However, when these breaks remain unrepaired by C-NHEJ they are transcriptionally silenced and relocalized to the nucleolar periphery where they can be recognized by the HR machinery. Altogether, different forms of chromatin regulate DNA repair pathway choice in a unique fashion.

Communication between DSB response and transcription

Although most heterochromatin is not transcribed, breaks in rDNA repeats indicated a link between DSB response and transcription silencing. A system was developed in U2OS cells to visualize the DSB response and the effect on nascent transcription simultaneously. Multiple breaks are introduced in a LacO cassette 4 kb upstream of an inducible YFP transcription unit in which the 3'-UTR (untranslated region) contains 24 repeats of a stem loop structure that is recognized by phage coat protein MS2 (159, 160). This enables real time visualization of the DSB introduced by mCherry-LacI-FokI and the nascent transcription through the expression of YFP-MS2. Introduction of a DSB upstream of the transcriptional start site effectively silences RNA Pol II-dependent transcription in an ATM and ubiquitin-dependent manner. Transcription was rapidly restored upon removal of FokI and DSB repair (161). This transcriptional reporter system also revealed that ATM-dependent silencing suppressed transcriptionally induced chromatin decondensation (160). The finding that the DSB response can suppress transcription associated chromatin decompaction seems contradictory, since DSBs themselves induce decompaction. However, it is likely that the pre-existing state of chromatin at the time of DSB induction influences the nature of the DSB response and the outcome of ATM signalling. Transcriptionally active regions are often more open as compared to inactive compacted DNA.

After transcription silencing, DNA repair occurs. Regions with actively transcribed genes were found to only associate with recruitment of the HR-protein Rad51 and not with the C-NHEJ protein XRCC4 using chromatin immunoprecipitation-sequencing (ChIP-seq). In these areas the transcription-elongation associated histone mark, histone H3 lysine 36 tri-methylation (H3K36me3) was present. In agreement with this, the methyltransferase placing this mark, SETD2, has been shown to be required for the recruitment of CtIP (CtBP-interacting protein) which in turn promotes DNA end resection and HR (162).

In a subset of these HR-prone regions, DSBs cluster together. Clustering of damaged genes occurs primarily during the G1 cell-cycle phase and coincides with delayed repair as has been shown by capture Hi-C. The study revealed that DSBs induced in active genes

are prone to be repaired by HR in post replicative cells, but to be refractory for repair in the G1 phase (163). Interestingly, this behaviour is similar to other large scale DSB mobility events in heterochromatin mentioned before that are associated with persistent or 'difficult' DSBs (137, 139, 152, 153, 164). The reasons underlying repair deficiency at active genes in G1 remain unknown. Maybe, DSBs cluster to be prepared for faithful repair. Since the other available pathway that can accommodate resected/processed ends in G1 is A-EJ, this might be too detrimental for the cell given the high mutation rate associated with this pathway. Clustering may help to inhibit such error prone repair pathway to sequester DSBs from the rest of the genome, while awaiting a more appropriated cell cycle phase (163). Overall it emerges that spatial positioning of DSBs in the highly compartmentalized nucleus may have significant implications for fidelity and choice of repair pathways.

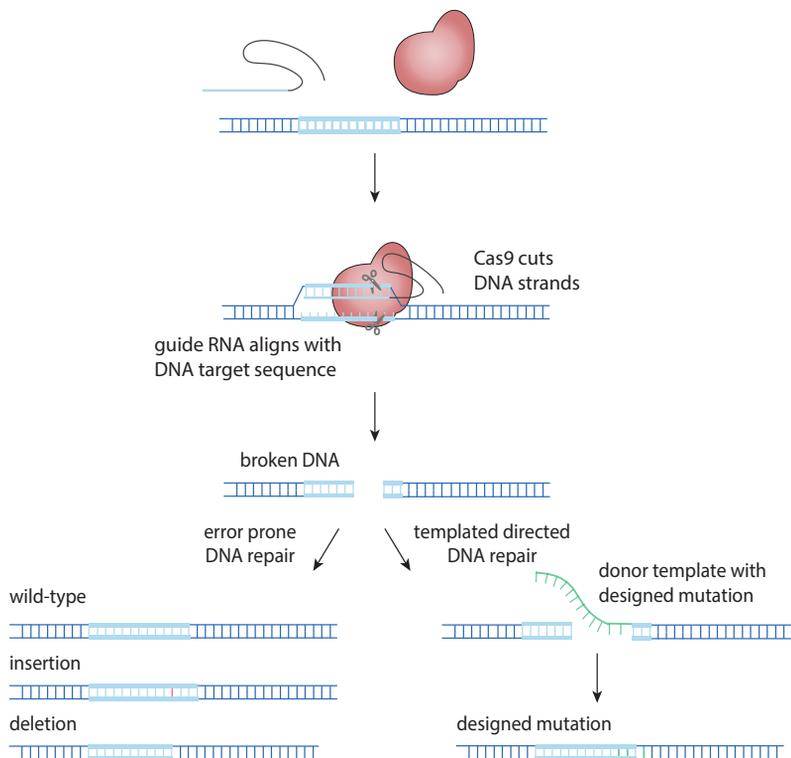


Figure 6: Genome editing with CRISPR/Cas9. The Cas9 nuclease is directed to target site by its sgRNA and introduces a double stranded break. The break is repaired by one of two mechanisms: 1) Non-homologous end joining which can creates random insertions or deletions at the targeted site or 2) Homology directed repair which creates precise changes based on template DNA.

PART 4: CRISPR/Cas9 – DSB Repair – Chromatin Interplay

REPAIR FIDELITY OF CRISPR INDUCED BREAKS

How can repair by endogenous pathways result in edits in the DNA after a Cas9-induced DSB? The C-NHEJ pathway ligates the two broken ends together and has no built-in mechanism of restoring the original sequence around the DSB. Therefore, small mistakes such as insertion or deletion mutations of various lengths can be introduced at the targeted location during repair. When these indels land in the coding or regulatory region of a gene they may lead to functional knockouts due to disruption of the reading frame of a gene, the promotor region, binding sites for transcription factors or enhancer regions (165). Using the homology directed repair (HDR) pathway, designed (point) mutations or specific sequences can be inserted by recombination of the target locus with exogenously delivered DNA donor templates (166, 167) (FIGURE 6).

Several methods have been developed to monitor the induction and re-joining of DNA DSBs in the genome. Direct detection of DSBs includes comet assay (168) or pulsed-field gel electrophoresis (121). These techniques lack sensitivity and are unable to monitor DSB repair where only a few DSBs are induced or remain. Other strategies use immunofluorescence against DSB markers such as γ H2AX, combined with microscopy, flow cytometry or chromatin immunoprecipitation (169-171). These techniques are limited by the requirement to use fluorescent proteins or luciferase-based readouts as a substitute for DSB repair activity. The difficulty with tracking of molecular components is that it is unknown how their accumulation and dissociation at the break site relates to the actual process of repairing of the DNA break. Recently, an alternative approach was reported in which next-generation sequencing was used to study DSB formation and DNA repair. This method does not depend on the expression of reporter genes and provides a direct read-out for repair and has the power to study multiple sites at the same time (172-176). Several computational tools have become available to analyse the sequence data (177-179).

In a systematic study, the repair outcomes of 223 CRISPR targets were monitored in the human genome. It was shown that at some sites one or two repair events were dominant, while at other locations a wide variety of repair events took place at lower frequency. After Cas9-induced DSB the pattern of DNA repair at each target site appeared not to be random and was consistent between experimental replicates, cell lines or reagent delivery methods (180). Using different reporter cell lines and inhibitors it was demonstrated that multiple repair pathways can resolve a single Cas9-mediated DSB. From these experiments, it has been suggested that the presence and polarity of the

overhanging structure is a critical determining factor for the pathway choice of double-strand break repair (180-182). This assumption was supported by the observation that where a Cas9 nickase mutant produced a staggered DSB, it resulted in different repair products (181) (FIGURE 3B). In addition, micro-homologies were found in the DNA sequence neighbouring the DSB where larger deletions were introduced during repair in the absence of DNA-PKcs (180). The presence of microhomologies and the type of DNA ends clearly have implications for the choice of repair pathway and offer opportunities to improve the ability to steer genome editing outcome. However, for a particular target site or CRISPR variant, the repair outcome results in a complex mixture of multiple mutations. It is a matter of balancing between the preferences for repair pathways in order to optimize the desired DNA editing outcome.

CRISPR IN CHROMATIN CONTEXT

Although the CRISPR system has been optimized *in vitro* in test tubes, most applications of the CRISPR are *in vivo* in cells and animals. There are still discrepancies between the activity of CRISPR/Cas9 on episomal targeted DNA or genomic DNA, suggesting that like for DNA repair, chromatin structure influences the working of CRISPR.

Accessibility

Cas9 nuclease activity has been shown to correlate with the absence of repressive histone marks and increased accessibility. This was demonstrated by a reporter locus construct harbouring an array of tetO elements that can switch from compact to relaxed chromatin through doxycycline (Dox) dependent release of tTR-KRAB fusion protein. Binding of KRAB proteins triggers recruitment of chromatin remodelling factors such as KAP-1 and HP-1 resulting in epigenetic silencing. This system and similar reporter assays revealed that in a closed chromatin conformation CRISPR/Cas9 nuclease yields less edited target sites than in relaxed chromatin (183, 184).

A library-on-library approach demonstrates a similar correlation. Hereby the activity of CRISPR/Cas9 was evaluated with a library of sgRNAs of ~1400 endogenous target sites, compared to CRISPR/Cas9 activity of a lenti virus library of corresponding target sites integrated mostly in open chromatin. The target sites that had different editing frequencies in the two libraries were often found in region of low DNA accessibility (185). Furthermore, in zebrafish CRISPR/Cas9 mutagenesis efficiency was found to be positively correlated with chromatin accessibility at different stages of development (186). Additionally, single molecule imaging studies have demonstrated that dCas9 explores euchromatin more frequently than it does heterochromatin (187). All these studies together indicate that Cas9 nuclease is less active in more compacted chromatin.

At a smaller scale, chromatin folding by the nucleosomes restricts the activity of CRISPR/Cas9. Detailed biochemical studies with a variety of nucleosomal templates and *in vivo* studies using Mnase occupancy, demonstrated that the intrinsic stability of the histone-DNA

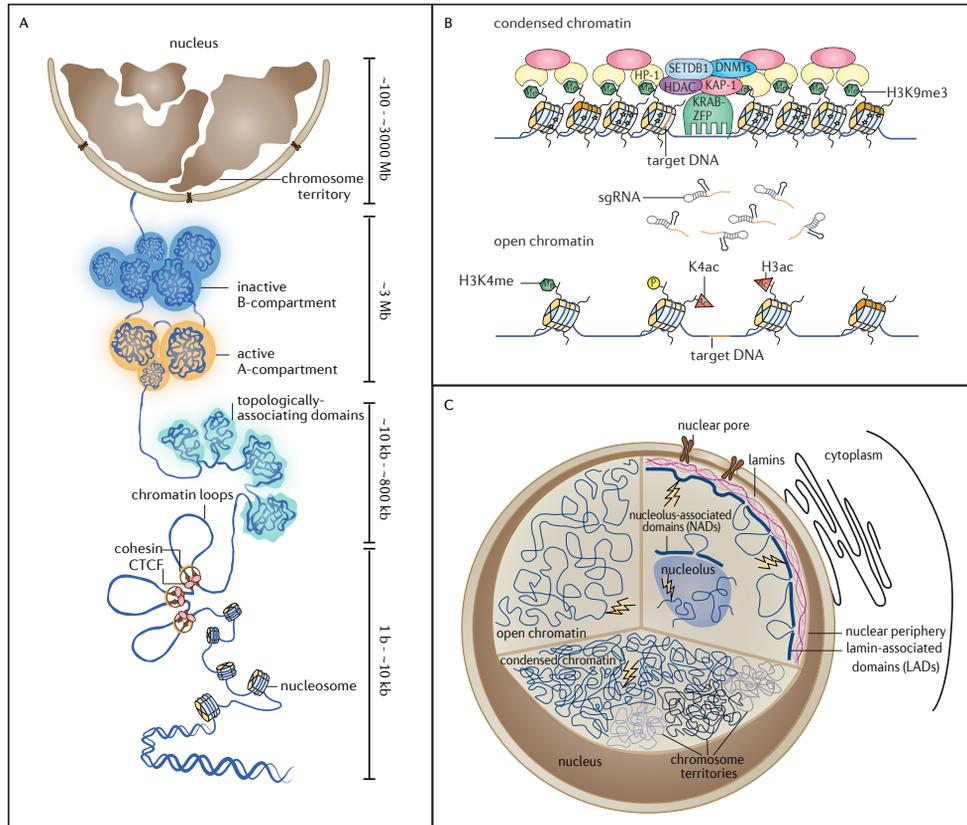


Figure 7: Nuclear architecture and DNA damage. (A) A scheme illustrating the hierarchical structure of interphase chromatin adapted from Razin et al. (195). Chromosome territories (at the top of the picture) are partitioned into A- and B-compartments formed by long-range spatial interactions between distant genome loci and containing active and repressed genome regions, respectively. At a sub-megabase level, chromatin is folded into topologically-associating domains, TADs, commonly interpreted as self-interacting globular structures the positions of which are largely conserved across cell types. The internal structure of TADs is represented by arrays of so-called loop domains formed by spatial contacts between CTCF/cohesin-binding sites. (B) Figure elements adapted from (191, 196, 197). Several epigenetic regulators define the chromatin state of cells. Relevant epigenetic marks include histone modifications, DNA methylation and incorporation of different core histone variants (yellow and orange cylinders) that alter accessibility of the DNA (dark blue line). The main histone marks, the active H3K4me3 and the repressive H3K9me3 are positively regulated by specific histone methyltransferases (HMTs; including SETDB1) and negatively regulated by the respective histone demethylases (HDMs). The methyl group is indicated as a green hexagon (Me). Histone acetylation also marks active chromatin, and the acetyl group (the red triangle, Ac) can be added through histone acetyltransferase (HATs) and removed by histone deacetylases (HDACs). Phosphorylated histone residues are often associated with gene activation (yellow circle, P). DNA methylation (yellow stars) is typically present in heterochromatin (marked by H3K9me3 and HP-1). DNA can be hypermethylated, as a result of the action of DNA methyltransferases (DNMTs). In euchromatic regions DNA is generally unmethylated. The chromatin make-up of a region can influence the efficiency of CRISPR guide RNA. (C) Adapted from Stratigi et al. (198). DNA double strand breaks and repair in various structures in the nucleus that affect the DNA mobility and repair kinetics, e.g. damage in lamin associated domains, near nuclear pore or at the nucleolus. The radial distribution of chromosome territories in the nucleus as well as the level of chromatin compaction affect the DNA accessibility to damage and the DNA repair kinetics.

interactions, the location of the target site within the nucleosome and the action of chromatin remodelling enzymes play critical roles in regulating the activity of *SpCas9* (FIGURE 7B). Target sites located in DNA that is wrapped around a nucleosome are subjected less to digestion than sites in the linker DNA between nucleosomes. The activity could be recovered when the nucleosome was relocated by remodelling enzymes (188, 189).

To improve CRISPR/Cas9 activity in compact chromatin regions, chromatin decondensation or derepression by chromatin-factor drugs such as histone deacetylase (HDAC) inhibitors or DNA methyltransferase could be a strategy. However, such an approach may affect the cells in unintentional or undesired ways. A localized approach could be beneficial, for example in combination with a Cas9 orthologue or TALE fused to a decondensation effector protein.

PART 5: Outline of Thesis

CRISPR/Cas9 is powerful technology that has greatly changed the scientific field for genome editing and has the potential to have an impact on gene therapy/genome modification in future. Despite the broad application, the process of repair of Cas9-induced DSBs has been only partially characterized. It is clear that both the sequence and location are important for guide efficacy, but it is not known how long it takes before an individual Cas9-induced DSB is repaired, how error-prone this process is and what the influence of chromatin is on these aspects. To reconcile current discrepancies it will be important to develop systems whereby DSBs can be induced within different chromatin states in the same biological system to determine how this influences chromatin dynamics. However, at present methods to track CRISPR/Cas9 induced DSB repair with high specificity and resolution throughout the genome in time are lacking. Therefore a toolbox of methods is developed to study the fidelity and kinetics of repair CRISPR/Cas9 induced DSBs; which repair pathways are involved and how the chromatin status affect these processes.

CHAPTER 2 introduces TIDE, a method for quantitative detection of insertions and deletions after repair of a targeted DSB.

CHAPTER 3 describes TIDER, a method based on the TIDE algorithm to quantitate the number of homologous directed repair events driven by a donor template.

CHAPTER 4 specifies the major procedures and nuances that help improve the TIDE and TIDER methods.

CHAPTER 5 presents quantitative modelling of the accumulating indels after Cas9-induced DSB at a single locus in the genome to study the kinetics and fidelity of cutting and repair.

CHAPTER 6 describes a strategy to monitor the chromatin effect on CRISPR/Cas9 induced DSB repair. A variant of the TRIP assay was designed to track DSB repair at multiple loci in the genome in parallel.

CHAPTER 7 closes with a general discussion of results presented in this thesis and highlights the direction of future research.

REFERENCES

1. Freedman, B.S., Brooks C.R., Lam A.Q., Fu H., Morizane R., Agrawal V., Saad A.F., Li M.K., Hughes M.R. *et al.*, Modelling kidney disease with CRISPR-mutant kidney organoids derived from human pluripotent epiblast spheroids. *Nat Commun* 6, 8715 (2015).
2. Bellin, M., Casini S., Davis R.P., D'Aniello C., Haas J., Ward-van Oostwaard D., Tertoolen L.G., Jung C.B., Elliott D.A. *et al.*, Isogenic human pluripotent stem cell pairs reveal the role of a KCNH2 mutation in long-QT syndrome. *EMBO J* 32, 3161-3175 (2013).
3. Rouet, P., Smih F., Jasin M., Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol* 14, 8096-8106 (1994).
4. Plessis, A., Perrin A., Haber J.E., Dujon B., Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics* 130, 451-460 (1992).
5. Porteus, M., Genome Editing: A New Approach to Human Therapeutics. *Annu Rev Pharmacol Toxicol* 56, 163-190 (2016).
6. Silva, G., Poirot L., Galetto R., Smith J., Montoya G., Duchateau P., Paques F., Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr Gene Ther* 11, 11-27 (2011).
7. Miller, J., McLachlan A.D., Klug A., Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J* 4, 1609-1614 (1985).
8. Kim, Y.G., Cha J., Chandrasegaran S., Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* 93, 1156-1160 (1996).
9. Urnov, F.D., Rebar E.J., Holmes M.C., Zhang H.S., Gregory P.D., Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11, 636-646 (2010).
10. Boch, J., Scholze H., Schornack S., Landgraf A., Hahn S., Kay S., Lahaye T., Nickstadt A., Bonas U., Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509-1512 (2009).
11. Christian, M., Cermak T., Doyle E.L., Schmidt C., Zhang F., Hummel A., Bogdanove A.J., Voytas D.F., Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757-761 (2010).
12. Jiang, W., Bikard D., Cox D., Zhang F., Marraffini L.A., RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 31, 233-239 (2013).
13. DiCarlo, J.E., Norville J.E., Mali P., Rios X., Aach J., Church G.M., Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res* 41, 4336-4343 (2013).
14. Li, J.F., Norville J.E., Aach J., McCormack M., Zhang D., Bush J., Church G.M., Sheen J., Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat Biotechnol* 31, 688-691 (2013).
15. Cong, L., Ran F.A., Cox D., Lin S., Barretto R., Habib N., Hsu P.D., Wu X., Jiang W. *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819-823 (2013).
16. Mali, P., Yang L., Esvelt K.M., Aach J., Guell M., DiCarlo J.E., Norville J.E., Church G.M., RNA-guided human genome engineering via Cas9. *Science* 339, 823-826 (2013).

17. Jinek, M., East A., Cheng A., Lin S., Ma E., Doudna J., RNA-programmed genome editing in human cells. *Elife* 2, e00471 (2013).
18. Hsu, P.D., Lander E.S., Zhang F., Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262-1278 (2014).
19. Wang, H., Yang H., Shivalila C.S., Dawlaty M.M., Cheng A.W., Zhang F., Jaenisch R., One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153, 910-918 (2013).
20. Cho, S.W., Kim S., Kim J.M., Kim J.S., Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* 31, 230-232 (2013).
21. Wiedenheft, B., Sternberg S.H., Doudna J.A., RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331-338 (2012).
22. Fineran, P.C., Charpentier E., Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* 434, 202-209 (2012).
23. Horvath, P., Barrangou R., CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167-170 (2010).
24. Barrangou, R., Fremaux C., Deveau H., Richards M., Boyaval P., Moineau S., Romero D.A., Horvath P., CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709-1712 (2007).
25. Deltcheva, E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602-607 (2011).
26. Jinek, M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821 (2012).
27. Garneau, J.E., Dupuis M.E., Villion M., Romero D.A., Barrangou R., Boyaval P., Fremaux C., Horvath P., Magadan A.H. *et al.*, The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67-71 (2010).
28. Bolotin, A., Quinquis B., Sorokin A., Ehrlich S.D., Clustered regularly interspaced short palindromic repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151, 2551-2561 (2005).
29. Mojica, F.J., Diez-Villasenor C., Garcia-Martinez J., Soria E., Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60, 174-182 (2005).
30. Pourcel, C., Salvignol G., Vergnaud G., CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151, 653-663 (2005).
31. Brouns, S.J., Jore M.M., Lundgren M., Westra E.R., Slijkhuis R.J., Snijders A.P., Dickman M.J., Makarova K.S., Koonin E.V. *et al.*, Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960-964 (2008).
32. Gasiunas, G., Barrangou R., Horvath P., Siksnys V., Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 109, E2579-2586 (2012).
33. Esvelt, K.M., Mali P., Braff J.L., Moosburner M., Yaung S.J., Church G.M., Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* 10, 1116-1121 (2013).
34. Hou, Z., Zhang Y., Propson N.E., Howden S.E., Chu L.F., Sontheimer E.J., Thomson J.A., Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc Natl Acad Sci U S A* 110, 15644-15649 (2013).
35. Najm, F.J., Strand C., Donovan K.F., Hegde M., Sanson K.R., Vaimberg E.W., Sullender M.E., Hartenian E., Kalani Z. *et al.*, Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. *Nat Biotechnol* 36, 179-189 (2018).
36. Wang, T., Wei J.J., Sabatini D.M., Lander E.S., Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343, 80-84 (2014).
37. Shalem, O., Sanjana N.E., Hartenian E., Shi X., Scott D.A., Mikkelsen T., Heckl D., Ebert B.L., Root D.E., Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84-87 (2014).
38. Zhou, Y., Zhu S., Cai C., Yuan P., Li C., Huang Y., Wei W., High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 509, 487-491 (2014).
39. Doench, J.G., Hartenian E., Graham D.B., Tothova Z., Hegde M., Smith I., Sullender M., Ebert B.L., Xavier R.J. *et al.*, Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* 32, 1262-1267 (2014).

40. Barrangou, R., Birmingham A., Wiemann S., Beijersbergen R.L., Hornung V., Smith A., Advances in CRISPR-Cas9 genome engineering: lessons learned from RNA interference. *Nucleic Acids Res* 43, 3407-3419 (2015).
41. Gilbert, L.A., Horlbeck M.A., Adamson B., Villalta J.E., Chen Y., Whitehead E.H., Guimaraes C., Panning B., Ploegh H.L. *et al.*, Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* 159, 647-661 (2014).
42. Shalem, O., Sanjana N.E., Zhang F., High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet* 16, 299-311 (2015).
43. Kampmann, M., Horlbeck M.A., Chen Y., Tsai J.C., Bassik M.C., Gilbert L.A., Villalta J.E., Kwon S.C., Chang H. *et al.*, Next-generation libraries for robust RNA interference-based genome-wide screens. *Proc Natl Acad Sci U S A* 112, E3384-3391 (2015).
44. Malina, A., Katigbak A., Cencic R., Maiga R.I., Robert F., Miura H., Pelletier J., Adapting CRISPR/Cas9 for functional genomics screens. *Methods Enzymol* 546, 193-213 (2014).
45. Chen, S., Sanjana N.E., Zheng K., Shalem O., Lee K., Shi X., Scott D.A., Song J., Pan J.Q. *et al.*, Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* 160, 1246-1260 (2015).
46. Wang, T., Birsoy K., Hughes N.W., Krupczak K.M., Post Y., Wei J.J., Lander E.S., Sabatini D.M., Identification and characterization of essential genes in the human genome. *Science* 350, 1096-1101 (2015).
47. Evers, B., Jastrzebski K., Heijmans J.P., Grernrum W., Beijersbergen R.L., Bernards R., CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat Biotechnol* 34, 631-633 (2016).
48. Housden, B.E., Valvezan A.J., Kelley C., Sopko R., Hu Y., Roesel C., Lin S., Buckner M., Tao R. *et al.*, Identification of potential drug targets for tuberous sclerosis complex by synthetic screens combining CRISPR-based knockouts with RNAi. *Sci Signal* 8, rs9 (2015).
49. Parnas, O., Jovanovic M., Eisenhaure T.M., Herbst R.H., Dixit A., Ye C.J., Przybylski D., Platt R.J., Tirosh I. *et al.*, A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* 162, 675-686 (2015).
50. Korkmaz, G., Lopes R., Ugalde A.P., Nevedomskaya E., Han R., Myacheva K., Zwart W., Elkon R., Agami R., Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol* 34, 192-198 (2016).
51. Rajagopal, N., Srinivasan S., Kooshesh K., Guo Y., Edwards M.D., Banerjee B., Syed T., Emons B.J., Gifford D.K. *et al.*, High-throughput mapping of regulatory DNA. *Nat Biotechnol* 34, 167-174 (2016).
52. Gilbert, L.A., Larson M.H., Morsut L., Liu Z., Brar G.A., Torres S.E., Stern-Ginossar N., Brandman O., Whitehead E.H. *et al.*, CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442-451 (2013).
53. Hilton, I.B., D'Ipollito A.M., Vockley C.M., Thakore P.I., Crawford G.E., Reddy T.E., Gersbach C.A., Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 33, 510-517 (2015).
54. Kearns, N.A., Pham H., Tabak B., Genga R.M., Silverstein N.J., Garber M., Maehr R., Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods* 12, 401-403 (2015).
55. Qi, L.S., Larson M.H., Gilbert L.A., Doudna J.A., Weissman J.S., Arkin A.P., Lim W.A., Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173-1183 (2013).
56. Chen, B., Gilbert L.A., Cimini B.A., Schnitzbauer J., Zhang W., Li G.W., Park J., Blackburn E.H., Weissman J.S. *et al.*, Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479-1491 (2013).
57. Maeder, M.L., Linder S.J., Cascio V.M., Fu Y., Ho Q.H., Joung J.K., CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* 10, 977-979 (2013).
58. Perez-Pinera, P., Kocak D.D., Vockley C.M., Adler A.F., Kabadi A.M., Polstein L.R., Thakore P.I., Glass K.A., Ousterout D.G. *et al.*, RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods* 10, 973-976 (2013).
59. Konermann, S., Brigham M.D., Trevino A., Hsu P.D., Heidenreich M., Cong L., Platt R.J., Scott D.A., Church G.M. *et al.*, Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 500, 472-476 (2013).

60. Cheng, A.W., Wang H., Yang H., Shi L., Katz Y., Theunissen T.W., Rangarajan S., Shivalila C.S., Dadon D.B. *et al.*, Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res* 23, 1163-1171 (2013).
61. Taylor, J., Woodcock S., A Perspective on the Future of High-Throughput RNAi Screening: Will CRISPR Cut Out the Competition or Can RNAi Help Guide the Way? *J Biomol Screen* 20, 1040-1051 (2015).
62. Hsu, P.D., Scott D.A., Weinstein J.A., Ran F.A., Konermann S., Agarwala V., Li Y., Fine E.J., Wu X. *et al.*, DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31, 827-832 (2013).
63. Fu, Y., Foden J.A., Khayter C., Maeder M.L., Reyon D., Joung J.K., Sander J.D., High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 31, 822-826 (2013).
64. O'Geen, H., Yu A.S., Segal D.J., How specific is CRISPR/Cas9 really? *Curr Opin Chem Biol* 29, 72-78 (2015).
65. Davis, K.M., Pattanayak V., Thompson D.B., Zuris J.A., Liu D.R., Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat Chem Biol* 11, 316-318 (2015).
66. Semenova, E., Jore M.M., Datsenko K.A., Semenova A., Westra E.R., Wanner B., van der Oost J., Brouns S.J., Severinov K., Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* 108, 10098-10103 (2011).
67. Wiedenheft, B., van Duijn E., Bultema J.B., Waghmare S.P., Zhou K., Barendregt A., Westphal W., Heck A.J., Boekema E.J. *et al.*, RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A* 108, 10092-10097 (2011).
68. Pattanayak, V., Lin S., Guilinger J.P., Ma E., Doudna J.A., Liu D.R., High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* 31, 839-843 (2013).
69. Ran, F.A., Hsu P.D., Lin C.Y., Gootenberg J.S., Konermann S., Trevino A.E., Scott D.A., Inoue A., Matoba S. *et al.*, Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154, 1380-1389 (2013).
70. Fu, Y., Sander J.D., Reyon D., Cascio V.M., Joung J.K., Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* 32, 279-284 (2014).
71. Cradick, T.J., Fine E.J., Antico C.J., Bao G., CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res* 41, 9584-9592 (2013).
72. Cho, S.W., Kim S., Kim Y., Kweon J., Kim H.S., Bae S., Kim J.S., Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* 24, 132-141 (2014).
73. Montague, T.G., Cruz J.M., Gagnon J.A., Church G.M., Valen E., CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* 42, W401-407 (2014).
74. Xu, H., Xiao T., Chen C.H., Li W., Meyer C.A., Wu Q., Wu D., Cong L., Zhang F. *et al.*, Sequence determinants of improved CRISPR sgRNA design. *Genome Res* 25, 1147-1157 (2015).
75. Farboud, B., Meyer B.J., Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. *Genetics* 199, 959-971 (2015).
76. Nielsen, S., Yuzenkova Y., Zenkin N., Mechanism of eukaryotic RNA polymerase III transcription termination. *Science* 340, 1577-1580 (2013).
77. Mali, P., Aach J., Stranges P.B., Esvelt K.M., Moosburner M., Kosuri S., Yang L., Church G.M., CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* 31, 833-838 (2013).
78. Shen, B., Zhang W., Zhang J., Zhou J., Wang J., Chen L., Wang L., Hodgkins A., Iyer V. *et al.*, Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods* 11, 399-402 (2014).
79. Slaymaker, I.M., Gao L., Zetsche B., Scott D.A., Yan W.X., Zhang F., Rationally engineered Cas9 nucleases with improved specificity. *Science* 351, 84-88 (2016).
80. Kleinstiver, B.P., Pattanayak V., Prew M.S., Tsai S.Q., Nguyen N.T., Zheng Z., Joung J.K., High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529, 490-495 (2016).
81. Chen, J.S., Dagdas Y.S., Kleinstiver B.P., Welch M.M., Sousa A.A., Harrington L.B., Sternberg S.H., Joung J.K., Yildiz A. *et al.*, Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 550, 407-410 (2017).
82. Casini, A., Olivieri M., Petris G., Montagna C., Reginato G., Maule G., Lorenzin F., Prandi D., Romanel A. *et al.*, A highly specific SpCas9 variant is identified by *in vivo* screening in yeast. *Nat Biotechnol* 36, 265-271 (2018).

83. Thyme, S.B., Akhmetova L., Montague T.G., Valen E., Schier A.F., Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nat Commun* 7, 11750 (2016).
84. Jensen, K.T., Floe L., Petersen T.S., Huang J., Xu F., Bolund L., Luo Y., Lin L., Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. *FEBS Lett* 591, 1892-1901 (2017).
85. Dang, Y., Jia G., Choi J., Ma H., Anaya E., Ye C., Shankar P., Wu H., Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol* 16, 280 (2015).
86. Frit, P., Barboule N., Yuan Y., Gomez D., Calsou P., Alternative end-joining pathway(s): bricolage at DNA breaks. *DNA Repair (Amst)* 17, 81-97 (2014).
87. Harper, J.W., Elledge S.J., The DNA damage response: ten years after. *Mol Cell* 28, 739-745 (2007).
88. Rogakou, E.P., Pilch D.R., Orr A.H., Ivanova V.S., Bonner W.M., DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* 273, 5858-5868 (1998).
89. Dickey, J.S., Redon C.E., Nakamura A.J., Baird B.J., Sedelnikova O.A., Bonner W.M., H2AX: functional roles and potential applications. *Chromosoma* 118, 683-692 (2009).
90. Bekker-Jensen, S., Mailand N., Assembly and function of DNA double-strand break repair foci in mammalian cells. *DNA Repair (Amst)* 9, 1219-1228 (2010).
91. Mailand, N., Bekker-Jensen S., Fastrup H., Melander F., Bartek J., Lukas C., Lukas J., RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. *Cell* 131, 887-900 (2007).
92. Chapman, J.R., Taylor M.R., Boulton S.J., Playing the end game: DNA double-strand break repair pathway choice. *Mol Cell* 47, 497-510 (2012).
93. Symington, L.S., Gautier J., Double-strand break end resection and repair pathway choice. *Annu Rev Genet* 45, 247-271 (2011).
94. Bothmer, A., Robbiani D.F., Feldhahn N., Gazumyan A., Nussenzweig A., Nussenzweig M.C., 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *J Exp Med* 207, 855-865 (2010).
95. Bunting, S.F., Callen E., Wong N., Chen H.T., Polato F., Gunn A., Bothmer A., Feldhahn N., Fernandez-Capetillo O. et al., 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* 141, 243-254 (2010).
96. Uematsu, N., Weterings E., Yano K., Morotomi-Yano K., Jakob B., Taucher-Scholz G., Mari P.O., van Gent D.C., Chen B.P. et al., Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. *J Cell Biol* 177, 219-229 (2007).
97. Mari, P.O., Florea B.I., Persengiev S.P., Verkaik N.S., Bruggenwirth H.T., Modesti M., Gigliamari G., Bezstarosti K., Demmers J.A. et al., Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proc Natl Acad Sci U S A* 103, 18597-18602 (2006).
98. Costantini, S., Woodbine L., Andreoli L., Jeggo P.A., Vindigni A., Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK. *DNA Repair (Amst)* 6, 712-722 (2007).
99. Nick McElhinny, S.A., Snowden C.M., McCarville J., Ramsden D.A., Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol Cell Biol* 20, 2996-3003 (2000).
100. Yano, K., Morotomi-Yano K., Wang S.Y., Uematsu N., Lee K.J., Asaithamby A., Weterings E., Chen D.J., Ku recruits XLF to DNA double-strand breaks. *EMBO Rep* 9, 91-96 (2008).
101. Yoo, S., Dynan W.S., Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. *Nucleic Acids Res* 27, 4679-4686 (1999).
102. Calsou, P., Frit P., Humbert O., Muller C., Chen D.J., Salles B., The DNA-dependent protein kinase catalytic activity regulates DNA end processing by means of Ku entry into DNA. *J Biol Chem* 274, 7848-7856 (1999).
103. Cary, R.B., Peterson S.R., Wang J., Bear D.G., Bradbury E.M., Chen D.J., DNA looping by Ku and the DNA-dependent protein kinase. *Proc Natl Acad Sci U S A* 94, 4267-4272 (1997).
104. Weterings, E., van Gent D.C., The mechanism of non-homologous end-joining: a synopsis of synopsis. *DNA Repair (Amst)* 3, 1425-1435 (2004).
105. DeFazio, L.G., Stansel R.M., Griffith J.D., Chu G., Synapsis of DNA ends by DNA-dependent protein kinase. *EMBO J* 21, 3192-3200 (2002).
106. Weterings, E., Verkaik N.S., Bruggenwirth H.T., Hoeijmakers J.H., van Gent D.C., The role of DNA dependent protein kinase in synapsis of DNA ends. *Nucleic Acids Res* 31, 7238-7246 (2003).

107. Wang, H., Shi L.Z., Wong C.C., Han X., Hwang P.Y., Truong L.N., Zhu Q., Shao Z., Chen D.J. *et al.*, The interaction of CtIP and Nbs1 connects CDK and ATM to regulate HR-mediated double-strand break repair. *PLoS Genet* 9, e1003277 (2013).
108. Dobbs, T.A., Tainer J.A., Lees-Miller S.P., A structural model for regulation of NHEJ by DNA-PKcs autophosphorylation. *DNA Repair (Amst)* 9, 1307-1314 (2010).
109. Ding, Q., Reddy Y.V., Wang W., Woods T., Douglas P., Ramsden D.A., Lees-Miller S.P., Meek K., Autophosphorylation of the catalytic subunit of the DNA-dependent protein kinase is required for efficient end processing during DNA double-strand break repair. *Mol Cell Biol* 23, 5836-5848 (2003).
110. Shibata, A., Conrad S., Birraux J., Geuting V., Barton O., Ismail A., Kakarougkas A., Meek K., Taucher-Scholz G. *et al.*, Factors determining DNA double-strand break repair pathway choice in G2 phase. *EMBO J* 30, 1079-1092 (2011).
111. Gu, J., Lu H., Tsai A.G., Schwarz K., Lieber M.R., Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4-DNA ligase IV complex: influence of terminal DNA sequence. *Nucleic Acids Res* 35, 5755-5762 (2007).
112. Nicolette, M.L., Lee K., Guo Z., Rani M., Chow J.M., Lee S.E., Paull T.T., Mre11-Rad50-Xrs2 and Sae2 promote 5' strand resection of DNA double-strand breaks. *Nat Struct Mol Biol* 17, 1478-1485 (2010).
113. Zhu, Z., Chung W.H., Shim E.Y., Lee S.E., Ira G., Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell* 134, 981-994 (2008).
114. Mimitou, E.P., Symington L.S., Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature* 455, 770-774 (2008).
115. Huertas, P., Jackson S.P., Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *J Biol Chem* 284, 9558-9565 (2009).
116. Chen, L., Nievera C.J., Lee A.Y., Wu X., Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair. *J Biol Chem* 283, 7713-7720 (2008).
117. Yuan, J., Chen J., N terminus of CtIP is critical for homologous recombination-mediated double-strand break repair. *J Biol Chem* 284, 31746-31752 (2009).
118. Xia, B., Sheng Q., Nakanishi K., Ohashi A., Wu J., Christ N., Liu X., Jasin M., Couch F.J. *et al.*, Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Mol Cell* 22, 719-729 (2006).
119. Zhang, F., Ma J., Wu J., Ye L., Cai H., Xia B., Yu X., PALB2 links BRCA1 and BRCA2 in the DNA-damage response. *Curr Biol* 19, 524-529 (2009).
120. Zelensky, A., Kanaar R., Wyman C., Mediators of homologous DNA pairing. *Cold Spring Harb Perspect Biol* 6, a016451 (2014).
121. DiBiase, S.J., Zeng Z.C., Chen R., Hyslop T., Curran W.J., Jr., Iliakis G., DNA-dependent protein kinase stimulates an independently active, nonhomologous, end-joining apparatus. *Cancer Res* 60, 1245-1253 (2000).
122. Iliakis, G., Wang H., Perrault A.R., Boecker W., Rosidi B., Windhofer F., Wu W., Guan J., Terzoudi G. *et al.*, Mechanisms of DNA double strand break repair and chromosome aberration formation. *Cytogenet Genome Res* 104, 14-20 (2004).
123. Wang, H., Perrault A.R., Takeda Y., Qin W., Wang H., Iliakis G., Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic Acids Res* 31, 5377-5388 (2003).
124. Kabotyanski, E.B., Gomelsky L., Han J.O., Stamato T.D., Roth D.B., Double-strand break repair in Ku86- and XRCC4-deficient cells. *Nucleic Acids Res* 26, 5333-5342 (1998).
125. McVey, M., Lee S.E., MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* 24, 529-538 (2008).
126. Nussenzweig, A., Nussenzweig M.C., A backup DNA repair pathway moves to the forefront. *Cell* 131, 223-225 (2007).
127. Iliakis, G., Backup pathways of NHEJ in cells of higher eukaryotes: cell cycle dependence. *Radiother Oncol* 92, 310-315 (2009).
128. Mao, Z., Bozzella M., Seluanov A., Gorbunova V., Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst)* 7, 1765-1771 (2008).
129. Goodarzi, A.A., Noon A.T., Deckbar D., Ziv Y., Shiloh Y., Lohrich M., Jeggo P.A., ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. *Mol Cell* 31, 167-177 (2008).
130. Noon, A.T., Shibata A., Rief N., Lohrich M., Stewart G.S., Jeggo P.A., Goodarzi A.A., 53BP1-dependent robust localized KAP-1 phosphorylation is essential for heterochromatic DNA double-strand break repair. *Nat Cell Biol* 12, 177-184 (2010).

131. Ziv, Y., Bielopolski D., Galanty Y., Lukas C., Taya Y., Schultz D.C., Lukas J., Bekker-Jensen S., Bartek J. *et al.*, Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. *Nat Cell Biol* 8, 870-876 (2006).
132. Ayoub, N., Jeyasekharan A.D., Bernal J.A., Venkitaraman A.R., HP1-beta mobilization promotes chromatin changes that initiate the DNA damage response. *Nature* 453, 682-686 (2008).
133. Riballo, E., Kuhne M., Rief N., Doherty A., Smith G.C., Recio M.J., Reis C., Dahm K., Fricke A. *et al.*, A pathway of double-strand break rejoining dependent upon ATM, Artemis, and proteins locating to gamma-H2AX foci. *Mol Cell* 16, 715-724 (2004).
134. Kim, J.A., Kruhlak M., Dotiwala F., Nussenzweig A., Haber J.E., Heterochromatin is refractory to gamma-H2AX modification in yeast and mammals. *J Cell Biol* 178, 209-218 (2007).
135. Murga, M., Jaco I., Fan Y., Soria R., Martinez-Pastor B., Cuadrado M., Yang S.M., Blasco M.A., Skoultchi A.I. *et al.*, Global chromatin compaction limits the strength of the DNA damage response. *J Cell Biol* 178, 1101-1108 (2007).
136. Janssen, A., Breuer G.A., Brinkman E.K., van der Meulen A.I., Borden S.V., van Steensel B., Bindra R.S., LaRocque J.R., Karpen G.H., A single double-strand break system reveals repair dynamics and mechanisms in heterochromatin and euchromatin. *Genes Dev* 30, 1645-1657 (2016).
137. Chiolo, I., Minoda A., Colmenares S.U., Polyzos A., Costes S.V., Karpen G.H., Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. *Cell* 144, 732-744 (2011).
138. Cremer, T., Cremer M., Chromosome territories. *Cold Spring Harb Perspect Biol* 2, a003889 (2010).
139. Nagai, S., Dubrana K., Tsai-Pflugfelder M., Davidson M.B., Roberts T.M., Brown G.W., Varela E., Hediger F., Gasser S.M. *et al.*, Functional targeting of DNA damage to a nuclear pore-associated SUMO-dependent ubiquitin ligase. *Science* 322, 597-602 (2008).
140. Jakob, B., Splinter J., Conrad S., Voss K.O., Zink D., Durante M., Lobrich M., Taucher-Scholz G., DNA double-strand breaks in heterochromatin elicit fast repair protein recruitment, histone H2AX phosphorylation and relocation to euchromatin. *Nucleic Acids Res* 39, 6489-6499 (2011).
141. Kruhlak, M.J., Celeste A., Dellaire G., Fernandez-Capetillo O., Muller W.G., McNally J.G., Bazett-Jones D.P., Nussenzweig A., Changes in chromatin structure and mobility in living cells at sites of DNA double-strand breaks. *J Cell Biol* 172, 823-834 (2006).
142. Falk, M., Lukasova E., Gabrielova B., Ondrej V., Kozubek S., Chromatin dynamics during DSB repair. *Biochim Biophys Acta* 1773, 1534-1545 (2007).
143. Soutoglou, E., Dorn J.F., Sengupta K., Jasin M., Nussenzweig A., Ried T., Danuser G., Misteli T., Positional stability of single double-strand breaks in mammalian cells. *Nat Cell Biol* 9, 675-682 (2007).
144. Jakob, B., Splinter J., Durante M., Taucher-Scholz G., Live cell microscopy analysis of radiation-induced DNA double-strand break motion. *Proc Natl Acad Sci U S A* 106, 3172-3177 (2009).
145. Torres-Rosell, J., Sunjevaric I., De Piccoli G., Sacher M., Eckert-Boulet N., Reid R., Jentsch S., Rothstein R., Aragon L. *et al.*, The Smc5-Smc6 complex and SUMO modification of Rad52 regulates recombinational repair at the ribosomal gene locus. *Nat Cell Biol* 9, 923-931 (2007).
146. Ciabrelli, F., Cavalli G., Chromatin-driven behavior of topologically associating domains. *J Mol Biol* 427, 608-625 (2015).
147. Almouzni, G., Probst A.V., Heterochromatin maintenance and establishment: lessons from the mouse pericentromere. *Nucleus* 2, 332-338 (2011).
148. Fadloun, A., Eid A., Torres-Padilla M.E., Mechanisms and dynamics of heterochromatin formation during mammalian development: closed paths and open questions. *Curr Top Dev Biol* 104, 1-45 (2013).
149. Politz, J.C., Scalzo D., Groudine M., Something silent this way forms: the functional organization of the repressive nuclear compartment. *Annu Rev Cell Dev Biol* 29, 241-270 (2013).
150. Kind, J., Pagie L., Ortazobkoyun H., Boyle S., de Vries S.S., Janssen H., Amendola M., Nolen L.D., Bickmore W.A. *et al.*, Single-cell dynamics of genome-nuclear lamina interactions. *Cell* 153, 178-192 (2013).
151. Lemaitre, C., Grabarz A., Tsouroula K., Andronov L., Furst A., Pankotai T., Heyer V., Rogier M., Attwood K.M. *et al.*, Nuclear position dictates DNA repair pathway choice. *Genes Dev* 28, 2450-2463 (2014).

152. Tsouroula, K., Furst A., Rogier M., Heyer V., Maglott-Roth A., Ferrand A., Reina-San-Martin B., Soutoglou E., Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. *Mol Cell* 63, 293-305 (2016).
153. van Sluis, M., McStay B., A localized nucleolar DNA damage response facilitates recruitment of the homology-directed repair machinery independent of cell cycle stage. *Genes Dev* 29, 1151-1163 (2015).
154. Chan, F.L., Wong L.H., Transcription in the maintenance of centromere chromatin identity. *Nucleic Acids Res* 40, 11178-11188 (2012).
155. Aymard, F., Bugler B., Schmidt C.K., Guillou E., Caron P., Briois S., Iacovoni J.S., Daburon V., Miller K.M. et al., Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat Struct Mol Biol* 21, 366-374 (2014).
156. Daugaard, M., Baude A., Fugger K., Povlsen L.K., Beck H., Sorensen C.S., Petersen N.H., Sorensen P.H., Lukas C. et al., LEDGF (p75) promotes DNA-end resection and homologous recombination. *Nat Struct Mol Biol* 19, 803-810 (2012).
157. Harding, S.M., Boiarsky J.A., Greenberg R.A., ATM Dependent Silencing Links Nucleolar Chromatin Reorganization to DNA Damage Recognition. *Cell Rep* 13, 251-259 (2015).
158. Warmerdam, D.O., van den Berg J., Medema R.H., Breaks in the 45S rDNA Lead to Recombination-Mediated Loss of Repeats. *Cell Rep* 14, 2519-2527 (2016).
159. Janicki, S.M., Tsukamoto T., Salghetti S.E., Tansey W.P., Sachidanandam R., Prasanth K.V., Ried T., Shav-Tal Y., Bertrand E. et al., From silencing to gene expression: real-time analysis in single cells. *Cell* 116, 683-698 (2004).
160. Shanbhag, N.M., Rafalska-Metcalf I.U., Balane-Bolivar C., Janicki S.M., Greenberg R.A., ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks. *Cell* 141, 970-981 (2010).
161. Kakarougkas, A., Ismail A., Chambers A.L., Riballo E., Herbert A.D., Kunzel J., Lobrich M., Jeggo P.A., Downs J.A., Requirement for PBAF in transcriptional repression and repair at DNA breaks in actively transcribed regions of chromatin. *Mol Cell* 55, 723-732 (2014).
162. Pfister, S.X., Ahrabi S., Zalmas L.P., Sarkar S., Aymard F., Bachrati C.Z., Helleday T., Legube G., La Thangue N.B. et al., SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. *Cell Rep* 7, 2006-2018 (2014).
163. Aymard, F., Aguirrebengoa M., Guillou E., Javierre B.M., Bugler B., Arnould C., Rocher V., Iacovoni J.S., Biernacka A. et al., Genome-wide mapping of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes. *Nat Struct Mol Biol* 24, 353-361 (2017).
164. Marnef, A., Legube G., Organizing DNA repair in the nucleus: DSBs hit the road. *Curr Opin Cell Biol* 46, 1-8 (2017).
165. Lieber, M.R., The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* 79, 181-211 (2010).
166. Rudin, N., Sugarman E., Haber J.E., Genetic and physical analysis of double-strand break repair and recombination in *Saccharomyces cerevisiae*. *Genetics* 122, 519-534 (1989).
167. Choulika, A., Perrin A., Dujon B., Nicolas J.F., Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. *Mol Cell Biol* 15, 1968-1973 (1995).
168. Wang, Y., Xu C., Du L.Q., Cao J., Liu J.X., Su X., Zhao H., Fan F.Y., Wang B. et al., Evaluation of the comet assay for assessing the dose-response relationship of DNA damage induced by ionizing radiation. *Int J Mol Sci* 14, 22449-22461 (2013).
169. Forment, J.V., Walker R.V., Jackson S.P., A high-throughput, flow cytometry-based method to quantify DNA-end resection in mammalian cells. *Cytometry A* 81, 922-928 (2012).
170. Seo, J., Kim S.C., Lee H.S., Kim J.K., Shon H.J., Salleh N.L., Desai K.V., Lee J.H., Kang E.S. et al., Genome-wide profiles of H2AX and gamma-H2AX differentiate endogenous and exogenous DNA damage hotspots in human cells. *Nucleic Acids Res* 40, 5965-5974 (2012).
171. Bocker, W., Iliakis G., Computational Methods for analysis of foci: validation for radiation-induced gamma-H2AX foci in human cells. *Radiat Res* 165, 113-124 (2006).
172. Soong, C.P., Breuer G.A., Hannon R.A., Kim S.D., Salem A.F., Wang G., Yu R., Carriero N.J., Bjornson R. et al., Development of a novel method to create double-strand break repair fingerprints using next-generation sequencing. *DNA Repair (Amst)* 26, 44-53 (2015).

173. Canela, A., Sridharan S., Sciascia N., Tubbs A., Meltzer P., Sleckman B.P., Nussenzweig A., DNA Breaks and End Resection Measured Genome-wide by End Sequencing. *Mol Cell* 63, 898-911 (2016).
174. Crosetto, N., Mitra A., Silva M.J., Bienko M., Dojer N., Wang Q., Karaca E., Chiarle R., Skrzypczak M. *et al.*, Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat Methods* 10, 361-365 (2013).
175. Zhou, Z.X., Zhang M.J., Peng X., Takayama Y., Xu X.Y., Huang L.Z., Du L.L., Mapping genomic hotspots of DNA damage by a single-strand-DNA-compatible and strand-specific ChIP-seq method. *Genome Res* 23, 705-715 (2013).
176. Hendel, A., Kildebeck E.J., Fine E.J., Clark J., Punjya N., Sebastiano V., Bao G., Porteus M.H., Quantifying genome-editing outcomes at endogenous loci with SMRT sequencing. *Cell Rep* 7, 293-305 (2014).
177. Boel, A., Steyaert W., De Rocker N., Menten B., Callewaert B., De Paepe A., Coucke P., Willaert A., BATCH-GE: Batch analysis of Next-Generation Sequencing data for genome editing assessment. *Sci Rep* 6, 30330 (2016).
178. Guell, M., Yang L., Church G.M., Genome editing assessment using CRISPR Genome Analyzer (CRISPR-GA). *Bioinformatics* 30, 2968-2970 (2014).
179. Pinello, L., Canver M.C., Hoban M.D., Orkin S.H., Kohn D.B., Bauer D.E., Yuan G.C., Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol* 34, 695-697 (2016).
180. van Overbeek, M., Capurso D., Carter M.M., Thompson M.S., Frias E., Russ C., Reece-Hoyes J.S., Nye C., Gradia S. *et al.*, DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol Cell* 63, 633-646 (2016).
181. Bothmer, A., Phadke T., Barrera L.A., Margulies C.M., Lee C.S., Buquicchio F., Moss S., Abdulkarim H.S., Selleck W. *et al.*, Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat Commun* 8, 13905 (2017).
182. Friend, L.E., Jasin M., Krawczyk P.M., Assaying break and nick-induced homologous recombination in mammalian cells using the DR-GFP reporter and Cas9 nucleases. *Methods Enzymol* 546, 175-191 (2014).
183. Chen, X., Rinsma M., Janssen J.M., Liu J., Maggio I., Goncalves M.A., Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Res* 44, 6482-6492 (2016).
184. Daer, R.M., Cutts J.P., Brafman D.A., Haynes K.A., The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells. *ACS Synth Biol* 6, 428-438 (2017).
185. Chari, R., Mali P., Moosburner M., Church G.M., Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat Methods* 12, 823-826 (2015).
186. Chen, Y., Zeng S., Hu R., Wang X., Huang W., Liu J., Wang L., Liu G., Cao Y. *et al.*, Using local chromatin structure to improve CRISPR/Cas9 efficiency in zebrafish. *PLoS One* 12, e0182528 (2017).
187. Knight, S.C., Xie L., Deng W., Guglielmi B., Witkowsky L.B., Bosanac L., Zhang E.T., El Beheiry M., Masson J.B. *et al.*, Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* 350, 823-826 (2015).
188. Isaac, R.S., Jiang F., Doudna J.A., Lim W.A., Narlikar G.J., Almeida R., Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *Elife* 5, (2016).
189. Horlbeck, M.A., Witkowsky L.B., Guglielmi B., Replogle J.M., Gilbert L.A., Villalta J.E., Torigoe S.E., Tjian R., Weissman J.S., Nucleosomes impede Cas9 access to DNA *in vivo* and *in vitro*. *Elife* 5, (2016).
190. Lopes, R., Korkmaz G., Agami R., Applying CRISPR-Cas9 tools to identify and characterize transcriptional enhancers. *Nat Rev Mol Cell Biol* 17, 597-604 (2016).
191. Gaspar-Maia, A., Alajem A., Meshorer E., Ramalho-Santos M., Open chromatin in pluripotency and reprogramming. *Nat Rev Mol Cell Biol* 12, 36-47 (2011).
192. Shiloh, Y., The ATM-mediated DNA-damage response: taking shape. *Trends Biochem Sci* 31, 402-410 (2006).
193. Gatti, M., Pinato S., Maiolica A., Rocchio F., Prato M.G., Aebbersold R., Penengo L., RNF168 promotes noncanonical K27 ubiquitination to signal DNA damage. *Cell Rep* 10, 226-238 (2015).
194. Iliakis, G., Murmann T., Soni A., Alternative end-joining repair pathways are the ultimate backup for abrogated classical non-homologous end-joining and homologous recombination repair: Implications for the formation of chromosome translocations. *Mutat Res Genet Toxicol Environ Mutagen* 793, 166-175 (2015).

195. Razin, S.V., Ulianov S.V., Gene functioning and storage within a folded genome. *Cell Mol Biol Lett* 22, 18 (2017).
196. Ecco, G., Imbeault M., Trono D., KRAB zinc finger proteins. *Development* 144, 2719-2729 (2017).
197. Fischle, W., Tip60-ing the balance in DSB repair. *Nat Cell Biol* 11, 1279-1281 (2009).
198. Stratigi, K., Chatzidoukaki O., Garinis G.A., DNA damage-induced inflammation and nuclear architecture. *Mech Ageing Dev* 165, 17-26 (2017).

Chapter 2

EASY QUANTITATIVE ASSESSMENT OF GENOME
EDITING BY SEQUENCE TRACE DECOMPOSITION

Eva K. Brinkman, Tao Chen, Mario Amendola and Bas van Steensel

Nucleic Acids Research, 42, e168, 2014

ABSTRACT

The efficacy and the mutation spectrum of genome editing methods can vary substantially depending on the targeted sequence. A simple, quick assay to accurately characterize and quantify the induced mutations is therefore needed. Here we present TIDE, a method for this purpose that requires only a pair of PCR reactions and two standard capillary sequencing runs. The sequence traces are then analysed by a specially developed decomposition algorithm that identifies the major induced mutations in the projected editing site and accurately determines their frequency in a cell population. This method is cost-effective and quick, and it provides much more detailed information than current enzyme-based assays. An interactive web tool for automated decomposition of the sequence traces is available. TIDE greatly facilitates the testing and rational design of genome editing strategies.

INTRODUCTION

Genome editing tools such as TAL effector nucleases, zinc finger nucleases and RNA-guided endonucleases (RGENs) enable targeted mutagenesis of a selected DNA sequence in genomes of many species (1, 2). In each of these methods, introduction of an endonuclease with programmable sequence specificity into a pool of cells leads to a precisely defined DNA double-strand break (DSB), which, when repaired by non-homologous end joining, results in a mixture of unaltered and mutated DNA. The latter consists primarily of a diversity of short deletions and (more rarely) short insertions that are centered round the break site (3-5). Cells with a mutation of interest then need to be cloned in order to establish a stable mutant line.

In order to implement this approach, it is usually necessary to test the efficacy of the programmable nuclease, which can vary dramatically depending on the sequence that is targeted. For example, with RGENs one typically needs to test several single-guide RNAs (sgRNAs) that are predicted to target the endonuclease to a gene of interest (6). Commonly used assays to verify the efficacy of programmable nucleases are the enzymatic Surveyor and T7 endonuclease I cleavage assays (7, 8), which detect small sequence changes. These methods are, however, semi-quantitative and suffer from high background signals when sequence polymorphisms are present. More importantly, these enzymatic assays do not provide insight into the nature and the diversity of the mutations that are introduced. This information is particularly useful if one needs to establish a clonal cell line with a specific editing outcome, such as a defined deletion size that causes a specific frame shift in an open reading frame of interest, or that generates a subtle sequence change in a regulatory element. To determine the frequency of the desired editing event in the pool of cells, one can amplify the targeted genomic region by PCR, clone individual DNA molecules in a bacterial vector, and analyse 50-100 clones by sequencing. This approach is labour-intensive, time-consuming and relatively costly. Next generation sequencing around the induced break site (9) is a powerful alternative, but is also expensive and usually takes several weeks in most research environments.

Here, we present a simple, rapid and cost-effective strategy that accurately quantifies the editing efficacy and simultaneously identifies the predominant types of insertions and deletions (indels) in the targeted pool of cells. The method, named TIDE (Tracking of Indels by DEcomposition), requires only two parallel PCR reactions followed by a pair of standard capillary sequencing analyses. The two resulting sequencing traces are then analysed using specially designed software that we provide as a simple web tool (<http://tide.nki.nl>).

RESULTS

The TIDE method

In the first step of TIDE, a stretch of about 500-1500 basepairs around the editing site is PCR amplified from genomic DNA isolated from the cell pool that was treated with the targeted nuclease. A parallel PCR amplifies the same stretch of DNA from a control cell pool lacking the nuclease or sgRNA. Both PCR products are then directly subjected to conventional capillary ('Sanger') sequencing, a basic technology that is available in most laboratories. In the DNA sample from the cells expressing the targeted nuclease, the sequence trace after the break site consists of a mixture of signals derived from unmodified DNA and sequences that are each shifted by a different number of nucleotides due to insertions and deletions (FIGURE 1A).

Based on the quantitative sequence trace data, the TIDE software first visualizes the proportion of aberrant base signals along the sequence traces in an intuitive graph (FIGURE 1B, STEP 1). This enables the user to visually inspect the sequence deviation caused by the targeted nuclease, and verify that the break site is located as expected.

Subsequently, the TIDE software decomposes the composite sequence trace into its individual components by means of multivariate non-negative linear modelling, with the control sequence trace serving as a template to model the individual indel components. This decomposition results in an estimate of the relative abundance of every possible indel within a chosen size range (FIGURE 1B, STEP 2). The software provides the R^2 value as a goodness-of-fit measure, and calculates the statistical significance for each indel. In the TIDE webtool, the sequence segment used for decomposition can be interactively adjusted, which is helpful in case the sequence traces are locally of poor quality.

Finally, for insertions of a single basepair, the relative frequency of the four possible bases is deduced from the model, which is of interest if one wishes to obtain a +1 mutation of a particular sequence (FIGURE 1B, STEP 3). For longer insertions this base-calling is computationally more complicated and currently not implemented.

In vitro proof of principle

In order to test our approach, we first constructed a series of artificial samples consisting of wild-type DNA mixed with DNA carrying various indels in a broad range of relative concentrations. We then performed standard capillary sequencing and fed the resulting data into the TIDE algorithm. The constituents of the mixes could be identified and quantified with great accuracy. In a mixture of wild-type and +1 insertion DNA our algorithm was able to detect the insertion quantitatively with a sensitivity down to ~2.5%

(FIGURE 2A; SUPPLEMENTARY FIGURE S1), and generally predicted the correct base (FIGURE 2A, INSET; SUPPLEMENTARY FIGURE S1). Even a -15 deletion could be reliably detected when mixed 1:10 with wild-type DNA. (FIGURE 2B; SUPPLEMENTARY FIGURE S2A). All constituents in mixtures of wild-type DNA with either five or eight different indels were identified by TIDE (FIGURE 2B; SUPPLEMENTARY FIGURE S2B-C). These *in vitro* simulations show that sequence trace decomposition can accurately identify and quantify the constituent indels in a mixture.

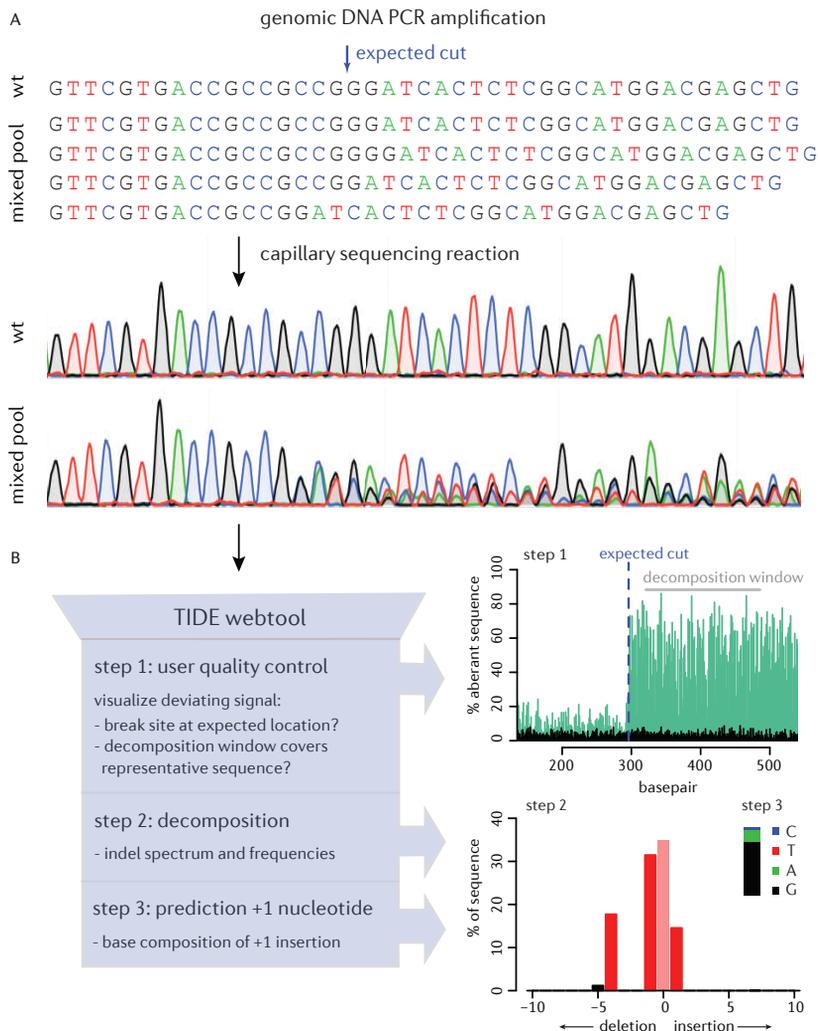


Figure 1: Assessment of genome editing by sequence trace decomposition. (A) Due to imperfect repair after cutting by a targeted nuclease, the DNA in the cell pool consists of a mixture of indels, which yields a composite sequence trace after the break site. (B) Overview of TIDE algorithm and output, which consists of three main steps: 1) Visualization of aberrant sequence signal in control (black) and treated sample (green), the expected break site (vertical dotted line) and the region used for decomposition (grey bar); 2) Decomposition yielding the spectrum of indels and their frequencies; 3) Inference of the base composition of +1 insertions. See main text and <http://tide.nki.nl> for explanation.

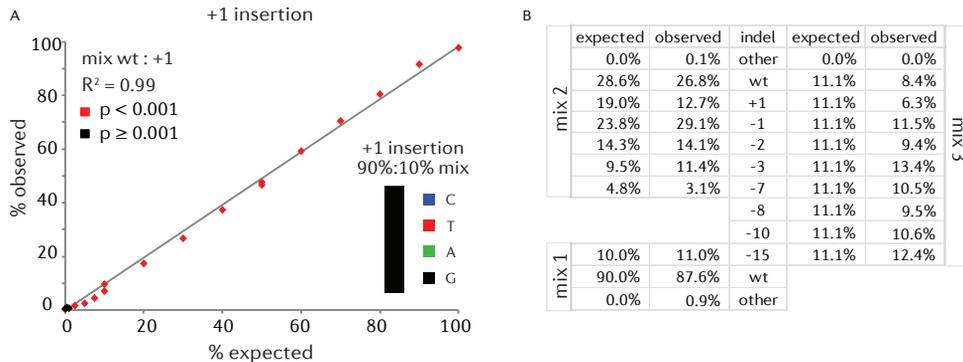


Figure 2: Proof-of-principle of TIDE. (A) A DNA fragment carrying a +1 insertion was mixed in indicated relative amounts with a corresponding wild-type DNA fragment (horizontal axis), after which the +1 insertion content was determined by TIDE (vertical axis) using the default search for indels with a size range of 0..10. Inset: relative abundance of the inserted nucleotide in a wt, +1 mix (90%:10%). See SUPPLEMENTARY FIGURE S1 for the complete decomposition results. (B) TIDE decomposition of various complex mixtures of wild-type DNA with DNA carrying a range of indels. See also SUPPLEMENTARY FIGURE S2A-C.

Application of TIDE to CRISPR/Cas9 edited DNA sequences

We then tested this approach on a pool of human K562 cells carrying random integrations of the green fluorescent protein (GFP) reporter gene. We transfected these cells with the RGEN Cas9 together with a sgRNA designed to target the GFP gene, or without the sgRNA as a control. TIDE determined that 34.2% of GFP sequences in the sgRNA-treated cell pool carried an indel, with 23.2% being a -1 deletion (FIGURE 3A). The composite sequence started at the expected break site (FIGURE 3B), confirming correct targeting by the sgRNA. The +1 insertions consisted almost exclusively of a G nucleotide on the forward strand (FIGURE 3A, INSET), indicating that the choice of the inserted nucleotide is non-random. Sequencing of the opposite DNA strand yielded virtually identical quantitative results (SUPPLEMENTARY FIGURE S3A), indicating that the assay is highly robust.

To independently validate these results, we cloned and sequenced 84 individual DNA molecules from the same PCR product. This revealed a similar spectrum of indels, in which the frequency of each indel is generally not significantly different from the TIDE calculations (FIGURE 3C). However, some larger indels with frequencies below ~2% were not significantly detected by TIDE. All +1 insertions of the individual clones consisted of a G nucleotide in the forward strand, confirming the computational inference.

All the significant mutations found by TIDE are predicted to lead to frame shifts yielding a non-functional truncated GFP protein. In agreement with this finding FACS analysis shows a 38.4% loss of GFP-positive cells in the pool of cells expressing the sgRNA compared to the pool lacking the sgRNA (FIGURE 3D). Thus, the calculation of gene editing efficiency by sequence trace decomposition is in close agreement with the observed frequency of the phenotype.

Finally, we used our approach to test different sgRNAs designed to target endogenous genes in human or *Drosophila* cell lines (FIGURE 3E-H). Interestingly, the results show that different sgRNAs resulted in distinct indel spectra. For example, a sgRNA targeting the NDC1 gene produces roughly equal amounts of +1 and -1 indels, while a sgRNA targeting the LBR gene produces mainly +1 insertions and a few bigger deletions including a more pronounced -7 deletion. Overall, small indels (+1 and -1) appear to be the most common mutations induced by Cas9, which is in agreement with other studies (3-5). Individual indels were detected at estimated frequencies down to ~1%. Again, the results were nearly identical when the opposite strand was sequenced (SUPPLEMENTARY FIGURE S3B-C). In addition, +1 insertions were typically dominated by one specific nucleotide, which was identical to one of the two terminal nucleotides of the break site. Which of the two neighbouring nucleotides is duplicated appears to vary. For example, one sgRNA guide targeting NDC1 resulted in different +1 insertions in K562 and RPE cells (FIGURE 3E-F). How the DSB repair machinery chooses the inserted bases remains to be further elucidated.

DISCUSSION

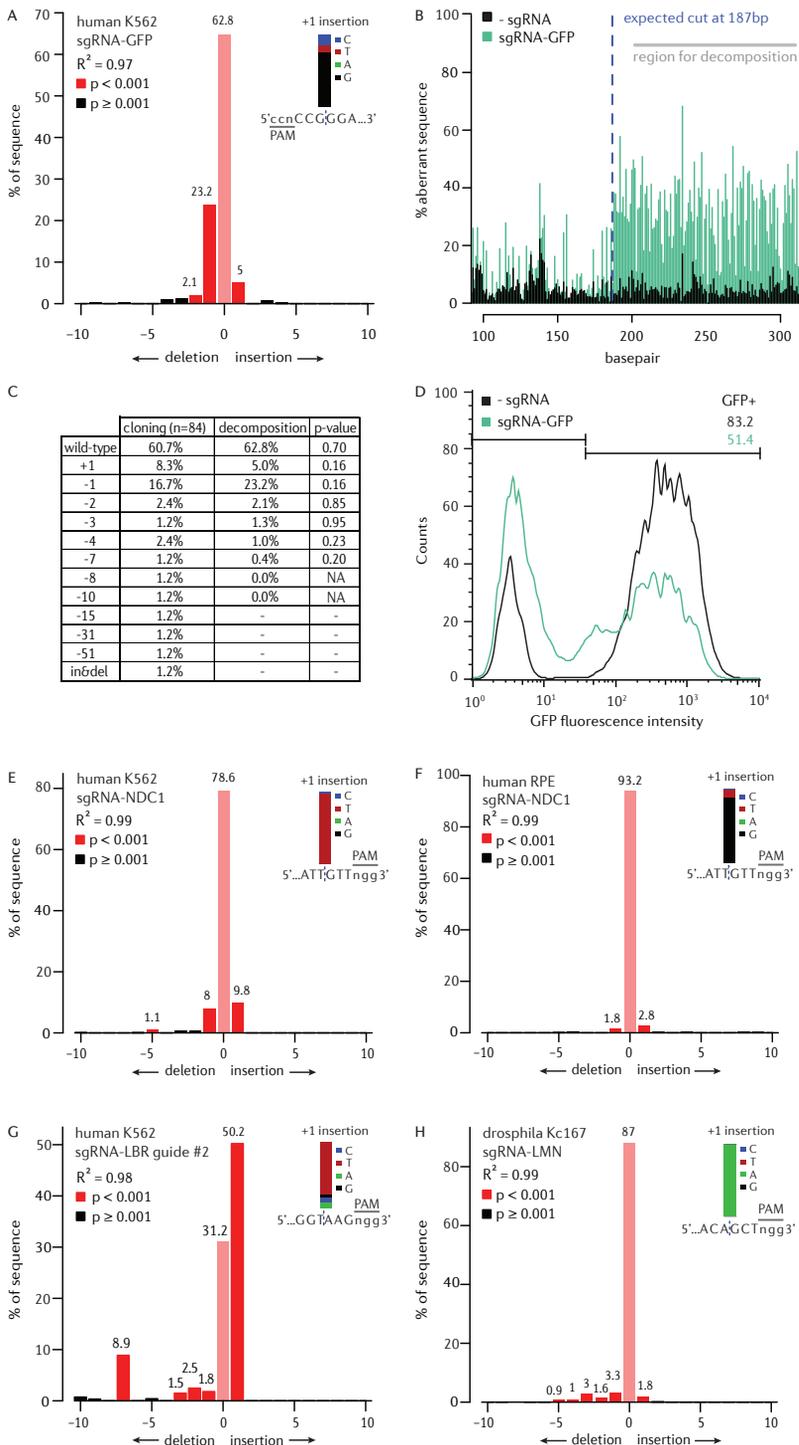
Advantages of TIDE

Genome editing techniques like CRISPR, TALENs and ZFPs are now widely used to alter specific sequences in genomes of cultured cells. However, the efficacy and the spectrum of mutations vary greatly depending on the RGEN target site and cells used (10, 11 and this report). Hence, a fast and cost-effective approach to determine the efficacy of RGENs is essential to optimize the genome editing strategies. TIDE takes advantage of the fact that non-homologous end joining repair of DNA DSBs leaves an indel at the break site. By decomposition of the quantitative sequence trace data, the TIDE software identifies and quantifies these indels. This allows researchers to quickly determine the efficiency of the RGENs and rationally estimate the number of cell clones that must be picked and screened in order to obtain a clonal line with a particular indel of interest.

Figure 3: Application of TIDE to in vivo edited DNA sequences. (A-D) A pool of human K562 cells expressing GFP treated with Cas9 alone (control) and cells treated with Cas9 and a GFP targeting sgRNA (sample) were analysed by: TIDE (A-B), sequence analysis of 84 cloned DNA fragments (C) and flow cytometry (D). (A) Indel spectrum determined by TIDE. Inset shows the estimated composition of the inserted base for the +1 insertion. (B) Aberrant nucleotide signal of the sample (green) compared to that of the control (black). Blue dotted line indicates the expected cutting site. Grey horizontal bar shows the region used for decomposition. (C) Comparison of indel occurrences in cloned DNA fragments ($n = 84$) to frequencies estimated by TIDE, with p -values according to Pearson's chi-squared test. Decomposition was limited to indels of size 0..10, hence larger indels could not be detected. (D) Distributions of GFP fluorescence intensities of Cas9 and Cas9+sgRNA treated cells, measured by flow cytometry. The percentage of GFP-positive cells is indicated in the top right corner within indicated histogram gate. (E-H) TIDE analysis of various endogenous genes (NDC1, LBR, LMN) targeted with RGENs in human cell lines (K562, RPE) and in a *Drosophila* cell line (Kc167). Insets: prediction of the inserted base for +1 insertions.

Easy quantitative assessment of genome editing by sequence trace decomposition

2



(legend on previous page)

Attractive features of TIDE are the low costs and the fact that it requires only two standard PCRs and two capillary sequencing runs. Hands-on time is therefore limited, and results can be obtained in one or two days. We found that TIDE is capable of detecting insertions and deletions with a sensitivity up to ~1-2% across various target regions in a pool of cells. The method is highly robust, as indicated by the strong correspondence between the decomposition results from forward and reverse sequence traces. Good agreement of TIDE results with the sequence composition of a set of individually cloned DNA molecules underscores the reliability.

Comparison to other methods

Several other methods have been used to assess genome editing efficacies. Cloning and sequencing of 50-100 individual DNA molecules provides an accurate characterization of the indel spectrum, but this is obviously more labor-intensive and 25-50 times more expensive than TIDE. Next-generation sequencing (NGS) of bulk PCR products, followed by analysis using software such as CRISPR-GA (9) provides a highly detailed estimate of the indel spectrum, but this method is only cost-effective if a large number of samples are multiplexed; moreover, in most research institutes NGS takes several weeks.

For the creation of gene knockouts, homology based integration of a selectable marker gene at the induced break site can facilitate the isolation of the desired clonal line (12). This approach is time-consuming and still requires moderately effective DSB induction and hence prior characterization of the efficacy of the genome editing tool is desirable. Finally, assays that employ mismatch-detecting enzymes (7, 8) require a similar amount of hands-on time as TIDE, but only estimate the overall mutation frequency and do not characterize the spectrum of indels. Moreover, polymorphisms in the vicinity of the break site are expected to cause high background signals in these enzymatic assays, unlike TIDE. Thus, TIDE offers a cost-effective and accurate alternative strategy for the rapid testing of genome editing efficacy.

Potential limitations of TIDE

Naturally, the reliability of TIDE depends on the purity of the PCR products and the quality of the sequence reads. Decomposition results with a low R^2 must be interpreted with caution. As a rule of thumb, we recommend to aim for a background signal of aberrant sequences before the break site $<10\%$ (both control and test sample), and $R^2 > 0.9$ for the decomposition result. Sequencing of the opposite strand is recommended to confirm the results. Highly repetitive sequences around the target site may in some instances hamper the decomposition. Incorrect alignments can be detected when the quality plot shows an aberrant sequence signal that is not located at the expected break site. The decomposition window can be adjusted in order to avoid repetitive regions.

Additional applications

While we demonstrated the utility here for Cas9-based mutagenesis, TIDE should also be applicable to other genome editing tools that are based on targeted DSB induction (1, 2). Because of the quantitative nature of the results, TIDE may also be used to study mechanisms of DSB repair. For example, we observed that the spectrum of insertions and deletions varies among various target sites and cell types. It will be interesting to employ TIDE to investigate how DSB repair is affected by sequence context or local chromatin environment. Another application would be to determine differential effects of various indels at a gene of interest on cellular fitness. With TIDE, the relative abundance of indels can be followed over time in a growing population of cells treated with RGENs. An increase or decrease of wild-type sequence or particular indel over time could be an indication that the targeted sequence is lethal. In summary, TIDE will be a valuable tool for a broad diversity of research involving genome editing methods.

MATERIAL & METHODS

Cell culture and transfection

K562 cells (ATCC) were cultured in RPMI 1640 (Gibco) supplemented with 10% fetal bovine serum (FBS, HyClone®), 1% penicillin/streptomycin. A pool of K562 cells stably expressing GFP was generated by transduction with the lentivirus construct pCCLsin.PPT.hPGK.GFP.pre (13). Because the lentiviral construct integrates randomly, the distribution of GFP expression levels is broad. This cell pool also includes cells that were not transduced and do not express GFP at all. For transient transfection with CRISPR vectors, 1×10^6 K562 cells were resuspended in Nucleofector® Solution V (Lonza) with 1 µg plasmid DNA, and electroporated in an Amaxa 2D Nucleofector using program T-016. In case of LBR editing, a clonal K562 line stably transformed with Cas9 was used.

Human retinal pigment epithelial (RPE) cells were cultured in a 1:1 mixture of DMEM (Gibco) with Nutrient F12 (Gibco) supplemented with 10% fetal bovine serum (FBS, HyClone®), 1% penicillin/streptomycin. CRISPR vectors were transfected with 5 µL Lipofectamine® 2000 Reagent (Invitrogen) and 2.5 µg plasmid DNA in 250 µL antibiotic-free medium (Gibco).

Kc167 cells were cultured in BPYE media with 5% heat-inactivated FBS and 1% penicillin/streptomycin. 1×10^6 cells were electroporated with 1 µg each of Cas9 and sgRNA expression plasmid using a BioRad Gene Pulser II (450µF, 86V).

Constructs

For human cells, expression vector PX330 (Addgene plasmid 42230) encoding Cas9

and chimeric guide RNA was used (14). The LBR guides were cloned into expression vector pBluescript with the sgRNA cassette of PX330 and transfected into the K562 line stably transformed with Cas9. For *Drosophila* cells, Cas9 expression vector pBS-Hsp70-Cas9 (Addgene plasmid 46294) was used in combination with pU6-BbsI-chiRNA construct (Addgene plasmid 45946) (15). The sgRNAs were designed using CRISPR design (<http://crispr.mit.edu/>) (16) and CHOPCHOP (<https://chopchop.rc.fas.harvard.edu/>) (17).

The following sgRNA sequences were used:

GFP guide	5' CATGCCGAGAGTGATCCCGG 3'
NCD1 guide	5' GGATAGTTGCAAGTATTGTT 3'
CD59 guide	5' CAAGGAGGGTCTGTCTGTT 3'
LMN guide	5' GTCTGCTCGATGACACAGCT 3'
LBR guide #1	5' GAAATTTGCCGATGGTGAAG 3'
LBR guide #2	5' GCCGATGGTGAAGTGGTAAG 3'

For the cloning of individual DNA fragments from the edited GFP gene, PCR products were ligated in Zero Blunt® vector (Invitrogen) using standard procedures.

PCR

Genomic DNA (~1 x 10⁶ cells) was isolated three days after transfection using the ISOLATE II Genomic DNA Kit of Bioline. PCR reactions were carried out with 50 ng genomic DNA in MyTaq™ Red mix (Bioline) according to manufacture instructions. PCR conditions were 1 min at 95°C (1x), followed by 15 sec at 95°C, 15 sec at 55°C, and 1 min 72°C (25-30x). The PCR products were purified using the ISOLATE II PCR and Gel Kit (Bioline).

The following primer pairs spanning the target site were used (FW: forward; RV: reverse):

GFP FW	5' GCAGAAGAACGGCATCAAGGT 3'
GFP RV	5' AGCAGCGTATCCACATAGCG 3'
NCD1 FW	5' CCACCACCCCTCATACAAAG 3'
NCD1 RV	5' CTGCCCAAAGGAAAAACAAA 3'
LMN FW	5' ACATGTCTGAGCAAATCCCGA 3'
LMN RV	5' CTCTGTCTGTTTGATGCGGC 3'
LBR FW	5' GTAGCCTTTCTGGCCCTAAAAT 3'
LBR RV	5' AAATGGCTGTCTTTCCCAGTAA 3'

Sanger sequencing

Purified PCR samples (100ng) were prepared for sequencing using 4 µL of BigDye® terminator v3.1 (Applied Biosystems®) and 5pM primer in final volume of 20 µL. PCR program: 1 min at 96°C (1x), followed by 30 sec at 96°C, 15 sec at 50°C, and 4 min at 60°C (30x), and finishing with 1 min incubation at 4°C (1x). Samples were analysed by an Applied Biosystems 3730xl DNA Analyzer.

Flow cytometry

K562-GFP cells were collected 8d after nucleofection and directly analysed for fluorescence using a BD FACSCalibur. Viable cells were gated on size and shape using forward and side scatter. The GFP expression was measured using a 488 nm laser for excitation.

TIDE software

TIDE code was written in R, version 3.1.1. TIDE requires as input a control sequence data file (e.g. obtained from cells transfected without RGEN Cas9), a sample sequence data file (e.g. DNA from a pool of cell treated with RGEN Cas9) and a character string representing the sgRNA sequence (20 nt). The sequencing data files (.abif or .scf format) are imported into TIDE using the R Bioconductor package *sangerseqR* (version 1.0.0) (18). Additional parameters have default settings but can be adjusted if necessary. The web interface was constructed using the *shiny* R package, with some code adapted from the Poly Peak Parser web tool (<http://spark.rstudio.com/yostlab/PolyPeakParser/>). The latter is a genotyping tool that can identify heterozygous short indels in sequence traces, but it cannot resolve sequences with complex indel mixtures (19).

TIDE first aligns the sgRNA sequence to the control sequence to determine the position of the expected Cas9 break site. Next, the control sequence region upstream of the break site is aligned to the experimental sample sequence in order to determine any offset between the two sequence reads. Alignments are done using standard Smith-Waterman local alignment implemented in the *BioStrings* package in Bioconductor. From here on, the software uses the peak heights for each base, as determined by the sequence analysis software provided by the manufacturer of the capillary sequencing equipment (we used 3730 Series Data Collection Software V4 and Sequencing Analysis Software V6). TIDE uses these peak heights to determine the relative abundance of aberrant nucleotides over the length of the whole sequence trace.

Note that there is a 25% chance that an identical nucleotide will be found in the composite sequence trace when compared to the wild-type sequence at the same position, since only 4 different nucleotides are available. This means that the average maximum aberrant sequence signal of 75% actually represents 100% of aberrant sequence trace. The plot of this aberrant sequence signal allows the user to gauge the quality of the sequence data, verify the expected cut site, and interactively select the region used for decomposition.

The decomposition is conducted on a sequence segment downstream of the break site. By default it spans from s+5bp downstream of the break to s+5bp from the end of the shortest sequence read, with s being the maximum indel size in bp. Sequence trace models of all possible deletions and insertions of sizes {0..n} (n is by default set to 10) are constructed from the control sample trace by shifting all peaks by the appropriate number of positions

to the left or right, respectively. This is done for each of the four bases, after which the vectors of the four bases are concatenated so that the decomposition is done for all bases combined. Next, the sequence trace from the mutated DNA sample is assumed to be a linear combination of the wild-type and the modelled indel traces. This combination is then resolved by standard non-negative linear modelling, for which we used the R package *npls*. R^2 is calculated to assess the goodness of fit. The p-value associated with the estimated abundance of each indel is calculated by a two-tailed t-test of the variance-covariance matrix of the standard errors. In order to account for systematic differences between the sequence trace intensities of the control and mutated DNA, the fitting parameters are then multiplied by a constant factor such that their sum equals R^2 .

Lastly, to model insertions, the TIDE software estimates the relative frequency with which each of the four nucleotides is introduced immediately after the break site. This is done by removal of the aggregate of estimated signals of mutants that have smaller number of insertions (including non-mutated and deletions). While this can be done for all insert sizes, TIDE currently only estimates the nucleotide composition of +1 insertions, which are the most frequently observed insertions.

ACKNOWLEDGEMENT

We thank Lauren Tracey, Laura Brückner and Bastiaan Evers for sharing reagents, genomic DNA and sequence data; William Peters for assistance with setting up the web tool; several NKI colleagues for software testing; Fred van Leeuwen and members of the BvS lab for helpful suggestions.

AUTHOR CONTRIBUTION

EKB designed the study, performed experiments, wrote code, analyzed data, wrote the manuscript. TC wrote code; BvS designed and supervised the study, wrote the manuscript.

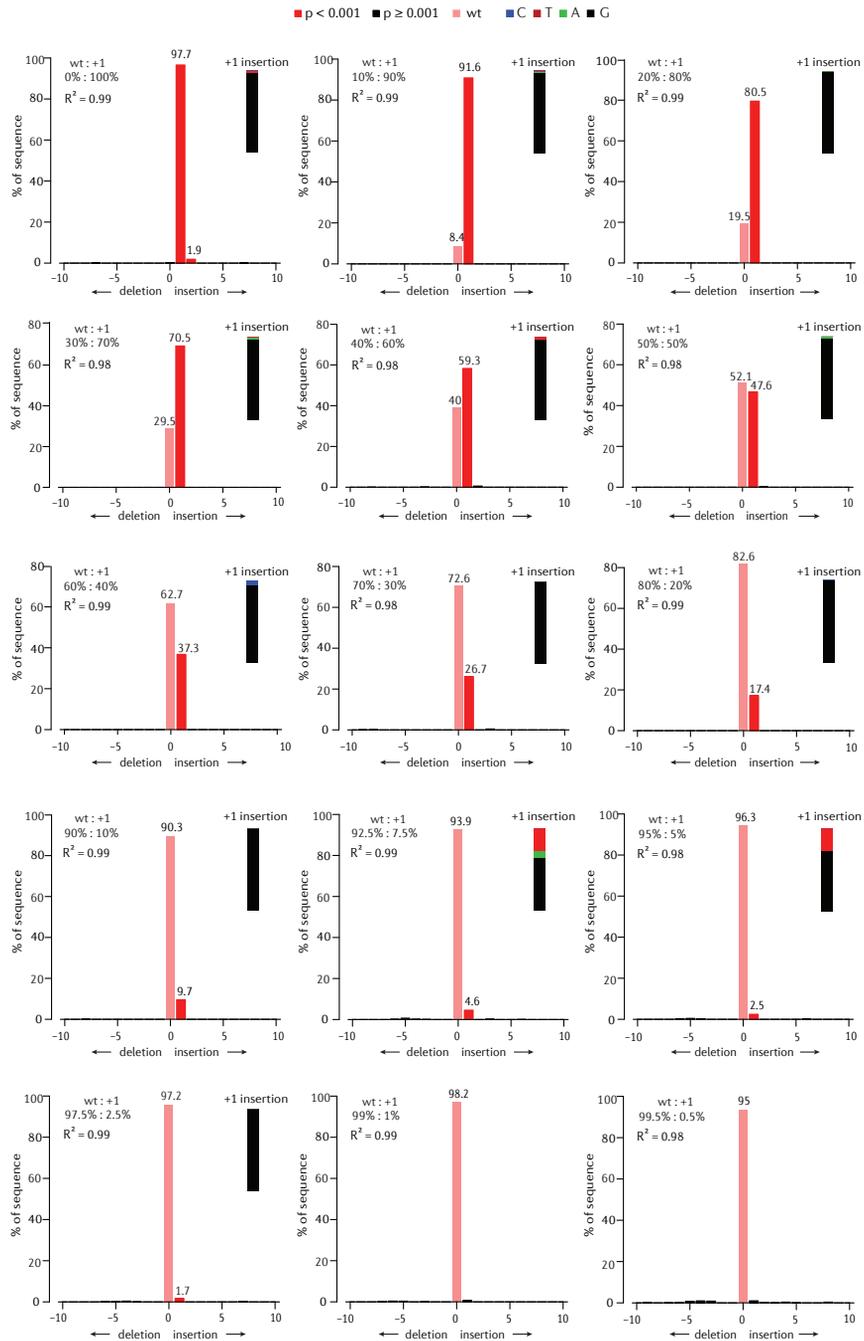
FUNDING

This work was supported by an European Research Council Advanced grant [grant number 293662] and NWO ZonMW-TOP. Funding for open access charge: Netherlands Organization for Scientific Research.

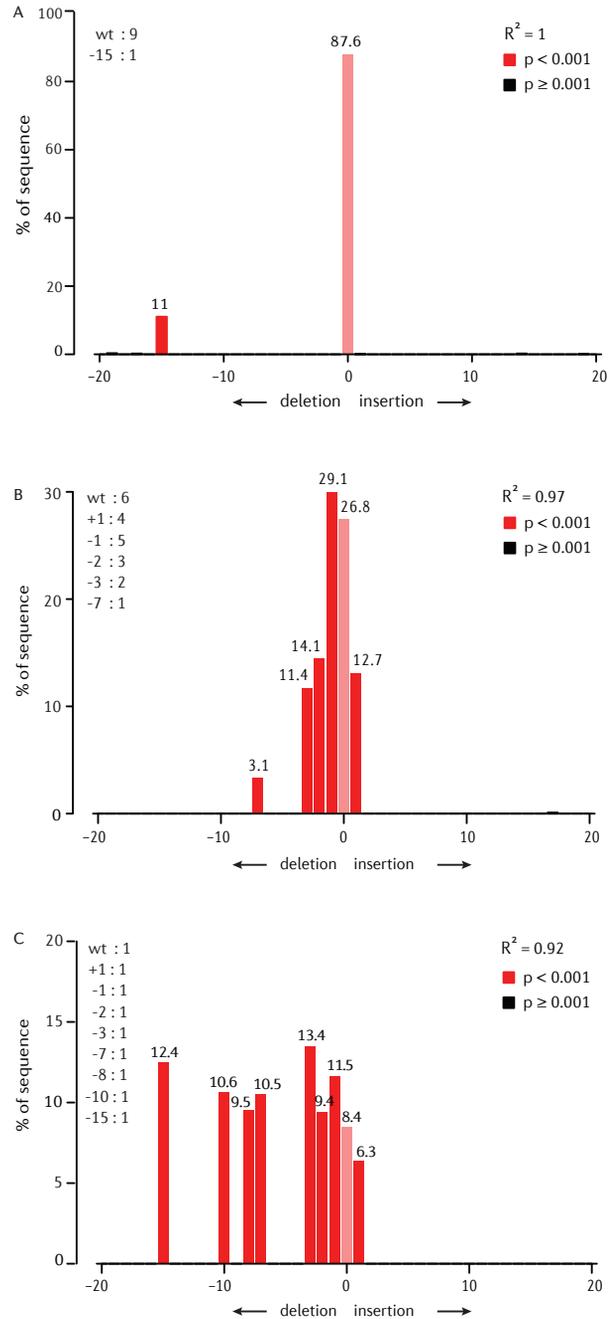
REFERENCES

- Gaj, T., Gersbach C.A., Barbas C.F., 3rd, ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* 31, 397-405 (2013).
- Kim, H., Kim J.S., A guide to genome engineering with programmable nucleases. *Nat Rev Genet* 15, 321-334 (2014).
- Yang, L., Guell M., Byrne S., Yang J.L., De Los Angeles A., Mali P., Aach J., Kim-Kiselak C., Briggs A.W. *et al.*, Optimization of scarless human stem cell genome editing. *Nucleic Acids Res* 41, 9049-9061 (2013).
- Wang, T., Wei J.J., Sabatini D.M., Lander E.S., Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343, 80-84 (2014).
- Canver, M.C., Bauer D.E., Dass A., Yien Y.Y., Chung J., Masuda T., Maeda T., Paw B.H., Orkin S.H., Characterization of Genomic Deletion Efficiency Mediated by CRISPR/Cas9 in Mammalian Cells. *J Biol Chem*, (2014).
- Mali, P., Esvelt K.M., Church G.M., Cas9 as a versatile tool for engineering biology. *Nat Methods* 10, 957-963 (2013).
- Qiu, P., Shandilya H., D'Alessio J.M., O'Connor K., Durocher J., Gerard G.F., Mutation detection using Surveyor nuclease. *Biotechniques* 36, 702-707 (2004).
- Kim, H.J., Lee H.J., Kim H., Cho S.W., Kim J.S., Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res* 19, 1279-1288 (2009).
- Guell, M., Yang L., Church G.M., Genome editing assessment using CRISPR Genome Analyzer (CRISPR-GA). *Bioinformatics*, (2014).
- Hsu, P.D., Scott D.A., Weinstein J.A., Ran F.A., Konermann S., Agarwala V., Li Y., Fine E.J., Wu X. *et al.*, DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31, 827-832 (2013).
- Fu, Y., Sander J.D., Reyon D., Cascio V.M., Joung J.K., Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* 32, 279-284 (2014).
- Maresca, M., Lin V.G., Guo N., Yang Y., Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Res* 23, 539-546 (2013).
- Amendola, M., Venneri M.A., Biffi A., Vigna E., Naldini L., Coordinate dual-gene transgenesis by lentiviral vectors carrying synthetic bidirectional promoters. *Nat Biotechnol* 23, 108-116 (2005).
- Cong, L., Ran F.A., Cox D., Lin S., Barretto R., Habib N., Hsu P.D., Wu X., Jiang W. *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819-823 (2013).
- Gratz, S.J., Cummings A.M., Nguyen J.N., Hamm D.C., Donohue L.K., Harrison M.M., Wildonger J., O'Connor-Giles K.M., Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics* 194, 1029-1035 (2013).
- Ran, F.A., Hsu P.D., Wright J., Agarwala V., Scott D.A., Zhang F., Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8, 2281-2308 (2013).
- Montague, T.G., Cruz J.M., Gagnon J.A., Church G.M., Valen E., CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* 42, W401-407 (2014).
- Hill, J.T., Demarest B. (2014).
- Hill, J.T., Demarest B.L., Bisgrove B.W., Su Y.C., Smith M., Yost H.J., Poly Peak Parser: Method and software for identification of unknown indels using Sanger sequencing of PCR products. *Dev Dyn*, (2014).

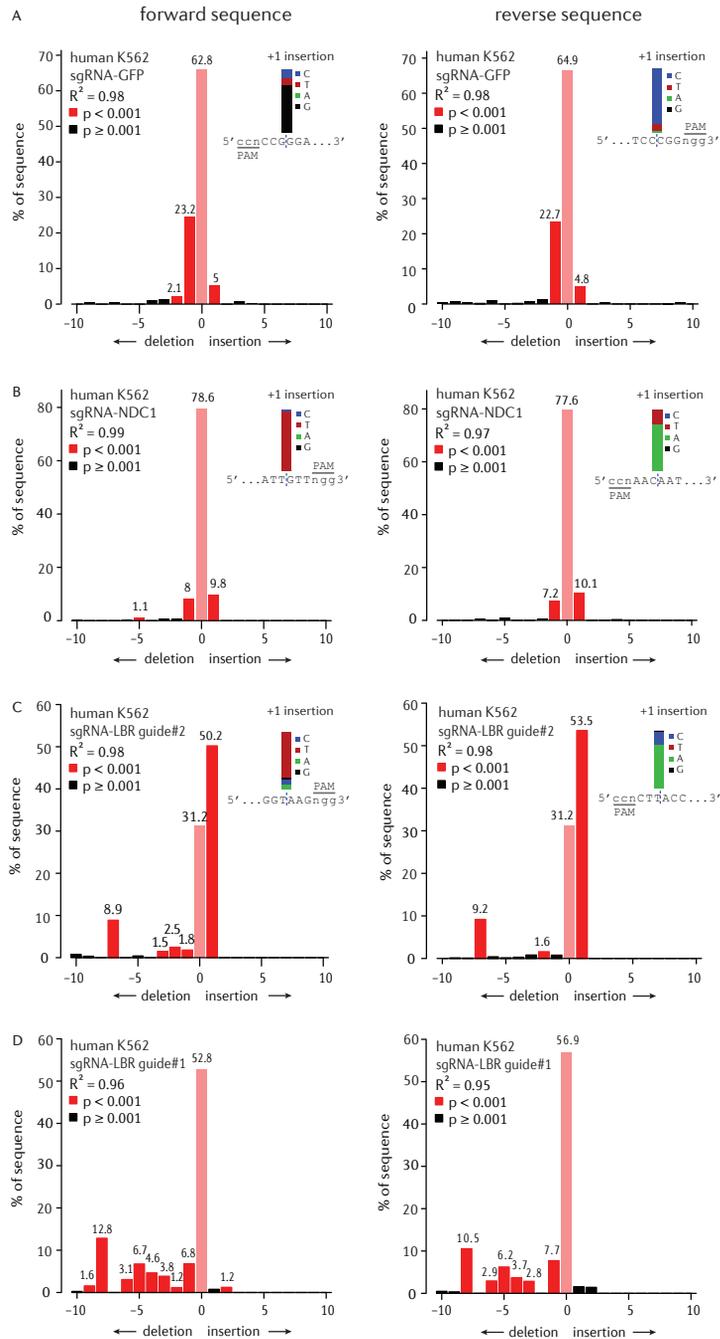
SUPPLEMENTARY DATA



Supplementary Figure S1: Specificity of TIDE indel detection of *in vitro* mixed samples. TIDE indel spectra of several samples consisting of wild-type DNA mixed with DNA with a +1 insertion in different ratios. See also FIGURE 2A.



Supplementary Figure S2: TIDE indel spectra of in vitro composite samples. TIDE was applied to several artificial samples consisting of wild-type DNA mixed with DNA carrying various indels in a broad range of relative concentrations. (A) Mix of (-15, 0) with ratio 1:9. (B) Mix of (+1, 0, -1, -2, -3, -7) with ratios 4:6:5:3:2:1. (C) Mix of (+1, 0, -1, -2, -3, -7, -8, -10, -15) all with the same concentration. See also FIGURE 2B.



Supplementary Figure S3: Comparison of TIDE indel decomposition of forward and reverse sequence reads. Indel spectra for targeted mutation in the genes (A) GFP, (B) NDC1 and (C) LBR guide#2 (D) LBR guide#1, based on sequence reads from either the forward (left panels) or reverse (right panels) strand. Insets: estimated nucleotide composition of +1 insertions; note that the forward and reverse reads should yield complementary nucleotides.

Chapter 3

EASY QUANTIFICATION OF TEMPLATE-DIRECTED
CRISPR/CAS9 EDITING

Eva K. Brinkman, Arne N. Kousholt, Tim Harmsen, Christ Leemans,
Tao Chen, Jos Jonkers and Bas van Steensel

Nucleic Acids Research, 46, e58, 2018

ABSTRACT

Template-directed CRISPR/Cas9 editing is a powerful tool for introducing subtle mutations in genomes. However, the success rate of incorporation of the desired mutations at the target site is difficult to predict and therefore must be empirically determined. Here, we adapted the widely used TIDE method for quantification of templated editing events, including point mutations. The resulting TIDER method is a rapid, cheap and accessible tool for testing and optimization of template-directed genome editing strategies. A free web tool for TIDER data analysis is available at <http://tide.nki.nl>.

INTRODUCTION

The CRISPR system for genome editing has become one of the most popular techniques in molecular biology. CRISPR endonucleases such as Cas9 can cleave genomic DNA with high precision, and due to error-prone repair mechanisms this can result in small insertions or deletions (indels) (1-3). Alternatively, precisely designed small nucleotide changes can be incorporated near the break site by providing a donor template (4, 5), such as a single-stranded oligodeoxynucleotide (ssODN) (4, 6). By homology directed repair (HDR), the DNA of the donor template is exchanged with the genomic DNA, and thereby the desired mutations are introduced (7, 8). Such precise editing offers the possibility to create and study specific mutations, or to correct disease-causing nucleotide variants (5, 9).

A current limitation of this template-directed strategy is that the efficacy is unpredictable and often low. Because error-prone non-templated repair pathways are active besides HDR, various indels are often introduced at the target site instead of the desired mutation. Moreover, a substantial fraction of the target sequence may remain unaltered. Thus, exposing a pool of cells to CRISPR and a donor template yields a complex mixture of cells with wild-type DNA, indels and the designed mutation, with unpredictable ratios (10-12). A quick and easy assay to determine these ratios is of key importance, particularly if one wants to estimate how many cells are to be cloned from the pool in order to obtain at least one clonal line with the desired mutation.

High throughput sequencing of DNA around the induced break site is a powerful tool to analyse the mutation spectrum (13), but is also expensive and requires substantial computational analysis. The frequently used Tracking of Indels by DEcomposition (TIDE) method (14) is much simpler and cheaper, as it requires only two standard Sanger capillary sequencing reactions and an easy-to-use web tool for data analysis. However, in its present form TIDE is not suitable for templated genome editing, because it can only detect overall indel frequencies and no nucleotide substitutions or specifically designed indels. To overcome this limitation we developed TIDER (Tracking of Insertions, DEletions and Recombination events), a redesigned version of TIDE. TIDER can estimate the incorporation frequency of any type of template-directed mutations (including point mutations) and separate it from the background spectrum of additional indels. The corresponding TIDER web tool is freely accessible at <http://tide.nki.nl>.

RESULTS

Quantification of template-directed CRISPR/Cas9 editing - the TIDER method.

To positively identify and quantify template-directed editing, we redesigned the TIDE method. The original TIDE protocol requires two capillary sequencing traces from a DNA stretch around the editing site: one test sample (DNA from cells treated with targeted nuclease) and one control (e.g. DNA from mock transfected cells). Indels are then quantified by computational decomposition of the mixture of sequences in the test sequence trace, using the control sequence for comparison. TIDER requires one additional capillary sequencing trace. This 'reference' trace is derived from a pure DNA sample that carries the designed base pair changes as present in the donor template. Such a reference trace can be generated readily from commercially synthesized DNA or from DNA obtained by a simple two-step PCR procedure as outlined in Methods section. The latter approach requires slightly more hands-on time, but is typically quicker and cheaper. Sequence traces derived from either source performed equally well in TIDER (see below). In addition, a text string representing the sequence of the sgRNA is used as input to determine the expected break site. To determine the individual sequence variants in the DNA of a cell pool, the algorithm decomposes the sequence trace of the experimental sample by multivariate non-negative linear modelling (FIGURE 1). For this, it uses the control and reference traces to construct a set of models of all likely outcomes of the cutting and repair process: wild-type sequence, all possible random indels at the break site, and the desired sequence as result of HDR. All of these models are collectively fitted to the experimental sample

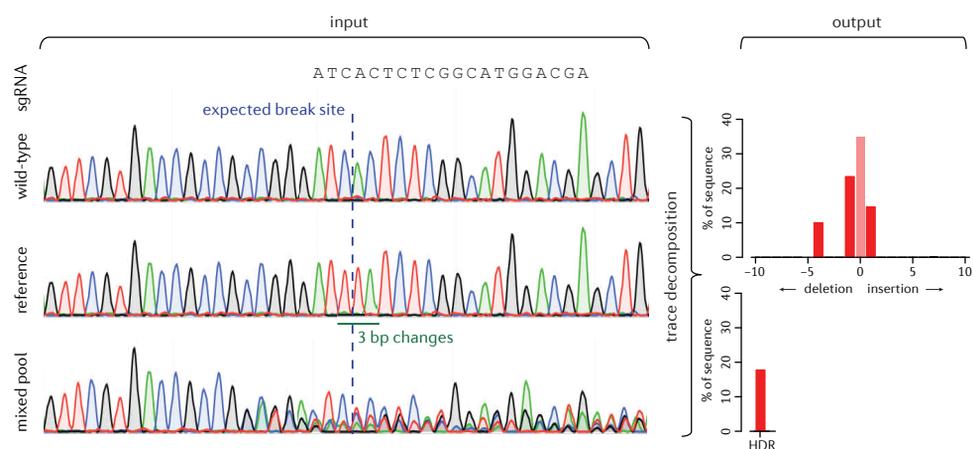


Figure 1: Assessment of homologous direct repair by sequence trace decomposition. Overview of TIDER algorithm and output. The introduction of designed mutations by homology directed repair with a donor template results in defined changes in a sequence trace. Due to NHEJ repair also insertions and deletions arise at the targeted break site. All these mutations yield in a composite sequence trace. As input a sgRNA sequence string and three sequences are required: 1) wild-type control, 2) reference file with designed mutations in the used donor template and 3) composite test sample. Trace decomposition yields the spectrum of indels and the HDR events with their frequencies (See main text and <http://tide.nki.nl> for explanation).

trace. As the construction of the models and the fitting of the data are more complicated than for TIDE, the software code was extensively modified for TIDER and subjected to a large series of tests (see below). The TIDER software provides an R^2 value as a goodness-of-fit measure, and calculates the statistical significance of the detected HDR events. Additionally, it generates a set of quality control plots that enable the user to verify the expected break site, and to visually inspect the sequence changes resulting from the editing process (SUPPLEMENTARY FIGURE S1).

In vitro validation of TIDER.

To test the performance of TIDER, we initially mimicked the occurrence of HDR events *in vitro* by mixing DNA carrying defined sequence variants. First, we combined ‘wild-type’ DNA with ‘mutant’ DNA carrying a single base pair change in various ratios. We performed standard capillary sequencing and analysed the resulting data with the TIDER software. The algorithm was able to detect the single base pair change quantitatively with a sensitivity down to ~5% at a p-value cut-off of 0.01 (FIGURE 2A, B; SUPPLEMENTARY FIGURE S2A). Only very small amounts of false-positive indels were scored across the entire range of mixing ratios (SUPPLEMENTARY FIGURE S3). No statistically significant signal was detected when a reference sequence with a different point mutation was used, attesting to the specificity of TIDER for one particular mutation (FIGURE 2B, purple triangles). More complex mixtures consisting of wild-type DNA, DNA carrying various indels and DNA with a single base pair change could also be resolved accurately (FIGURE 2A, C; SUPPLEMENTARY FIGURE S4). In this particular experiment the proportion of the designed mutant was somewhat overestimated at low mixing ratios, but with increasing ratios the estimates were accurate. Results were nearly identical for reference DNA generated by full synthesis or by the two-step PCR procedure (compare FIGURE 2B-C & SUPPLEMENTARY FIGURE S2B-C). In another mixing experiment with a different complex pool and a different mutant, the accuracy was substantially higher (SUPPLEMENTARY FIGURE S2D-F), presumably because this mutant differed at four base pair positions from the wild-type DNA instead of one position.

A potentially more challenging scenario is when the templated mutation is a small deletion. During the repair process, other (non-templated) deletions of the same size may arise. We tested the ability of TIDER to discriminate the designed deletion from alternative deletions of the same size. When we mixed wild-type DNA with varying amounts of a -4 deletion, TIDER correctly determined the deletion with high specificity as ‘designed’ when DNA carrying this deletion was used for the reference trace (FIGURE 2A, D). Similar results were obtained with four different ‘designed’ -1 deletions, although in two instances a small fraction was scored as non-templated deletion (FIGURE 2A, E). Therefore, in the presence of only a small designed deletion (-1, -2) near the expected break site the designed mutation may be underestimated somewhat. In general, however, TIDER does not mistake a ‘designed’ deletion for a non-templated deletion of the same size.

As a more stringent *in vitro* test, we generated several mutant sequences with a +1 insertion at various positions relative to the break site (FIGURE 2A), and mixed each of these ‘designed’ mutant DNAs with a complex pool of DNA that contained ~39% of ‘natural’ +1 insertions. TIDER analysis resolved the composition of the mixtures with high accuracy (FIGURE 2A, F). Sequencing of the opposite DNA strand yielded very similar results (SUPPLEMENTARY FIGURE S5), illustrating the robustness of the approach. This experiment illustrates that the presence of a non-templated insertion generally does not compromise the detection of the designed insertion of the same size. Together, these *in vitro* mixing experiments show that sequence trace decomposition can in most cases accurately identify and quantify ‘designed’ mutations (base pair substitutions as well as small deletions and insertions) in a complex background of indels caused by imperfect repair.

Comparison of TIDER to next generation sequencing.

We tested TIDER in a series of *in vivo* experiments in which we subjected specific genomic sequences to templated editing in mouse embryonic stem (mES) cells and human retinal pigment epithelium (RPE-1) cells. We co-transfected these cells with Cas9, a sgRNA and a corresponding ssODN carrying three or four nucleotide substitutions. As control the ssODN was omitted. To verify the TIDER results, we sequenced the same samples by next generation sequencing (NGS). For 5 out of 5 tested sgRNA/ssODN combinations we found that the NGS results are similar to the TIDER estimations (FIGURE 3A-E). Moreover, in cells treated with sgRNA in the absence of a donor template, TIDER detects almost no HDR events, while the non-templated indel spectra are again highly similar to those determined by NGS (FIGURE 3F; SUPPLEMENTARY FIGURE S6). Furthermore, in one set of editing experiments involving a complex set of templated nucleotide substitutions, application of TIDER with different window settings, combined with the data visualization tool, reproducibly revealed that one nucleotide substitution 40 bp upstream of the break site was less efficiently incorporated than the more proximal substitutions; this result was confirmed by NGS (SUPPLEMENTARY FIGURE S7). We conclude that TIDER can reliably estimate the frequency of HDR events in a background of non-templated indels in genomic DNA from pools of cells.

DISCUSSION

Advantages of TIDER.

TIDER is a simple and rapid assay to evaluate the efficacy of templated editing. Like TIDE, TIDER requires only standard capillary sequencing, thereby offering a widely accessible, cheap and rapid alternative to NGS. TIDER is much more quantitative and informative than the Surveyor and T7 endonuclease I cleavage assays (15, 16), which are unable to discriminate between the designed mutation and randomly induced indels.

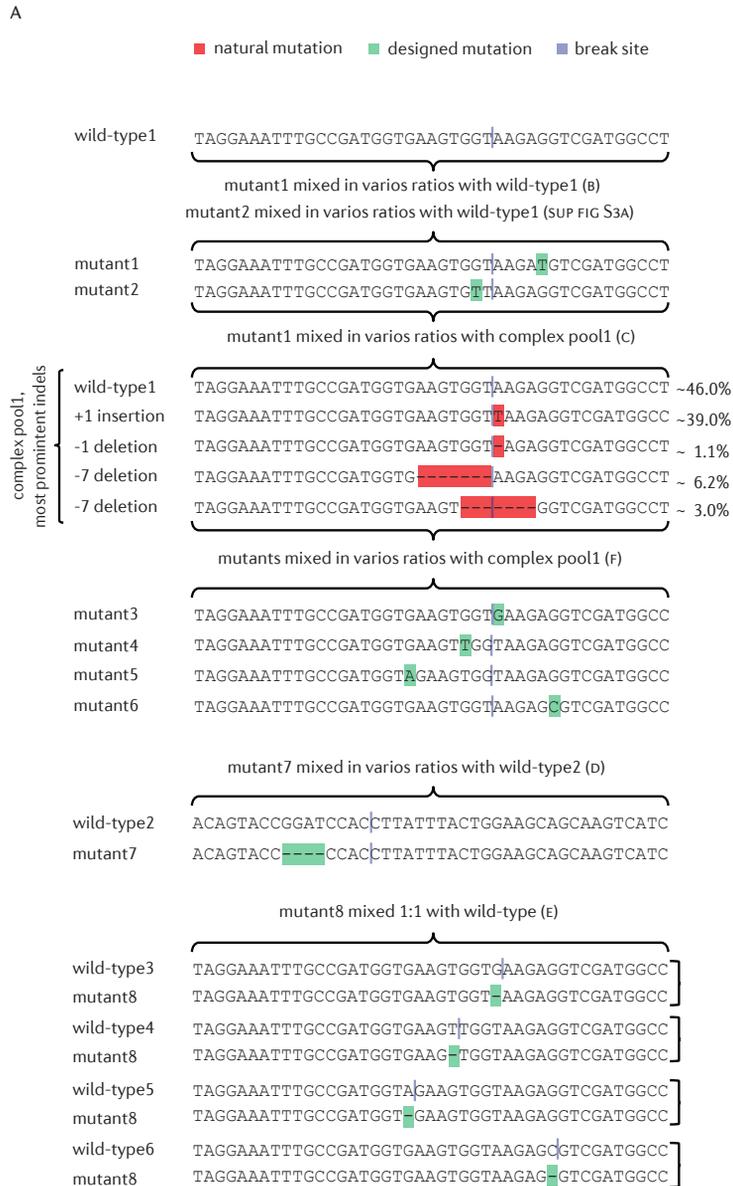
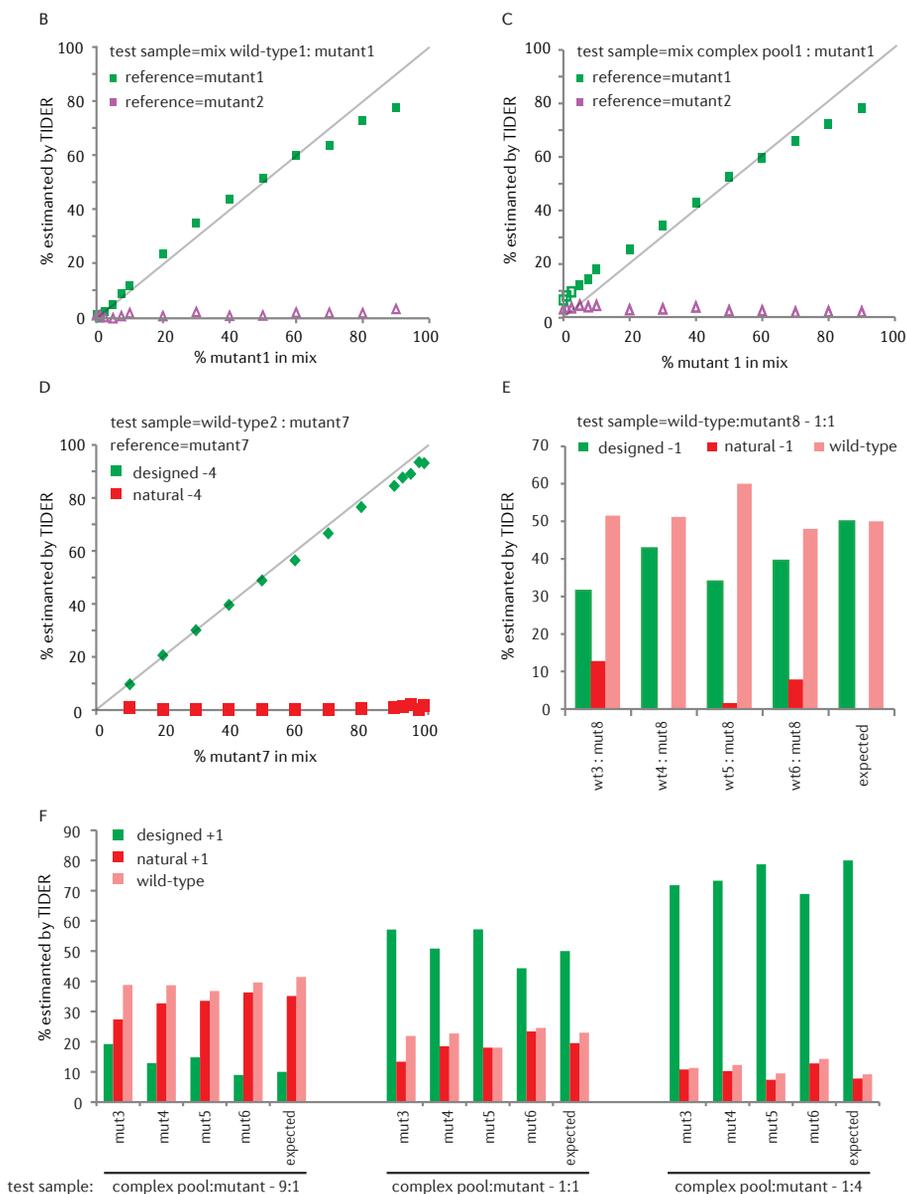


Figure 2: TIDER decomposition of in vitro mixes of DNA. Template-directed genome editing experiments in a pool of cells were simulated by in vitro mixing of DNA fragments carrying specific mutations with a corresponding wild-type DNA fragment, or with a complex pool of DNA fragments carrying different indels. (A) DNA mixtures that were tested. Letters in parentheses refer to the panels that show the corresponding TIDER results. Only the relevant sequences of the tested DNA fragments are shown; the total length of the fragments was 529 bp. 'Designed' mutations are indicated in green, 'natural' indels in red. Virtual Cas9 break sites used in these analyses are marked in dark blue. The complex pool is DNA from a pool of cells treated with Cas9 and sgRNA; it contains wild-type DNA as well as various indels introduced by NHEJ, of which the relative amounts are indicated. (B-C) PCR product with mutation1 was mixed in indicated relative amounts (horizontal axis) with wild-type DNA or with the complex pool. The proportion of mutant DNA was determined by TIDER (vertical axis) using either correct reference (mutant1, green squares) or incorrect reference (mutant2, purple triangles). See SUPPLEMENTARY FIGURES S3 and S4 for the complete decomposition results.

(figure and legend continued on next page)



(D) Same as (B-C), but for wild-type2 mixed at various ratios with mutant7 that carries a -4 deletion. Green diamonds: estimated 'designed' -4 deletions as in the reference file. Red squares: estimated 'natural' -4 deletions (i.e. all deletions of size 4 that overlap with or are immediately adjacent to the break site). (E) 1:1 mixtures of mutant8 and wild-type3-6. For the TIDER analysis mutant8 was used as reference and the respective break sites were chosen as indicated in (A); hence in each analysis mutant8 carries a 'designed' -1 deletion relative to the wild-type DNA. The percentages of the designed -1, natural -1 (other deletions of size 1) and wild-type DNA as estimated by TIDER are shown. The expected percentages are depicted in the last column. (F) TIDER analyses of mixtures of the complex DNA pool with each of mutant3-6 at three different ratios (9:1, 1:1, and 1:4). Bar graphs show percentages of the designed +1, natural +1 (other insertions of size 1) and wild-type DNA as estimated by TIDER. Expected percentages are depicted in the last column of each mixture set. In all analyses in (B-F) default TIDER settings were used (size range 0-10 for deletions and 0-5 for insertions).

Next-generation sequencing of bulk PCR products, followed by analysis using software such as CRISPResso (17) provides a highly detailed estimate of the introduced mutations, but this method is only cost-effective when large numbers of samples are multiplexed. Moreover, in many research institutes NGS is only available through a shared facility that may have turn-around times that are much longer than Sanger sequencing.

TIDER is primarily designed to determine the efficacy of templated genome editing. It complements TIDE, which can only detect non-templated indels. TIDER can estimate the incorporation frequency of any type of template-directed mutations, including point mutations, and distinguish them from the background spectrum of additional indels. While TIDER can also quantify the latter, TIDE is more suitable for the assessment of non-templated editing experiments because it is slightly simpler in experimental design. The TIDER web tool is freely available through <http://tide.nki.nl/>.

Possible limitations.

Because the TIDER algorithm analyses individual peak heights in the input sequence traces, the accuracy of TIDER relies on the quality of the PCR products and the sequence reads. This is particularly relevant when the difference between the wild-type and reference sequence is small, e.g., in case of single-nucleotide differences. In such cases we recommend that the results are verified with a sequence trace of the opposite strand. The TIDER web tool provides graphical feedback as well as an R^2 value as means to estimate the reliability of the analysis. We generally recommend that R^2 is above 0.9. While the default settings of the web tool are suited for most purposes, parameter settings can be adjusted interactively to optimize the performance. Note that the TIDER algorithm cannot resolve HDR events that have acquired an additional non-templated indel. However, it has been reported that the frequency of such double templated/non-templated mutations is low when a PAM disrupting mutation is included (17, 18).

MATERIAL & METHODS

Cell culture and transfection

Human retinal pigment epithelial (hTERT RPE-1, ATCC CRL-4000) cells were cultured in DMEM (Gibco 31966) supplemented with 10% fetal bovine serum (FBS, HyClone®). Mouse embryonic stem cells (mESCs) were cultured as described (19). Briefly, mESCs were expanded and maintained on sub-lethally irradiated mouse embryonic fibroblast feeder cells in LIF supplemented medium. Prior to transfection, cells were seeded on gelatin-coated plates and cultured in Buffalo Rat Liver cell (BRL) conditioned medium supplemented with LIF (ESG1107, Merck (Millipore)).

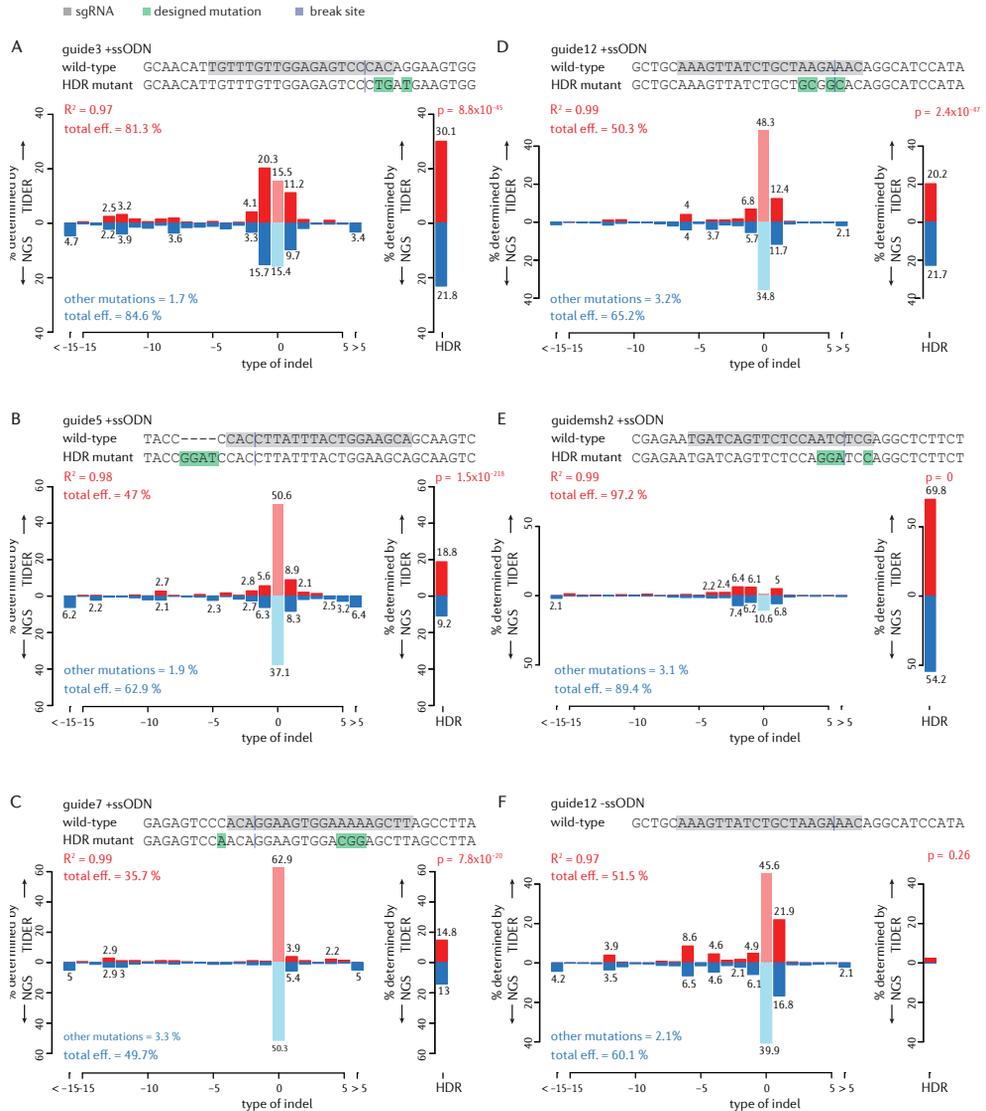


Figure 3: Application of TIDER to in vivo edited DNA sequences. Comparison of TIDER and NGS analyses of various mutations introduced by template-directed Cas9 editing in human cell line RPE (A-D,F) and mouse ES cells (E). In each panel (A-E), a pool of cells was treated with Cas9, a targeting sgRNA and a ssODN carrying three to four mutations. Panel (F) shows a control experiment corresponding to (D) in which the ssODN was omitted. Additional control experiments corresponding to (A-C) are shown in SUPPLEMENTARY FIGURE S6. In each panel, the top sequence corresponds to wild-type, with the sgRNA sequence highlighted in grey and the expected cut site marked by a vertical line; the bottom sequence indicates the designed mutant, with mutated nucleotides highlighted in green. Bar graphs show the estimated percentage of successfully edited DNA molecules (right-hand plot; ‘HDR’) and of indels of the indicated size (left-hand plot). Upward axes show TIDER estimates; downward axes show the NGS estimates based on the same DNA sample. Pale red and blue bars indicate proportions of wild-type (non-mutated) sequence. R^2 values indicate the goodness-of-fit score for the TIDER estimates; ‘total eff.’ indicates the total according to TIDER (top) and NGS (bottom); ‘other mutations’ are all non-indel, non-designed mutations as detected by NGS (and which cannot be detected by TIDER). For TIDER, the decomposition was limited to deletions of sizes 0-15 and insertions of sizes 0-5. For NGS, at least 2×10^4 reads were analysed in each experiment.

The desired mutations were introduced in hTERT RPE-1 according to the RNP CRISPR approach of IDT. The sgRNAs were designed using CRISPR design tools of Benchling or MIT tool (20). In brief, 1×10^5 cells were seeded out the day before transfection in 12-well dish in 750 μ L medium with 1 μ M final concentration DNA-PKcs inhibitor NU7441 (Cayman). 3 μ L of 10 μ M sgRNA and 3 μ L of 10 μ M Cas9 protein were mixed in optiMEM (Life Technologies) to final volume of 125 μ L and incubated in for 5 min at RT. 4.5 μ L of this Cas9/sgRNA mix, 1.5 μ L of 10 μ M ssODN (Ultramer IDT) and 4.5 μ L Lipofectamine RNAiMAX (Invitrogen) were added to 240 μ L optiMEM. Mixture was incubated at RT for 20 min before adding to the cells. The next day the medium was changed, and 2 days after transfection the cells were harvested for analysis of the genomic DNA.

mESCs were seeded 2 days before transfection at a density of 5×10^4 cells in each well of a 6-well dish. 250ng of a PX330 derived vector (Addgene #42230, with an added puromycin resistance cassette) and 2.25 μ g of ssODN were added to 250 μ L optiMEM. 6.25 μ L of Mirus TRANSIT LT-1 was added to this mixture and mixed by pipetting. After incubation for 15 min at RT, the solution was added dropwise to the cells. One day after transfection, the cells were reseeded on gelatin coated plates in BRL medium containing 3.6 μ g/mL puromycin. 2 days after reseeded, the medium was replaced without puromycin, and 4 days later the cells were harvested for genomic DNA extraction.

PCR control & test sample

Genomic DNA was isolated 2 days (RPE cells) or 7 days (mESC cells) after transfection using either the ISOLATE II Genomic DNA Kit (Bioline) or lysis buffer (100 mM Tris (pH 8.5), 50 mM EDTA, 40 mM NaCl, 0.2% SDS and 100 μ g/mL proteinase K) for 2 hours at 55°C followed by 45 min incubation at 85°C and DNA precipitation by addition of 2.5 volumes of 100% ethanol, followed by 30 min centrifugation at 14,000 RPM at 4°C. After washing of pellets with 70% ethanol, pellets were dissolved in TE buffer by overnight incubation at 55°C. PCR reactions were carried out with 50 ng genomic DNA in MyTaq™ Red mix (Bioline) according to manufacturer's instructions using primers a & b (10 μ M) as listed below. PCR thermocycling scheme: 1 min at 95°C (1x), followed by 15 sec at 95°C, 10-20 sec at 55-60°C, and 10-20 sec 72°C (25-35x). The PCR products were purified using the ISOLATE II PCR and Gel Kit (Bioline). The Msh2 target site was amplified with Taq polymerase (MRC Holland) using the following PCR program: 2 min 94°C (1x), followed by 30 sec at 94°C, 3 sec at 53.8°C and 40 sec at 72°C (37x) and 5 min at 72°C (1x).

The following sgRNA sequences were used:

guide_msh2	5' GATCAGTTCTCCAATCTCG 3'
guide3	5' TGTTTGTGGAGAGTCCCAC 3'
guide5	5' TGCTCCAGTAAATAAGGTG 3'
guide7	5' AAGCTTTTCCACTTCCTGT 3'
guide8	5' TAAGCTTTTCCACTTCCTG 3'
guide12	5' AAAGTTATCTGCTAAGAAAC 3'
guide_lbr2	5' GCCGATGGTGAAGTGGTAAG 3'

The following ssODNs sequences were used:

guide_msh2	960 msh2- 1187r-rflp-s	5' TAGTCTCTCCTCCTGGTAAAACGCATTCTTTGGTCCAATCTGA ATCAGAAGAGCCTGGATCCTGGAGAACTGATCATTCTCGGGAACTC ACACAAGCTTAGCTTCCTCTGGGTGGA*A*T 3'
guide3	ssODN9	5' GTGTTTCATAGATTCTCAGAGGATTAACAGCAAGCAACATTGTT TGTTGGAGAGTCCCTGATGAAGTGGAAAAAGCTTAGCCTTACTTTGT TCTGCTTTAGCATGGCAACAATCTCTTAG 3'
guide5	ssODN7	5' CATCTTCCACAAAATTTCTGGTGATAGATGACTTGCTGCTTCC AGTAAATAAGGTGGATCCGGTACTGTACTTTAAAGATGTCACTTCAA GTGTAGACTCATTGCTCTGTATATTGGTT 3'
guide7, guide8	ssODN8	5' ACTACTTACCACTAAGAGATTGTTGCCATGCTAAAGCAGAACAA AGTAAGGCTAAGCTCCGTCCACTTCCTGTTGGACTCTCCAACAAACA ATGTTGCTTGCTGTTTAACTCCTCTGAGAA 3'
guide12	ssODN13	5' TAAATTACTTATATAAGACTCCCTGAAAAAACCCTCTGGCTG CAAAGTTATCTGCTGCGGCACAGGCATCCATATACAGAGATGAAAAT GATGATTTTCAAGTAGAGAAGAAAAGAAT 3'

* denote PTO linkages

The following primer pairs spanning the target site were used (a: forward; b: reverse):

	primer a	primer b
guide_msh2	5' TCTTTCTCAGTTTGAAGACATCC 3'	5' GGGGTATTTTACATGAAGG 3'
guide3, guide7, guide8	5' GAGCTGGGCATCTTCACTTTA 3'	5' CCAGGCTGCCAAGTCTTTAT 3'
guide5	5' TTTCCCTGGGTTTACCTTT3'	5' CCCATTACAGTACACCATACT 3'
guide12	5' AGAAGGAGCTTTTCAAGATTATGG 3'	5' CCATGCAGTTTCACTTGAACG 3'
guide_lbr2	5' GTAGCCTTTCTGGCCCTAAAAT3'	5' AAATGGCTGTCTTTCCAGTAA 3'

The following primer pairs spanning the to be edited site were used (c: reverse; d: forward):

	primer c*	primer d*
guide_msh2	5' TAGTCTCTCCTCCTGGTAAAACGCATTCC TTTGGTCCAATCTGAATCAGAGAGCCTGGATC CTGGAGAAGTATCATTCTCGGGAACTCACA CAAGCTTAGCTTCCTCTGGGTGGAAT 3'	5' ATTCCACCCAGAGGAAGCTAAGCTTGTGT GAGTTCCTCCGAGAATGATCAGTTCTCCAGGAT CCAGGCTCTTGTGATTGATTGACCAAAAGG AATGCGTTTTACCAGGAGGAGACTA 3'
guide3	5' CTAAGCTTTTTCCACTTCATCAGGGACTC TCCAACAAACAATGTT 3'	5' GTGTTTATAGATTCTCAGAGGATTAACA GCAAGCAACATTGTTTGTGGAGAGTCCCTGA TGAAGTGGAAAAGCTTAGCCTTACTTTGTTC TGCTTTAGCATGGCAACAATCTCTTAG 3'
guide7, guide8	5' ACTACTTACCACTAAGAGATTGTTGCCAT GCTAAAGCAGAACAAAGTAAGGCTAAGCTCCG TCCACTTCCTGTTGGACTCTCCAACAAACAAT GTTGCTTGCTGTTAATCCTCTGAGAA 3'	5' GTTGAGAGTCCAACAGGAAGTGGACGGA GCTTAGCCTTACTTTG 3'
guide5	5' CATCTCCACAAAATTTCTGGTGATAGA TGACTTGCTGCTTCCAGTAAATAAGGTGGATC CGGTACTGTACTTTAAAGATGTCACTTCAAGT GTAGACTCATTGCTCCTGTATATGTT 3'	5' GTGACCTCTTTAAAGTACAGTACATCACC TTATTTACTGGAAGC 3'
guide12	5' CTGTATATGGATGCCTGTGCCGAGCAGA TAACTTTGCGAGCCAG 3'	5' TAAATTACTTATATAAGACTCCCCTGAAA AAACCACTCTGGCTGCAAGTTATCTGCTGCG GCACAGGCATCCATATACAGAGATGAAAATGA TGATTTTCAAGTAGAGAAGAAAAGAAAT 3'
guide_lbr2_1	5' GCCATCGACGCTCTTACCACT 3'	5' AGTGGTAAGAGCGTCGATGGC 3'
guide_lbr2_2	5' CTTACCACTTCTACCATCGCAAAT 3'	5' ATTTGCCGATGGTAGAAGTGGTAAG 3'
guide_lbr2_3	5' CGACCTCTTACCAACTTACCATCG 3'	5' CGATGGTGAAGTTGGTAAGAGGTCG 3'
guide_lbr2_4	5' GCCATCGACCTCTTACCACT 3'	5' AGTGGTGAAGAGGTCGATGGCC 3'

*Note that some of primers c,d are similar to the used ssODN oligo

PCR Reference sample

The reference sequence was generally generated in a 2-step PCR reaction. Two complementary primers (primers c & d) were designed that carried the designed mutations as present in the donor template. Two standard PCR reactions were done with 50 ng wild-type genomic DNA in MyTaq™ Red mix (Bioline) using primers a & c and primers b & d. PCR thermocycling scheme: 1 min at 95°C (1x), followed by 15 sec at 95°C, 15 sec at 55°C, and 20 sec 72°C (25-30x). The two PCR products were purified using the ISOLATE II PCR and Gel Kit (Bioline). Next, the resulting two PCR amplicons (each 1 µL) were combined with 48 µL buffer (10 mM Tris, 50 mM NaCl, 1 mM EDTA) and denatured for 5 min 95°C and cooled down to 25°C (0.1 °C/sec). Of this mixture 3 µL was subsequently used as template in a PCR reaction with MyTaq™ Red mix (Bioline) with primers a & b, starting with an extension step as follows:

15 sec at 72°C (1x), followed by 15 sec at 95°C, 15 sec at 55°C, and 20 sec 72°C (25-30x). The PCR products were purified using the ISOLATE II PCR and Gel Kit (Bioline).

Sanger sequencing

Purified PCR samples (100 ng) were prepared for sequencing using 4 µL of BigDye® terminator v3.1 (Applied Biosystems®) and 5 pmol primer in final volume of 20 µL. Thermocycling program: 1 min at 96°C (1x), followed by 30 sec at 96°C, 15 sec at 50°C, and 4 min at 60°C (30x), and finishing with 1 min incubation at 4°C (1x). Sequence traces were generated on an Applied Biosystems 3730xl DNA Analyzer, running 3730 Series Data Collection Software V4 and Sequencing Analysis Software V6.

Next generation sequencing

The following primers sequences were used for NGS:

msh2_F	EB514	5' ACACTCTTTCCCTACACGACGCTCTCCGATCTTAACGCTGTTGGAGTTGG GTATGTGG 3'
msh2_R	EB515	5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCAAGGAAATACAGGGG AAGG 3'
guide5_F	EB524	5' ACACTCTTTCCCTACACGACGCTCTCCGATCTATAGCGGACAGAACCAA TATACAGGACAA 3'
guide5_R	EB525	5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACAAACCACCATATTTA AGGAATTA 3'
guide12_F	EB532	5' ACACTCTTTCCCTACACGACGCTCTCCGATCTCTACGTAATTCCTACC TGTAAGTTATTTATG 3'
guide12_R	EB533	5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCATGCAGTTTCACTTGA ACGA 3'
guide3_7_8_F	EB534	5' ACACTCTTTCCCTACACGACGCTCTCCGATCTACGGATTTTTCATTGGTT TCTGTGTTCA 3'
guide3_7_8_R	EB535	5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTTTGCTATATTGAAAC TCTTTTGGGA 3'

PCR was performed in two steps with genomic DNA as template; PCR1 with ~50 ng genomic DNA and site specific barcoded primers. PCR2 used 2 µL of each PCR1 product with Illumina PCR Index Primers Sequences 1-12. Each sample was generated with a unique combination of a barcode and index. Both PCR reactions were carried out with 25 µL MyTaq™ Red mix (Bioline), 4 µM of each primer and 50 µL final volume in a 96 well plate. PCR conditions were 1 min at 95°C, followed by 15 sec at 95°C, 15 sec at 58°C and 1 min at 72°C (15x). 20 µL of 8 samples were pooled and 100 µL was loaded onto a 1% agarose gel. PCR product was cut from gel to remove the primer dimers and cleaned with ISOLATE II PCR and Gel Kit (Bioline). The isolated samples were sequenced by Illumina MiSeq.

NGS data analysis

In order to identify insertions and deletions, the distance between a fixed sequence ~50 nt upstream of the break site and ~50 nt downstream of the break site was determined. Insertions and deletions have a distance longer or shorter than wild-type, respectively. For each of the remaining reads a window of 50 nucleotides (from -25 to +25 relative to the expected break site) was compared to the corresponding nucleotide sequence strings of the control and reference sequences. Windows with zero or one mismatches compared to the control sequence were counted as wild-type reads. Subsequently, remaining reads with zero or one mismatches compared to the reference sequence were counted as HDR reads. All other reads are counted as other mutations. Reads in which we could not find a match with the constant parts are discarded. Finally, for each sample, the ratio of each mutation type over the total of reads is calculated.

TIDER software

TIDER is built upon the previous published TIDE software (14). TIDER code was written in R, version 3.3.2. TIDER requires as input a control sequence trace file (e.g. obtained from cells transfected without Cas9), a sample sequence trace file (e.g. DNA from a pool of cell treated with Cas9 and donor template), a reference sequence trace file (e.g. DNA from the donor template) and a character string representing the sgRNA sequence (20 nt). We advise to sequence a stretch of DNA ~700 bp enclosing the designed editing site. The projected break site should be located preferably ~200 bp downstream from the sequencing start site. The sequencing data files (.abif or .scf format) are parsed using R Bioconductor package *sangerseqR* (version 1.10.0) (21). Additional parameters have default settings but can be adjusted if necessary. The web interface was constructed using the *shiny* R package (version 1.0.0).

Briefly, the algorithm consists of the following steps. Both the test sample and the reference sequence are first aligned to the control sample sequence using standard Smith-Waterman local alignment implemented in the *BioStrings* package (version 2.42.1) in Bioconductor (22). Subsequent calculations are done using the peak heights of the four bases for each position in the aligned sequence trace data. Next, for each position, the absolute peak height of each base is converted to a relative peak height by dividing it by the sum of the peak heights of all four bases at that position. All subsequent calculations are done using these relative peak heights.

In contrast to TIDE, the decomposition window of TIDER spans by default from 20 bp upstream of the break to 80 bp downstream from the break. This window can be interactively adjusted, but it should contain all nucleotides that are edited. Within this window, sequence trace models are constructed of all possible indel occurrences that

may realistically be expected, i.e., deletions of sizes $\{0..n\}$ and insertions of sizes $\{0..m\}$ that overlap with or are immediately adjacent to the break site. For example, to model all possible -4 deletions, 5 different sequence trace models are constructed; and to simulate all possible insertions of size 3, $4^3 = 64$ trace models are constructed. By default, n is set to 10, and m to 5. For deletions, the model traces are simply constructed from the control trace by deleting the values at the corresponding positions. For insertions, the average value of the same nucleotide occurrence within the whole sequence trace is used. The break site is assumed to be between the 17th and 18th bases in the sgRNA target sequence (3 bp before the PAM) (23). The sequence trace models are constructed accordingly for each of the four bases, after which the vectors of the four bases are concatenated, so that each model consists of a single vector. Subsequently, control sequence model, all indel models and the reference sequence model are combined into a single decomposition matrix. To avoid doublet models, in case the reference consists of an insertion or deletion at the break site, the identical simulated insertion or deletion is removed from the decomposition matrix.

The decomposition is subsequently performed in two iterations. First, the sequence trace from the test sample is assumed to be a linear combination of the wild-type trace, the modelled indel traces and the reference trace. This combination is decomposed by standard non-negative linear modelling, for which we used the R package *npls* (version 1.4). After this first trace decomposition, all sequence variants with an estimated frequency of exactly 0 are removed, and the decomposition is repeated with the remaining models.

Next, the frequencies of the various traces of same deletion or insertion size are summed. R^2 is calculated to assess the goodness of fit. The p-value associated with the estimated abundance of the reference trace is calculated by a two-tailed t-test of the variance-covariance matrix of the standard errors. Finally, the fitting coefficients (frequencies) are multiplied by a constant factor such that their sum equals R^2 .

Plots for visual inspection of sequence traces

TIDER uses the relative peak heights to determine the abundance of aberrant nucleotides by subtracting the peak heights of the highest control nucleotide over the length of the whole sequence trace of either the test sample or reference. Then, the highest peaks in the reference and the peaks in the control at the same location that are not the highest are identified (the designed base pair changes). Of these positions the corresponding nucleotide peak signal in the control and test sample are plotted to show the relative incorporation of the donor template. The plots of these sequence signals allows the user to check the quality of the sequence data, inspect proper alignment, verify the expected cut site, and interactively select the region used for decomposition.

TIDER settings and constrains

For TIDER, we have empirically determined an optimal decomposition window of 100 bp for most applications, but this can be interactively adjusted.

In case the designed mutation consists of an insertion larger than +1, TIDER does not consider natural insertions of the same size, because we found the decomposition to become less robust, and because we and others have rarely observed natural insertions larger than +1 (14, 24).

It has been reported that the incorporation of donor template is less efficient when the designed point mutations are further away from the break site (25). This may confound TIDER estimates when such distal mutations are combined with mutations close to the break site. This is also what we observed (SUPPLEMENTARY FIGURE S7). By comparing different settings for the decomposition window and by visual inspection of the TIDER plots it is possible to infer such biases.

ACKNOWLEDGEMENT

We thank William Peters and Rubayte Rahman for assistance with setting up the web tool; NKI colleagues for software testing; members of the BvS lab for helpful suggestions.

AUTHOR CONTRIBUTIONS

EKB designed the study, performed experiments, wrote code, analyzed data, wrote the manuscript. ANK and TH performed *in vivo* editing experiments; CL and TC wrote code; JJ: supervised ANK; BvS supervised the study, wrote the manuscript.

FUNDING

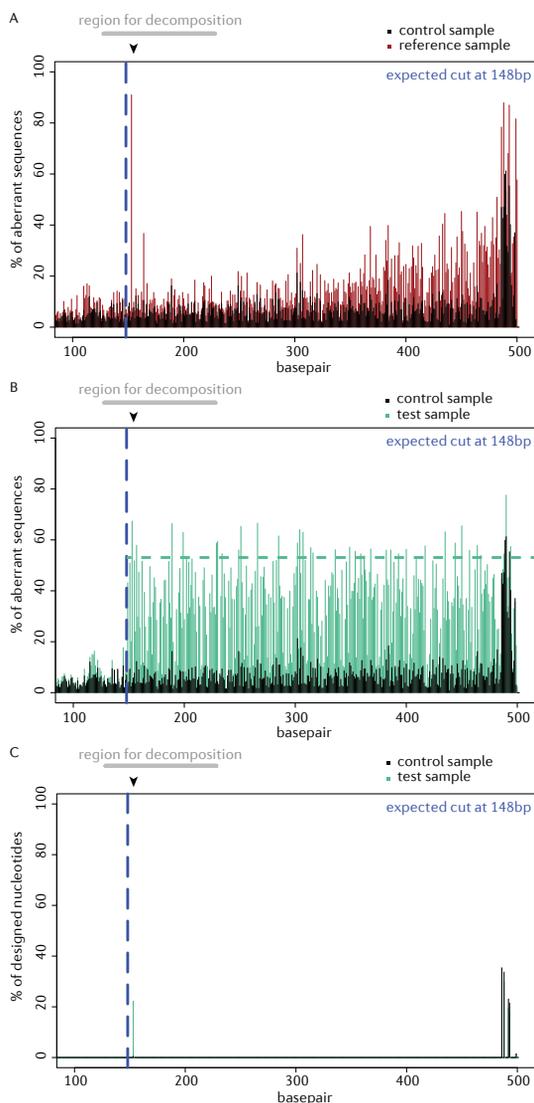
This work was supported by the Netherlands Organization for Scientific Research ZonMW-TOP grants 91210033 to H. te Riele and 91211061 to BvS; European Research Council Advanced Grant 694466 to BvS; the Danish Council for independent Research (ANK), Medical Sciences grant DFF-4183-00599 to JJ and ANK; and The Danish Cancer Society grant R91-A7351-14-S9 to ANK. Funding for open access charge: European Research Council grant AdG 694466.

REFERENCES

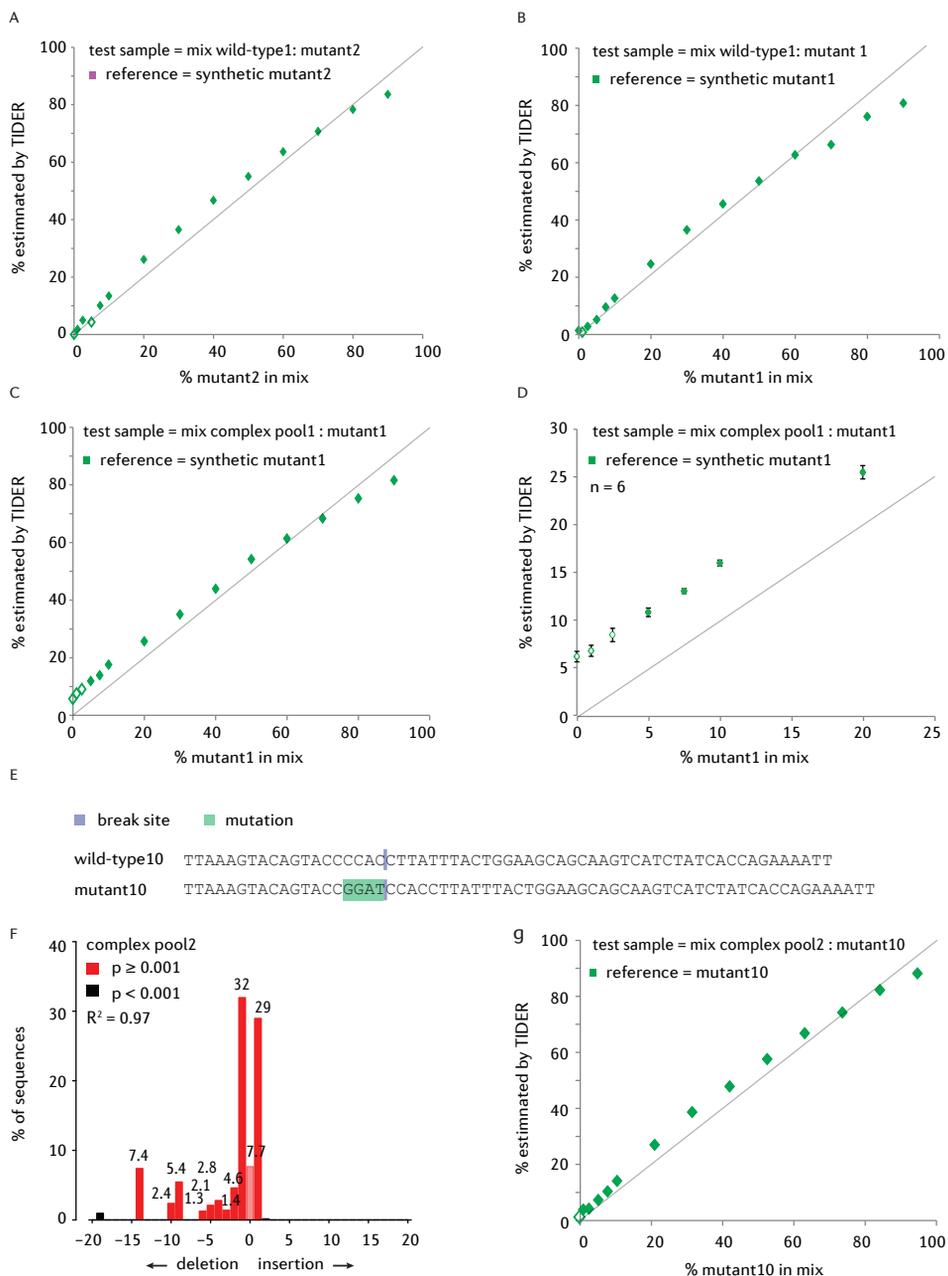
- Mali, P., Esvelt K.M., Church G.M., Cas9 as a versatile tool for engineering biology. *Nat Methods* 10, 957-963 (2013).
- Cong, L., Ran F.A., Cox D., Lin S., Barretto R., Habib N., Hsu P.D., Wu X., Jiang W. *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819-823 (2013).
- Jinek, M., East A., Cheng A., Lin S., Ma E., Doudna J., RNA-programmed genome editing in human cells. *Elife* 2, e00471 (2013).
- Yoshimi, K., Kunihiro Y., Kaneko T., Nagahora H., Voigt B., Mashimo T., ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. *Nat Commun* 7, 10431 (2016).
- Inui, M., Miyado M., Igarashi M., Tamano M., Kubo A., Yamashita S., Asahara H., Fukami M., Takada S., Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system. *Sci Rep* 4, 5396 (2014).
- Chen, F., Pruett-Miller S.M., Huang Y., Gjoka M., Duda K., Taunton J., Collingwood T.N., Frodin M., Davis G.D., High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Methods* 8, 753-755 (2011).
- Storici, F., Snipe J.R., Chan G.K., Gordenin D.A., Resnick M.A., Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing. *Mol Cell Biol* 26, 7645-7657 (2006).
- Wang, Z., Zhou Z.J., Liu D.P., Huang J.D., Double-stranded break can be repaired by single-stranded oligonucleotides via the ATM/ATR pathway in mammalian cells. *Oligonucleotides* 18, 21-32 (2008).
- Ma, H., Marti-Gutierrez N., Park S.W., Wu J., Lee Y., Suzuki K., Koski A., Ji D., Hayama T. *et al.*, Correction of a pathogenic gene mutation in human embryos. *Nature*, (2017).
- Richardson, C.D., Ray G.J., DeWitt M.A., Curie G.L., Corn J.E., Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* 34, 339-344 (2016).
- Horlbeck, M.A., Witkowsky L.B., Guglielmi B., Replogle J.M., Gilbert L.A., Villalta J.E., Torigoe S.E., Tjian R., Weissman J.S., Nucleosomes impede Cas9 access to DNA *in vivo* and *in vitro*. *Elife* 5, (2016).
- Yang, L., Guell M., Byrne S., Yang J.L., De Los Angeles A., Mali P., Aach J., Kim-Kiselak C., Briggs A.W. *et al.*, Optimization of scarless human stem cell genome editing. *Nucleic Acids Res* 41, 9049-9061 (2013).
- Guell, M., Yang L., Church G.M., Genome editing assessment using CRISPR Genome Analyzer (CRISPR-GA). *Bioinformatics*, (2014).
- Brinkman, E.K., Chen T., Amendola M., van Steensel B., Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* 42, e168 (2014).
- Babon, J.J., McKenzie M., Cotton R.G., The use of resolvases T4 endonuclease VII and T7 endonuclease I in mutation detection. *Mol Biotechnol* 23, 73-81 (2003).
- Qiu, P., Shandilya H., D'Alessio J.M., O'Connor K., Durocher J., Gerard G.F., Mutation detection using Surveyor nuclease. *Biotechniques* 36, 702-707 (2004).
- Pinello, L., Canver M.C., Hoban M.D., Orkin S.H., Kohn D.B., Bauer D.E., Yuan G.C., Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol* 34, 695-697 (2016).
- Paquet, D., Kwart D., Chen A., Sprout A., Jacob S., Teo S., Olsen K.M., Gregg A., Noggle S. *et al.*, Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* 533, 125-129 (2016).
- Aarts, M., te Riele H., Subtle gene modification in mouse ES cells: evidence for incorporation of unmodified oligonucleotides without induction of DNA damage. *Nucleic Acids Res* 38, 6956-6967 (2010).
- Hsu, P.D., Scott D.A., Weinstein J.A., Ran F.A., Konermann S., Agarwala V., Li Y., Fine E.J., Wu X. *et al.*, DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31, 827-832 (2013).
- Hill, J.T., Demarest B. (2014).
- Huber, W., Carey V.J., Gentleman R., Anders S., Carlson M., Carvalho B.S., Bravo H.C., Davis S., Gatto L. *et al.*, Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* 12, 115-121 (2015).
- Jinek, M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821 (2012).

24. van Overbeek, M., Capurso D., Carter M.M., Thompson M.S., Frias E., Russ C., Reece-Hoyes J.S., Nye C., Gradia S. *et al.*, DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol Cell* 63, 633-646 (2016).
25. Beumer, K.J., Trautman J.K., Mukherjee K., Carroll D., Donor DNA Utilization During Gene Targeting with Zinc-Finger Nucleases. *G3* (Bethesda) 3, 657-664 (2013).

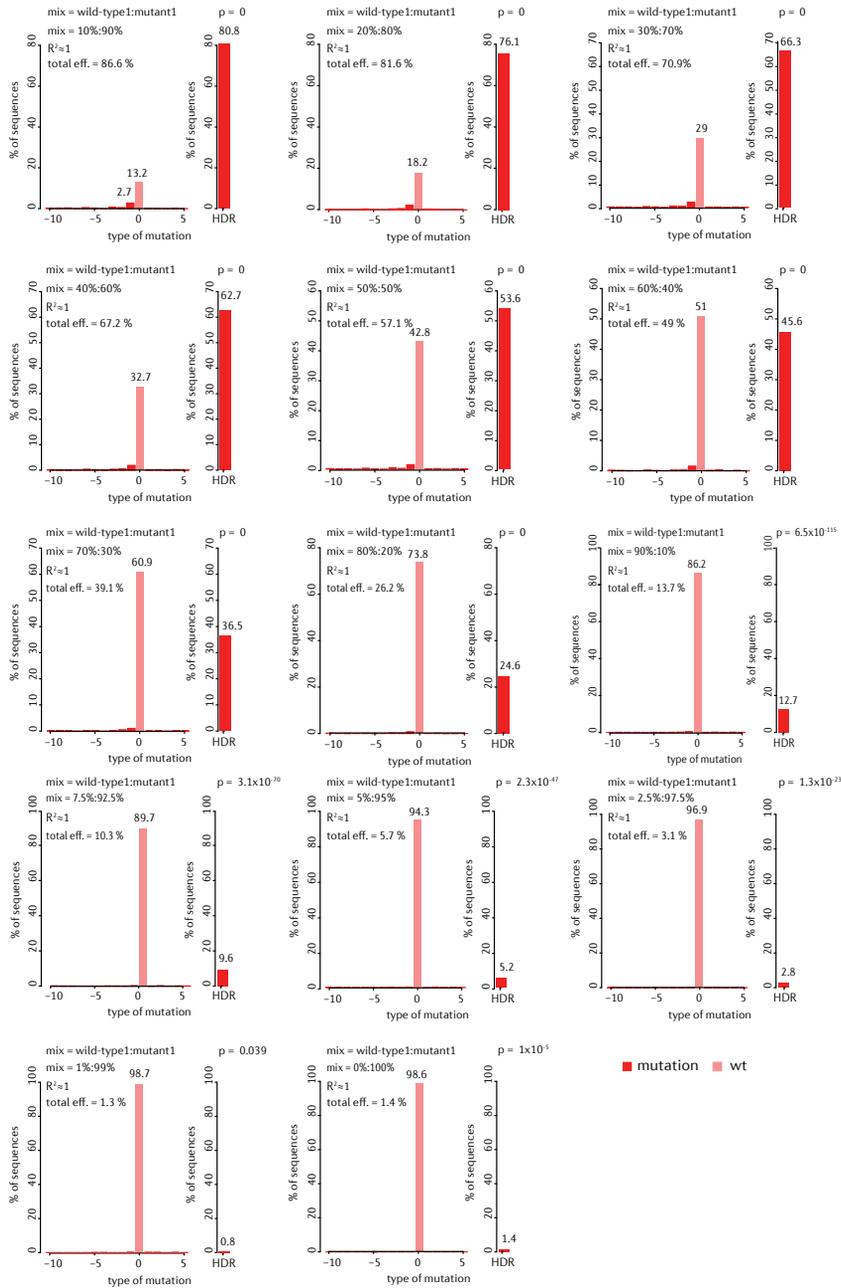
SUPPLEMENTARY DATA



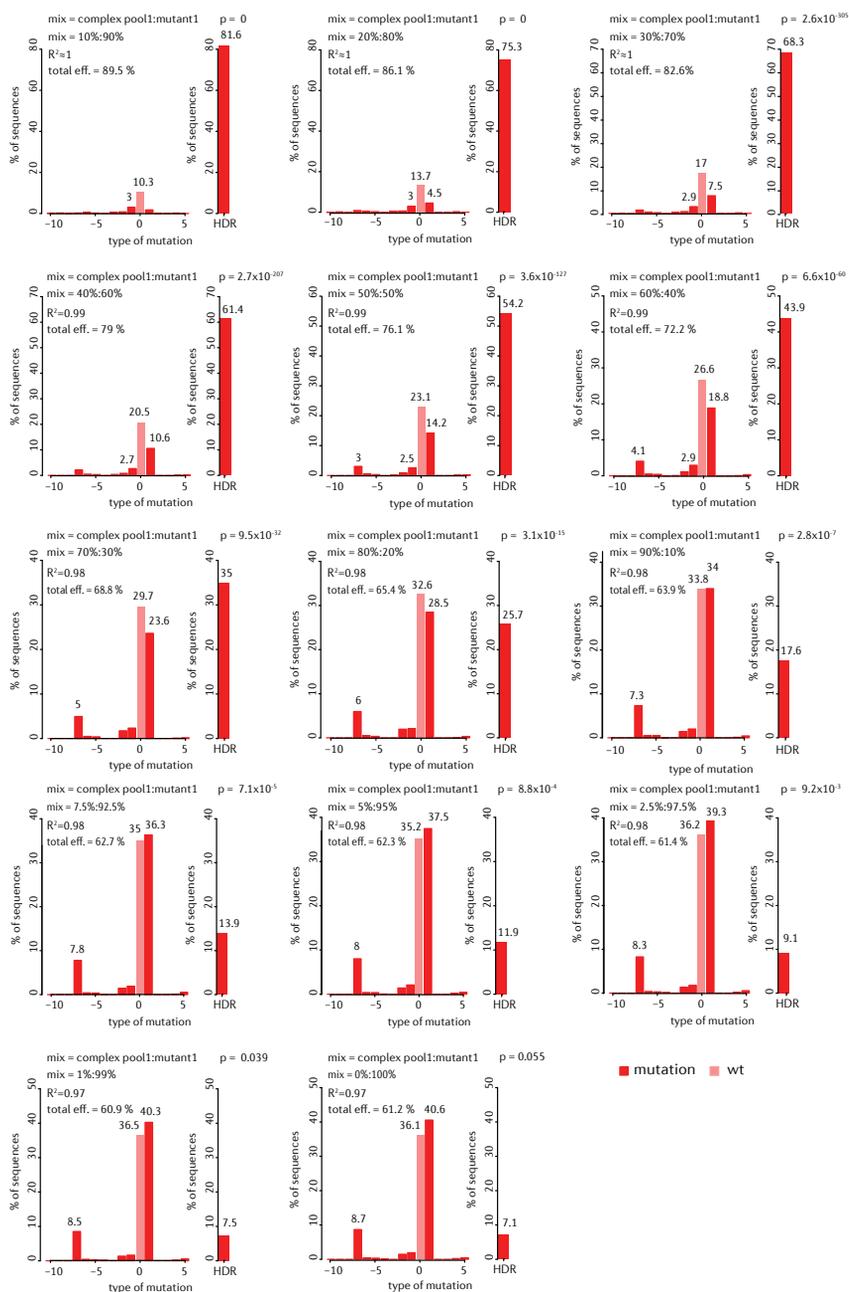
Supplementary Figure S1: Plots for visual inspection of TIDER analysis. Plots show for each position around the expected break site (vertical blue dashed line) how much the nucleotide signals in the sequencing traces deviate from the expected signals according to the control sample sequence. For each position, a value of 0% indicates that the detected nucleotide did not deviate from the control sequence; 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). Vertical dashed blue line marks the expected break site; arrowheads mark nucleotides that are expected to be substituted. (A) Comparison of the reference (red) and control (black) traces. Note the high scores at the positions of the altered nucleotides. Near the end of the sequencing trace are also high scores, but these are typically due to the lower quality of the trace towards the end. Fluctuations in the control signal reflect local variation in the quality of the sequence trace. (B) Comparison of the experimental test sample (green) and control (black) traces. Immediately after the break site a consistently elevated signal is observed in the experimental sample (horizontal green dashed line), which is due to insertions and deletions at the break site. The starting position of this elevated signal may be used to verify that breaks were induced at the expected location. (C) Masked plot of (B), showing only the expected mutated nucleotide(s) in the experimental test sample. The substituted nucleotide (arrowhead) again shows elevated signals, but not as much as in (A) because the editing efficiency is not 100%.



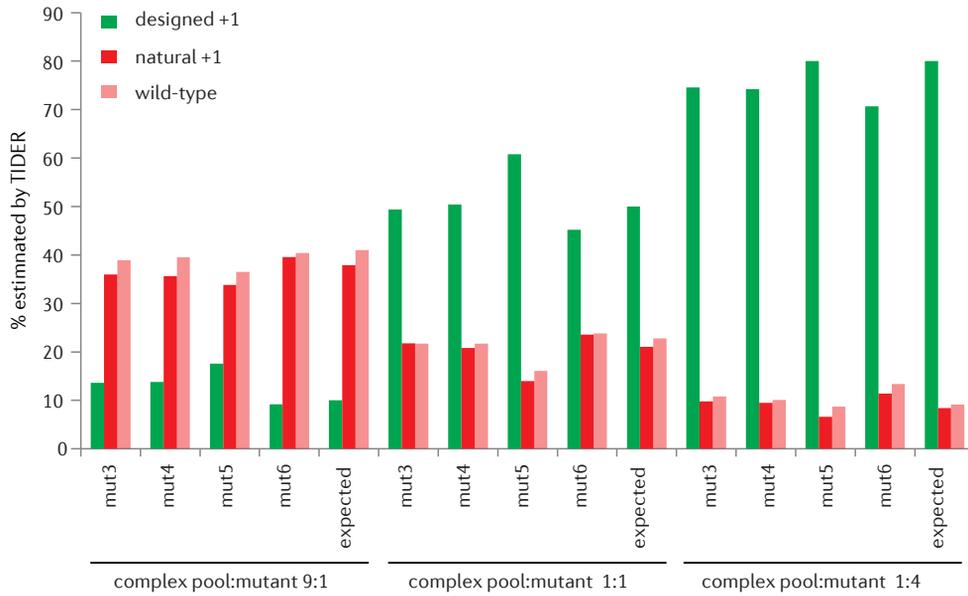
Supplementary Figure S2: TIDER analyses of additional in vitro mixes. (A-C) Proportions of mutant DNA as determined by TIDER for various mixtures of DNA as indicated. Reference DNA was generated from fully synthetic DNA, rather than by the 2-step PCR mutagenesis procedure. Mixtures in (B-C) correspond to those in FIGURE 2B-C. (D) Estimate of the reproducibility of TIDER: mixtures with low ratios of reference, similar as in (C) were independently sequenced 6 times, divided over two separate sequencing runs. Mean ± SD is plotted. (E-G) Additional mixing experiment with a different mutant and a different complex DNA pool. (E) Sequences, mutations and break site. (F) Indel spectrum of complex pool2 as determined by conventional TIDE. (G) TIDER estimates from samples with various mixing ratios.



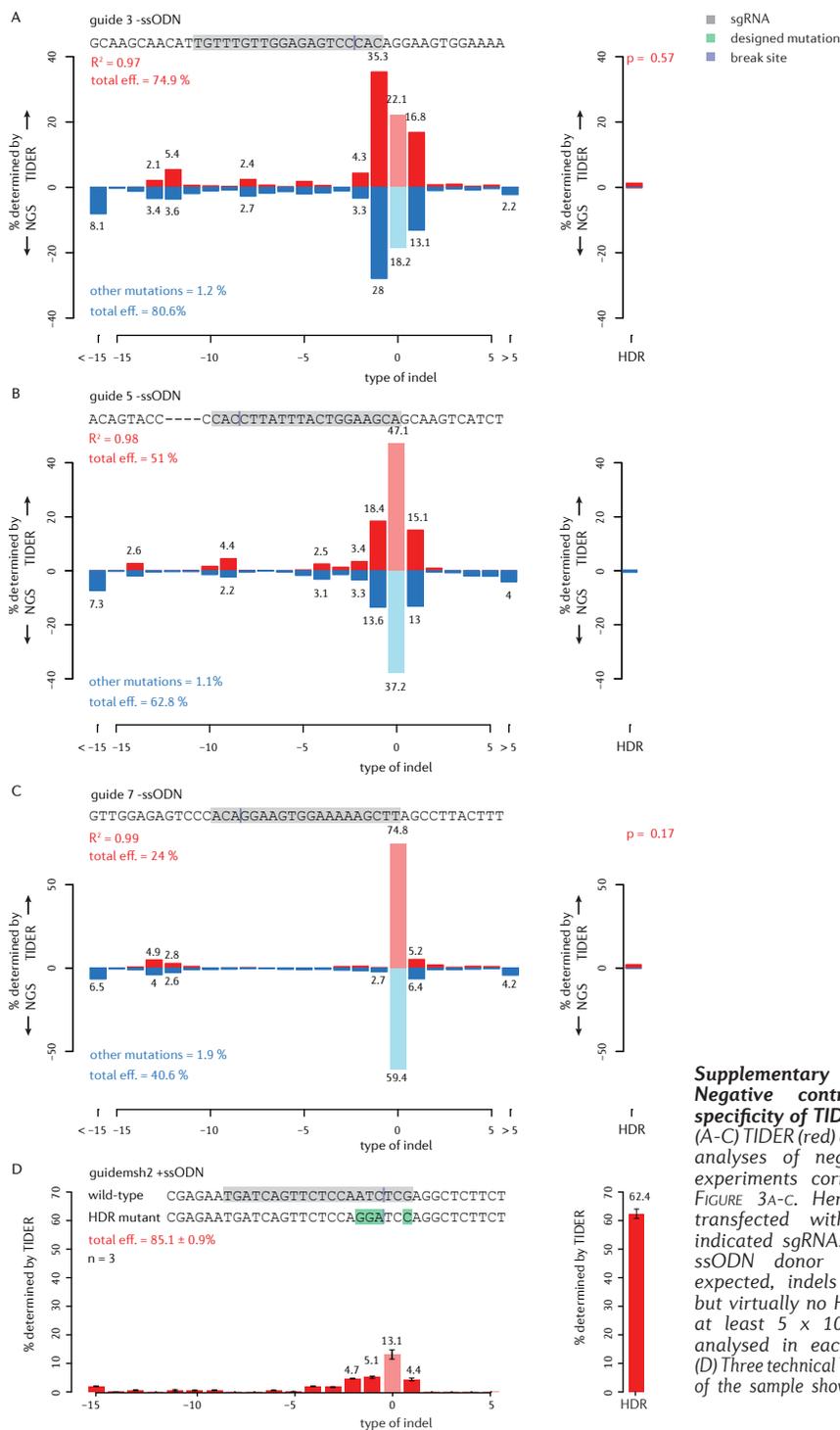
Supplementary Figure S3: Specificity of TIDER indel detection in vitro mixed samples. TIDER indel and HDR spectra of samples consisting of wild-type DNA mixed with DNA with a HDR reference in different ratios. Samples correspond to individual data points in FIGURE 2B; input mixing ratios are indicated. In each panel, bar graphs show the estimated percentage of successfully edited DNA molecules (right-hand plot; 'HDR') and of indels of the indicated size (left-hand plot). R^2 values indicate the goodness-of-fit score for the TIDER estimates; 'total eff' indicates all mutations combined (HDR plus indels).



Supplementary Figure S4: Specificity of TIDER indel detection in in vitro mixed samples with complex DNA pool. TIDER spectra of samples corresponding to individual data points in FIGURE 2c. See legend of SUPPLEMENTARY FIGURE S3 for explanation.

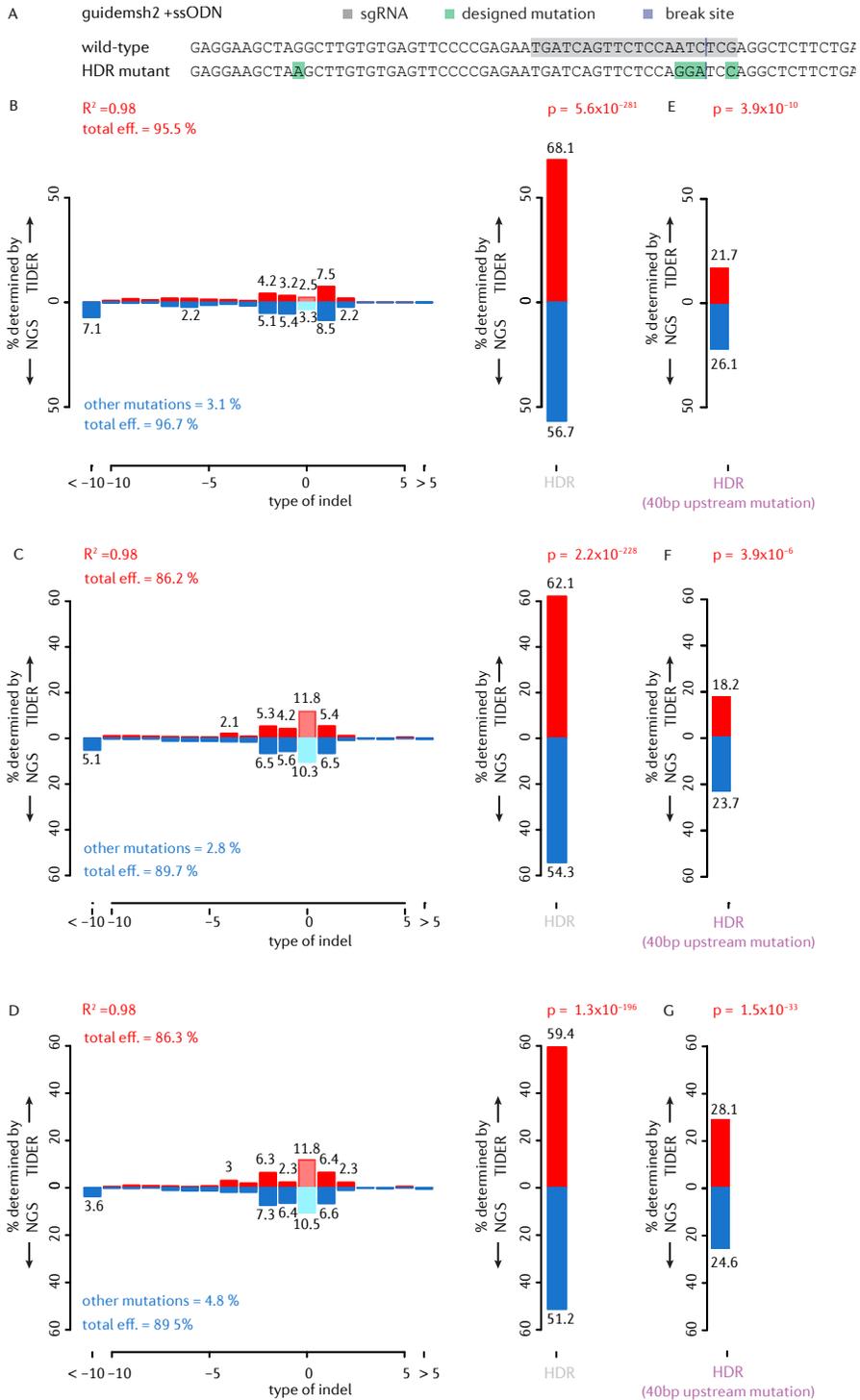


Supplementary Figure S5: TIDER analysis of reverse complement sequencing data. Wild-type and mutant sequences with +1 insertion at or surrounding break site. Mutant sequences of a were mixed in vitro with DNA carrying various indels including natural +1 insertion in different ratios. Reverse complement reads of same in vitro mixtures as shown in FIGURE 2H. The used reference is indicated and the expected ratio is plotted in the last column of each mixture. Designed +1, natural +1 and wild-type as determined by TIDER are shown.

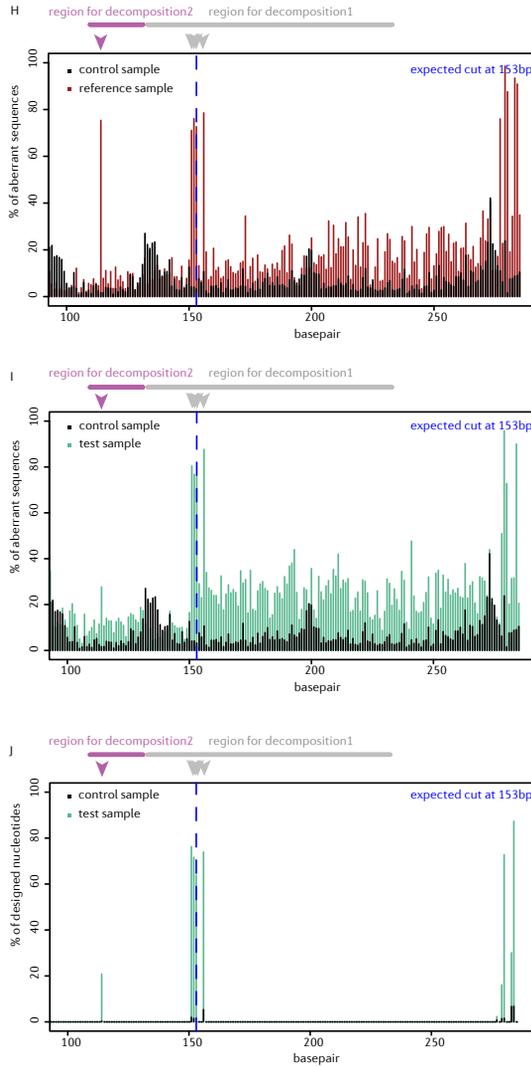


Supplementary Figure S6. Negative controls confirm specificity of TIDER estimates. (A-C) TIDER (red) and NGS (blue) analyses of negative control experiments corresponding to FIGURE 3A-C. Here, cells were transfected with Cas9 and indicated sgRNAs, but without ssODN donor template. As expected, indels are detected but virtually no HDR. For NGS, at least 5×10^4 reads were analysed in each experiment. (D) Three technical TIDER replicates of the sample shown in FIGURE 3E.

Chapter 3



(figure and legend continued on next page)



Supplementary Figure S7: Reproducible and robust TIDER analysis of a complex mutagenesis design. Three additional biological replicates and further analysis of the mutagenesis experiment of FIGURE 3E. (A) Extended sequences showing an additional designed substitution 40 bp upstream of the break site (B-D) Comparison of TIDER and NGS analyses of three independent biological replicates when analysis is restricted to decomposition window 133..233. Data representation as in FIGURE 3E. For NGS, at least 2×10^5 reads were analysed in each experiment. (E-G) Comparison of TIDER and NGS estimates of HDR frequency based the same data as in (B-D), but with a decomposition window of 110..130. TIDER and NGS consistently indicate that the incorporation of the distal mutation is ~2.5-fold less frequent than that of the designed mutations near the break site. (H-J) Quality plots of the test sample depicted in (B). Grey bar: decomposition window 133..233; purple bar: 110..130. TIDER visualization of the sequence trace data confirms this interpretation, as the distal (purple arrowhead) and proximal (grey arrowheads) mutated nucleotide peaks are present in the reference sample (H, red), but the distal peak is weaker compared to the proximal mutated nucleotide peaks in the test sample (I J, green). See legend of SUPPLEMENTARY FIGURE S1 for explanation quality plots. In all panels the decomposition was limited to deletions of sizes 0-10 and insertions of sizes 0-5.

Chapter 4

RAPID QUANTITATIVE EVALUATION OF CRISPR
GENOME EDITING BY TIDE AND TIDER

Eva K. Brinkman and Bas van Steensel

Springer Nature: Methods in Molecular Biology, in press

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 is 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 strands 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 analysed by the TIDE or TIDER algorithm (available at <http://tide.nki.nl>). 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.

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 (2-4). 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 (10-12). To study a mutation of interest, clonal mutant lines need to be isolated from the cell pool. Because this is a very labour-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 analysed 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>.

MATERIALS

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 webtools (e.g. <http://crispr.mit.edu/>, <https://chopchop.rc.fas.harvard.edu/>, <https://www.deskgen.com/>).

DNA Purification Buffers and Solutions

Usually, one to three 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 , BIO-52067) 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).

PCR amplification of control & 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 500-1500bp 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
3. PCR mastermix (example makes 50 μ L):

21-x μ L	H ₂ O
2 μ L	primer a (10 μ M stock)
2 μ L	primer b (10 μ M stock)
x μ L	genomic DNA (~50 ng)
25 μ L	2x pre-mix of buffer, Taq polymerase and dNTPs (e.g. BioLine MyTaq™, BIO-25044)

PCR program:

Step	Temperature	Time (min:sec)	Number of cycles
Initial denaturation	95 °C	1:00	1
Denaturation	95 °C	0:15	25-30
Annealing	55-58 °C	0:15	
Extension	72 °C	0:10	
Final Extension	72 °C	1:00	1
	4 °C	hold	

4. Check an aliquot of the PCR product on 1-2% agarose gel. A sharp single band should be visible.
5. Purify the PCR product using a kit according to manufacturer's instructions (e.g. BioLine ISOLATE II PCR and Gel Kit, BIO-52060).

Two-step PCR amplification of reference DNA (TIDER only)

1. Genomic DNA
2. PCR primers. Similar primers as used for control & experimental sample plus two specific primers carrying the designed mutation(s) as present in the donor template (SEE NOTE 13). For the primers carrying the mutation(s), it is advised to include at least 10 complementary nucleotides on the 3' side of the mutation(s).
3. PCR mastermix (example makes 50 μ L):

	PCR mix1	PCR mix2
21-x μ L	H ₂ O	H ₂ O
2 μ L	primer a (10 μ M stock)	primer d (10 μ M stock)
2 μ L	primer c (10 μ M stock)	primer b (10 μ M stock)
x μ L	genomic DNA (~50 ng)	genomic DNA (~50 ng)
25 μ L	2x pre-mix of buffer, Taq polymerase and dNTPs (e.g. BioLine MyTaq™, BIO-25044)	2x pre-mix of buffer

PCR program:

Step	Temperature	Time (min:sec)	Number of cycles
Initial denaturation	95 °C	1:00	1
Denaturation	95 °C	0:15	25
Annealing	55-58 °C	0:15	
Extension	72 °C	0:10	
Final Extension	72 °C	1:00	1
	4 °C	hold	

4. Purify PCR product using kit and manufacture instructions (e.g. BioLine ISOLATE II PCR and Gel Kit, BIO-52060)
5. Anneal the following two PCR products for 1 minute 95°C, cool down to 20°C (0.1 °C/sec).

48 μ L	annealing buffer (=10 mM Tris, 50mM NaCl, 1mM EDTA)
1 μ L	PCR mix1
1 μ L	PCR mix2

- Extend the annealed products and amplify the joined product.

18 µL	H ₂ O
2 µL	primer a (10 µM stock)
2 µL	primer b (10 µM stock)
3 µL	annealed oligo mix
25 µL	2x pre-mix of buffer, Taq polymerase and dNTPs (e.g. BioLine MyTaq™, BIO-25044)

4

PCR program:

Step	Temperature	Time (min:sec)	Number of cycles
Initial denaturation	95 °C	1:00	1
Denaturation	95 °C	0:15	25
Annealing	55-58 °C	0:15	
Extension	72 °C	0:10	
Final Extension	72 °C	1:00	1
	4 °C	hold	

- Check the PCR product on 1-2% agarose gel. A sharp single band should be visible.
- Purify the PCR product using a kit and manufacturer's instructions (e.g. BioLine ISOLATE II PCR and Gel Kit, BIO-52060).

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.

- Purified PCR samples (100 ng)
- PCR primers. Similar primers can be used as for the amplification of control & experimental sample
- PCR mastermix (example makes 20 µL):

15.5-x µL	H ₂ O
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, 4337456)

PCR program:

Step	Temperature	Time (min:sec)	Number of cycles
Initial denaturation	96 °C	1:00	1
Denaturation	96 °C	0:30	30
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 3730xl DNA Analyzer). Sequence trace files must be saved in .ab1 or .scf format.

Equipment

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

Software

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

METHODS

Control and experimental sample generation

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. 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 (FIGURE 1A,B). 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 (FIGURE 2A).

Reference sample generation (TIDER only)

TIDER is required for genome editing experiments in the presence of a donor template. In addition to the control and experimental sample trace, 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 (FIGURE 2E). There are two paths to obtain the reference sequence as described below.

2-step PCR

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) (FIGURE 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 NOTE 13).

DNA synthesis

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 13).

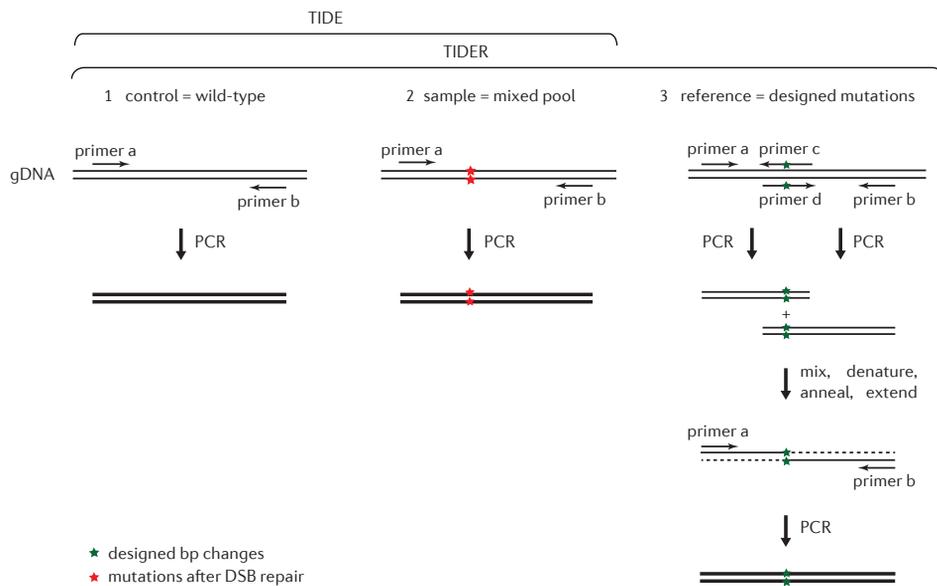


Figure 1 Method to generate the required input samples for TIDE and TIDER. Control and test samples can be obtained by PCR using primers spanning the CRISPR target site (primers a,b). The reference sequence (TIDER only) can be created in a similar way as site directed mutagenesis (16) (See Methods for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

Web tool

The PCR products of the control, optional reference, and experimental sample are processed by conventional Sanger sequencing. 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>). In addition, a character string representing the guide RNA sequence (20 nt) is required as input (SEE NOTE 2,9). 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 NOTE 3,4). Subsequently, two output plots are generated: one plot that can help with quality control, and one that displays the indel/HDR spectrum.

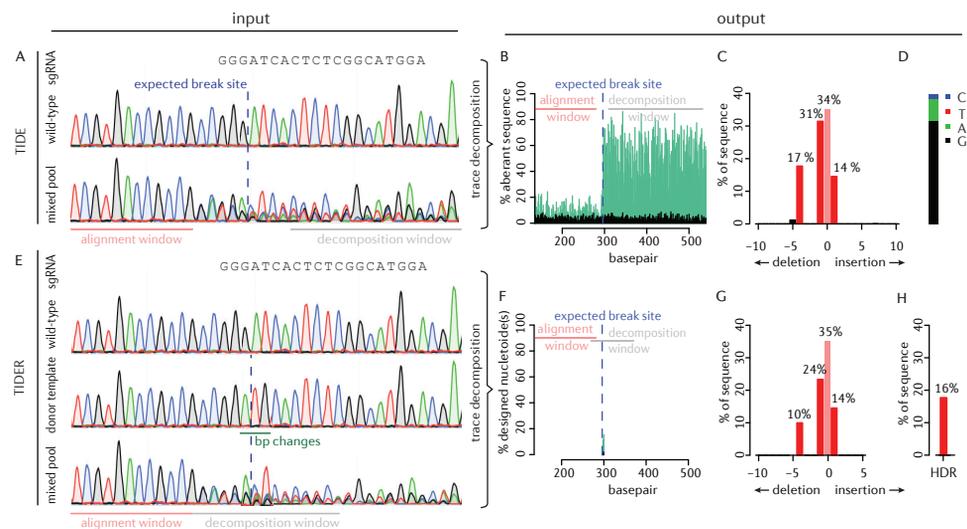


Figure 2: Overview of TIDE and TIDER algorithm. Due to imperfect repair (and repair by homology directed repair with a donor template) after cutting by a targeted nuclease, the DNA in the cell pool consists of a mixture of indels (and designed mutations). The various introduced mutations in the pool 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 (grey 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 (grey bar) as used in TIDER are represented. (G-H) Decomposition gives the spectrum of indels (G) and the HDR events (H) with their frequencies.

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 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) (FIGURE 2B). The percentages of aberrant nucleotides in the control should be low along the whole sequence trace. However, the experimental sample consists of 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 four 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 NOTE 1,5,8).

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 (FIGURE 2F).

Mutation detection by decomposition

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 5). To perform the decomposition, a set of sequence trace models are generated 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 peaks 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 modelling the combination of trace models that can best explain the composite sequence trace in the experimental sample is determined (FIGURE 2C) (SEE NOTE 6). An R^2 value is calculated as a measure of the goodness of fit (SEE NOTE 7,8), and the statistical significance of the detection of each indel is calculated.

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 NOTE 5,12). 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 11).

The reliability of TIDE and TIDER depends on the quality of the input samples (SEE NOTE 10). 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 12).

Sequence determination of the +1 insertion (TIDE only)

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 (FIGURE 2D). For longer insertions this base-calling is computationally complicated and currently not implemented.

NOTES

1. *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 towards the end (FIGURE 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 (FIGURE 3B). Such region might confound the decomposition of the sequence trace.
2. *Nuclease type.* TIDE(R) is currently designed for regular Cas9. But it can be used to analyse data from another nuclease, by entering in the webtool 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 (FIGURE 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.
3. *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 (FIGURE 3C).
4. *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 (FIGURE 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).

5. *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 fulfil 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 Material section 'PCR amplification of control & experimental sample DNA'. 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.
6. *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 analysed again by TIDE(R). A diploid cell gives a percentage of ~50% per mutation.
7. *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 10). 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 1). 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 modelled, which may result in a low R^2 score. The size range of indels that are modelled can be manually changed to larger number to test if this improves the fit (FIGURE 3E).
8. *Overall efficiency.* The overall efficiency refers to the estimated total fraction of DNA with mutations around the break site. It is calculated as $R^2 \times 100 - \% \text{ wild-type}$.
9. *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 (FIGURE 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.
10. *Poor sequence quality.* When the sequence has poor quality overall, TIDE(R) will yield poor results with a low R^2 value (SEE NOTE 8) 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

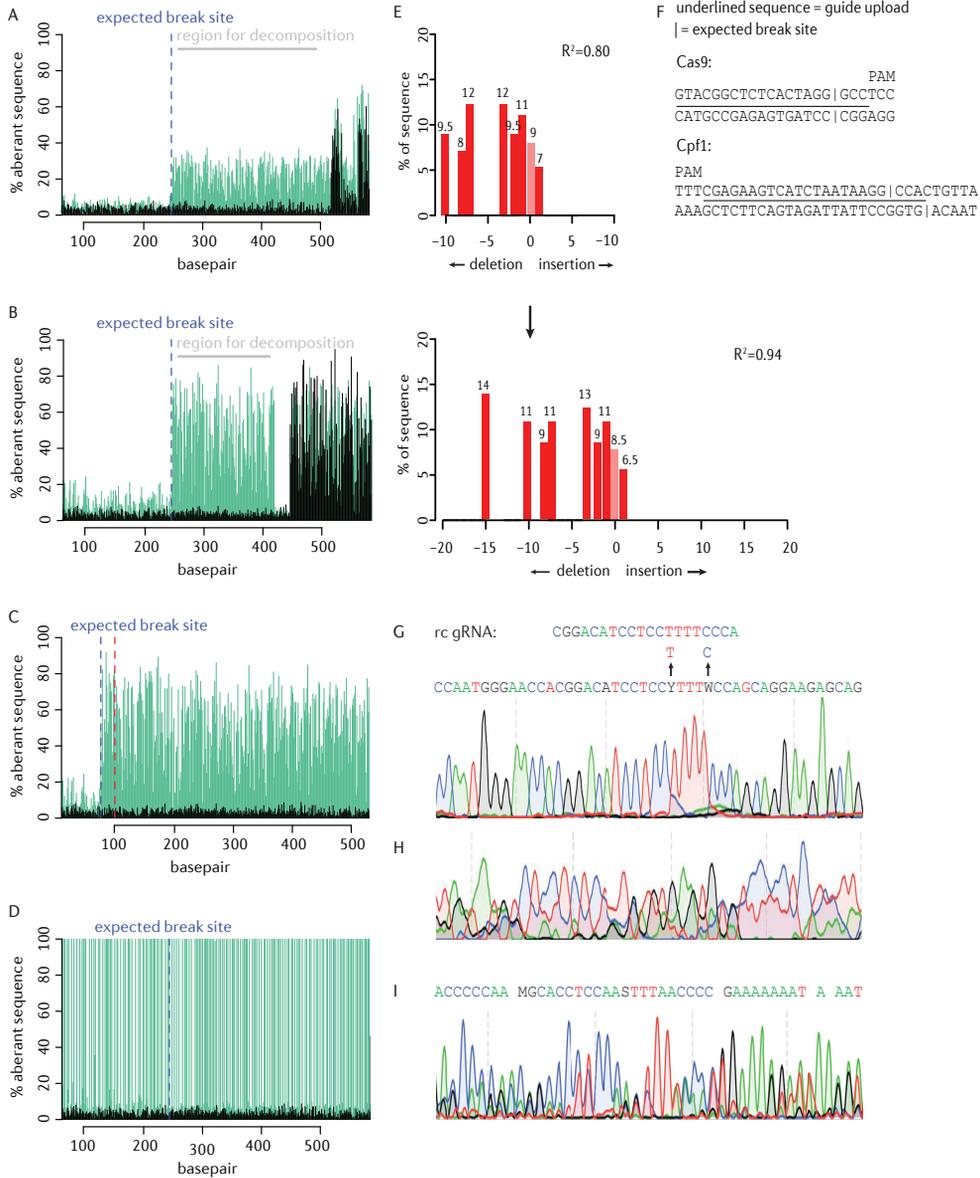


Figure 3: Troubleshooting with TIDE and TIDER. All parameters in TIDE(R) have default settings but can be adjusted if necessary. Different settings is often a remedy to solve error messages. (A-I) Examples of most common error messages with the recommended setting changes. (A-B) Avoid the decomposition 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 are 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).

the experimental sample, before and after the break site (SEE NOTE 1). It is recommended to check the chromatograms of the samples (FIGURE 3H) for poor sequencing quality. If so, these samples cannot be analysed 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 (FIGURE 3I). TIDE(R) gives a warning when the spacing between the nucleotides in the chromatogram of the sequence trace are 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).

11. *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).
12. *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.
13. *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.

ACKNOWLEDGMENTS

We thank Marcel de Haas, Stefano Manzo and Ruben Schep for critical reading of the manuscript.

AUTHOR CONTRIBUTIONS

EKB and BvS wrote the manuscript.

FUNDING

This work was supported by the Netherlands Organization for Scientific Research ZonMW-TOP grant 91211061, and European Research Council Advanced Grant 694466. 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

1. Brinkman, E.K., Chen T., de Haas M., Holland H.A., Akhtar W., van Steensel B., Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. *Molecular cell* 70, 801-813 e806 (2018).
2. Cong, L., Ran F.A., Cox D., Lin S., Barretto R., Habib N., Hsu P.D., Wu X., Jiang W. *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)* 339, 819-823 (2013).
3. Jinek, M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)* 337, 816-821 (2012).
4. Mali, P., Esvelt K.M., Church G.M., Cas9 as a versatile tool for engineering biology. *Nature methods* 10, 957-963 (2013).
5. Yoshimi, K., Kunihiro Y., Kaneko T., Nagahora H., Voigt B., Mashimo T., ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. *Nature communications* 7, 10431 (2016).
6. Inui, M., Miyado M., Igarashi M., Tamano M., Kubo A., Yamashita S., Asahara H., Fukami M., Takada S., Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system. *Scientific reports* 4, 5396 (2014).
7. Storici, F., Snipe J.R., Chan G.K., Gordenin D.A., Resnick M.A., Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing. *Mol Cell Biol* 26, 7645-7657 (2006).
8. Wang, Z., Zhou Z.J., Liu D.P., Huang J.D., Double-stranded break can be repaired by single-stranded oligonucleotides via the ATM/ATR pathway in mammalian cells. *Oligonucleotides* 18, 21-32 (2008).
9. Ma, H., Marti-Gutierrez N., Park S.W., Wu J., Lee Y., Suzuki K., Koski A., Ji D., Hayama T. *et al.*, Correction of a pathogenic gene mutation in human embryos. *Nature*, (2017).
10. Richardson, C.D., Ray G.J., DeWitt M.A., Curie G.L., Corn J.E., Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature biotechnology* 34, 339-344 (2016).
11. Horlbeck, M.A., Witkowsky L.B., Guglielmi B., Replogle J.M., Gilbert L.A., Villalta J.E., Torigoe S.E., Tjian R., Weissman J.S., Nucleosomes impede Cas9 access to DNA *in vivo* and *in vitro*. *eLife* 5, (2016).
12. Yang, L., Guell M., Byrne S., Yang J.L., De Los Angeles A., Mali P., Aach J., Kim-Kiselak C., Briggs A.W. *et al.*, Optimization of scarless human stem cell genome editing. *Nucleic acids research* 41, 9049-9061 (2013).
13. Brinkman, E.K., Chen T., Amendola M., van Steensel B., Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic acids research* 42, e168 (2014).
14. Brinkman, E.K., Kousholt A.N., Harmsen T., Leemans C., Chen T., Jonkers J., van Steensel B., Easy quantification of template-directed CRISPR/Cas9 editing. *Nucleic acids research*, (2018).
15. Green, M.R., Sambrook, Joseph, *Molecular Cloning A Laboratory Manual*. (2012), vol. 4.
16. Kunkel, T.A., Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proceedings of the National Academy of Sciences of the United States of America* 82, 488-492 (1985).
17. van Overbeek, M., Capurso D., Carter M.M., Thompson M.S., Frias E., Russ C., Reece-Hoyes J.S., Nye C., Gradia S. *et al.*, DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Molecular cell* 63, 633-646 (2016).
18. Dorsett, Y., Zhou Y., Tubbs A.T., Chen B.R., Purman C., Lee B.S., George R., Bredemeyer A.L., Zhao J.Y. *et al.*, HCoDES reveals chromosomal DNA end structures with single-nucleotide resolution. *Mol Cell* 56, 808-818 (2014).
19. Beumer, K.J., Trautman J.K., Mukherjee K., Carroll D., Donor DNA Utilization During Gene Targeting with Zinc-Finger Nucleases. *G3* 3, 657-664 (2013).

Chapter 5

KINETICS AND FIDELITY OF THE REPAIR OF
CAS9-INDUCED DOUBLE-STRAND DNA BREAKS

Eva K. Brinkman, Tao Chen, Marcel de Haas, Hanna A. Holland,
Waseem Akhtar and Bas van Steensel

Molecular Cell, 70, 803-813, 2018

ABSTRACT

The RNA-guided DNA endonuclease Cas9 is a powerful tool for genome editing. Little is known about the kinetics and fidelity of the double-strand break (DSB) repair process that follows a Cas9 cutting event in living cells. Here, we developed a strategy to measure the kinetics of DSB repair for single loci in human cells. Quantitative modelling of repaired DNA in time series after Cas9 activation reveals variable and often slow repair rates, with half-life times up to ~10 h. Furthermore, repair of the DSBs tends to be error-prone. Both classical and microhomology-mediated end joining pathways contribute to the erroneous repair. Estimation of their individual rate constants indicates that the balance between these two pathways changes over time and can be altered by additional ionizing radiation. Our approach provides quantitative insights into DSB repair kinetics and fidelity in single loci, and indicates that Cas9-induced DSBs are repaired in an unusual manner.

INTRODUCTION

The CRISPR/Cas9 nuclease system is a powerful tool for genome editing due to its efficient targeting of specific sequences in the genome (1). Cas9 endonuclease is directed by a guide RNA to a specific target site in the genome, where it induces a single double-strand break (DSB) (2-4). The break is subsequently repaired by the cellular DNA repair mechanisms that can introduce mutations in the target sequence (5). This application of Cas9 has become widely popular to generate mutant alleles of genes and regulatory elements of interest. Despite the broad application, the process of repair of Cas9-induced DSBs has been only partially characterized. For example, it is not known how long it takes before an individual Cas9-induced DSB is repaired, and how error-prone this process is.

Eukaryotic cells have two main pathways for DSB repair: classical non-homologous end joining (C-NHEJ) and homologous recombination (HR) (5). A large proportion of DSBs is repaired by C-NHEJ, which directly re-joins the two DNA ends. This type of repair is thought to be mostly perfect but may lead to insertions or deletions (indels) at the break site (6). However, estimates of the frequency at which these indels occur are a matter of debate. The accuracy of re-joining is subject of debate, and is thought to depend on the nature of the damaging agent and the end structures of the formed DSBs (6, 7). In contrast, HR is highly precise because it utilizes a homologous template sequence to restore the DNA sequences around the DSB (8). Apart from these two main pathways, there are alternative end joining (A-EJ) pathways that are thought to be highly mutagenic. One of these is microhomology-mediated end-joining (MMEJ), which uses short sequence homologies near the two ends, leading to characteristic small deletions (9). Current evidence indicates that multiple pathways can contribute to the repair of Cas9-induced DSBs (10, 11), but the interplay and the relative contributions have remained largely uncharacterized. Moreover, the fidelity of these pathways in the context of Cas9-induced breaks is still largely unclear.

Related to this, it is still unknown how quickly a Cas9-induced break is repaired. DSBs have been measured by comet assays (12) or pulsed-field gel electrophoresis, typically (13) after exposure to high doses of ionizing radiation that cause hundreds of DSBs per cell. Time courses of such bulk measurements in mammalian cells have consistently shown that DSBs are generally repaired with a half-life of 10-60 minutes (13-18). Computational modelling of such datasets has indicated that a small sub-fraction of DSBs is re-joined more slowly, over several hours (14, 15, 19, 20). A-EJ pathways are thought to be largely responsible for this population of more slowly repaired DSBs (21, 22). However, such bulk measurements of DSBs lack the sensitivity that is required to follow DSB repair at single loci.

Repair rates have also been inferred from immunofluorescent labelling of DSB markers

such as γ H2AX, which form transient foci at DSBs (23, 24). In irradiated mammalian cells, clearance of such foci typically takes on to three hours (24-26). Note that marker proteins at foci may linger at the DSB site after re-joining of the two DNA ends, hence the actual repair may be quicker, as is strongly suggested by the direct DSB measurements mentioned above. So far, the rate of repair of individual Cas9-induced DSBs has remained unknown due to limitations of the available methods.

In principle, accumulation of mutations over time can be used to infer repair rates. A crude estimate based on indel detection suggested that about 15 hours were necessary to repair the majority of the Cas9-induced lesions (27). However, precise quantitative kinetics of actual re-joining of DNA ends after the induction of a DSB at a single defined genomic location are missing. One challenge is that perfectly repaired junctions are indistinguishable from DNA that was never broken, which may lead to systematic errors in the rate constants. Furthermore, when sequence-specific nucleases are used to create a break, indels may only arise after multiple cycles of breaking and perfect repair, and thus the final mutation rate may grossly overestimate the error rate of individual repair events (7). Finally, with such nucleases the accumulation of indels is also dependent on the cutting rate. Here, we tackle these problems by a combination of mathematical modelling and highly accurate measurements of indel accumulation, focusing on Cas9-induced DSBs.

RESULTS

A kinetic model of DSB repair

We approached the process of repair of a Cas9-induced DSB as a simple three-state model (FIGURE 1A). In this model, the intact state is the original unbroken DNA sequence that can be recognized by the Cas9/sgRNA complex. After introduction of a DSB by this complex the DNA enters a reversible broken state. This state may be repaired perfectly, after which the DNA is susceptible to another round of cutting. Alternatively, an error-prone repair mechanism may introduce a small insertion or deletion (indel) at the break site. The latter results in an irreversible indel state that can no longer be recognized properly by the sgRNA and therefore cannot be cut again by Cas9 (see below for validation of this assumption). Hence, in this model there are three reaction steps: cutting, perfect repair, and mutagenic repair, each having a specific rate constant that we refer to as k_c , k_p and k_m , respectively (FIGURE 1A).

Our aim was to determine these key descriptors of the repair process for individual Cas9/sgRNA target loci. For this purpose, we captured the model in a set of ordinary differential equations (ODEs) describing the three reaction steps, with the rate constants as

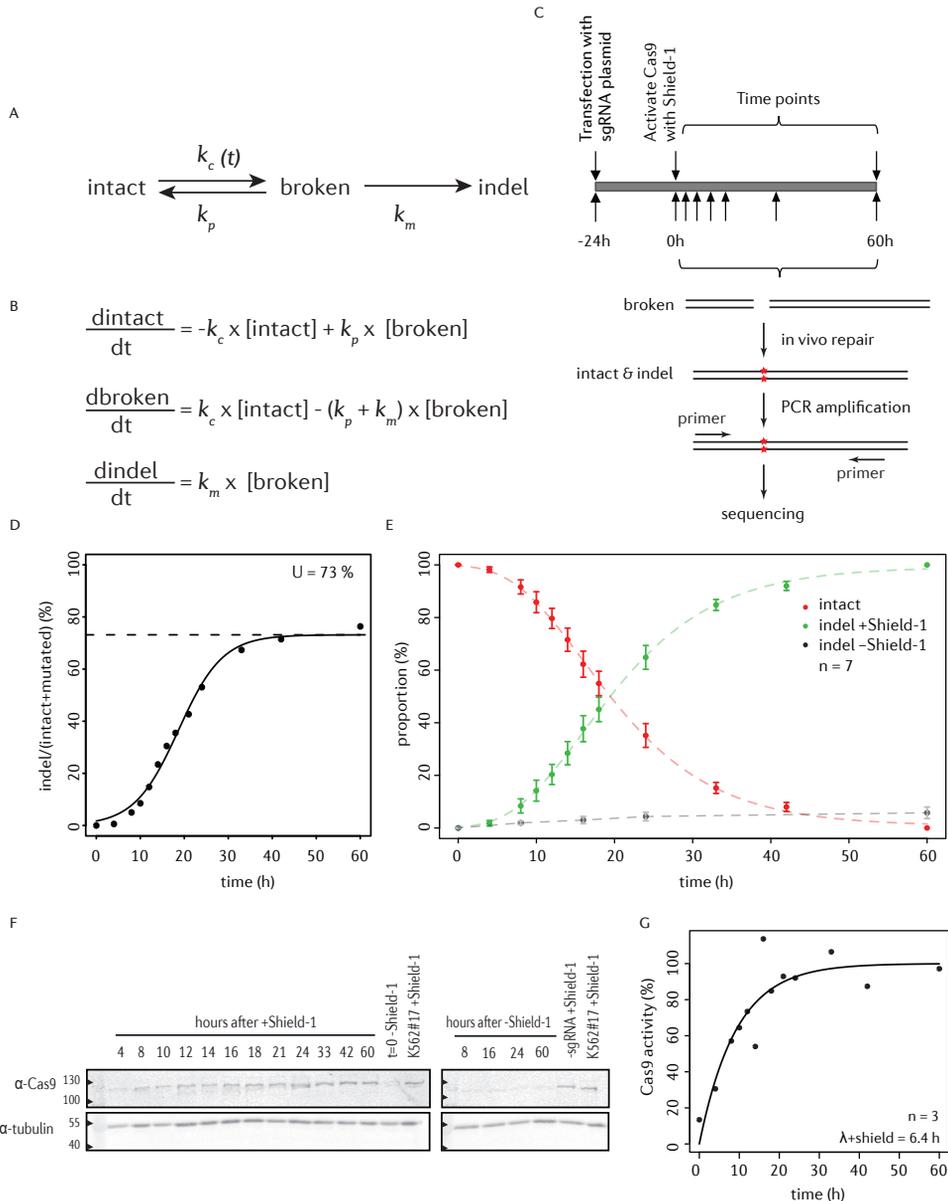


Figure 1: Quantitative analysis of Cas9-induced DSBs. (A) Proposed model of DSB repair based on stochastic transitions between intact, broken and indel states. k_p and k_m are rate constants of perfect and mutagenic repair, respectively; k_c is the rate constant of cutting by Cas9. The latter depends on Cas9 activation and is therefore denoted as $k_c(t)$. (B) ODEs describing the three reaction steps, with rate constants as parameters. (C) Outline of the experimental strategy; see main text. (D) Representative time course experiment, showing gradual accumulation of indels. A sigmoid curve was fitted to the data to determine the plateau level at late time points (dashed line), which reflects the transfection efficiency. (E) Relative proportions of intact (red) and indel (green) fractions at the LBR2 locus over time. The data points are normalized on to the total indel fraction to correct for the variation in transfection efficiency. Indel fraction in absence of Shield-1 is shown in grey. Average of 7 independent experiments; error bars represent the standard deviation and the dashed lines show fitted sigmoid curves.

(legend continued on next page)

parameters (FIGURE 1B). With this ODE model it is possible to simulate the relative abundance of the three states over time in a pool of cells after activation of Cas9. Such simulations show that activation of Cas9 generally leads to a gradual loss of intact DNA, a transient increase in the broken DNA state, and a gradual increase in the indel state until virtually all DNA is converted to the latter. However, the shape of each curve is determined by the rate constants (SUPPLEMENTARY FIGURE S1A-C). We reasoned that it should therefore be possible to determine the rate constants by fitting the ODE model to actual time series measurements of one or all three states after Cas9 induction.

Quantification of Cas9 cutting and repair rates

We set out to measure the accumulation of indels in cells in which a specific DSB was introduced by Cas9. To control the timing of DSB formation we established a clonal K562 cell line (K562#17) with a stably integrated construct that encodes a tightly controlled inducible Cas9 nuclease. To switch Cas9 activity on and off, we fused a ligand-responsive destabilizing domain (28) to Cas9. With the small ligand Shield-1, Cas9 can then be reversibly stabilized for transient DSB-induction. K562 cells are capable of activating DNA damage response upon DSBs, although the G1 checkpoint is affected due to a mutated *TP53* gene (29).

We transiently transfected K562#17 with a plasmid encoding a sgRNA targeting the *LBR* gene (sgRNA-LBR2). We previously found that this sgRNA effectively induces indels (30). Twenty-four hours after transfection we stabilized Cas9 by adding Shield-1. Flow cytometry analysis showed that cells 16 hours after damage had an ~10% increase in G2 population suggesting a modest check point activation (SUPPLEMENTARY FIGURE S5). We collected cells at various time points up to 60 hours after Cas9 induction, amplified a ~300-bp region around the sgRNA target site by PCR and subjected the products to next generation sequencing (NGS) to determine the intact and indel fractions (FIGURE 1C).

The results show a gradual accumulation of indels over time (FIGURE 1D), indicating that DSBs were introduced and repaired imperfectly. Towards the end of the time course the indel frequency reached a plateau of ~70%. This value corresponds approximately to the mean transfection efficiency (mean \pm S.D. = $67.6 \pm 11.9\%$, $n=7$), determined as the proportion of cells that express GFP after transfection with a GFP-expressing plasmid (SUPPLEMENTARY FIGURE S1D-E). We therefore assumed that the plateau value of ~70% is good representation of the total proportion of cells that received the sgRNA and underwent DSB induction and repair. After normalization for this transfection efficiency,

(F) Western blot analysis of Cas9 presence as a function of time. Tubulin was used as loading control. (G) The intensities of Cas9 antibody signal were determined by densitometry from time points of three individual western blots and normalized to a sample incubated for 60 hours with Shield-1 (lanes labelled 'K562#17+Shield-1' in (F)). An ODE fit was performed to determine the activity score of Cas9 in time.

the data were highly reproducible over 7 independent replicate experiments (FIGURE 1E). Indel accumulation was dependent on stabilization of Cas9 by Shield-1 (FIGURE 1E).

The sigmoid appearance of the measured indel time curves suggested a delayed onset of indel accumulation. This may be explained by delayed activation of Cas9 at the beginning of the time series, as indicated by Western blot analysis of Cas9 (FIGURE 1F). Assuming that the cutting activity of Cas9 is proportional to its abundance, we modified the computational model to incorporate the gradual increase of Cas9 levels as determined from the Western blot signals (FIGURE 1G). We assume that the transfected cells continuously express sgRNA over the duration of the time courses, because after transfection with a GFP-expressing vector the proportion of GFP-positive cells did not decline for at least 120 hours (SUPPLEMENTARY FIGURE S1F).

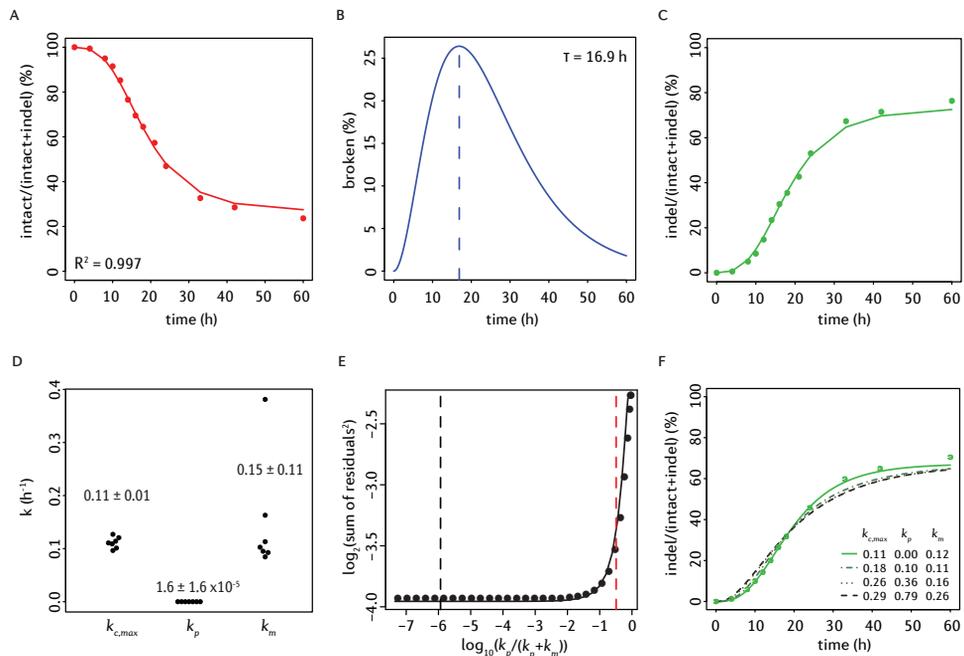


Figure 2: Estimation of cutting and repair rates. Representative time series traces of the intact (A), broken (B) and indel (C) fractions. Measured data (dots) are overlaid with the ODE model fit (solid lines). The percentage of intact and indel traces are relative to the total. The broken fraction is estimated by the model on the basis of the intact and indel measurements. τ indicates time of the largest amount of broken DNA. (D) Distributions of rate constants from 7 independent experiments (values indicate mean \pm SD). (E) Estimate of the upper confidence bound of the proportion of perfect repair. Fitting residual errors are shown for various fixed $k_p/(k_p+k_m)$ ratios (7 combined time series). Black dashed line marks the optimal fit; red dashed line marks the $k_p/(k_p+k_m)$ ratio above which the fit becomes significantly worse than the optimal fit ($P < 0.1$, one-sided F-test). This corresponds to $k_p/(k_p+k_m) = 0.28$. For each $k_p/(k_p+k_m)$ ratio, $k_{c,\text{max}}$ and k_m were adjusted to produce the best fit. (F) Examples of model fitting with various combinations of k_p and k_m forced to values that would correspond to models of rapid cycling of cutting and perfect repair (dashed curves). Green dots are measured values, green curve shows unrestrained optimal fit.

Cutting and repair rate constants in the LBR gene

Next, we fitted the set of ODEs to the measured indel curves for sgRNA-LBR2 (FIGURE 2A-D). Based on 7 independent replicate experiments, this yielded a cutting rate $k_c = 0.11 \pm 0.01 \text{ h}^{-1}$ at maximum Cas9 expression. This corresponds to a cutting half-life (i.e., the time that would be required to cut 50% of the available target sequences in the absence of repair) of ~6 hours. Cutting by Cas9/sgRNA-LBR2 in our system is thus a rather slow process.

Our model fit estimated the rate constant for imperfect repair to be $k_m = 0.15 \pm 0.11 \text{ h}^{-1}$. This corresponds to a half-life of broken DNA of 4.6 hours. Surprisingly, the rate constant for perfect repair was estimated to be $k_p = 1.6 \pm 1.6 \times 10^{-5} \text{ h}^{-1}$, which is about ten thousand times slower than imperfect repair. This suggests that virtually all repair events at this locus result in the formation of indels, while perfect repair is very rare.

Robustness of the model

On average, the goodness of fit between the model and the measured data was $R^2 = 0.995$. We considered that the parameter estimates could be strongly influenced by the modelling of the Cas9 induction. To test this, we also modelled a simple step function with various time delays of Cas9 activation. Although the results were quantitatively slightly different (SUPPLEMENTARY FIGURE S1G-J), the main conclusion remained that repair at this locus is slow and error-prone.

While these results point to robustness of the modelling, we were surprised to find the extremely low rate of perfect repair. To check whether the model with the low perfect repair rate is indeed the most optimal fit to the data, we conducted a ‘parameter sweep’ survey in which we imposed different fixed perfect/mutagenic repair ratios. Analysis of the fitting residuals indicated that low $k_p/(k_p+k_m)$ ratios ($\sim 10^{-6}$) indeed yield the best fit (FIGURE 2E-F). However, the difference with higher ratios was rather minor, and statistical testing revealed that only at $k_p/(k_p+k_m)$ ratios > 0.28 the fit became significantly poorer ($P < 0.1$, F-test) than at the initially estimated ratio of 10^{-6} . We therefore conservatively conclude that the contribution of perfect repair at this locus can be at most 28%, although lower ratios are more likely to be correct. The other parameter values (k_c , k_m and the predicted broken fraction) showed only minor variation within this range (SUPPLEMENTARY FIGURE S2A-E), further attesting to the robustness of the model.

From this parameter sweep analysis follows that the indel accumulation curve is not compatible with a very rapid cycle of cutting and perfect repair with only occasional imperfect repair. In such a scenario $k_p/(k_p+k_m)$ would be close to 1. As mentioned above, according to the F-test this yields a significantly poorer fit compared to lower $k_p/(k_p+k_m)$ ratios (FIGURE 2E-2F). Increased cutting rates also lead to poorer fits (SUPPLEMENTARY FIGURE S2F).

Experimental validation

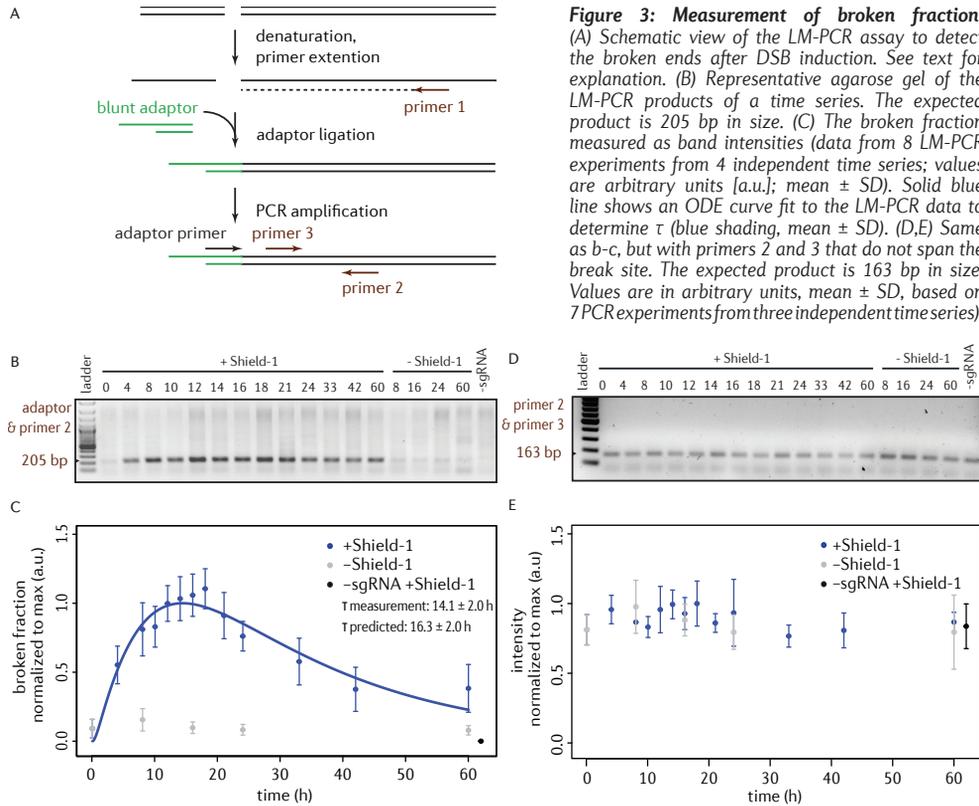
We validated the underlying assumptions and results of the modelling with several independent biological assays. First, we tested our assumption that the reverse reaction from the indel state to the broken state cannot occur (FIGURE 1A). For this purpose we isolated several clonal cell lines that had acquired one or more indels in the target site and lacked wild-type sequences (SUPPLEMENTARY FIGURE S4A). We then re-transfected three of these cell clones with sgRNA-LBR2 and again activated Cas9. Despite prolonged re-exposure to Cas9/sgRNA-LBR2, we could not detect any change in the indels present in each clone (SUPPLEMENTARY FIGURE S4B). We conclude that the target site, once it has acquired an indel, is not recognized again by the same sgRNA.

Second, we verified the kinetics of the broken state as predicted by the computational model (which was based on intact and indel frequencies only). The model predicts that the broken fraction peaks at 16.4 ± 1.7 h, with a maximum of $19.9 \pm 6.7\%$ broken DNA (mean \pm SD, $n = 7$; FIGURE 2B). To verify this, we established a variant of the ligation-mediated PCR assay (LM-PCR) for the quantification of DNA breaks at a defined location (31, 32). In this assay, we first denature the DNA and subject it to a primer extension reaction using a primer near the break site. This ensures that all cleavage sites are converted into blunt ends, even if resection of the broken ends has occurred. Next, an adaptor is ligated to the blunted DNA end, followed by PCR with one primer near the break site and a second primer that is complementary to the adaptor sequence (FIGURE 3A). When analysed on agarose gels, the samples from cells treated with sgRNA and Shield-1 yielded a band of the expected size (FIGURE 3B), and the band intensity was sufficiently linear with input across the measured range (SUPPLEMENTARY FIGURE S3F). Analysis of several time series showed that the band intensity increased until 14.1 ± 2.0 h (mean \pm SD, $n = 4$) hours after Cas9 induction, and then decreased again (FIGURE 3B-C). This is in agreement with the peak time of the broken state as predicted by the model fitting (cf. FIGURE 2B). As expected, a control primer pair not spanning the break site showed stable signals over the time course (FIGURE 3D-E; SUPPLEMENTARY FIGURE S3G). We obtained similar results when probing the other end of the same DSB (15.6 ± 2.2 h, $n = 3$; SUPPLEMENTARY FIGURE S3B-E, H-I).

Furthermore, the parameter sweep shows that when $k_p/(k_p+k_m)$ approaches 1, the predicted amount of broken DNA at peak time becomes so low ($<1\%$) that we would be unable to measure it by ligation-mediated PCR (SUPPLEMENTARY FIGURE S2D). Thus, a model consisting of a very rapid cycle of cutting and perfect repair with only occasional imperfect repair is not compatible with our ability to detect broken DNA.

Two repair pathways active at one locus

Next, we took a closer look at the indels that were generated. Repair of Cas9-induced DSBs produces non-random indel patterns that are specific for the sgRNA (11, 30). In our



experiments sgRNA-LBR2 yielded predominantly a deletion of 7 bp or an insertion of 1 bp (FIGURE 4A). After 60 hours of Cas9 induction the +1 insertions reached a frequency of $59.6 \pm 4.2\%$, while the -7 deletions accumulated to $18.0 \pm 2.3\%$ of the total indel pool (FIGURE 4C). Analysis of 20 clonal lines derived from single cells in which Cas9/sgrNA-LBR2 had been transiently active indicated that the +1 insertion is the predominant mutation but can co-occur with the -7 deletion on other alleles in the same cell (SUPPLEMENTARY FIGURE S4A).

Different indels can be the result of different repair pathways, e.g. C-NHEJ or MMEJ (11, 33). To explore whether this may be the case for the -7 and +1 indels, we added an inhibitor of DNA-PKcs (NU7441) to the cells. DNA-PKcs plays an essential role in the C-NHEJ pathway but not in MMEJ (34). We found that in the presence of $1 \mu\text{M}$ NU7441 the proportion of the -7 deletion events increased by 3-fold while the +1 insertion diminished by about 2-fold (FIGURE 4A-C). The presence of NU7441 inhibitor did not affect cell viability (SUPPLEMENTARY FIGURE S4C). These results indicate that the +1 insertion is the result of C-NHEJ, while the -7 deletion is not. MMEJ makes use of microhomologies near the broken ends (18, 35). Sequence analysis revealed that the -7 deletion fraction consists of two types of deletions

that occur in an approximate ratio of 1:2 (FIGURE 4D). Both types can be explained by recombination through 3-nucleotide microhomologies, which strongly points to MMEJ as the responsible pathway for the formation of the -7 indel. We conclude that at least two different repair pathways are active and lead to distinct types of mutations at one specific break site.

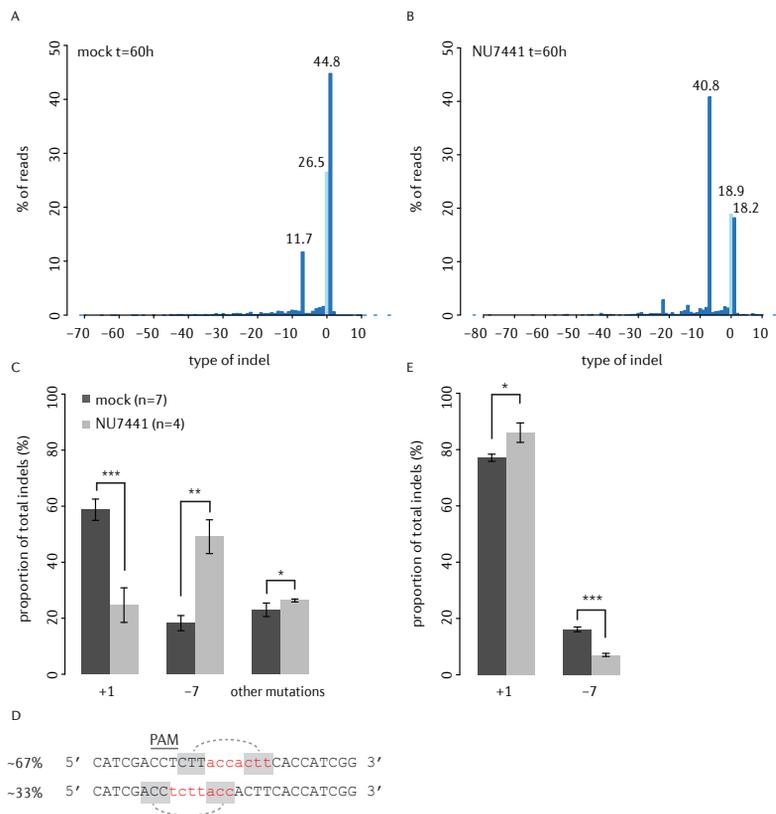


Figure 4: Multiple repair pathways active at one locus. (A,B) The spectrum of indels and their frequencies at the LBR2 locus at time point $t = 60$ h, in cells cultured without (A) or with (B) $1 \mu\text{M}$ NU7441. A representative experiment is shown. Light-blue bar: wild-type sequence; dark-blue bars: indels. (C) Frequencies of -7 and +1 indels in the presence (black, $n=7$) and absence (grey, $n=4$) of NU7441. All series are normalized to the total indel fraction. Asterisks indicate P -values according to Student's t -test: * $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$. (D) The -7 deletions consist of two types; red nucleotides mark the deleted DNA. Shaded nucleotides show possible models for microhomology-mediated repair. Percentages indicate the proportion of observed -7 sequence reads. (E) TIDE analysis of +1 insertion and -7 deletion indels after exposure of the cells to 10 Gy of IR at the time of Cas9 induction.

Interestingly, we found that the ratio between the +1 and -7 indels changes in favour of the +1 insertion when 10 Gy of ionizing radiation (IR) damage was administered just prior to Cas9 induction. In particular, the -7 deletion fraction decreased about 2-fold (FIGURE 4E). Cell viability did not differ between the control and irradiated sample (SUPPLEMENTARY FIGURE S4D). The shift in pathway utilization can be either due to additional breaks elsewhere in genome (which may sequester components of the MMEJ pathway), a cell cycle arrest, or a combination thereof.

Delayed activity of MMEJ

Having established that two types of indels are largely the result of separate pathways, we decided to study the kinetics of these pathways in more detail by tracking the +1 and -7 indel frequencies over time. We modified the ODE model by incorporating separate k_m rate constants for each type of indel (FIGURE 5A). We then fitted this model to the +1 and -7 indel time series data. This yielded a good fit of the +1 curve (mean $R^2 = 0.99$), but a poorer fit of the -7 curve (mean $R^2 = 0.93$; FIGURE 5C). The deviation of the fitted curve is mostly due to a delay in the -7 indel appearance (FIGURE 5B). As a consequence, our estimate of the k_m for the -7 indel is less accurate, but we can conclude that the MMEJ pathway exhibits a delayed onset compared to C-NHEJ. Thus, the k_{-7} rate appears to increase over time, rather than be constant. Possibly the MMEJ pathway is only activated when C-NHEJ fails to repair a DSB, as has been proposed previously (13, 21, 35).

We then tested whether the model could be improved by explicitly including such a delayed onset. For simplicity we assumed a linear increase in the production rate of the -7 deletion over time, starting with a rate $k_{-7,t=0h}$ immediately after Cas9 induction and increasing to $k_{-7,t=60h}$ at the end of the time course (FIGURE 5D). Indeed, a better fit of the -7 indel curve (mean $R^2 = 0.99$) was obtained (FIGURE 5E). The model fitting estimates k_{-7} to be nearly zero at the onset of Cas9 induction, while at the end of the time course it is $0.09 \pm 0.09 \text{ h}^{-1}$, which approaches the activity of the +1 indel repair ($0.12 \pm 0.10 \text{ h}^{-1}$; FIGURE 5F). These results strongly suggest gradual activation of MMEJ over time.

Interplay between C-NHEJ and MMEJ

We wondered whether this gradual increase in MMEJ rate is somehow due to competition with the C-NHEJ pathway. To test this, we performed time course experiments in the presence of NU7441 and repeated the computational modelling. As expected, in the presence of the inhibitor k_{+1} is reduced dramatically (~6-fold; FIGURE 5F; SUPPLEMENTARY TABLE 3), while the cutting rate k_c as well as the perfect repair rate k_p are virtually unaltered. Strikingly, in the presence of NU7441 $k_{-7,t=0h}$ became about 7-fold higher than in the absence of the inhibitor, while $k_{-7,t=60h}$ remained largely unaffected. Thus, inhibition of DNA-PKcs leads to a more rapid engagement of MMEJ soon after the DSB is introduced.

However, in the presence of NU7441, MMEJ does not fully compensate for the loss of C-NHEJ. The total repair rate is lower, as k_m is reduced and k_p remains close to zero (FIGURE 5F). A logical prediction is that it takes more time for DSBs to be repaired. The model as well as actual measurements show that in the presence of NU7441 the peak time of DSBs is delayed (compare FIGURE 3C & 5C), although this is not statistically significant for the measured data.

Cutting and repair rates in three other loci

To investigate whether the rate constants are locus-specific, we performed kinetics

experiments for three additional loci. We designed a sgRNA (sgRNA-LBR8) targeting another sequence in the *LBR* gene that is 169 bp upstream from sgRNA-LBR2; a previously reported sgRNA to target the *AAVS1* gene (4); and a sgRNA that targets an intergenic locus on chromosome 11 (chr11). Each target locus shows a different indel spectrum after 60 hours (SUPPLEMENTARY FIGURE S6B, H, K). For each sgRNA we conducted multiple time series measurements (FIGURE 6A-H).

For the three new loci the indel frequency did not fully reach a plateau at the end of the time course (FIGURE 6C, E, G). This compromised the robustness of the rate constant estimates, since our algorithm used this plateau to estimate the transfection efficiency of individual experiments. We therefore employed a slightly modified computational strategy. We combined the data of all replicate experiments and performed parameter fitting in two iterations. In the first iteration, we conducted a parameter sweep for the transfection efficiency. We chose the efficiency that yielded the lowest residuals (see Methods). In the second iteration, we used this value and applied a standard 1,000-fold bootstrapping approach in order to estimate confidence intervals for each parameter of the kinetic model. The results are summarized in FIGURE 6A-H & SUPPLEMENTARY TABLE 4.

The confidence intervals of the fitted rate constants are relatively large, and hence the results should be interpreted with some caution. Nevertheless, for the LBR2 locus, this alternative computational strategy yielded parameter estimates that were similar to those from the original strategy. Across the four loci, the estimated cutting rates (k_c) varied only about 2-fold. The overall repair rates (perfect and imperfect repair combined) vary over an ~8-fold range and correspond to half-lives of approximately 8.8 h (LBR8), 1.4 h (AAVS1) and 10.7 h (chr11), as compared to 3.9 h for the LBR2 locus. These values indicate that repair rates of Cas9-induced DSBs are variable and often slow.

Like for LBR2, the estimated rates of perfect repair (k_p) are very low for AAVS1 and chr11. In contrast, perfect repair in locus LBR8 is frequent, making up 76% of all repair events. For LBR8 the fitted models predicted a slower clearance of broken DNA compared to LBR2 (FIGURE 6B, D). Directly measured time courses of the LBR8 broken state are in agreement with this, with a τ of 28.9 ± 17.8 h for LBR8 compared to 14.7 ± 2.1 h for LBR2 (mean \pm SD; measurements for both ends combined; $P = 0.005$, Wilcoxon test) (SUPPLEMENTARY FIGURE S6D-F). Together, these results indicate that the error rate of the repair of Cas9-induced DSBs is locus-dependent and often high.

Tight binding of Cas9 after cutting may explain erroneous and slow repair

We sought an explanation for the relatively erroneous and slow repair that we observed. It was reported that the Cas9 can remain attached to the broken DNA ends after cutting *in vitro* (36), but it is not known how general this behaviour is. We therefore tested this

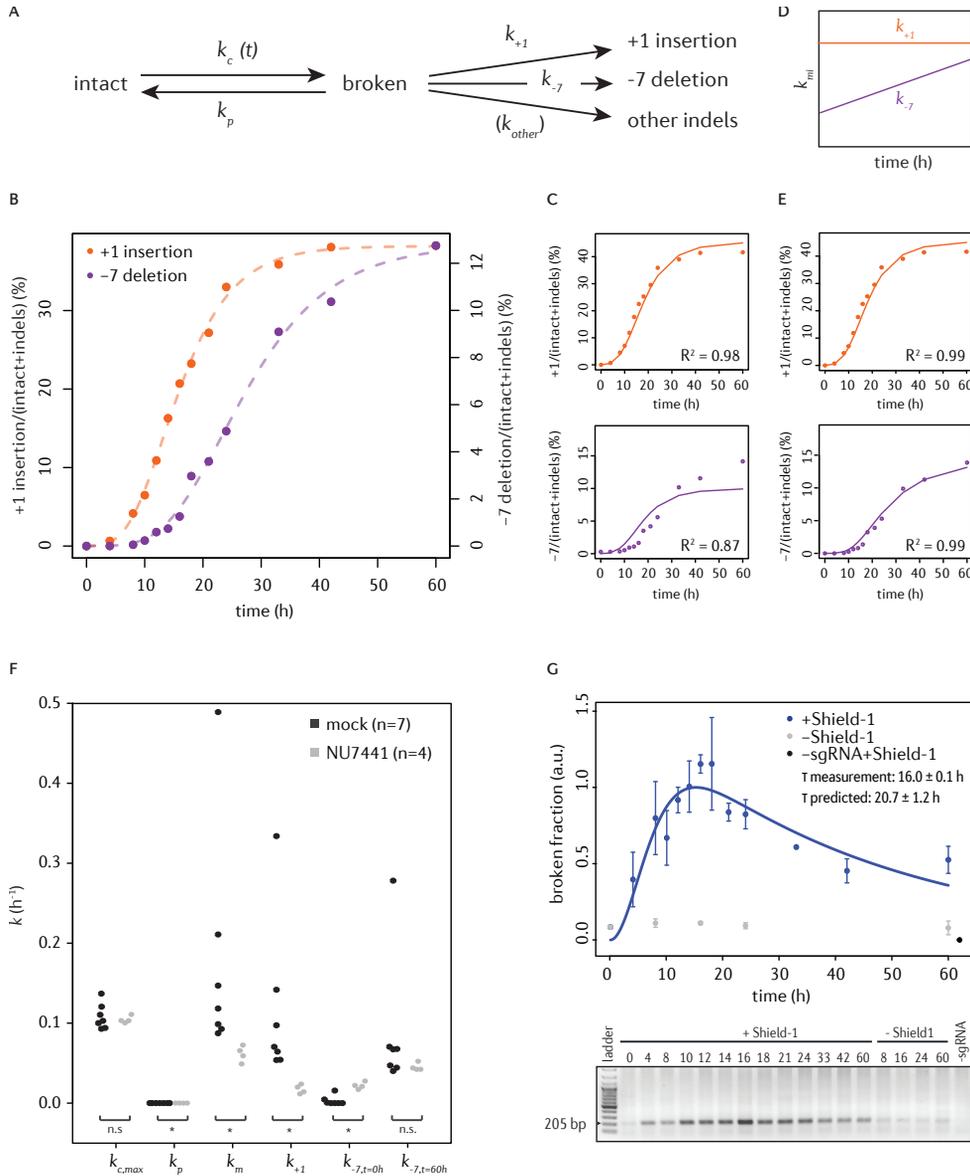


Figure 5: MMEJ has slower repair kinetics than C-NHEJ. (A) Kinetic model of Cas9-induced DSB repair assuming that each type of indel is generated with a specific repair rate. (B) +1 insertion and -7 deletion accumulation in a representative time series. Dashed lines show a Gompertz sigmoid fit. (C) Same measurements as in (B) with the multi-indel ODE model fit (solid lines) for the +1 (top) and -7 (bottom) indels. (D) Cartoon representation of hypothesized time dependency of the k_{+1} and k_{-7} rates added to the fitted model in (E). (E) Time dependent ODE model fit of the +1 insertion (top) and -7 deletion (bottom) as proposed in (D). (F) Rate constants without (black) and with (grey) NU7441. The total mutagenic repair rate is represented by k_m ; rate constant for the +1 insertion is k_{+1} and rate constants for the -7 deletion are $k_{-7,t=0h}$ and $k_{-7,t=60h}$ at the start and end of the time series, respectively. n indicates the number of time series. Asterisks indicate P -values according to Wilcoxon test: * $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$. (G) Broken fraction measurements in presence of NU7441 of 2 independent time series, similar to FIGURE 3B-C.

directly for three DNA loci for which we had determined the kinetic rate constants. By PCR we first produced double-stranded DNA fragments of 600-1,000 bp consisting of precisely the same sequences as the three loci. We then incubated each fragment *in vitro* with the respective Cas9/sgRNA complex to induce DSBs, and investigated the reaction products by agarose electrophoresis (FIGURE 6I, J).

The PCR product treated with Cas9/sgRNA-LBR2 showed the expected digestion products, but the smallest fragment was underrepresented while a broad smear indicated aberrant migration of the DNA. After heat denaturation of Cas9/sgRNA the smear disappeared and the smaller digestion product became clearly visible. For sgRNA-AAVS1 and sgRNA-chr11 even more pronounced effects were observed: without denaturation the digested DNA appeared largely unbroken and the bands were shifted upwards, but after heat treatment it became clear that most of the DNA was in fact correctly digested. Together, these results indicate that the Cas9/sgRNA complex remains bound to the DNA ends after cutting, even when incubated overnight (FIGURE 6J). In the case of sgRNA-LBR2 it appears that this binding occurs primarily at one DSB end, while for the other two sgRNAs Cas9 remains bound to both ends. As has been suggested (37), post-cutting adherence of Cas9 to the DNA ends may impair the repair process, which could explain the slow and erroneous repair.

Repair fidelity after double cutting

Finally, we investigated the fidelity of DSB repair upon induction of Cas9 in combination with two co-transfected sgRNAs that target adjacent sequences. If the two cuts are made simultaneously then the intermediate fragment may be lost, after which the two remaining ends are joined by the DSB repair machinery. We amplified the resulting junctions by PCR and sequenced them in order to determine the error rate of the repair process, i.e., the frequency at which indels occur at the junction. Such a double-cut strategy has been used previously in combination with I-SceI (35, 38). Importantly, once the two ends are joined perfectly, they cannot be cut again because the new junction is not recognized by either of the two sgRNAs. This assay therefore complements our kinetic modelling of single-cut repair, in which cycles of repeated cutting and perfect repair were theoretically possible.

We designed five sgRNAs targeting a second DSB site ~110-300 bp upstream or downstream of the sgRNA-LBR2 target site. One of these was sgRNA-LBR8. We co-transfected each sgRNA together with sgRNA-LBR2, induced Cas9 expression and harvested cells after 60 hours. PCR amplification followed by next generation sequencing uncovered all intermediates and end-products that could be expected, such as junctions resulting from excision events as well as indels at one or both of the two cutting sites (FIGURE 7A-B).

In those DNA molecules in which excision was successfully followed by repair, re-joining had occurred with highly varying degrees of fidelity. Depending on the combination of sgRNAs

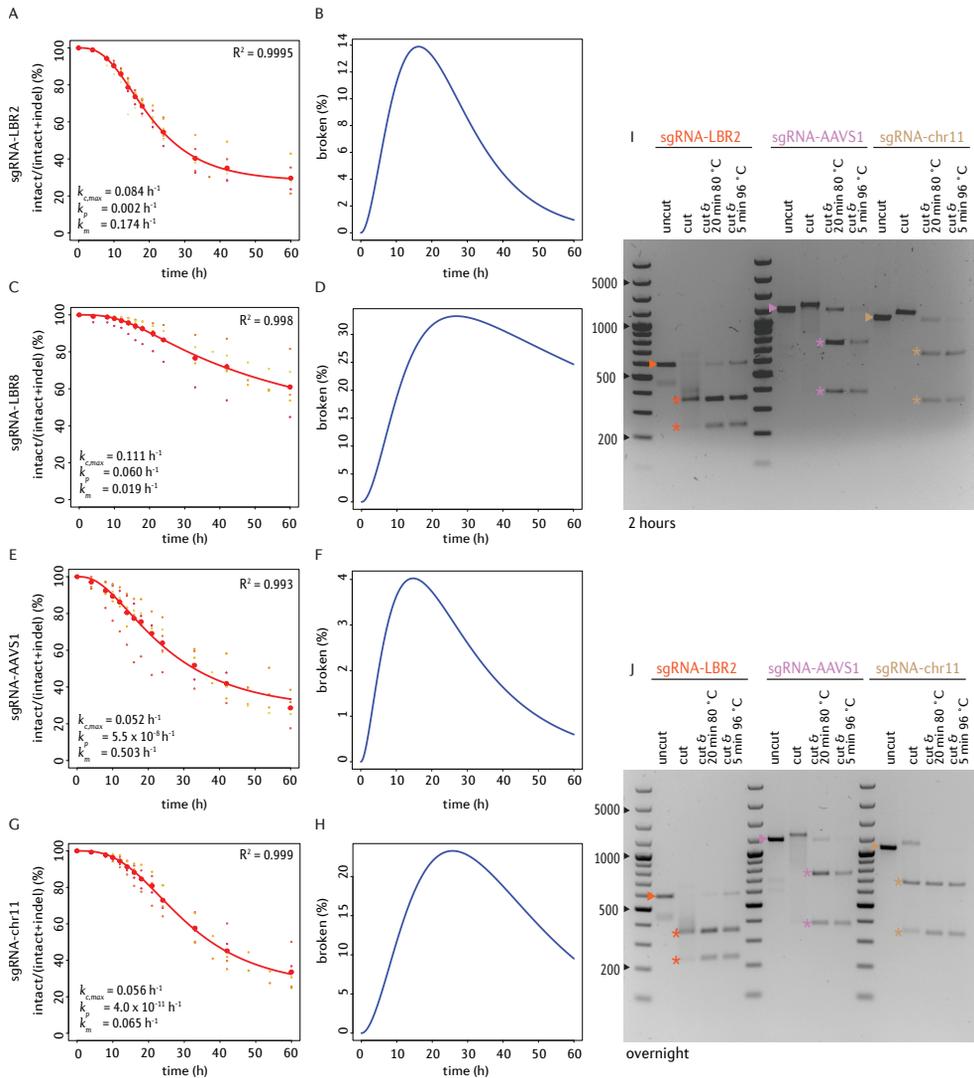


Figure 6: ODE modelling of DSBs at additional loci; Cas9 remains bound to broken ends. (A-H) Time series of Cas9 cutting and repair at four loci (LBR2 (A,B), LBR8 (C,D), AAVS1 (E,F), intergenic region on chromosome 11 (G,H)). (A,C,E,G) Indel abundance of all individual replicates are shown in small dots where each replicate has a unique color. The mean of each time point is shown as big red dots. Fitted models are shown as red curves. (B,D,F,H) Predicted broken fraction curves. (I,J) In vitro digestion of DNA fragments by Cas9 and one of the sgRNAs targeting LBR (orange), AAVS1 (purple) or intergenic region on chromosome 11 (brown), respectively. DNA was incubated for either 2 hours (I) or overnight (J) in the absence (uncut) or presence (cut) of Cas9/sgRNA. In some samples Cas9 was subsequently denatured by two different heat treatment protocols as indicated. The expected band for the intact DNA is marked by an arrowhead and the expected digestion products are marked by asterisks.

tested, we found that 25-95% of these junctions were imperfect (FIGURE 7C), as indicated by the occurrence of indels that ranged from 1 to ~20 bp in size. In 4 out of 5 tested double-cut combinations we detected at least 2-fold more excised fragments than indels at single

break sites (FIGURE 7B). This implies that the second cut typically occurred before the first cut had been repaired. This is in agreement with our modelling results indicating that repair of Cas9-induced breaks is often a relatively slow process.

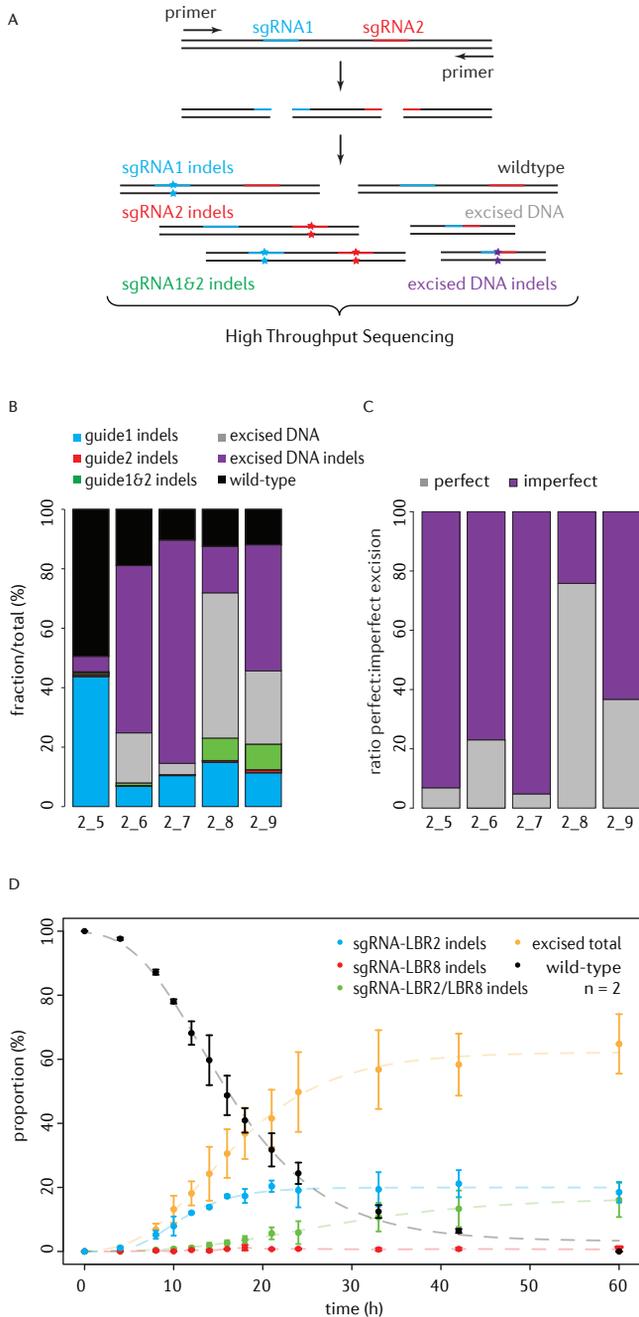


Figure 7: DSB repair fidelity after double cutting. (A) Repair products that may result from Cas9 in combination with two sgRNAs that target adjacent sequences. All products are amplified by PCR and detected by next generation sequencing. (B) Quantification of the various repair products 60 hours after Cas9 induction in the presence of sgRNA-LBR2 and one of 5 sgRNAs (labelled 5 thru 9) that target a sequence within 110-300 bp from the LBR2 site. (C) Same as (B), but highlighting the relative proportions of perfect and imperfect junctions among the repaired excision events. (D) Relative proportions of indels at sgRNA-LBR2 target site (blue), indels at sgRNA-LBR8 target sites (red), indel at both at sgRNA-LBR2 and sgRNA-LBR8 target sites (green), excised DNA (yellow) and wild-type (black) fractions over time. The data points are normalized to the total mutation fraction to correct for the variation in transfection efficiency. Average of 2 independent experiments; error bars represent the standard deviation. Dashed lines show fitted sigmoid curves.

Because the combination sgRNA-LBR2/sgRNA-LBR8 yielded an exceptionally high frequency of perfect excisions (~75%), we studied these repair events in more detail by generating time course data (FIGURE 7D). This revealed that accumulation of excisions took approximately 15 hours to reach 50% of the plateau level. This underscores again that repair of breaks induced by Cas9 at these sites is slow. Furthermore, indels at the LBR8 site alone were virtually undetectable throughout the time course, while indels at both sites or at the LBR2 site alone did accumulate. In part, this may be explained by the relatively high frequency of perfect repair that we estimated for LBR8 in the single-cut kinetics experiments. Another, not mutually-exclusive, explanation may be that repair of the LBR2 site is quicker than at the LBR8 site, which is also consistent with our parameter estimates based on single cuts (SUPPLEMENTARY TABLE S4).

DISCUSSION

DNA repair kinetics

Our measurements and modelling of the kinetics of re-joining of broken DNA ends after a Cas9-induced lesion indicate that the rate of DSB repair is variable and relatively slow. Moreover, our results indicate that the repair process tends to be error-prone. A recent study using a novel inducible Cas9 system also reported diverse kinetics profiles, but did not determine rate constants (39).

Across the four tested loci, we observed repair half-life times of 1.4, 3.9, 8.8 and 10.7 h. The latter three of these estimates are much slower than previous estimates based on bulk detection of DSBs after ionizing irradiation, which rarely exceeded 1 hour (see Introduction). Most of these previous studies were performed after a pulse of ionizing radiation ranging from 1-40 Gy, which creates dozens to hundreds of DSBs. Here, we introduce only a few DSBs per cell. We cannot rule out that large numbers of simultaneous DSBs accelerate the repair process. However, this seems not very likely because in the presence of 10 Gy IR damage in addition to the Cas9-induced DSB, we see only a small shift in repair pathway choice towards the faster C-NHEJ pathway.

Another factor that could contribute to the difference in repair rates is the possible adherence of Cas9 to the DNA ends after cleavage, which could prevent access by the repair machinery (37). In accordance with previous studies (36, 37, 40) our *in vitro* data show that Cas9 remains tightly bound to one or both DNA ends after cutting, and detachment could only be achieved by protein denaturation. *In vivo*, catalytically inactive Cas9 was also found to tightly bind to its target DNA (41), with a dwell time of about 2 hours (42).

DNA repair fidelity

It has been suggested that the genome in a human cell may be hit by as many as 10-50 DSBs per day (6, 43). Yet, in the genome of skin cells of a 55-year old individual only about 2000 small indels were detected by deep sequencing (44). In yeast and fly genomes indels accumulate at frequencies of roughly 10–11/base/generation (45, 46). These data suggest that repair of naturally occurring DSBs is highly precise. Ligation of mammalian V(D)J recombination junctions has also been found to have low error rates (7, 47). In this light, our estimated error rates in the range of 20-100% per break event seem rather high. This raises the possibility that repair of Cas9-induced DSBs is not representative for naturally occurring DSBs.

Repair fidelity could be affected by the adherence of Cas9 to broken ends, but also by the structure of the DNA ends themselves. High precision of C-NHEJ was found when DNA breaks were introduced by the I-SceI nuclease (35). I-SceI makes a staggered cut and leaves a 3' overhang, while Cas9 generates blunt ends (2-4). Blunt ends were shown *in vitro* and in wild-type yeast to be preferentially joined imprecisely (48, 49). Variants of Cas9 protein that generate DSBs with different overhangs (the nickases N863A and D10A) also resulted in differences in repair (10, 50). Thus, the high rates of imperfect repair that we observe may in part be related to the blunt ends created by Cas9. Furthermore, the local chromatin environment of the targeted loci may affect the repair outcome, because recruitment of repair factors is dependent on the chromatin environment (51).

The low rates of perfect repair that we observed also imply that HR plays only a minor role in the repair of Cas9-induced breaks, even though K562 cells are proficient in HR (52). HR frequencies in combination with Cas9 have been estimated to be up to 4-15% (10, 50), but these frequencies are likely to represent an overestimate because they were measured in the presence of a donor template that was either closely linked in *cis* or provided in excess by transient transfection. This also indicates that the accuracy of our broken fraction assay is unlikely to be compromised by substantial amounts of extensively resected ends that could escape detection in this assay.

Repair fidelity at double DSBs

The proportion of perfect junctions in our double-cut assays was also variable, and depending on the precise combination of sgRNAs used. In combination with sgRNA-LBR2, we observed the lowest error rate for sgRNA-LBR8, which also triggers mostly perfect repair when used in a single-cut kinetics assay. We note that repair of two nearby breaks may be more complex than of a single break. For example, the first break may trigger local chromatin changes such as phosphorylation of H2AX, which may in turn alter events at the second site, such as cutting rate and the recruitment of specific repair complexes.

Furthermore, Cas9 may linger on one or both of the ends that are to be joined. This may depend on the orientation of the PAM sites, or on other sequence features of the sgRNA used, but we have not been able to identify a predictive feature in the pairs of sgRNAs that we tested. Others have observed that sgRNA pairs resulted in high levels of precision repair (53, 54), although Geisinger *et al.* considered some single nucleotide indels as perfect repair and may have underestimated error rates.

Different repair pathways

We found that at one single genomic location (LBR2) both C-NHEJ and MMEJ can repair DSBs with different kinetics. Overall, MMEJ operates with lower rates than C-NHEJ. This is in agreement with previous findings that indicate that C-NHEJ is the primary repair pathway (55). We observe that the lower rate is mainly because MMEJ exhibits a delayed onset compared to C-NHEJ, rather than a reduced activity. Possibly the C-NHEJ system initially prevents access of the MMEJ pathway to the DSB; only after several hours, if C-NHEJ has failed to repair the break, the MMEJ pathway is allowed to engage. In contrast, upon inhibition of DNA-PKcs the MMEJ has immediate access and an increased rate of activity. These results are consistent with a previously proposed model in which MMEJ acts as a backup system for C-NHEJ (13, 21, 35).

Modeling of DSB repair kinetics and fidelity: outlook

Our methodology to determine rate constants may be improved further. We found that activation of Cas9 was relatively slow. We accommodated this by taking the gradual accumulation of Cas9 activity into account, but instant activation of Cas9 would simplify and improve the modelling. A recently reported chemically inducible variant of Cas9 (39) may serve this purpose. Stable integration of a sgRNA-expressing vector, rather than transient transfection, would ensure that 100% of the cells express the sgRNA, thus removing the need to include transfection efficiency as a fitting parameter. This would also circumvent any dilution of intranuclear plasmids due to cell division, ensuring that the expression level of the sgRNA remains stable over prolonged time course experiments. However, stable transfection will only be feasible with an inducible sgRNA expression system with very low background activity, otherwise indels may accumulate before the start of the time course measurements.

Our measurements and modelling did not include possible differences in pathway activity over the cell cycle. The modelling also did not take into account that cells with a DSB undergo a transient cell cycle arrest, although it seems unlikely that this has a major impact on the estimates of the kinetic parameters (see Methods). Certainly, we did not sample the full diversity of sequence contexts and chromatin environments, and it will be interesting to investigate the repair kinetics and fidelity of many more loci in the genome.

MATERIAL & METHODS

Cell culture and transfection

We established K562#17, which is a clonal cell line of the female K562 cells (American Type Culture Collection) stably expressing DD-Cas9. K562#17 cells were cultured in RPMI 1640 (Life Technologies) supplemented with 10% fetal bovine serum (FBS, HyClone®), 1% penicillin/streptomycin. Mycoplasma tests were negative. For transient transfection, 6×10^6 K562 cells were resuspended in self-made transfection buffer (100 mM KH_2PO_4 , 15 mM NaHCO_3 , 12 mM MgCl_2 , 8 mM ATP, 2 mM glucose (pH 7.4)) (56). After addition of 3.0 μg plasmid DNA, the cells were electroporated in an Amaxa 2D Nucleofector using program T-016. DD-Cas9 was induced with a final concentration of 500 nM Shield-1 (Aobious). For kinetics experiments, 18×10^6 cells were transfected and divided over 12-well plates, one well for each time point and each well carrying 1×10^6 cells. Cas9 was activated 24 hours after nucleofection and cells were collected at the indicated time points after Cas9 induction. As controls, cells without Shield-1 were collected at various time points. DNA-PKcs inhibitor NU7441 (Cayman) (final concentration 1 μM) or DMSO (control) was added to K562#17 at the same time when the cells were supplemented with Shield-1 to induce DD-Cas9. 10 Gy of IR was administered by Cs source Gammacell®40 Exactor (Best Theratronics).

Constructs

The sgRNA oligos (SUPPLEMENTARY TABLE 1) were cloned into expression vector pBluescript with the sgRNA cassette of PX330 (Addgene plasmid 42230) and transfected into K562#17. The sgRNAs were designed using CHOPCHOP (57). The pLenti-Cas9-T2A-Neo expression vector (58) was a kind gift of Dr. Bastiaan Evers, NKI. In the expression vector, the ubiquitin promoter was exchanged for the hPGK promoter and a destabilization domain (DD) (28) was added at the N-terminus of the Cas9 gene, to generate DD-Cas9.

Cell viability

Cell viability was measured using a CellTiter-Blue Cell Viability Assay (Promega). CellTiter-Blue Reagent was 1:5 diluted in RPMI 1640 (Life Technologies) supplemented with 10% fetal bovine serum, 20 μL of this diluted CellTiter-Blue Reagent was added to 100 μL cell suspension. After a 3 hours incubation at 37°C in a 96 well tissue culture plate, fluorescence ($560_{\text{Ex}}/590_{\text{Em}}$) was measured on an EnVision Multilabel Plate Reader (Perkin Elmer). Results of one CellTiter-Blue Cell Viability Assay per experiment are shown. Cell viability was measured 45 hours after Shield-1.

Next Generation Sequencing

Cells were collected by centrifugation (300xg, 5 min) and the genomic DNA was isolated using the ISOLATE II Genomic DNA Kit (Bioline). PCR was performed in two steps; PCR1

with ~100 ng genomic DNA and site specific barcoded primers (SEE SUPPLEMENTARY TABLE 2). PCR2 used 2 μ L of each PCR1 product with Illumina PCR Index Primers Sequences 1-12. Each sample was generated with a unique combination of a barcode and index. Both PCR reactions were carried out with 25 μ L MyTaq™ Red mix (Bioline), 4 μ M of each primer and 50 μ L final volume in a 96 well plate. PCR conditions were 1 min at 95°C, followed by 15 sec at 95°C, 15 sec at 58°C and 1 min at 72°C (15x). 20 μ L of 8 samples were pooled and 100 μ L was loaded onto a 1% agarose gel. PCR product was cut from gel to remove the primer dimers and cleaned with ISOLATE II PCR and Gel Kit (Bioline). The isolated samples were sequenced by Illumina MiSeq. In this study, we only amplified the sgRNA on-target sequences. The effect of possible off-target activity of Cas9/sgRNA was ignored and is considered to be equal between the different experiments.

LM-PCR

Genomic DNA (350 ng, determined by Qubit assay, Thermo Fisher Scientific) was incubated with 0.1 μ M extension primer (EB479 or EB551) (SEE SUPPLEMENTARY TABLE 2) and Phusion High Fidelity DNA polymerase (Thermo Fisher Scientific) for 5 min at 95°C, 30 sec at 55°C and 30 sec at 72°C, to extend the non-blunt DNA ends near the break site. Subsequently, 0.16 μ M dsDamID adaptor (59) was ligated to the blunted broken ends at 16°C overnight using T4 ligase (5 U/ μ L, Roche). Ligase was heat inactivated for 20 min at 65°C. To detect broken DNA a PCR was performed on the adaptor-broken DNA ligation product with 3 μ M adaptor primer, EB486 and broken primer (EB487 or EB553). Note that this assay may not detect a fraction of the broken ends that have undergone large resection near the break site to prepare for HR. In parallel an internal standard PCR was performed with the same samples using 3 μ M primers EB488 and EB487 or EB555 and EB553 that are both located downstream of the sgRNA break site. PCR conditions were 4 min at 95°C, followed by 33 cycles of 10 sec at 95°C, 10 sec at 58°C, and 10 sec at 72°C. During preparation the samples were always kept at 4°C. The PCR products were analysed on a 1.5% agarose gel and imaged with ChemiDoc™ Imaging Systems (BioRad). A box with a fixed volume was set over the appropriate LM-PCR band for each sample and the intensities within each box were determined in arbitrary units. The signal without sgRNA was considered background and was subtracted from all other time point samples. To correct for the arbitrary intensity values between different time series, each sample was divided by the sum of all values in one experiment. Subsequently the data was scaled to the maximum value within the time series. For the input PCR two boxes were set with similar volume at and under the PCR band for each sample. These values were subtracted to correct for background signal, divided by the sum of all values in one experiment and scaled to the maximum value within the time series. The mean of biological replicates is shown \pm standard deviation (SD). The peak time (t), when the broken fraction reaches a maximum, is determined by fitting the band intensity values of the time points with the expected curve shape from the three-state Ordinary Differential Equations (ODE) model.

TIDE method

The TIDE method was performed as described in (30). Briefly, PCR reactions were carried out with ~100 ng genomic DNA in MyTaq™ Red mix (Bioline) and purified using the ISOLATE II PCR and Gel Kit (Bioline). About 100-200 ng DNA from purified PCR samples was prepared for sequencing using BigDye terminator v3.1. Samples were analysed by an Applied Biosystems 3730x1 DNA Analyzer. The data obtained was analysed using the TIDE software (<http://tide.nki.nl>). The decomposition window used for TIDE was set to indels of size 0-10 bp.

In vitro digestion with Cas9

PCR fragments of the target regions were amplified with MyTaq™ (Bioline) according to manufacturer's instructions. SEE SUPPLEMENTARY TABLE 2 for the used primers. *In vitro* transcribed sgRNA was generated by T7 promoter driven transcription using the Ribomax kit (Promega) and the RNA was purified with the MEGAclear kit (Life Technologies). 1.5 µL Cas9 (NEB), 50 µg sgRNA and 200 ng PCR fragment were incubated for 2 hours or overnight at 37°C. For denaturation of the Cas9, samples were either incubated for 20 min at 80°C or for 3 min at 96°C and slowly cooled down to 20°C (1°C/min).

Flow cytometry

K562#17 cells were collected 1 day after nucleofection and directly analysed for fluorescence using a BD FACSCalibur. Viable cells were gated on size and shape using forward and side scatter.

For cell cycle profiles, cells were fixed with 5 mL of 70% ethanol at 4°C overnight. After fixation, cells were washed with PBS and then incubated in PBS containing propidium iodide (PI) and RNase for 30 min at 37°C in the dark. GFP & PI expression were measured using a 488 nm laser for excitation.

Western blotting

Whole-cell extracts of ~0.5 x 10⁶ cells were prepared by washing cultures in PBS and lysing with 50 µL lysis buffer (Tris pH 7.6, 10% SDS, Roche proteinase inhibitor). Samples were pulse sonicated for 2 min and protein concentrations were determined using the Pierce BCA protein assay (Thermo Fisher Scientific). Samples containing 40 µg of total protein were separated by SDS-PAGE on 8% acrylamide gels and transferred to a nitrocellulose membrane through electroblotting. The membranes were blocked for 1 hour in PBS/Tween-20 0.1% containing 5% low fat milk. After washing twice with PBS/Tween20 0.1% the membranes were incubated with 1:2,000 α-Cas9 7A9 (Thermo Fisher Scientific) or γ-tubulin (T6557, Sigma) for 2 hours at room temperature with mild shaking. Subsequently, the membranes were washed again and incubated with a secondary antibody, 1:10,000 α-mouse IR800 (Li-Cor) at 37°C for 1 hour with shaking. The antibody was

detected by Odyssey scanner. The gel was analysed by setting boxes with a fixed volume over the signals in each sample. The intensity within a set box was determined by Image Studio Version 2.0. Sample $t = 60$ h was loaded on every gel and used to normalize for the intensity of the signal in the different gels. The mean of three biological replicates is shown \pm standard deviation (SD).

The relative activity of CRISPR/Cas9 is calculated by fitting the DD-Cas9 protein abundance from the quantified signals. The model employed for fitting describes stabilization of

$$R(t) = 1 - 2^{-\frac{t}{hl}} \quad (1)$$

DD-Cas9 upon introduction of Shield-1, with the unit-less relative activity:

where hl is the protein half-life of DD-Cas9 with Shield-1 determined as 6.4 h, and t is time after introduction of Shield-1. The model assumes that DD-Cas9 is very unstable without Shield-1, which is confirmed by the virtual absence of indel accumulation in the absence of stabilization (FIGURE 1E).

NGS data analysis

In each sequence read, the distance between a fixed sequence at the start and at the end are determined and used to calculate a score, defined as the difference between the measured distance in the read and the expected distance in a wild-type sequence. *Insertions* and *deletions* have score >0 and <0 , respectively. A *point mutation* has score=0, but some bases in the sgRNA target site differ from the wild-type sequence. The *intact* type specifies reads identical to wild-type sequences.

Per time point, the ratio of each type over the total of reads is calculated. We observed only 2-6.5% sequence reads in which we could not find a match with the constant parts and we discarded these reads in subsequent analyses. The called *point mutations* (score = 0) showed a very similar kinetic profile as the *intact* sequence (wild-type), indicating that they are mostly sequencing errors. We therefore assigned them as *intact* sequence in the analysis. *Insertion* and *deletion* levels at time point $t = 0$ h were considered as background and subtracted from all time points.

In the double-cut assays, paired end sequencing was performed. The forward and reverse read were matched by the unique sequence ID of a pair of reads. The deletion events were divided into two types: (i) perfectly excised DNA, and (ii) excised DNA with an indel when the deletion was larger than the expected excised product or up to 5 nucleotides smaller.

To determine the untransfected fraction, a standard sigmoid fit was applied to the

$$y(t) = 1 - \frac{U}{1 + e^{-a(t-b)}} \quad (2)$$

time lapse curve of *intact* sequences according to the following equation. Where a , b and U are parameters that determine the shape of the curve. The fitting is done by the *nls* package of R by minimization of the Gaussian dispersion. U describes the asymptotic plateau, that is, the untransfected fraction

Mathematical modeling

I. Modeling the kinetics of total indels

The three fractions of *intact* (P), *broken* (B) and *total indels* (M) of a locus at any given time (t) after cutting by CRISPR/Cas9 must adhere to the principle of conservation

$$P(t) + B(t) + M(t) = 1 \quad (3)$$

We assumed the cutting and repair kinetics is of first-order and the overall activity of CRISPR/Cas9 is proportional to the abundance of Cas9 that is modelled in equation (1). As shown in the diagram (FIGURE 1A) and taking equation (3) into account, the kinetics of the *intact* fraction (P) and the total *indels* (M) is determined by a nonlinear ODE

$$\begin{cases} \frac{dP(t)}{dt} = -k_c R(t)P(t) + k_p B(t) = -k_c R(t)P(t) + k_p(1 - P(t) - M(t)) \\ \frac{dM(t)}{dt} = k_m B(t) = k_m(1 - P(t) - M(t)) \end{cases} \quad (4)$$

whereby k_c stands for the maximal Cas9 cutting rate, k_p for perfect repair rate and k_m for mutagenic repair rate that gives rise to indels, all of which are in h^{-1} .

Because broken DNA is not detected in amplicons across the break site, the measured intact fraction (P_r) is the ratio between the abundance of intact sequences and the sum of the abundance of intact and indel sequences. Including the untransfected fraction (U) from equation (2) as a part in the intact fraction, the measured intact fraction is

$$P_r(t) = \frac{U + P(t)}{U + P(t) + M(t)} \quad (5)$$

Taking equations (1), (2) and (4) together and a set of randomly chosen initial values for k_c, k_p, k_m , the $P(t)$, $B(t)$ and $M(t)$ were modelled. $P_r(t)$ was calculated with equation (5), to demonstrate simulation outcomes as shown in SUPPLEMENTARY FIGURE 1. Then, we estimated the optimal values of kinetic parameters by minimizing the difference between the modelled $P_r(t)$ and experimentally measured intact fraction using the Levenberg-Marquardt algorithm (LMA) using Package *FME* in R. The estimated rates were used to deduce the $P(t)$, $B(t)$ and $M(t)$ fractions according to equation (3). The time point of the highest value of $B(t)$ is τ .

The time course spans multiple cell divisions. Because cells with broken DNA may enter a transient cell cycle arrest, this could lead to under-representation of the *broken* fraction,

as it provides cells without broken DNA with a relative growth advantage in the cell pool. However, because of three reasons combined, we believe that the impact of this bias on our parameter estimates is minor. First, we found that the percentage of arrested cells is no more than ~10% at 16 hours after Cas9 activation, and even less at earlier and later time points (SUPPLEMENTARY FIGURE S5). Second, because K562 cells are approximately tetraploid (SUPPLEMENTARY FIGURE S4), arrested cells are likely to also harbour intact or indel DNA; hence a cell cycle arrest will not only reduce the *broken* fraction, but also the *intact* and *indel* fractions. Third, our kinetic modelling is based on the measured $indel/(indel+intact)$ ratios, not on measurements of broken DNA. A modest underrepresentation of broken DNA in the cell pool does not significantly affect this ratio. For FIGURE 6, we performed the fitting for each sgRNA, using the measurements from all time series combined. We first inferred the transfection efficiency from these data by conducting a series of model fits while varying the transfection efficiency value from 60 to 90% in steps of 1%. The transfection efficiency value that yielded the lowest sum of residuals squared was chosen. For LBR2, LBR8 and chr11 this resulted in estimated transfection efficiencies of 72%, 69% and 75%, respectively, which is in close agreement with efficiencies observed by flow cytometry analysis (see Results). For AAVS1 this value was ~85% but the residual errors did not converge to a clear minimum; we therefore decided to use the average estimate of the other three loci, i.e. 72%. Next, the model fitting was carried out on the data points of all replicates combined, using a standard 1,000-fold bootstrapping by random sampling of the data with replacement.

II. Testing the robustness of the modelled perfect repair rate at LBR2

For the parameter sweep analysis testing the robustness of the rate of perfect repair, k_p (FIGURE 2E, corresponding text), we restricted the ratio of k_p to the total repair rate (k_p+k_m) as a sweeping factor

$$V(i) = \frac{k_p}{k_p + k_m} = \frac{iW_0}{iW_0 + (1 - W_0)2^{i-1}} \quad (6)$$

varying from $W_0=0.75$ downwards by i times from 1 to 32, spanning 10 orders of magnitude. The incentive of introducing i as a term of multiplication is to test more carefully at high k_p and a broad dynamic range with virtually equal step size at the low end. Therefore, we can represent k_p as

$$k_p = \frac{k_m V(i)}{1 - V(i)} = \frac{i2^{1-i}W_0k_m}{(1 - W_0)} \quad (7)$$

Taking equations (1), (2), (4), (5) and (7) together, we performed the model fitting with LMA. Given a fixed ratio $V(i)$ of perfect repair, the F-test was applied to examine the statistical significance for the difference in performance between restricted and optimal fittings, by

the standard approach that takes the degree of freedom, which is the number of time points of the experiment minus 2 (the number of parameters in the model) and the fold difference in the deviance of fittings. A cut-off at $p = 0.1$ is applied to determine the upper bound of deviance and thereafter the upper bound of the ratio of perfect repair (V) is calculated accordingly (FIGURE 2E).

Taking a similar approach of parameter sweep, we tested the robustness of k_c , and calculated the corresponding deviation of model fit (SUPPLEMENTARY FIGURE S2F).

III. Modeling the kinetics of individual indels

To model +1 (M_{+1}) and -7 (M_{-7}) indels individually, we introduced the kinetic terms k_{+1} and k_{-7} correspondingly for each mutant. As shown in the diagram (FIGURE 5A), the kinetics of individual indels can be written as

$$\begin{cases} \frac{dM_{+1}(t)}{dt} = k_{+1}B(t) = k_{+1}(1 - P(t) - M(t)) \\ \frac{dM_{-7}(t)}{dt} = k_{-7}B(t) = k_{-7}(1 - P(t) - M(t)) \end{cases} \quad (8)$$

Similar to the measured intact fraction (P_r), the measured fraction of individual indels ($M_{r,+1}, M_{r,-7}$) can be represented as

$$\begin{cases} M_{r,+1}(t) = \frac{M_{+1}(t)}{U + P(t) + M(t)} \\ M_{r,-7}(t) = \frac{M_{-7}(t)}{U + P(t) + M(t)} \end{cases} \quad (9)$$

Taking equations (1), (2), (4), (5) and (8) together, we estimated the kinetic parameters by minimizing a gradient of differences between the modelled $P_r(t)$, $M_{r,+1}(t)$ and $M_{r,-7}(t)$ and experimentally measured sequence data of the intact, +1 and -7 fraction by LMA.

IV. Adjustment of the modeling of the -7 indel kinetics

Following the modelling of individual indels, we discovered inconsistency between the fitted curve of the -7 indel and the experimental data (FIGURE 5C), suggesting the repair rate for the -7 indel is not a constant. Assuming a linear change of -7 repair rate over time, we adjusted the model by introducing a starting rate ($k_{-7,t=0h}$) and an end rate ($k_{-7,t=60h}$) (FIGURE 5D), and the non-constant -7 repair rate can be described as

$$k_{-7}(t) = k_{-7,t=0h} + \frac{k_{-7,t=60h} - k_{-7,t=0h}}{T}t \quad (10)$$

whereby T is 60 h, the duration of experiment.

Adjusting equation (8) by equation (10), we have

$$\begin{cases} \frac{dM_{+1}(t)}{dt} = k_{+1}(1 - P(t) - M(t)) \\ \frac{dM_{-7}(t)}{dt} = (k_{-7,t=0h} + \frac{k_{-7,t=60h} - k_{-7,t=0h}}{T}t)(1 - P(t) - M(t)) \end{cases} \quad (11)$$

Taking equations (1), (2), (4), (5), (9) and (11) together, we performed the model fitting with LMA.

Data and Software Availability

Raw image files are deposited on Mendeley Data (<https://data.mendeley.com/datasets/wg4ssg7pfw/draft?a=536407f5-2248-4cef-a949-bf7e709ebf17>).

Sequence files are deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE113129 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE113129>).

Code is provided at https://github.com/vansteensellab/DSB_Repair_Kinetics.

ACKNOWLEDGMENTS

We thank Bastiaan Evers for sharing plasmids prior to publication; Jeroen van den Berg for help with experiments; Hein te Riele and members of our laboratory for critical reading of the manuscript; the NKI Genomics and flow cytometry core facilities for technical assistance.

AUTHOR CONTRIBUTIONS

EKB designed the study, performed experiments, wrote code, analyzed data, wrote the manuscript. TC designed and conducted mathematical modeling, wrote code, analyzed data. MdH optimized and performed experiments to measure DSBs, Cas9 western blots, *in vitro* Cas9 experiments. HAH contributed to studying the effects of IR and NU7441. WA wrote code. BvS designed and supervised the study, analyzed data, wrote the manuscript.

FUNDING

This work was supported by a ZonMW-TOP grant and ERC Advanced Grant 293662 (to BvS). WA was supported by an NWO-ALW grant awarded to M. van Lohuizen. Oncode Institute is supported by the Dutch Cancer Society (KWF).

REFERENCES

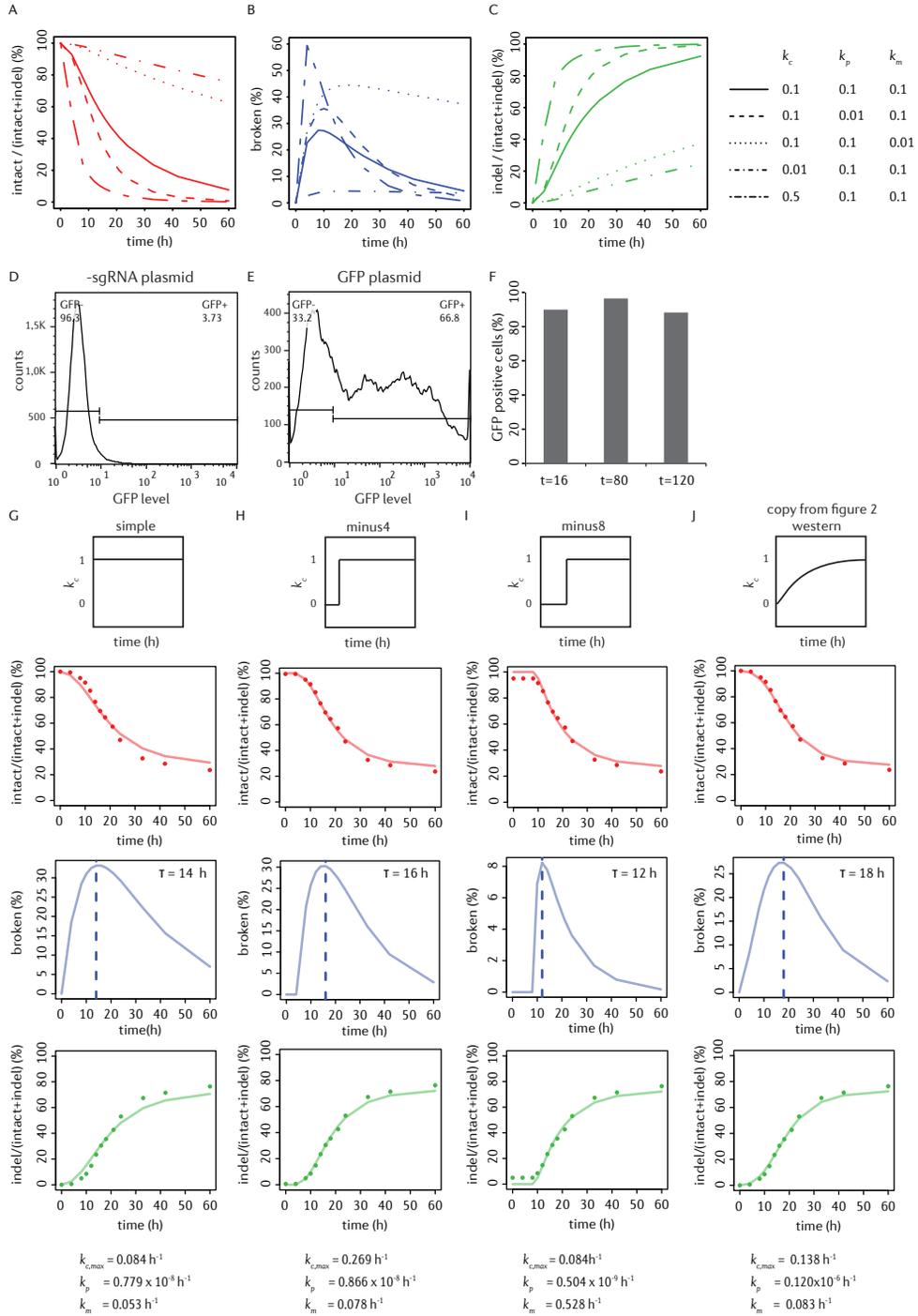
- Sander, J.D., Joung J.K., CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* 32, 347-355 (2014).
- Cong, L., Ran F.A., Cox D., Lin S., Barretto R., Habib N., Hsu P.D., Wu X., Jiang W. *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)* 339, 819-823 (2013).
- Jinek, M., East A., Cheng A., Lin S., Ma E., Doudna J., RNA-programmed genome editing in human cells. *eLife* 2, e00471 (2013).
- Mali, P., Yang L., Esvelt K.M., Aach J., Guell M., DiCarlo J.E., Norville J.E., Church G.M., RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)* 339, 823-826 (2013).
- Jasin, M., Haber J.E., The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. *DNA repair* 44, 6-16 (2016).
- Lieber, M.R., The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* 79, 181-211 (2010).
- Betermier, M., Bertrand P., Lopez B.S., Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet* 10, e1004086 (2014).
- Greene, E.C., DNA Sequence Alignment during Homologous Recombination. *The Journal of biological chemistry* 291, 11572-11580 (2016).
- McVey, M., Lee S.E., MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends in genetics : TIG* 24, 529-538 (2008).
- Bothmer, A., Phadke T., Barrera L.A., Margulies C.M., Lee C.S., Buquicchio F., Moss S., Abdulkarim H.S., Selleck W. *et al.*, Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nature communications* 8, 13905 (2017).
- van Overbeek, M., Capurso D., Carter M.M., Thompson M.S., Frias E., Russ C., Reece-Hoyes J.S., Nye C., Gradia S. *et al.*, DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Molecular cell* 63, 633-646 (2016).
- Wang, Y., Xu C., Du L.Q., Cao J., Liu J.X., Su X., Zhao H., Fan F.Y., Wang B. *et al.*, Evaluation of the comet assay for assessing the dose-response relationship of DNA damage induced by ionizing radiation. *International journal of molecular sciences* 14, 22449-22461 (2013).
- DiBiase, S.J., Zeng Z.C., Chen R., Hyslop T., Curran W.J., Jr., Iliakis G., DNA-dependent protein kinase stimulates an independently active, nonhomologous, end-joining apparatus. *Cancer research* 60, 1245-1253 (2000).
- Metzger, L., Iliakis G., Kinetics of DNA double-strand break repair throughout the cell cycle as assayed by pulsed field gel electrophoresis in CHO cells. *Int J Radiat Biol* 59, 1325-1339 (1991).
- Nunez, M.I., Villalobos M., Olea N., Valenzuela M.T., Pedraza V., McMillan T.J., Ruiz de Almodovar J.M., Radiation-induced DNA double-strand break rejoining in human tumour cells. *Br J Cancer* 71, 311-316 (1995).
- Schwartz, J.L., Rotmensch J., Giovanazzi S., Cohen M.B., Weichselbaum R.R., Faster repair of DNA double-strand breaks in radioresistant human tumor cells. *Int J Radiat Oncol Biol Phys* 15, 907-912 (1988).
- Stenerlow, B., Karlsson K.H., Cooper B., Rydberg B., Measurement of prompt DNA double-strand breaks in mammalian cells without including heat-labile sites: results for cells deficient in nonhomologous end joining. *Radiation research* 159, 502-510 (2003).
- Wang, M., Wu W., Wu W., Rosidi B., Zhang L., Wang H., Iliakis G., PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic acids research* 34, 6170-6182 (2006).
- Cucinotta, F.A., Pluth J.M., Anderson J.A., Harper J.V., O'Neill P., Biochemical kinetics model of DSB repair and induction of gamma-H2AX foci by non-homologous end joining. *Radiation research* 169, 214-222 (2008).
- Woods, M.L., Barnes C.P., Mechanistic Modelling and Bayesian Inference Elucidates the Variable Dynamics of Double-Strand Break Repair. *12*, e1005131 (2016).
- Wang, H., Perrault A.R., Takeda Y., Qin W., Wang H., Iliakis G., Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic acids research* 31, 5377-5388 (2003).
- Sharma, S., Javadekar S.M., Pandey M., Srivastava M., Kumari R., Raghavan S.C., Homology and enzymatic requirements of microhomology-dependent alternative end joining. *Cell Death Dis* 6, e1697 (2015).

23. Forment, J.V., Walker R.V., Jackson S.P., A high-throughput, flow cytometry-based method to quantify DNA-end resection in mammalian cells. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* 81, 922-928 (2012).
24. Shibata, A., Conrad S., Birraux J., Geuting V., Barton O., Ismail A., Kakarougkas A., Meek K., Taucher-Scholz G. *et al.*, Factors determining DNA double-strand break repair pathway choice in G2 phase. *EMBO J* 30, 1079-1092 (2011).
25. Leatherbarrow, E.L., Harper J.V., Cucinotta F.A., O'Neill P., Induction and quantification of gamma-H2AX foci following low and high LET-irradiation. *Int J Radiat Biol* 82, 111-118 (2006).
26. Woodbine, L., Haines J., Coster M., Barazzuol L., Ainsbury E., Sienkiewicz Z., Jeggo P., The rate of X-ray-induced DNA double-strand break repair in the embryonic mouse brain is unaffected by exposure to 50 Hz magnetic fields. *Int J Radiat Biol* 91, 495-499 (2015).
27. Kim, S., Kim D., Cho S.W., Kim J., Kim J.S., Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome research* 24, 1012-1019 (2014).
28. Banaszynski, L.A., Chen L.C., Maynard-Smith L.A., Ooi A.G., Wandless T.J., A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* 126, 995-1004 (2006).
29. Law, J.C., Ritke M.K., Yalowich J.C., Leder G.H., Ferrell R.E., Mutational inactivation of the p53 gene in the human erythroid leukemic K562 cell line. *Leuk Res* 17, 1045-1050 (1993).
30. Brinkman, E.K., Chen T., Amendola M., van Steensel B., Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic acids research* 42, e168 (2014).
31. Dai, S.M., Chen H.H., Chang C., Riggs A.D., Flanagan S.D., Ligation-mediated PCR for quantitative *in vivo* footprinting. *Nature biotechnology* 18, 1108-1111 (2000).
32. Garrity, P.A., Wold B.J., Effects of different DNA polymerases in ligation-mediated PCR: enhanced genomic sequencing and *in vivo* footprinting. *Proceedings of the National Academy of Sciences of the United States of America* 89, 1021-1025 (1992).
33. Hicks, W.M., Kim M., Haber J.E., Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science (New York, N.Y.)* 329, 82-85 (2010).
34. Perrault, R., Wang H., Wang M., Rosidi B., Iliakis G., Backup pathways of NHEJ are suppressed by DNA-PK. *J Cell Biochem* 92, 781-794 (2004).
35. Guirouilh-Barbat, J., Rass E., Plo I., Bertrand P., Lopez B.S., Defects in XRCC4 and KU80 differentially affect the joining of distal nonhomologous ends. *Proceedings of the National Academy of Sciences of the United States of America* 104, 20902-20907 (2007).
36. Sternberg, S.H., Redding S., Jinek M., Greene E.C., Doudna J.A., DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62-67 (2014).
37. Richardson, C.D., Ray G.J., DeWitt M.A., Curie G.L., Corn J.E., Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature biotechnology* 34, 339-344 (2016).
38. Mao, Z., Bozzella M., Seluanov A., Gorbunova V., Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA repair* 7, 1765-1771 (2008).
39. Rose, J.C., Stephany J.J., Valente W.J., Trevillian B.M., Dang H.V., Bielas J.H., Maly D.J., Fowler D.M., Rapidly inducible Cas9 and DSB-ddPCR to probe editing kinetics. *Nature methods* 14, 891-896 (2017).
40. Shibata, M., Nishimasu H., Kodera N., Hirano S., Ando T., Uchihashi T., Nureki O., Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. *Nature communications* 8, 1430 (2017).
41. Knight, S.C., Xie L., Deng W., Guglielmi B., Witkowsky L.B., Bosanac L., Zhang E.T., El Beheiry M., Masson J.B. *et al.*, Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science (New York, N.Y.)* 350, 823-826 (2015).
42. Ma, H., Tu L.C., Naseri A., Huisman M., Zhang S., Grunwald D., Pederson T., CRISPR-Cas9 nuclear dynamics and target recognition in living cells. *The Journal of cell biology* 214, 529-537 (2016).
43. Vilenchik, M.M., Knudson A.G., Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. *Proceedings of the National Academy of Sciences of the United States of America* 100, 12871-12876 (2003).
44. Martincorena, I., Roshan A., Gerstung M., Ellis P., Van Loo P., McLaren S., Wedge D.C., Fullam A., Alexandrov L.B. *et al.*, Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science (New York, N.Y.)* 348, 880-886 (2015).

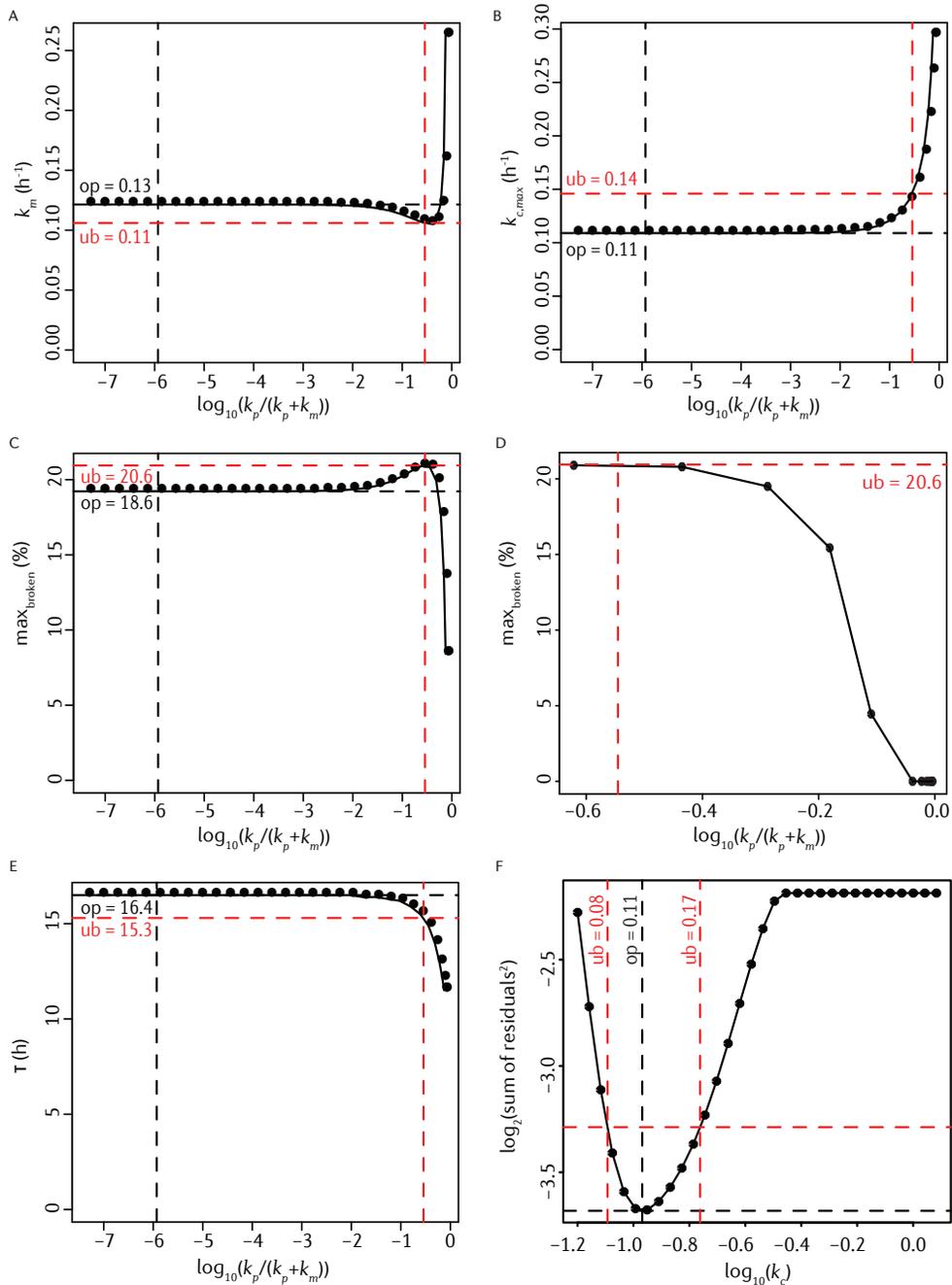
45. Keightley, P.D., Trivedi U., Thomson M., Oliver F., Kumar S., Blaxter M.L., Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome research* 19, 1195-1201 (2009).
46. Zhu, Y.O., Siegal M.L., Hall D.W., Petrov D.A., Precise estimates of mutation rate and spectrum in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 111, E2310-2318 (2014).
47. Schatz, D.G., Swanson P.C., V(D)J recombination: mechanisms of initiation. *Annu Rev Genet* 45, 167-202 (2011).
48. Boulton, S.J., Jackson S.P., *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J* 15, 5093-5103 (1996).
49. Schar, P., Herrmann G., Daly G., Lindahl T., A newly identified DNA ligase of *Saccharomyces cerevisiae* involved in RAD52-independent repair of DNA double-strand breaks. *Genes Dev* 11, 1912-1924 (1997).
50. Vriend, L.E., Prakash R., Chen C.C., Vanoli F., Cavallo F., Zhang Y., Jasin M., Krawczyk P.M., Distinct genetic control of homologous recombination repair of Cas9-induced double-strand breaks, nicks and paired nicks. *Nucleic acids research* 44, 5204-5217 (2016).
51. Kalousi, A., Soutoglou E., Nuclear compartmentalization of DNA repair. *Curr Opin Genet Dev* 37, 148-157 (2016).
52. Voit, R.A., Hendel A., Pruett-Miller S.M., Porteus M.H., Nuclease-mediated gene editing by homologous recombination of the human globin locus. *Nucleic acids research* 42, 1365-1378 (2014).
53. Canver, M.C., Bauer D.E., Dass A., Yien Y.Y., Chung J., Masuda T., Maeda T., Paw B.H., Orkin S.H., Characterization of genomic deletion efficiency mediated by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *The Journal of biological chemistry* 292, 2556 (2017).
54. Geisinger, J.M., Turan S., Hernandez S., Spector L.P., Calos M.P., *In vivo* blunt-end cloning through CRISPR/Cas9-facilitated non-homologous end-joining. *Nucleic acids research* 44, e76 (2016).
55. Chiruvella, K.K., Liang Z., Wilson T.E., Repair of double-strand breaks by end joining. *Cold Spring Harbor perspectives in biology* 5, a012757 (2013).
56. Hendel, A., Kildebeck E.J., Fine E.J., Clark J.T., Punjya N., Sebastiano V., Bao G., Porteus M.H., Quantifying genome-editing outcomes at endogenous loci with SMRT sequencing. *Cell reports* 7, 293-305 (2014).
57. Montague, T.G., Cruz J.M., Gagnon J.A., Church G.M., Valen E., CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic acids research* 42, W401-407 (2014).
58. Prahallad, A., Heynen G.J., Germano G., Willems S.M., Evers B., Vecchione L., Gambino V., Lieftink C., Beijersbergen R.L. *et al.*, PTPN11 is a Central Node in Intrinsic and Acquired Resistance to Targeted Cancer Drugs. *Cell reports* 12, 1978-1985 (2015).
59. Vogel, M.J., Peric-Hupkes D., van Steensel B., Detection of *in vivo* protein-DNA interactions using DamID in mammalian cells. *Nature protocols* 2, 1467-1478 (2007).

Supplementary Figure S1. (A-C) Simulations of the ODE model showing the relative abundance of the three states (intact, broken, indel) over time for various sets of rate constants. (D,E) Representative flow cytometry plots of a transfection of a control plasmid (D) or GFP expression plasmid (E) in same experiment as sgRNA-LBR2 transfection for a time series. Cells were collected 48 hours after transfection with indicated plasmid. Histograms of GFP fluorescence intensity is plotted. Proportions of GFP-negative and GFP-positive cells are indicated. (F) Separate flow cytometry experiment of a transfection of GFP plasmid, imaged 16, 80 or 120 hours after transfection ($n=1$). Note that the proportion of GFP-positive cells is virtually stable. (G-J) Various tested ODE models that differ in the assumed onset and accumulation curve of Cas9 activity over time: (G) instant onset at $t = 0$ h; (H) instant onset at $t = 4$ h; (I) instant onset at $t = 8$ h; (J) gradual onset quantified by Western blots. The latter is the model used in all analyses. Relative abundances of the intact and indel fractions are plotted (dots) together with the model fit (solid lines). Broken fraction is estimated from the model.

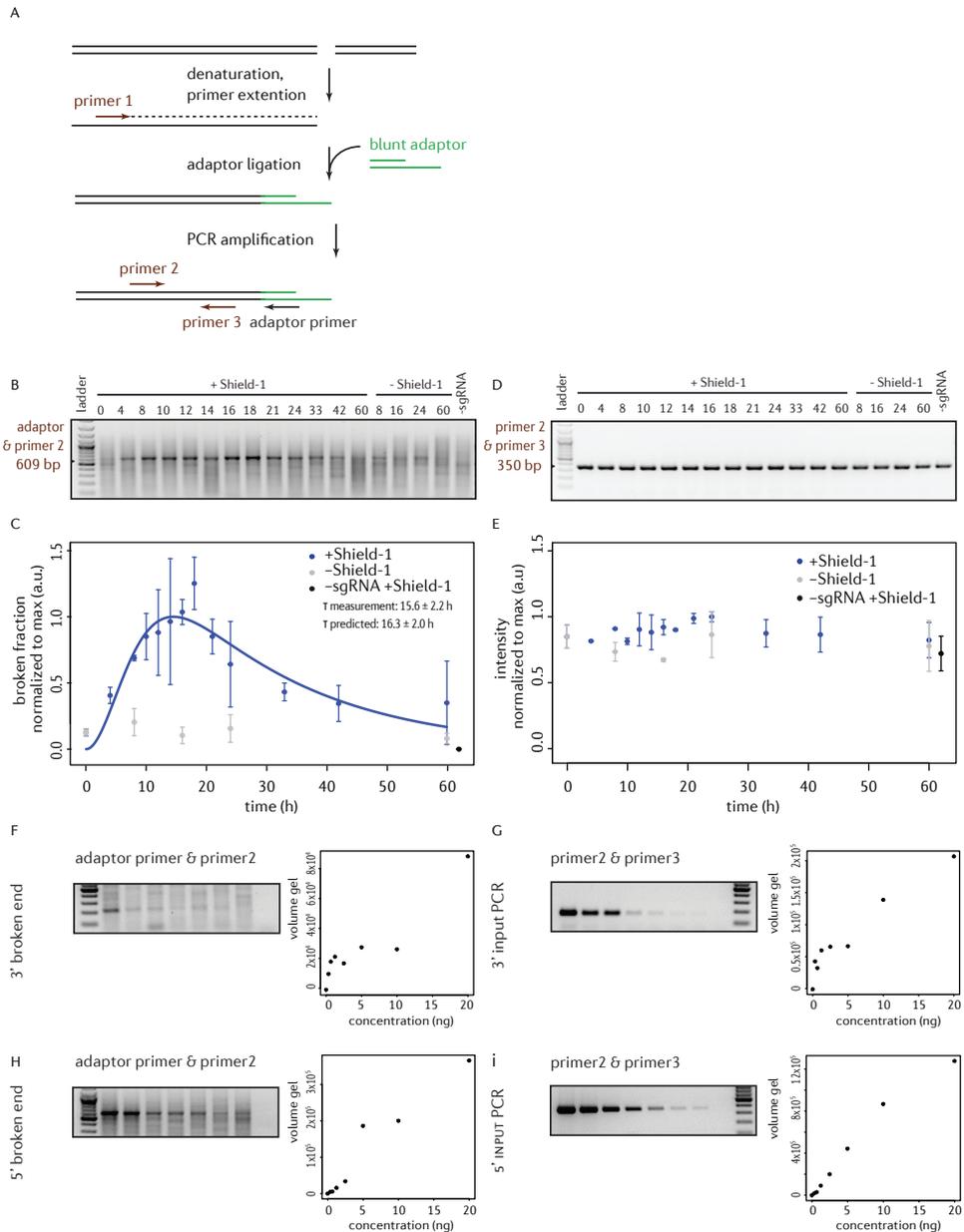
SUPPLEMENTARY DATA



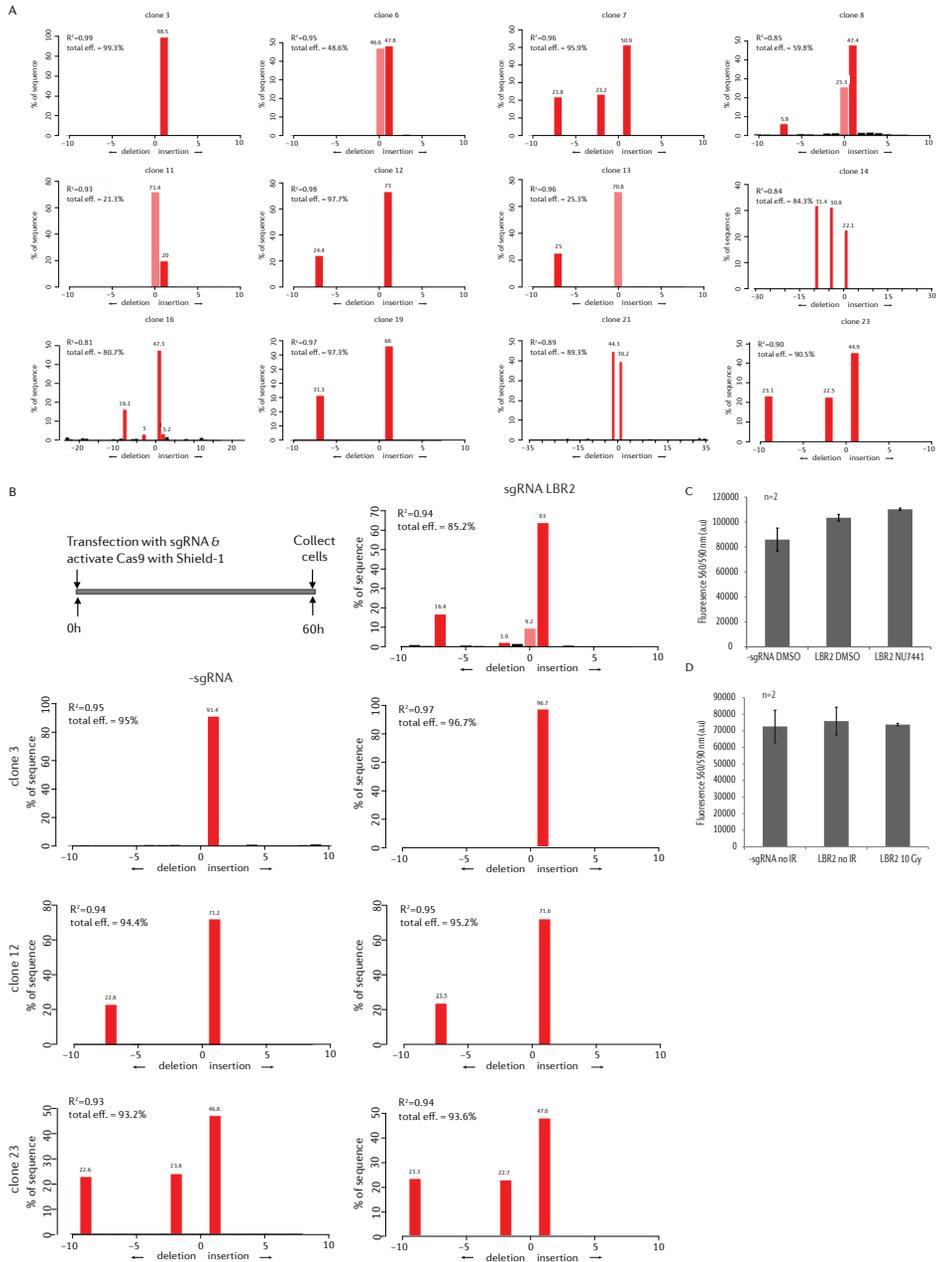
(legend on previous page)



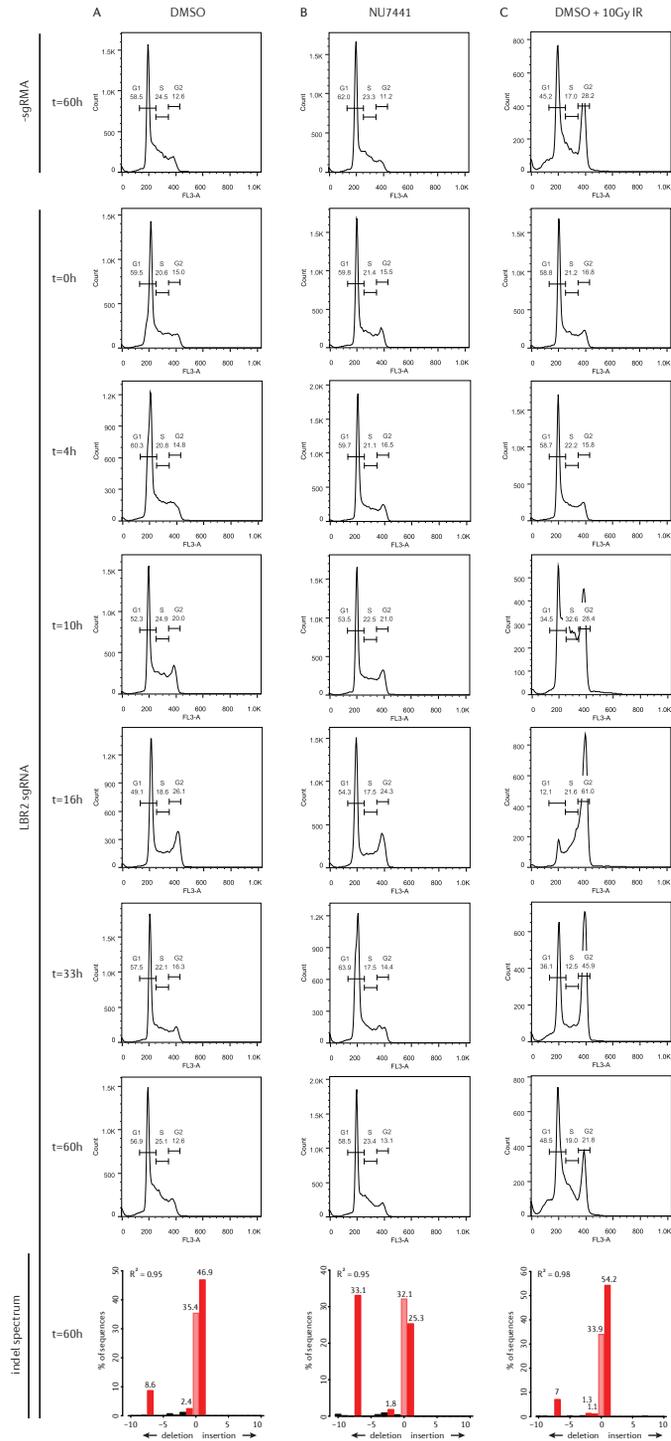
Supplementary Figure S2. Changes of the parameters $k_{c,max}$ (A), k_m (B), the amount predicted broken fraction (C,D) and τ (E) in a parameter sweep survey in which we imposed different fixed perfect/mutagenic repair ratios. Optimal fit is shown in black and significant upper bound in red ($P < 0.1$, F-test). Within this confidence interval the parameters show only modest changes. (F) Residuals of the fit to the data points are plotted for a parameter sweep survey of a fixed $k_{c,max}$ and changing k_p and k_m rates.



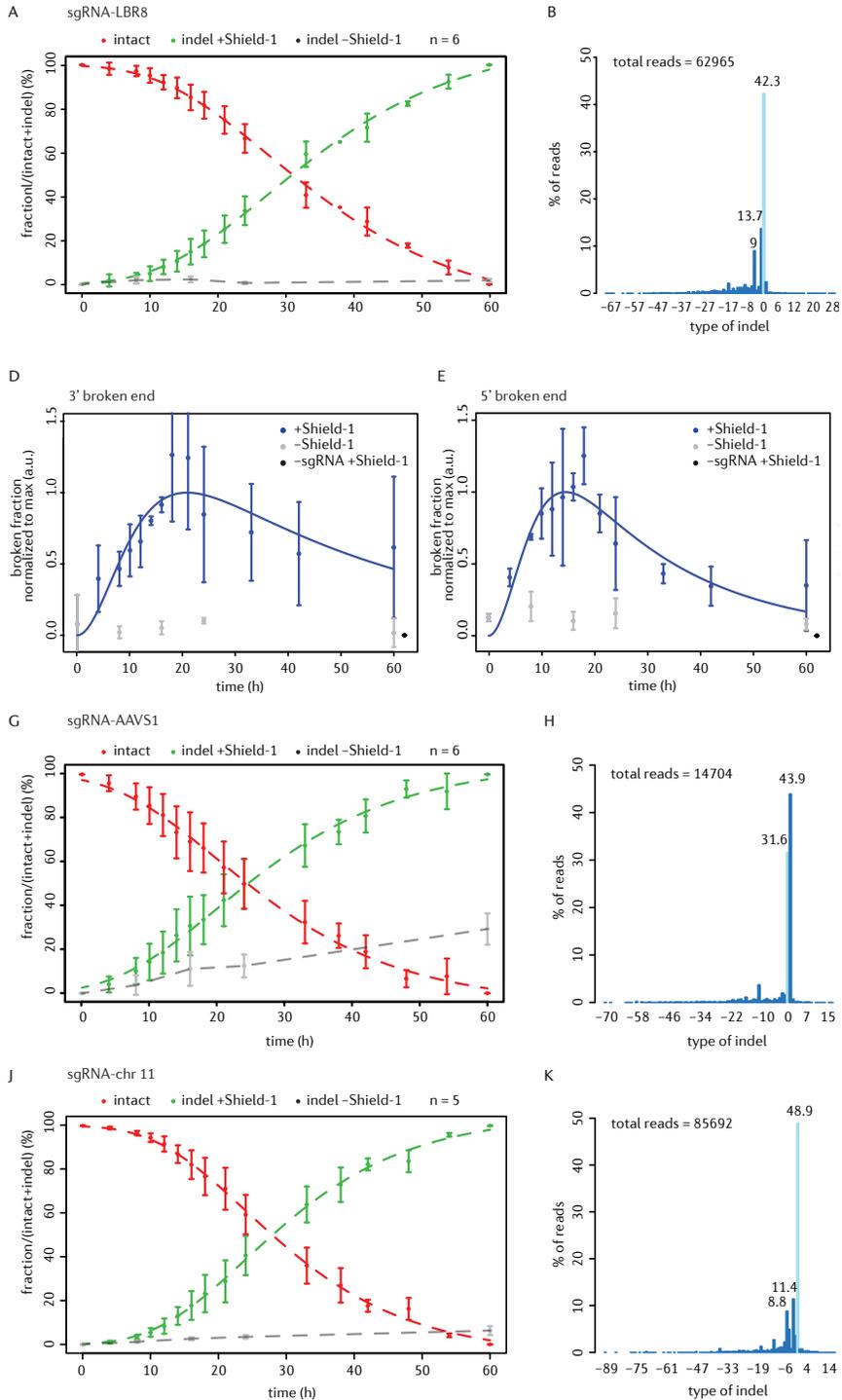
Supplementary Figure S3. (A) Schematic view of the LM-PCR assay to detect the 5 prime broken ends after DSB induction. (B) Representative agarose gel of the LM-PCR products of a time series. The expected product is 609 bp in size. (C) The broken fraction measured as band intensities (data from three LM-PCR experiments spanning three different time series; values are mean \pm SD). Solid blue line shows an ODE curve fit to the LM-PCR data to determine τ (blue shading, mean \pm SD). (D) Representative agarose gel of the input PCR products of a time series. The expected product is 350 bp in size. (E) The input PCR fraction measured as band intensities (data from 2 PCR experiments spanning 2 different time series; values are mean \pm SD). (F-I) Test of the linearity of the used primer pairs, indicated in the figure. $T=16$ h sample was diluted to range of concentrations. Agarose gel and quantification is plotted.



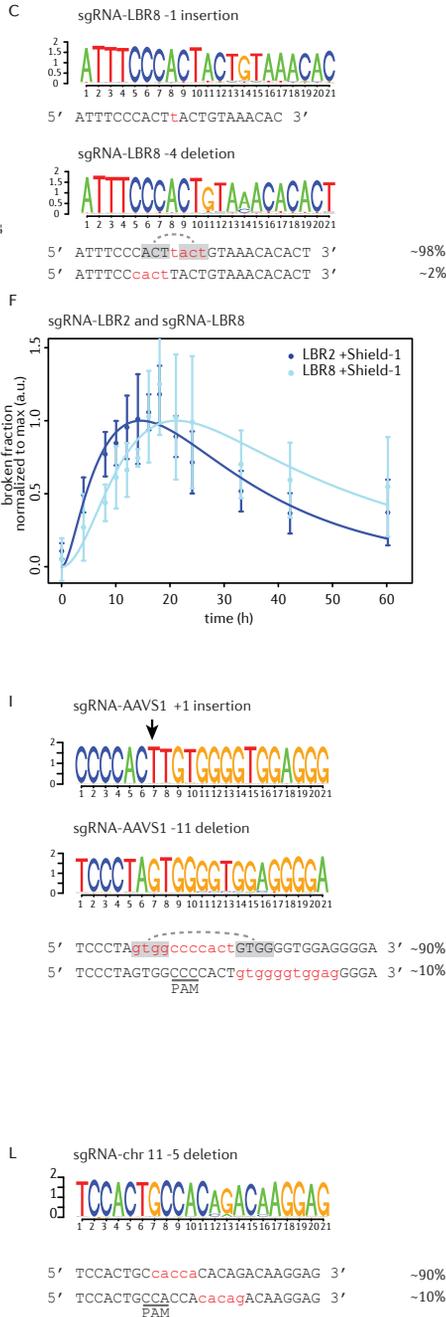
Supplementary Figure S4. (A) Indel spectra determined by TIDE in twelve out of twenty cell clones derived from sgRNA-LBR2 treated cells that gained a mutation. Note that K562 cells are tri- to tetraploid, hence the individual cell clones can have multiple peaks for allele specific mutations. (B) Three cell clones with already acquired indels by sgRNA-LBR2 were re-transfected with or without a sgRNA-LBR2 expressing plasmid. The panels show the distribution and frequencies of indels as determined by TIDE 60 hours after Cas9/sgRNA induction. (C) Cell viability assay for the cells 48 hours after addition of Shield-1 and either DMSO or NU7441. (D) Cell viability assay for the cells without a sgRNA, with sgRNA-LBR2 and with additional damage by 10 Gy. Data in (C,D) are mean \pm standard deviation from two technical replicates.



Supplementary Figure S5. Flow cytometry DNA content profiles of K562#17 transfected with sgRNA-LBR2 in the absence and presence of NU7441 to investigate whether a cell cycle arrest is triggered by Cas9-induced DSB. As a positive control, cells were treated with 10 Gy of ionizing radiation (right-hand column). Time points after Cas9 induction are indicated. Horizontal axis (FL3-A) indicates DNA content. Bottom row shows indel spectra of the same cell pools at t = 60 h, confirming that DSBs were formed and repaired in these cells.



(figure and legend continued on next page)



Supplementary Figure S6. Time series experiments of three additional loci; an additional sequence in the LBR gene using sgRNA-LBR8 (A-E), AAVS1 gene (G-I) and an intergenic region on chromosome 11 (J-L). (A,G,I) Relative fractions of intact (red) and indel (green) as a function of time. The dashed lines show sigmoid fits of the data points. Indel fraction in absence of Shield-1 is shown in grey. *n* indicates the number of time series per locus and error bars represent the SD. (B,H,K) Distribution of the type of indels as determined by next generation sequencing at *t* = 60 h. (C,I,L) Nucleotide compositions of particularly abundant indels. (D,E) Broken fraction detection in a time series experiment in presence of sgRNA-LBR8 of the 3' (D) and 5' end (E) of the break. (F) Comparison of measured broken fractions for sgRNA-LBR2 (*n*=7) and sgRNA-LBR8 (*n*=5) (average of all 3' and 5' measurements combined).

Supplementary Table 1, Related to Methods

Name	Sequence	Location (hg38)	Targeting strand
LBR guide #2	5' GCCGATGGTGAAGTGGTAAG 3'	Chr 1: 225424038- 225424057	-
LBR guide #5	5' GTATTTTAGTGATCAGCCTG 3'	Chr 1: 225424155- 225424174	-
LBR guide #6	5' AGGCTACATTCAATCTCATT 3'	Chr 1: 225424215- 225424234	+
LBR guide #7	5' GAGATTGAATGTAGCCTTTC 3'	Chr 1: 225424212- 225424231	-
LBR guide #8	5' AGAGTGTGTTTACAGTAAGT 3'	Chr 1: 225423869- 225423888	-
LBR guide #9	5' GTGTGAGCTTCTTGGGAACA 3'	Chr 1: 225423715- 225423734	+
Intergenic guide	5' TGGTCTCCTGTCTGTGTGG 3'	Chr 11: 5561098- 5561117	+
AAVS1 (4)	5' GTCCCTCCACCCACAGTG 3'	Chr 19: 55115771- 55115790	-

Supplementary Table 3, Related to FIGURE 5. Fitted parameter values for the LBR2 locus in the absence or presence of NU7441. Parameters were fitted for +1 as well as -7 indels, including different start ($t=0h$) and end ($t=60h$) rates for the -7 deletion to model the delayed accumulation. Average values \pm standard deviations are shown for n independent time courses, each fitted individually.

	LBR2 without NU7441 (n=7)	LBR2 with NU7441 (n=4)
$k_{c,max}$ (h^{-1})	0.11 ± 0.02	0.10 ± 0.004
k_p (h^{-1})	$1.3 \pm 2.3 \times 10^{-6}$	$8.6 \pm 5.5 \times 10^{-6}$
k_m (h^{-1})	0.17 ± 0.12	0.06 ± 0.01
k_{+1} (h^{-1})	0.11 ± 0.08	0.02 ± 0.006
$k_{-7,t=0h}$ (h^{-1})	$3.0 \pm 5.8 \times 10^{-3}$	0.02 ± 0.004
$k_{-7,t=60h}$ (h^{-1})	0.09 ± 0.09	0.05 ± 0.005

Supplementary Table 2, Related to Methods

Name	Number	Oligo*
Adaptor primer	EB486	5' GCGTGGTCGCGGCCGAGGA 3'
Extension primer 3'	EB479	5' TGGGTGGTTGGCAGAGTTAC 3'
Broken primer 3'	EB487	5' GAATGTAGCCTTTCTGGCCCTAAAATCCTG 3'
Standard primer 3'	EB488	5' TCCTACTTGGCATTCTTCTATAATTAACCTGA 3'
Extension primer 5'	EB551	5' CCCTTGGGCATGGAATATAA 3'
Broken primer 5'	EB553	5' CCTTCCAGCACTTGGCTGACTGTGT 3'
Standard primer 5'	EB555	5' GATTGAGCTCTTGCTTTGGGTCACATAC 3'
LBR2-fw	EB279	5' AAATGGCTGTCTTTCCCAGTAA 3'
LBR2-rv	EB361	5' ACGCAGTGGCTAAATCATCC 3'
AAVS1-fw	EB326	5' GCTTCTCCTCTTGGGAAGTGTAA 3'
AAVS1-rv	EB415	5' TTTCTGTCTGCAGCTTGTGG 3'
chr11-fw	EB333	5' AGGAAGACGATGGAGAAGACAG 3'
chr11-rv	EB416	5' CTTTCTGCCCCATGTTGATT 3'
LBR2-5-6-7-fw-bc	EB386-EB392, EB396-EB398, EB417-EB421	5' ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNTA AAGCTGGGAGGTGCTGTGC 3'
LBR8-fw-bc	EB511, EB518-EB519	5' ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNAG CTCAATCCTCTGCCTTCA 3'
Lbr9-fw-bc	EB512	5' ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTAGGTAG TAAACCCAGGGGACCAAC 3'
LBR2-5-6-7-8-9-rv1	EB393	5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACGCAGT GGCTAAATCATCC 3'
LBR2-5-6-7-8-9-rv2	EB449	5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTAGCCT TTCTGGCCCTAAAAT 3'
LBR8-rv	EB517	5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTGTG GAAAAAGACGAAT 3'
AAVS1-fw-bc	EB451-EB460	5' ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNAA GGAGGAGGCCTAAGGATG 3'
AAVS1-rv	EB450	5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTGCTT TCTTTGCCTGGAC 3'
Chr11-fw-bc	EB462-EB471	5' ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNCA GCATGGAGAGGAAAAGGT 3'
Chr11-rv	EB461	5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTAACCTG AGTTCATTGAGGGTT 3'
Illumina-fw	EB354	5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACA CGACGCTCTTCCGATCT 3'
Illumina-rv		5' CAAGCAGAAGACGGCATACGAGATNNNNNNNGTACTGGAGT TCAGACGTGTGCTCTTCCGATC 3'

*N's are the barcode or index sequences

Supplementary Table 4, Related to FIGURE 2 & FIGURE 6. Summary of estimated parameter values. (a) Average values \pm standard deviations for the LBR2 locus, based on 7 independent time courses, each fitted individually. k_p fraction show the proportion of perfect repair of the overall repair rate ($k_p/(k_p+k_m)$). (b) Values estimated by a single model fit using the all data points from n datasets (see Methods). 5-95% confidence margins are indicated in brackets below each value and were obtained by bootstrapping (1,000 cycles).

	LBR2 ^a (n=7)	LBR2 ^b (n=7)	LBR8 ^b (n=6)	AAVS1 ^b (n=6)	chr11 ^b (n=5)
$k_{c,max}$ (h ⁻¹)	0.11 \pm 0.01	0.084 [0.077 – 0.187]	0.111 [0.017 – 0.246]	0.053 [0.005 – 0.153]	0.066 [0.047 – 0.121]
k_p (h ⁻¹)	1.6 \pm 1.6 $\times 10^{-5}$	0.002 [1.1 $\times 10^{-10}$ – 0.053]	0.060 [2.0 $\times 10^{-10}$ – 0.091]	2.8 $\times 10^{-7}$ [3.3 $\times 10^{-9}$ – 1.901]	2.4 $\times 10^{-7}$ [2.9 $\times 10^{-11}$ – 0.014]
k_m (h ⁻¹)	0.15 \pm 0.11	0.174 [0.050 – 0.217]	0.019 [0.005 – 0.226]	0.443 [0.150 – 8.850]	0.051 [0.022 – 0.100]
$k_p/(k_p+k_m)$	1.1 $\times 10^{-4}$	0.011 [9.6 $\times 10^{-10}$ – 0.377]	0.758 [3.8 $\times 10^{-9}$ – 0.805]	1.1 $\times 10^{-7}$ [1.4 $\times 10^{-8}$ – 0.580]	6.2 $\times 10^{-10}$ [5.8 $\times 10^{-10}$ – 0.261]

Chapter 7

GENERAL DISCUSSION

ABSTRACT

When DNA breaks occur in the genome, DNA repair is activated. Cells employ various mechanisms to repair breaks, such as the classical non-homologous end joining pathway, the alternative end joining pathway or the homology directed repair. The choice for a particular repair pathway is regulated, however not all contributing factors are completely understood. Chromatin structure and its dynamic changes seem to play a role in the choice and kinetics of DNA repair. Currently, genome editing tools that utilize DNA repair mechanisms of the host cell to introduce mutations in the genome are widely used by researchers. In anticipation of adapting these techniques for therapeutic use, it is important to unravel the mechanism behind the choice and kinetics of repair. We review the work in this thesis and outline new research strategies.

UNDERSTANDING CRISPR/CAS9

CRISPR is a perfect example that studying a biological process can (unexpectedly) lead to a landslide of practical applications and emphasizes the importance of fundamental research. The now well-known CRISPR was described for the first time quite recently (1993) (1). Mojica *et al.* found that previously reported dissimilar repeats in genomes of prokaryotes, actually shared some repetitive motifs. These regions were named Clustered Regularly Interspaced Short Palindromic Repeats (2). Later it was reported that the spacer regions between the repeating CRISPR units corresponded to pre-existing sequences from bacteriophage genomes. The presence of external DNA elements in bacteria protected them from infection by spacer-carrying phage strains. Therefore CRISPR was suggested to function as an adaptive immune system against bacteriophages (3, 4). Several years of study led to the discovery of all essential elements of the CRISPR/Cas9 system, including; crRNAs, tracrRNA, DNA target molecule, PAM sequence and the Cas9 nuclease that actually introduces the DSB (5-9). By using the natural properties of the CRISPR system it was proven that Cas9 could be experimentally redirected to a locus of choice just by changing the crRNA spacer sequence (10, 11). The CRISPR/Cas9 system was then engineered rapidly to function in mammalian cells and it was shown that DNA at targeted locations in genome could be successfully edited (12, 13).

Applications of this technique continue to evolve at lightning speed. At present, the CRISPR/Cas9 technology has passed the National Institutes of Health's first safety review for human use and a small clinical trial for cancer treatment may get under way in the United States before the end of 2018 (14). Approval from the Food and Drug Administration is still pending. However, ethical questions about the best use of CRISPR and for what case are still unsettled. Many scientific questions remain to be solved; e.g. uncertainties concerning off-target or unintended effects during gene editing and safe delivery of the system into every cell. Solutions for these issues are essential for successful gene editing treatment.

To contribute to the applicability and understanding of the CRISPR system, we developed several methods to rapidly monitor genome editing experiments based on sequence trace decomposition (15, 16). In addition, we developed an approach to measure single-locus DSB repair kinetics after Cas9-induced breaks and to access the contribution of different repair pathways (17). We continued to measure the repair pathway preference at thousands of loci in the genome to investigate how the chromatin is influencing the repair process. For this discussion, we elaborate on how these tools can be utilized to gain more insights in the CRISPR system.

CRISPR AS A TOOL TO STUDY DNA DAMAGE RESPONSE

The development of CRISPR as a genome editing tool would have never developed so fast if we did not have an extensive knowledge of DNA repair in cells already. Since homologous sequences were known to be used to repair DSBs, cells can be provided with exogenous DNA homologous to an endonuclease target site to be used for homology directed DNA repair (HDR) (18-23). In addition, it was found that DSBs introduced by designer nucleases based on zinc finger proteins were repaired in a similar way as by locus-specific HDR (22, 23). Furthermore, it was observed that in absence of a homology template, the target site exhibited insertion or deletion mutations introduced by the non-homologous end joining (NHEJ) repair pathway (24). These studies showed that both DSB-induced repair pathways (HDR and NHEJ) are capable of editing the genome at a DNA break site (25-28).

7

No detectable rapid cycling of break and repair

Nowadays many groups have started to use CRISPR as a tool to study DNA repair at precise locations in the genome. For example by targeting specific heterochromatin repeats in the genome to study repair in these chromatin domains (29, 30). A drawback for using Cas9 to study DNA repair is that active Cas9 will keep cutting the target DNA continuously. As a consequence it cannot be excluded that in S and G2 phase of the cell cycle Cas9 may cut the target site in both sister chromatids at the same time, thereby frustrating DNA repair by HR. In addition, in case of a perfect repair event, the repaired break becomes a substrate for Cas9 again, resulting in a cycle of cutting and perfect repair events that only terminates when an imperfect repair occurs (31). Such a scenario is not representative for naturally occurring DSBs.

Nonetheless, our quantitative modelling of CRISPR repair kinetics of four independent endogenous loci revealed that CRISPR induced breaks are predominantly repaired in an error prone fashion. Our data does not support evidence for a model with rapid cycling between repetitive cutting and repair. Although, the estimated kinetics for perfect repair events in the LBR8 locus may allow several rounds of cutting and repair the contribution is low considering that the estimated half-lives for perfect repair are ~10 h and ~6 h for cutting. All in all, our model is most consistent with slow repair of Cas9 induced DSB (both for mutagenic and perfect repair).

Slow repair after Cas9-induced breaks

In our approach we consistently find repair kinetics of Cas9 induced DSBs to be considerably slower (average half-life of ~6 h) than observed in studies with bulk DSBs introduced by ionizing irradiation, where repair rarely exceeded 1 hour (32-41). The half-life times of repair foci are found to vary from 1-3 h (42-45), though these observations may overestimate the duration of actual DSB presence since foci may linger on the DNA after

the DSB is re-joined. To verify our findings that repair rates of DSBs are slow after Cas9 cutting, we carried out a double-cut experiment in which two proximal DNA sites were targeted by CRISPR in parallel. For an excision repair event (i.e. in which the intermediate part has been removed after repair) both guide target sites must be broken before repair occurs, which is more likely to happen in case where repair is slower than cutting. Indeed we observed a slow but dominant accumulation of the excision product; another indication that repair is slow. Several independent observations strengthen our conclusion that Cas9 induced DSB repair can take several hours. Sternberg *et al.* reported that *in vitro* Cas9 remains bound to the broken ends. Also in living cells catalytically inactive Cas9 has been found to have a residence time of up to 2 hours (46-48). Therefore a plausible explanation for slow repair of Cas9 induced DSBs might be that the long residence time of the Cas9 complex at the broken ends has a decisive inhibiting influence on the kinetics of the DNA repair process. Live cell imaging of repair proteins after a Cas9 induced break may validate at protein level the timing of recruitment to the break site. However the K562 cell line used for our kinetics study proved to be difficult to generate conclusive immunofluorescence images because of large numbers of background foci.

Collectively, our data give insights into the kinetics of DSB repair. CRISPR allows the study of DNA repair in a controlled manner at precisely defined locations in the genome. However, the kinetic data cannot be directly compared to the kinetics of DNA breaks introduced by other damaging agents, since the structure of the DNA ends is likely a determining factor for the Cas9 induced DSB repair process.

Broken fraction detection

Our quantitative modelling approach to determine the rate constants of breaking and repair proved to be robust and is validated by independent assays (17). Although, improvements can be made for more accurate modelling if and when the broken ends fraction could be quantified in a sensitive way. The semi quantitative LM-PCR we used here is optimized for agarose gel detection. However next-generation sequencing (NGS) of these samples suffered from high background adaptor signals, making it unsuitable for quantitation. The background may be the result of the processing of genomic DNA during isolation, where often additional technical DNA breaks are introduced. These breaks are labelled and amplified with our adaptor and primers thereby generating non-informative background signals.

In the last few years, new techniques have been developed that are based on NGS to detect both endonuclease-induced DSBs and complex genome-wide DSBs over background sequence reads by labelling the breaks *in situ*. For example the BLESS (49), End-Seq (50) and DSBcapture (51). The drawback of these techniques is that they require large amounts of input material (typically, in the order of millions of cells) and are labour-intensive. Recently, a new variant assay was developed: BLISS, 'Breaks Labelling *in situ* and Sequencing'.

This assay requires fewer cells and provides more quantitative data by uniquely labeling of all individual molecules with a UMI (unique molecule identifier) (52). Using this method for our time series samples may provide a more sensitive view of the status of broken ends in time which may lead to a more robust modelling.

PERFORMANCE OF CRISPR EDITING IN THE NUCLEUS

To control and improve the working of CRISPR editing, it is important to understand the mechanism behind the variation in cutting and repair kinetics of Cas9-induced DSBs. We have observed that the estimated ratios of perfect/imperfect repair rates vary between the four tested loci. Remarkably, even sgRNAs located only 169 bp apart (LBR2 and LBR8) exhibit marked different rates for perfect repair, suggesting that the fidelity of repair depends on the target sequence. This was strengthened by the observations that each target site is repaired in a unique but non-random manner. Even guides that are only 4 bp apart (LBR1 and LBR2) produce a completely different indel spectrum. The reason why these specific indel patterns arise is still a mystery. Multiple factors can play a role, e.g. sequence composition, presence of proteins or nucleosome positioning. It would of interest to be able to predict this phenomenon, because prior knowledge may greatly enhance effective sgRNA design. Our TIDE and TIDER webtools have become increasing popular in the last few years. Presently, the web tool hosts >1,000 analysis sessions per week, with a cumulative count of >100,000. By collecting data from these analyses, a large data set could be readily available for training algorithms to discover trends for specific sequences and indel patterns.

Various repair pathways activated by Cas9-induced DSBs

When studying the DNA repair in time by sequencing the repair outcome at nucleotide level, we found that the DSBs induced by Cas9/sgRNA-LBR2 are repaired in a way that leaves two major indels: a +1 insertion and a -7 deletion. We could link these specific indels to the activity of two individual repair pathways, respectively C-NHEJ and MMEJ. This finding gave us a great opportunity to study choice of repair pathway by reading the +1 insertion and -7 deletion under various conditions, which can be tested by a simple PCR reaction followed by sequencing.

Besides the +1 insertion and -7 deletion mutations, various different indels are introduced during LBR2 repair although at low frequency. Some of these indels may arise due to technical noise caused by sequencing errors, others that steadily increase in time probably originate from a biological process. Cluster analysis of our indel accumulation data in time may reveal whether these low abundant mutations behave similarly or differently than the

established kinetics of the C-NHEJ and MMEJ pathways. This may connect other indels to the aforementioned repair pathways or reveal new kinetic behaviours that originate from additional (unknown) repair processes.

Effect of chromatin status on Cas9-induced DSB repair

The way a DSB is repaired is regulated by multiple factors. Evidence shows that the chromatin context where breaks occur is one of these regulation factors (29, 53-55). With our newly designed pathway reporter assay explained above we aimed to investigate whether the way of repair is affected by the local chromatin status of the break site. To answer this question we integrated transposons carrying the CRISPR target site for guide LBR2 together with a unique barcode at multiple locations in the genome using the TRIP methodology (56) creating the DSB TRIP assay. The sgRNA to target LBR2 was transfected in cells carrying the complex TRIP pool to induce the breakage and repair of DNA at thousands individual locations in the genome in parallel. Since all integrated transposons carry the same target site, the position of the transposon in the genome and nucleus are the only variable. Like for the endogenous LBR2 locus the ratio of +1 and -7 indel frequencies reflects the preference for a repair pathway for many different loci spread over the genome at the same time. We observed that the variation in C-NHEJ-specific +1 and MMEJ-specific -7 indel frequencies is greater than expected on the basis of endogenous LBR2 site, suggesting a position effect for repair.

Comparing our data to chromosomal features revealed that reporters integrated in nuclear lamina associated domains (LADs) generally have a lower frequency of indels than reporters integrated in DNA positioned away from the lamina (inter-LADs, iLADs). LADs typically feature heterochromatin properties, e.g. they have low gene density, are transcriptionally silent or express at low levels and replicate late (57, 58). In previous TRIP experiments, where the GFP gene was randomly integrated in the genome, a 5- to 6-fold reduction of expression was found in LADs versus inter-LADs (56). Apparently LADs have a transcriptional repressive environment that also affect CRISPR cutting and repair rates. This corresponds to the results reported by Guell *et al.*, who finds a positive correlation between total indels and DNaseI-accessible regions (59).

Moreover, our data suggest that nuclear position also influences the preference for a repair pathway. Although the C-NHEJ pathway is genome-wide dominant, in LADs the contribution of MMEJ repair appears to be higher when compared to iLADs that have a relatively higher contribution of the C-NHEJ pathway. A less pronounced but a similar effect was observed in DNA covered with H3K27me3, which is associated with a different type of repressed chromatin. This pathway preference shift may be due to a different recruitment of repair factors in the chromatin micro-environment (29, 53-55, 60, 61). For example a DSB induced by endonuclease I-SceI in a mammalian locus near the nuclear

lamina was unable to recruit the HR associated factors BRCA1 and Rad51. Proteins of the C-NHEJ and A-EJ pathway were associated with the break (53). Our data confirms this in genome wide fashion and quantified the contribution of the C-NHEJ and MMEJ pathway in LADs and iLADs.

Upon damage, the chromatin near the break will undergo massive remodelling involving various histone modifications (62). Recently, it was shown that these modifications vary depending on which repair pathways is activated as shown by monitoring 20 chromatin features at 80 euchromatic breaks (63). To complement these finding, we showed that the pre-existing chromatin state also influences the preference for a repair pathway, whereby the classical heterochromatin versus euchromatin seem to have the biggest influence.

To verify our findings with the published data, it would be interesting to monitor the recruitment of repair pathway specific proteins by immunofluorescence in the TRIP cell pool. Furthermore, determining the rate constants of cutting and repair at all the integrated transposons would provide us insights how the chromatin state influences Cas9-induced DSB repair. The challenge with this assay is to obtain sufficient coverage for each transposon position in the cell pool at each time point. Therefore, the number of positions that can be followed in parallel is a limiting factor. As a proof of principle it would be informative to monitor the variation in repair kinetics of the C-NHEJ and MMEJ pathway specific indels for a handful of loci at the same time. The use of robotics to automatically collect time-course samples would greatly improve the collection of data of DSB repair at multiple locations in the genome in parallel and make it possible to increase the throughput. Our DSB TRIP experiment already gave us a glance on the influence of chromatin status on repair pathway preference (LAD vs iLAD). By further developing this technique, the complexity can be increased in the future. In addition, extending these studies with a broader selection of different sgRNAs targeting the integrated transposon may help to draw more general conclusions.

7

OUTLOOK TO UNRAVEL MECHANISTIC INSIGHTS IN CRISPR-INDUCED BREAKS

High and low complexity TRIP pools

The complex cell pool we have generated by TRIP covers thousands of reporters at different locations in the genome making it an ideal platform to search for correlations between repair and known chromosomal characteristics, such as chromatin states. However, some practical limitations are connected with TRIP pools. Due to divergent growth rates between cells, the cell population will drift and an unequal representation of barcodes

arises after several cell divisions. To obtain sufficient coverage of a repair distribution at all barcodes high sequencing depth is required. Hence for some experiments, low complexity or dedicated TRIP pools offer a possibility to more specific questions. Therefore we isolated unique cells from the complex TRIP pool that have been grown into a collection of cell lines with a small subset of specific transposon integrations. These cell lines with specific reporter integrations may be of interest for a screening assay of potential drugs or siRNAs targeting specific chromatin factors that change chromatin status. (FIGURE 1). For example, a dedicated pool carrying a set of ten reporters evenly distributed over LADs and iLADs is transfected with the LBR2 sgRNA to generate a DSB repair at the target site in the presence of drugs or siRNAs of interest. Changes in pathway preference ratio can be easily be detected and related to potential LAD or iLAD disrupting drugs/siRNAs and may provide mechanistic leads.

Tethering TRIP

Our data revealed that there is a clear difference for repair pathway preference in LAD and iLADs. To dissect how chromatin state influences pathway balance further experiments.

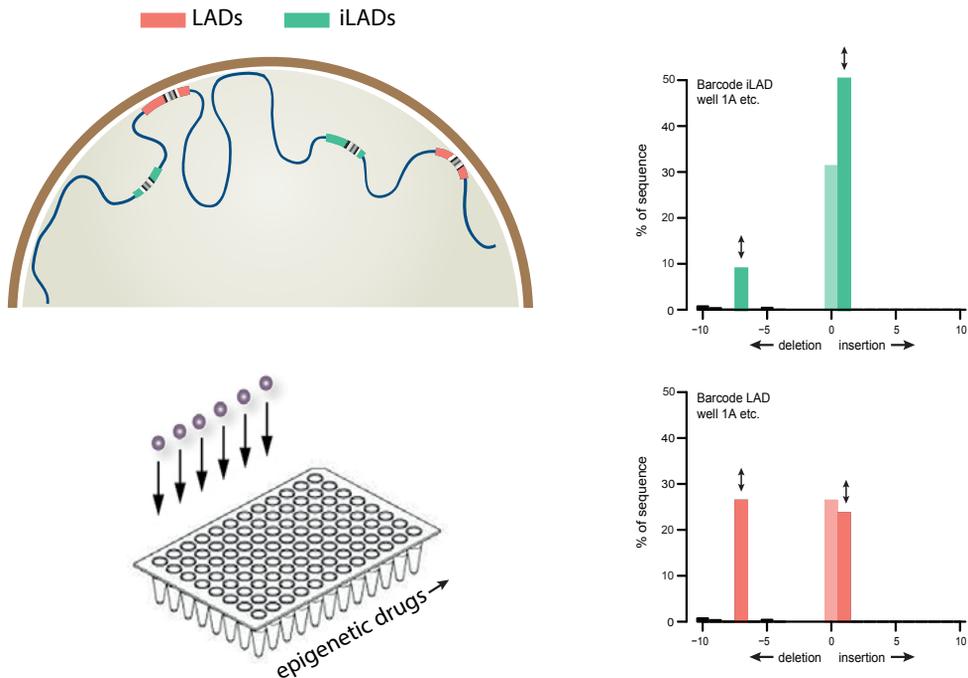


Figure 1: Repair pathway choice reporter assay. The repair products at the endogenous or integrated target site of sgRNA-LBR2 can function as a readout for DNA repair choice. The distinct +1 insertion and -7 deletions are shown to be specific for C-NHEJ and MMEJ pathways respectively. Cells can be screened for the change in +1/-7 ratio in various conditions (e.g. in presence of epigenetics drugs). With barcoded integrated sgRNA-LBR2 target sites different chromatin states can be separated (e.g. LADs vs iLADs).

are necessary. Tethering TRIP is a variant technology that has been already explored in our lab to study effects of the chromosome environment on the regulatory activity of HP-1 (64). Here, a transposon was constructed with multiple Gal4 binding sites upstream of a promoter that drives GFP expression. This configuration offers the possibility to tether Gal4-fusion proteins to the integrated transposons. Binding of the Gal4-HP-1 protein was found to repress the GFP expression in all integrated transposon reporters, proving that HP-1 is a potent and universal transcriptional repressor (64). This tethering strategy has not yet been combined with the DSB TRIP, which would give the possibility to change features in the local chromatin environment around the DSB-site. Direct effects of factors that define a certain chromatin position on DNA repair pathway choice can then be tested.

To make use of the already available TRIP cell pools, which have sgRNA-LBR2 target sites but lack Gal4 tethering sites, TALE-fusion or nuclease dead CRISPR-fusion proteins targeting the reporter transposon can be used instead (FIGURE 2). Unfortunately, for this particular DSB TRIP pool dCas9 based fusion proteins are not possible candidates as they will interfere with the Cas9 that is used to introduce the DSB. Orthologues like dCpf1 could be considered in the future (65). Short catalytically dead guide RNAs with MS2-binding loops may also be explored to tether MS2-fusions to the reporter transposons (66).

Interesting candidates to recruit to the transposon would be proteins known to be present in LADs such as the transcriptional repressor KRAB, or the methyl transferase, G9A, that writes the H3K9me2 mark (67, 68). The presence of KRAB will result in the spreading of repressive histone modification H3K9me3 (69). G9A is identified as a regulator of NL-interactions (70). Tethering these proteins to the transposons are likely to change the local chromatin environment and may enhance NL interactions of the integration in iLADs. Questions that could be answered are: does the generation of an artificial LAD changes the repair pathway choice? Is the localization of the DNA near the NL causal for pathway choice or are other regulatory elements near the integration site needed? Conversely, detachment from the NL was described for genes in LADs upon activation by the recruitment of an activator domain to the promoter (71). Similar movement was observed by the tethering of DELQPASIDP, a fortuitously discovered peptide that causes chromatin decondensation but no gene activation (71, 72). Delivering this synthetic peptide to the transposons may detach the LADs to the nuclear interior and the influence of this movement on repair preference can be studied.

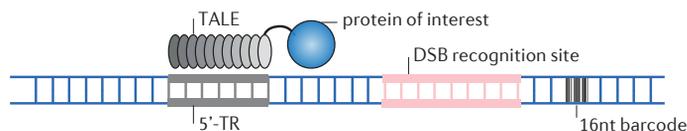


Figure 2: Thousands of reporters integrated in parallel coupled with targeted recruitment of TALE-fused proteins. A transcription activator-like effector is fused to protein of interest and designed to recognize a constant part of the transposon cassette used by the PiggyBac transposase. The protein fusion will be directed to all integrated transposons in the TRIP assay.

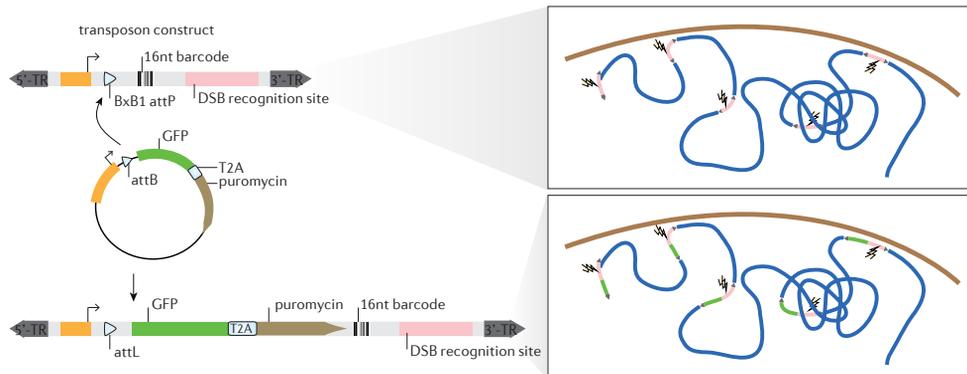


Figure 3: TRIP with landing pad for DNA integration. Schematic representation of the loading of a new feature into an integrated transposon. The TRIP cassette contains the 5' and 3' TR necessary for transposon integration, an attP attachment site (landing pad, LP) for the BxB1 recombinase, a constitutive promoter and CRISPR target site. The cargo plasmid carries GFP reporter and a puromycin selection marker that are co-expressed using a 2A translation skip peptide. BxB1-mediated recombination occurs between the LP's attP site and the cargo's attB site, yielding a similar TRIP cell pool with extra functionality (e.g. transcription unit).

Viability TRIP

Several other interesting questions can be studied using the DSB TRIP technology. Vulnerable break sites in the genome can be identified. TRIP cells with and without inflicted damage can be grown for a couple of generations. Afterwards, the barcodes of the transposon integrations can be counted. The dropout barcodes in the damaged condition may have integrations at sites that cannot be repaired or drive the cell in (permanent) cell cycle arrest or apoptosis.

Transcription - DSB TRIP

Evidence has been reported that the DSB response is differentially regulated in the proximity of a transcription unit. Transcription near a break site is silenced and the break is often associated with proteins of the HR pathway (55, 73, 74). To gain more insights on this process another TRIP variation could provide new insight on the effect of transcription on DNA repair. Two identical libraries can be generated that contain a similar CRISPR target sequence but one transposon reporter is carrying an additional GFP expression cassette. However, each TRIP cell pool has random transposon integration in the genome, while in this experiment identical locations are preferred in the two libraries. A high complexity library will deal with this to some extent, but it will be better to introduce a landing pad (LP) containing a recombination site and a selectable marker in the transposon (75) (FIGURE 3). In this way one TRIP library can be constructed without GFP cassette but with a LP. Using the recombination sites of the LP the CRISPR target sequence can be accompanied with a GFP expression cassette. Two different TRIP libraries would be created but with identical transposon integration sites. The pathway preference of a DSB near the transcription unit can be studied, but also the overall repair efficiency compared to the sites that do not have a transcription unit in close proximity.

Traffic light reporter screen

To discover new players in the various repair pathways, our sensitive LBR2-based reporter may also be used in combination with a CRISPR knock out or mutagenesis screen. For this approach a variant of the traffic light reporter will be integrated in a preferably haploid cell line like HAP-1 (76). The reporter is constructed of a promoter upstream of the LBR2 target sequence and drives the expression of a series of fluorescent proteins that are out of frame (FIGURE 4A). The first fluorescent reporter is GFP carrying a 1 nucleotide out of frame mutation, second is mCherry with an out of frame mutation of 2 nucleotides. Then the haploid cell with the integrated reporter will be exposed to either a retroviral gene-trap screen (77) or a knock out CRISPR screen (78). Hereby, each cell will be knocked out for a specific active gene. In brief, retroviral gene-trap will unbiasedly inactivate genes upon insertion in the transcription unit. A CRISPR knock out screen may introduce frame-shift mutations in the genes designed in the sgRNA library.

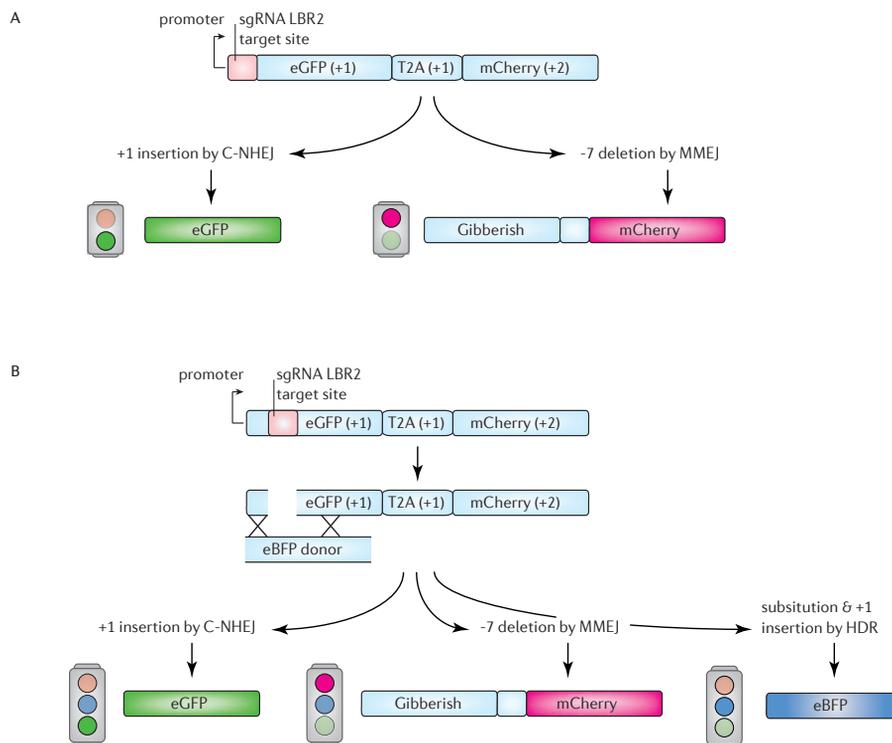


Figure 4: Screen for proteins that influence repair pathways choice. Adapted from Certo et al. (88). Schematic diagram of reporter. The arrow represents the promoter and initial reading frame. The target site of sgRNA-LBR2 is placed after the promoter and the different genome engineering outcomes following the introduction of a DSB will restore the expression of a specific fluorescent protein. (A) If the break undergoes C-NHEJ via a +1 insertion the eGFP sequence is restored and the cell will fluoresce green. If the break undergoes microhomology-mediated end joining resulting in a -7 deletion and a frameshift to the +2 reading frame, eGFP will be translated out of frame and the T2A and mCherry sequences will be translated in frame causing the cells to fluoresce red. (B) The system in (A) can be adjusted to detect three repair pathways. A donor template carrying a single Y66H amino acid substitution is added to the cells. With the incorporation of donor template via HDR, the resulting protein has a shift in the fluorescence excitation and emission spectra of the protein, converting eGFP to eBFP.

To read the effect of the inactivation of certain genes on the preference for repair by C-NHEJ or MMEJ pathway, the mutagenized cell pool would be exposed to Cas9 and the sgRNA LBR2 to induce repair by C-NHEJ and MMEJ in the traffic light reporter cassette. This will result in a mixture of GFP or mCherry positive cells. For example a hit in a gene affecting in C-NHEJ pathway will result in a decrease in the GFP while a hit in a gene of the MMEJ pathway will decrease the red signal. FACS sorting on the colours and subsequent DNA sequencing of the gene-trap or sgRNAs of the distinct cell pools will give a depletion of cells with respectively knockouts in genes involved in C-NHEJ or MMEJ repair. These dropouts may reveal potential new proteins essential in one of the two repair pathways. The traffic light reporter can be even extended with a third colour e.g. BFP surrounding the break site that is non-functional due to truncation, similar as the DR-GFP assay (79). This protein can be restored by repair with a donor template (FIGURE 4B).

CURRENT DEVELOPMENTS OF CRISPR

In this thesis, we used an inducible system that limits the activity of Cas9 to a certain period of time. However, the regulation of Cas9 activation systems is relatively slow (80, 81). More recently, inducible Cas9s have been developed using chemicals or light, which have the potential to quickly activate Cas9 (82, 83). Another interesting option is to stop or inhibit the nuclease activity of Cas9 after a period of time by the use of anti-Cas9 molecules. Anti-CRISPR proteins have been extracted from phages that have an active counter-defense system. These proteins inactivate the CRISPR-Cas-interference complex in a sequence-independent way (84). A drawback is that only anti-proteins have been reported for the NmeCas9 orthologue thus far (85). Moreover, administration of anti-proteins by means of transfection is not very suitable for a quick shutdown of Cas9.

Orthologues of CRISPR proteins are also interesting to study how they behave in time, how they are affected by the chromatin state and what repair pathway preferences they have. Various new CRISPR systems have been discovered in the last few years. Zetsche *et al.* describe the Cpf1 protein, which is part of a different CRISPR system, but also makes use of RNA-guided target cleavage (86). This variant CRISPR system has already been adjusted for gene editing in human cells. Interestingly, Cpf1 uses only a crRNA but does not need tracrRNA (86). Furthermore, sequence analysis showed that Cpf1 contains a RuvC-like cleavage domain and lacks an NHN nuclease domain and therefore generates staggered DSBs (86). In addition to Cpf1, new CRISPR effector proteins have been discovered, e.g. C2C1, C2C2, C2C5 (87). These findings expand the opportunities for genome editing using targeted endonucleases. Perhaps these CRISPR variants do not linger at the DNA ends after generation of a DSB and/or exhibit different repair kinetics than SpCas9. Since

Cpf1 generates a staggered overhang (86) HDR repair may be more efficient than for Cas9-nickases. With the systems we developed these questions can be studied.

Altogether, CRISPR is an extremely versatile technology. Application of CRISPR has the potential to engineer food, correct genetic disorders, eliminate microbes that cause disease, eradicate dangerous pests or even bring extinct animal species back to life. There is still a lot of progress to be made to achieve this and technological and ethical barriers have yet to be overcome. CRISPR has already had a tremendous impact in improving the development of model systems, both for cell lines and transgenic animals. It enables targeted regulation of gene expression and can change a single DNA base. We are only beginning to see the full potential of CRISPR and much more is to be expected. Better understanding of the cutting and repair processes involved in CRISPR-mediated genome editing will further expand the number of tools in the laboratory and bring us closer to goals of clinical use.

7

REFERENCES

1. Mojica, F.J., Juez G., Rodriguez-Valera F., Transcription at different salinities of *Haloflex mediterranei* sequences adjacent to partially modified PstI sites. *Mol Microbiol* 9, 613-621 (1993).
2. Jansen, R., Embden J.D., Gastra W., Schouls L.M., Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 43, 1565-1575 (2002).
3. Mojica, F.J., Diez-Villasenor C., Garcia-Martinez J., Soria E., Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60, 174-182 (2005).
4. Pourcel, C., Salvignol G., Vergnaud G., CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151, 653-663 (2005).
5. Bolotin, A., Quinquis B., Sorokin A., Ehrlich S.D., Clustered regularly interspaced short palindromic repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151, 2551-2561 (2005).
6. Brouns, S.J., Jore M.M., Lundgren M., Westra E.R., Slijkhuis R.J., Snijders A.P., Dickman M.J., Makarova K.S., Koonin E.V. et al., Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960-964 (2008).
7. Marraffini, L.A., Sontheimer E.J., CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843-1845 (2008).
8. Garneau, J.E., Dupuis M.E., Villion M., Romero D.A., Barrangou R., Boyaval P., Fremaux C., Horvath P., Magadan A.H. et al., The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67-71 (2010).
9. Deltcheva, E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602-607 (2011).
10. Gasiunas, G., Barrangou R., Horvath P., Siksnys V., Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 109, E2579-2586 (2012).
11. Jinek, M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821 (2012).
12. Cong, L., Ran F.A., Cox D., Lin S., Barretto R., Habib N., Hsu P.D., Wu X., Jiang W. et al., Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819-823 (2013).

13. Mali, P., Yang L., Esvelt K.M., Aach J., Guell M., DiCarlo J.E., Norville J.E., Church G.M., RNA-guided human genome engineering via Cas9. *Science* 339, 823-826 (2013).
14. Stadtmauer, E. (University of Pennsylvania 2018).
15. Brinkman, E.K., Chen T., Amendola M., van Steensel B., Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* 42, e168 (2014).
16. Brinkman, E.K., Kousholt A.N., Harmsen T., Leemans C., Chen T., Jonkers J., van Steensel B., Easy quantification of template-directed CRISPR/Cas9 editing. *Nucleic Acids Res*, (2018).
17. Brinkman, E.K., Chen T., de Haas M., Holland H.A., Akhtar W., van Steensel B., Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. *Mol Cell* 70, 801-813 e806 (2018).
18. Rudin, N., Sugarman E., Haber J.E., Genetic and physical analysis of double-strand break repair and recombination in *Saccharomyces cerevisiae*. *Genetics* 122, 519-534 (1989).
19. Plessis, A., Perrin A., Haber J.E., Dujon B., Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics* 130, 451-460 (1992).
20. Rouet, P., Smih F., Jasin M., Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol* 14, 8096-8106 (1994).
21. Choulifa, A., Perrin A., Dujon B., Nicolas J.F., Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. *Mol Cell Biol* 15, 1968-1973 (1995).
22. Bibikova, M., Carroll D., Segal D.J., Trautman J.K., Smith J., Kim Y.G., Chandrasegaran S., Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* 21, 289-297 (2001).
23. Bibikova, M., Beumer K., Trautman J.K., Carroll D., Enhancing gene targeting with designed zinc finger nucleases. *Science* 300, 764 (2003).
24. Bibikova, M., Golic M., Golic K.G., Carroll D., Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* 161, 1169-1175 (2002).
25. Canny, M.D., Moatti N., Wan L.C.K., Fradet-Turcotte A., Krasner D., Mateos-Gomez P.A., Zimmermann M., Orthwein A., Juang Y.C. *et al.*, Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. *Nat Biotechnol* 36, 95-102 (2018).
26. Maruyama, T., Dougan S.K., Truttmann M.C., Bilate A.M., Ingram J.R., Ploegh H.L., Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol* 33, 538-542 (2015).
27. Chu, V.T., Weber T., Wefers B., Wurst W., Sander S., Rajewsky K., Kuhn R., Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol* 33, 543-548 (2015).
28. Robert, F., Barbeau M., Ethier S., Dostie J., Pelletier J., Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome Med* 7, 93 (2015).
29. Tsouroula, K., Furst A., Rogier M., Heyer V., Maglott-Roth A., Ferrand A., Reina-San-Martin B., Soutoglou E., Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. *Mol Cell* 63, 293-305 (2016).
30. Warmerdam, D.O., van den Berg J., Medema R.H., Breaks in the 45S rDNA Lead to Recombination-Mediated Loss of Repeats. *Cell Rep* 14, 2519-2527 (2016).
31. Richardson, C.D., Ray G.J., Bray N.L., Corn J.E., Non-homologous DNA increases gene disruption efficiency by altering DNA repair outcomes. *Nat Commun* 7, 12463 (2016).
32. DiBiase, S.J., Zeng Z.C., Chen R., Hyslop T., Curran W.J., Jr., Iliakis G., DNA-dependent protein kinase stimulates an independently active, nonhomologous, end-joining apparatus. *Cancer Res* 60, 1245-1253 (2000).
33. Iliakis, G., Wang H., Perrault A.R., Boecker W., Rosidi B., Windhofer F., Wu W., Guan J., Terzoudi G. *et al.*, Mechanisms of DNA double strand break repair and chromosome aberration formation. *Cytogenet Genome Res* 104, 14-20 (2004).
34. Stenerlow, B., Karlsson K.H., Cooper B., Rydberg B., Measurement of prompt DNA double-strand breaks in mammalian cells without including heat-labile sites: results for cells deficient in nonhomologous end joining. *Radiat Res* 159, 502-510 (2003).
35. Kinner, A., Wu W., Staudt C., Iliakis G., Gamma-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. *Nucleic Acids Res* 36, 5678-5694 (2008).

36. Metzger, L., Iliakis G., Kinetics of DNA double-strand break repair throughout the cell cycle as assayed by pulsed field gel electrophoresis in CHO cells. *Int J Radiat Biol* 59, 1325-1339 (1991).
37. Nunez, M.I., Villalobos M., Olea N., Valenzuela M.T., Pedraza V., McMillan T.J., Ruiz de Almodovar J.M., Radiation-induced DNA double-strand break rejoining in human tumour cells. *Br J Cancer* 71, 311-316 (1995).
38. Schwartz, J.L., Mustafi R., Beckett M.A., Weichselbaum R.R., Prediction of the radiation sensitivity of human squamous cell carcinoma cells using DNA filter elution. *Radiat Res* 123, 1-6 (1990).
39. Wang, M., Wu W., Wu W., Rosidi B., Zhang L., Wang H., Iliakis G., PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Res* 34, 6170-6182 (2006).
40. Wu, W., Wang M., Mussfeldt T., Iliakis G., Enhanced use of backup pathways of NHEJ in G2 in Chinese hamster mutant cells with defects in the classical pathway of NHEJ. *Radiat Res* 170, 512-520 (2008).
41. Wu, W., Wang M., Wu W., Singh S.K., Mussfeldt T., Iliakis G., Repair of radiation induced DNA double strand breaks by backup NHEJ is enhanced in G2. *DNA Repair (Amst)* 7, 329-338 (2008).
42. Leatherbarrow, E.L., Harper J.V., Cucinotta F.A., O'Neill P., Induction and quantification of gamma-H2AX foci following low and high LET-irradiation. *Int J Radiat Biol* 82, 111-118 (2006).
43. Redon, C.E., Dickey J.S., Bonner W.M., Sedelnikova O.A., gamma-H2AX as a biomarker of DNA damage induced by ionizing radiation in human peripheral blood lymphocytes and artificial skin. *Adv Space Res* 43, 1171-1178 (2009).
44. Shibata, A., Conrad S., Birraux J., Geuting V., Barton O., Ismail A., Kakarougkas A., Meek K., Taucher-Scholz G. *et al.*, Factors determining DNA double-strand break repair pathway choice in G2 phase. *EMBO J* 30, 1079-1092 (2011).
45. Woodbine, L., Haines J., Coster M., Barazzuol L., Ainsbury E., Sienkiewicz Z., Jeggo P., The rate of X-ray-induced DNA double-strand break repair in the embryonic mouse brain is unaffected by exposure to 50 Hz magnetic fields. *Int J Radiat Biol* 91, 495-499 (2015).
46. Ma, H., Tu L.C., Naseri A., Huisman M., Zhang S., Grunwald D., Pederson T., CRISPR-Cas9 nuclear dynamics and target recognition in living cells. *J Cell Biol* 214, 529-537 (2016).
47. Richardson, C.D., Ray G.J., DeWitt M.A., Curie G.L., Corn J.E., Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* 34, 339-344 (2016).
48. Sternberg, S.H., Redding S., Jinek M., Greene E.C., Doudna J.A., DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62-67 (2014).
49. Crosetto, N., Mitra A., Silva M.J., Bienko M., Dojer N., Wang Q., Karaca E., Chiarle R., Skrzypczak M. *et al.*, Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat Methods* 10, 361-365 (2013).
50. Canela, A., Sridharan S., Sciascia N., Tubbs A., Meltzer P., Sleckman B.P., Nussenzweig A., DNA Breaks and End Resection Measured Genome-wide by End Sequencing. *Mol Cell* 63, 898-911 (2016).
51. Lensing, S.V., Marsico G., Hansel-Hertsch R., Lam E.Y., Tannahill D., Balasubramanian S., DSBCapture: in situ capture and sequencing of DNA breaks. *Nat Methods* 13, 855-857 (2016).
52. Yan, W.X., Mirzazadeh R., Garnerone S., Scott D., Schneider M.W., Kallas T., Custodio J., Wernersson E., Li Y. *et al.*, BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat Commun* 8, 15058 (2017).
53. Lemaître, C., Grabarz A., Tsouroula K., Andronov L., Furst A., Pankotai T., Heyer V., Rogier M., Attwood K.M. *et al.*, Nuclear position dictates DNA repair pathway choice. *Genes Dev* 28, 2450-2463 (2014).
54. Chiolo, I., Minoda A., Colmenares S.U., Polyzos A., Costes S.V., Karpen G.H., Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. *Cell* 144, 732-744 (2011).
55. Aymard, F., Bugler B., Schmidt C.K., Guillou E., Caron P., Briois S., Iacovoni J.S., Daburon V., Miller K.M. *et al.*, Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat Struct Mol Biol* 21, 366-374 (2014).

56. Akhtar, W., de Jong J., Pindyurin A.V., Pagie L., Meuleman W., de Ridder J., Berns A., Wessels L.F., van Lohuizen M. *et al.*, Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* 154, 914-927 (2013).
57. Guelen, L., Pagie L., Brassat E., Meuleman W., Faza M.B., Talhout W., Eussen B.H., de Klein A., Wessels L. *et al.*, Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453, 948-951 (2008).
58. Peric-Hupkes, D., Meuleman W., Pagie L., Bruggeman S.W., Solovei I., Brugman W., Graf S., Flicek P., Kerkhoven R.M. *et al.*, Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* 38, 603-613 (2010).
59. Chari, R., Mali P., Moosburner M., Church G.M., Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat Methods* 12, 823-826 (2015).
60. Janssen, A., Breuer G.A., Brinkman E.K., van der Meulen A.I., Borden S.V., van Steensel B., Bindra R.S., LaRocque J.R., Karpen G.H., A single double-strand break system reveals repair dynamics and mechanisms in heterochromatin and euchromatin. *Genes Dev* 30, 1645-1657 (2016).
61. Goodarzi, A.A., Noon A.T., Deckbar D., Ziv Y., Shiloh Y., Loblrich M., Jeggo P.A., ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. *Mol Cell* 31, 167-177 (2008).
62. Rossetto, D., Truman A.W., Kron S.J., Cote J., Epigenetic modifications in double-strand break DNA damage signaling and repair. *Clin Cancer Res* 16, 4543-4552 (2010).
63. Clouaire, T., Rocher V., Lashgari A., Arnould C., Aguirrebengoa M., Biernacka A., Skrzypczak M., Aymard F., Fongang B. *et al.*, Comprehensive Mapping of Histone Modifications at DNA Double-Strand Breaks Deciphers Repair Pathway Chromatin Signatures. *Mol Cell*, (2018).
64. Brueckner, L., van Arensbergen J., Akhtar W., Pagie L., van Steensel B., High-throughput assessment of context-dependent effects of chromatin proteins. *Epigenetics Chromatin* 9, 43 (2016).
65. Zhang, X., Wang J., Cheng Q., Zheng X., Zhao G., Wang J., Multiplex gene regulation by CRISPR-ddCpf1. *Cell Discov* 3, 17018 (2017).
66. Ye, L., Wang C., Hong L., Sun N., Chen D., Chen S., Han F., Programmable DNA repair with CRISPRa/i enhanced homology-directed repair efficiency with a single Cas9. *Cell Discov* 4, 46 (2018).
67. Margolin, J.F., Friedman J.R., Meyer W.K., Vissing H., Thiesen H.J., Rauscher F.J., 3rd, Kruppel-associated boxes are potent transcriptional repression domains. *Proc Natl Acad Sci U S A* 91, 4509-4513 (1994).
68. Tachibana, M., Sugimoto K., Nozaki M., Ueda J., Ohta T., Ohki M., Fukuda M., Takeda N., Niida H. *et al.*, G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev* 16, 1779-1791 (2002).
69. Groner, A.C., Meylan S., Ciuffi A., Zangger N., Ambrosini G., Denervaud N., Bucher P., Trono D., KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS Genet* 6, e1000869 (2010).
70. Kind, J., Pagie L., Ortobozkoyun H., Boyle S., de Vries S.S., Janssen H., Amendola M., Nolen L.D., Bickmore W.A. *et al.*, Single-cell dynamics of genome-nuclear lamina interactions. *Cell* 153, 178-192 (2013).
71. Therizols, P., Illingworth R.S., Courilleau C., Boyle S., Wood A.J., Bickmore W.A., Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* 346, 1238-1242 (2014).
72. Chuang, C.H., Carpenter A.E., Fuchsova B., Johnson T., de Lanerolle P., Belmont A.S., Long-range directional movement of an interphase chromosome site. *Curr Biol* 16, 825-831 (2006).
73. Shanbhag, N.M., Rafalska-Metcalf I.U., Balane-Bolivar C., Janicki S.M., Greenberg R.A., ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks. *Cell* 141, 970-981 (2010).
74. Kakarougkas, A., Ismail A., Chambers A.L., Riballo E., Herbert A.D., Kunzel J., Loblrich M., Jeggo P.A., Downs J.A., Requirement for PBAF in transcriptional repression and repair at DNA breaks in actively transcribed regions of chromatin. *Mol Cell* 55, 723-732 (2014).
75. Gaidukov, L., Wroblewska L., Teague B., Nelson T., Zhang X., Liu Y., Jagtap K., Mamo S., Tseng W.A. *et al.*, A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Res* 46, 4072-4086 (2018).
76. Kuhar, R., Gwiazda K.S., Humbert O., Mandt T., Pangallo J., Brault M., Khan I., Maizels N., Rawlings D.J. *et al.*, Novel fluorescent genome editing reporters for monitoring DNA repair pathway utilization at endonuclease-induced breaks. *Nucleic Acids Res* 42, e4 (2014).

77. Carette, J.E., Guimaraes C.P., Wuethrich I., Blomen V.A., Varadarajan M., Sun C., Bell G., Yuan B., Muellner M.K. *et al.*, Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. *Nat Biotechnol* 29, 542-546 (2011).
78. Malina, A., Katigbak A., Cencic R., Maiga R.I., Robert F., Miura H., Pelletier J., Adapting CRISPR/Cas9 for functional genomics screens. *Methods Enzymol* 546, 193-213 (2014).
79. Nakanishi, K., Cavallo F., Brunet E., Jasin M., Homologous recombination assay for interstrand cross-link repair. *Methods Mol Biol* 745, 283-291 (2011).
80. Zetsche, B., Volz S.E., Zhang F., A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol* 33, 139-142 (2015).
81. Wang, T., Wei J.J., Sabatini D.M., Lander E.S., Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343, 80-84 (2014).
82. Rose, J.C., Stephany J.J., Valente W.J., Trevillian B.M., Dang H.V., Bielas J.H., Maly D.J., Fowler D.M., Rapidly inducible Cas9 and DSB-ddPCR to probe editing kinetics. *Nat Methods* 14, 891-896 (2017).
83. Zhou, X.X., Zou X., Chung H.K., Gao Y., Liu Y., Qi L.S., Lin M.Z., A Single-Chain Photoswitchable CRISPR-Cas9 Architecture for Light-Inducible Gene Editing and Transcription. *ACS Chem Biol* 13, 443-448 (2018).
84. Maxwell, K.L., The Anti-CRISPR Story: A Battle for Survival. *Mol Cell* 68, 8-14 (2017).
85. Pawluk, A., Amrani N., Zhang Y., Garcia B., Hidalgo-Reyes Y., Lee J., Edraki A., Shah M., Sontheimer E.J. *et al.*, Naturally Occurring Off-Switches for CRISPR-Cas9. *Cell* 167, 1829-1838 e1829 (2016).
86. Zetsche, B., Gootenberg J.S., Abudayeh O.O., Slaymaker I.M., Makarova K.S., Essletzbichler P., Volz S.E., Joung J., van der Oost J. *et al.*, Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759-771 (2015).
87. Shmakov, S., Abudayeh O.O., Makarova K.S., Wolf Y.I., Gootenberg J.S., Semenova E., Minakhin L., Joung J., Konermann S. *et al.*, Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell* 60, 385-397 (2015).
88. Certo, M.T., Ryu B.Y., Annis J.E., Garibov M., Jarjour J., Rawlings D.J., Scharenberg A.M., Tracking genome engineering outcome at individual DNA breakpoints. *Nat Methods* 8, 671-676 (2011).

Addendum

Summary

Nederlandse Samenvatting

Abbreviations

Curriculum Vitae

List of Publications

PhD Portfolio

Acknowledgement

SUMMARY

Each individual cell contains DNA, the carrier of the hereditary information that is written in a four-letter code. The DNA code contains different elements. For example, a gene is a stretch of DNA that codes for a protein, which executes a function inside or outside the cell. When the DNA code of a gene is translated, a protein will be produced. Other parts of DNA contain different types of information, like a code for the regulation of protein expression or the structural folding of the DNA. All this information together makes the DNA two meters in size in mammalian cells. To fit it into the nucleus of a cell, it has to be tightly folded at several levels. The folding and the presence of proteins and modifications on the DNA (together called chromatin) influence the function of the DNA. For example, tightly folded DNA is less accessible to be read than more loosely folded DNA.

Changing the genetic code in cells can lead to changed functionalities. Engineered changes could for example lead to beneficial characteristics to farm crops or may disable disease causing genes in cancer cells. The development of new technology enables genome editing at selected locations in the genomes of many species. Most popular in the field of molecular biology to engineer changes are proteins such as CRISPR/Cas9, that break DNA at a programmable location. The CRISPR system consists of a guide RNA that recognizes a specific sequence in the genome and escorts the Cas9 nuclease to this site, which in turn introduces a precisely defined DNA double-strand break (DSB). In response the cell will repair this break. The main repair mechanism in mammalian cells is non-homologous end joining (NHEJ) which joins broken ends together. This repair is not always accurate whereby extra DNA letters (nucleotides) can be inserted or deleted at the site of the break, causing a mutation, which are referred to as indels. However, the efficacy and the mutation spectrum can vary substantially depending on the efficiency of introducing a DSB and on the way it is repaired. CHAPTER 1 provides an overview of CRISPR/Cas9 genome editing and the factors that affect the efficiency of this process.

In genome editing experiments with CRISPR/Cas9 it is important to know the frequency and the nature of the introduced indel mutations. Since this is difficult to predict, a simple and quick assay to accurately characterize and quantify the induced mutations is needed. In CHAPTER 2 we report the TIDE method. This is a fast and easily applicable method that accurately identifies and quantifies indels after CRISPR/Cas9 editing. TIDE requires only a pair of PCR reactions and two standard Sanger sequencing runs, which is the technique to determine the order of nucleotides of a piece of DNA. These sequence traces are then analysed by a specially developed decomposition algorithm to identify the individual indel mutations.

To generate specific mutations by CRISPR/Cas9 genome editing, the DSB can also be repaired with the help of a donor DNA. This is a designed sequence that is almost identical to the target location except from some desired mutations. In the presence of the donor DNA the designed mutations can be introduced in the genome during repair. This process is mediated by a different repair mechanism called homology directed repair (HDR) that makes use of the overlapping sequence between genomic DNA and the donor DNA for repair. This very precise editing strategy is becoming more popular, due to its property to create and study specific mutations, or to correct disease-causing nucleotide variants. In CHAPTER 3 we explain our second method, called TIDER, that can estimate the incorporation frequency of any type of template-directed mutations.

For both methods presented in CHAPTER 2 & 3 we developed interactive web tools that greatly facilitate the rapid testing and rational design of genome editing strategies. TIDE and TIDER make it easy to test the efficacy of the designed guide and to determine type and proportion of mutations that can be expected. CHAPTER 4 provides a detailed procedure and optimization steps to overcome potential problems when using these two methods.

The process of cutting of DNA by CRISPR/Cas9 and the subsequent repair of the DSB is still poorly understood. It is unknown how fast these events occur and how accurate DSB repair is. We developed a strategy to directly measure the kinetics and fidelity of DNA repair at single loci in human cells (CHAPTER 5). To do so, we constructed a cell line with a tightly controlled inducible Cas9 that made it possible to track the formation of DSBs and their repair in time for a locus of choice. Intact, broken and repaired DNA fractions are quantified by a combination of next generation sequencing and an assay to detect non-repaired broken ends. The resulting data are analysed in a mathematical model that estimates the key parameters for repair kinetics and fidelity. Our approach revealed that repair rates varied, with half-life times up to ~10 h, indicating that the repair is slow. Moreover, our data show that repair of the DSBs tends to be error-prone. Interestingly, we found that the spectrum of introduced indels is highly dependent on the sequence that is targeted. The same target sequence acquired two specific indels each representing the pathway that repairs the DSB. By analyzing the sequence and using an inhibitor we were able to designate the pathways, as NHEJ and microhomology-mediated end joining (MMEJ). Moreover, we have established that these pathways have different repair kinetics. Therefore, this target sequence can be used as a reporter for the preference of repair pathways.

In CHAPTER 6 we investigate the effect of DNA-protein packaging on Cas9 cutting and subsequent repair. We inserted the pathway reporter sequence at ~1500 different genomic locations. Our data indicate that the chromatin state of the target sequence influenced the indel frequency and pathway preference. We found that DSBs positioned

close to the periphery of the cell nucleus (LADs) are more often repaired by the MMEJ pathway compared to DSBs found closer to the centre of the cell nucleus (inter-LADs). In the presence of donor DNA, the participation of the MMEJ pathway is mostly reduced, although DSBs in LADs are still repaired with higher frequency by MMEJ than inter-LADs. These quantitative analyses contribute to a better understanding of DSB repair and the impact of chromatin on this process.

In CHAPTER 7 the results are discussed in relation to the latest developments in the field of genome editing. New research strategies are presented that can contribute to a better understanding of the influences of the chromatin environment on the process of double-strand break repair.

NEDERLANDSE SAMENVATTING

Elke individuele cel bevat DNA, de drager van de erfelijke informatie die is geschreven in een vierletterige code (het totaal aan DNA in de kern wordt genoom genoemd). De DNA-code beschrijft verschillende elementen. Een gen is bijvoorbeeld een stuk DNA dat een eiwit beschrijft, een molecuul dat een functie in of buiten de cel uitvoert. Wanneer de DNA-informatie van een gen wordt gelezen, produceert de cel het eiwit. Andere delen van het DNA bevatten weer andere informatie, zoals de code voor de regulatie van de eiwitexpressie of ruimtelijk vouwing van het DNA. Al deze informatie samen maakt dat het DNA twee meter lang is. Om dit in de celkern te passen wordt het strak opgevouwen op meerdere niveaus. De vouwing en de aanwezigheid van eiwitten en modificaties op het DNA (samen chromatine genoemd) hebben invloed op de functie van het DNA. Bijvoorbeeld, strak opgevouwen DNA is minder gemakkelijk te lezen dan losser gevouwen DNA.

Wijzigen van de DNA-code in cellen kan leiden tot gewijzigde functies. Deze veranderingen kunnen bijvoorbeeld leiden tot gunstige eigenschappen voor landbouwgewassen of kunnen ziekmakende genen in kankercellen uitschakelen. De ontwikkeling van nieuwe technologieën maakt het mogelijk het DNA van verschillende organismen op geselecteerde locaties aan te passen. De populairste techniek in de moleculaire biologie om genetische aanpassingen te maken is CRISPR/Cas9. Het CRISPR-systeem bestaat uit een gids-RNA dat een specifieke volgorde in het genoom (sequentie) herkent en de Cas9-nuclease die daar een nauwkeurig gedefinieerde dubbel-strengsbreuk (DSB) aanbrengt. Als reactie zal de cel deze breuk herstellen. In zoogdiercellen is het meest gebruikte reparatiemechanisme non-homologous end joining (NHEJ), die de gebroken uiteinden met elkaar verbindt. Deze reparatie is niet altijd nauwkeurig waarbij extra DNA-letters (nucleotiden) kunnen worden ingevoegd (inserties) of worden verwijderd (deleties) op de plaats van de breuk. Dit soort mutaties worden aangeduid als indels. Echter, de effectiviteit en het mutatiespectrum kunnen aanzienlijk variëren afhankelijk van de efficiëntie waarmee een DSB wordt geïntroduceerd en de manier waarop deze wordt gerepareerd. HOOFDSTUK 1 geeft een overzicht van de genoombewerking door CRISPR/Cas9 en de factoren die van invloed zijn op de efficiëntie van deze methode.

Bij experimenten om het genoom te bewerken met CRISPR/Cas9 is het belangrijk om de frequentie van de geïntroduceerde indel-mutaties te kennen. Omdat die moeilijk te voorspellen is, is een eenvoudige en snelle test nodig om de geïnduceerde mutaties nauwkeurig te karakteriseren en te kwantificeren. In HOOFDSTUK 2 beschrijven we de TIDE-methode. Dit is een snelle en gemakkelijk toepasbare methode die indels nauwkeurig identificeert en kwantificeert na een CRISPR/Cas9-bewerking. TIDE vereist slechts een paar PCR-reacties en twee standaard Sanger-sequencing reacties, een techniek is om de



volgorde van nucleotiden van een stuk DNA af te lezen. De sequentiepatronen worden vervolgens geanalyseerd door een speciaal ontwikkeld ontledingsalgoritme om de individuele mutaties te identificeren.

Voor het genereren van specifieke mutaties door middel van CRISPR/Cas9-genomebewerking kan een geïntroduceerde DSB ook hersteld worden met behulp van donor-DNA. Dit is een ontworpen sequentie die, op enkele gewenste mutaties na, bijna identiek is aan de locatie waar de breuk wordt gemaakt, de doelwitlocatie. In de aanwezigheid van het donor-DNA kunnen de ontworpen mutaties tijdens de reparatie in het genoom worden geïntegreerd. Dit proces verloopt via een ander reparatiemechanisme dat *homology directed repair (HDR)* wordt genoemd, en het gebruikt daarbij de overlappende sequentie tussen het genoom en het donor-DNA. Deze zeer nauwkeurige strategie is de laatste tijd steeds populairder geworden, vanwege de mogelijkheid om specifieke mutaties te creëren en ze te bestuderen of om ziekteverwekkende mutaties te corrigeren. In HOOFDSTUK 3 bespreken we een tweede methode, TIDER, die de frequentie kan schatten van elke ontworpen mutatie(s).

Voor beide methoden die in HOOFDSTUK 2 & 3 zijn gepresenteerd hebben we bijbehorende interactieve webtools ontwikkeld die door de snelle analyse rationeel de ontwerpstrategieën voor het bewerken van een genoom vergemakkelijken. TIDE en TIDER maken het eenvoudig om de werkzaamheid van een ontworpen gids-RNA te testen en om het type, aantal en frequentie van de verwachte mutaties te bepalen. HOOFDSTUK 4 biedt gedetailleerde procedures en optimalisatiestappen om mogelijke problemen bij het gebruik van deze twee methoden te ondervangen.

Het is nog niet precies duidelijk hoe het proces van het knippen van het DNA door CRISPR/Cas9 en de daaropvolgende reparatie van de DSB verloopt. Hoe snel vinden deze gebeurtenissen plaats en hoe nauwkeurig is het DSB-herstel? We hebben een strategie ontwikkeld om de kinetiek en betrouwbaarheid van DNA-herstel op enkele locaties in menselijke cellen direct te meten (HOOFDSTUK 5). Hiervoor hebben we een cellijn gemaakt met een nauwkeurig gecontroleerde, induceerbare Cas9 die het mogelijk maakte de vorming van DSB's en hun herstel in de tijd te volgen in een locatie naar keuze. Percentages met intact, gebroken en hersteld DNA worden gekwantificeerd door een combinatie van DNA-sequencingtechnologie en een test om niet-gerepareerde, gebroken uiteinden te detecteren. De resultaten worden geanalyseerd in een wiskundig model dat de belangrijkste parameters voor reparatiekinetiek en -nauwkeurigheid schat. Onze aanpak onthulde dat reparatiesnelheden varieerden per doelwitsequentie, met halfwaardetijden tot ~10 uur, wat aangeeft dat de reparatie traag is. Bovendien laten onze gegevens zien dat reparatie van de DSB's vaak foutief is. Interessant genoeg vonden we dat het spectrum van

geïntroduceerde indels tijdens reparatie in hoge mate afhankelijk is van de sequentie van het doelwit van de CRISPR. Dezelfde doelsequentie verwierf twee specifieke indels, die elk een apart mechanisme representeerde waarmee de breuk wordt hersteld. Door de indel-sequenties te bestuderen en een remmer te gebruiken konden we de gebruikte routes toekennen aan NHEJ en de microhomology-mediated end joining (MMEJ). Bovendien hebben we vastgesteld dat deze routes een verschillende reparatiekinetiek hebben. Daarom kan deze doelwitsequentie gebruikt worden als een verslaggever (reporter) van de voorkeur voor deze twee reparatiemechanismen.

In HOOFDSTUK 6 onderzoeken we het effect van DNA-eiwitpakking op het knippen door Cas9 en de daaropvolgende reparatie. We hebben de reportersequentie van de reparatiemechanismen ingevoegd op ~1500 verschillende locaties in het genoom. De resultaten geven aan dat de chromatinestatus de indel-frequentie en de voorkeur van de herstelroute beïnvloedt. We hebben ontdekt dat DSB's die dichtbij de periferie van de celkern geïmponeerd zijn (LAD's), vaker worden hersteld door de MMEJ-route dan DSB's die dichtbij het centrum van de celkern worden gevonden (inter-LADs). In aanwezigheid van donor-DNA wordt de deelname aan het MMEJ-route meestal verminderd, hoewel DSB's in LAD's nog steeds met een hogere frequentie door MMEJ worden gerepareerd dan in inter-LAD's. Deze kwantitatieve analyses dragen bij aan een beter begrip van DSB-reparatie en de impact van chromatine op dit proces.

In HOOFDSTUK 7 worden de resultaten besproken in het licht van de nieuwste ontwikkeling op het gebied van genoombewerking. Nieuwe onderzoeksstrategieën worden gepresenteerd die kunnen bijdragen aan een beter begrip van de invloed van de chromatineomgeving op het proces van herstel van een breuk in de dubbele streng.



ABBREVIATIONS

3C	chromosome conformation capture
4-OHT	4-hydroxytamoxifen
A	adenine
A-EJ	alternative end joining
ATM	ataxia telangiectasia-mutated protein
ATR	ATM and Rad3-related kinase
BLESS	<i>in situ</i> breaks labelling, enrichment on streptavidin, and next-generation sequencing
BLISS	breaks labelling <i>in situ</i> and sequencing
bp	base pair
C	cytosine
Cas	CRISPR-associated
ChIP	chromatin immunoprecipitation
C-NHEJ	classical non-homologous end joining
CRISPR	clustered regularly interspaced short palindromic repeat
crRNA	CRISPR RNA
CtIP	CtBP-interacting protein
dCas9	catalytically dead Cas9
DDR	DNA damage response
DNA-PK	DNA-dependent protein kinase
DNMT	DNA methyltransferase
Dox	doxycycline
DSB	double-strand break
FACS	fluorescence activated cell sorting
G	guanine
gDNA	genomic DNA
GFP	green fluorescent protein
gRNA	guide RNA
H3K4me2	histone H3 dimethylation on lysine 4
H3K9me2	histone H3 dimethylation on lysine 9
H3K9me3	histone H3 trimethylation on lysine 9
H3K27me3	histone H3 trimethylation on lysine 27
H3K36me3	histone H3 trimethylation on lysine 36
H4K20	histone H4 lysine 20
HAT	histone acetyltransferase
HDAC	histone deacetylase
HDM	histone demethylase
HDR	homology directed repair
HMT	histone methyltransferase
HP-1	heterochromatin protein 1
HR	homologous recombination
iLAD	inter lamin associated domain

indel	insertions and deletions
IR	ionizing radiation
IRES	internal ribosome entry site
KAP-1	KRAB-associated protein 1
KI	knock in
KO	knock out
KRAB	Kruppel-associated box domain
LAD	lamina associated domain
LP	landing pad
LM-PCR	ligation-mediated PCR
Mb	megabases
MMEJ	microhomology-mediated end joining
MRN	Mre11, Rad50, Nbs1
NGS	next generation sequencing
NL	nuclear lamina
<i>Nm</i>	<i>Neisseria meningitidis</i>
PAM	protospacer adjacent motif
PB	PiggyBac
PuroR	puromycin resistance cassette
RGENs	RNA-guided endonucleases
RNAi	RNA interference
RPA	replication protein A
RPR	repair pathway reporter
<i>Sa</i>	<i>Staphylococcus aureus</i>
sgRNA	single guide RNA
<i>Sp</i>	<i>Streptococcus pyogenes</i>
ssDNA	single-stranded DNA
ssODN	single-stranded oligodeoxynucleotides
SSTR	single-stranded template repair
<i>St</i>	<i>Streptococcus thermophiles</i>
T	thymine
TALEN	transcription-activator-like effector nuclease
TIDE	tracking of indels by decomposition
TIDER	tracking of insertions, deletions and recombination events
TRIP	thousands of reporters integrated in parallel
U	uracil
UMI	unique molecular identifier
UTR	untranslated region
tracrRNA	transactivating CRISPR RNA
XLF	XRCC4-like factor
XRCC4	X-ray cross-complementing protein 4
wt	wild-type
ZFN	zinc finger nuclease

CURRICULUM VITAE

Eva Karina Brinkman was born March 5, 1987 in Vlaardingen, the Netherlands. She was awarded a Atheneum degree at S.G. Spieringshoek in Schiedam in 2005. Eva started to study Life Science & Technology at the Delft University of Technology and University Leiden in 2005. During her studies Eva did internships in the group of Dr. Claude Backendorf at the Molecular Genetics group at the Leiden Institute of Chemistry. She studied the dynamics of chromatin-modifying enzyme BRD2 in living cells. Then continued with her Master internship at the Molecular Radiation Biology group of Prof. dr. Roland Kanaar at the Department of Cell Biology & Genetics at the Erasmus Medical Center. Here, Eva investigated the differentially regulated UV damage response in the cell cycle. For a company internship she went to DSM Food Specialties in Delft to work on the improvement of microorganisms that degrade plant biomass for a more efficient production of second generation biofuels. After that she participated in 2010 in a scientific biology student competition, iGEM, where my team finished as one of the finalists, earned the best presentation award at MIT, USA and won the UfD-E.ON teamwork award at the TU Delft, the Netherlands. She completed her Master degree in September 2011. That year she also co-founded Science Matters, a company that is engaged in knowledge transfer of life sciences. Projects have been developed a.o. H2020 Marie Skłodowska-Curie grants, Boijmans van Beuningen museum and the Rathenau Instituut. In 2012 she joined the group of Prof. dr. Bas van Steensel at the Netherlands Cancer Institute Antoni van Leeuwenhoek in Amsterdam. Together with co-inventors Tao Chen and Bas van Steensel, Eva developed the TIDE and TIDER software and granted an exclusive license hereof to Desktop Genetics in 2018.

LIST OF PUBLICATIONS

1. **Brinkman, E.K.**, van Steensel B., Rapid quantitative evaluation of CRISPR genome editing by TIDE and TIDER. Springer Nature Methods in Molecular Biology, (in press)
2. Chen, Y., Zhang Y., Wang Y., Zhang L., **Brinkman E.K.**, Adam S.A., Goldman R., van Steensel B., Ma J. *et al.*, Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. The Journal of cell biology, (2018).
3. **Brinkman, E.K.**, Chen T., de Haas M., Holland H.A., Akhtar W., van Steensel B., Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. Molecular cell 70, 801-813 e806 (2018).
4. **Brinkman, E.K.**, Kousholt A.N., Harmsen T., Leemans C., Chen T., Jonkers J., van Steensel B., Easy quantification of template-directed CRISPR/Cas9 editing. Nucleic acids research, (2018).
5. Janssen, A., Breuer G.A., **Brinkman E.K.**, van der Meulen A.I., Borden S.V., van Steensel B., Bindra R.S., LaRocque J.R., Karpen G.H., A single double-strand break system reveals repair dynamics and mechanisms in heterochromatin and euchromatin. Genes Dev 30, 1645-1657 (2016).
6. **Brinkman, E.K.**, Chen T., Amendola M., van Steensel B., Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic acids research 42, e168 (2014).
7. Warmerdam, D.O., **Brinkman E.K.**, Marteiijn J.A., Medema R.H., Kanaar R., Smits V.A., UV-induced G2 checkpoint depends on p38 MAPK and minimal activation of ATR-Chk1 pathway. J Cell Sci 126, 1923-1930 (2013).
8. **Brinkman, E.K.**, Schipper K., Bongaerts N., Voges M.J., Abate A., Wahl S.A., A toolkit to enable hydrocarbon conversion in aqueous environments. J Vis Exp, e4182 (2012).



PHD PORTFOLIO

Name of the PhD student: ir. Eva Karina Brinkman
Erasmus MC Dept: Cell Biology
Graduate School Onderzoekschool Oncologie Amsterdam
PhD period: 2012-2018
Promoter: Prof.dr. Bas van Steensel

1. PhD training

Courses

2013 OOA Course: 'Introduction to R and statistics', Amsterdam
2013 Summer School: 'Chromatin and Systems Biology', Spetses
2014 OOA Course: 'Replication stress and chromosomal instability', Amsterdam
2015 OOA Course: 'Epigenetics and non-coding RNA', Amsterdam
2015 EMBO Practical Course: 'Analysis of High-Throughput Sequencing Data',
Cambridge
2016 6th Winterschool of TRR8, Kleinwarsental

Seminars and workshops

2012-2014 OOA PhD student retreat (poster/oral presentation)
2012-2018 NKI research club meetings (oral presentation)
2012-2018 Weekly discussion meetings (oral presentation)
2012-2018 International NKI seminar, lunch meetings with speaker
2014, 2017 1st and 2nd NKI CRISPR meeting (oral presentation)
2016 Erasmus MC & Leids Universitair Medisch Centrum DNA repair meeting
(oral presentation)

(Inter-)national conferences

2012-2018 Cancer Genomics Center annual meeting, Amsterdam
2012-2018 Dutch Chromatin Day, the Netherlands (oral presentation/2e place poster prize)
2013 5th EMBO meeting, Amsterdam
2014 11th EMBL Conference Transcription and Chromatin, Heidelberg (poster)
2015 IMB DNA repair & Genome Stability in a Chromatin Environment, Mainz (poster)
2016 10th Quinquennial Conference on Responses to DNA damage, Egmond aan Zee
(1e place poster prize)
2016 Genome engineering: The CRISPR-Cas Revolution, Cold Spring Harbor
(oral presentation)

- 2017 Functional Analysis of Sequence Variants in Hereditary Breast-and Ovarian Cancer Genes, Amsterdam (poster)
- 2017 4th International TRR81 Symposium on Chromatin Changes in Differentiation and Malignancy, Egmond aan Zee (poster)
- 2017 Keystone Symposia meeting on Genomic Instability and DNA Repair, Santa Fe (poster)
- 2018 Innovation for Health – premier event in Health & Life Sciences, Rotterdam (shared 2e place poster pitch)
- 2018 EMBO workshop: Chromatin dynamics and nuclear organization in genome maintenance, Strasbourg (shared 1e place poster prize)
- 2018 Systems Epigenetics: Towards Precision Cancer Medicine, Amsterdam (oral presentation)

2. Teaching

- 2016 Supervision Master student Oncology of VUmc Amsterdam (8 months of guidance in lab work, presenting, writing)
- 2016-2017 Introduction–day first year Bachelor students Molecular Life Sciences of Utrecht University (oral presentation)
- 2016-2017 EMBO Course on Genome Engineering: CRISPR/Cas, Heidelberg (invited speaker)
- 2018 Project pitch for second year Biomedical research student of Hogeschool Avans Breda (oral presentation)



