Published in final edited form as:

Trends Biotechnol. 2015 February ; 33(2): 132–140. doi:10.1016/j.tibtech.2014.12.001.

Quantifying On and Off-Target Genome Editing

Ayal Hendel^{1,3}, Eli J Fine^{2,3}, Gang Bao², and Matthew H Porteus^{1,*}

¹Department of Pediatrics, Stanford University, Stanford, CA 94305, USA

²Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA 30332, USA

Abstract

Genome editing with engineered nucleases is a rapidly growing field, thanks to transformative technologies that allow researchers to precisely alter genomes for numerous applications, including basic research, biotechnology, and human gene therapy. While the capability to make precise and controlled changes at specified sites throughout the genome has grown tremendously in recent years, we still lack a comprehensive and standardized battery of assays for measuring the different genome editing outcomes created at endogenous genomic loci. Here, we review the existing assays for quantifying on and off-target genome editing and describe their utility in advancing the technology. We also highlight unmet assay needs for quantifying on and off-target genome editing outcomes and discuss their importance for the genome editing field.

Keywords

Gene editing; Gene targeting; Homologous recombination; non-homologous end-joining; ZFNs; TALENs; CRISPR/Cas9; RNA guided endonucleases; RGENs

Targeted genome-editing using engineered nucleases

Genome editing with engineered nucleases is a transformative technology for targeted modification of essentially any genomic DNA sequence [1]. The engineered nucleases generate genomic site-specific double-stranded breaks (DSBs), which then can be resolved by the cell's endogenous DNA repair mechanisms [2]. These cellular DNA repair mechanisms of non-homologous end-joining (NHEJ) and homologous recombination (HR) can be exploited to introduce the desired genomic alteration (See Glossary) [3]. Genome editing by NHEJ generally results in small insertions and/or deletions (indels) at the site of the break. If the DSB is within the coding region of a gene, the insertions/deletions can create a frameshift mutation [4] that may knockout gene function. One can create a defined deletional event by using engineered nucleases to create two DSBs on the same

©2014 Elsevier Ltd. All rights reserved.

*Correspondence: mporteus@stanford.edu (M.H.P.).

³These authors contributed equally to this work

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

chromosome. These defined deletions can be small (tens of base pairs) or large (several megabases) and can be used, not only to knock out gene coding sequences, but also to knock out critical regulatory regions or other non-coding genetic elements [5].

In genome editing by HR, an exogenous DNA donor is introduced along with the engineered nuclease. The provided donor has homologous sequences flanking the DSB: >400 base pairs of homology [6] in the case of a plasmid or 70–130 bp in the case of single-stranded oligodeoxynucleotides (ssODNs) [7]. For reasons that remain unclear, the cell's HR machinery will use the supplied donor sequence as template for repair, thereby creating precise nucleotide changes at or near the site of the break [8]. The donor DNA can be used to introduce precise nucleotide substitutions or deletions, endogenous gene labeling, and targeted transgene addition [9]. In contrast to the random indels created by NHEJ, genome editing by HR gives precise nucleotide resolution to the editing event.

Currently, there are four principal families of engineered nucleases used for genome editing: Zinc Finger Nucleases (ZFNs) [10], Transcription Activator-Like Effector Nucleases (TALENs) [11], Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR/Cas9) or RNA-guided endonucleases (hereafter called "RGENs") [12, 13], and engineered meganucleases [14]. In addition, hybrid versions of these platforms have also been reported including mega-TALs and RGEN-FokI nuclease fusions [15–17]. Despite the fact that these engineered nuclease platforms are structurally different, they share the ability to introduce site-specific DSBs and therefore can induce the desired targeted genomic modification through the mechanisms described above. One intriguing difference, which may subtly affect the genome-editing outcome, is that each platform creates a slightly different type of DSB. Engineered meganucleases create breaks with precise 3' overhangs [18], ZFNs create breaks with precise 5' overhangs [19], and TALENs create breaks with 5' overhangs, for the most part [20], but may also create 3' overhangs. Finally, standard RGENs create blunt breaks [21] but can be designed to create DSBs with either 3' or 5' overhangs when used in a paired "nickase" configuration [22, 23]. The frequency of indels at the site of DSBs that have 3' overhangs can be substantially increased by the co-expression of TREX2 [24], a 3' single-strand exonuclease [24].

Each of these platforms has their own potential advantages and disadvantages, a full discussion of which goes beyond the scope of this review. Briefly, however, engineered meganucleases may have the greatest specificity but are generally the most difficult to re-engineer to recognize novel target sites. ZFNs were the critical platform in the early development of genome editing but highly active and specific ZFNs are also relatively difficult to engineer. TALENs can be engineered using multiple different methods, which usually require 1–2 weeks of relatively sophisticated molecular biology. In general, ~70% of engineered TALEN pairs show reasonable on-target activity when designed using appropriate tools [25], and the platform generally has greater specificity than ZFNs. Finally, the RGEN's are the simplest to engineer requiring only basic molecular biology skills with a success rate of 30–66% [26]. The specificity of RGEN's in comparison to ZFNs and TALENs remains under active investigation; comparison assays have yielded mixed results. It is clear, however, that one can improve the specificity by screening multiple nucleases for a given genetic locus—an approach particularly amenable to RGENs, given their ease of

construction. ZFNs, TALENs, and RGENs can all be purchased from commercial sources for research purposes. Although we have outlined some general guidelines, based on published results from multiple labs, as well as our own extensive experience using nucleases from all four platforms, these guidelines are not exhaustive. It is important to note that every situation is unique. Variables such as the specific nuclease, target site, cell type, method of delivery, among others, can all interact to determine which nuclease platform is best suited for a particular purpose.

Measuring efficacy of genome editing

An important aspect of evaluating the activity of a given nuclease or a new nuclease platform is quantitative measurement of both on-target and off-target activity. These quantitative measurements are crucial when either optimizing the genome editing process at a specific site or moving the process into a new system, such as a new delivery method or cell type.

Analysis of Genome Editing by Mutagenic NHEJ

The insertion or deletion mutations that are introduced by NHEJ, subsequent to a DSB caused by an engineered nuclease, can range in size from a few or to tens of nucleotides. Detecting these mutations, for example through a gel-based assay, can provide an indirect measurement of the cleavage potential of the engineered nuclease. In the gel-based mutation detection assay (Figure 1A), the most common protocols use the CEL-I nuclease or T7 endonuclease I (T7E1) [27] [28] [29]. The gel-based mutation detection assay is very rapid and economical, but it has a detection limit of ~1–2% nuclease-modified alleles and cannot be used to reveal the mutation spectra induced by the engineered nuclease. Other methods for detection of engineered nuclease-induced mutation include fluorescent PCR assay [30], DNA melting analysis[31] and restriction fragment length polymorphism (RFLP) analysis[32].

Mutagenic NHEJ events can also be measured by DNA sequencing. In sequencing based assays, the PCR amplicons are sequenced either by Sanger techniques or next generation techniques [33] [34]. The sequencing based approaches are more sensitive, detecting mutation frequencies reliably to 0.01%, and give direct information about the sequence content of the indel in contrast to gel-based assays. It should be noted that PCR based-methods for measuring mutagenic NHEJ can lead to underestimates of on-target activity. Nucleases can introduce large deletions that span beyond the boundaries of the PCR amplicon, and these events therefore would not be detected. Similarly, a large insertion at the on-target site makes it less likely that the sequence will be amplified and therefore would not be detected.

Although the sequencing method is more sensitive and gives more detailed information about the genome editing event, these assays are indirect measures of nuclease activity, as they depend on the mutagenic propensity of the repair machinery in the cell type being used. Thus, in order to compare nuclease activities, the same cell type must be used in order to control for the differences in fidelity of DNA repair mechanisms across cell types. Many of the cell types used to evaluate the activity of engineered nucleases were empirically

identified as useful due to high mutagenic activity (i.e. U2OS or HEK-293T cells), ease of delivery of the nuclease (K562 cells), or both. In general, a highly active nuclease should have greater than 25% indel activity in U2OS or HEK-293T cells under optimal delivery conditions.

Analysis of targeted genome-editing using donor DNA through HR

It is more challenging to measure genome editing by HR than by mutagenic NHEJ. The assay needs to distinguish between HR-mediated events and NHEJ-mediated events, which will occur simultaneously in the cell population. The two events may occur at a relatively equal frequency, but in certain cell types and under certain conditions, the NHEJ mediated events may predominate [6]. One method to quantitatively measure HR mediated outcomes is via an RFLP assay, by introducing synonymous SNPs that create a restriction enzyme site within the donor DNA (Figure 1B) [32]. Targeting efficiency is calculated on the basis of the ratio of cleaved product to total product. Under optimal conditions, RFLP assays for quantifying HR mediated outcomes have a sensitivity of 1–2%, like the enzymatic assays for measuring mutagenic NHEJ events.

An alternative approach for estimating the efficiency of engineered nuclease mediated gene targeting through HR is by using donor DNA that includes a constitutively active reporter gene in between the homology arms [35]. Upon successful homologous recombination, the reporter gene is stably integrated into the targeted locus. This way, targeted cells can be analyzed using flow cytometry (FACS) analysis. A critical control in this assay is to introduce the donor DNA without a nuclease in order to determine the background frequency of random integration. Generally, an active nuclease will increase the frequency of reporter gene positive cells by greater than 10-fold over the background random integration frequency. If there is a single copy of the target in the cell, this assay is an accurate measure of allele targeting frequency. However, if there are multiple copies of the target, the targeted integration assay can sometimes underestimate the allele targeting frequency.

A number of assays have now been developed for simultaneous measurement of mutagenic NHEJ and gene targeting by HR using donor DNA. The Traffic Light Reporter (TLR) system (Figure 1C), which uses a reporter cell line, was the first system used to simultaneously measure mutagenic NHEJ and gene targeting by HR using donor DNA [36, 37]. Briefly, the system is based on using a fluorescent marker to score for either mutagenic NHEJ (mCherry) or HR-mediated genome editing (GFP). The TLR can be used to provide a simple, rapid and quantitative readout. Although the TLR is a very sensitive assay for measuring genome editing outcomes, the need to generate a reporter locus prevents measurement at endogenous target loci. Thus far, no one has reported the use of the TLR assay in human primary cells.

High-throughput sequencing of endogenous loci makes it possible to develop assays that overcome these limitations. Illumina [7] and 454 [38] sequencing have been used to measure HR and NHEJ outcomes when ssODNs or plasmids with short homology arms are used as donor templates. But the read-length limitations of these platforms do not allow analysis of longer arms of homology, which drive more efficient HR and provide the flexibility to target

long gene cassettes. As an important side note, genome editing using ssODNs does not occur mechanistically through the canonical HR machinery, and the mechanism by which editing by ssODN occurs remains under investigation [39]. In order to overcome the read-length limitation, Hendel *et al.* [6] developed a new method for measuring genome editing outcomes at endogenous loci using single molecule real time (SMRT) DNA sequencing (Figure 1D), which provides average read lengths of 8.5 kb. This technique allows for analysis of gene editing frequencies when using donor templates with long arms of homology. In this study, Hendel *et al.* demonstrated that the rate of genome editing by HR is improved by using homology arms 400 base pairs or longer, which exceeds the current reliable read length capability of Illumina based methods. The SMRT DNA sequencing strategy offers three principle advantages over other currently available techniques: i) Sensitive measurement of genome editing in any cell type, including primary stem cells, without the need to make a stable reporter cell line, ii) measurement of modifications at endogenous loci regardless of transcriptional status, and iii) long sequencing read-lengths that allow insight into a wide range of DNA repair outcomes when donor templates with long arms of homology are used. Table 1 summarizes the features, pros and cons of the typical assays for quantifying on-target genome editing.

Consequences and types of off-target activity

Genome editing tools can be engineered to make extremely well-defined alterations to the intended target genomic locus, but one potential complication is that the engineered nuclease will create other, unintended genomic changes. This “off-target” activity of the nucleases occurs fundamentally because the nucleases lack perfect specificity, i.e. they are able to bind to sequences other than the intended DNA target. Engineered nucleases have different relative affinities for different nucleotide sequences and thus the overall specificity of an engineered nuclease is the relative affinity for its intended target site with respect to its affinity for the population of other off-target sites in the genome. There are various strategies for mitigating the effects of off-target activity by trying to improve specificity of the nucleases (Box 1).

The most common manifestation of off-target activity is as a small indel after a DSB is repaired in a mutagenic fashion by NHEJ. The size of the indel can depend on the nuclease platform used and the duration that the nuclease is expressed. However, off-target activity has consequences at larger scales and can alter DNA in other ways (Box 2). Other documented alterations include incorporation into the genome of exogenously supplied DNA (such as a donor template, contaminant bacterial DNA remaining after plasmid production [6], nuclease expression vectors, or bovine DNA presumably from fetal bovine serum in culture media), deletion of large (i.e. many kilobases) stretches of chromosomal sequence [40, 41], inversions [42], incorporation of stretches of sequence (~100 bp) from alternate locations in the human genome [6], or gross chromosomal translocations [43].

Population Based Functional Assays for Off-Target Effects

In addition to identifying specific off-target changes induced by engineered nucleases using bioinformatics and sequencing, there are several assays that can measure the functional

toxicity of engineered nuclease expression without having to predict potential off-target sites (Box 2). These assays include measuring whether the nucleases induce overt cellular apoptosis [41], whether the expression of the nuclease causes the cells have a decreased replicative fitness (either because of cell death or cell cycle arrest) compared to cells not expressing the nuclease [44, 45], whether the nuclease perturbs the cell cycle characteristics of the population [41], or whether the nucleases create extra DSBs in the genome [45–47]. All of these assays provide different approaches to evaluating the functional toxicity of an engineered nuclease delivered into a specific cell type using a specific method.

Cancerous Transformation and Clonal Expansion

For clinical applications, understanding the exact location and number of alterations to the genome may be less critical than assessing whether the off-target effects of the nuclease cause the cell to transform into a cancerous phenotype and begin rapid clonal expansion. A common *in vitro* test is the soft agar transformation assay [48], which involves plating single cells that have been treated with nuclease to determine what fraction of the cell population has gained the ability to rapidly expand in an anchorage-independent manner. Another method, which can be applied either *in vitro* or *in vivo*, involves integrating a randomized DNA barcode sequence into the genome of cells through viral delivery or nuclease mediated integration. Each cell contains a unique barcode, thus allowing the diversity and composition of barcodes remaining in the cell population to be tracked over time. If a single barcode (or a small set of them) begins to dominate the results, this indicates that a cell has undergone rapid clonal expansion, possibly due to off-target effects of the nuclease causing a transformation to a phenotype that shares characteristics with malignant transformation [49].

Predicting Off-Target Site Locations

Unlike the pseudo-random nature of retro or lenti-viral integration sites, nuclease off-target activity is presumed to be a direct result of the nuclease binding a nearby sequence with some level of homology to the intended target, and subsequently inducing a DSB. Theoretically, the entire genome could be considered a potential off-target site, but in practice the binding energy for the engineered nuclease across the entire genome would be so low that bioinformatic-based predictions will not be useful. This principle also suggests that given the large size of the genome, an occasional indel might be created at a very divergent target site but that this rare event would not be reproducible from cell to cell. It has not been conclusively ruled out that low levels of completely non-specific nuclease-DNA interactions could cause infrequent mutagenic events. However, empirically, several studies comparing untreated cells to cells transfected with the Cas9 nuclease or a FokI:Cas9 fusion (but no sgRNA) have found no evidence of increased mutagenesis at any sites interrogated [17, 50].

Therefore, rather than a broad distribution of off-target activity evenly spread throughout the genome, nucleases tend to induce off-target activity at certain locations or “hot-spots”, known as “off-target sites”. In general, off-target activity is consistent in frequency and location for a given nuclease in a given cell type, and can share substantial similarity across

different cell types (of the same species) [51]. This allows an “off-target profile” of known genomic locations to be constructed by compiling data from multiple assays that can then be used to guide measurements under the desired experimental conditions.

All approaches for identifying off-target activity follow the same basic pattern (Figure 2). First, an *in vitro*, cellular, or *in silico* technique is used to predict locations in the genome that might be prone to off-target activity, which yields a list of potential off-target sites. Next, cells are treated with the nuclease and genomic DNA is harvested for analysis. Finally, potential sites are PCR amplified and analyzed for mutations using enzyme or sequencing based assays. Sites with a mutation rate above the background levels observed in mock-treated cells are termed *bona fide* off-target sites.

Several elegant methods have been devised to predict nuclease off-target sites based on characterizations of the nucleases *in vitro* or in cells, but these methods are technically challenging, thus limiting their broader utility. Systematic Evolution of Ligands by eXponential Enrichment (SELEX) has been used to identify the sequences individual nucleases prefer to bind *in vitro* [52–54]. The genome can subsequently be searched for near matches to these sequences to predict potential off-target sites. A semi-randomized library of oligonucleotides can be exposed to nucleases to identify the sequences that are able to be cleaved *in vitro* then the genome can be searched for exact matches to those sequences [55, 56]. Alternatively, a machine learning approach can be applied to the full breadth of sequences that the nucleases cleaved in order to predict likely sites of off-target activity [57, 58].

Although examining nucleases *in vitro* allows for a more comprehensive and controlled analysis of the specificity of the DNA binding domains, the genomic context of a putative site in cells (e.g. chromatin accessibility, methylation, histone modifications) is known to affect the rate of potential off-target activity. Similarly, the specific *in vitro* conditions of these assays (salt content and concentration, temperature, duration of assay, etc.) will skew results towards the identification of certain types of sites. Therefore, assays have been developed which take place inside of cells in order to predict off-target activity in a given cell type. One assay relies on trapping integrase-deficient lentiviruses (IDLVs) or adeno-associated viruses (AAVs) at the site of DSBs and subsequently using integration mapping techniques to determine where in the genome the IDLVs or AAVs are found, presumably at locations of nuclease activity [59, 60]. Another *in cellula* approach is to use chromatin immunoprecipitation to pull down the nuclease protein, followed by sequencing the DNA fragments to which the nuclease was bound (ChIP-Seq) and mapping those fragments to their originating locations in the genome [61–63]. Unfortunately, these *in cellula* assays have generally had less success locating *bona fide* off-target sites because they either are less sensitive to detecting lower levels of off-target activity [59] or yield very high false-positive prediction rates [61–63].

Because of the difficulty of implementing the *in vitro* and *in cellula* predictive techniques, *in silico* prediction tools are widely used. Early *in silico* tools relied on simple metrics such as overall homology between the intended nuclease target and other sites in the genome, which led to poor predictions that failed to uncover *bona fide* off-target sites [64–67]. However,

newer tools that incorporate biological principles and factors driving the nuclease-DNA interactions into their prediction algorithms have had much greater success in locating novel off-target sites. The PROGNOS tool [68] (<http://baolab.bme.gatech.edu/Research/BioinformaticTools/prognos.html>) has been used to locate *bona fide* off-target sites for several different ZFNs and TALENs [41, 68, 69]. A tool for RGEN off-target prediction has also been developed (<http://crispr.mit.edu>) based on observed trends in RGEN specificity [70]. Recently the COSMID tool has been created to allow searching for RGEN off-target sites containing RNA or DNA bulges (<http://crispr.bme.gatech.edu>).

Analysis of Clonal Populations

If the desired genome editing outcome is a clonal population of cells, then characterizing off-target activity will require a slightly different strategy and additional methods. But it is important to note that a lack of off-target effects observed in a clonal population does not imply that a nuclease would not display off-target activity in a non-clonal population. The most straightforward approach is to use an off-target location prediction method as described above, but then use a SURVEYOR or T7E1 assay or simple Sanger sequencing, rather than deep sequencing, to look for off-target effects. Array comparative genomic hybridization (aCGH) is a method that can examine the whole genome of a clonal population to look for copy number variations (duplications or deletions) of regions 1 kilobase or larger that may have been induced by the nucleases [71]. As the costs of DNA sequencing have decreased, it has become feasible to perform whole exome or whole genome sequencing (WGS) of a clonal population of edited cells. Several studies have documented the generation of clones that contained no nuclease-induced alterations other than the intended modification throughout the entire genome [64] or exome [44]. However, WGS has several caveats. A high level of bioinformatics expertise is necessary to discern signal from noise in the assembly of a genome, especially when looking for relatively small mutations that may have been introduced by NHEJ. Most WGS protocols are not designed with the detection of gross chromosomal rearrangements in mind, thus these may be missed, though custom protocols for detecting structural variants are being developed [72]. Because many spontaneous mutations occur in the generation of a clonal population from the parental cells, it can be difficult to determine with certainty which mutations are “natural” and which occurred as an off-target effect of the nuclease [73]. As sequencing costs continue to drop, it may become possible in the near future to easily and accurately sequence thousands of clones in order to create a complete and unbiased assessment of an off-target profile, but this remains infeasible with current technology.

An important consideration in all of these approaches to identifying small indels at engineered nuclease off-target sites is the genomic context of the process. Recent whole genome sequencing studies have shown that every individual has on the order of several hundred thousand unique indels in their genome. Moreover, the creation of indels is a process that occurs continually in cells in response to the challenges the genome faces from its own metabolic by-products (i.e. reactive oxygen species) and external stimuli. Thus, it may be the case that the indels created by engineered nucleases are a mere rounding error compared to the background number that are present and being continually created. Nonetheless, the assessment of indels at identified off-target sites does provide an important

quantitative measure of the relative specificity of optimized variants of the same nuclease or as a method to compare the relative specificity of two different nucleases.

Concluding remarks and future perspectives

The recent progress in the field of genome engineering is enabling scientists from diverse disciplines to precisely modify essentially any locus of interest, allowing researchers to advance basic science, biotechnology and human gene therapy. Nevertheless, more work needs to be done to streamline the development of genome editing projects to a wider range of applications. Improving methods for delivery of engineered nucleases and donor DNA, enhancing nuclease activity in specific cell types and loci, increasing nuclease specificity and reducing toxicity, among others, are some challenges that need to be addressed. Therefore, an important technological goal is to have a comprehensive battery of assays for quantifying on and off-target genome editing outcomes and rates. Although this need is currently addressed by a variety of methods as described in this review, there is still need for new assays that would complement the existing ones (Box 3). SMRT sequencing is a fast and cost-effective way to evaluate the outcomes of individual genome editing experiments, but it remains relatively expensive, thus making it difficult for large-scale efforts to assess on-target genome editing outcomes. A major step forward would be to develop a cost-effective, high throughput assay for large scale quantification of on-target HR and NHEJ genome editing outcomes, either through cost reduction of SMRT sequencing, increased read-lengths of Illumina sequencing, or a novel approach. This will enable screening for new conditions aimed at increasing HR over NHEJ rates for precise genome modifications.

As for assays related to quantifying off-target genome editing, an unbiased assay to detect novel chromosomal rearrangements that could be applied to bulk cell populations is needed. Whole genome sequencing currently lacks the sensitivity to detect such rare events in large populations of cells. Although we described a few assays for assessing whether the off-target effects of the nuclease can cause the cell to transform into a cancerous phenotype and begin rapid clonal expansion, this area of investigation should be expanded, particularly for human gene therapy studies. In summary, the current and future assays for quantifying on- and off-target genome editing will continue to allow the scientists of the rapidly expanding genome engineering community—who come from a large variety of scientific backgrounds—to refine and expand the utilities of these technologies for the benefit of society.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health as a Nanomedicine Development Center Award [PN2EY018244 to GB and MHP]. A.H. was supported in part by the Fund-A-Fellow postdoctoral fellowship research grant award from the Myotonic Dystrophy Foundation (MDF) and the Dean's postdoctoral fellowship award at the Stanford School of Medicine. E.J.F. was supported by the National Science Foundation Graduate Research Fellowship [DGE-1148903]. M.H.P. is supported as the Laurie Kraus Lacob Faculty Scholar in Pediatric Translational Research. We would like to thank Niraj Punjya for his assistance in generating the figures for this paper. We apologize for the need to omit many relevant citations due to space limitations.

REFERENCES

1. Baker M. Gene-editing nucleases. *Nature methods*. 2012; 9:23–26. [PubMed: 22312637]

2. Carroll D. Genome engineering with targetable nucleases. *Annual review of biochemistry*. 2014; 83:409–439.
3. Jasin M, Rothstein R. Repair of strand breaks by homologous recombination. *Cold Spring Harbor perspectives in biology*. 2013; 5:a012740. [PubMed: 24097900]
4. Bibikova M, et al. Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics*. 2002; 161:1169–1175. [PubMed: 12136019]
5. Bauer DE, et al. An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science*. 2013; 342:253–257. [PubMed: 24115442]
6. Hendel A, et al. Quantifying genome-editing outcomes at endogenous loci with SMRT sequencing. *Cell reports*. 2014; 7:293–305. [PubMed: 24685129]
7. Yang L, et al. Optimization of scarless human stem cell genome editing. *Nucleic acids research*. 2013; 41:9049–9061. [PubMed: 23907390]
8. Rouet P, et al. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*. 1994; 91:6064–6068. [PubMed: 8016116]
9. McMahon MA, et al. Gene editing: not just for translation anymore. *Nature methods*. 2012; 9:28–31. [PubMed: 22205513]
10. Porteus MH, Carroll D. Gene targeting using zinc finger nucleases. *Nat Biotechnol*. 2005; 23:967–973. [PubMed: 16082368]
11. Bogdanove AJ, Voytas DF. TAL effectors: customizable proteins for DNA targeting. *Science*. 2011; 333:1843–1846. [PubMed: 21960622]
12. Gaj T, et al. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*. 2013; 31:397–405. [PubMed: 23664777]
13. Mali P, et al. Cas9 as a versatile tool for engineering biology. *Nat Methods*. 2013; 10:957–963. [PubMed: 24076990]
14. Silva G, et al. Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Current gene therapy*. 2011; 11:11–27. [PubMed: 21182466]
15. Boissel S, et al. megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic acids research*. 2014; 42:2591–2601. [PubMed: 24285304]
16. Guilinger JP, et al. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature biotechnology*. 2014; 32:577–582.
17. Tsai SQ, et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature biotechnology*. 2014; 32:569–576.
18. Chevalier BS, Stoddard BL. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic acids research*. 2001; 29:3757–3774. [PubMed: 11557808]
19. Smith J, et al. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic acids research*. 2000; 28:3361–3369. [PubMed: 10954606]
20. Humbert O, et al. Targeted gene therapies: tools, applications, optimization. *Critical reviews in biochemistry and molecular biology*. 2012; 47:264–281. [PubMed: 22530743]
21. Jinek M, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012; 337:816–821. [PubMed: 22745249]
22. Mali P, et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology*. 2013; 31:833–838.
23. Ran FA, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*. 2013; 154:1380–1389. [PubMed: 23992846]
24. Certo MT, et al. Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. *Nature methods*. 2012; 9:973–975. [PubMed: 22941364]
25. Lin Y, et al. SAPTA: a new design tool for improving TALE nuclease activity. *Nucleic Acids Research*. 2014; 42:e47–e47. [PubMed: 24442582]
26. Doench JG, et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotech* advance online publication. 2014

27. Guschin DY, et al. A rapid and general assay for monitoring endogenous gene modification. *Methods in molecular biology*. 2010; 649:247–256. [PubMed: 20680839]
28. Qiu P, et al. Mutation detection using Surveyor nuclease. *BioTechniques*. 2004; 36:702–707. [PubMed: 15088388]
29. Kim HJ, et al. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome research*. 2009; 19:1279–1288. [PubMed: 19470664]
30. Kim H, et al. Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nature methods*. 2011; 8:941–943. [PubMed: 21983922]
31. Dahlem TJ, et al. Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. *PLoS genetics*. 2012; 8:e1002861. [PubMed: 22916025]
32. Urnov FD, et al. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*. 2005; 435:646–651. [PubMed: 15806097]
33. Kim Y, et al. TALENs and ZFNs are associated with different mutation signatures. *Nature methods*. 2013; 10:185. [PubMed: 23396284]
34. Chen S, et al. A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. *Nucleic acids research*. 2013; 41:2769–2778. [PubMed: 23303782]
35. Moehle EA, et al. Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104:3055–3060. [PubMed: 17360608]
36. Certo MT, et al. Tracking genome engineering outcome at individual DNA breakpoints. *Nature methods*. 2011; 8:671–676. [PubMed: 21743461]
37. Porteus M. Seeing the light: integrating genome engineering with double-strand break repair. *Nature methods*. 2011; 8:628–630. [PubMed: 21799496]
38. Qi Y, et al. Increasing frequencies of site-specific mutagenesis and gene targeting in Arabidopsis by manipulating DNA repair pathways. *Genome research*. 2013; 23:547–554. [PubMed: 23282329]
39. Davis L, Maizels N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111:E924–E932. [PubMed: 24556991]
40. Cradick TJ, et al. CRISPR/Cas9 systems targeting β -globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Research*. 2013; 41:9584–9592. [PubMed: 23939622]
41. Mussolino C, et al. TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic Acids Research*. 2014; 42:6762–6773. [PubMed: 24792154]
42. Lee HJ, et al. Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome research*. 2012; 22:539–548. [PubMed: 22183967]
43. Torres R, et al. Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR–Cas9 system. *Nature Communications*. 2014;5.
44. Ousterout DG, et al. Reading Frame Correction by Targeted Genome Editing Restores Dystrophin Expression in Cells From Duchenne Muscular Dystrophy Patients. *Molecular Therapy*. 2013
45. Pruett-Miller SM, et al. Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. *Molecular therapy : the journal of the American Society of Gene Therapy*. 2008; 16:707–717. [PubMed: 18334988]
46. Guo J, et al. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *Journal of molecular biology*. 2010; 400:96–107. [PubMed: 20447404]
47. Pruett-Miller SM, et al. Attenuation of zinc finger nuclease toxicity by small-molecule regulation of protein levels. *PLoS genetics*. 2009; 5:e1000376. [PubMed: 19214211]
48. Steinberg, P. Steinberg, P. *High-Throughput Screening Methods in Toxicity Testing*. John Wiley & Sons, Inc; 2013. Automated Soft Agar Colony Formation Assay for the High-Throughput Screening of Malignant Cell Transformation; p. 309-316.
49. Porter SN, et al. Lentiviral and targeted cellular barcoding reveals ongoing clonal dynamics of cell lines in vitro and in vivo. *Genome biology*. 2014; 15:R75. [PubMed: 24886633]

50. Kuscu C, et al. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature biotechnology*. 2014; 32:677–683.
51. Fu Y, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*. 2013; 31:822–826.
52. Perez EE, et al. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotech*. 2008; 26:808–816.
53. Tesson L, et al. Knockout rats generated by embryo microinjection of TALENs. *Nature Biotechnology*. 2011; 29:695–696.
54. Hockemeyer D, et al. Genetic engineering of human pluripotent cells using TALE nucleases. *Nature Biotechnology*. 2011; 29:731–734.
55. Pattanayak V, et al. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Meth*. 2011; 8:765–770.
56. Pattanayak V, et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology*. 2013; 31:839–843.
57. Sander JD, et al. In silico abstraction of zinc finger nuclease cleavage profiles reveals an expanded landscape of off-target sites. *Nucleic Acids Research*. 2013
58. Guilinger JP, et al. Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nature Methods*. 2014; 11:429–435. [PubMed: 24531420]
59. Gabriel R, et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotech*. 2011; 29:816–823.
60. Petek LM, et al. Frequent endonuclease cleavage at off-target locations in vivo. *Molecular therapy : the journal of the American Society of Gene Therapy*. 2010; 18:983–986. [PubMed: 20216527]
61. Kuscu C, et al. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature Biotechnology*. 2014; 32:677–683.
62. Wu X, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnology*. 2014; 32:670–676.
63. O'Geen H, et al. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *bioRxiv*. 2014
64. Ding Q, et al. A TALEN Genome-Editing System for Generating Human Stem Cell-Based Disease Models. *Cell Stem Cell*.
65. Sebastiano V, et al. In Situ Genetic Correction of the Sickle Cell Anemia Mutation in Human Induced Pluripotent Stem Cells Using Engineered Zinc Finger Nucleases. *STEM CELLS*. 2011; 29:1717–1726. [PubMed: 21898685]
66. Huang P, et al. Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotech*. 2011; 29:699–700.
67. Watanabe T, et al. Non-transgenic genome modifications in a hemimetabolous insect using zinc-finger and TAL effector nucleases. *Nature Communications*. 2012;3.
68. Fine EJ, et al. An online bioinformatics tool predicts zinc finger and TALE nuclease off-target cleavage. *Nucleic Acids Research*. 2013
69. Abarrategui-Pontes C, et al. Codon Swapping of Zinc Finger Nucleases Confers Expression in Primary Cells and In Vivo from a Single Lentiviral Vector. *Current Gene Therapy*. 2014; 14:12.
70. Hsu PD, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*. 2013; 31:827–832.
71. Lindsay CR, Roth DB. An unbiased method for detection of genome-wide off-target effects in cell lines treated with zinc finger nucleases. *Methods in molecular biology (Clifton, NJ)*. 2013; 1114:353–369.
72. Mijuškovi M, et al. A Streamlined Method for Detecting Structural Variants in Cancer Genomes by Short Read Paired-End Sequencing. *PLoS ONE*. 2012; 7
73. Veres A, et al. Low Incidence of Off-Target Mutations in Individual CRISPR-Cas9 and TALEN Targeted Human Stem Cell Clones Detected by Whole-Genome Sequencing. *Cell Stem Cell*. 2014; 15:27–30. [PubMed: 24996167]

74. Gaj T, et al. Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nature Methods*. 2012; 9:805–807. [PubMed: 22751204]
75. Ran FA, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*. 2013; 154:1380–1389. [PubMed: 23992846]
76. Guilinger JP, et al. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotech*. 2014; 32:577–582.

Glossary

Donor DNA

Donor DNA is an exogenous DNA segment that, when provided to cells along with engineered nucleases, can cause site specific genomic modification, cassette addition or tagging of endogenous genes. The donor DNA should have homology to the genomic target. Most donor DNA constructs contain flanking homology arms of lengths between 400 and 800 base pairs. Longer homology arms can be used, but come at the cost of increased vector size, which can decrease the efficiency of delivery via electroporation and of viral vector packaging. Traditionally donor DNA is provided to cells as plasmid DNA but also can be delivered using adeno-associated virus (AAV) or integrase-deficient lentiviruses (IDLVs) vectors. Single-stranded DNA oligonucleotides (ssODNs) as short as 40 bases, but usually 100 base pairs or longer, have been successfully used as DNA donors albeit with reduced efficiency

Engineered Meganucleases

Engineered nucleases derived from a family of naturally occurring endonucleases that have a sequence specific recognition domain

Engineered nuclease

Engineered nucleases are artificial enzymes that can be programmed to induce site specific double-strand breaks. The ability to cut defined sites in the genome enables efficient editing of genetic information. Currently, there are four principal families of engineered nucleases used for genome editing: Zinc Finger Nucleases (ZFNs) [10], Transcription Activator-Like Effector Nucleases (TALENs), Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR/Cas9) or RNA-guided endonucleases (hereafter called “RGENs”), and engineered meganucleases. In addition, hybrid versions of these platforms have also been reported including mega-TALs and RGEN-FokI nuclease fusions

HR mediated genome editing

The cell has two basic ways of repairing a double-strand break induced by engineered nucleases: non-homologous end-joining (NHEJ) or homologous recombination. Homologous recombination uses donor DNA as a template to repair the break in a “copy and paste” type mechanism. By providing an appropriately designed donor DNA, precise small or large modifications can be made to the genome (HR mediated genome editing)

NHEJ mediated genome editing

The cell has two basic ways of repairing a double-strand break induced by engineered nucleases: non-homologous end-joining (NHEJ) or homologous recombination. NHEJ functionally ligates the two ends of broken DNA together. While NHEJ is mostly error-free, it can occasionally result in insertion or deletion of DNA at the site of the break resulting in mutations at a specific site in the genome (NHEJ mediated genome editing). Insertions/deletions created by NHEJ in this way are uncontrolled in their nucleotide content and are relatively random

Off-Target Effects

Nucleases lack perfect specificity and can therefore bind and cut locations in the genome other than their intended target site, leading to unwanted genomic modifications

RNA-guided engineered nucleases (RGENs)

Engineered nucleases composed of the Cas9 endonuclease and a guide RNA

Transcription Activator-Like Effector Nucleases (TALENs)

Engineered nucleases composed of the FokI type IIS restriction enzyme catalytic domain and Transcription Activator-Like Effector (TALE) DNA-binding domains

Zinc Finger Nucleases (ZFNs)

Engineered nucleases composed of the FokI type IIS restriction enzyme catalytic domain and zinc-finger DNA-binding domains

Box 1**Mitigating off-target activity**

Off-target activity can be simplistically modeled as a function of four factors, which can be varied in an experimental setting to optimize the outcome. From basic biochemistry, two of these factors are the concentration of nuclease and the amount of time the genome is exposed to the nuclease. Studies have found that changing these parameters can improve the “specificity ratio”—defined as the ratio of on-target activity to off-target activity [70, 74]. Another factor is the number of sites in the genome homologous to the intended target site. Because sequence of a genome is not random, different DNA “words” can have dramatically different numbers (over several orders of magnitude) of homologous sites in a genome. Several online tools exist to assist in locating a target sequence within your gene of interest that is highly unique throughout the genome [68, 70]. The final factor is the intrinsic level of DNA specificity of the nuclease. Different Cas9 variants and TALEN designs [58, 75, 76] have been developed that offer improved specificity ratios (although sometimes at the cost of lower on-target activity).

Box 2**Larger scale consequences of off-target activity and methods of analysis****Effects of Off-Target Activity on cells**

The functional off-target activity of a nuclease that disrupts normal cell function can be observed over a relatively short time-span (several days). If off-target activity is significant, then this can manifest as cytotoxicity, leading to apoptosis, which can be assayed using fluorescent dyes such as 7-AAD and propidium iodide (PI) or dye-conjugates such as AnnexinV, and quantified by microscopy or flow cytometry [41]. The proliferative potential of cells expressing nucleases can be quantified by a competition assay, in which cells transfected with a nuclease have to compete with cells that are not transfected with the nuclease in the same culture [44, 45]. To directly measure the effect of nuclease expression on the cell cycle, the FUCCI system can quantify by flow cytometry whether nuclease expression perturbs the cell cycle parameters of a population, particularly by inducing arrest in G2 [41]. Finally, nuclease induction of extra DSBs, the major cause of cellular toxicity, can be directly measured using flow cytometry (i.e. mean fluorescent intensity) or fluorescent microscopy (i.e. counting the number of foci) by staining for γ H2AX or 53BP1 foci [45–47].

Gross chromosomal rearrangements

Perhaps the most concerning type of off-target effects are gross chromosomal rearrangements, as they are most clearly associated with malignant transformation. For a rearrangement of known type and location, such as the deletion or inversion of the sequence between two nuclease cut sites on the same chromosome, PCR primers can be designed to test for that rearrangement qualitatively or quantitatively [41]. A more unbiased assay for detecting novel chromosomal rearrangements has not yet been developed. An assay that could be applied to bulk cell populations would be of great use to the field, as whole genome sequencing currently lacks the sensitivity to detect rare events in large populations of cells.

Box 3**Outstanding questions**

- How can we quantify on-target NHEJ and HR genome editing outcomes using DNA sequencing in a high throughput cost-effective way at a large scale?
- How can user-friendly, accurate, and accessible software be developed and standardized to allow non-specialists to process deep sequencing data for measurement of on-target and off-target editing?
- How can we quantify gross chromosomal rearrangements following engineered nuclease treatment of bulk cell populations?
- How can *in cellula* off-target sites be predicted in a sensitive, unbiased, and accurate manner?
- How can we predict whether off-target effects of a given nuclease will lead to adverse clinical outcomes (i.e. transformation of cells to cancerous phenotype)?

Highlights

- Genome editing will transform biomedical research and patient treatment.
- Assays for on and off-target events can be shared across multiple nuclease platforms.
- Sequencing based methods provide the richest data to detect on-target genome editing.
- Understanding off-target genome editing requires sequencing and functional assays.

Figure 1. Assays for quantifying on-target genome editing outcomes

(A) Schematic of mutation detection assay using CEL-I/T7E1 enzyme to measure NHEJ.

(B) Schematic of RFLP assay to measure HR. **(C)** Schematic of traffic light reporter assay to measure both NHEJ and HR. **(D)** Schematic of SMRT sequencing to measure both NHEJ

and HR at endogenous loci. For more details, refer to the main text.

Locating Nuclease Off-target Activity

Figure 2. Locating nuclease off-target activity
Off-target sites are predicted through either *in vitro*, cellular, or *in silico* techniques, which generate a list of potential off-target sites. Genomic DNA is harvested from cells treated with nucleases, the sites are amplified, tested and a subset validated as *bona fide* off-target sites.

Table 1

Assays to characterize nuclease on-target activity.

	Mutation detection assay with CEL-I/T7EI	RFLP Gel Assay	Fluorescent Gene Addition	Traffic Light Reporter	Single Cell Clone Analysis	NextGen Sequencing	SMRT Sequencing
Measures HDR	No	Yes	Yes	Yes	Yes	Yes	Yes
Highly Sensitive (<2%)	No	No	Yes	Yes	Possible	Yes	Yes
Measures NHEJ	Yes	No	No	Yes	Yes	Yes	Yes
Uses Clinically Applicable Donor DNA	Not relevant	Yes	No	No	Yes	Yes	Yes
Measures Endogenous Locus	Yes	Yes	Yes	No	Yes	Yes	Yes
High-Throughput	Moderate	Moderate	Yes	Yes	No	Yes	Yes
Allows Long Homology Arms	Not relevant	Yes	Yes	Yes	Yes	No	Yes

In defining “clinically applicable” donor DNA, we refer to a donor molecule that is free of elements unapproved for human application (such as fluorescent proteins) and is homologous to the therapeutic gene target (not an artificial reporter locus).

Table 2

Assays to characterize nuclease off-target activity.

Assay Category	Assay Name	* Specialized laboratory technique(s) required?	Bioinformatics required?
Bulk Off-Target Activity	Viability Staining	No	No
	GFP Loss	No	No
	Cell Cycle Disruption	No	No
Phenotype Transformation	Soft Agar	No	No
	Clonal Barcoding	Yes	Yes
Off-Target Site Prediction	<i>In silico</i> prediction tools	No	Yes
	SELEX	Yes	Yes
	In vitro cleavage	Yes	Yes
	IDLV Trapping	Yes	Yes
	ChIP-Seq	Yes	Yes
Analysis of Clonal Populations	aCGH	Yes	Yes
	Whole Genome/Exome Sequencing	Yes	Yes

* In defining laboratory techniques, we assume that researchers working with nucleases are familiar with basic cell transfection, detecting expression of fluorescent proteins, and cell culture protocols, but not more specialized procedures such as viral production or immunoprecipitation.