Cell Reports
Resource

Quantifying Genome-Editing Outcomes at Endogenous Loci with SMRT Sequencing

Ayal Hendel,[1,4] Eric J. Kildebeck,[1,4] Eli J. Fine,[2,4] Joseph T. Clark,[1] Niraj Punjya,[1] Vittorio Sebastiano,[3] Gang Bao,[2] and Matthew H. Porteus[1,*]
[1]Department of Pediatrics, Stanford University, Stanford, CA 94305, USA
[2]Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA 30332, USA
[3]Department of Obstetrics and Gynecology, Stanford University, Stanford, CA 94305, USA
[4]These authors contributed equally to this work
*Correspondence: mporteus@stanford.edu
http://dx.doi.org/10.1016/j.celrep.2014.02.040

SUMMARY

Targeted genome editing with engineered nucleases has transformed the ability to introduce precise sequence modifications at almost any site within the genome. A major obstacle to probing the efficiency and consequences of genome editing is that no existing method enables the frequency of different editing events to be simultaneously measured across a cell population at any endogenous genomic locus. We have developed a method for quantifying individual genome-editing outcomes at any site of interest with single-molecule real-time (SMRT) DNA sequencing. We show that this approach can be applied at various loci using multiple engineered nuclease platforms, including transcription-activator-like effector nucleases (TALENs), RNA-guided endonucleases (CRISPR/Cas9), and zinc finger nucleases (ZFNs), and in different cell lines to identify conditions and strategies in which the desired engineering outcome has occurred. This approach offers a technique for studying double-strand break repair, facilitates the evaluation of gene-editing technologies, and permits sensitive quantification of editing outcomes in almost every experimental system used.

INTRODUCTION

Genome editing with engineered nucleases is a transformative technology for efficiently modifying essentially any genomic sequence of interest (McMahon et al., 2012). This technology utilizes engineered nucleases to generate site-specific double-strand breaks (DSBs) at desired genomic locations followed by resolution of DSBs using the endogenous cellular repair mechanisms of nonhomologous end-joining (NHEJ) and homology directed repair (HDR) (Porteus and Carroll, 2005). A variety of desired genetic modifications can be achieved with this approach, including mutation of a specific site through mutagenic NHEJ and precise change of a genomic sequence to a new sequence through HDR. There are currently four principal families of engineered nucleases used for gene editing: Zinc Finger Nucleases (ZFNs) (Porteus and Carroll, 2005), Transcription Activator-Like Effector Nucleases (TALENs) (Bogdanove and Voytas, 2011), Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR/Cas9) or RNA-guided endonucleases (hereafter called "RGENs") (Gaj et al., 2013; Mali et al., 2013), and engineered meganucleases (Silva et al., 2011). The rapid development of these technologies is allowing for the precise alteration of genomes for numerous applications, including plant engineering (Li et al., 2012), generation of cell lines for basic science (Soldner et al., 2011), human gene therapy (Urnov et al., 2005), and industrial applications (Fan et al., 2012).

When a new set of gene-editing reagents is developed for a custom application, the activity levels of nucleases and the frequency of the desired gene-editing event at the target locus must be determined and often need to be optimized for the specific cell type and system being used. This need has previously been met by a variety of methods including gel-based assays to measure mutagenic NHEJ (Guschin et al., 2010), gene addition of fluorescent reporters to measure HDR (Porteus and Baltimore, 2003; Stark et al., 2004), analysis of large numbers of single-cell clones, and the use of optimization assays to measure NHEJ and HDR at engineered reporter loci (Certo et al., 2011). Although each of these assays have their utility, each have important limitations including a lack of sensitivity required for difficult applications (gel-based assays), the use of an indirect rather than a direct measure of genome editing (targeted gene addition of fluorescent reporters), and the need to generate reporter cell lines (Traffic Light Reporter system). The Traffic Light Reporter (TLR) system (Certo et al., 2011) is the only one of these assays that allows simultaneous measurement of NHEJ and HDR by expressing GFP in cells that undergo HDR-mediated correction of a GFP gene and expressing mCherry in cells with NHEJ-induced frameshift mutations. Although the TLR is a very sensitive assay for measuring DSB repair (DSBR) pathway choice, the need to generate a fluorescent reporter locus precludes measurement at endogenous target loci and thus far has prevented the use of the TLR in human primary cells. High-throughput sequencing of endogenous loci overcomes these limitations, but the range of outcomes that can be measured is limited by sequencing read lengths. Illumina (Yang et al.,

2013b) and 454 (Qi et al., 2013) sequencing have recently been used to measure HDR and NHEJ outcomes when single-stranded oligodeoxynucleotides (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 that drive more efficient HDR and provide the flexibility to target long gene cassettes.

Here, we present a method for measuring genome-editing outcomes at endogenous loci using single-molecule real-time (SMRT) DNA sequencing, which provides read lengths approaching 15 kb and is an affordable approach that can be widely used. This technique allows for analysis of gene-editing frequencies when donor templates with long arms of homology are used, which is a common strategy to increase HDR efficiency in primary cells and for the addition of large gene inserts. Using this method, we were able to measure simultaneous frequencies of NHEJ and HDR in primary cells and cell lines with greatly improved detection sensitivity. We describe the use of SMRT sequencing analysis to measure genome-editing outcomes and rare large insertions generated by TALENs, ZFNs, and RGENs at the endogenous *IL2RG*, *HBB*, and *CCR5* loci. In addition, we use this system to quantify the effect of varying different parameters on the frequency of different gene-editing outcomes.

RESULTS

Measurement of Gene-Editing Outcomes at the Endogenous *IL2RG* Locus

To develop a method for quantitatively and rapidly measuring the different gene alterations occurring at an endogenous locus of interest, we used a highly active TALEN pair to stimulate DSBs at the endogenous IL-2 receptor common γ-chain gene (*IL2RG*), mutations in which are responsible for the congenital primary immunodeficiency SCID-X1 (Kildebeck et al., 2012; Shaw and Kohn, 2011). For the introduction of precise sequence alterations at this locus, we designed a donor template with approximately 400 bp arms of homology 5′ and 3′ of the TALEN cut site (Figure 1A). Within the 3′ arm of homology, we introduced seven point mutations that, upon successful HDR, are stably integrated into the *IL2RG* gene and prevent binding and cleavage by the TALEN pair (Figures 1A and 1B). To measure the frequency of mutagenic NHEJ and HDR with this system, we developed a strategy based on single-molecule real-time (SMRT) DNA sequencing, a high-throughput sequencing technology capable of analyzing long DNA fragments. First, the *IL2RG* locus was amplified using a forward primer that is 5′ and outside the start of the 5′ homology arm and a reverse primer that is downstream of the TALEN pair target site (Figure 1C). With this approach, nonintegrated and randomly integrated donor templates are not amplified, removing common sources of background noise. The SMRT DNA sequencing technology allows for the determination of DNA sequence from individual DNA templates (Eid et al., 2009; Roberts et al., 2013). Single-molecule read lengths approaching 15 kb were reached in this study, with an average read length approaching 3 kb. For DNA fragments shorter than the read limit of the polymerase, improved sequence accuracy (frequently reaching an average Phred QV score of 40, denoting 99.99% accuracy) is achieved by iteratively sequencing the same circular DNA template (Travers et al., 2010) (Figure 1C).

To induce sequence alterations in *IL2RG*, we expressed the *IL2RG* TALENs from plasmid DNA with or without introduction of donor DNA and then analyzed cell populations by SMRT DNA sequencing. Following transfection with TALENs alone, we detected unmodified alleles and alleles with deletions or insertions indicative of mutagenic NHEJ (Figure 1D). When cells were transfected with both the TALENs and donor DNA, we detected unmodified alleles, alleles with deletions or insertions, and alleles with the seven point mutations precisely integrated into *IL2RG* by HDR. Notably, no alleles were detected with both the seven point mutations and indels indicative of NHEJ, validating the ability of the point mutations to prevent TALEN cleavage of HDR-modified alleles. High frequencies of "on-target" *IL2RG* modification were observed in K562 cells under these conditions, with 18% of alleles mutated by NHEJ and 17% of alleles precisely modified by HDR (Figure 1D). Because the PCR strategy being used, where one primer is outside of the donor template arm of homology, essentially no background signal was detected from amplification of nonintegrated or randomly integrated donor template. Control experiments (either mock or donor only transfections) had low background frequencies of NHEJ and HDR reads resulting from PCR or DNA sequencing errors, which ranged from 0.00% to 0.03% for individual samples with average frequencies of 0.007% NHEJ and 0.001% HDR (Tables S1–S6). This low level of background noise, coupled with the high-throughput nature of this approach, produces a high level of sensitivity and creates possibilities for studying rare DNA repair events.

Reliability of SMRT Sequencing Analysis at a Single Endogenous Locus

To validate the accuracy of the SMRT DNA sequencing strategy, we compared our high-throughput results with standard gel-based assays and single-cell clone analysis of K562 cells. First, we used a restriction fragment length polymorphism (RFLP) assay to measure the frequency of HDR by measuring the presence of an AflII restriction site that is created when the seven point mutations within the donor template are precisely incorporated into the target locus (Figure 2A). Using the RFLP assay, the AflII restriction site was detected in an average of 14.3% of alleles normalized for transfection efficiency compared to 16.8% of alleles by SMRT sequencing analysis of the same populations. The most commonly used methods for determining the frequency of NHEJ measure any small deletion or insertion events, which is confounded by sequence alterations introduced by HDR. To independently determine the true frequency of alleles modified by NHEJ and HDR, we grew single-cell clones from a representative sample. Analysis of these clones showed that 11.3% of alleles had undergone mutagenic NHEJ and 11.1% of alleles had been precisely modified by HDR, compared to frequencies of 11.2% and 11.0%, respectively, as measured by SMRT sequencing analysis of the same population (Figure 2B); SMRT frequencies represent the total population frequency, not normalized to transfection efficiency, in order to directly compare to the clonal analysis. To confirm the reproducibility of SMRT sequencing analysis, we

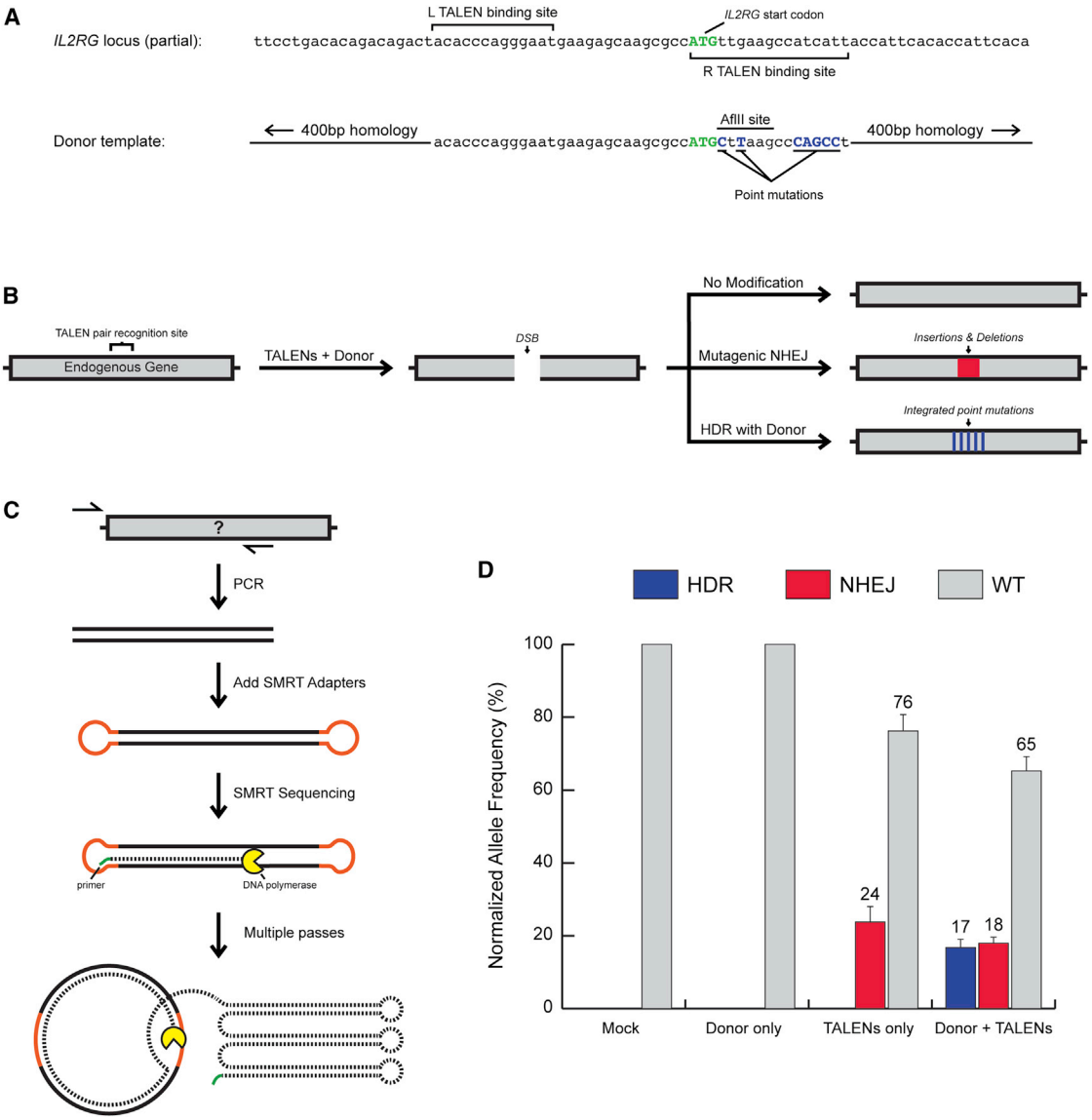

Figure 1. Measuring Gene Editing at an Endogenous Locus with SMRT Sequencing

(A) Sequence of the TALEN target site at the *IL2RG* locus and the *IL2RG* donor template. The donor template harbors seven point mutations that, when integrated into *IL2RG*, create silent mutations and an AflII restriction site. These substitutions alter the right TALEN binding site and provide a signature for alleles precisely modified by HDR.

(B) Diagram of gene editing at an endogenous locus. TALENs create a double-strand break (DSB), which can lead to no modification, insertion or deletion mutations, or integration of point mutations from the donor template.

(C) Schematic of SMRT DNA sequencing analysis. The endogenous locus is amplified by PCR, with at least one primer outside the arms of homology of the donor template, and SMRT adapters are added to PCR amplicons. Individual DNA molecules are sequenced by SMRT sequencing, with read lengths averaging ~3 kb in length and approaching ~15 kb.

(D) Measurement of gene-editing outcomes at the *IL2RG* locus in K562 cells. Modification frequencies are normalized to transfection efficiency.

Data for graph are from Table S1. Bars represent three independent biological replicates; error bars, SD.

analyzed a single targeted population eight times and found SDs for NHEJ and HDR of 0.66% and 0.79%, respectively (Figure 2C). This experimental variance between samples was only slightly higher than the expected statistical variance for the number of sequences analyzed, demonstrating the reliability of this approach (Figure 2D).

Quantification of Gene Editing at the *IL2RG* Locus in Primary Cells

For gene-editing applications, moving from known conditions in commonly used cell lines to more difficult experimental platforms, such as induced pluripotent stem cells or primary cells, poses a significant challenge. Using the gene-editing tools

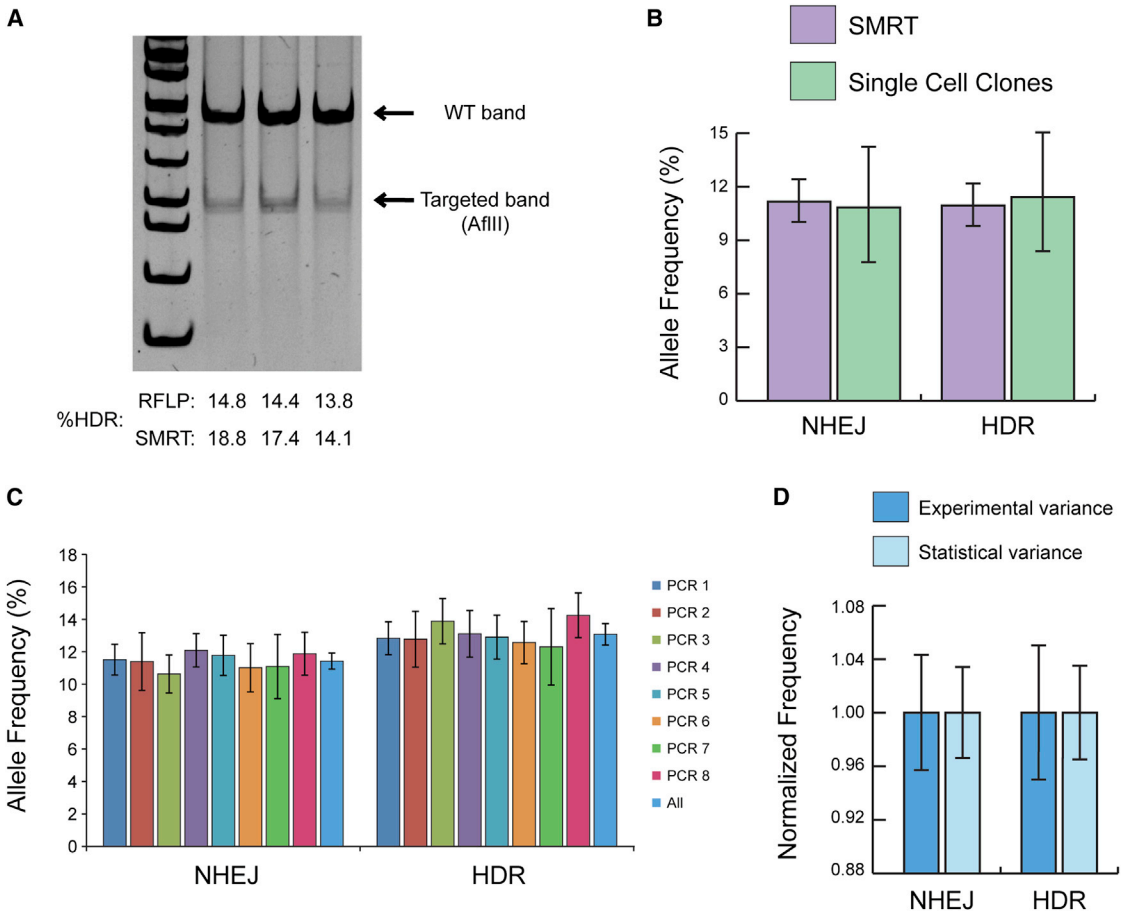

Figure 2. Reliability of SMRT Sequencing Analysis for Measuring Gene-Editing Outcomes at an Endogenous Locus

(A) RFLP analysis of K562 cells targeted with 1 μg of each TALEN and 5 μg donor in triplicate. The frequency of HDR in each sample as measured by RFLP and SMRT sequencing analysis is shown.

(B) Quantification of NHEJ and HDR frequencies in single-cell clones grown from a representative population of K562 cells. Error bars represent 90% confidence intervals.

(C) A representative sample of targeted K562 cells was analyzed by SMRT sequencing eight separate times to determine the variability introduced by PCR, SMRT library synthesis, and sequencing. Error bars represent 90% confidence intervals.

(D) Quantification of the observed experimental variation compared to the expected statistical variation for the number of sequences analyzed for the eight replicates. Error bars for experimental variation represent SD. Error bars for statistical variation represent 68% confidence intervals (corresponding to the fraction of the normal distribution covered by ±1 SD).

previously described, we next tested our ability to measure gene-editing events in $CD34^+$ hematopoietic stem/progenitor cells (HSPCs) and human embryonic stem cells (hESCs), both of which are difficult to target, but important cell types for basic research and gene therapy. After introducing TALENs and donor DNA into $CD34^+$ HSPCs, SMRT sequencing analysis showed frequencies of mutagenic NHEJ and HDR of 7% and 1%, respectively, at the endogenous *IL2RG* locus (Figure 3A). In hESCs, which commonly require enrichment of targeted clones due to low gene-editing efficiencies, addition of our gene-editing reagents resulted in mutagenic NHEJ and HDR frequencies of 0.10% and 0.14%, respectively (Figure 3B). Control hESC samples transfected with only donor DNA showed background frequencies of 0.02% NHEJ and no HDR, illustrating the very low level of background noise for this technique. Our transfection efficiency for these hESC populations was approximately 20%, suggesting that with enrichment for transfected cells we would generate modification frequencies of 0.5%–0.7%. These numbers are consistent with the numbers published by Soldner et al. (2011), who showed that after sorting for highly transfected hESCs targeting frequencies of 0.4%–0.8% were obtained. Importantly, because *IL2RG* is silent in both of these cell types, these results demonstrate the ability of this approach to provide quantitative and sensitive measures of gene editing and DNA damage repair at a silent endogenous locus in primary cells.

Analysis of Gene Editing with TALENs, RGENs, and ZFNs at the Endogenous *IL2RG*, *HBB*, and *CCR5* Loci

The extraordinary expansion of gene-editing technologies over recent years has created a plethora of opportunities for

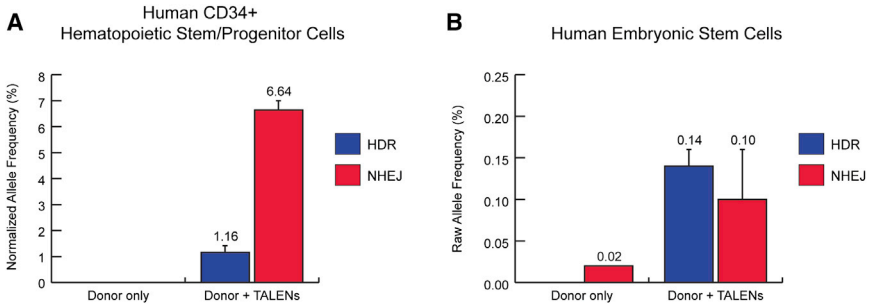

Figure 3. Measurement of Genome Editing at an Endogenous Locus in Human Primary Cells

(A) Measurement of gene-editing outcomes at *IL2RG* in CD34$^+$ HSPCs using the high-expression TALEN plasmids (Supplemental Experimental Procedures). Data for graph are from Table S2. Bars represent three independent biological replicates; error bars, SD.

(B) Measurement of gene-editing outcomes at *IL2RG* in hESCs using the high-expression TALEN plasmids (Supplemental Experimental Procedures). Data for graph are from Table S3. Bars represent three independent biological replicates; error bars, SD.

researchers attempting to modify genomes. With the introduction of TALENs and RGENs, it is now possible to generate tens to hundreds of candidate nucleases in a matter of weeks, or even days, and target multiple genomic sites simultaneously (Briggs et al., 2012; Cong et al., 2013; Kim et al., 2013; Reyon et al., 2012; Yang et al., 2013a). Via simultaneous analysis of different genomic sites and conditions in a single SMRT sequencing run, this approach has the potential to rapidly expedite the process of characterizing nuclease activities and optimizing targeting parameters. To determine the ability of this method to measure the activities of different classes of nucleases at multiple genomic sites, we treated cells with TALENs, RGENs, and ZFNs designed to target the *IL2RG*, *HBB*, and *CCR5* genes and analyzed gene modification.

To test the relative activity of TALENs and RGENs at the *IL2RG* locus, we first constructed an RGEN with a target site overlapping the target site of the *IL2RG* TALENs (Figure 4A). Expression of the RGEN in K562 cells generated targeted mutations in 37% of *IL2RG* alleles compared to only 13% of alleles using TALENs. Introduction of donor template DNA with either the RGEN or TALENs produced alleles modified by mutagenic NHEJ and alleles precisely modified by HDR. As expected, the more active RGEN stimulated a higher level of HDR than the TALENs with 33% and 22% of alleles harboring the integrated SNPs, respectively. Moving to a different genomic locus, we used a TALEN pair and an RGEN targeting the *HBB* gene, mutations in which are responsible for sickle cell anemia and thalassemia (Gallagher, 2013) (Figure 4B). At *HBB*, the RGEN again produced significantly higher frequencies of gene disruption than the TALENs and stimulated higher frequencies of HDR upon the introduction of donor template. When expressed with donor template, the *HBB* RGEN mutated 41% of *HBB* alleles, whereas the *IL2RG* RGEN mutated 21% of *IL2RG* alleles, suggesting that more DSBs were being created at the *HBB* locus. Despite this increase in mutagenesis, the simultaneous level of HDR was only 14% at *HBB* compared to 33% at *IL2RG*. Thus, total modification levels at *HBB* and *IL2RG* were highly similar, 55% and 54%, respectively, but the ratio of HDR to NHEJ was markedly lower at *HBB* (0.34:1) than at *IL2RG* (1.6:1) (Figures 4A and 4B). This large difference in the efficiency of precise gene targeting suggests that there could be intrinsic differences between these loci affecting their ability to participate in plasmid-mediated gene targeting by HDR.

To further confirm the utility of SMRT sequencing analysis to measure targeted genomic modifications for multiple classes of nucleases, we compared the activity of previously reported ZFNs and TALENs designed to target the *CCR5* locus (Mussolino et al., 2011; Perez et al., 2008). As seen at *IL2RG* and *HBB*, addition of nucleases led to targeted disruption of the endogenous gene by NHEJ and the further addition of donor template DNA stimulated targeted integration through HDR (Figure 4C). The *CCR5*-specific ZFNs, variants of which are currently being used in clinical trials for HIV (Perez et al., 2008), produced higher levels of modification at the endogenous *CCR5* locus than TALENs designed to an overlapping target site. It is important to note that in this study stable modifications in K562 cells were measured for TALENs, RGENs, and ZFNs 14 days posttransfection. The absolute genome-editing frequencies reported here are thus somewhat different than published results for previously described nucleases where nuclease activity was measured at different time points, in different cell types, and with different nuclease levels (Mussolino et al., 2011; Perez et al., 2008; Voit et al., 2013). Nuclease-induced NHEJ is typically measured with high nuclease levels 3 days posttransfection to detect maximal NHEJ levels, but these modifications decrease over time due to toxicity (Doyon et al., 2010, 2011; Kim et al., 2009; Lombardo et al., 2007) (Figure S1). Instead of measuring NHEJ and HDR separately or with different transfection conditions, SMRT DNA sequencing provides a simple alternative for comparing stable NHEJ and HDR frequencies simultaneously.

Optimization of Gene Targeting Parameters

In addition to comparing different nuclease platforms, we have also used the SMRT DNA sequencing approach to study different variables that might affect genome-editing outcomes. To explore how varying the dose of TALENs affects gene-editing frequencies, we measured frequencies of NHEJ and HDR at *IL2RG* while progressively decreasing the amount of TALENs transfected in K562 cells. Keeping the amount of donor DNA constant and titrating down the amount of TALENs by 100-fold, we saw a progressive decrease in both mutagenic NHEJ and HDR events, whereas their relative frequencies remained largely unchanged (Figure 5A). Using this approach, we were

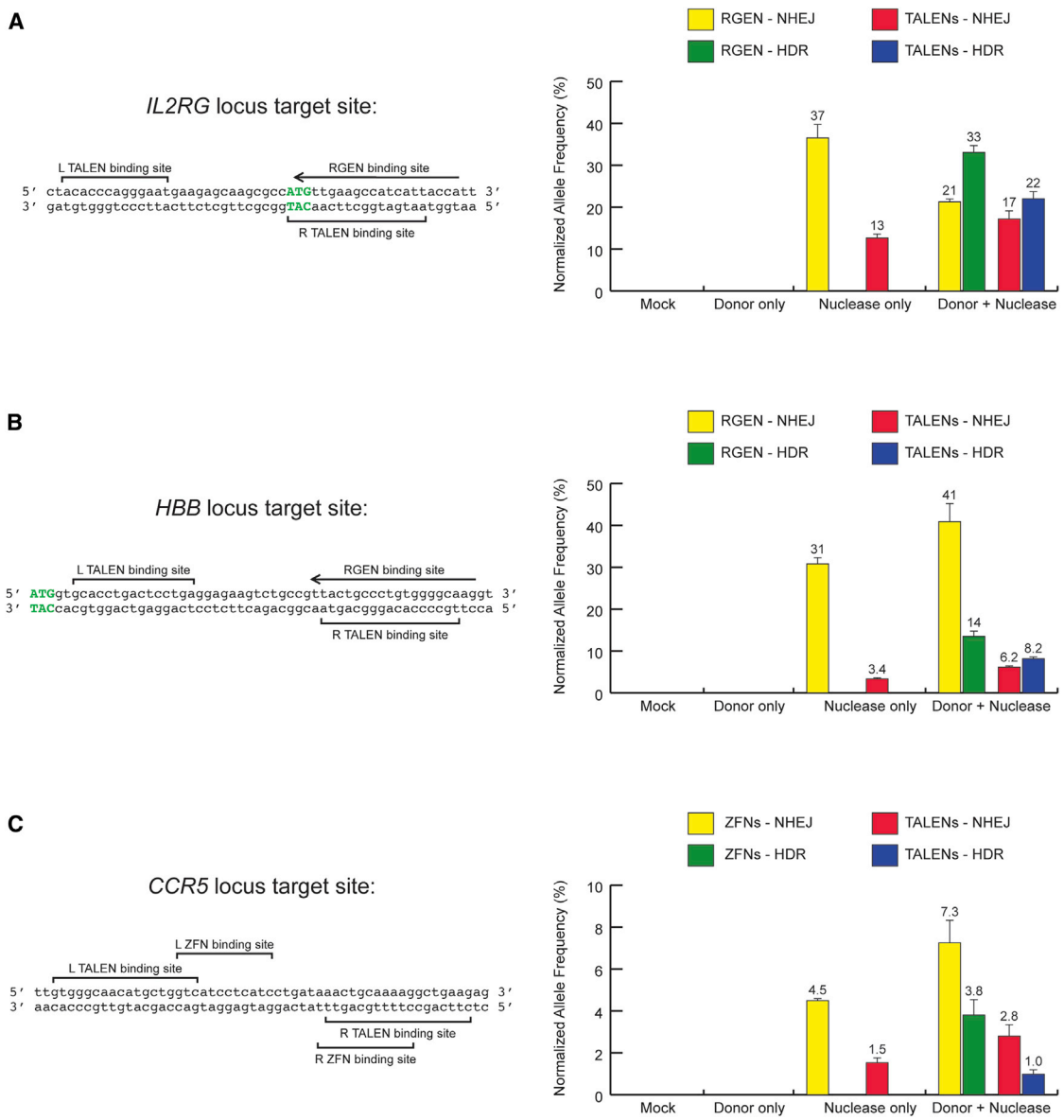

Figure 4. Measuring Gene Editing with Different Engineered Nuclease Platforms at Different Genomic Targets

(A) Left: *IL2RG* target site for TALENs and RGEN guide sequence. The *IL2RG* start codon is shown in green. Right: modification of the *IL2RG* locus in K562 cells. Data for graph are from Table S4.

(B) Left: *HBB* target site for TALENs and RGEN guide sequence. The *HBB* start codon is shown in green. Right: modification of the *HBB* locus in K562 cells. Data for graph are from Table S5.

(C) Left: *CCR5* target site for TALENs and ZFNs in exon 3. Right: modification of the *CCR5* locus in K562 cells. Data for graph are from Table S6. For Cell Assay results, see also Figure S1. Bars represent three independent biological replicates; error bars, SD.

able to reliably detect gene-editing outcomes at frequencies ranging from >20% at high TALEN levels to ≤0.1% at very low TALEN levels. Even at modification frequencies of 0.1%–0.4%, relative activity levels were easily distinguished with this approach (Table S7). Next, to test conditions for maximizing the frequency of HDR at *IL2RG*, we investigated the effect of the amount of donor template DNA on gene modification. Keeping the amount of TALENs constant and titrating the amount of donor DNA, we saw the overall level of modification at *IL2RG* remain relatively constant, whereas the total level of HDR rose from 1.6% to 17.8% at optimal levels (Figure 5B; Table S8). With this rise in the contribution of HDR, the ratio of HDR to NHEJ increased from 0.12 to 1.37 with increasing abundance of donor DNA (Figure 5C).

Another important variable for efficient HDR-mediated DNA repair is the length of the homologous regions in donor DNA, which has been shown to vary between species and in different cell types of the same species (Beumer et al., 2013; Orlando

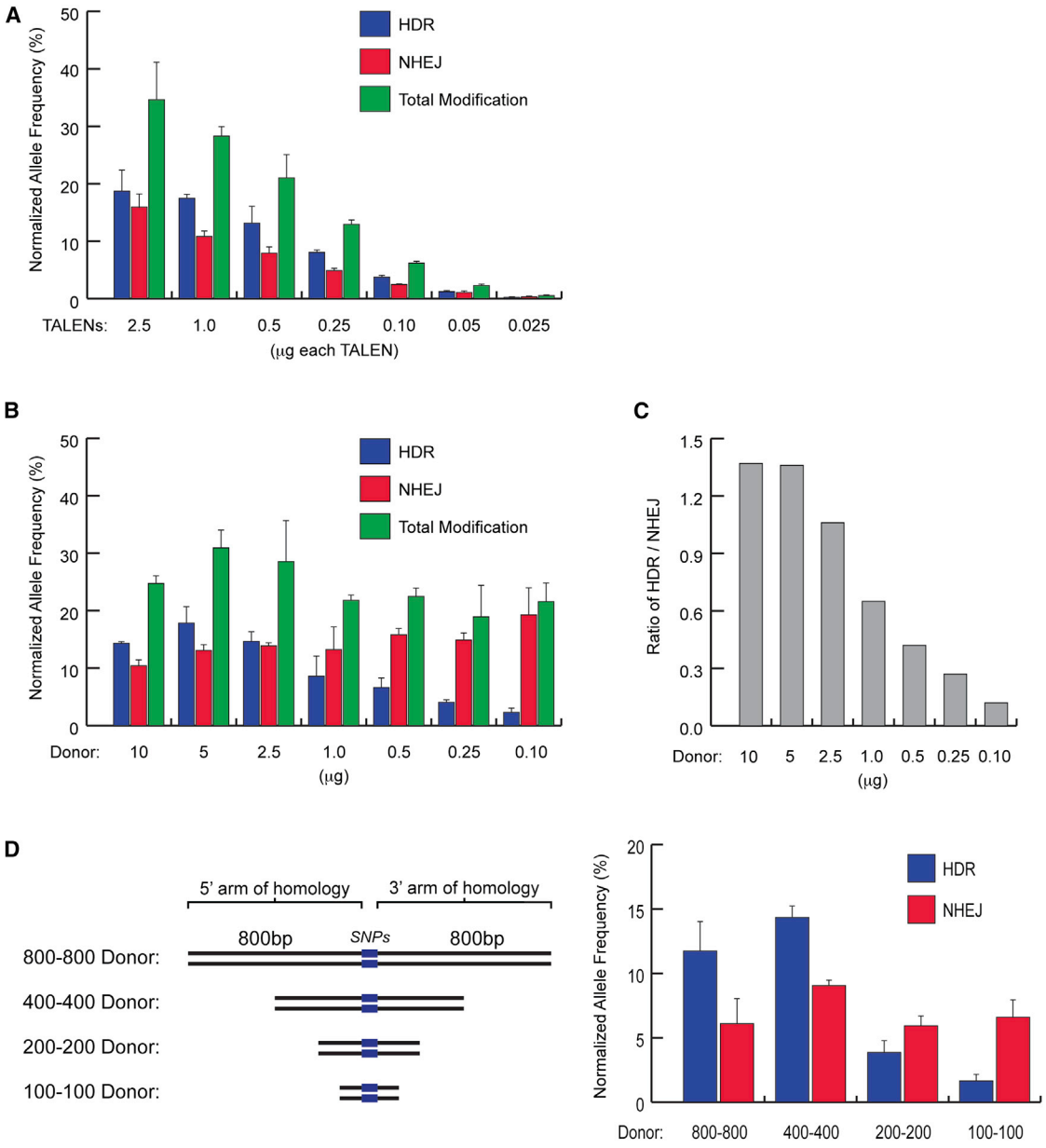

Figure 5. Optimization of Gene-Editing Parameters at *IL2RG* with SMRT Sequencing

(A) Titration of TALEN amount in K562 cells with amount of donor DNA held constant at 5 μg. Data for graph are from Table S7.

(B) Titration of donor DNA amount in K562 cells with TALEN DNA amount held constant at 1 μg each TALEN. Data for graph are from Table S8.

(C) Ratio of HDR to NHEJ for samples in (B).

(D) Left: schematic of donor templates with varying arm of homology lengths. Right: quantification of effect of homology arm length on gene-editing frequencies in K562 cells. Data for graph are from Table S9. Bars represent three independent biological replicates; error bars, SD.

et al., 2010; Waldman, 2008). To determine the effect of homology arm length on HDR efficiency with plasmid donors, we constructed a series of donor templates with a range of homology arm lengths from 800 to 100 bp (Figure 5D; Table S9). At the *IL2RG* locus in K562 cells, homology arms 100 or 200 bp in length were found to be significantly less effective for HDR than plasmid donor templates with 400 or 800 bp homology arms. As would be predicted, the homology arm length did not change the frequency of mutagenic NHEJ. In this cell type, 400 bp arms of homology actually resulted in the same levels of HDR as more commonly used 800 bp arms, suggesting that for gene targeting in some human cell types maximal levels of HDR may be achieved with relatively short 400 bp arms of homology. For all areas of optimization, however, the specific setting of the cell type and the chromosomal locus under investigation should be taken into consideration.

SMRT Sequencing of Genome-Editing Outcomes Reveals Genomic and Plasmid DNA Sequences Captured into Targeted Sites

A unique feature of the SMRT sequencing method is the combination of high-throughput with long sequence read lengths. This combination allowed us to see rare mutations including large insertions and deletions hundreds of base pairs (bp) in length. Analysis of mutations at the *IL2RG*, *HBB*, and *CCR5* loci showed a wide range of insertions and deletions ranging from +334 to −412 bp. Interestingly, when we BLASTed the inserted sequences that were >30 bp against NCBI nonredundant nucleotide databases, we were able to identify sequences originating from the same chromosome as the target site, nonhomologous chromosomes, plasmid DNA, and the *E. coli* genome. Representative samples of such insertion events can be seen in Figure 6. At each end of an insertion event, there is a junction with the chromosomal segment on one side and the inserted segment on the other side. Processing of the chromosomal sides of the junctions can be tracked by examining the sequences at these sites. NHEJ-mediated DSBR commonly involves deletion of a few nucleotides. Surprisingly, of the 42 long insertions we found only seven that involved deletions of nucleotides. The remaining 35 did not involve even single-nucleotide deletions. This finding suggests that the insertion event may have a role in protecting chromosomes from harmful deletions during DSBR. Microhomology is known to drive some small insertion and deletion events during NHEJ. Analysis of the sequences flanking the external sources of these long insertion events did not reveal any trends, suggesting that these events were not driven by flanking microhomologous sequences (Supplemental Results).

DISCUSSION

Although the simplicity and flexibility of engineering TALENs and RGENs has transformed gene editing over recent years, many questions remain about DSBR processes and gene targeting in different cell types and at different genomic sites. Here, we present a rapid, accurate, and sensitive strategy for analyzing gene-editing outcomes and DSBR pathway choice at endogenous loci in potentially any cell type using any type of engineered nuclease. The SMRT DNA sequencing strategy offers three principal advantages over currently available techniques: (1) sensitive measurement of genome editing in any cell type, including primary stem cells, without the need to make a stable reporter cell line, (2) measurement of modifications at endogenous loci regardless of transcriptional status, and (3) 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.

Without generating reporter cell lines, we used the SMRT DNA sequencing strategy to measure gene-editing outcomes in $CD34^+$ HSPCs, hESCs, and K562 cells. Measurement at the *IL2RG* locus was not inhibited by the lack of transcription of this gene in $CD34^+$ HSPCs and hESCs, demonstrating the ability of this method to provide quantitative and sensitive analysis of silent endogenous loci. Epigenetic status is known to affect all DNA-metabolism processes including transcription, replication, and repair (Cavalli and Misteli, 2013; Papamichos-Chronakis and Peterson, 2013). The importance of transcriptional activation and epigenetic status for gene-editing efficiency is still largely unknown, but epigenetic modification was recently shown to impact DSBR pathway choice (Daboussi et al., 2012; Kuhar et al., 2014; Valton et al., 2012a, 2012b; van Rensburg et al., 2013). The SMRT DNA sequencing strategy could be used to further study how chromatin status influences DSBR pathway choice and gene-editing efficiency by providing analysis in a broader range of cell types in which the chromatin state of the targeted site is known. One other potentially important variable for gene-editing efficiency, particularly when working between different cell types, is the method of delivery, and this strategy could be used to quantitatively measure the impact of using different delivery methods including electroporation-based techniques, viral-based strategies, and lipid or nanoparticle-based methods.

Genome editing with engineered nucleases can be used to create many types of changes to a genome, and any site within an organism's genome is now a potential target. The versatility of this approach, combined with the ease of synthesizing new nucleases, creates a need for a method to evaluate different types of nucleases at different genomic locations. In this study, we used SMRT DNA sequencing analysis to measure genome editing at sites within the *IL2RG*, *HBB*, and *CCR5* genes using the three most widely used classes of engineered nucleases. For the specific nucleases we investigated in K562 cells, we found that at both the *IL2RG* and *HBB* loci the RGEN generated significantly higher frequencies of mutagenic NHEJ than TALENs designed to overlapping sites. When transfected alone, the *IL2RG*-specific and *HBB*-specific RGENs created very similar levels of mutagenic NHEJ, suggesting that a similar amount of DSBs are being created at the two loci. Despite this, the *HBB*-specific RGEN stimulated a significantly lower frequency of HDR than that seen at *IL2RG*. Whether this difference in repair pathway utilization is the result of different chromatin status or the sequence composition of the target sites and corresponding donor DNA is still unclear, but this technique could be applied to further elucidate how spatial parameters affect DNA repair. Additionally, moving between genomic loci we have encountered single nucleotide polymorphisms (SNPs) that confound measurement of NHEJ using standard mismatch detection assays (Guschin et al., 2010), including one at *HBB* in K562 cells. The ability to use SMRT DNA sequencing to quantify mutation frequencies even in the presence of SNPs is another advantage of this system.

In addition to the gamut of nucleases and target sites, molecular and genetic strategies to influence DSBR pathway choice can play a significant role in achieving a desired outcome or minimizing unwanted outcomes. By titrating the amount of donor template in K562 cells, we were able to optimize conditions for generating HDR events and alter the ratio of HDR to NHEJ significantly. Furthermore, the long read lengths of SMRT DNA sequencing allowed us to measure gene-editing outcomes using donor templates with 800 bp arms of homology. These data demonstrate the advantage of using long arms of homology to stimulate higher frequencies of HDR with plasmid donors, and this technique could further be used to directly compare

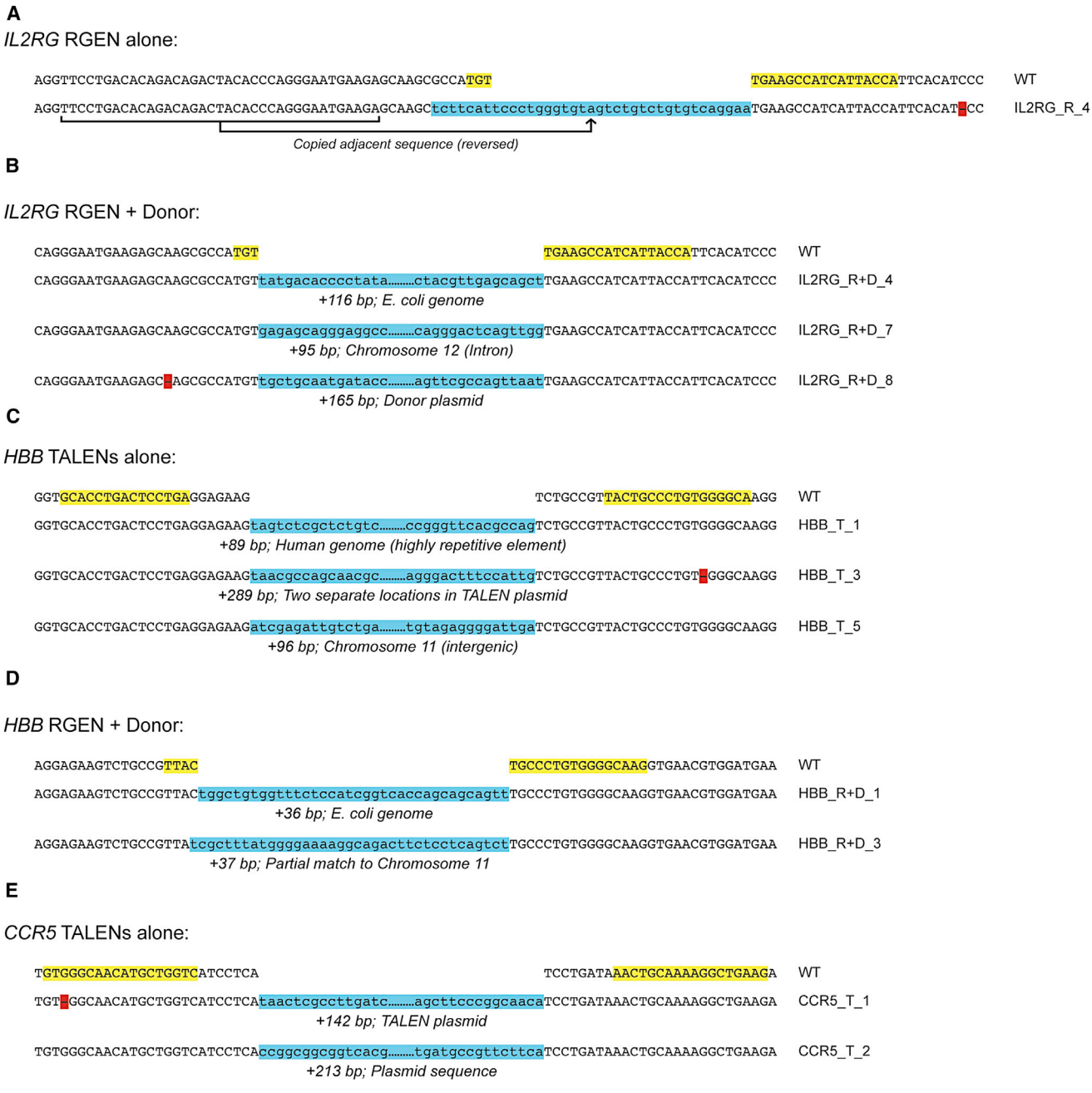

Figure 6. DNA Repair by Insertion of Large Sequences from Various Sources

(A) An event from the *IL2RG* RGEN-only transfections where the insert is an exact repeat—in the inverse orientation—of a sequence near the cleavage site.

(B) Reads from the *IL2RG* RGEN and donor transfections containing inserts derived from the *E. coli* genome, an intronic sequence in chromosome 12, and a donor plasmid.

(C) Several representative SMRT reads from the *HBB* TALENs-only transfections were recovered containing inserts derived from a repetitive genomic element, one of the TALEN plasmids, and a region in chromosome 11 on the opposite arm from the *HBB* gene.

(D) Reads from the *HBB* RGEN and donor transfections containing inserts derived from the *E. coli* genome and a region in chromosome 11 on the opposite arm from the *HBB* gene.

(E) Reads from the *CCR5* TALENs-only transfections containing inserts derived from a TALEN plasmid and an unknown plasmid.

Wild-type (WT) sequences are shown with nuclease binding sites highlighted in yellow. Inserted sequences in the SMRT reads are lowercase and highlighted in blue with the size and source listed below. Deleted bases, likely resulting from sequencing errors, are represented by hyphens highlighted in red. Sequence identifiers are provided to cross-reference the Supplemental Results for further information: IL2RG_R_4, IL2RG_R+D_4, IL2RG_R+D_7, IL2RG_R+D_8, HBB_T_1, HBB_T_3, HBB_T_5, HBB_R+D_1, HBB_R+D_3, CCR5_T_1, CCR5_T_2.

gene-editing outcomes with different donor architectures including plasmids, minicircle DNA, viral vectors, and ssODNs (Chen et al., 2003, 2011; Lombardo et al., 2007). Beyond targeted gene editing, this method offers an experimental system for studying DNA repair pathway utilization when DSBs occur at endogenous genomic loci following manipulation of DNA repair genes. As was shown in this study, and previous studies, parameters like nuclease properties, donor template architecture, cell type, and the site being modified can influence DSBR. The SMRT DNA sequencing method thus opens up possibilities for studying DSBR with engineered nuclease-induced breaks, where previous work has focused significantly on breaks induced by I-SceI at defined sites within reporter cell lines.

One area where application of SMRT DNA sequencing is challenging is the quantification of gene modifications that result in differently sized alleles, such as when entire linear donor templates are "captured" by NHEJ at "on-target" and "off-target" DSBs (Cristea et al., 2013; Gabriel et al., 2011). PCR bias when amplifying WT and modified loci with significantly different sizes can favor shorter alleles and confound quantification, which is further affected by loading bias for shorter molecules in the SMRT sequencing cells. For analysis of large gene inserts mediated by HDR, we have overcome this obstacle using embedded primers that distinguish between targeted and WT allele sequences while producing similar PCR amplicon sizes (Voit et al., 2014). Simultaneous measurement of amplicons with different lengths has also been achieved by adding a size standard ladder to the SMRT sequencing reaction, and a similar strategy could be used for quantification of large gene additions or NHEJ-mediated integrations of the donor template (Loomis et al., 2013).

By analyzing thousands of alleles within cell populations modified by TALENs and CRISPRs, this technique revealed the presence of rare insertional events where large stretches of DNA from other sources were integrated at nuclease cleavage sites. Choosing a cutoff of >30 bp to exclude sequences generated by DNA polymerase, we analyzed these inserted sequences and found matches to the donor template and nuclease expression plasmids introduced for gene targeting, sequences from nearby chromosomal sites, sites on other chromosomes, and sites within the *E. coli* genome that may have originated from trace impurities from the plasmid purification process. The presence of these events highlights the importance of minimizing the amount of exogenous DNA added for gene targeting and illustrates the potential for SMRT DNA sequencing to measure large, rare sequence alterations at sites throughout the genome.

The recent explosion in custom gene-editing technologies is ushering in a new age of genome engineering where scientists across fields of study and using different organisms and cell types can precisely modify essentially any locus they desire. Here, we show that SMRT DNA sequencing provides a simple, rapid, quantitative, and sensitive strategy for measuring genome-editing outcomes with different cell lines, at any endogenous loci, including transcriptionally silent loci, and using multiple nuclease platforms. Moreover, our strategy offers an approach for studying DNA repair pathway utilization when DNA breaks occur within genomic sites that have been difficult to study using previous methodologies. With the flexibility to evaluate engineered nucleases and targeting constructs directly at desired loci without the development of reporter systems, SMRT DNA sequencing can streamline the development of genome-editing projects and hasten the expansion of these technologies to a wider range of applications.

EXPERIMENTAL PROCEDURES

Construct Assembly

IL2RG TALENs were synthesized (GenScript) using the Δ152 N-terminal domain and the +63 C-terminal domain as previously described (Miller et al., 2011) and fused to the FokI nuclease domain and cloned into pcDNA3.1 (Invitrogen). *HBB* TALENs were previously described in Voit et al. (2014). *CCR5* TALENs containing the same RVDs as previously described in Mussolino et al. (2011) were made using a Golden Gate cloning strategy (Cermak et al., 2011) and cloned with the same N and C termini and nuclease domain into pcDNA3.1. *CCR5* ZFNs were previously described in Perez et al. (2008). For generating RGEN expression vectors, the bicistronic expression vector (pX330, provided by Dr. Feng Zhang, and also available through Addgene #42230) expressing Cas9 and sgRNA were digested, and the linearized vector was gel purified. Oligo pairs for the *IL2RG* and *HBB* sites (Table S10) were annealed, phosphorylated, and ligated to linearized vectors.

The *IL2RG*, *HBB*, and *CCR5* targeting vectors were constructed by PCR amplifying arms of homology from the corresponding loci using genomic DNA isolated from K562 cells. The point mutations that, upon successful homologous recombination, would be stably integrated into the genome and prevent binding and cleavage by the engineered nucleases were added as part of the PCR primers used to generate the arms of homology. The homology arms were then cloned into a ~2,900 base pair vector based on pBluescript SK^+ using standard cloning methods. The full sequences of all donor constructs used in this study are provided in the Plasmid Sequences section of the Supplemental Information.

Cell Culture

K562 cells (ATCC) were maintained in RPMI 1640 (HyClone) supplemented with 10% bovine growth serum, 100 units/ml penicillin, 100 μg/ml streptomycin, and 2 mM L-glutamine. Human $CD34^+$ hematopoietic stem/progenitor cells (HSPCs) were purchased from Lonza (2M-101B) and thawed per the manufacturer's instructions. $CD34^+$ HSPCs were maintained in X-VIVO15 (Lonza) supplemented with SCF (100 ng/ml), TPO (100 ng/ml), Flt3-Ligand (100 ng/ml), IL-6 (100 ng/ml), and StemRegenin1 (0.75 μM). hESC line H1 (WiCell) was maintained in feeder-free culture conditions in mTeSR1 (STEMCELL Technologies) on a thin layer of Matrigel (Becton Dickinson). Cultures were passaged every 3–5 days enzymatically with Accutase (Innovative Cell Technologies). Cells were transfected between passage 45 and 47.

Transient Transfection for Genome Editing

K562 cells (1×10^6) were transfected with 2 μg TALEN-encoding plasmid and 5 μg donor plasmid (unless otherwise indicated) by nucleofection (Lonza) with program T-016 and a nucleofection buffer containing 100 mM KH2PO4, 15 mM $NaHCO_3$, 12 mM MgCl2 × 6H20, 8 mM ATP, 2 mM glucose (pH 7.4). $CD34^+$ HSPCs (4×10^5) were nucleofected with an Amaxa 4D Nucleofector with the P3 Primary Cell Nucleofector Kit (V4XP-3032) and program EO-100 per the manufacturer's instructions. H1 cells (1×10^6) were transfected with 0.5 or 2.5 μg of each TALEN-encoding plasmid and 4 μg donor plasmid (unless otherwise indicated) by nucleofection (Lonza) with an Amaxa 4D Nucleofector (program B-105) with the P3 Primary Cell Nucleofector Kit (V4XP-3032) and following manufacturer's instructions.

Flow Cytometry

Samples were collected 72 hr after nucleofection and analyzed for fluorescence using an Accuri C6 flow cytometer. GFP expression was measured using a 488 nm laser for excitation and a 530/30 band-pass filter for detection.

Restriction Fragment Length Polymorphism Assay

Restriction fragment length polymorphism assay was performed as previously described in Chen et al. (2011). Briefly, Genomic DNA was extracted from transfected cells with DNeasy Blood & Tissue Kit (QIAGEN). Genomic DNA was then PCR amplified with primers flanking the donor target region (see Table S10 for PCR primer sequences). The amplification was carried out with Accuprime polymerase (Invitrogen), using the following cycling conditions: 95°C for 5 min for initial denaturation; 30 cycles of 95°C for 30 s, 67°C for 45 s, and 68°C for 120 s; and a final extension at 68°C for 5 min. PCR products were digested with 20 U of AflII at 37°C for ~2 hr and resolved with PAGE.

Single-Cell Clone Analysis

Single-cell cloning was performed by flow cytometry cell sorting on a BD FACSAria. Genomic DNA was isolated from single clones using the QIAGEN DNeasy kit (QIAGEN). The *IL2RG* target region was PCR amplified with Accuprime polymerase (Invitrogen) and the following cycling conditions: 95°C for 5 min for initial denaturation; 30 cycles of 95°C for 30 s, 67°C for 45 s and 68°C for 120 s; and a final extension at 68°C for 5 min. PCR amplicons were sequenced using standard Sanger sequencing. Sequences were analyzed using the ApE plasmid editor by M. Wayne Davis.

SMRT Sequencing

Genomic DNA containing *IL2RG* alleles was harvested from cultured K562, $CD34^+$ HSPC, and hESC samples using the DNeasy Blood & Tissue Kit (QIAGEN). *IL2RG* alleles were amplified using the primers in Table S10 with Accuprime polymerase (Invitrogen) and the following cycling conditions: 95°C for 5 min for initial denaturation; 30 cycles of 95°C for 30 s, 67°C for 45 s, and 68°C for 60 s; and a final extension at 68°C for 5 min for the K562 samples and 95°C for 5 min for initial denaturation; 30 cycles of 95°C for 30 s, 67°C for 45 s, and 68°C for 90 s; and a final extension at 68°C for 5 min for the HSPC and hESC samples. Sequencing libraries were constructed, as previously described (Travers et al., 2010), using the DNA Template Prep Kit 1.0 (Pacific Biosciences). SMRTbell libraries contained amplicons that were pooled together, with different barcodes appended to allow multiplex analysis. Purified, closed circular SMRTbell libraries were annealed with a sequencing primer complementary to a portion of the single-stranded region of the hairpin. For all SMRTbell libraries, annealing was performed at a final template concentration between 30 and 60 nM, with a 20-fold molar excess of sequencing primer. All annealing reactions were carried out at 80°C for 2 min, with a slow cool to 25°C at a rate of 0.1°C/s. Annealed templates were stored at −20°C until polymerase binding. DNA polymerase enzymes were stably bound to the primed sites of the annealed SMRTbell templates using the DNA Polymerase Binding Kit 2.0 (Pacific Biosciences). SMRTbell templates (3 nM) were incubated with 6 nM of polymerase in the presence of phospholinked nucleotides at 30°C for 2 hr. Following incubation, samples were stored at 4°C. Sequencing was performed within 72 hr of binding using final on plate concentration of 0.3 nM. Each sample was sequenced as previously described (Rasko et al., 2011) using DNA Sequencing Kit 2.0 (Pacific Biosciences). Sequencing data collection was performed on the PacBio RS (Pacific Biosciences) using C2/C2 chemistry and movies of 55 min in each case.

SMRT Analysis Pipeline

The SMRT Sequencing Analysis pipeline was implemented in Strawberry Perl and utilizes the NCBI BLAST software as well as the mEmboss Needleman-Wunsch pairwise alignment algorithm. All components of the pipeline were run on a standard Windows PC and are available for download (https://sourceforge.net/projects/tdna-getsmart/). Further details and description of the pipeline are available in Supplemental Experimental Procedures SMRT Analysis Pipeline.

Statistical Analysis

To calculate confidence intervals, t statistics were calculated as previously described (Pattanayak et al., 2011). We calculated 90% confidence intervals by determining the upper and lower bounds of the mutation rates that would yield p values of 0.05, and 66% confidence intervals were calculated with a target p value of 0.32.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Results, Supplemental Experimental Procedures, one figure, and ten tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.02.040.

AUTHOR CONTRIBUTIONS

A.H., E.J.K., J.T.C., and N.P. performed experiments. E.J.F. analyzed all SMRT sequencing data. V.S. performed experiments with hESCs. G.B. and M.H.P. directed the research. A.H. and E.J.K. wrote the manuscript with help from all authors.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health as an NIH Nanomedicine Development Center Award (PN2EY018244 to M.H.P. and G.B.). This work was also supported by NIH grant R01 AI097320 from the NIAID to M.H.P. M.H.P. also thanks the Laurie Krauss Lacob Faculty Scholar Award for ongoing support. 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 NSF Graduate Research Fellowship under Grant No. DGE-1148903. We thank Dr. Norma Neff, Ben Passarelli, Jad Kanbar, and Pacific Biosciences for help with the SMRT sequencing technology. We thank Prof. Stephen Quake, Prof. Michael Snyder, and Prof. Joseph Puglisi for letting us using their PacBio RS instruments. We thank Shondra Miller, John Eid, Jonas Korlach, Aaron Straight, and Porteus and Bao lab members for helpful comments and discussion.

Received: December 4, 2013
Revised: January 16, 2014
Accepted: February 26, 2014
Published: March 27, 2014

REFERENCES

Beumer, K.J., Trautman, J.K., Mukherjee, K., and Carroll, D. (2013). Donor DNA utilization during gene targeting with zinc-finger nucleases. G3 *3*, 657–664.

Bogdanove, A.J., and Voytas, D.F. (2011). TAL effectors: customizable proteins for DNA targeting. Science *333*, 1843–1846.

Briggs, A.W., Rios, X., Chari, R., Yang, L., Zhang, F., Mali, P., and Church, G.M. (2012). Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. Nucleic Acids Res. *40*, e117.

Cavalli, G., and Misteli, T. (2013). Functional implications of genome topology. Nat. Struct. Mol. Biol. *20*, 290–299.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., and Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. *39*, e82.

Certo, M.T., Ryu, B.Y., Annis, J.E., Garibov, M., Jarjour, J., Rawlings, D.J., and Scharenberg, A.M. (2011). Tracking genome engineering outcome at individual DNA breakpoints. Nat. Methods *8*, 671–676.

Chen, Z.Y., He, C.Y., Ehrhardt, A., and Kay, M.A. (2003). Minicircle DNA vectors devoid of bacterial DNA result in persistent and high-level transgene expression in vivo. Mol. Ther. *8*, 495–500.

Chen, F., Pruett-Miller, S.M., Huang, Y., Gjoka, M., Duda, K., Taunton, J., Collingwood, T.N., Frodin, M., and Davis, G.D. (2011). High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat. Methods *8*, 753–755.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Cristea, S., Freyvert, Y., Santiago, Y., Holmes, M.C., Urnov, F.D., Gregory, P.D., and Cost, G.J. (2013). In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. Biotechnol. Bioeng. *110*, 871–880.

Daboussi, F., Zaslavskiy, M., Poirot, L., Loperfido, M., Gouble, A., Guyot, V., Leduc, S., Galetto, R., Grizot, S., Oficjalska, D., et al. (2012). Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. Nucleic Acids Res. *40*, 6367–6379.

Doyon, Y., Choi, V.M., Xia, D.F., Vo, T.D., Gregory, P.D., and Holmes, M.C. (2010). Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. Nat. Methods *7*, 459–460.

Doyon, Y., Vo, T.D., Mendel, M.C., Greenberg, S.G., Wang, J., Xia, D.F., Miller, J.C., Urnov, F.D., Gregory, P.D., and Holmes, M.C. (2011). Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat. Methods *8*, 74–79.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., et al. (2009). Real-time DNA sequencing from single polymerase molecules. Science *323*, 133–138.

Fan, L., Kadura, I., Krebs, L.E., Hatfield, C.C., Shaw, M.M., and Frye, C.C. (2012). Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. Biotechnol. Bioeng. *109*, 1007–1015.

Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaeppel, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedman, G., et al. (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat. Biotechnol. *29*, 816–823.

Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. *31*, 397–405.

Gallagher, P.G. (2013). Disorders of red cell volume regulation. Curr. Opin. Hematol. *20*, 201–207.

Guschin, D.Y., Waite, A.J., Katibah, G.E., Miller, J.C., Holmes, M.C., and Rebar, E.J. (2010). A rapid and general assay for monitoring endogenous gene modification. Methods Mol. Biol. *649*, 247–256.

Kildebeck, E., Checketts, J., and Porteus, M. (2012). Gene therapy for primary immunodeficiencies. Curr. Opin. Pediatr. *24*, 731–738.

Kim, H.J., Lee, H.J., Kim, H., Cho, S.W., and Kim, J.S. (2009). Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Res. *19*, 1279–1288.

Kim, Y., Kweon, J., Kim, A., Chon, J.K., Yoo, J.Y., Kim, H.J., Kim, S., Lee, C., Jeong, E., Chung, E., et al. (2013). A library of TAL effector nucleases spanning the human genome. Nat. Biotechnol. *31*, 251–258.

Kuhar, R., Gwiazda, K.S., Humbert, O., Mandt, T., Pangallo, J., Brault, M., Khan, I., Maizels, N., Rawlings, D.J., Scharenberg, A.M., and Certo, M.T. (2014). Novel fluorescent genome editing reporters for monitoring DNA repair pathway utilization at endonuclease-induced breaks. Nucleic Acids Res. *42*, e4.

Li, T., Liu, B., Spalding, M.H., Weeks, D.P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat. Biotechnol. *30*, 390–392.

Lombardo, A., Genovese, P., Beausejour, C.M., Colleoni, S., Lee, Y.L., Kim, K.A., Ando, D., Urnov, F.D., Galli, C., Gregory, P.D., et al. (2007). Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat. Biotechnol. *25*, 1298–1306.

Loomis, E.W., Eid, J.S., Peluso, P., Yin, J., Hickey, L., Rank, D., McCalmon, S., Hagerman, R.J., Tassone, F., and Hagerman, P.J. (2013). Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. Genome Res. *23*, 121–128.

Mali, P., Esvelt, K.M., and Church, G.M. (2013). Cas9 as a versatile tool for engineering biology. Nat. Methods *10*, 957–963.

McMahon, M.A., Rahdar, M., and Porteus, M. (2012). Gene editing: not just for translation anymore. Nat. Methods *9*, 28–31.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., et al. (2011). A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. *29*, 143–148.

Mussolino, C., Morbitzer, R., Lütge, F., Dannemann, N., Lahaye, T., and Cathomen, T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res. *39*, 9283–9293.

Orlando, S.J., Santiago, Y., DeKelver, R.C., Freyvert, Y., Boydston, E.A., Moehle, E.A., Choi, V.M., Gopalan, S.M., Lou, J.F., Li, J., et al. (2010). Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. Nucleic Acids Res. *38*, e152.

Papamichos-Chronakis, M., and Peterson, C.L. (2013). Chromatin and the genome integrity network. Nat. Rev. Genet. *14*, 62–75.

Pattanayak, V., Ramirez, C.L., Joung, J.K., and Liu, D.R. (2011). Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. Nat. Methods *8*, 765–770.

Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., Wang, N., Lee, G., Bartsevich, V.V., Lee, Y.L., et al. (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat. Biotechnol. *26*, 808–816.

Porteus, M.H., and Baltimore, D. (2003). Chimeric nucleases stimulate gene targeting in human cells. Science *300*, 763.

Porteus, M.H., and Carroll, D. (2005). Gene targeting using zinc finger nucleases. Nat. Biotechnol. *23*, 967–973.

Qi, Y., Zhang, Y., Zhang, F., Baller, J.A., Cleland, S.C., Ryu, Y., Starker, C.G., and Voytas, D.F. (2013). Increasing frequencies of site-specific mutagenesis and gene targeting in Arabidopsis by manipulating DNA repair pathways. Genome Res. *23*, 547–554.

Rasko, D.A., Webster, D.R., Sahl, J.W., Bashir, A., Boisen, N., Scheutz, F., Paxinos, E.E., Sebra, R., Chin, C.S., Iliopoulos, D., et al. (2011). Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. N. Engl. J. Med. *365*, 709–717.

Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D., and Joung, J.K. (2012). FLASH assembly of TALENs for high-throughput genome editing. Nat. Biotechnol. *30*, 460–465.

Roberts, R.J., Carneiro, M.O., and Schatz, M.C. (2013). The advantages of SMRT sequencing. Genome Biol. *14*, 405.

Shaw, K.L., and Kohn, D.B. (2011). A tale of two SCIDs. Sci. Transl. Med. *3*, 97ps36.

Silva, G., Poirot, L., Galetto, R., Smith, J., Montoya, G., Duchateau, P., and Pâques, F. (2011). Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. Curr. Gene Ther. *11*, 11–27.

Soldner, F., Laganière, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S., et al. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell *146*, 318–331.

Stark, J.M., Pierce, A.J., Oh, J., Pastink, A., and Jasin, M. (2004). Genetic steps of mammalian homologous repair with distinct mutagenic consequences. Mol. Cell. Biol. *24*, 9305–9316.

Travers, K.J., Chin, C.S., Rank, D.R., Eid, J.S., and Turner, S.W. (2010). A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. *38*, e159.

Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., and Holmes, M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature *435*, 646–651.

Valton, J., Daboussi, F., Leduc, S., Molina, R., Redondo, P., Macmaster, R., Montoya, G., and Duchateau, P. (2012a). 5′-Cytosine-phosphoguanine (CpG) methylation impacts the activity of natural and engineered meganucleases. J. Biol. Chem. *287*, 30139–30150.

Valton, J., Dupuy, A., Daboussi, F., Thomas, S., Maréchal, A., Macmaster, R., Melliand, K., Juillerat, A., and Duchateau, P. (2012b). Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. J. Biol. Chem. *287*, 38427–38432.

van Rensburg, R., Beyer, I., Yao, X.Y., Wang, H., Denisenko, O., Li, Z.Y., Russell, D.W., Miller, D.G., Gregory, P., Holmes, M., et al. (2013). Chromatin

structure of two genomic sites for targeted transgene integration in induced pluripotent stem cells and hematopoietic stem cells. Gene Ther. *20*, 201–214.

Voit, R.A., McMahon, M.A., Sawyer, S.L., and Porteus, M.H. (2013). Generation of an HIV resistant T-cell line by targeted "stacking" of restriction factors. Mol. Ther. *21*, 786–795.

Voit, R.A., Hendel, A., Pruett-Miller, S.M., and Porteus, M.H. (2014). Nuclease-mediated gene editing by homologous recombination of the human globin locus. Nucleic Acids Res. *42*, 1365–1378.

Waldman, A.S. (2008). Ensuring the fidelity of recombination in mammalian chromosomes. Bioessays *30*, 1163–1171.

Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., and Jaenisch, R. (2013a). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell *154*, 1370–1379.

Yang, L., Guell, M., Byrne, S., Yang, J.L., De Los Angeles, A., Mali, P., Aach, J., Kim-Kiselak, C., Briggs, A.W., Rios, X., et al. (2013b). Optimization of scarless human stem cell genome editing. Nucleic Acids Res. *41*, 9049–9061.