Off-Target Assessment of CRISPR-Cas9 Guiding RNAs in Human iPS and Mouse ES Cells

E-Pien Tan, Yilong Li, Martin Del Castillo Velasco-Herrera, Kosuke Yusa, and Allan Bradley*

Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom

Received 22 October 2014; Accepted 30 October 2014

Summary: The CRISPR-Cas9 system consists of a site-specific, targetable DNA nuclease that holds great potential in gene editing and genome-wide screening applications. To apply the CRISPR-Cas9 system to these assays successfully, the rate at which Cas9 induces DNA breaks at undesired loci must be understood. We characterized the rate of Cas9 off-target activity in typical Cas9 experiments in two human and one mouse cell lines. We analyzed the Cas9 cutting activity of 12 gRNAs in both their targeted sites and ~90 predicted off-target sites per gRNA. In a Cas9-based knockout experiment, gRNAs induced detectable Cas9 cutting activity in all on-target sites and in only a few off-target sites genome-wide in human 293FT, human-induced pluripotent stem (hiPS) cells, and mouse embryonic stem (ES) cells. Both the cutting rates and DNA repair patterns were highly correlated between the two human cell lines in both on-target and off-target sites. In clonal Cas9 cutting analysis in mouse ES cells, biallelic Cas9 cutting was observed with low off-target activity. Our results show that off-target activity of Cas9 is low and predictable by the degree of sequence identity between the gRNA and a potential off-target site. Off-target Cas9 activity can be minimized by selecting gRNAs with few off-target sites of near complementarity.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited. Additional Supporting Information may be found in the online version of this article.

Key words: mammal; organism; genetics; process

INTRODUCTION

Site-specific DNA nucleases are an important class of molecular tools that can generate double-strand DNA breaks at virtually any locus in a target genome. Because of the flexibility of their sequence specificity, they have enabled a wide range of applications in the field of genetic modification (Joung and Sander, 2013; Urnov et al., 2010). The CRISPR-Cas system consists of naturally occurring clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins. The system represents one of the most recently developed site-specific DNA nuclease system. The CRISPR-Cas systems are unique adaptive defense mechanisms that have evolved in bacteria and archaea to degrade foreign DNA from invading viruses or plasmids (Makarova et al., 2011). The catalytic activity of these systems is provided by a DNA nuclease that is guided to a target sequence by a ~20-bp guide RNA oligonucleotide complementary to the target sequence (Mali et al., 2013b). The CRISPR-Cas system therefore resembles earlier generations of site-specific DNA nucleases, e.g., zinc finger nucleases (ZFNs) or transcription activator-like effector nucleases (TALENs), in that the DNA nuclease is targetable to a specific locus in a complex genome.

There are three major types of CRISPR-Cas system with further divisions into numerous subtypes and variants (Makarova et al., 2011). Among these, the type II CRISPR-Cas9 system from Streptococcus pyogenes has been adapted into a genome-editing tool that allows mutations to be efficiently introduced into mammalian cells (Cho et al., 2013; Cong et al., 2013; Mali et al., 2013c). The system requires three minimal
components: the CRISPR-associated nuclease (Cas9), a CRISPR RNA (crRNA), and a partially complementary trans-activating crRNA (tracrRNA) (Deltcheva et al., 2011). The system has been simplified by joining the two RNA components into a single chimeric molecule known as a guiding RNA (gRNA) (Mali et al., 2013c). In the native system, the Cas9 nuclease is directed to genomic loci complementary to the 20-nucleotide guide sequence specified by the crRNA. This system has been used to generate DNA double-strand breaks at high efficiency in numerous specific sites in a variety of genomes, indicating the versatility of the system.

Cas9 targeting is theoretically limited by only the presence of a protospacer adjacent DNA motif (PAM) sequence (NGG or NAG) at the 3’ end of the 20-nucleotide target sequence (Hsu et al., 2013; Pattanayak et al., 2013); it is thus possible to generate genomewide DNA double-stranded breaks with a theoretical frequency of one in every eight nucleotides. Compared to earlier site-specific nucleases such as ZFNs or TALENs, Cas9 is much more straightforward to use as it only requires the generation of a crRNA with complementarity to the desired target locus which eliminates the costly and time-consuming protein engineering required with ZFNs and TALENs.

In addition to cleaving an intended target locus, Cas9 may also induce double-strand breaks at unintended “off-target sites” (Fu et al., 2013; Hsu et al., 2013; Pattanayak et al., 2013). Such off-target activity is an important experimental consideration that may confound results by disrupting genes elsewhere in the genome. It is thus imperative to understand the rate and nature of such Cas9 off-target activity. The level of off-target activity has been assessed in several studies where the Cas9 system was used to modify the genomes of various mammalian cells and organisms (Cong et al., 2013; Hwang et al., 2013; Mali et al., 2013c). For example, Hwang et al. found that zebrafish embryos modified by Cas9–gRNA complexes exhibited similar lethality to those modified by ZFNs and TALENs (Hwang et al., 2013). A recent study of transcriptional repression in Escherichia coli with a catalytically inactive mutant Cas9 reported efficient repression of targeted genes with no observable off-target effects (Qi et al., 2013). These studies suggest that there is a low toxicity level of Cas9 at the organismal level.

Previous findings suggest that the specificity of Cas9-mediated DSBs arises from the complementarity between a gRNA and its target site in the nucleotides adjacent to the PAM sequence (3’ end of the gRNA). This is in contrast to the mismatches at the non-PAM end (5’ end of the gRNA), which are tolerated (Cong et al., 2013; Jiang et al., 2013; Jinek et al., 2012; Semenova et al., 2011). Recently, three studies have been conducted to directly assess the rate of off-target activities in the CRISPR-Cas9 system (Fu et al., 2013; Hsu et al., 2013; Pattanayak et al., 2013). Pattanayak et al. analyzed the off-target activities of eight gRNA–Cas9 complexes in 10^{12} potential DNA sequences via in vitro selection and high-throughput sequencing. They identified 51 off-target sites for two gRNAs that were enriched in the in vitro selection and performed genomic polymerase chain reaction (PCR) and high-throughput sequencing to look for in vitro cutting in a human embryonic kidney cell line (293FT). From these, they identified a total of five off-target sites. Their model also suggests that Cas9 specificity extends past the seed sequence and that a shorter, less active version of a gRNA is more specific than a longer, more active version of the gRNA (Pattanayak et al., 2013). Hsu et al. generated approximately 700 gRNA variants with mismatches in most positions with respect to four original guide sequences (approximately 175 gRNA variants per target gRNA) and assessed their ability to generate insertions and/or deletions (indels) at the target site in 293FT cells. They reported that mismatches between a gRNA and its target DNA are tolerated at different positions along the length of the gRNA (Hsu et al., 2013). Fu et al. tested six gRNAs and analyzed their corresponding potential off-target sites (ranging from 46 to 64) in U2OS cells by the T7 endonuclease I assay. They found a total of 24 off-target sites with cutting efficiencies ranging from 5.6 to 125% compared with the on-target site. As these off-target sites contained up to five mismatches and many were mutagenized at comparable efficiencies to the on-target site, they concluded that Cas9-induced double-stranded breaks can be nonspecific and highly promiscuous (Fu et al., 2013). All three studies mentioned above investigated the specificity of gRNAs as opposed to the off-target propensity of a given gRNA in the context of living cells. It is unclear how these experimentally derived rules can be translated into real off-target cutting rates in living cells.

In this study, we directly evaluated the rates of off-target activities of the Cas9 system in human-induced pluripotent stem (hiPS) cells, a human embryonic kidney cell line (293FT), and a mouse embryonic stem (ES) cell line. We comprehensively studied genome-wide off-target cutting rates by analyzing a total of 12 gRNAs and their corresponding 95 predicted off-target sites (up to five mismatches) for each gRNA. Furthermore, we performed off-target activity analysis in single mouse ES cell clones to understand off-target cutting during CRISPR-mediated gene-targeting experiments.

**RESULTS**

**Off-Target Activity of hCas9 in 293FT, hiPS, and Mouse ES Cells**

Seven gRNAs were generated that target four different genes within the human genome: SLC35A2 (SLCA...
and SLCB), ATP6AP2 (ATPA and ATPB), CDK19 (CDKA and CDKB), and AKT2 (AKT2). Four gRNAs were generated for the mouse Slc35a2 (mSlcA), Atp6ap2 (mAtpA), and Cdk19 (mCdA and mCdB) genes. To assess the ability of each gRNA to generate double-strand DNA breaks in their respective cell types, we transfected each gRNA into the corresponding human or mouse cell lines independently: 293FT, hiPS, or mouse ES cells along with a hCas9 expression vector. Three days after transfection, each locus was amplified by PCR and the products were mixed and sequenced on the MiSeq platform. Indel rates, ranging from 2.7 to 13.4%, were observed for all 12 tested gRNAs at their respective on-target sites (Table 1).

The off-target activities of each gRNA were assessed by locus-specific PCR and MiSeq analysis at 95 potential off-target sites per gRNA (Supporting Information Fig. 1 and Supporting Information Table 2). In total, we examined 624 potential off-target sites in 293FT cells, 717 potential sites in hiPS cells, and 354 potential sites in mouse ES cells (Supporting Information Tables 1 and 3–5). At the significance threshold of 0.05 (q-value < 0.05), we found six significant off-target sites in 293FT cells, two significant off-target sites in hiPS cells, and one significant off-target site in mouse ES cells (Fig. 1 and Table 1).

Five of the six significant off-target sites found in the human cell lines were off-targets of the AAVS1 gRNA. The remaining significant off-target site came from the CDKB gRNA. However, close analysis of the indel pattern at this off-target site reveals that this is a false positive because the same deletion peaks are observed in both the control and experimental samples (Supporting Information Fig. 2). This pattern was not observed in true on and off-target sites, where several prominent indel size peaks were always present in the experimental but not in the control conditions (Supporting Information Figs. 4–6). None of the remaining gRNAs tested produced any significant off-target cutting rates. The two significant off-target sites found in hiPS cells (AAVS1_14 and AAVS1_11) were also found in 293FT cells. Four of five identified off-target sites contained three mismatches against the AAVS1 gRNA target sequence. The remaining identified off-target site against the AAVS1 gRNA target sequence had four mismatches. Out of these, two sites contain one mismatch in the seed region (12 nucleotides adjacent to the PAM sequence), whereas the remaining three sites contain two mismatches in the seed region (Table 1).

Three of the gRNAs tested in mouse ES cells did not yield any significant off-target sites. The only off-target site found had one mismatch against the mAtpA gRNA target sequence in the seed region. The cutting efficiencies of the off-target sites were always lower than those of their corresponding on-target sites (Table 1). We further observe that the cutting efficiencies for both on-target and genuine off-target sites are highly correlated between the two human-derived cell lines with a Pearson correlation coefficient of 0.7 (Fig. 2a).

We next investigated the indel patterns at all on-target and genuine off-target sites by analyzing the distribution of indel sizes at these loci (Supporting Information Fig. 3). All loci show a distinct indel distribution pattern, which was very similar in the different human cell lines, implying a conserved repair mechanism. Out of the 18 cut sites analyzed, 17 had a prominent deletion peak and only one, the on-target site of SLCB gRNA, had a dominant 1-bp insertion peak. The most prominent deletion peak was often associated with 2- to 4-bp microhomologies (Fig. 2b).

We next investigated whether point mutations, rather than indels, could be generated by the activity of CRISPR-Cas9 or the DNA-repair induced by Cas9-mediated DNA cutting. We compared the allelic proportions of 8,064 nucleotides in 293FT cells, 9,216 nucleotides in hiPS cells, and 4,068 nucleotides in mouse ES cells around the on- and potential off-target cutting sites between experimental and control conditions and found no evidence of altered allelic proportions (Supporting Information Table 6). Similarly, of the 960 nucleotides analyzed from the mouse ES cell colonies, all had retained the major allele (Supporting Information Table 6).

Clonal Analysis of Off-Target Sites in Mouse ES Cells

We next sought to understand whether both on-target cutting and off-target damage could occur in the same cell. Off-target activity is of specific interest in the context of directed genetic manipulation such as gene targeting. Using the four gRNAs described for mouse genes, we paired each with a targeting vector and generated targeted clones. We analyzed between 48 and 96 independent clones per gRNA and found that CRISPR-guided gene targeting in mouse ES cells occurred at efficiencies of 4–31% (Table 2).

We selected 15 targeted clones per gRNA and analyzed on-target and off-target cutting by next-generation sequencing of the targeted locus and an additional 19 potential off-target sites. Only two targeted clones with mAtpA gRNA were analyzed (Table 2). No reads were obtained at the on-target site of all mSlcA and mAtpA targeted clones (Fig. 3c, Supporting Information Fig. 7); this is expected because these genes are on the X chromosome. In contrast, Cdk19 is an autosomal gene and analysis of the targeted clones generated with mCdA and mCdB gRNAs revealed that two and four clones were homozygously targeted, respectively (Fig. 3c, Table 2, Supporting Information Fig. 7). The remaining colonies that did not yield any MiSeq reads were colonies that have been successfully targeted on one allele.
## Table 1
### On- and Off-Target Cas9 Cutting Sites

| gRNA       | Target ID | Target locus                  | Target sequence                  | PAM | Human 293FT Cutting efficiency % | Human iPS cells Cutting efficiency % |
|------------|-----------|--------------------------------|----------------------------------|-----|----------------------------------|--------------------------------------|
| AAVS1 On-target | 19:55627117–55627139 | G G G G C C A C T A G G G A C A G G A T TGG | 2.83E–96 | 7.7 | 5.03E–115 | 5.3 |
| AAVS1 14       | 6:36765484–36765486 | G G G G C C A C T A G G G A C A G G A T TGG | 6.27E–33 | 3.3 | 6.65E–45 | 2.6 |
| AAVS1 11       | 13:106612912–106612934 | G G G G C C A A T T A G G G A C A G G A T TGG | 2.66E–20 | 2.1 | 1.35E–02 | 0.4 |
| AAVS1 13       | 19:16174987–16175009 | G G G G C T T C T A A G G A C A G G A T TGG | 1.31E–06 | 1.3 | NS | NS |
| AAVS1 34       | 19:45608588–45650880 | G G G G C T G C C A G G A C A G G C T TGG | 6.57E–06 | 0.9 | NS | NS |
| AAVS1 7        | 21:42892942–42892964 | G G G G C C C C T G G G A C A G A A T G G G A C G | 2.18E–02 | 0.6 | NS | NS |
| AAVS1 34       | 19:40762954–40762976 | G T A T T C A C C T G A A A T G A G G C AGG | ND | ND | 6.70E–81 | 4.0 |
| mATPA On-target | X:12604941–12604963 | G C A G T T G G C T C C A G G G A C A G G A T TGG | 7.92E–74 | 2.7 | 3.0 | 5.36E–81 | 9.9 |
| mATPA 2        | 14:52540780–52540802 | G C A G T T G G C T C C A G G G A C A G G A T TGG | 1.15E–30 | 1.0 | NS | NS |
| mCdA On-target | 10:40346161–40346183 | G G A G G T A T G G A T G C C A T C G | 2.18E–69 | 5.2 | NS | NS |
| mCdB On-target | X:7884390–7884412 | G G G C T G T G T C C T C G G G A C G G G C T TGG | 7.13E–119 | 3.8 | 3.77E–204 | 9.9 |
| mSlA On-target | X:84767091–84767113 | G C T G C T C T C T C G C A C A G A G A A G G | 6.29E–85 | 7.7 | 2.32E–95 | 4.9 |

This table shows all the sites with statistically significant cutting rates. All designed on-target sites and seven additional off-target sites had significant cutting.

FDR, false discovery rate; ND, not determined; NS, not significant.
and contained a large deletion (>250 bp) on the other allele. Fifteen colonies were targeted on one allele and disrupted by a small deletion on the other allele: 10 for mCdA and five for mCdB. Two of the mCdA targeted clones were found to contain three alleles: a targeted allele, a wild-type allele, and a Cas9-mediated indel-containing allele. It is likely that these two clones were mosaic, reflecting the timing of genetic alterations in their clones. Out of the 15 analyzed clones for mCdB, only one clone was heterozygous (targeted on one allele and wild-type on the other allele).

We next sought to analyze the off-target cutting efficiencies of the four gRNAs in a clonal setting. Consistent with the pooled analysis, we found that all analyzed clones targeted using mCdA, mCdB, or mSlcA gRNAs and their corresponding targeting vectors did not display any significant off-target cutting. However, for both of the mAtpA targeted clones, we observed off-target cutting at the same site (mAtpA_2) identified in the pooled off-target site analysis experiment. For these two clones, our data indicate that both alleles of this locus were cut (Fig. 3c, Supporting Information Fig. 7). We further observed that mAtpA colony 2 had an additional hemizygous deletion at the off-target site mAtpA_4.

DISCUSSION

The CRISPR-Cas9 system is a highly efficient genome modification tool that has potential applications throughout biology and medicine. Recently, the system has been used to generate genome-wide mutations in mammalian cells (Koike-Yusa et al., 2014; Shalem et al., 2014; Wang et al., 2014). Successful screens conducted in such knockout libraries have identified genes involved in resistance to Clostridium septicum alpha-toxin and resistance to 6-thioguanine in mouse ES cells, genes essential for cell viability in pluripotent stem cells and genes involved in resistance to 6-thioguanine and...
the DNA topoisomerase II in a human chronic myeloid leukemia cell line. For the purposes of genome modification, it is crucial that the performances of individual gRNAs are specific to the on-target site with limited off-target events. Recent research suggests that the CRISPR-Cas9 system has different off-target cutting frequencies in mammalian cells (Fu et al., 2013; Hsu et al., 2013; Mali et al., 2013a, Pattanayak et al., 2013). It has been suggested that high GC content gRNAs are associated with high off-target cutting events (Fu et al., 2013). However, Koike-Yusa et al. tested a gRNA (55% GC content) targeting the Piga gene and reported only two off-

FIG. 2. Cas9 cutting efficiency and repair patterns are consistent between human cell lines. (a) Estimated cutting efficiencies at Cas9 on- and off-target sites. The rate of cutting is highly correlated between the two cell lines with a Pearson correlation of 0.70. The correlation increases to 0.92 if CDKA_1 is discarded, which had highly disparate cutting rates in the replicate experiments in 293FT and hiPS cells (Supporting Information Fig. 3). (b) Indel size distribution at both replicates of human 293FT and hiPS cells at CDKB on-target (left) and AAVS1_14 off-target sites (right). Negative values indicate sizes of deletions while positive values indicate sizes of insertions. The sequence context at the most frequent deletion (representing the highest peak in both sites) is shown above the barplots. The cutting site is shown with a |, the deleted sequence is in lowercase red alphabets, and identified microhomology is shown in bold.
target cleavages out of 275 potential off-target sites analyzed (Koike-Yusa et al., 2014). It is as yet unclear what factors govern the specificity of a particular gRNA.

We examined 12 gRNAs in three cell lines, all of which could successfully target Cas9 to the intended target locus and induce DNA double-strand breaks (Table 1). Furthermore, the indel rates were highly correlated between the two human cell lines in both on-target and genuine off-target sites. Our results are consistent with a model in which the Cas9 cutting activity at any given site is primarily determined by sequence complementarity between gRNA and the DNA.

**Table 2**

**gRNA-Mediated Targeting Efficiency in Mouse ES Cells**

| gRNA ID  | Chr. | No. of clones analyzed | No. of targeted clones | Targeting efficiency (%) | Genotype of analyzed clones |
|----------|------|------------------------|------------------------|--------------------------|---------------------------|
| mAtpA    | X    | 48                     | 2                      | 4                        | – – –                     |
| mSlcA    | X    | 96                     | 19                     | 20                       | – – –                     |
| mCdA     | 10   | 86                     | 30                     | 31                       | 0 13 2                    |
| mCdB     | 10   | 84                     | 24                     | 29                       | 1 10 4                    |

This table shows the targeting efficiency for the four loci in mouse ES cells.

T/+, one targeted allele, one wild-type allele; T/Δ, one targeted allele, one Cas9-mediated indel-containing allele; T/T, two targeted alleles.

**FIG. 3.** Cas9 on- and off-target cutting analysis in targeted mouse ES cells. (a) General design of the targeting vectors, (b) Genotyping design for targeted clones, and (c) PCR genotyping results and cutting patterns for each of the four studied gRNAs. mAtpA, mCdA, mCdB, and mSlcA refer to the on-target site of each locus, respectively. No off-target Cas9 activity was found in mCdA, mCdB, or mSlcA, whereas off-target activity was detected alongside targeting of the mAtpA gene. Both targeted clones exhibited cutting of the mAtpA_2 site and one also showed cutting of mAtpA_4.
sequence on the site rather than cell type-specific epigenomic architecture. However, based on data generated by the ENCODE consortium (UCSC browser), we found that gene expression and DNase hypersensitive sites at the on-target and genuine off-target sites were very similar between the two cell lines (data not shown) (Rosenbloom et al., 2013). Therefore, further studies specifically targeting Cas9 to epigenetically diverse loci are needed to examine the extent to which Cas9 cutting activity is dictated by genetic or epigenetic factors.

Of the eight gRNAs tested in human cells, only one had significant off-target activity. The AAVS1 gRNA appeared to be particularly promiscuous, cutting sites with three base mismatches. Overall, our data show that gRNAs have a higher cutting efficiency at the on-target site compared to genuine off-target sites. Most potential off-target sites are not cut at a measurable rate at all. Of the 617 and 709 potential off-target sites that were analyzed in human 293FT and hiPS cells, respectively, just five and two were found to be cut and all of these were targeted by just one of eight gRNAs tested. All but one significant off-target cutting took place at sites with three or fewer mismatches compared with the on-target site. In total, 1,254 potential off-target sites with four to five mismatches were successfully tested in 293FT and hiPS cells. Out of these, one (0.08%) showed significant off-target cutting (Supporting Information Table 1). Of the 72 potential off-target sites with ≤3 mismatches tested in 293FT and hiPS cells, six (8.3%) showed significant off-target cutting.

Analysis of mouse ES cells produced broadly similar off-target estimates. In the pooled analysis, three of the gRNAs produced no significant off-target cutting and just one, mATpA, had one off-target site. This site differed from the on-target site by a single nucleotide. Out of the 10 studied potential off-target sites with ≤3 mismatches, off-target Cas9 activity was detected at two sites (one from the pooled analysis and an additional one from the clonal analysis). Echoing the off-target cutting results in human cells, no Cas9 activity was detected in any of the 341 potential off-target sites with four to five mismatches. Taken together, our experiment suggests that in the genomic setting, the CRISPR-Cas9 system has relatively few actual off-target sites in both the human and the mouse genome and that these sites are cut less efficiently than the genuine target site. Off-target cutting appears to be a property of just a few of the designed gRNAs and most gRNAs show no off-target cutting at all.

Interestingly, our results suggest that a simple rule of minimizing the number of genomic sites where a gRNA aligns with ≤4 mismatches may be effective in reducing the total frequency of Cas9-induced off-target cutting.

The indels introduced at repaired DNA double-strand breaks, whether on-target or off-target sites, were non-uniform in size with clear site-specific preferences that were consistent between the two human cell lines. We have previously observed these non-uniform indel patterns in mouse ES cells (Koike-Yusa et al., 2014). These observations suggest that the same DNA double-strand break repair mechanism is operative in these cell types. Further studies will be needed to assess whether these repair patterns can be generalized to other human/mouse cell types. The presence of microhomology at the most preferred indel is consistent with alternative nonhomologous end joining (NHEJ) as the preferred repair mechanism in these cells, as reported previously (Bennardo et al., 2008). Given this repair mechanism, it might be possible to predict deletion patterns on the basis of the target site sequence. Consequently, carefully designed gRNAs might have better chances of producing out-of-frame deletions. Interestingly, although most preferred indels at repair sites involved deletions, there were also preferred peaks of insertions, most notably at the human SLC4 on-target site (Supporting Information Fig. 3).

The simplest way to produce homozygous mutants for either forward genetic screens or functional studies is to sequentially target both alleles of a gene. Two rounds of conventional gene targeting, with two targeting vectors carrying two different selectable markers, are therefore required to generate homozygous mutants. Serial gene targeting is both time-consuming and labor-intensive because individual clones have to be picked and genotyped at every stage. Our success in isolating biallelic targeted clones demonstrates that Cas9-mediated targeting can achieve efficient knockout of both alleles in mouse ES cells. In addition, our results suggest that off-target indels were infrequently generated by the four gRNAs used for gene targeting in mouse ES cells with one exception. While off-target cleavages can occur through imperfect hybridization between the gRNA and the target site, most gRNAs do not cleave off-target sites including closely matched ones. These observations are in line with a recent study that could not find any likely off-target cutting induced by two gRNAs, each of which was assayed in three human pluripotent stem cell clones at the genome-wide level (Veres et al., 2014). This raises an interesting possibility that background somatic indels may in fact far outnumber Cas9 induced off-target indels in cell colonies generated in Cas9 experiments (Veres et al., 2014). Ongoing efforts are showing promising results that engineered variants of gRNA or Cas9 could dramatically reduce the rate of off-target mutations (Cho et al., 2014; Fu et al., 2014). Carefully designed gRNAs, with no off-target sites in exons, are useful genetic tools that allow quick and easy generation of homozygous mutant cells.

MATERIALS AND METHODS

Plasmid Construction

The target gRNA expression constructs were ordered as individual 455-bp gBlocks from IDT and cloned into the pCR-BluntII TOPO vector (Invitrogen) (Mali et al.,
mixed with 500 l.

The targeting vectors for the four mouse gRNAs were designed as follows. The genomic region from each side of the gRNA target site for each gene was PCR amplified from genomic DNA from the JM8 cell line (Pettitt et al., 2009). The length of each homology arm is approximately 600–700 bps. The PCR products were purified with QIAquick PCR Purification Kit (Qiagen, Alameda, CA) before being ligated to a PGK-puromycin-PGKpA construct (see Fig. 3).

Cell Culture and Transfection

Male JM8 mouse ES cells (JM8; Pettitt et al., 2009) were cultured at 37°C with 5% CO2 in M15: Knockout DMEM supplemented with 15% fetal calf serum (FCS), 2 mM l-glutamine, 50 U/ml penicillin, 40 µg/ml streptomycin, 100 µM β-mercaptoethanol, and 1,000 U/ml of recombinant mouse leukemia inhibitory factor (LIF; Millipore). Cell media were changed daily unless otherwise specified. ES cells were cultured on a layer of η-irradiated (60 Gray) SNL76/7 STO feeder fibroblasts (Ramirez-Solis et al., 1993). 293FT cells (Invitrogen) were cultured in DMEM supplemented with 10% FCS and 1% GlutaMax. HiPS cells (S7RE14) were cultured in media as described on a layer of mitotically inactivated MEFs (Yusa, 2013). All cell lines used in this study were negative for mycoplasma contamination.

For potential gRNA off-target cutting analysis, transient transfection of mouse ES cells and 293FT cells were carried out using Lipofectamine LTX (Invitrogen). For transfection of mouse ES cells, 1 µg of hCas9 expression vector, 1 µg of gRNA expression vector, 2.5 µl of PLUS reagent, and 7.5 µl of LTX reagent were mixed with 500 µl of OptiMEM (Invitrogen) media and incubated at room temperature for 5 min. Subsequently, 1 × 10⁶ ES cells suspended in 500 µl of OptiMEM and 1,000 U/ml LIF were mixed with 500 µl of the DNA:-PLUS:LTX mixture and plated directly onto one well of a six-well plate pre-plated with STO fibroblasts. These cells were incubated at 37°C for 1 h whereupon 3 ml of M15 was added directly onto the well. For transfection of 293FT cells, 1 × 10⁶ 293FT cells were plated onto one well of a gelatinized six-well plate 1 day before transfection. The next day, 1 µg of hCas9 expression vector, 1.5 µg of gRNA expression vector, 2.5 µl of PLUS reagent, and 7.5 µl of LTX reagent were mixed with 500 µl of OptiMEM media and incubated at room temperature for 5 min. This DNA:-PLUS:LTX mixture was mixed with 1.5 ml of OptiMEM media and added directly onto the pre-plated adherent cells. Cells were incubated at 37°C for 3 h whereupon the transfection media were replaced with the usual culture media. HiPS cells were transfected with 2.5 µg hCas9 and 2.5 µg of gRNA expression vector using Amaxa Human Stem Cell Nucleofector Kit 2 Mix per manufacturer’s instructions. Approximately 2 × 10⁶ hiPS cells were used per transfection condition. Transfected mouse ES cells were cultured for 3 days, trypsinized, and allowed to separate from STO feeder fibroblasts before genomic DNA extraction. 293FT cells were cultured for 3 days before genomic DNA extraction, whereas transfected hiPS cells were cultured for 4 days before genomic DNA extraction. All transfections were performed in biological duplicates. Transfection efficiencies averaged 80–90% for 293FT cells, 60–70% for mouse ES cells, and 50–60% for hiPS cells based on the fraction of GFP-positive cells observed by fluorescence microscopy. This was determined by parallel transfections of the relevant cells with GFP.

Selection of Potential Off-Target Sites

Guide RNA sequences (20 bp) were mapped to the respective human or mouse target genome (GRCh37 and GRChm38, respectively) using BWA to exhaustively (BWA option “-N”) find alignments with no gaps and with up to five mismatches in the non-PAM region (Li and Durbin, 2009). Subsequently, mapped sites that were followed by the PAM sequence (NGG) were identified as potential off-target sites. These sites were then cross-referenced to repetitive sequences and segmental duplications (UCSC Repeat Masker and Segmental Duplication tables) using BedTools (Quinlan and Hall, 2010). Segmental duplication data were available for the human reference genome GRCh37 and the mouse reference genome GRCm37; the latter was therefore translated into GRCm38 coordinates with the UCSC LiftOver utility. For the eight human gRNAs, a total of 9,264 potential off-target sites with up to five mismatches were identified (Supporting Information Table 2). For the four mouse gRNAs, a total of 3,589 potential off-target sites with up to five mismatches were identified (Supporting Information Table 2). From these potential off-target sites, we manually chose 95 sites per gRNA for off-target cutting analysis. In our choice of off-target sites, we aimed to fulfill the following criteria: an equal representation of sites containing 0–5 mismatches in the non-PAM region, an equal representation of sites containing 0–5 mismatches in the 12 nucleotides adjacent to the PAM sequence (5’ end of the gRNA), an equal representation of sites in exonic (CCDS set) and nonexonic regions, an equal representation of sites containing mismatches in every position along the 20 nucleotide target sequence, and sites that were amenable to PCR amplification (Supporting Information Fig. 1).

PCR and Deep Sequencing to Assess Targeting Specificity

PCR primers for the selected potential off-target sites were designed with Batch Primer3 (You et al., 2008).
Each locus was individually amplified using genomic DNA extracted from the relevant transfected cells. For each independent cell line, the relevant cells transfected with GFP and hCas9 expression cassette were used as controls. PCRs were performed with Q5 Hot-Start HiFi Polymerase (NEB) under the following conditions: 98°C for 30 s, 30 cycles of 98°C for 10 s, 62°C for 15 s and 72°C for 20 s, and 72°C for 2 min. PCR products of all potential off-target loci for each cell type were subsequently pooled and purified with QIAquick PCR Purification Kit (Qiagen).

Five hundred nanograms of the purified PCR products were ligated to Illumina adaptors with NEBNext DNA Library Prep Master Mix (NEB) per manufacturer’s instructions (Quail et al., 2012). Adaptor-ligated products were then subjected to PCR enrichment with KAPA HiFi HotStart ReadyMix under the following PCR conditions: 98°C for 30 s, six cycles of 98°C for 10 s, 66°C for 15 s and 72°C for 20 s, and 72°C for 5 min. The PCR products were purified with Agencourt AMPure XP beads (Beckman) at a PCR product:bead ratio of 1:0.7. The purified libraries were quantified and sequenced using a 250-bp paired-end cycle on the Illumina MiSeq sequencing.

Off-Target Cutting Analysis

A custom reference genome was generated by selecting coordinates 150 bps on either side of each potential off-target gRNA-binding site. Raw, unfiltered, sequencing reads were aligned to this custom reference genome using BLAT (Kent, 2002). Using these alignments, we counted the number of reads mapping to each potential off-target site with or without indels overlapping the predicted cutting position (three nucleotides adjacent to the PAM sequence) ± 3 bps. Reads with indels >100 bps were discarded, while the remaining reads were kept for downstream counting. Read counts were computed separately for each cell line, transfection condition, and biological replicate. Indels were counted independently. For example, if a read contained an insertion and a deletion independently, this read would contribute two indels to the final counts. Standard random sampling tests such as chi-squared test or binomial test only take into account of the variance arising from read sampling but ignore all other experimental variance. We therefore chose to use the beta-binomial model that corrects for overdispersion in the indel rates. Beta-binomial model likelihoods were computed using R and the R library VGAM (R Development Core Team, 2013; Yee, 2010).

An overdispersion parameter $\rho$ was estimated separately for each cell line. As the underlying indel rate of each cell line and transfection condition is assumed to be identical between biological replicates, the variance in indel rates is assumed to be due to the sum of experimental and sampling variance; this is modelled by $\rho$. We estimated the maximum likelihood value $\rho$ numerically for each cell line, and used it to compute the beta-binomial $P$-values for the alternative hypothesis of indel rate_{experiment} > indel rate_{control} at the potential off-target sites. False discovery rates ($q$-values) were computed from the $P$-values using the Benjamini–Hochberg method (Benjamini and Hochberg, 1995).

For analysis of single nucleotide mutations induced by hCas9, we used Samtools mpileup to generate a pileup of each analyzed on-target and potential off-target cutting position ±6 bp. A custom Perl script was written to compute the total counts of each allele at each position in each sample.

Clonal Analysis of Off-Target Sites in Mouse ES Cells

For clonal analysis of potential gRNA off-target sites, 1 × 10^6 mouse ES cells were transfected with 1 μg of hCas9 expression vector, 1 μg of gRNA expression vector, and 1 μg of the uncut corresponding targeting vector with Lipofectamine LTX (Invitrogen) per manufacturer’s instructions. Transfected cells were allowed to recover in M15 for 3 days before being plated for puromycin selection at low density on 10-cm dishes pre-plated with STO fibroblasts. Eight days later, fully formed ES cell colonies were picked. These colonies were grown for two generations on gelatinized 96-well plates, which removes STO fibroblast contamination, before genomic DNA extraction. Subsequently, clones were genotyped to identify clones with targeted modification (hereafter, targeted clones). For each gRNA, 15 clones were selected from the available targeted clones and subjected to downstream off-target analysis.

For each gRNA and individual clone, one on-target and 19 potential off-target sites were analyzed. The 19 off-target sites were chosen from the 95 potential off-target sites of the pooled experiment and comprised the top 19 potentially significant off-target sites from the first data analysis of the pooled experiment. PCR products of all 20 sites analyzed for each individual clone were pooled and 250-bp paired-end Illumina MiSeq sequencing was performed. Read alignment, indel counting, and statistical analyses were performed as described for the pooled experiment. However, $\rho$ was not estimated. Instead, the $\rho$ estimated for mouse ES cells in the pooled analysis was used. False discovery rates were computed as in the pooled analysis.

ACKNOWLEDGMENTS

The authors thank C. Spagnuolo for transfecting some hiPS cells and the Sanger Institute DNA pipeline for sequence analyses.
LITERATURE CITED

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate—A practical and powerful approach to multiple testing. J R Stat Soc Ser B-Methodol 57: 289–300.

Bennardo N, Cheng A, Huang N, Stark JM. 2008. Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. PLoS Genet 4:e1000110.

Cho SW, Kim S, Kim JM, Kim JS. 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31:230–232.

Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, Kim JS. 2014. Analysis of off-target effects of CRISPR-Cas-derived RNA-guided endonucleases and nickases. Genome Res 24:132–141.

Cong L, Ran FA, Cox D, Lin S, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823.

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471:602–607.

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 31:822–826.

Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol 32:279–284.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 31:827–832.

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK. 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31:227–229.

Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31:233–239.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821.

Joung JK, Sander JD. 2013. TALENs: A widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol 14:49–55.

Kent WJ. 2002. BLAT—The BLAST-like alignment tool. Genome Res 12:656–664.

Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera MD, Yusa K. 2014. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat Biotechnol 32: 267–273.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Kooin EV. 2011. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 9:467–477.

Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM. 2013a. Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol 31:833–838.

Mali P, Esvelt KM, Church GM. 2013b. Cas9 as a versatile tool for engineering biology. Nat Methods 10: 957–963.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013c. RNA-guided human genome engineering via Cas9. Science 339:823–826.

Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol 31:839–843.

Pettitt SJ, Liang Q, Rairdan XY, Moran JL, Prosser HM, Beier DR, Lloyd KC, Bradley A, Skarnes WC. 2009. Agouti C57BL/6N embryonic stem cells for mouse genetic resources. Nat Methods 6:493–495.

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152:1173–1183.

Quail MA, Otto TD, Gu Y, Harris SR, Skelly TF, McQuillan JA, Swerklow HP, Oyola SO. 2012. Optimal enzymes for amplifying sequencing libraries. Nat Methods 9:10–11.

Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

R Development Core Team. 2013. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ramirez-Solis R, Davis AC, Bradley A. 1993. Gene targeting in embryonic stem cells. Methods Enzymol 225: 855–878.

Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG, Lee BT, Barber GP, Harte RA, Diekhans M, Long JC, Wilder SP, Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ. 2013. ENCODE data in the UCSC Genome Browser: Year 5 update. Nucleic Acids Res 41(Database issue):D56–D63.
Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJ, Severinov K. 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci USA 108:10098–10103.

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343:84–87.

Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. 2010. Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11:636–646.

Veres A, Gosis BS, Ding Q, Collins R, Ragavendran A, Brand H, Erdin S, Talkowski ME, Musunuru K. 2014. 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 15:27–30.

Wang T, Wei JJ, Sabatini DM, Lander ES. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. Science 343:80–84.

Yee TW. 2010. The VGAM package for categorical data analysis. J Stat Softw 32:1–34.

You FM, Huo N, Gu YQ, Luo MC, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD. 2008. BatchPrimer3: A high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9:253.

Yusa K. 2013. Seamless genome editing in human pluripotent stem cells using custom endonuclease-based gene targeting and the piggyBac transposon. Nat Protoc 8:2061–2078.