Cell Stem Cell
Brief Report

Low Incidence of Off-Target Mutations in Individual CRISPR-Cas9 and TALEN Targeted Human Stem Cell Clones Detected by Whole-Genome Sequencing

Adrian Veres,[1,2] Bridget S. Gosis,[1] Qiurong Ding,[1] Ryan Collins,[3] Ashok Ragavendran,[3] Harrison Brand,[3] Serkan Erdin,[3] Chad A. Cowan,[1,2,5] Michael E. Talkowski,[3,4,5] and Kiran Musunuru[1,2,5,6,*]
[1]Department of Stem Cell and Regenerative Biology, Harvard University and Harvard Stem Cell Institute, Cambridge, MA 02138, USA
[2]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA
[3]Molecular Neurogenetics Unit, Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA
[4]Department of Neurology, Harvard Medical School, Boston, MA 02115, USA
[5]Broad Institute, Cambridge, MA 02142, USA
[6]Division of Cardiovascular Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA
*Correspondence: kiranmusunuru@gmail.com
http://dx.doi.org/10.1016/j.stem.2014.04.020

SUMMARY

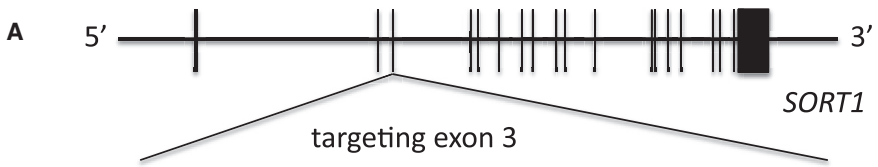
Genome editing has attracted wide interest for the generation of cellular models of disease using human pluripotent stem cells and other cell types. CRISPR-Cas systems and TALENs can target desired genomic sites with high efficiency in human cells, but recent publications have led to concern about the extent to which these tools may cause off-target mutagenic effects that could potentially confound disease-modeling studies. Using CRISPR-Cas9 and TALEN targeted human pluripotent stem cell clones, we performed whole-genome sequencing at high coverage in order to assess the degree of mutagenesis across the entire genome. In both types of clones, we found that off-target mutations attributable to the nucleases were very rare. From this analysis, we suggest that, although some cell types may be at risk for off-target mutations, the incidence of such effects in human pluripotent stem cells may be sufficiently low and thus not a significant concern for disease modeling and other applications.

Clustered regularly interspaced short palindromic repeat (CRISPR)-associated (Cas) systems and transcription activator-like effector nucleases (TALENs) are recently developed genome-editing tools that target desired genomic sites in mammalian cells (Miller et al., 2011; Hockemeyer et al., 2011; Cong et al., 2013; Mali et al., 2013b; Cho et al., 2013; Jinek et al., 2013). The most commonly employed CRISPR-Cas system, derived from *Streptococcus pyogenes*, uses Cas9 nuclease that complexes with a guide RNA that hybridizes a 20 nt DNA sequence (protospacer) immediately preceding an NGG motif (protospacer-associated motif or PAM), resulting in a double-strand break (DSB) 3 bp upstream of the NGG (Jinek et al., 2012). TALENs bind as a pair on sequences surrounding a genomic site, positioning a dimer of FokI nuclease domains in order to generate a DSB at the site. The introduction of a DSB at a specified genomic site allows for modification of the site via either nonhomologous end-joining (NHEJ), which typically introduces an insertion or deletion (indel), or homology-directed repair, which can be exploited to knock in a point mutation or insert a desired sequence at the site.

One important application of genome-editing technology is disease modeling (Musunuru, 2013). The ability to generate isogenic wild-type and mutant clones for phenotypic comparison would enable rigorous functional genetic studies. However, both CRISPR-Cas9 and TALENs have been demonstrated to produce off-target effects; i.e., mutagenesis at sites in the genome other than the desired on-target site (Hockemeyer et al., 2011; Mussolino et al., 2011; Fu et al., 2013; Hsu et al., 2013; Mali et al., 2013a; Pattanayak et al., 2013; Cradick et al., 2013; Cho et al., 2014). These studies have largely focused on sites with high sequence similarity to the on-target site and have documented mutagenesis rates as high as 77% for CRISPR-Cas9 and 1% for TALENs at individual off-target sites. Relatively unexplored is whether CRISPR-Cas9 or TALENs produce off-target effects at sites with low sequence similarity to the on-target site. Although the nucleases might have poor affinity and a low probability of generating a mutation at any given single site in the genome, they may nonetheless generate a sizeable number of nonspecific mutations across the billions of base pairs of the genome in any single cell. This would significantly confound the validity of disease-modeling studies that rely upon genome-edited clones.

To date, most studies of nuclease off-target effects have been performed in aggregated pools of transformed or immortalized cultured human cells such as human embryonic kidney 293T and K562 cells that are not well suited for disease modeling. Therefore, we decided to study nuclease off-target effects generated in a "real-world" application of genome editing centered on human pluripotent stem cell (hPSC) clones being actively used for biological studies (e.g., Ding et al., 2013a).

We assessed the degree of genome-wide off-target mutagenesis in hPSC clones targeted with either CRISPR-Cas9 or TALENs. We performed whole-genome sequencing at high coverage (60× target coverage) of ten cell lines, including nine



TALEN on-target site - chr1:109910022-109910068 (reverse)

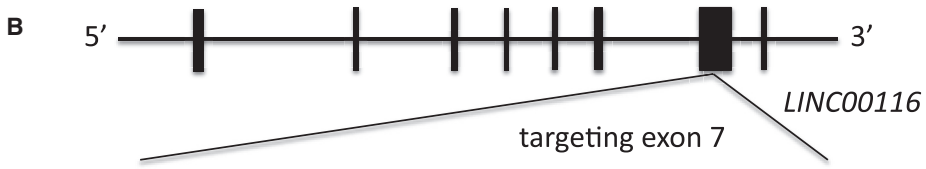
HUES 9	GGTAATTATGACTTTTGGACAGTCCAAGCTATATCGAAGGTGAGATC	wild-type
	CCATTAATACTGAAAACCTGTCAGGTTCGATATAGCTTCCACTCTAG	
clone B	GGTAATTATGACTTTTGGAC----C-AGCTATATCGAAGGTGAGATC	5bp del
	GGTAATTATGACTTTTGGAC----CAAGCTATATCGAAGGTGAGATC	4bp del
clone C	GGTAATTATGACTTTTGGACAGTC-AAGCTATATCGAAGGTGAGATC	1bp del
	GGTAATTATGACTTTTG-----TCCAAGCTATATCGAAGGTGAGATC	5bp del

TALEN off-target site - chr4:126910465-126910511

HUES 9	G**A**TA**C**TT**G**TGACTTTGGTGATACAGAACAAGGAT**T**GAA**T**GTG**TT**ATC	wild-type
	C**T**AT**G**AA**C**ACTGAAACCACTATGTCTTGTTCCTA**A**CTT**A**CAC**AA**TAG	
clone B	GATACTTGTG----------ACAAGGATTGAATGTGTTATC	16bp del

CRISPR on-target site - chr1:109910022-109910068 (reverse)

HUES 9	GGTAATTATGACTTTTGGACAGTCCAAGCTATATCGAAGGTGAGATC	wild-type
	CCATTAATACTGAAAACCTGTCAGGTTCGATATAGCTTCCACTCTAG	
clones E, F	GGTAATTATGACTTTTGGACAGTCCAAGCTATA-CGAAGGTGAGATC	1bp del
	GGTAATTATGACTTTTGGACAGTCCAAGCTATA-CGAAGGTGAGATC	1bp del



CRISPR on-target site - chr2:110970033-110970079 (reverse)

HUES 9	ACTGCAGTTGTCCGTGCTAGTAGCCTTCGCTTCTGGAGTACTCCTGG	wild-type
	TGACGTCAACAGGCACGATCATCGGAAGCGAAGACCTCATGAGGACC	
clone H	ACTGCAGTTGTCCGTGCTAGTAGCCTTCGCcTTCTGGAGTACTCCTGG	1bp ins
	ACTGCAGTTGTCCGTGCTAGTAGCCTTCGC*TTCTGGAGTACTCCTGG	262bp ins
clone I	ACTGCAGTTGTCCGTGCTAGTAGCCTTC-----TGGAGTACTCCTGG	5bp del
	ACTGCAGTTGTCCGTGCTAGTAGCCTTC-----TGGAGTACTCCTGG	5bp del

(legend on next page)

Table 1. Numbers of Unique On-Target and Candidate Off-Target Indels, SVs, and SNVs in TALEN and CRISPR-Cas9 Targeted Clones

	SORT1 TALENs			*SORT1* CRISPR-Cas9			*LINC00116* CRISPR-Cas9		
Clones	A	B	C	D	E	F	G	H	I
On-target indels	—	2	2	—	1[a]	1[a]	—	1	1[a]
On-target SVs	—	—	—	—	—	—	—	1[b]	—
Likely off-target indel	—	1	—	—	—	—	—	—	—
Other candidate off-target indels	2	1	2	4	4	2	3	5	4
Candidate off-target SVs	—	—	—	—	—	1	—	—	—
SNVs	64	115	142	55	94	74	111	127	112

See also Figure S1 and Tables S1–S3.
[a]Homozygous for indels.
[b]261 bp duplicated insertion.

clones we had previously generated with genome editing (Ding et al., 2013a, 2013b) (Figure 1): the human embryonic stem cell line HUES 9: three HUES 9 clones exposed to TALENs targeting the *SORT1* gene with one clone remaining wild-type in both alleles (clone A) and two clones bearing indels in both *SORT1* alleles (clones B and C), three HUES 9 clones exposed to CRISPR-Cas9 targeting the same site in the *SORT1* gene with one wild-type clone (clone D) and two clones bearing indels in both *SORT1* alleles (clones E and F), and three HUES 9 clones exposed to CRISPR-Cas9 targeting the *LINC00116* gene with one wild-type clone (clone G) and two clones bearing indels in both *LINC00116* alleles (clones H and I). All of the HUES 9 clones were derived from the same stock of parental HUES 9 cells. Notably, we had found the targeting efficiency of the *SORT1* TALENs to be 11%, in contrast to CRISPR-Cas9 for *SORT1*, which was 76%, The targeting efficiency of CRISPR-Cas9 for *LINC00116* was 57% (Ding et al., 2013b).

Upon obtaining the whole-genome sequencing data, we assessed the clones for small indels, single-nucleotide variants (SNVs), and structural variants (SVs), which include chromosomal inversions, rearrangements, duplications, and deletions (Supplemental Experimental Procedures available online). We largely focused on the identification of small indels and SVs because they comprise virtually all of the mutations introduced by NHEJ. After filtering for the small indels most likely to be true positives and to be potential off-target mutations (rather than mutations that arose in the parental cell pool) and confirmation with Sanger sequencing, we identified a total of 28 such indels across the nine experimental clones in comparison to the parental HUES 9 cells as the reference. Notably, all of the previously known on-target indels (seven in total) were correctly identified by the whole-genome sequencing and filtering (Tables 1 and S1). One of the 28 off-target indels was a frameshift in the coding sequence of *ZDHHC11* (in clone I). None of the other indels lay in either the coding sequence of a gene or the expressed sequence of an annotated noncoding RNA.

None of the indels in CRISPR-Cas9 clones were within 100 nt of a potential off-target site as predicted by sequence similarity—up to six mismatches—with the on-target site, and none lay near sequences that matched the on-target sites better than would be expected by chance (Figure S1). Moreover, none of the indels lay within 100 nt of a sequence perfectly matching the last 10 nt of the protospacer with an adjacent PAM site (NGG as well as NAG, which has also been shown to be tolerated) (Hsu et al., 2013; Pattanayak et al., 2013). Furthermore, we paid special attention to the indels that lay within 5 bp upstream of a potential PAM site (Table S1), where CRISPR-Cas9-mediated DSBs would be expected to occur. Although the majority of indels had a potential PAM site, none of the adjacent sequences matched the on-target site better than would be expected by chance (Figure S1).

One of the indels in a TALEN clone was located between two potential off-target binding sites as predicted by sequence similarity with the on-target sites—one with three mismatches and another with four mismatches—with the binding sites being 17 bp apart, within the optimal range for generating a DSB with TALENs of this type (Ding et al., 2013a) (Figure 1A). None of the other TALEN clone indels were optimally positioned near a pair of degenerate TALEN binding sites (up to five mismatches with the on-target site), and none lay near sequences that matched the on-target sites better than would be expected by chance (Figure S1).

None of the SVs and SNVs that passed our filtering criteria in CRISPR-Cas9 clones was within 100 nt of a predicted off-target site. None of the variants in TALEN clones were optimally positioned near a pair of degenerate TALEN binding sites. We detected 894 unique SNVs across the nine clones (average of 100 per clone) in comparison to the parental HUES 9 cell line (Table 1). The SV analysis revealed two structural variants unique to an individual clone: a 5.5 kb deletion on chromosome 6 in clone F and a 261 bp segment of chromosome 4 inserted within the *LINC00116* CRISPR-Cas9 on-target site on chromosome 2 in clone H (Tables 1 and S2). Sanger sequencing confirmed that both alleles of the chromosome 4 region were intact in clone H, signifying a duplicated insertion into the chromosome 2 on-target site rather than a balanced translocation. We speculate that, because of microhomology, the chromosome 4 region was used as a repair template for a DSB at the on-target site.

Just one of the detected variants—a TALEN clone indel—seems certain to be a nuclease-mediated off-target effect. It is probable that some, if not all, of the other indels and SVs reflect clonal heterogeneity within the original stock of HUES 9 cells. Previous studies have documented mutagenesis occurring during the derivation and expansion of hPSCs (Hussein et al., 2011; Gore et al., 2011; Howden et al., 2011; Yusa et al., 2011).

Figure 1. On- and Off-Target Mutations

(A) HUES 9 clones targeted in the *SORT1* gene with TALENs or CRISPR-Cas9.

(B) HUES 9 clones targeted in the *LINC00116* gene with CRISPR-Cas9. For TALEN targeted clones, the boxes indicate the TALEN on-target and off-target binding sequences. For CRISPR-Cas9 targeted clones, the boxes indicate the 20 bp sequence matching the protospacer and the 3 bp PAM. For the on-target sites, deletions and insertions in the two alleles of each clone are indicated. For the off-target site, the mismatches with the TALEN on-target binding sequences are indicated in bold, and the deletion in one allele of the clone is indicated.

Furthermore, each clone harbored a sizable number of unique SNVs, which would not be predicted to result from NHEJ. Nonetheless, with a maximum of just two to five confirmed events in each individual clone, our results suggest that nuclease-mediated off-target effects of CRISPR-Cas9 and TALENs do not intrinsically cause a large degree of indiscriminate nonspecific mutagenesis across the genome.

We note the limitations of this study. Even with whole-genome sequencing at high coverage, it is likely that some variants in the clones were not detected given the limitations of short-read sequencing. The small number of sequenced clones targeted at just two loci prevents generalization to all hPSC clones targeted with any CRISPR-Cas9 or TALENs of any configuration by any methodology. Furthermore, our results are not relevant to therapeutic applications targeting up to millions of cells at a time, where rare events may have deleterious consequences.

We do note that clonal heterogeneity may represent a more serious obstacle to the generation of truly isogenic cell lines than nuclease-mediated off-target effects, given that each of our clones harbored a very small number of unique indels and SVs (two to five) in comparison to a relatively larger number of unique SNVs (average of 100) that likely arose spontaneously in culture. This suggests that even if one had a genome-editing tool in hand with perfect specificity, targeted clones would still be likely to harbor some differences elsewhere in the genome. Rigorous studies will require whole-genome sequencing of the clones used for experiments to fully characterize their mutational profiles, or they will need to include multiple clones for each experimental condition to ensure that potential confounding by any single mutation in a clone is minimized.

ACCESSION NUMBERS

The NCBI Sequence Read Archive accession number for the whole-genome sequencing data reported in this paper is SRP039576.

SUPPLEMENTAL INFORMATION

Supplemental Information contains Supplemental Experimental Procedures, one figure, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2014.04.020.

ACKNOWLEDGMENTS

This work was supported in part by the Harvard Presidential Scholars Fund of the Harvard Medical School MD/PhD Program (A.V.); grants R00-HL098364 (K.M.), R01-HL118744 (K.M.), R01-DK097768 (K.M. and C.A.C.), U01-HL100408 (C.A.C.), and R00-MH095867 (M.E.T.) from the National Institutes of Health (NIH); the Harvard Stem Cell Institute (K.M.); and Harvard University (Q.D. and K.M.). We thank Vamsee Pillalamarri, Carrie Hanscom, and the staff of the MGH Genomics and Technology Core and NextGen Sequencing Core for technical assistance.

Received: January 11, 2014
Revised: March 13, 2014
Accepted: April 24, 2014
Published: July 3, 2014

REFERENCES

Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. *31*, 230–232.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. *24*, 132–141.

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.

Cradick, T.J., Fine, E.J., Antico, C.J., and Bao, G. (2013). CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. Nucleic Acids Res. *41*, 9584–9592.

Ding, Q., Lee, Y.K., Schaefer, E.A., Peters, D.T., Veres, A., Kim, K., Kuperwasser, N., Motola, D.L., Meissner, T.B., Hendriks, W.T., et al. (2013a). A TALEN genome-editing system for generating human stem cell-based disease models. Cell Stem Cell *12*, 238–251.

Ding, Q., Regan, S.N., Xia, Y., Oostrom, L.A., Cowan, C.A., and Musunuru, K. (2013b). Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. Cell Stem Cell *12*, 393–394.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. *31*, 822–826.

Gore, A., Li, Z., Fung, H.L., Young, J.E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M.A., Kiskinis, E., et al. (2011). Somatic coding mutations in human induced pluripotent stem cells. Nature *471*, 63–67.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C., et al. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nat. Biotechnol. *29*, 731–734.

Howden, S.E., Gore, A., Li, Z., Fung, H.L., Nisler, B.S., Nie, J., Chen, G., McIntosh, B.E., Gulbranson, D.R., Diol, N.R., et al. (2011). Genetic correction and analysis of induced pluripotent stem cells from a patient with gyrate atrophy. Proc. Natl. Acad. Sci. USA *108*, 6537–6542.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

Hussein, S.M., Batada, N.N., Vuoristo, S., Ching, R.W., Autio, R., Närvä, E., Ng, S., Sourour, M., Hämäläinen, R., Olsson, C., et al. (2011). Copy number variation and selection during reprogramming to pluripotency. Nature *471*, 58–62.

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

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. Elife *2*, e00471.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. *31*, 833–838.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013b). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

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.

Musunuru, K. (2013). Genome editing of human pluripotent stem cells to generate human cellular disease models. Dis. Model. Mech. *6*, 896–904.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. *31*, 839–843.

Yusa, K., Rashid, S.T., Strick-Marchand, H., Varela, I., Liu, P.Q., Paschon, D.E., Miranda, E., Ordóñez, A., Hannan, N.R., Rouhani, F.J., et al. (2011). Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. Nature *478*, 391–394.