Template plasmid integration in germline genome-edited cattle.

Alexis L. Norris[1*], Stella S. Lee[1*], Kevin J. Greenlees[1], Daniel A. Tadesse[1], Mayumi F. Miller[1], Heather Lombardi[1†]

[1] Center for Veterinary Medicine, Food and Drug Administration, Rockville, MD 20855

[*] Contributed equally

[†] Corresponding author: Heather.Lombardi@fda.hhs.gov

Author contributions:

A.L.N. and S.S.L. developed the strategy; A.L.N. conducted the experiments; A.L.N., S.S.L., and D.A.T. analyzed the data; and A.L.N, S.S.L, K.J.G., D.A.T., M.F.M., and H.L. wrote the manuscript.

Competing Financial Interests:

The authors declare no competing financial interests.

Acknowledgments:

We acknowledge Recombinetics, Inc. for generating the animals and sharing the sequencing data through NCBI SRA. We appreciate Drs. Alison Van Eenennaam, C. Titus Brown, and Tamer Mansour for helpful discussions concerning our discovery of the template plasmid integration. We thank Drs. Mike Mikhailov and Fu-Jyh Luo for technical expertise, as well as Drs. Harlan Howard, Andrew Fidler, and Kelly Underwood for veterinary expertise.

Abstract

We analyzed publicly available whole genome sequencing data from cattle which were germline genome-edited to introduce polledness. Our analysis discovered the unintended heterozygous integration of the plasmid and a second copy of the repair template sequence, at the target site. Our finding underscores the importance of employing screening methods suited to reliably detect the unintended integration of plasmids and multiple template copies.

Article

As genome editing technology evolves, so does our understanding of the unintended alterations it produces, both in form and frequency. Several sequencing-based methods have been developed to screen for off-target errors (GUIDE-Seq[1], SITE-Seq[2], CIRCLE-Seq[3], DISCOVER-Seq[4]), and long-read sequencing of the target site can be used to detect on-target errors. Each screening approach carries assumptions and biases that may allow alterations of unexpected types to go undetected. Recent examples of previously unexpected alterations are complex genomic rearrangements at or near the target site in mammalian genome editing experiments[5, 6]. The complex rearrangements included insertions, deletions, inversions, and translocations that were difficult to detect by standard PCR and DNA sequencing methods.

In this study, we analyzed the target site of publicly-available whole genome sequencing data[7] from genome-edited calves to confirm the intended edit and to screen for potentially undetected on-target errors (**Supplementary Methods**). The calves were genome-edited[7], using transcription activator-like effector nucleases (TALENs) and a repair template for homology-directed repair (HDR)[8], to introduce the Celtic polled allele (P_c), a variant that produces the hornlessness (polled) trait in cattle. The P_c variant, common in some cattle breeds[9], is a 212-bp duplication in place of a 10-bp sequence in an intergenic region on chromosome 1 (chr1:2,429,000-2,429,500; bosTau9). The variant follows an autosomal dominant inheritance, but the mechanism underlying the association with polled trait is unknown.

Given that the repair template plasmid was delivered in the pCR2.1-TOPO plasmid (**Fig. 1a**), we included the plasmid backbone sequence in our comparison of the sequencing reads with the bovine reference genome. In our analysis, we discovered the presence of the full-length plasmid backbone in both genome-edited calves (**Supplemental Fig. 1**). While one allele contained the intended edit, identical to the naturally-occurring P_c variant (**Fig. 1b**),

close inspection revealed integration of the plasmid and a second copy of the repair template sequence at the target site in the other allele in both calves (**Fig. 1c**). The plasmid-containing allele (denoted P_c*) was found to have been inserted continuous with the template ends, producing a duplication of the template and two novel bovine-plasmid junctions (**Supplemental Fig. 2**). No off-target insertions of the plasmid or the template were detected.

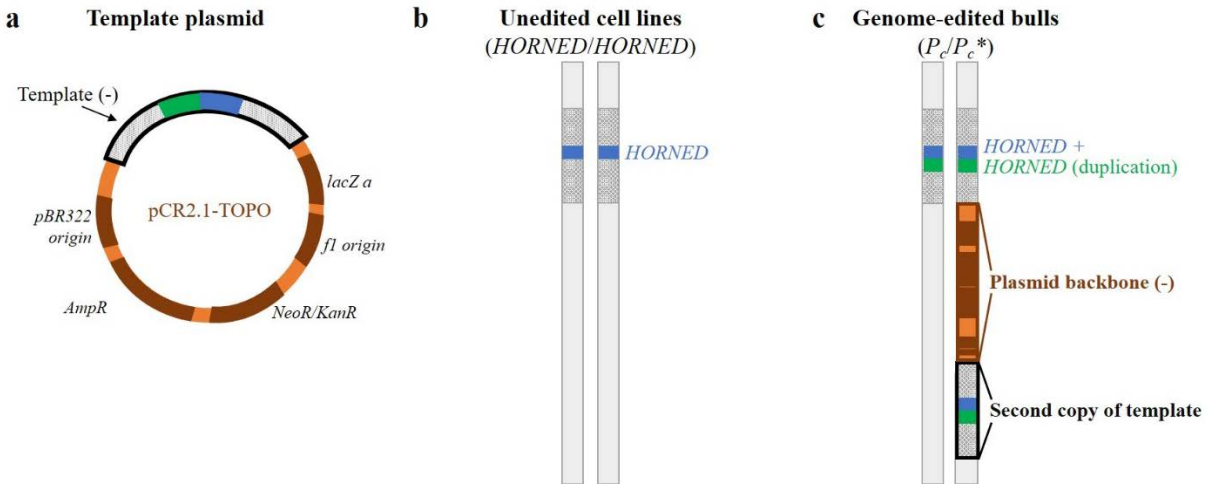
Previously, the template plasmid integration was not detected[7]. Probable reasons include: the plasmid backbone was not included in the sequence alignment, elevated noise at the target locus, limited signal of the sequencing data, and PCR conditions insensitive to detect the integrations. The noise was elevated due to the complex sequence context that obscured the integration: (1) the P_c variant itself is a duplication of the reference sequence (*HORNED* allele) in place of a 10-bp sequence, and (2) the target locus is highly repetitive, potentially masking rearrangements. The signal is limited by the sequencing depth of 20 reads for each DNA base, on average. In an ideal scenario, heterozygosity would result in 10 reads identifying the plasmid insertion but given that the plasmid DNA sequence is not in the reference genome, the plasmid reads would remain unmapped. The template plasmid integration was not detected by PCR genotyping[7] due to the following: (1) the expected PCR amplicons, correctly sized for the P_c variant (212-bp duplication in place of a 10-bp sequence), were produced, (2) the primers were not designed to amplify the plasmid, (3) the amplicons produced by the template plasmid integration were prohibitively large, and (4) the qualitative nature of the assay was insensitive to the increased number of template copies (**Supplementary Methods**, **Supplemental Fig. 3, and Supplemental Table 1**).

Next, we performed a literature search of template plasmid integration at the target site in genome editing experiments, to determine the prevalence of this class of unintended alterations. We found that while there are reports, the template plasmid integration is often not a major finding, and thus we suspect that the integration errors are under reported or

78 overlooked. Template plasmid integration events are known to occur with zinc finger
79 nucleases (ZFNs) at both the target and off-target sites[10, 11]. Using HEK-293 cells, Olsen, *et*
80 *al.*[10] showed that transfection of plasmid alone resulted in plasmid integration at a rate of
81 $28x10^{-5}$ per cell; the addition of one ZFN increased the frequency to $55x10^{-5}$ per cell and
82 two ZFNs further increased the frequency to $99x10^{-5}$ per cell. Work by Dickinson *et al.*[12],
83 using CRISPR/Cas9 with a template plasmid in *C. elegans*, reported the integration of a
84 second copy of the template at the target site. Additional publications using CRISPR/Cas9
85 with double stranded DNA (dsDNA) repair templates showed that the dsDNA templates can
86 form multimers that integrate into the target site in fish[13] and mice[14].

87 Our discovery highlights a potential blind spot in standard genome editing screening
88 methods. In light of our finding, we propose modifications to current screening methods to
89 enable detection of plasmid integration and integration of multiple template copies. The
90 alignment of sequencing data should include both the reference genome and plasmid
91 sequences. PCR genotyping should incorporate plasmid-specific primers. Methods to detect
92 increased copies of the template and unintended integration of the template plasmid include
93 long-range PCR conditions, quantitative PCR (e.g., digital droplet PCR), Southern blot, and
94 long-read sequencing (e.g., Nanopore, PacBio). The application of suitable screening
95 methods will provide a more precise measure of the prevalence of template plasmid
96 integration events and drive improvements to genome editing, to the benefit of the field.

97 **Figures**



98

99 **Fig. 1: Template plasmid integration at the target site of genome-edited calves.**

100 Genomic structure of the template plasmid (**a**), unedited parental cell lines (**b**), and the
101 genome-edited calves (**c**). (**a**) The repair template, containing the P_c sequence and flanking
102 homology arms, is inserted in the pCR2.1 plasmid in an antisense orientation at the TOPO
103 cloning site. (**b**) The unedited parental cell lines are homozygous for *HORNED*. (**a**) The
104 genome-edited calves are heterozygous: one chromosome contains the intended edit (P_c),
105 while the other chromosome harbors template plasmid integration, in addition to the
106 intended edit.

References

1. Tsai, S.Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33**, 187-197 (2015).
2. Cameron, P. et al. Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat Methods* **14**, 600-606 (2017).
3. Tsai, S.Q. et al. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat Methods* **14**, 607-614 (2017).
4. Wienert, B. et al. Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science* **364**, 286-289 (2019).
5. Shin, H.Y. et al. CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat Commun* **8**, 15464 (2017).
6. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* **36**, 765-771 (2018).
7. Carlson, D.F. et al. Production of hornless dairy cattle from genome-edited cell lines. *Nat Biotechnol* **34**, 479-481 (2016).
8. Tan, W. et al. Efficient nonmeiotic allele introgression in livestock using custom endonucleases. *Proc Natl Acad Sci U S A* **110**, 16526-16531 (2013).
9. Medugorac, I. et al. Bovine polledness--an autosomal dominant trait with allelic heterogeneity. *PLoS One* **7**, e39477 (2012).
10. Olsen, P.A., Gelazauskaite, M., Randol, M. & Krauss, S. Analysis of illegitimate genomic integration mediated by zinc-finger nucleases: implications for specificity of targeted gene correction. *BMC Mol Biol* **11**, 35 (2010).
11. Radecke, S., Radecke, F., Cathomen, T. & Schwarz, K. Zinc-finger nuclease-induced gene repair with oligodeoxynucleotides: wanted and unwanted target locus modifications. *Mol Ther* **18**, 743-753 (2010).
12. Dickinson, D.J., Ward, J.D., Reiner, D.J. & Goldstein, B. Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination. *Nat Methods* **10**, 1028-1034 (2013).
13. Gutierrez-Triana, J.A. et al. Efficient single-copy HDR by 5' modified long dsDNA donors. *Elife* **7** (2018).
14. Skryabin, B.V., Gubar, L., Seeger, B., Kaiser, H., Stegemann, A., Roth, J., Meuth, S. G., Pavenstädt, H., Sherwood, J., Pap, T., Wedlich-Söldner, R., Sunderkötter, C., Schwartz, Y.B., Brosius, J., Rozhdestvensky, T. S. Pervasive head-to-tail insertions of DNA templates mask desired CRISPR/Cas9-mediated genome editing events. *Preprint at:* https://www.biorxiv.org/content/10.1101/570739v1 (2019).