BRIEF REPORT

Sequencing data from Massachusetts General Hospital shows Cas9 integration into the genome, highlighting a serious hazard in gene-editing therapeutics [version 1; peer review: 1 approved with reservations]

Sandeep Chakraborty

Celia Engineers, Mumbai, India

Abstract
The ability to edit a specific gene within our genomes using guided-nucleases (Cas9/ZFN/TALEN - CaZiTa) presents huge opportunities for curing many genetic disorders. Delivery of this ‘drug’ within cells is a critical step for such therapies. The ability of recombinant adeno-associated virus (rAAV) to enter cells makes it a perfect choice as a vector for gene therapy. A plasmid comprising the rAAV, the CaZiTa, guide RNAs (for CRISPR) is expected to enter the cell, edit the target gene(s), remain episomal, and thus fade away with time. However, the rather obvious danger of integration of the plasmid into the genome, if the episomal hypothesis is incorrect, is under-reported. A recent report has highlighted that bacterial genes from a plasmid were integrated into bovine genomes. Massachusetts General Hospital has recently published data on CRISPR edits (Accid:PRJNA563918), noting ‘high levels of AAV integration (up to 47%) into Cas9-induced double-strand breaks’. However, there is no mention of Cas9 integration. Here, the same data from Massachusetts General Hospital shows Cas9 integration in the exact edit sites provided for two genes - TMC1 and DMD. Also, there is a mis-annotation of one sample as ‘no gRNA’, since Cas9 integrations have been detected in that sample. This is an important distinction between AAV and CaZiTa integration: while AAV integration can be tolerated, Cas9 integration is a huge, and unacceptable, danger.

Keywords
Gene-editing, CRISPR-Cas, AAV, plasmid integration
Introduction

Nuclease based gene-editing techniques (Cas9/ZFN/TALEN - CaZiTa) introduce a double stranded break at a specified location with the guide of DNA-binding proteins (ZFN/TALEN) or RNA (CRISPR). Delivery within cells is a critical step for such gene-therapies. The ability of recombinant adeno-associated virus (rAAV) to enter cells makes it a perfect choice as a vector for gene therapy. A plasmid comprising the rAAV, CaZiTa, guide RNAs (for CRISPR), etc. is expected to enter the cell, edit, and remain episomal. However, the rather obvious danger of integration of the plasmid into the genome is ignored. A recent pre-print has highlighted that bacterial genes from the template plasmid (pCR2.1-TOPO) has integrated into bovine genomes.

Recently, Massachusetts General Hospital published data on CRISPR edits (Accid:PRJNA563918), and concluded that ‘AAV integration should be recognised as a common outcome for applications that utilize AAV for genome editing’. However, there was no report of Cas9 integration. Here, data showing Cas9 integrating in the exact edit sites is shown. The same caveat applies to all three CaZiTa gene-therapies.

Methods

Sequencing data was obtained from BioProject accession number, PRJNA563918, using SRA download tools (https://github.com/ncbi/sra-tools/wiki). This project aims at finding “the genomic consequences of transduction with AAV vectors encoding CRISPR-Cas nucleases”.

The reads were uniquified, split into open-reading frames using the getorf (version:6.6.0.0) program from the EMBOSS suite (http://emboss.sourceforge.net/download/).

Command: “getorf -find 1 -sequence <infile> -outseq <output> -minsize 10”.

Exact 10-kmer were aligned to the Cas9 protein (Accid:AGZO1981.1, 1417aa). Subsequently, these reads were then aligned to the gene of interest - TMC1 (Accid:NG_008213.1) or DMD (Accid:NM_00134034.1). The online BLAST interface using the default settings (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to generate images of integration with the genes of interest. The number of reads with integration is conservative, since reads with less than a 10-kmer match were ignored. The AAV genome used was Accid:MK163936.1. Sequences that integrate Cas9/AAV with TMC1 and DMD are provided in Extended data: TMC1.cas9.fa/TMC1.aav.fa and DMD.cas9.fa/DMD.aav.fa, respectively.

Results and discussion

Cas9 integration in transmembrane channel like 1 (TMC1)

TMC1, a transmembrane protein, is required for proper functioning of cochlear hair cells, and has been implicated in hearing loss and prelingual deafness. In mice, this gene is located in chr19 (Accid:NG_008213.1)). Table 1 lists the samples sequenced for targeting this gene. As expected, the non-injected controls have no Cas9 reads. The sample SRR10068671 (marked with triple asterisks) has probably been mis-annotated as one with “no gRNA”, since Cas9 integrations are detected in that sample. From the Cas9 integration site, the guide RNA (gRNA) for the TMC1 gene can be deduced to be “CATGGAATATGTCCTCCTGTGGGA“, although this information is not yet available. The sequences for these reads are provided as extended data. As a specific example, the sequence (SRR10068639.63180.1) encodes the 26aa peptide=SPEKLLMYHHPQTYQKLKLIMEQYG from Cas9 and is merged with the TMC1 gene (Figure 1).

Cas9 integration in dystrophin

Duchenne muscular dystrophy (DMD), a genetic disorder associated with progressive muscle degeneration (dystrophy), is caused by aberrations in the dystrophin protein, located on chromosome X in humans and mice. According to the data (Accid:PRJNA563918), different exons and introns have been targeted using two different variations of the Cas9 - SpCas9 and SaCas9 - which differ in certain characteristics. SaCas9 does seem empirically to have more samples without any integration, but that might just be random. The sequences for these reads are provided as extended data. As a specific example, the sequence (SRR10068622.33932.1) encodes the 12aa peptide=LDAHLSIQTG from Cas9 and is merged with the DMD genome (Figure 2).

AAV integration in TMC1 and DMD

The integration of AAV into the genome has already been noted. This has been replicated in this study as well. The AAV genome used was Accid:MK163936.1. The reads can be found in Extended data: TMC1.aav.fa (N=5000) and DMD.aav.fa (N=14000).

Conclusions

Gene-editing based therapies provide revolutionary hope for curing many diseases. However, ensuring safety of any such endeavours must be paramount to avoid doing more harm. Plasmid integration is one such potential hazard. Recently, Massachusetts General Hospital reported high-levels of in vivo AAV into the genome while providing sequencing data (Accid:PRJNA563918), and concluded that ‘AAV integration should be recognized as a common outcome for applications that utilize AAV for genome editing’. In this study, in addition to AAV, Cas9 integration is shown in the same samples. The same caveat applies to gene-therapies using CaZiTa. Off-target edits (OTE) are an important aspect of CRISPR-cas gene-editing. Such integrations, found by targeting amplification sequencing, will only get worse due to OTEs, which are hard to detect, although this information is not yet available. The sequences for these reads are provided as extended data. As a specific example, the sequence (SRR10068639.63180.1) encodes the 26aa peptide=SPEKLLMYHHPQTYQKLKLIMEQYG from Cas9 and is merged with the TMC1 gene (Figure 1).
Table 1. Integration of Cas9 into the edited gene. The paired reads are clubbed into a single file, and then uniquified. Open reading frames from these are then matched to Cas9 proteins for 10 aa, and subsequently these reads are matched to the gene of interest (TMC1 or DMD), helping identify the exact edit site. The sample SRR10068671 (marked with triple asterisks) has probably been mis-annotated as one with “no gRNA”, since Cas9 integrations are detected in that sample. These integrations happen both for SpCas9 and SaCas9. From the Cas9 integration site, the gRNA for the TMC1 gene can be deduced to be “CATGGTAATGTCCCTCCTG”, although this information is not yet available. CN: cortical neuron; CA: cochlea, adult.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Accid</th>
<th>Description</th>
<th>NREADS</th>
<th>NUNIQ</th>
<th>NCas9</th>
<th>NTarget</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRR10068641</td>
<td>In vivo CA non injected control</td>
<td>140156</td>
<td>12184</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>SRR10068642</td>
<td>In vivo CA non injected control</td>
<td>160880</td>
<td>9511</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>SRR10068670</td>
<td>CN neonatal no gRNA</td>
<td>164390</td>
<td>7079</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>SRR10068671</td>
<td>CN neonatal no gRNA</td>
<td>259600</td>
<td>10794</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>SRR10068648</td>
<td>CN neonatal 2</td>
<td>243658</td>
<td>16919</td>
<td>90</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>TMC1</td>
<td>SRR10068659</td>
<td>CN neonatal 1</td>
<td>220840</td>
<td>14033</td>
<td>108</td>
<td>98</td>
</tr>
<tr>
<td>SRR10068666</td>
<td>CN neonatal 2</td>
<td>230024</td>
<td>20975</td>
<td>772</td>
<td>682</td>
<td></td>
</tr>
<tr>
<td>SRR10068667</td>
<td>CN neonatal 1</td>
<td>208190</td>
<td>17984</td>
<td>598</td>
<td>546</td>
<td></td>
</tr>
<tr>
<td>SRR10068636</td>
<td>In vivo CA 4</td>
<td>187904</td>
<td>16498</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>SRR10068638</td>
<td>In vivo CA 3</td>
<td>117948</td>
<td>7866</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>SRR10068639</td>
<td>In vivo CA 2</td>
<td>222014</td>
<td>15268</td>
<td>10</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>SRR10068640</td>
<td>In vivo CA 1</td>
<td>136336</td>
<td>9321</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>SRR10068621</td>
<td>In vivo exon 53</td>
<td>324432</td>
<td>13477</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>SRR10068622</td>
<td>In vivo exon 53</td>
<td>285824</td>
<td>12335</td>
<td>7</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>SRR10068623</td>
<td>In vivo intron51-53</td>
<td>382044</td>
<td>18977</td>
<td>312</td>
<td>238</td>
<td></td>
</tr>
<tr>
<td>SRR10068624</td>
<td>In vivo intron51-53</td>
<td>370728</td>
<td>19000</td>
<td>266</td>
<td>238</td>
<td></td>
</tr>
<tr>
<td>SRR10068625</td>
<td>In vivo intron 53</td>
<td>289552</td>
<td>10673</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>DMD</td>
<td>SRR10068627</td>
<td>In vivo intron 53</td>
<td>253460</td>
<td>9596</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>SpCas9</td>
<td>SRR10068628</td>
<td>In vivo intron 51</td>
<td>388128</td>
<td>16116</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>SRR10068629</td>
<td>In vivo intron 51</td>
<td>267736</td>
<td>10710</td>
<td>7</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>SRR10068630</td>
<td>In vivo intron51-53</td>
<td>336368</td>
<td>12750</td>
<td>48</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>SRR10068631</td>
<td>In vivo intron51-53</td>
<td>472724</td>
<td>20169</td>
<td>48</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>DMD</td>
<td>SRR10068632</td>
<td>In vivo intron 53</td>
<td>221456</td>
<td>7786</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SaCas9</td>
<td>SRR10068633</td>
<td>In vivo intron 53</td>
<td>237768</td>
<td>8136</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SRR10068634</td>
<td>In vivo intron 51</td>
<td>428600</td>
<td>17698</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Specific example of a sequence in the TMC1 gene which has Cas9 integrated. The sequence (SRR10068639.63180.1) encodes the 26aa peptide=SPEKLLMYHIDPQTYQKLKIMEQYG from Cas9 and is merged with the TMC1 gene.
danger. After integration, individuals may be expressing Cas9, and that is fraught with genotoxic danger.

Data availability

Raw sequence reads from Massachusetts General Hospital, Accession number PRJNA563918, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA563918/

Extended data

Zenodo: Sequencing data from Massachusetts General Hospital shows Cas9 integration into the genome, highlighting a serious hazard in gene-editing therapeutics, https://doi.org/10.5281/zenodo.3460305

This project contains the following extended data:
- DMD.aav.fa
- DMD.cas9.fa
- TMC1.aav.fa
- TMC1.cas9.fa

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Figure 2. Specific example of a sequence in the DMD gene which has Cas9 integrated. The sequence (SRR10068622.33932.1) encodes the 12aa peptide=LDATLIHQSITG from Cas9 and is merged with the DMD gene.

References


18. Chakraborty S: Inconclusive studies on possible CRISPR-Cas off-targets should moderate expectations about enzymes that have evolved to be non-specific. J Biosci. 2018; 43(2): 225–228. Published Abstract | Publisher Full Text


Open Peer Review

Current Peer Review Status:  

Review Version 1

Reviewer Report 06 February 2020

https://doi.org/10.5256/f1000research.22812.r59005

© 2020 Chatrikhi R. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rakesh Chatrikhi

Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Chakraborty studied sequencing data on CRISPR-induced DNA edits from Massachusetts General Hospital to show Cas9 integration into the genome. The author analyzed previously published data on CRISPR-induced DNA edits to highlight integration of Cas9 into the genome, which was not reported in the previously published study that generated the sequencing data. This study highlights risks involved in using gene-editing techniques such as CRISPR-Cas9 based system and raises important caveats that need to be considered when using the technology for therapeutics. However, the author needs to address the following points to strengthen the manuscript:

1. The author needs to quantitatively compare the levels of AAV and Cas9 integration and whether it is statistically significant. This is important because the author is using the same samples and sequencing data to discuss Cas9 integration into the genome, which was not reported earlier and thus required to effectively highlight the results in this study. The author needs to make a clear comparison of AAV and Cas9 integration from the data in the Results and Discussion section.

2. The author needs to support the argument that Cas9 integration is a serious hazard and is a huge unacceptable danger. This is important because the author argues that Cas9 integration is hazardous throughout the manuscript (title, abstract and conclusion). The author needs to include multiple references to previous work in the literature that have shown or highlighted the toxicity of Cas9 integration into the genome and that it is hazardous.

3. In the Conclusions section, the author needs to briefly discuss the alternative approaches that could be used to avoid Cas9 integration. For example, the author needs to elaborate on the use of purified Cas9 ribonucleoproteins as a substitute to plasmids. The author included a sentence on this topic in the Conclusions section, but need to elaborate on the advantages of using Cas9 ribonucleoproteins. For example, in addition to avoiding Cas9 integration, this method has higher efficiency (Farboud B et al., J Vis Exp, 2018), fewer off-targets (Liang X et
al., Journal of Biotech, 2015; Kim S et al., Genome Res, 2014 - Ref 25 in this study)\textsuperscript{1,2}.

4. The author needs to include the rationale for focusing on the genes TMC1 and DMD. Were those genes studied in the previous study that generated the data (Ref 7)? This needs to be clearly stated. If there are more genes studied in Ref 7, then the author needs to include the rationale of choosing these two genes over other genes in the study.

5. The Introduction section could also be strengthened by briefly discussing if there are any similar studies carried out previously in the literature studying the level of Cas9 integration into the genome and their limitations.

References

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

\textbf{Competing Interests:} No competing interests were disclosed.

\textbf{Reviewer Expertise:} T cell Biology; RNA Biology; Cancer Biology; Molecular Biology; Regulation of gene expression; Protein-RNA interactions; Biochemistry and Molecular Biophysics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.
The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com