Reviewer Report

Title: GuideMaker: Software to design CRISPR-Cas guide RNA pools in non-model genomes

Version: Original Submission **Date:** 7/16/2021

Reviewer name: Kornel Labun

Reviewer Comments to Author:

In this study, "GuideMaker: Software to design CRISPR-Cas guide RNA pools in non-model genomes", Poudel et al. provides a software for sgRNAs design, focusing on genome wide screens. The tool uses the original strategy of finding off-targets with the use of Hierarchical Navigable Small World graphs trying to provide fast running times for the all vs all comparison. Additional novelty is introduced with proximity filters towards features of interest, and filters for restriction sites inside the guide RNA. What's more, the tool creates control guide RNAs which is mandatory for pooled screens. I applaud the selection of the license as all versions of the GuideMaker are available under a Creative Commons CC0 1.0 Universal Public Domain Dedication. Below I list some of my comments and suggestions.
Comments & suggestions:
1. I tested the website and the tool, not finding any bugs and errors. Website is well made, congratulations!
2. Name of the tool: GuideMaker is not self-explanatory for what it is specialized for, which is pooled design. In the future consider naming your tools more distinctly as I am afraid that currently the tool will be buried under hundreds of other GuideSomething tools.
3. Authors also claim to support Cas13 (page 3 line 65), but don't mention anything more specific about it. I mention that because design for RNA is vastly different from design for DNA and it should be explained how the tool designs for RNA.
4. From my understanding the tool offers highly discriminatory settings towards off-target search for a quick resolution of the all vs all comparison problem, however authors ignore that CRISPR off-targets are not defined by the hamming distance, but levenshtein distance. This was proven already by many studies e.g. Tsai et al. 2015. I recommend that authors embrace this issue in the paper and explain why their design may be suitable, and for what kind of studies it would be alright to use hamming distance vs levenshtein distance instead of ignoring the problem.
5. Study could gain prominence by showing a couple figures and describing how the grid-optimization parameters were selected. This would be especially important for everyone that wants to use this tool for nonbacterial gnomes (page 6, lines 128-131). Although script for optimization is included, it would be good to see what are the tradeoffs.
6. I believe that Figure 4 and all other AVX2 vs nonAVX2 comparisons are not interesting enough to include multiple times. AVX2 improvements are nice, but the tool is already plenty fast, and running time of 250 vs 220 seconds does not matter for normal users. Similarly the number of cores does not seem to influence tool speed above 8 cores and one figure should be enough to explain that. Tool claims very fast running times, but does not compare to the running times of other similar tools for the design of the pooled screens, this could highlight its superiority.

7. CHOPCHOP is a general tool for the design of pooled screens while here it is used as a pooled screen tool due to its configurability. Additionally, CHOPCHOP also supports all PAM and all species, but on its python version available here https://bitbucket.org/valenlab/chopchop/src/master/, website supports only some genomes due to slow process of index building for bowtie.
8. Comparisons to CHOPCHOP focus on the guides found, but I don't understand why consensus ratio between the tools should matter. What is more important is whether GuideMaker does indeed not filter any guides that are preferable for each gene (e.g. by CHOPCHOP ranking) and whether its hamming based filter is good enough to not cause significant unknown off-target effects (levenshtein distance off-targets not found by hamming distance filter). All it takes is one bulge and the hamming distance will become large, while levenshtein distance can even be as low as 1.
9. It is not clear to me why the tool can't be used with large genomes, filtering on the 11bp seed and hamming distance should be plenty fast for also very large genomes. Could it be that the tool should support other input, not only genbank file format?

Level of Interest

Please indicate how interesting you found the manuscript: Choose an item.

Quality of Written English

Please indicate the quality of language in the manuscript: Choose an item.

Declaration of Competing Interests

Please complete a declaration of competing interests, considering the following questions:

- Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?
- Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?
- Do you hold or are you currently applying for any patents relating to the content of the manuscript?
- Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?
- Do you have any other financial competing interests?
- Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

Choose an item.

To further support our reviewers, we have joined with Publons, where you can gain additional credit to further highlight your hard work (see: https://publons.com/journal/530/gigascience). On publication of this paper, your review will be automatically added to Publons, you can then choose whether or not to claim your Publons credit. I understand this statement.

Yes Choose an item.