Reviewer’s report

Title: AMLVaran: A software approach to implement variant analysis of targeted NGS sequencing data in an oncological care setting

Version: 0 Date: 24 Oct 2019

Reviewer: Enrico Glaab

Reviewer's report:

The authors present a new web-based software, AMLVaran, to facilitate reproducible variant analysis of targeted NGS sequencing data in a clinical setting. The software provides a very comprehensive coverage of common workflows for targeted NGS analysis, can be customized to work with different variant calling and scoring approaches, and generates standardized reports for clinical assessment. As a test case, the authors apply the approach to three datasets from AML and MDS disease studies, reporting a high predictive performance for this setting.

Overall, although there are many other competitive software tools available, the authors present a useful new software with a variety of practical features not contained in other related tool sets. The manuscript is clearly structured and well written, and I have tested demo logins of the web-based software, which worked as described for all tests conducted. My comments mainly refer to the missing GitHub repository (or missing hyperlink for this repository), the guidelines on best practices recommendations for the user, the evaluation and generation of predictions, and the applicability beyond cancer data (see Major comments below).

Major comments:

1. In the supplementary "QuickStart Guide" on page 9, the sentence "The source code for AMLVaran is provided via GitHub at ..." contains an empty place holder for the GitHub link, i.e. the link is missing. When searching on GitHub for "AMLVaran" I could not find any corresponding repository. The authors need to ensure that the GitHub pages are indeed available and accessible, and that a correct hyperlink is provided.

2. While the software is very flexible and comprehensive and the QuickStart guide is useful, a clinical user will not know which variant calling pipeline and scoring approach to use in practice. For this purpose, the authors should provide recommendations or best practice guidelines, e.g. justified by previous benchmarks or practical experience, to help the less experienced user with the set-up of a suitable analysis pipeline.

3. The sensitivity and PPV results on the example datasets are very convincing; however, since the authors mention that discrepancies between AMLVaran and appreci8 (and also the ground truth?) reflected borderline cases that were reported by only two or three out of eight variant callers, the question arises whether the software is also able to adjust the scoring/confidence for a called variant based on comparing the results from multiple calling pipelines and predicting variant presence probabilities as output (rather than discrete presence/absence predictions). The authors should point out whether a probabilistic analysis can be achieved with the given software and whether this improves the evaluation (e.g. comparing the AUROC), or whether and how a corresponding extension could be
implemented on the basis of the available software.

4. The authors have tested their approach only on cancer datasets, but there are significant differences between the analysis of somatic and germline variants. Since the manuscript indicates a general applicability of the software, the authors should ideally also present an example use case on another disease (or limit the focus explicitly on cancer applications, and clarify already in the abstract that it cannot be used e.g. for family-based germline sequencing).

Minor comments:

1. For the figures, a LaTeX error ("[scale=0.33]Graphics/Fig2" etc.) is shown instead of the image or a reference to the actual figures at the end of the manuscript.

2. The authors mention that technically the software could also be used for WES analysis, but have they tested how the significantly extended data size would affect the performance (even if only a single variant calling approach is used)? If WES analysis is feasible with a reasonable performance and usability, is then also WGS analysis possible and practical? A short comment on the practical usability of the tool for WES and WGS analysis, and whether this has been tested, should be provided.

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:
1. 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?

2. 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?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. 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?

5. Do you have any other financial competing interests?

6. 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.

I agree to the open peer review policy of the journal