Peer Review File

Article information: https://dx.doi.org/10.21037/pcm-21-29

Review Comments

Reviewer A

In the article 'Enhancing deep targeted sequencing for detecting rare subclonal mutations through best practices' Dan Li et al. have developed a web portal to define the best practices to detecting rare subclonal mutation in cancer; moreover, they examine how these mutations are spread throughout different kind of samples.

The introduction is well documented, and the methods are correct and well written, carrying out a set of experiments using different samples. The authors have conducted a thorough investigation better to understand the role of the accurate and robust in the precision oncology. Although the methods and the code of the SEQC2 are public, I recommend that the authors cite the web portal in the manuscript.

This work is exceptionally well organized and is easy to read, scientifically sound, so I earnestly recommend publishing it.

Reply: Thank you kindly for your affirmation and thoughtful comments. The SEQC2 consortium is working hard to organize all data and code into a data portal to encourage and enhance access for the entire research community. Despite our great enthusiasm for this endeavor, it remains in-progress as of authoring this reply. We mention the plan for this task in page 12 line 267.

Changes in the text: N/A

Reviewer B

The authors discuss the accomplishments of the SEQC2 Oncopanel Sequencing Working Group. The issues addressed are of high importance to the field of precision oncology, and the authors have done an impressive job summarizing

the key issues and efforts to address these issues.

Reply: Thank you kindly for your encouraging words. Please see the replies for each question below.

Minor comments:

1. I find the establishing of a reference sample fascinating, and I think the findings are intuitive and informative. Can the authors discuss future improvements to creating reference DNA-based samples? For instance, I wonder if mixing more samples from normal individuals will also generate a greater diversity of low frequency variants that can be used to rule out non-disease associated variants that were found in normal individuals? Or mixing more samples from older normal individuals to rule out mutations due to clonal hematopoiesis?

Reply: Thank you for these insightful comments and questions. The main purpose of generating a reference sample by mixing 10 cancer cell lines is to increase the density of variants involved and thus enhance comprehensive assessments of oncopanels, experimental protocols, sequencing instruments, bioinformatics pipelines, and such. We believe it may be more effective to build databases (dbSNP for example) for collecting those non-disease or aging related mutations. This interesting point seems out of the scope of this manuscript.

Changes in the text: N/A

2. In the absence of a reference sample for RNA expressed mutations and fusions, can the reference DNA sample be used to call low frequency copy number variants, fusions, and other genomic aberrations (inversions, translocations etc)?

Reply: Our reference DNA-based samples were in fact developed from the same cell lines that were used to make Agilent Universal Human Reference RNA sample (UHRR). This design choice is mentioned in the manuscript. Conveniently, the matching RNA reference samples can be used for the RNA-based evaluation of expressed mutations and fusions. These studies are in progress as of writing this reply. In summary, we would recommend using RNA sequencing data for these kinds of analyses.

Changes in the text: N/A

I recommend the manuscript for publication with minor text edits (more discussion).

Reviewer C

This manuscript gives an overview on the valuable efforts of "The Sequencing Quality Control Phase 2 (SEQC2) consortium.

In my opinion this is a review article. The manuscript does not contain original research data. In this respect the title is misleading and does not reflect the main content of the manuscript. The title states "... detecting rare subclonal mutations through best practices". However detection of rare mutations is not the main focus of the manuscript. Furthermore, the best practices are not described.

I would like to suggest to designate the manuscript as a review article, change the title and focus on the summary and discussion of SEQC2 consortium activities.

Reply: Thank you for reviewing our manuscript and for sharing your helpful suggestions. We agree that this is a review article, summarizing the achievements of the SEQC2 Oncopanel Sequencing Working Group and discussing our ongoing studies and perspectives. Following your suggestion, we modified the title to better suit this review format.

Changes in the text: *we changed the title to "FDA-led consortium studies advance quality control of targeted NGS assays for precision oncology".*

Reviewer D

This manuscript describes the most recent accomplishments of the SEQC2 Oncopanel Sequencing Working Group. All the related studies were described and they detailed the bioinformatics frameworks used and importantly provide best practice recommendations, which will be a great help for the scientists working on cancer precision medicine. As the authors concluded the reference samples and performance assessment frameworks of the SEQC2 Oncopanel Sequencing Working Group will enable proficiency testing and routine performance monitoring of oncopanels beyond establishing analytical validity.

Reply: Your insights and comments are greatly appreciated, thank you for your

time and your review.
Changes in the text: N/A

Reviewer E

This paper summarizes the guidelines for detecting rare subclonal mutations by NGS. Authors provided recommendations based on the results of SEQC2 studies. They sited their original papers and provided guidelines and recommendations as a review.

The essence of the study results (real data) might be better to be included in this paper (for example, lines 148, 160, 170, 174, and 177). Similar line of paper seems to be published elsewhere, therefore, it is better to be clarified what is the particular topics or position of this paper? From the title, it is focused on the detection of rare subclonal mutations (including ctDNA?) but the content involves more general issues. If the "rare subclonal mutations" is the target of this paper, the introduction part should be more focused on them (or change the title to fit the content).
Reply: This is a well-thought-out point, and we appreciate your suggestion. Indeed, this paper aims to review the accomplishments of SEQC2 Oncopanel Sequencing Working Group for the quality control studies (5 papers) of targeted NGS assays to advance precision oncology. We agree that the target of this paper is not detecting rare subclonal mutations and the title has been changed to fit the content.
Changes in the text: *the title has been changed to "FDA-led consortium studies advance quality control of targeted NGS assays for precision oncology".*

(General points to consider)

Please consider adding the following points for providing further information and discussion from the SEQC2.

How to consider the relationship between error rate of NGS platforms and false

positive rate of the assay? The former is included in the latter but how to distinguish them? Are there any methods to evaluate sequence error in hardware, which should be different among each platform?

We agree that the sequence error is a critical issue for variant detection, especially for the rare-frequency variants. Various methods have been developed to evaluate and reduce the sequence errors. For example, some bioinformatic quality control algorithms can identify the sequence errors based on the sequencing quality score. Molecular barcodes can be used to track the read replicates and reduce sequence errors. PacBio adopts a method to sequence the circular molecule template multiple times to remove some sequence errors. Some synthetic controls such as gBlocks developed by Integrated DNA Technologies (IDT) are pre-designed DNA-fragment spike-ins for NGS. These kinds of controls can usually be used to measure and quantify technical bias including the rate of amplification errors.

Just as you mentioned, the sequence errors are platform dependent. They are a contributing factor for the false positive variant calls. However, it is difficult to decipher the complex relationship between sequence errors and false positive calls, as for each experiment many other factors (panel target regions, bioinformatics pipelines, etc.) are involved.

Changes in the text: N/A

Is there any guidance to calculate cut-off value for VAF based on the false-positive rate of the assay?

Reply: The false positive rates according to the VAF ranges were different across panels. For examples, with a 2.5% VAF cutoff, the FP rates ranged from about 1 to 4 FP calls per Mb. There was no FP call from any panel when the VAF was over 10%. On the other hand, with higher VAF cutoff, we would lose some sensitivity. So, we concluded in our study of comprehensive solid tumor testing oncopanels that the cut-off value for VAF should be set according to the purpose of the study.

Changes in the text: this sentence is added in page 7 line 143: "*With the default VAF threshold of each panel, FP rates were lower than 10.5 per Mb. The majority of FP calls were in the low VAF ranges (<5%).*"

Similary, how to evaluate software errors such as misalignments or variant call? Are there any reference data to evaluate them?

Reply: Unfortunately, we do not have a reference data suited for answering that question. We adopted the best practice pipelines provided/recommended by each panel vendor. The positive variants and false positive positions were carefully detected by taking the common results of many pipelines and applying critical thresholds. Software error was not the focus of our studies.

Changes in the text: N/A

Are there any reference materials available or in preparation for structural variants such as translocations, large in/del, gene amplification, etc.?

Reply: This is a good question. At this moment we do not have this kind of reference materials in development or a plan yet. We agree it may be a good topic for future work.

Changes in the text: N/A

Usefulness of control sample (in tumor/normal set) as negative control sequence for an evaluation of false positive rate.

Reply: Thank you for this good point. In our study of comprehensive solid tumor testing oncopanels, the normal sample (Sample B) was used to detect false positive calls: if a call was detected in Sample B but with VAF less than 10%, it was considered as a false positive call. This was one of our FP detection strategies. At this moment, we do not have a matching tumor and normal dataset involved as tumors are only available in limited quantities and thus can not be used as a reference sample

Changes in the text: N/A

How do the sensitive methods for mutation detection, such as NanoSeq, BotSeqS, and duplex Seq, affect the performance and guidance of rare mutation detection such as ctDNA?

Reply: We appreciate the reviewer's enthusiasm, as it would be valuable to measure the performance of additional methods for variant detection. Both our reference samples and the results will be/are public available. We hope that

inquisitive groups and researchers who are interested in this topic will obtain our sample and/or results and perform their own studies. Duplex Seq was implemented in one of the ctDNA assays evaluated in our ctDNA study. The assay was developed by Integrated DNA Technologies, Inc.
Changes in the text: N/A

How does the SEQC2 consider or prepare the reference materials for the structural variants such as large In/Del or translocations.
Reply: We do not have any plan at this moment. Thanks for your suggestion, we will keep this in mind as we plan future studies.
Changes in the text: N/A

Overall, we thank the reviewer E for carefully reading our paper and providing great suggestions, many of them point to exciting new directions for further studies.

(Specific comments in detail)
Reply: Again, many thanks for the comments. Please see the replies for each question.

L111 Does "1653 variants in 723 cancer genes" mean pathogenic mutations or involving simple SNPs?
Reply: These variants are not restricted to pathogenic mutations. SNPs may be involved. We did not distinguish them as this metric was at the gene level.
Changes in the text: N/A

L137 How much were the reproducibility, positive detection sensitivity and false positive rate?
Reply: The reproducibility, positive detection sensitivity and false positive rate vary across panels and VAF ranges. To provide more information in terms of these metrics, we added some descriptions to the text.
Changes in the text: We added the sentences, see page 6 and line 140 to page 7 line 147*: "The overall sensitivity was high across VAF ranges in all panels. It*

ranged from 87.1 to 98.3% for the lowest VAF range (1-2.5%). With the default VAF threshold of each panel, FP rates were lower than 10.5 per Mb. The majority of FP calls were in the low VAF ranges (<5%). Reproducibility varied across panels and usually was over 95%."

L141 What is the outcome of the study for evaluation of TMB?

Reply: We investigated some relevant factors that may affect TMB estimation and concluded/recommended that it is better to perform TMB estimation in the consensus targeted regions (CTR) with a targeted panel at least 1 Mb; a VAF cutoff of 5% should be applied. This is described in page 7 and line 155.

Changes in the text: N/A

L144 How much was the hard VAF threshold?

Reply: The false positive rates according to the VAF ranges were different across panels. For examples, with a 2.5% VAF cutoff, the FP rates ranged from about 1 to 4 FP calls per Mb. There was no FP call from any panel when the VAF was over 10%. Thus, we concluded that a hard VAF threshold will be helpful to control the FP rate without giving any specified VAF cutoff value.

Changes in the text: N/A

L131 Is the reference sample for TMB evaluation same as Fig2?

Reply: We assume you are talking about L151. Yes, this TMB evaluation was based on the reference sample we described in Fig 2. Only Sample A was used.

Changes in the text: N/A

L180 Is it possible to show the specific values for the high cell count and the DNA input?

Reply: We recommended high cell count and DNA input in our FFPE study based on the observation of their impacts on the false positive rate. With the data and panels we used, we see that experiments with cell count <16K (equivalent to 100ng DNA) usually showed elevated false positive rates. Obviously, no DNA extraction method could achieve 100% yield efficiency. Different panels may have different DNA extraction methods with variable efficiency and DNA input

requirements. It is not possible to give a specific value for the high DNA input amount across panels. Thus, we did not provide the specific values for others to refer to at this point.
Changes in the text: N/A

L206 Define UMI here instead of L221.
Reply: Good catch, we fixed this in the revised manuscript.
Changes in the text: This "unique molecular identifiers (UMIs)" definition was moved to the first instance (page 10 line 207).

L223 meaning of "untangle spike-in sample native template sequence reads in FASTQ" is unclear.
Reply: Thank you for bringing this to our attention, we reworded this for clarity.
Changes in the text: The sentence is changed to (page 10 line 223): "*The authors developed computational methods to distinguish the spike-in internal standard (IS) controls from the sample native template (NT) sequence reads in FASTQ files.*"

L259 Define "MNV"
Reply: Thank you for noticing this abbreviation with no definition, we spelled out this term in the revised manuscript.
Changes in the text: We replaced MNV as "multiple nucleotide variants", in page 12 line 258.

L335 What is "white blood cells proper"?
Reply: Thank you for your comment. We have updated the manuscript in the specified location, by removing the word "proper" and retaining the phrase "white blood cells".
Changes in the text: "proper" has been removed from page 15 line 334.

L389 What is "aka fixed tissues"?
Reply: The sentence has been updated, thank you for pointing it out.
Changes in the text: "FFPE" has been retained, and the phrase "aka fixed tissues" has been removed in page 18 line 389.