Bioinformatics

Decision Letter (BIOINF-2014-0985)

From: bioinformatics.editorialoffice@oup.com

To: pjflaherty@wpi.edu, patrick.flaherty@gmail.com

CC: yhe2@wpi.edu, pjflaherty@wpi.edu, patrick.flaherty@gmail.com

Subject: BIOINF-2014-0985 - Major Revision

Body: 02-Jan-2015
Manuscript ID: BIOINF-2014-0985
Title: RVD2: An ultra-sensitive variant detection model for low-depth heterogeneous next-generation sequencing data

Dear Dr. Flaherty,

The reviews of your manuscript are now in hand for Bioinformatics and can be found at the foot of this e-mail.

Based on the reports of the referees, the paper has been rejected for publication in its present form. However, the Associate Editor, Inanc Birol, considers that if the paper were substantially rewritten taking into account all the referees' comments it may become acceptable for publication.

We ask that major revisions are submitted within one month ideally, but the system will allow a revised paper to be submitted within 90 days of the original decision date.

Please summarise your changes for the editor indicating which changes you have made and which changes you do not wish to make and why. This can be done either in a Response to Reviewers file uploaded alongside your revised manuscript or through the Author Centre where you can enter your responses directly.

Please submit your revised version through the Author Centre by clicking on the purple button 'Click here to Submit a Revision' in the Bioinformatics ScholarOne Manuscripts web site (https://mc.manuscriptcentral.com/bioinformatics).

At this major revision stage we ask that you upload the following revised manuscript files:

EITHER: (i) A .doc or .rtf file of the revised manuscript, with all tables, figures, schemes and equations inserted in the document.

OR: (ii) All necessary LaTeX files that will be required by the typesetter (including bioinfo.cls, bib, .bst and .ps files).

Please can you mark-up the changes made after revision by using the track changes function or highlighting these in red text.

NOTE: Please upload your final version of supplementary materials without any changes marked. This should be in pdf or Word format, not LaTex.

On behalf of the Bioinformatics Associate Editor, Inanc Birol, I want to thank you for selecting Bioinformatics to present your work.

Best regards,
Alison Hutchins
Bioinformatics

Here are the Associate Editor's comments:

Although your reviewers are in principle supportive of your work, one of your reviewers also

raise some major concerns that we agree with.
If all the reviewer comments are comprehensively addressed, we will reconsider your work.

Here are the comments of the reviewers:

Reviewer: 1

Comments to the Author
The manuscript from Yuting He and Patrick Flaherty presents an update of the method named Rare Variant Detection (RVD), published last year. The main feature of RVD2 is enhancing the detection of SNVs on lower coverage conditions. The manuscript is well written and, overall, the authors mainly focus on deeply exposing the modifications of the former method version. Additionally the authors apply RVD2 to one remarkable control based on synthetic DNA and DNA-seq of one gene (PAXIP1) in a paired tumor/normal cell line (HCC1187/HCC1187BL). The discussion and conclusions is fair to the limitations of the method.

I cannot review the method itself due to a personal limitation to Bayesian statistic, therefore, my review is mainly focused on the application of the developed method.

I have two major concerns regarding the presented results. First, when using synthetic DNA to evaluate the sensitivity and sensibility of RVD2, much is dedicated to (currently) unrealistic coverages. For example, in the Figure1 and Table1 (Figure 3) the authors presents a sequence coverage of 5.000x and even 55.000x. Despite being very useful for the own authors as high hypothetical control, I think ~30x-100x coverage is more informative, and presenting these extremely high coverages can be misleading.

Despite having the whole genome sequencing for HCC1187 and HCC1187BL cell lines, the authors focus all presented analysis one a single gene. It's not clear why the authors used the coordinates chr7:154738059-154782774 (~45Kb), excluding the three first exons (four introns) and the last exon and most of the last intron, when, in fact, PAXIP1 has ~59Kb. No further explanation is given regarding why this specific region was chosen.

Moreover, instead of using DNA sequencing of primary tumors genomes, which are highly heterogeneous and present many rare (single nucleotide) variants, the authors used the DNA from a commercially available cell line, which is colonially expanded (genetically homogenous). Interestingly, chromosome 7, where the investigated gene is located, is tetraploid on HCC1187[1], therefore, one would expect to find only allele frequencies 0%, 25%, 50%, 75 and 100% on variants located in this chromosome. Figure 5 seems to be in agreement to this hypothesis, except for variants 49704 and 81769, which has marginally lower estimations of allele frequency.Therefore, the variants detected on HCC1187 are not actually rare variants (MAF=25%) and, as shown on Figure 3 (Table1), may be detected by most of currently available methods. Despite the limitations, the analysis is complete, but it would be more informative to apply RVD2 to publicly available (and validated) DNA-seq from primary tumors.

b) minor:
Synthetic DNA Sequence Data
The authors cite a paper from 2011 to refer for the Synthetic DNA sequencing. Explicit details on the technology, read length and number of reads generated would greatly ease readers understanding of this section.

Comparison to other methods
The authors greatly expose the comparison of sensibility, specificity, true positive and true negative rates. It seems that the caveat of the method is on memory consumption and processing time. It would be very informative to explicitly have a table with execution time and memory consumption of all utilized methods, preferably in a more realistic scenario, more specifically, a whole genome analysis.

[1] http://www.path.cam.ac.uk/~pawefish/BreastCellLineDescriptions/HCC1187.html

Reviewer: 2

Comments to the Author
The authors have developed a partially novel and nicely refined algorithm to maximize sensitivity for the detection of somatic mutations in a tissue that may have only a small

fraction of cells that harbor the mutations such as in a tumor. And in particular, they have looked at improving the probability of finding true positives and minimizing false positives in areas of sequence coverage that are less than optimal for doing so. Because of the practical and economic reality that nearly all data collected to detect the presence of causal cancer mutations (or even to examine somatic mutation occurrences in a non malignant situation) is always limited, the development of algorithms to do this is very important. The authors have done a first rate effort to follow good practices in developing, testing, and comparing their approach to those of others and the work and code is quite useable by others now.

Specific points to check on:

page 4 line 22; check the signs of the threshold boundary conditions in your hypothesis tests.. looks like one should be less than minus Tau..

on pre-processing methods (4.1); please check what you mean by the -C50 option of BWA.. That should be something else.

On figure 6 panel A; at position index 35, 36, 37 and position 50; please check that the symbols shown for mis-call versus mutant calls are as you intend them.. looks like two are reversed

---

Date Sent: 02-Jan-2015

Close Window