Reviewer's report

Title: Unraveling the structural variome using insights from next generation sequencing technology: tools and tips for a messy job

Version: 2 Date: 20 May 2013

Reviewer: Santhosh Girirajan

Reviewer's report:

Comments to the Authors: Pavlopoulos and colleagues review the advances in the field of genome sequencing and illustrate methods to analyze sequencing data with a focus on study structural variation. The review, with revisions, will be a good overview of the various sequencing technologies and projects. The authors present on the complexities involved in analyzing sequencing data and current methods for structural variant detection. I like the organization of the review but fear that the students reading it might get lost, especially with the citations and headings. The figures and tables are complete, comprehensive and valuable.

Major Compulsory Revisions

The abstract does not reflect the title. Either the abstract has to be re-written to suit the title and the main text or vice versa. Alternatively, one possible suggestion for an appropriate title, based on the abstract and the main text, is “Unraveling genomic variation from next generation sequencing data”.

Several citations are incorrect in the paragraph under Sequencing Techniques. For example, a review on gene networks is cited for SOLiD technology. The authors should cite the primary articles instead of obscure papers that just mention the technology. References 22-31 need to be changed and primary or most appropriate review (for example, cite a Church or Shendure paper for second generation sequencing) cited. I don’t think Sanger sequencing is next generation.

Minor Revisions

Next generation sequencing applications: The paragraph title is out of place since the previous title is way generalized. Definition of “next generation” should be made carefully in the previous section. ChIP-seq information is shortchanged. I think it would be good to remove RNAseq and ChIP-seq sections. The references are mostly wrong, again, please check all the references (reference 50 is wrong, Reference 57 is not correct).

Discretionary revisions

The headings, somehow, do not make sense. I wonder if the following headings would suit the text, in the order it is presented.

1. Introduction
2. Sequencing technologies
a. Applications of next generation sequencing technologies
b. DNA sequencing
   i. Types of sequence assemblies
   ii. File formats
   RNA sequencing (better to eliminate RNaseq - make a mention of other extensions by citing a review (for example, take one from Mike Snyder)
3. Genome annotation and variant calling
   a. SNP calling pipelines
   b. Structural variation detection
4. Discussion

Thanks!
Santhosh Girirajan

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**
I have no competing financial interests