Reviewer’s report

Title: Scalable and Cost-effective NGS Genotyping in the Cloud

Version: 1 Date: 7 May 2015

Reviewer: Sandro Morganella

Reviewer’s report:

The Manuscript describes GenomeKey, a workflow for the analysis of NGS data that uses COSMOS, a well-known python library, to implement a parallel version of the presented algorithm. The workflow was deployed on Amazon Web Service, and performance evaluation was performed on a comprehensive benchmark dataset containing both exomes and genome data. Results show that the presented approach enables the analysis of NGS data with a good accuracy and demonstrate that this workflow is highly scalable with respect to the size of the analyzed dataset.

The manuscript is an interesting contribution to the field, especially considering the fact that the presented workflow significantly improves the computational time of data-intensive analysis steps, such as genome alignment.

Minor Essential Revisions

(1) Authors should improve the readability of the manuscript avoiding very long sentences and the repetitive use of the word 'workflow'.

(2) Supplementary Fig2 is a very important part of the presented study and it should be added in the main manuscript.

(3) In Fig1 A, neither the legend nor the caption provide a description of the light blue dotted line.

(4) Page 4 "... 3055906 matching alleles and 7368 ..." thousand separator should be used

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 declare that I have no competing interests