Dear Editor,

We would like to present the manuscript titled “Scalable Clustering of Genotype Information using MapReduce” as “Review articles” in Briefings in Bioinformatics.

In the paper we use the task of clustering exon-sequencing derived genomic variants as the example to illustrate advantages and limitations of a Hadoop-based BigData analysis solutions. Specifically, we compare the clustering by the BigData library implementation, Mahout, against a comparable solution in R, iteratively scaling up on the number of individuals and genome-wide features used. Furthermore, we discuss the common misconception that Hadoop-based approaches are seamlessly scalable and do not need expert input for resource management and show the impacts of not tuning the resources to the properties of the problem.

We expect the interface and the discussed issues and guidelines around Hadoop-approaches to be a valuable resource for a wide range of Briefings in Bioinformatics readers.

Our software is freely available from https://github.com/BauerLab/GeMaIn.

Thank you in advance for your consideration.

Sincerely,

**Denis C. Bauer**

(on behalf of all authors)