A.C. Camargo
Cancer Center

**Israel Tojal da Silva, PhD**

Computational Biology and
Bioinformatics Group
Rua Taguá, 440, São Paulo SP 01508-010
☎ +55(11) 2189-5000 ext: 2941
itojal@cipe.accamargo.org.br
http://www.accamargo.org.br

Dear Editor,

We are submitting a manuscript entitled *signeR: An empirical Bayesian approach to mutational signature discovery*, which we would like to have considered for review by BIOINFORMATICS.

It is currently recognized that the accumulation of genomic alterations is one of the major causes of malignant transformation. However, the knowledge of the processes that trigger the involved somatic mutations still remains elusive. Recently, these processes have been identified with mutational signatures while analysing high throughput data via non-negative matrix factorisation (NMF) methods. These efforts have advanced current knowledge about mutagenesis and the development of cancers and have received substantial interest in the literature. However, current methods used for the identification of mutational signatures are strongly dependent upon initial conditions and address the estimation of the underlying number of signatures by using ad-hoc heuristics.

In this manuscript we present a new method for the statistical estimation of mutational signatures based on an empirical Bayesian treatment of the NMF model. Our method addresses the determination of the number of signatures directly as a model selection problem. In addition, we introduce two new concepts, of significant clinical relevance for evaluating the mutational profile: the differential exposure score and the posterior classification. The advantages brought by our approach are shown by the analysis of real and synthetic data sets. Our approach is robust to initial conditions and more accurate than competing alternatives which consider the same NMF model. It also estimates the correct number of signatures even when other methods fail.

We have made our code available at https://github.com/rvalieris/signeR. A version submitted to Bioconductor should be available shortly. Please also note that the supplementary material includes a list of instructions about how to install and run this code. Last, to aid the reviewing process in case there is any trouble during the installation process, we also set up a Rstudio server at http://143.107.223.173:8787/. Login instructions are described Supplementary File for Review Only.

Our work is grounded on a solid statistical basis and we believe that it is attractive to a broad spectrum of computational biology scientists and those of the cancer genomics field and therefore hope you will consider it for publication in BIOINFORMATICS.

June 21, 2016

Sincerely yours,
Rafael A. Rosales
Rodrigo Drummond
Renan Valieris
Emmanual Dias-Neto
Israel Tojal da Silva