JASA ACS Reproducibility Initiative - Author Contributions Checklist Form

The purpose of the Author Contributions Checklist (ACC) Form is to document the artifacts associated with a manuscript (e.g., code and data supporting the computational findings), and describe how to reproduce the findings. The final version of this document will be included as online supplemental material with the published paper and referenced in the abstract.

As of Sept. 1, 2016, the ACC Form must be included with all submissions to JASA ACS.

This document is the template that will be provided to authors; please replace the (non-bold) text below that provides guidance on how to fill out each item with the actual information for your manuscript.

Data

Abstract (Mandatory)

We used The Cancer Genome Atlas (TCGA) signature loading data. Alexandrov et al. (2019) identified 49 mutational signatures and corresponding signature loadings for 9493 cancer genomes across 32 cancer types.

Reference
Alexandrov, L. B., J. Kim, N. J. Haradhvala, M. N. Huang, A. W. Ng, Y. Wu, A. Boot, K. R. Covington, D. A. Gordenin, E. N. Bergstrom, S. M. A. Islam, N. Lopez-Bigas, L. J. Klimczak, J. R. McPherson, S. Morganella, R. Sabarinathan, D. A. Wheeler, V. Mustonen, , G. Getz, S. G. Rozen, and M. R. Stratton (2019). The repertoire of mutational signatures in human cancer. bioRxiv.

Availability (Mandatory)

The entire dataset titled as "*TCGA_WES_sigProfiler_SBS_signatures_in_samples*" can be downloaded from https://www.synapse.org/#!Synapse:syn11801497 as of December 2019. The link varies when the authors update their manuscript, but the name of the file is invariant so that one can find the dataset with the provided navigation.

Description (Mandatory if data available)

The entire dataset contains signature loadings for all 49 signatures and 9493 cancer genomes across all 32 cancer types, but in the manuscript, we reported results with respect to gastrointestinal cancers, colorectal, esophageal and stomach adenocarcinomas. Regarding these three cancer types, signature loadings corresponding to active signatures are provided in data supplementary .zip files. A detailed information is described in Tables-and-Figures.html document.

Code

Abstract (Mandatory)

An R package SKIT implements the computation of test statistics and corresponding p-values. To speed up for bootstrap computation, the main function skit() is written as C.

Description (Mandatory)

The SKIT package is included in the data supplementary .zip file and a manual including an example is also included in the same file. If the manuscript is accepted it will be available at https://dceg.cancer.gov/tools/analysis/skit.

Instructions for Use

Reproducibility (Mandatory)

An Rmarkdown "Tables and Figures.Rmd" contains code to reproduce results for TCGA data analysis. A compiled html version Tables-and-Figures.html describes a detailed workflow to generate the main results in Table 2, Figure S.3 and Figure 4. In order to create Figure 4, the R package 'circlize' will be required.

Please note that all the computations in the manuscript are obtained under R version 3.6.3 but they are reproducible under the latest R version 4.0.3 (as of December 2020).

The computation is conducted in a Dell Latitude 7490 laptop with Intel(R) Core(TM) i7-8650 CPU @ 1.90GHz. With this machine the entire run-time for the .Rmd file is less than a minute.