

Application ‘Methylation profile and survival analysis’

by Pawel Pytlak, Adrianna Sudol

Introduction

The Shiny application **Methylation profile and survival analysis** was built in order to explore the different relationships between the DNA methylation profile and the lifespan of the selected group of patients suffering from different cancers. The analysis is based on the data derived from the RTCGA.methylation and RTCGA.clinical platforms. The detailed information on this data can be found on the following websites: <https://github.com/RTCGA/RTCGA.methylation> and <https://github.com/RTCGA/RTCGA.clinical>.

Data description

We consider both methylation and clinical data sets with the information about the following cancers:

- Breast Invasive Carcinoma (BRCA)
- Colon Adenocarcinoma (COAD)
- Colon Adenocarcinoma and Rectum Adenocarcinoma combined (COADREAD)
- Glioblastoma Multiforme (GBM)
- Glioblastoma Multiforme and Brain Lower Grade Glioma combined (GBMLGG)
- Kidney Renal Clear Cell Carcinoma and Kidney Renal Papillary Cell Carcinoma combined (KIPAN)
- Kidney Renal Clear Cell Carcinoma (KIRC)
- Kidney Renal Papillary Cell Carcinoma (KIRP)
- Acute Myeloid Leukemia (LAML)
- Lung Adenocarcinoma (LUAD)
- Lung Squamous Cell Carcinoma (LUSC)
- Ovarian Serous Cystadenocarcinoma (OV)
- Rectum Adenocarcinoma (READ)
- Stomach Adenocarcinoma (STAD)
- Stomach Adenocarcinoma and Esophageal Carcinoma combined (STES)
- Uterine Corpus Endometrial Carcinoma (UCEC)

In terms of our analysis, we consider CpG sites, i.e. places in the DNA sequence where the methylation occurs. These CpG sites will be hereinafter referred to as *probes* (for more information, visit the following website: https://en.wikipedia.org/wiki/CpG_site). For the purpose of the application, our analysis was restricted to the most significant probes for each cancer type. We determined the top 200 significant probes for each cancer using the log-rank test, with the median of methylation in a given probe as a threshold value. As a result we obtained 2452 different probes which were significant to at least one cancer type.

Application description

After executing the application the user has access to the following panels:

- Introduction

A brief description of the application.

- Survival Curves

This panel enables the user to analyse Kaplan-Meier survival curves (for more information about the survival curves visit <http://www.statystyka.az.pl/analiza-przezycia-metoda-kaplana-meiera/>).

Firstly, the user has an opportunity to select a probe of interest using the *Choose a probe to analyse* option and to subsequently select the appropriate cancer with the *Choose a cancer type to analyse* option. At the end the user can specify the degree of methylation, which serves as a threshold value during the execution of the log-rank test. After specifying the particular option, a plot of the survival curves is drawn, as well as the results of the corresponding log-rank test.

- Significant markers

This panel depicts the list of the most significant probes for each cancer type. The user can choose an appropriate cancer type in the side panel, as well as the number of probes to be displayed.

- Common markers

In this panel the user has an opportunity to see the names of significant probes common for multiple cancer types, as well as probes significant for single cancer types. The *2 cancer types* section displays a list of probes significant to exactly two cancers, as well as the names of the cancer types. In the *3 cancer types* section a list of probes significant to three cancers is shown. The *4 cancer types* and *5 cancer types* sections act similarly. A probe common for 6 cancer types doesn't exist.

- Distribution of methylation

This panel depicts the distribution of methylation using different statistical plots. Firstly, in the side panel the user can select a probe of interest, a single cancer type or multiple cancer types, in order to compare the methylation of a given probe between the different cancers. After choosing certain options the user can analyse three plots: Cumulated histogram: a histogram created on the basis of cumulated data from all the chosen cancer types. Comparison histogram: a histogram displaying the distribution of methylation in the chosen cancer types. Comparison boxplots: boxplots for each chosen cancer type, displayed on a single plot in order to compare different cancer types simultaneously

Summary

The **Methylation profile and survival analysis** application enables the user to analyse how the methylation process affects the survival length of patients suffering from certain types of cancers. The user has an opportunity to explore thoroughly the methylation of particular CpG sites by using statistical tests and plots. Furthermore, the application contains tools for comparing the methylation in a single probe simultaneously for multiple cancer types.

Pawel Pytlak

Warsaw Univeristy of Technology, Faculty of Mathematics and Information Science

line Koszykowa 75

line Warsaw, Poland

pytlakp@student.mini.pw.edu.pl

Adrianna Sudol

Warsaw Univeristy of Technology, Faculty of Mathematics and Information Science

line Koszykowa 75

line Warsaw, Poland

sudola@student.mini.pw.edu.pl