mRNA Browser Documentation

by Karolina Wyszyńska, Krzysztof Rudaś

Introduction

Following documentation refers to the application **mRNA Browser** which examines an influence of various mRNA expressions on a propability of patients survival for 13 cancer types. All the data comes from *RTCGA* (Kosinski and Biecek, 2015) library. For any furter information about collected data visit: Multiplatform analysis of 12 cancer types to identify integrative subtypes.

Standard biostatistic and propabilistic methods such as Kaplan-Meier, logRank test or histogram were used to solve this problem. Main technical tool for creating was R environment. The authors used few libraries such as: *ggplot2* (Wickham, 2009) (data visualization), *survMisc* (Dardis, 2015) (biostatistic methods), *shiny* (RStudio, Inc, 2013) (libraries to create applications with).

Application's description

User that open the aplication sees a welcome screen with a short description, authors and version with the date of release. There is also a button that reffers to appropriate application. After pressing it there are three bookmarks to choose from:

- **Gene influence on different types of cancer** - it offers a comparison of significance of given gene (mRNA expressions) for four choosen cancer types,

- **Genes influences on specyfic type of cancer** - gives an avability to compare a significance of four choosen genes (mRNA expressions) for a given cancer type,

- **Detailed survival plot** - introduces the significance of a choosen gene (mRNA expression) for a choosen cancer type with more detailed information and posiibility of modyfying some parameters.

All of the bookmarks are described more precisely below.

There is also an **Instruction** button that will return the user to the welcome screen.

Gene influence on different types of cancer

In this bookmark a user can choose four out of thirteen cancer types form a list. Next the user selects interesting gene (mRNA expression). The effect of the choice are four survival plots showing Kaplan-Meier curve that appear on the right side. There are two curves on each plot. One of them (orange) refers to patients who had the mRNA expression over median and the second (green) refers to patients which had the mRNA expression below median. The median was counted from all values of mRNA expression for fixed cancer type and gene. Below the genes list user can also find a slider with x axis values. It allows to modify the range of years showed on the plots that are in user's special interest.

Reffering to the genes list again the authors want to mark that they were sorted in a special way. On the beggining they were sorted by increasing p-value counted using LogRank test for each gene (with the median and cancer type as stratum values). However the results were not satysfying, so they choose another method:

1. For each gene (mRNA expression) they counted its p-value in LogRank test for a fixed cancer type (one of all 13).

2. Then for each cancer type they ordered the genes form the most important to the least important and remembered the number of its place.

3. Next for each gene they sumed its places numbers and sorted all of the genes from the least sum to the greatest.

This method caused that genes which were really important for all (or nearly all) of the 13 cancer types would be seen on the top of the list.

Genes influences on specyfic type of cancer

The bookmark is analogical to the previous one. However this time the user can choose one cancer type and four genes (mRNA expressions) which influences he wants to compare. Four default genes are the most significant for the choosen cancer type. All of the genes list are sorted from the most to the lest significant gene for given cancer type. As an effect four plots are generated. They show Kaplan - Meier plots for choosen genes. Plots are build analogically to the ones placed in a previous boookmark.

Detailed survival plot

In the last bookmark user can make a more precisely analyse of significance of choosen gene (mRNA expression) for a fixed cancer type. On the beggining the user selects a gene and a cancer type that are interesting for him. Then a histogram plot is showed below which can be helpful with modyfing a stratum value. Thanks to it user can also see the propability distribution of mRNA expression or which values occured more frequenty than others.

On the right side the user sees a plot analogical to the ones seen in the previous bookmarks. The significant addition is that previously the stratum value was fixed on a median point (default value) and now the user can manipulate that using slider on the left. However manipulating this value can result in "fake" conclusions. Thus authors put some more additional information under the plot showing how many patients are considered in each of the stratum groups.

What is more there is also a possibility of downloading the plot that the user sees. After pressing a **Save Plot** button there is a link showed over it leading to the plot's png file.

Summary

The **mRNA Browser** is simple and easy to use application that helps to show and explore which genes (measured by mRNA expression) have significant influence on patients suffering from cancer survival. Clarity in operating, easy interpretation of plots and a wide spectrum of possibilities makes this application a very usefull tool in doctors, biostatisticians and other people working with onkology topic work.

Bibliography

C. Dardis. *survMisc: Miscellaneous Functions for Survival Data*, 2015. URL http://CRAN.R-project.org/package=survMisc. R package version 0.4.6. [p1]

M. Kosinski and P. Biecek. *RTCGA: The Cancer Genome Atlas Data Integration*, 2015. R package version 1.1.10. [p1]

RStudio, Inc. *Easy web applications in R.*, 2013. URL: http://www.rstudio.com/shiny/. [p1]

H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN 978-0-387-98140-6. URL http://had.co.nz/ggplot2/book. [p1]

Karolina Wyszyńska

wyszynskakarolina@gmail.com

Krzysztof Rudaś

krzysztof.rudas@gmail.com