Genes mutations

by Marlena Bielat, Małgorzata Dobkowska, Sebastian Gargas

Summary

Application Genes mutations allows to browse the results of analysis of mutations of genes from The Cancer Genome Atlas using RTCGA package. The main goal is to find which biomarkers significantly impact on patient survival in different types of cancers.

Data

All of the used sets for the following cancers comes from RTCGA repository:

- GBMLGG - GBM + LGG (Glioblastoma multiforme and Lower Grade Glioma),
- BRCA - Breast invasive carcinoma
- KIPAN - Pan-kidney cohort
- COADREAD - Colorectal adenocarcinoma
- STES - Stomach and Esophageal carcinoma
- GBM - Glioblastoma multiforme
- OV - Ovarian serous cystadenocarcinoma
- UCEC - Uterine Corpus Endometrial Carcinoma
- KIRC - Kidney renal clear cell carcinoma
- HNSC - Head-Neck Squamous Cell Carcinoma
- LUAD - Lung adenocarcinoma
- LGG - Lower Grade Glioma
- LUSC - Lung squamous cell carcinoma
- THCA - Thyroid carcinoma

The cancers from list corresponds to sets containing informations about at least 500 patients. For each cancer we choose genes on which mutation has occurred in at least 5% cases and then we find genes which have significant impact on patients survival. This significance was measured by p-value of log-rank test comparing survival time in two groups of patients divided by the presence of the mutation. The final list of biomarkers is the union of significant genes for each type of cancer. In our application we consider 535 different genes.

To analyze the impact of the gene mutation on survival time, we have chosen two features:

- presence of mutation - we divided patients in two groups: The first one refers to patients without the gene mutation. The second one refers to patients with at least one mutation on gene.
- type of mutation - we used variable Variant Classification, in which mutations are classified into 10 different types: Missense_Mutation, Silent, Frame_Shift_Del, Frame_Shift_Ins, In_Frame_Del, Nonsense_Mutation, RNA, Splice_Site, In_Frame_Ins, Nonstop_Mutation.

Summary of gene mutation

The table contains informations about the frequency and number of patients with the mutation of the selected gene among patients suffering from different types of cancers. It also includes information about the significance of mutations to patients survival. It is measured by the p-value of the log-rank test, which allow to compare the survival time of two samples: in our case that samples are patients with the mutation of the selected gene and patients without this mutation. The null hypothesis is that the two groups have the same distribution of survival time. If the p-value is less than 0.05 then we have grounds to

reject the null hypothesis at given significance level and the selected gene can be considered as a prognostic biomarker.

Survival curves: Presence of mutation

This tab displays Kaplan-Meier curves for the given gene and the given cancers. The survival curves are estimated for the two groups of patients: the first one refers to the patients with a mutation of the given gene and the second one refers to the group of patients without any mutation of this gene. The Kaplan-Meier method estimates the probability of survival at least until specified on the X axis time. Marks mean censored observations. The figures also show the p-value of the log-rank test, which is helpful when we want to find out if two survival curves can be considered as different. Genes mutations are rare events, so estimates are often based on only a few cases - it is worth to keep it in mind.

Co-occurring genes

The table presents informations about co-occurrence mutations. For the selected gene and selected cancers a list of all considered biomarkers is given with information about how often patients who have had a mutation in a given gene have also mutation on listed genes. In addition, we show number of patients suffering on specific types of cancers with mutations on both genes.

Survival curves: Variant Classification

This tab also displays Kaplan-Meier curves for the given gene and the given cancers for two classification variants - presence of missense mutation and presence of nonsense mutation. Similar to tab 'Survival curves: Presence of mutation', the figures show the p-value of the log-rank test, which examines the significance of a particular type of mutation.

Frequency of mutation types

The table contains informations about the occurrence frequency of different types of mutations for a selected gene and selected cancers. As you can see in majority of cancers the most frequent types of mutation are Missense Mutation and Nonsense Mutation. Groups with another types of mutations offten are to small, so testing the significance of a particular type of mutation is not justified. In addition, we again show the level of occurrence of mutations in a given gene.

(R Core Team, 2012)

Bibliography

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL http://www.R-project.org/. ISBN 3-900051-07-0. [p2]

Marlena Bielat, bielat.marlena@gmail.com
Małgorzata Dobkowska, gosia.dobkowska@tlen.pl
Sebastian Gargas, gargass@student.mini.pw.edu.pl
MiNI Warsaw University of Technology