

Protein expression

Agata Starzyk, Marta Czerwińska

Abstract

The main objective of the project was to show difference in mortality in dependence of protein expression. The authors have used a few methods of the results visualization: the Kaplan-Meier estimate, Box plot and Densities. To use this methods it was necessary to use libraries in program R-Studio: shiny-to create the application, ggplot2-to plot the graphs, survival-to compute survival curves.

Data

All data that are in our application come from site: <https://github.com/RTCGA/RTCGA.RPPA> . We have a 35 types of cancer and 285 protein types. Below is a list with full names of all types of cancer:

- ACC Adrenocortical carcinoma
- BLCA Bladder urothelial carcinoma
- BRCA Kidney renal clear cell carcinoma
- CESC Cervical and endocervical cancersHead and Neck squamous cell carcinoma
- CHOL Cholangiocarcinoma
- COAD Colon adenocarcinoma
- COADREAD Colorectal adenocarcinoma
- DLBC Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
- ESCA Esophageal carcinoma
- GBM Glioblastoma multiforme
- GBMLGG Glioma
- HNSC Head and Neck squamous cell carcinoma
- KICH Kidney Chromophobe
- KIPAN Pan-kidney cohort (KICH+KIRC+KIRP)
- KIRC Kidney renal clear cell carcinoma
- KIRP Kidney renal papillary cell carcinoma
- LGG Brain Lower Grade Glioma
- LUAD Lung adenocarcinoma
- LUSC Lung squamous cell carcinoma
- MESO Mesothelioma
- OV Ovarian serous cystadenocarcinoma
- PAAD Pancreatic adenocarcinoma
- PCPG Pheochromocytoma and Paraganglioma
- PRAD Prostate adenocarcinoma
- READ Rectum adenocarcinoma
- SARC Sarcoma
- SKCM Skin Cutaneous Melanoma
- STAD Stomach adenocarcinoma
- STES Stomach and Esophageal carcinoma
- TGCT Testicular Germ Cell Tumors
- THCA Thyroid carcinoma
- THYM Thymoma
- UCEC Uterine Corpus Endometrial Carcinoma

- UCS Uterine Carcinosarcoma
- UVM Uveal Melanoma

First information about application

The first tab that opens at the beginning is the Instructions. Everyone that visit the application can read about methods that have used or what it can do. For everyone that don't know what Kaplan-Meier estimate, Box plot or Densities are references with more informations about.

Cancer Types

The second bookmark is about all types of cancer. Here is information about full name of cancer and how many observation we have analyzed. In the top left corner is browser in which you can find chosen type of cancer.

KM

In this tab it is important to choose the type of protein expression. The visitor have possibility to choose only one type of protein. Afterwards choose at least one cancer type of 35 available. The application can compare different types of cancer but does not compare a different type of protein.

For each type of cancer we receive two curves in the plot area. One is the observations of patients that the protein expression are higher of median of all observation for chosen type of cancer, and the second curve shows observations that are lower of median. In each plot that have at least one observation, p-values is computed. For all plots p-value is also calculated and showed below plot area.

Box plot

In this section we have box plots. A boxplot splits the data set into quartiles. The body of the boxplot consists of a "box" (hence, the name), which goes from the first quartile (Q1) to the third quartile (Q3). Within the box, a vertical line is drawn at the Q2, the median of the data set. Two horizontal lines, called whiskers, extend from the front and back of the box. The front whisker goes from Q1 to the smallest non-outlier in the data set, and the back whisker goes from Q3 to the largest non-outlier.

Density

We included a density plots in our last tab. You can see on one plot all types of cancer for chosen protein type. The next advantage is that this tool allows you compare the distribution of every combination of cancer types for chosen protein.

Bibliography

1. Silvia von der Heyde¹, Johanna Sonntag², Daniel Kaschek³, Christian Bender⁴, Johannes Bues², Astrid Wachter¹, Jens Timmer^{3,5, 6}, Ulrike Korf², and Tim Beißbarth, RPPanalyzer Toolbox: An improved R package for analysis of reverse phase protein array data, <http://www.biotechniques.com/BiotechniquesJournal/2014/September/RPPanalyzer-Toolbox-An-improved-R-package-for-analysis-of-reverse-phase-protein-array-data-353895.html>
2. Arthur Allignol and Aurelien Latouche, CRAN Task View: Survival Analysis, <https://cran.r-project.org/web/views/Survival.html>

3. R Documentation, Box Plots, <https://stat.ethz.ch/R-manual/R-devel/library/graphics/html/boxplot.html>
4. R Documentation, Kernel Density Estimation, <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/density.html> BioTechniques - RPPanalyzer Toolbox: An improved R package for analysis of reverse phase protein... Analysis of large-scale proteomic data sets requires specialized software tools, tailored toward the requirements of individual approaches. Here we introduce an extension of an open-source software solution for analyzing... biotechniques.com