

Introduction

The following poster presents **RTCGA**: a family of R packages with data from The Cancer Genome Atlas Project (TCGA) study. TCGA is a comprehensive and coordinated effort to accelerate our understanding of the molecular basis of cancer through the application of genome analysis technologies, including large-scale genome sequencing^a. We converted selected datasets from this study into few separate packages that are hosted on Bioconductor. These R packages make selected datasets easier to access and manage. Data sets in RTCGA.data packages are large and cover complex relations between clinical outcomes and genetic background.

^a<http://cancergenome.nih.gov/>

These packages will be useful for at least three audiences: biostatisticians that work with cancer data; researchers that are working on large scale algorithms, for them RTGCA data will be a perfect blasting site; teachers that are presenting data analysis method on real data problems. In this poster we present applications of the statistical modeling methods to the cancer's subtypes basing on the available cancer data of over 8,000 patients.

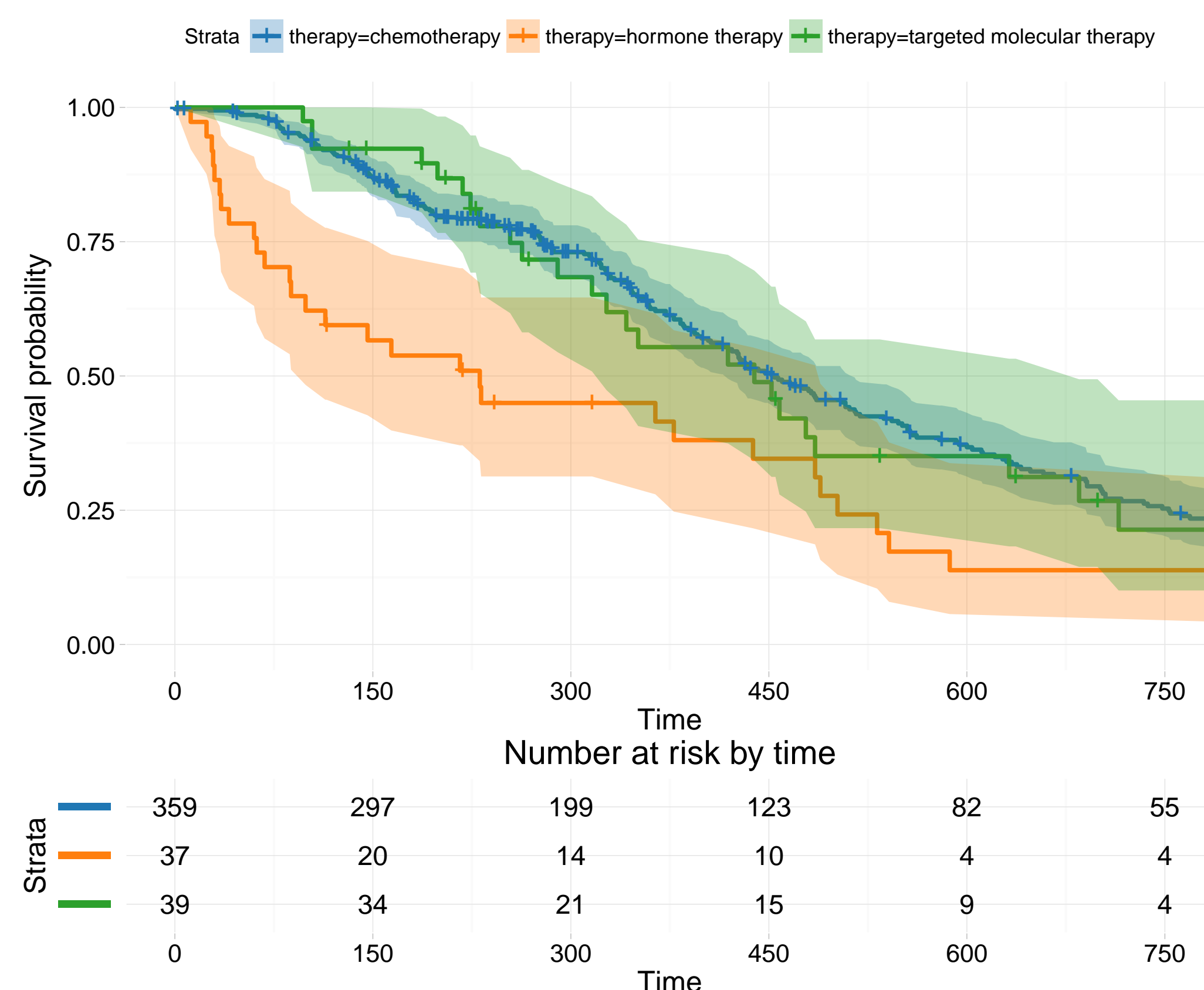
The Kaplan-Meier survival curves

RTCGA family is excellent when one researches in a field of survival analysis and genomics. Survival times are included in clinical datasets. This example plots Kaplan-Meier estimates of the survival functions for patients suffering from **Glioblastoma multiforme (GBM)**, divided into therapy type.

This can be reproduced in R with

```
GBM.clinical %>% rename(therapy = patient.drugs.drug.therapy_types.therapy_type) %>%
  . survivalTCGA(GBM.clinical, extract.cols = "therapy") %>%
  . filter(therapy != "immunotherapy") %>%
  . kmTCGA(explanatory.names = "therapy", xlim = c(0,750), break.time.by = 150,
  . risk.table.y.text = FALSE)
```

Survival Curves



Used R packages

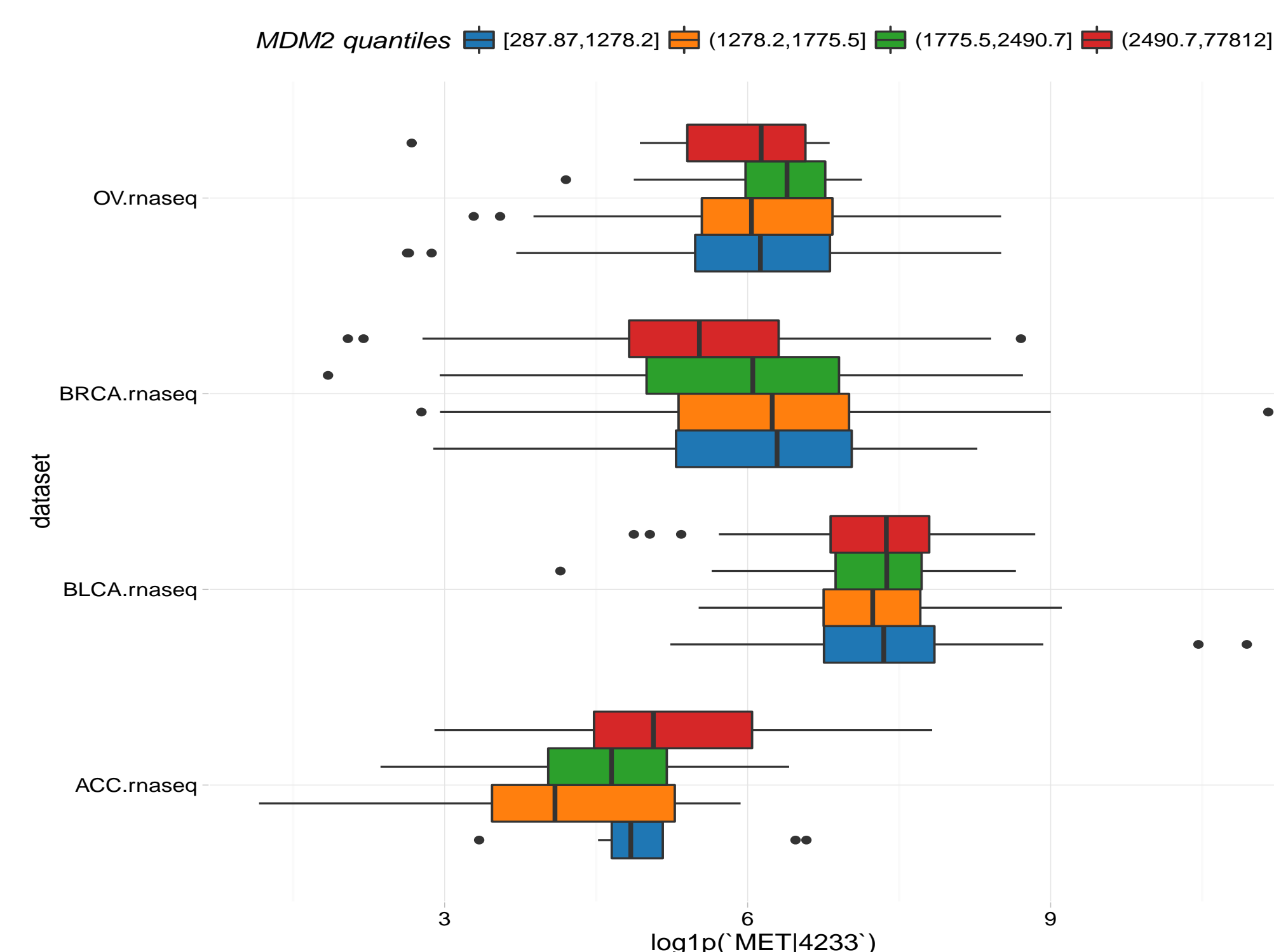
```
library(dplyr); library(RTCGA.clinical); library(RTCGA.rnaseq); library(RTCGA.mutations)
```

Boxplots of Genes' Expressions

RTCGA allows to create useful boxplot with few simple commands. Below is an example of boxplots for logarithm transformation for **MET** gene expression, divided on cancer types and 4 quartile levels of gene **MDM2**.

This can be reproduced in R with

```
expressionsTCGA(ACC.rnaseq, BLCA.rnaseq, BRCA.rnaseq, OV.rnaseq,
  . extract.cols = c("MDM2|4193", "MET|4233")) %>%
  . mutate(MDM2 = cut('MDM2|4193', include.lowest = TRUE, dig.lab = 5,
  . breaks = quantile('MDM2|4193', seq(0,1,0.25)))) %>%
  . boxplotTCGA(x = "dataset", y = "log1p('MET|4233')", fill = "MDM2",
  . legend.title = "MDM2 quantiles")
```



The Principal Components for RNASeq

It is simple to perform Principal Components Analysis of the RNAseq data for various cancer types and to plot 2 main components. Genes' expressions vary among **Adrenocortical carcinoma (ACC)**, **Cholangiocarcinoma (CHOL)**, **Glioma (GBM)**, **Pheochromocytoma + Paranganglioma (PCPG)** and **Uveal Melanoma (UVM)**.

This can be reproduced in R with

```
expressionsTCGA(ACC.rnaseq, CHOL.rnaseq, GBM.rnaseq, PCPG.rnaseq, UVM.rnaseq) %>%
  . pcaTCGA(group.names = "dataset")
```

