Ruprecht-Karls-Universität Heidelberg

Fakultät für Biowissenschaften

Bachelorstudiengang Molekulare Biotechnologie

Cancer Hallmark and Metabolic Pathways differ over Cancer types and in Prostate Adenocarcinoma patients

Data Science Project SoSe 2022

11 Juli 2022

Fabian Strobel, Lottida Phondeth, Laura Lange, Carla Welz

Contents

1 Introduction

1.1 vfbdjns

1.2 jutzhgrfedw

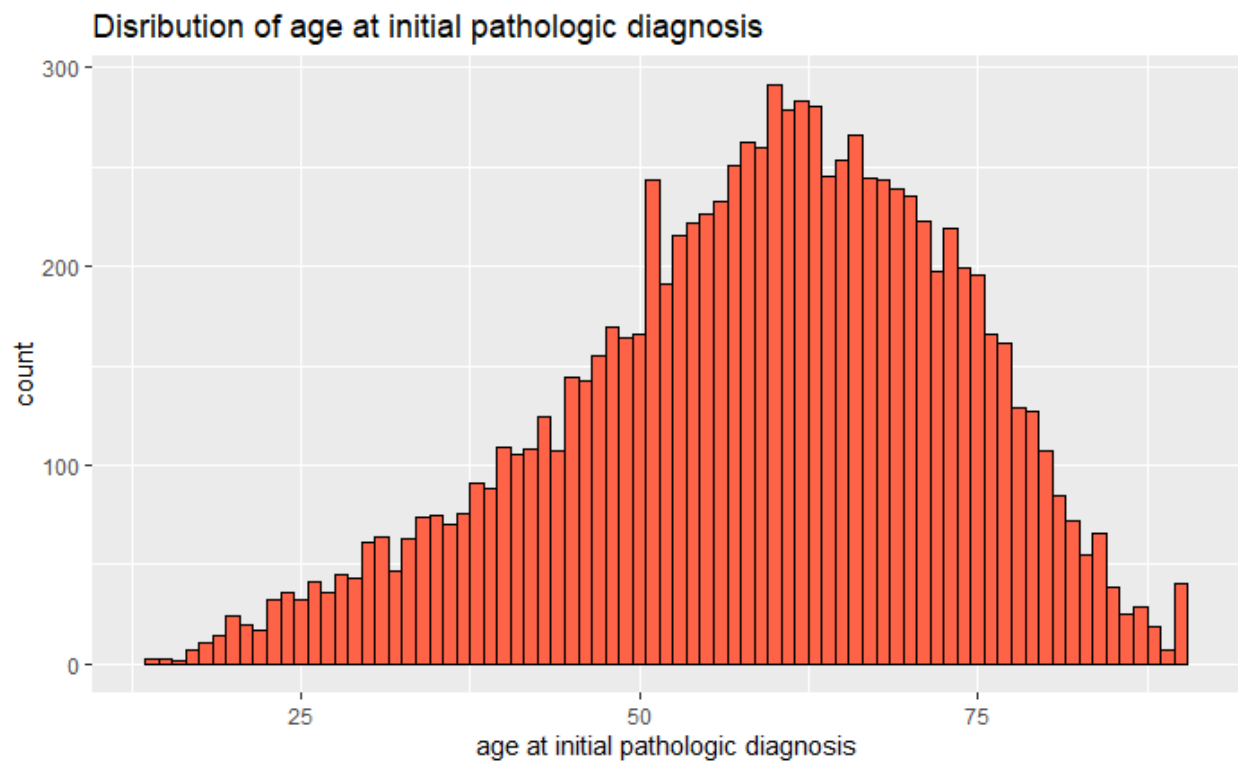Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R. (**10.3389/fmolb.2020.00053?**,)
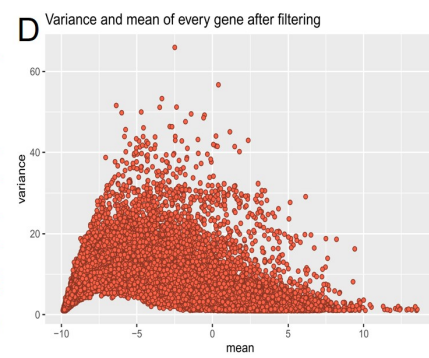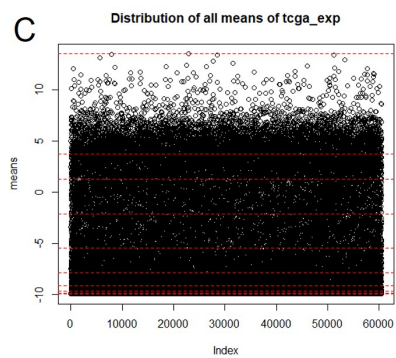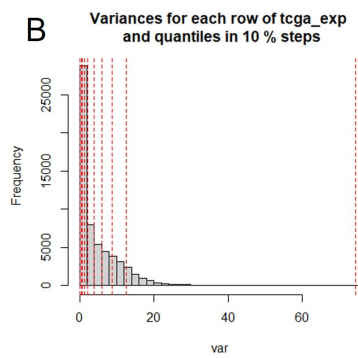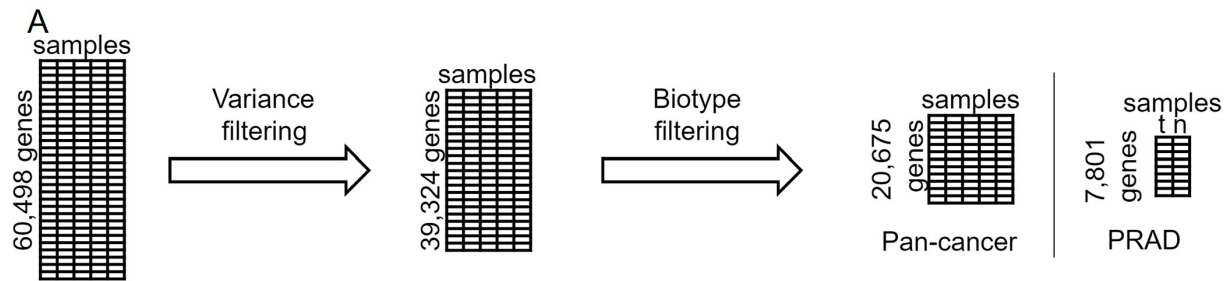

Figure 1: Caption for the picture.

2 Methods

2.1 Initial raw data

During our project, we used four given data sets. The first was an R-object consisting of a list of gene sets for cancer hallmarks. Second, a pan-cancer RNA-seq gene expression data frame for 9,741 patients of 33 various cancer types based on data generated by the "The Cancer Genome Atlas" Research Network: https://www.cancer.gov/tcga. In addition, there was an R-object containing 37 clinical annotations regarding the RNA-seq patients. And fourth, for a focused analysis of PRAD, an R-object with RNA-seq gene expression data of matched tumor tissue and normal tissue of 52 PRAD patients was used.

2.2 Preprocessing

The RNA-seq data came in a log2(TPM) format which served as normalization technique. The original pan-cancer data frame, which contained 60,498 genes, was preprocessed as follows (Figure 1A): After confirming the absence of missing values, the means and variances for all genes were calculated (Figure 1B, C). To remove rather constant genes across all cancer types, variance filtering was performed, were all genes with a variance below the 35 % quantile were discarded (Figure 1D).

A samples

60,498 genes

Variance filtering →

39,324 genes samples

Biotype filtering →

20,675 genes samples

Pan-cancer

7,801 genes samples t h

PRAD

B Variances for each row of tcga_exp and quantiles in 10 % steps

Frequency

var

C Distribution of all means of tcga_exp

means

Index

D Variance and mean of every gene after filtering

variance

mean

3   Results

3.1   uztrewertzu

3.2   kujzhtgrfe

4   Discussion

4.1   kjhtgrfedws

4.2   uztrwwret

5   Refrences