



R Shiny based interface for multiple testing

Cheonnam National University(Republic of Korea), Juseong Park

Participants

Presenter : Juseong Park

Authors :

Juseong Park (CNU, Master's course of Statistics)

Shinjune Kim (KREI)

Youngjae Oh (Korea Statistics)

Corresponding author : Jaesik Jeong (CNU, Prof. of Statistics)

Abstract

- To find the difference between metabolites in two groups in bioinformatics, multiple tests should be used, and one of the corresponding methods is to use FDR (False Discovery Rate) Control Methods.
- In order to obtain significant differences using the FDR Control Method in the field of metabolic data research, it is necessary to seek help from people in the field of statistics or to study statistics directly. This can be thought of as costly in some ways. We hope that these costs will be reduced with the R Shiny interface we created.
- Let's simulate the Schisandra data (mixed Chinese and Korean) with R Shiny interface for one representative example of 1D Method and 2D Method of FDR Control Method.

Background

Large-Scale Hypothesis table

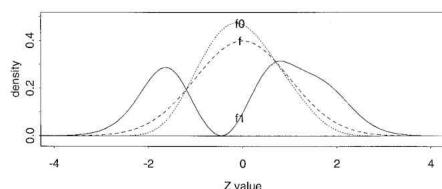
Unknown			Group		Observed				Unknown
Proportion	H_i	Actual	Normal	Abnormal	t_i	z_i	p_i	Decision	Density
π_1 Pr{non-null}	H_1	F	$\mu = \mu_{di}$ $\sigma = \sigma_{di}$	t_1	z_1	p_1		F	$f_1(z)$
	H_2	F		t_2	z_2	p_2		F	
	H_3	F		t_3	z_3	p_3		T	
	H_4	F		t_4	z_4	p_4		F	
	\vdots	\vdots		\vdots	\vdots	\vdots	\vdots	\vdots	
π_0 Pr{null}	$H_{N+\pi_1}$	T	$\mu = \mu_{ni}$ $\sigma = \sigma_{ni}$	$t_{N+\pi_1+1}$	$z_{N+\pi_1+1}$	$p_{N+\pi_1+1}$		T	$f_0(z)$
	$H_{N+\pi_1+2}$	T		$t_{N+\pi_1+2}$	$z_{N+\pi_1+2}$	$p_{N+\pi_1+2}$		T	
	\vdots	\vdots		\vdots	\vdots	\vdots	\vdots	\vdots	
	H_{N-2}	T		t_{N-2}	z_{N-2}	p_{N-2}		T	
	H_{N-1}	T		t_{N-1}	z_{N-1}	p_{N-1}		F	
	H_N	T		t_N	z_N	p_N		T	

Mixture dist. & Zero-Assumption

- As the assumption proposed by Bradley Efron, if the distribution of the entire data consists of the sum of the distributions of Null and Non-Null, it is assumed that only the probability of the distribution corresponding to Null exists near 0 (or mean) in the entire distribution

$$f(z) = \pi_0 f_0(z) + \pi_1 f_1(z) : \text{Mixture dist.}$$

$$f(z) = \pi_0 f_0(z) \text{ for } z \in Z = [c_1, c_2] : \text{Zero-Assumption}$$



- Efron et al. 2001 p.1155

- Example of Zero-Assumption

FDR Control Method

1D fdr : Bradley Efron et al.

- Estimate local fdr through empirical bayesian analysis : $fdr1d(z) = \pi_0 f_0(z) / f(z)$
- By estimating \hat{f} , \hat{f}_0 , $\hat{\pi}$ using Zero-Assumption, etc., $fdr1d(z)$ is estimated.
- There is also Ramos' method for 1D Method. (Description is omitted)

2D fdr : Alexander Ploner et al.

- Estimate two-dimensional local fdr (complementing fdr1d) : $fdr2d(z_1, z_2) = \pi_0 f_0(z_1, z_2) / f(z_1, z_2)$
- By estimating \hat{f} , \hat{f}_0 , $\hat{\pi}$, $fdr2d(z_1, z_2)$ is estimated. (Use t-value and variance as statistics)
- There is also Kim's method for 2D Method. (Description is omitted)

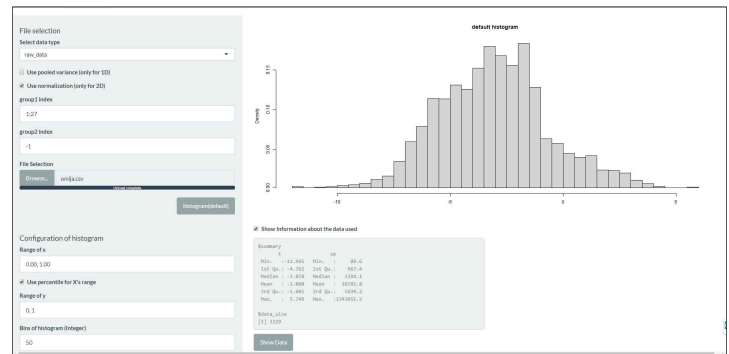
Use the R Shiny interface

- Data : schisandra data
- About 3200 gene intensities are matched by a mixture of 27 Chinese and 30 Korean data.
- We named **raw data** where all observations exist for each gene (feature).

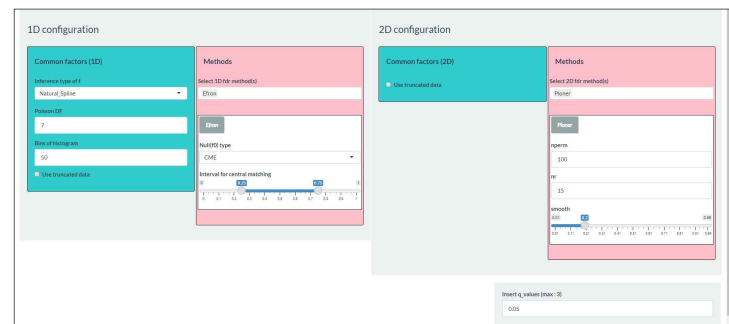
Steps to use the interface

- ① Data Load : Import data in csv file
- ② Method selection : Methodology and parameter setting to be analyzed
- ③ Result : Visualization information and significant gene (feature) search

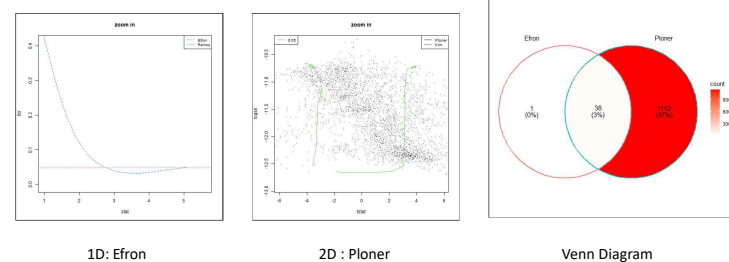
① Data Load



② Method selection



③ Result



Output : Plots and index of result.csv

Reference

- [1] Yoav Benjamini and Yoel Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological), 57(1):289–300, 1995.
- [2] Bradley Efron, Robert Tibshirani, John D Storey, and Virginia Tusher. Empirical bayes analysis of a microarray experiment. Journal of the American statistical association, 96(456):1151–1160, 2001.
- [3] Alexander Ploner, Stefano Calza, Arief Gusnanto, Yudi Pawitan. Bioinformatics, Volume 22, Issue 5, 1 March 2006, Pages 556–565

In RShiny, you can make the created app accessible through the Internet.
Here is my link QRcode. (It may disappear later)



This research was supported by the BK21 FOUR (Fostering Outstanding Universities for Research, NO.5120200913674) funded by the Ministry of Education(MOE, Korea) and National Research Foundation of Korea(NRF)