

A Practitioner’s Guide to Multiple Testing Error Rates

Jonathan Rosenblatt
Department of Statistics and Operations Research,
The Sackler Faculty of Exact Sciences,
Tel Aviv University
Israel

April 12, 2014

1 Introduction

Consider the problem of comparing two lists of same length. In the following, we will assume these lists consist of genes. A possible way of comparison would be to test the hypothesis that the lists’ ordering is arbitrary, versus some type of agreement. This is the purpose of this package, based on the work of Plaisier et al. [2010].

The proposed approach is to count the number of common genes in the first i and j elements of the first and second list respectively. As the count of common elements could be driven by chance, the significance of the observed count is computed assuming the hypothesis of completely random list orderings. As this is performed for all i and j , multiplicity kicks in.

The package offers both FWER control ¹ using permutation testing and FDR control using the B-Y procedure [Benjamini and Yekutieli, 2001] as proposed in the original work by Plaisier et al. [2010].

Remark 1.1. *FDR or FWER?*

Plaisier et al. [2010] recommend the control of the FDR over the different i and j s. Each i, j combination tests the null hypothesis of “arbitrary rankings of the two lists”, versus an alternative of “non-arbitrary ranking in the first i and j elements of the first and second list respectively”. FDR control is thus appropriate if concerned with the number of false i, j statements made.

¹For a general introduction to multiple testing error rates, see [Rosenblatt, 2013].

If only concerned with the existence of any non-arbitrariness, without claiming at which part of the lists it resides, then FWER control is more appropriate.

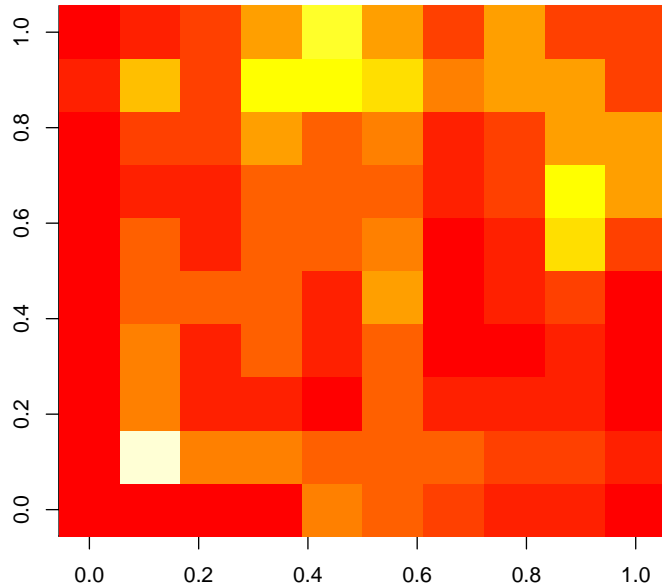
2 Workflow

We start with a sketch of the workflow. The details follow.

- Compute the significance of the gene overlap for all i and j first elements of the two lists.
- Correct the nominal significance levels for the multiple i s and j s.
- Report finding using the exported significance matrices and accompanying Venn diagrams.

```
> library(RRHO)
> # Create "gene" lists:
> list.length <- 100
> list.names <- paste('Gene',1:list.length, sep='')
> gene.list1<- data.frame(list.names, sample(100))
> gene.list2<- data.frame(list.names, sample(100))
> # Compute overlap and significance
> RRHO.example <- RRHO(gene.list1, gene.list2, BY=TRUE)

> # Examine Nominal (-log) pvalues
> image(RRHO.example$hypermat)
> # Note: If lattice is available try: levelplot(RRHO.example$hypermat)
```



```

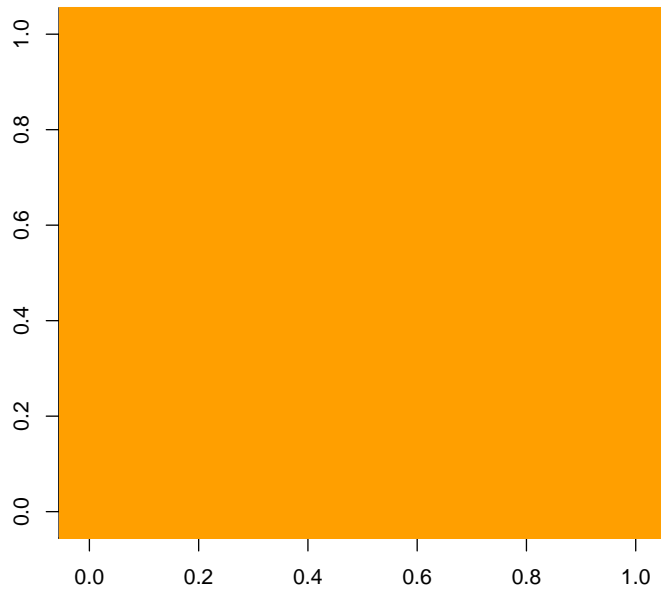
> pval.testing <- pvalRRHO(RRHO.example,50)
> # FWER corrected pvalue:
> pval.testing$pval

[1] 0.72

> # The sampling distribution of the minimum of the (-log) nominal p-values:
> xs<- seq(0,10,length=100)
> plot(Vectorize(pval.testing$FUN.ecdf)(xs)~xs, xlab='-log(pvalue)', ylab='ECDF', typ

> # Examine B-Y corrected pvalues
> # Note: probably nothing will be rejected in this example as the data is generated
> image(RRHO.example$hypermat.by)

```



References

- Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *ANNALS OF STATISTICS*, 29(4): 1165–1188, 2001.
- Seema B. Plaisier, Richard Taschereau, Justin A. Wong, and Thomas G. Graeber. Rank–rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Research*, 38(17):e169–e169, September 2010. ISSN 0305-1048, 1362-4962.
- Jonathan Rosenblatt. A practitioner’s guide to multiple testing error rates. arXiv e-print 1304.4920, April 2013. URL <http://arxiv.org/abs/1304.4920>.