gsmaRt v. 1.0 Package Vignette

Stephan Artmann <stephan-artmann@gmx.net>,
Mathias Fuchs

Göttingen 2011

Contents

1 Introduction

The 'gsmaRt' package (Gene Set Microarray Analysis) intends to bring together different functionalities for high-throughput gene set testing. This is based on random sampling from rotation data, as done for example by (Wu *et˜al.*, 2010). The test statistic can either be a two-sample Wilcoxon Test or a reimplementation of the test statistic described in (Goeman *et˜al.*, 2004). Multiple hypothesis testing correction is done as in (Tian *et˜al.*, 2005). A more detailed description shall soon be available (Fuchs and Artmann, pre.).

2 Usage

An example on how to use gsmaRt is given here.

In this synthetic experiment, 8 microarray replicates are present with three gene sets on each. Additionally, we need a corresponding matrix Y for mRNAs. Here we assume we have 20 mRNAs and 10 microarray replicates:

```
> #######################################
> ### Generate random expression data ###
> #######################################
> # Generate random mRNA expression data with 20 mRNAs
> # and 10 replicates
> Y = rnorm(200);
> dim(Y) = c(20,10);
> rownames(Y) = 1:20;
```

We use a rather simple phenotype vector.

```
> # Lets assume that the corresponding mRNA experiments had the following design:
> groups = (c(1,1,1,1,1,2,2,2,2,2));
```

Next, we need to know which gene is in which gene set, which is represented as a two-column data.frame, with genes in its first, gene sets in its second column.

```
> #####################
> ### Perform Test ###
> #####################
> library(gsmaRt)
> #Let gene set 1 contain mRNAs 1 to 9 and gene set 2 contain mRNAs 10 to 17.
> # Genes 18 to 20 are not contained anywhere. Gene set 3 is empty.
> miR = c(rep(1,9),c(rep(2,8)));
> mRNAs = 1:17;
> A = data.frame(mRNAs,miR); # Note that the gene sets MUST be in the second colum
> A

   mRNAs miR
1      1   1
2      2   1
3      3   1
4      4   1
5      5   1
6      6   1
7      7   1
8      8   1
9      9   1
10    10   2
11    11   2
12    12   2
13    13   2
14    14   2
15    15   2
16    16   2
17    17   2
```

Finally, the function 'analyse.gsmart' is called which does the testing.

```
> set.seed(1)
> P = analyse.gsmart(Y,groups,A)

[1] 1
[1] 2
```

```
[1] 1001    2
[1] "Estimated portion of true nulls = 0.0596988578416162"
[1] 1001    2
[1] "Estimated portion of true nulls = 1"

> P


$W
$W$tt
       orig marg.p.value   q.value q.least_mt_maj q.gr_mt_maj   ExpNrFD
2 0.1352504    0.7442557 0.04441595     0.04763969  0.04441595 0.04441595
1 1.6356409    0.5964036 0.04763969     0.04763969  0.04441595 0.09527938
  NrDiscoveries BH.q.value BY.q.value
2             1  0.7442557          1
1             2  0.7442557          1


$W$ttu
       orig marg.p.value   q.value q.least_mt_maj q.gr_mt_maj   ExpNrFD
1 1.6356409    0.5964036 0.04763969     0.04763969  0.04441595 0.09527938
2 0.1352504    0.7442557 0.04441595     0.04763969  0.04441595 0.04441595
  NrDiscoveries BH.q.value BY.q.value
1             2  0.7442557          1
2             1  0.7442557          1


$W$pi0
[1] 0.05969886



$G
$G$tt
       orig marg.p.value q.value q.least_mt_maj q.gr_mt_maj ExpNrFD
1 0.7332738    0.4335664   0.807          0.807       0.807   0.807
2 1.8392395    0.9530470   0.975          0.975       0.975   1.950
  NrDiscoveries BH.q.value BY.q.value
1             1  0.8671329          1
2             2  0.9530470          1


$G$ttu
       orig marg.p.value q.value q.least_mt_maj q.gr_mt_maj ExpNrFD
1 0.7332738    0.4335664   0.807          0.807       0.807   0.807
2 1.8392395    0.9530470   0.975          0.975       0.975   1.950
  NrDiscoveries BH.q.value BY.q.value
1             1  0.8671329          1
2             2  0.9530470          1
```

```
$G$pi0
[1] 1

>
>
```

References

Fuchs, M. and Artmann, S. gsmaRt: Gene Set Microarray Testing. *In preparation*

Goeman, J.~J., van~de Geer, S.~A., de~Kort, F., and van Houwelingen, H.~C. (2004). A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics*, **20**(1), 93 –99.

Tian L, Greenberg SA, Kong SW, Altschuler J, Kohane IS, Park PJ (2005). Discovering statistically significant pathways in expression profiling studies. *PNAS*, **102**(13544-13549).

Wu, D., Lim, E., Vaillant, F., Asselin-Labat, M., Visvader, J.~E., and Smyth, G.~K. (2010). ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics (Oxford, England)*, **26**(17).