

quasispectral

Dirk Moore

2019-05-13

Using quasispectral

To download the library, first install the “devtools” package, which will allow you to install packages from github; this only needs to be done once:

```
install.packages("devtools")
```

Next, install the `edgeR` package. To do this, you will need to first install the “biocLite” package, which allows you to access code from bioconductor:

```
source("https://bioconductor.org/biocLite.R")
```

Now you can get `edgeR` as follows:

```
biocLite("edgeR")
```

When asked, allow R to compile and install the package. This will take a few minutes.

Next, install the `fdrtools` package, which computes *q*-values:

```
`install.packages("fdrtools")`
```

Then install the `quasispectral` package from github:

```
devtools::install_github("mooredf22/quasispectral")
```

This will make the functions and data in the package available to you in R.

Next, load the libraries; the libraries `QuasiSeq` and `fdrtool` must have been previously installed from CRAN.

```
library(QuasiSeq)
library(fdrtool)
library(quasispectral)
```

The first few lines of the data may be examined:

```
head(spectralData_CLN1_late)
#>   geneName CLN1.1 CLN1.2 CLN1.3 CLN1.4 CLN1.5 CLN1.6 wt.1 wt.2 wt.3 wt.4
#> 1     Aga    149    142     67    135    150    142   135   155   159   202
#> 2    Arsa    577    542    314    601    569    590   602   639   617   599
#> 3    Arsb    466    482    258    409    465    465   624   599   645   688
#> 4    Arsg     68     65     41     55     69     63    87    84    46    98
#> 5    Arsk     59     60     22     39     45     54    75    75    80   107
#> 6   Asah1    806    905    533    729    748    827   865   812   887   861
```

```
#>   wt.5 wt.6
#> 1  195 181
#> 2  613 631
#> 3  662 667
#> 4  107  99
#> 5   77  79
#> 6  918 904
```

The first column is the gene name, the next 6 columns are spectral counts for mutant animals, and the final 6 columns are counts for wildtype animals.

Finally, call the package as follows, and look at the first few lines of output:

```
result.QLfit <- quasispectral(spectralDataAll=spectralData_CLN1_late, n.mut=6, n.wt=6)
#> Currently, bias reduction has not yet been implemented for Poisson model.
#> Spline scaling factor: 0.983018553187791
#> Warning in fdrtool(p.values, statistic = "pvalue"): There may be too few
#> input test statistics for reliable FDR calculations!
#> Step 1... determine cutoff point
#> Step 2... estimate parameters of null distribution and eta0
#> Warning: Censored sample for null model estimation has only size 8 !
#> Step 3... compute p-values and estimate empirical PDF/CDF
#> Step 4... compute q-values and local fdr
#> Step 5... prepare for plotting
```

```
QLfit.out <- result.QLfit$QLfit
head(QLfit.out)
#>   geneName CLN1.1 CLN1.2 CLN1.3 CLN1.4 CLN1.5 CLN1.6 wt.1 wt.2 wt.3 wt.4
#> 1    Aga    149    142     67    135    150    142    135    155    159    202
#> 2   Arsa    577    542    314    601    569    590    602    639    617    599
#> 3   Arsb    466    482    258    409    465    465    624    599    645    688
#> 4   Arsg     68     65     41     55     69     63     87     84     46     98
#> 5   Arsk     59     60     22     39     45     54     75     75     80    107
#> 6  Asah1    806    905    533    729    748    827    865    812    887    861
#>   wt.5 wt.6 coef.main      p.values p.values.bonf p.values.holm
#> 1  195 181 0.3129084 1.254920e-03 1.016486e-01 4.141237e-02
#> 2  613 631 0.1382393 5.303151e-03 4.295553e-01 1.198925e-01
#> 3  662 667 0.5354857 4.922626e-11 3.987327e-09 3.790422e-09
#> 4  107  99 0.4545213 3.430625e-03 2.778806e-01 8.919292e-02
#> 5   77  79 0.7465593 6.204245e-06 5.025439e-04 2.978038e-04
#> 6  918 904 0.1314973 1.497866e-02 1.000000e+00 2.845946e-01
#>   q.values.BH q.values.strimmer
#> 1 2.074460e-03 2.914893e-04
#> 2 7.159254e-03 1.029129e-03
#> 3 7.974655e-10 1.120724e-10
#> 4 4.875099e-03 7.007858e-04
```

```
#> 5 1.478070e-05      2.104575e-06  
#> 6 1.925828e-02      2.698284e-03
```

The first columns give the gene names and spectral count data that were input into the `quasispectral` function. The column “coef.main” is the log2 ratio of the estimated mean for wild type animals compared to the estimated mean for mutant animals.

To see the estimated proportion of null samples in the population, from `fdrtool`, look at `eta0`:

```
result.QLfit$eta0      # estimated proportion of null samples in population  
#> [1] 0.143748
```