Usage of MODA

Dong Li
dxl466@cs.bham.ac.uk
School of Computer Science, The University of Birmingham, UK

Date modified: 2016-10-17

In this example we embed parts of the examples from the MODA help page into a single document.

1 Module detection

First of all we conduct the experiment on the synthetic dataset which contains two expression profiles \( datExpr_1 \) and \( datExpr_2 \) with 500 genes, and each has 20 and 25 samples. Details of data generation can be found in supplementary file of MODA paper [1]. Basic module detection functions are provided by WGCNA [2].

```r
library(MODA)
##
data(synthetic)
ResultFolder = 'ForSynthetic' # where middle files are stored
CuttingCriterion = 'Density' # could be Density or Modularity
indicator1 = 'X' # indicator for data profile 1
indicator2 = 'Y' # indicator for data profile 2
specificTheta = 0.1 #threshold to define condition specific modules
conservedTheta = 0.1 #threshold to define conserved modules
##modules detection for network 1
intModules1 <- WeightedModulePartitionDensity(datExpr1,ResultFolder,
                                           indicator1,CuttingCriterion)
## ..done.
##modules detection for network 2
intModules2 <- WeightedModulePartitionDensity(datExpr2,ResultFolder,
                                           indicator2,CuttingCriterion)
## ..done.
```

which shows how to detect modules using hierarchical clustering with the optimal cutting height of dendrogram. The heatmap of correlation matrix of gene expression profile 1 may looks like Figure 1. Another package [3] has the similar function.

The selection of optimal cutting height for each expression profile would be stored under directory \( ResultFolder \). Take \( datExpr_1 \) in the synthetic data for example, a file named \( Partitions_X.pdf \) may looks like Figure 2.

At the same time, each module for each expression profile would be stored as plain text file, with the name indicator from \( indicator_1 \) and \( indicator_2 \). Each secondary directory under \( ResultFolder \) has the same name of condition name, e.g \( indicator_2 \), used to store differential analysis results.
The condition-specific networks can be specified by two vectors if there are more. There are three files under the secondary directory named by condition name: two text files of them are condition specific and conserved modules id

2 Network comparison

After the module detection for background network and all condition-specific networks, we can compare them using following function

\[
\text{CompareAllNets(RESULTFolder, intModules1, indicator1, intModules2, indicator2, specificTheta, conservedTheta)}
\]

The condition specific networks can be specified by two vectors if there are more. There are three files under the secondary directory named by condition name: two text files of them are condition specific and conserved modules id
from background network, and one pdf for showing how to determine these modules by two parameters specific\textit{Theta} and conserved\textit{Theta} based on a Jaccard index matrix. Theoretical details can be found in supplementary file of MODA paper. The figure may looks like Figure 3.

### 3 Biological explanation

Finally we can do gene annotation enrichment analysis with integrative tools like DAVID\textsuperscript{1} or Enrichr\textsuperscript{2}, to see whether a module gene list can be explained by existing biological process, pathways or even diseases.

### 4 Session info

- R version 3.3.1 (2016-06-21), x86_64-pc-linux-gnu
- Locale: LC_CTYPE=en_US.UTF-8, LC_NUMERIC=C, LC_TIME=en_US.UTF-8, LC_COLLATE=C, LC_MONETARY=en_US.UTF-8, LC_MESSAGES=en_US.UTF-8, LC_ADDRESS=C, LC_TELEPHONE=C, LC_MEASUREMENT=en_US.UTF-8, LC_IDENTIFICATION=C
- Base packages: base, datasets, grDevices, graphics, methods, stats, utils
- Other packages: MODA 1.0.0, knitr 1.14
- Loaded via a namespace (and not attached): AnnotationDbi 1.36.0, Biobase 2.34.0, BiocGenerics 0.20.0, BiocStyle 2.2.0, DBI 0.5-1, Formula 1.2-1, GO.db 3.4.0, Hmisc 3.17-4, IRanges 2.8.0, Matrix 1.2-7.1, RColorBrewer 1.1-2, RSQLite 1.0.0, Rcpp 0.12.7, S4Vectors 0.12.0, WGCNA 1.51, acepack 1.3-3.3, chron 2.3-47, cluster 2.0.5, codetools 0.2-15, colorspace 1.2-7, data.table 1.9.6, doParallel 1.0.10, dynamicTreeCut 1.63-1, evaluate 0.10, fastcluster 1.1.21, foreach 1.4.3, foreign 0.8-67, formatR 1.4, ggplot2 2.1.0, grid 3.3.1, gridExtra 2.2.1, gtable 0.2.0, igraph 1.0.1, impute 1.48.0, iterators 1.0.8, lattice 0.20-34, latticeExtra 0.6-28, magrittr 1.5, matrixStats 0.51.0, munsell 0.4.3, nnet 7.3-12, parallel 3.3.1, plyr 1.8.4, preprocessCore 1.36.0, rpart 4.1-10, scales 0.4.0, splines 3.3.1, stats4 3.3.1, stringi 1.1.2, stringr 1.1.0, survival 2.39-5, tools 3.3.1

\textsuperscript{1}https://david.ncifcrf.gov

\textsuperscript{2}http://amp.pharm.mssm.edu/Enrichr
References

