

Using BiomarkerR for identifying biomarker candidates and inferring networks

Michael Netzer

December 5, 2011

Contents

1	Overview	1
2	Data import	1
3	Paired Biomarker Identifier	1
4	Unpaired Biomarker Identifier	2
5	Infer network	3

1 Overview

Section 2 describes the required dataset structure and will give you an idea how to import an already existing example dataset. Sections 3 and 4 explain how to rank attributes based on the Paired and Unpaired Biomarker Identifier. Section 5 describes how to infer networks using pBI and uBI scores.

2 Data import

Datasets used in **BiomarkerR** assume rows as attributes and columns as samples (in order to easily handle ExpressionSets). The example dataset can be loaded by:

```
> library(BiomarkerR)
> data(BIdata)
```

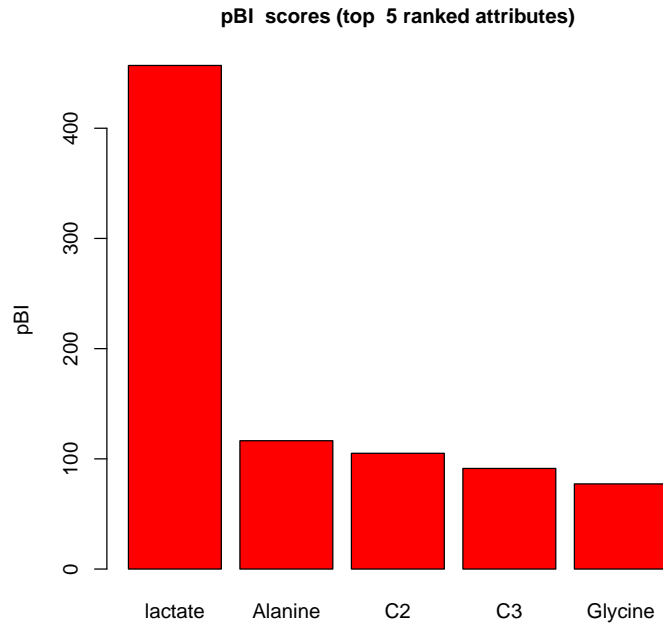
BIdata includes an unpaired dataset (ubi.data) with the associated class column (ubi.class). The paired dataset (pbi.data) comprises again the associated class column (pbi.class), and additionally the sample ids (pbi.id).

3 Paired Biomarker Identifier

The Paired Biomarker Identifier (pBI) [1] calculates a score for every attribute representing its discriminatory ability using dependent samples. Using our paired dataset the pBI scores are calculated as follows:

```
> pbi.scores <- pBI(dataset = pbi.data, classlabels = pbi.class,
+   referenceclasslabel = "rest", ids = pbi.id, useMedian = TRUE,
+   lambda = 100, plotScores = TRUE, numTopRankedToPlot = 5)
```

The resulting boxplot is



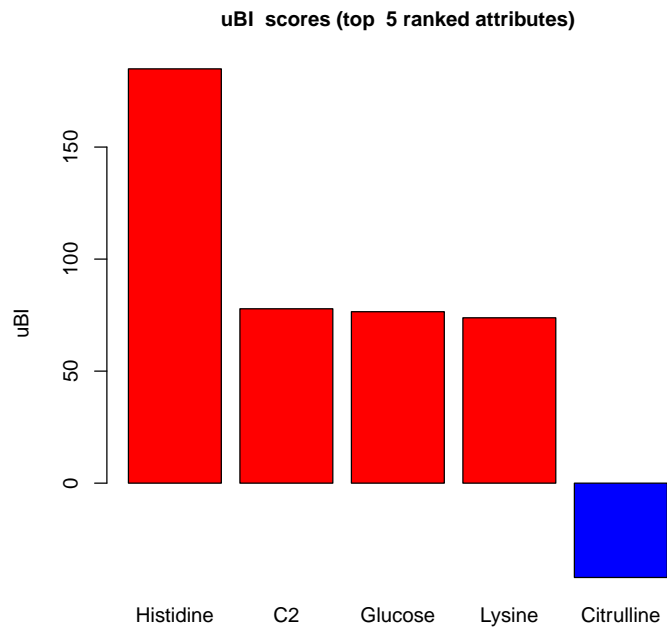
As reference class we use class "rest". Here we plot the top five ranked attributes (numTopRankedToPlot = 5).

4 Unpaired Biomarker Identifier

The Unpaired Biomarker Identifier (uBI) [1] calculates a score for every attribute representing its discriminatory ability using dependent samples. Using our unpaired dataset the uBI scores are calculated as follows:

```
> ubi.scores <- uBI(dataset = ubi.data, classlabels = ubi.class,
+   referenceclasslabel = "control", useMedian = TRUE, lambda = 100,
+   plotScores = TRUE, numTopRankedToPlot = 5)
```

The resulting boxplot is



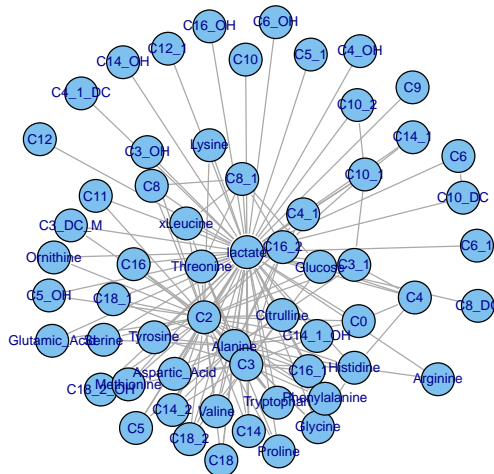
As reference class we use class "control". Here we plot the top five ranked attributes (`numTopRankedToPlot = 5`).

5 Infer network

Using our paired data we can infer and plot a network `g` by [4]:

```
> g <- pBIGraph(dataset = pbi.data, classlabels = pbi.class, referenceclasslabel = "rest",
+   ids = pbi.id, useMedian = TRUE, lambda = 100, threshold = "73",
+   plotGraph = FALSE, edge.file = NULL)
> plot(g, layout = layout.fruchterman.reingold)
```

The resulting plot is



If `edge.file` \neq NULL an edgfile will be created that can be imported in e.g. Cytoscape [2]. By default, the 90% percentile will be used as threshold for defining an edge. Finally, a graph will be created that can be further analyzed using e.g. the R package QuACN [3].

References

- [1] C~Baumgartner, GD~Lewis, M~Netzer, B~Pfeifer, and RE~Gerszten. A new data mining approach for profiling and categorizing kinetic patterns of metabolic biomarkers after myocardial injury. *Bioinformatics*, 26(14):1745–1751, Jul 2010.
- [2] MS~Cline, M~Smoot, E~Cerami, A~Kuchinsky, N~Landys, C~Workman, R~Christmas, I~Avila-Campilo, M~Creech, B~Gross, K~Hanspers, R~Isserlin, R~Kelley, S~Killcoyne, S~Lotia, S~Maere, J~Morris, K~Ono, V~Pavlovic, AR~Pico, A~Vailaya, PL~Wang, A~Adler, BR~Conklin, L~Hood, M~Kuiper, C~Sander, I~Schmulevich, B~Schwikowski, GJ~Warner, T~Ideker, and GD~Bader. Integration of biological networks and gene expression data using cytoscape. *Nat Protoc*, 2(10):2366–2382, 2007.
- [3] LAJ Mueller, KG~Kugler, A~Dander, A~Graber, and M~Dehmer. Quacn: an r package for analyzing complex biological networks quantitatively. *Bioinformatics*, 27(1):140–141, Jan 2011.
- [4] M~Netzer, KM~Weinberger, M~Handler, M~Seger, X~Fang, KG~Kugler, A~Graber, and C~Baumgartner. Profiling the human response to physical

exercise: A computational strategy for the identification and kinetic analysis of metabolic biomarkers. *J Clin Bioinforma*, accepted, 2011.