Detecting metabolite-transcript co-responses

Henning Redestig and Ivan G. Costa

March 25, 2011

1 Introduction

The trainmet package can be used to calculate association statistics (measures of co-responses) between metabolites and transcripts using combined profiling time-course experiments. An example of application is to find transcripts that behave similarly to the metablolites in a given pathway. With similar, directly similar or inverted response possibly at a time-lag. See Redestig and Costa (2011) for more detailed introduction.

1.1 Dependencies

You must have the script `train_rank.py` installed and working to be able to use the HMM; this package is only an interface to that script. The script provided with the package does not work out-of-the-box and must be installed separately. See instructions at[1].

You also need the `pls` package and are recommended to look into `KEGG.db` package and chip specific annotation packages from Bioconductor.

2 Example

As an example, we consider the galactose metabolism related metabolites measured in the 9 time-point high CO_2-stress dataset by [1]. We here demonstrate how to score transcripts after how strongly they co-respond to the glactose synthesis related metabolites.

First we get association statistics using classical Pearson correlation for all the glactose synthesis related metabolites, `mdat`, to all genes, `tdat`.

```
> library(trainmet)
> data(mdat)
> data(tdat)
> pearson <- assocStat(tdat, mdat, use = "pairwise")
```

for a three-state HMM we do (for this to work, you have to modify the path to `trank_rank.py` to fit your installation)

```
> hmm3 <- assocStat(tdat, mdat, "hmm", states = 3, trainrank = system.file("extdata",
+     "train_rank.py", package = "trainmet"))
```

then, we can summarize the scores to consensus statistics

```
> library(org.At.tair.db)
> members <- fData(tdat)$agi %in% get("00052", org.At.tairPATH2TAIR)
> pearsonS <- consStat(pearson, members, cv = TRUE)
> hmm3S <- consStat(hmm3, members, cv = TRUE)
> hmm3pearsonS <- consStat(cbind(hmm3, pearson), members, cv = TRUE)
```

```
> par(mfrow = c(1, 3))
> boxplot(pearsonS$ccthat ~ members, ylab = "Consensus Pearson",
+     xlab = "Transcript  in the galactose metabolism pathway")
> boxplot(hmm3S$ccthat ~ members, ylab = "Consensus HMM3", xlab = "Transcript  in the galactose m
> boxplot(hmm3pearsonS$ccthat ~ members, ylab = "Consensus HMM3+Pearson",
+     xlab = "Transcript  in the galactose metabolism cycle pathway")
```
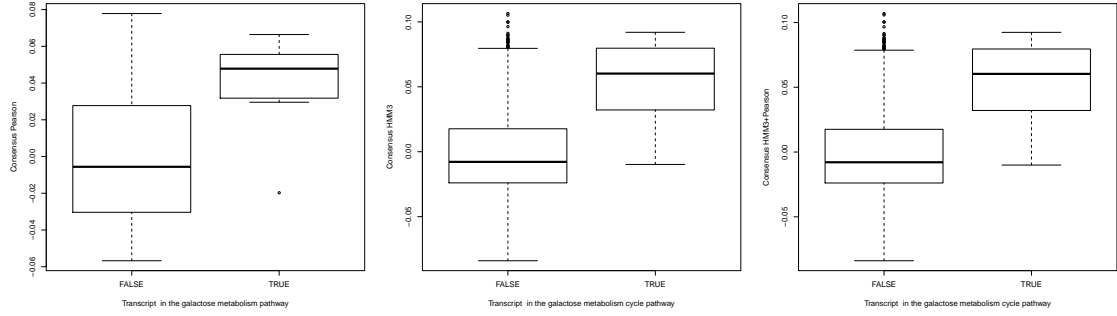

Figure 1:   In this example, HMM3 appears to give the best separation of the truly galactose metabolism associated transcripts from the other transcripts.

and then see how the differerent statistics score the true galactose metabolism members (Fig. 1).

The correlation loadings from the statistic-summarization gives a way to see which metabolite had the strongest relevance for separating the true galactose metabolism related genes (Fig. 2).

```
> par(mfrow = c(1, 3))
> barplot(pearsonS$ccp, las = 2, ylab = "Correlation loading")
> barplot(hmm3S$ccp, las = 2, ylab = "Correlation loading")
> barplot(hmm3pearsonS$ccp, las = 2, ylab = "Correlation loading")
```
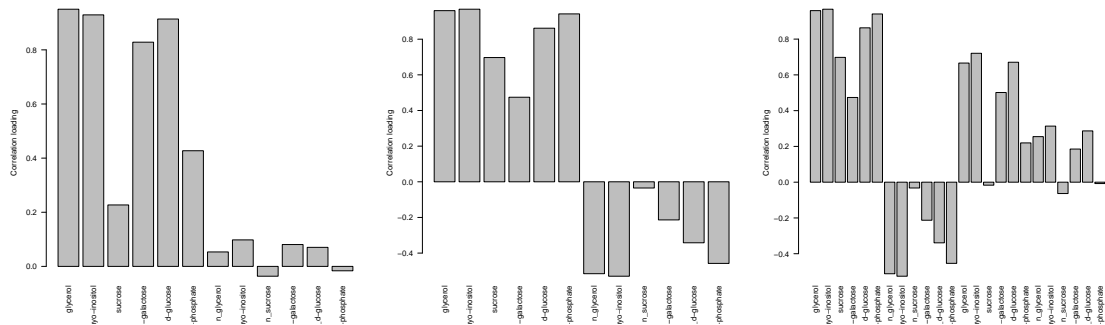

Figure 2:   The influence of different metabolites on the separation of truely glactose metabolism associated transcripts from other transcripts based. Note that since absolute values are considered, each metabolite is represented by bars, the positive mode and negative mod (prefixed with an n)

References

[1] Dutta, B., Kanani, H., Quackenbush, J., and Klapa, M. I. (2009). Time-series integrated omic analyses to elucidate short-term stress-induced responses in plant liquid cultures. Biotechnol

^{1}http://www.cin.ufpe.br/~igcf/Metabolites/scripts/hmm

```
> pearson <- assocStat(tdat, mdat, absolute = FALSE, use = "pairwise")
> colorLineplot(tdat, mdat, pearson, which(members), rows = 2,
+     cols = 3)
```
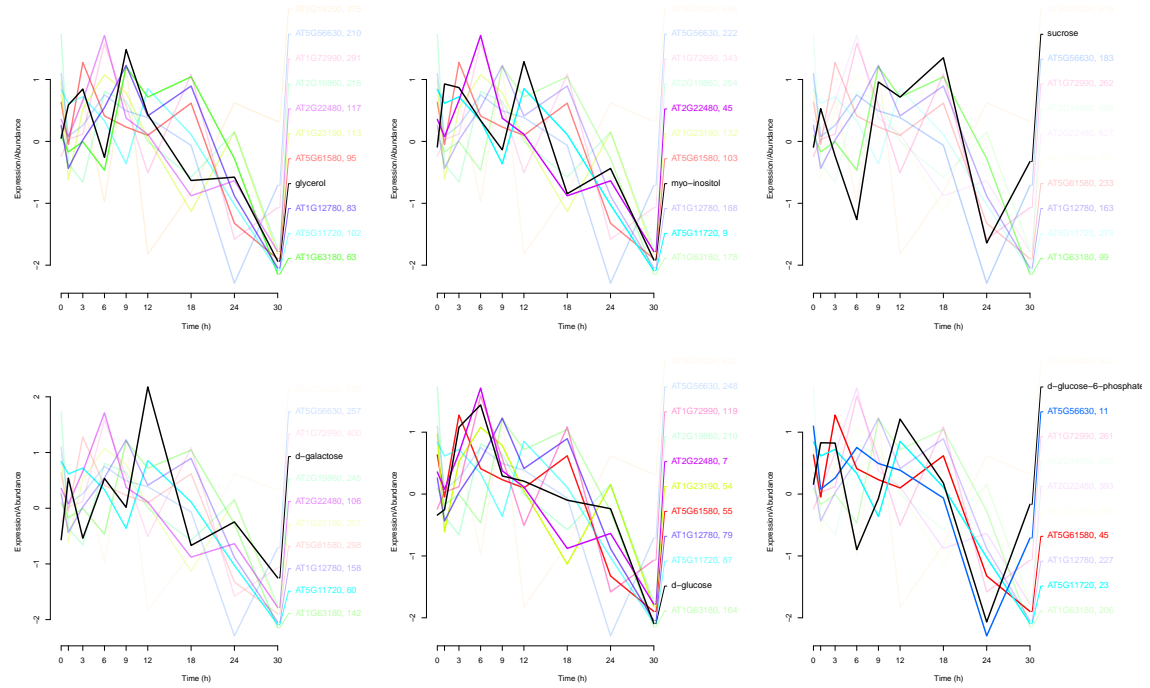

Figure 3: Line plot of the expression and abundance trajectories of galactose metabolism related genes and metabolites. Only transcripts in the pathway are plotted. Color intensity is proportional to the indicated rank (relative to the whole dataset).

Bioeng, 102(1), 264-279.

[2] Redestig, H. and Costa, I. G. (2011) Detection and interpretation of metabolite-transcript co-responses using combined profiling data. *Bioinformatics* submitted.