

tmod: Quickstart guide

January Weiner

2018-03-21

Abstract

The package `tmod` provides blood transcriptional modules described by Chaussabel et al. (2008) and by Li et al. (2014) as well as metabolic profiling clusters from Weiner et al. (2012). Furthermore, the package includes tools for testing the significance of enrichment of the modules as well as visualisation of the features (genes, metabolites etc.) and modules. This vignette is a quick start guide for the package.

For a more detailed manual, type “`tmodUserGuide()`”

This is a quick start guide. For a detailed manual, find the file “`tmodUserGuide.pdf`” in `tmod` documentation, or use the R command `tmodUserGuide()` to display the PDF file.

Basic data analysis

Loading an example data set

The following example shows a basic gene expression analysis and gene set enrichment analysis on the example expression data set included in `tmod`. The example is discussed in more detail in the `tmod` User Guide.

```
library(limma)
library(tmod)
data(Egambia)
design <- cbind(Intercept=rep(1, 30), TB=rep(c(0,1), each= 15))
E <- as.matrix(Egambia[, -c(1:3)])
fit <- eBayes( lmFit(E, design))
tt <- topTable(fit, coef=2, number=Inf,
  genelist=Egambia[, 1:3] )
```

Transcriptional module enrichment analysis

Hypergeometric test

```
fg <- tt$GENE_SYMBOL[tt$adj.P.Val < 0.05 & abs( tt$logFC ) > 1]
res <- tmodHGtest(fg=fg, bg=tt$GENE_SYMBOL)
res
```

##	ID	Title	b	B
## LI.M112.0	LI.M112.0	complement activation (I)	4	11
## LI.M11.0	LI.M11.0	enriched in monocytes (II)	4	20
## LI.M75	LI.M75	antiviral IFN signature	3	10
## LI.S4	LI.S4	Monocyte surface signature	3	10
## LI.S5	LI.S5	DC surface signature	4	34
## LI.M165	LI.M165	enriched in activated dendritic cells (II)	3	19
## LI.M4.3	LI.M4.3	myeloid cell enriched receptors and transporters	2	5
## LI.M16	LI.M16	TLR and inflammatory signaling	2	5

##	n	N	E	P.Value	adj.P.Val
## LI.M112.0	47	4826	37.3	2.48e-06	0.000858
## LI.M11.0	47	4826	20.5	3.41e-05	0.005907
## LI.M75	47	4826	30.8	9.91e-05	0.008569
## LI.S4	47	4826	30.8	9.91e-05	0.008569
## LI.S5	47	4826	12.1	2.96e-04	0.020465
## LI.M165	47	4826	16.2	7.52e-04	0.039413
## LI.M4.3	47	4826	41.1	9.11e-04	0.039413
## LI.M16	47	4826	41.1	9.11e-04	0.039413

The columns in the above table contain the following:

- **ID** The module ID. IDs starting with “LI” come from Li et al. (S. Li et al. 2014), while IDs starting with “DC” have been defined by Chaussabel et al. (Chaussabel et al. 2008).
- **Title** The module description
- **b** Number of genes from the given module in the fg set
- **B** Number of genes from the module in the bg set
- **n** Size of the fg set
- **N** Size of the bg set
- **E** Enrichment, calculated as $(b/n)/(B/N)$
- **P.Value** P-value from the hypergeometric test
- **adj.P.Val** P-value adjusted for multiple testing using the Benjamini-Hochberg correction

Mann-Whitney U test on an ordered list of genes

```
l <- tt$GENE_SYMBOL
res2 <- tmodUtest(l)
head(res2)
```

##	ID	Title	U	N1	AUC
## LI.M37.0	LI.M37.0	immune activation - generic cluster	352659	100	0.746
## LI.M37.1	LI.M37.1	enriched in neutrophils (I)	50280	12	0.870
## LI.S4	LI.S4	Monocyte surface signature	43220	10	0.897
## LI.M75	LI.M75	antiviral IFN signature	42996	10	0.893
## LI.M11.0	LI.M11.0	enriched in monocytes (II)	74652	20	0.777

```
## LI.M67      LI.M67      activated dendritic cells 28095  6 0.971
##              P.Value adj.P.Val
## LI.M37.0 1.60e-17 5.53e-15
## LI.M37.1 4.53e-06 6.57e-04
## LI.S4     6.85e-06 6.57e-04
## LI.M75    8.63e-06 6.57e-04
## LI.M11.0 9.49e-06 6.57e-04
## LI.M67    3.20e-05 1.81e-03
```

```
nrow(res2)
```

```
## [1] 25
```

The b, B, n, N and E columns in the output have been replaced by the following:

- **U** The Mann-Whitney U statistics
- **N1** Number of genes in the module
- **AUC** Area under curve – a measure of the effect size

CERNO test - a variant of Fisher's method for combining probabilities

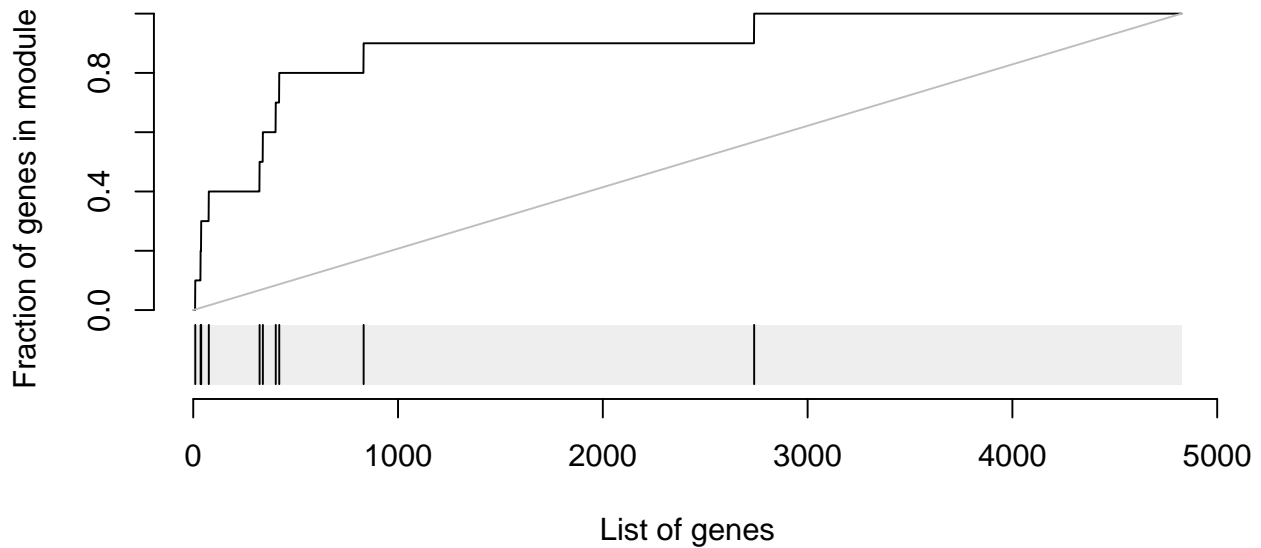
The CERNO test is actually much more practical than the U test for most purposes.

```
l <- tt$GENE_SYMBOL
res2 <- tmodCERNOtest(l)
head( res2 )
```

```
##              ID              Title cerno  N1  AUC
## LI.M37.0  LI.M37.0 immune activation - generic cluster 426.4 100 0.746
## LI.M11.0  LI.M11.0      enriched in monocytes (II) 113.8  20 0.777
## LI.S4     LI.S4      Monocyte surface signature  76.4  10 0.897
## LI.M112.0 LI.M112.0      complement activation (I)  73.7  11 0.846
## LI.M75    LI.M75      antiviral IFN signature  65.3  10 0.893
## LI.M16    LI.M16      TLR and inflammatory signaling 46.3   5 0.979
##              cES  P.Value adj.P.Val
## LI.M37.0  2.13 1.82e-18 6.31e-16
## LI.M11.0  2.85 5.26e-09 9.09e-07
## LI.S4     3.82 1.61e-08 1.85e-06
## LI.M112.0 3.35 1.72e-07 1.49e-05
## LI.M75    3.26 1.05e-06 7.19e-05
## LI.M16    4.63 1.25e-06 7.19e-05
```

Generating an evidence plot

```
evidencePlot(l, "LI.M75")
```



Working with multiple sets of comparisons

Working with limma

Generating a set of results from limma:

```
res.l <- tmodLimmaTest(fit, Egambia$GENE_SYMBOL)
length(res.l)
```

```
## [1] 2
```

```
names(res.l)
```

```
## [1] "Intercept" "TB"
```

```
head(res.l$TB)
```

```
##              ID              Title cerno  N1   AUC
## LI.M37.0  LI.M37.0 immune activation - generic cluster 414.3 100 0.726
## LI.M11.0  LI.M11.0      enriched in monocytes (II) 105.6  20 0.786
## LI.M112.0 LI.M112.0      complement activation (I)  75.6  11 0.867
## LI.S4      LI.S4      Monocyte surface signature  70.0  10 0.884
## LI.M75      LI.M75      antiviral IFN signature  66.1  10 0.865
## LI.M67      LI.M67      activated dendritic cells  50.4   6 0.971
##              cES  P.Value adj.P.Val
## LI.M37.0  2.07 4.57e-17 1.58e-14
## LI.M11.0  2.64 7.92e-08 9.67e-06
## LI.M112.0 3.44 8.39e-08 9.67e-06
## LI.S4      3.50 1.84e-07 1.59e-05
## LI.M75      3.31 7.78e-07 5.38e-05
```

```
## LI.M67      4.20 1.21e-06  6.97e-05
```

Comparing tests across experimental conditions

Generate a summary object from a list containing tmod results:

```
head(tmodSummary(res.l), 5)
```

##	ID	Title	AUC	Intercept
## LI.M11.0	LI.M11.0	enriched in monocytes (II)	0.815	
## LI.M112.0	LI.M112.0	complement activation (I)	NA	
## LI.M118.0	LI.M118.0	enriched in monocytes (IV)	NA	
## LI.M124	LI.M124	enriched in membrane proteins	0.881	
## LI.M127	LI.M127	type I interferon response	NA	
##	q.Intercept	AUC.TB	q.TB	
## LI.M11.0	0.000114	0.786	9.67e-06	
## LI.M112.0	NA	0.867	9.67e-06	
## LI.M118.0	NA	0.838	2.85e-03	
## LI.M124	0.011487	NA	NA	
## LI.M127	NA	0.945	1.04e-02	

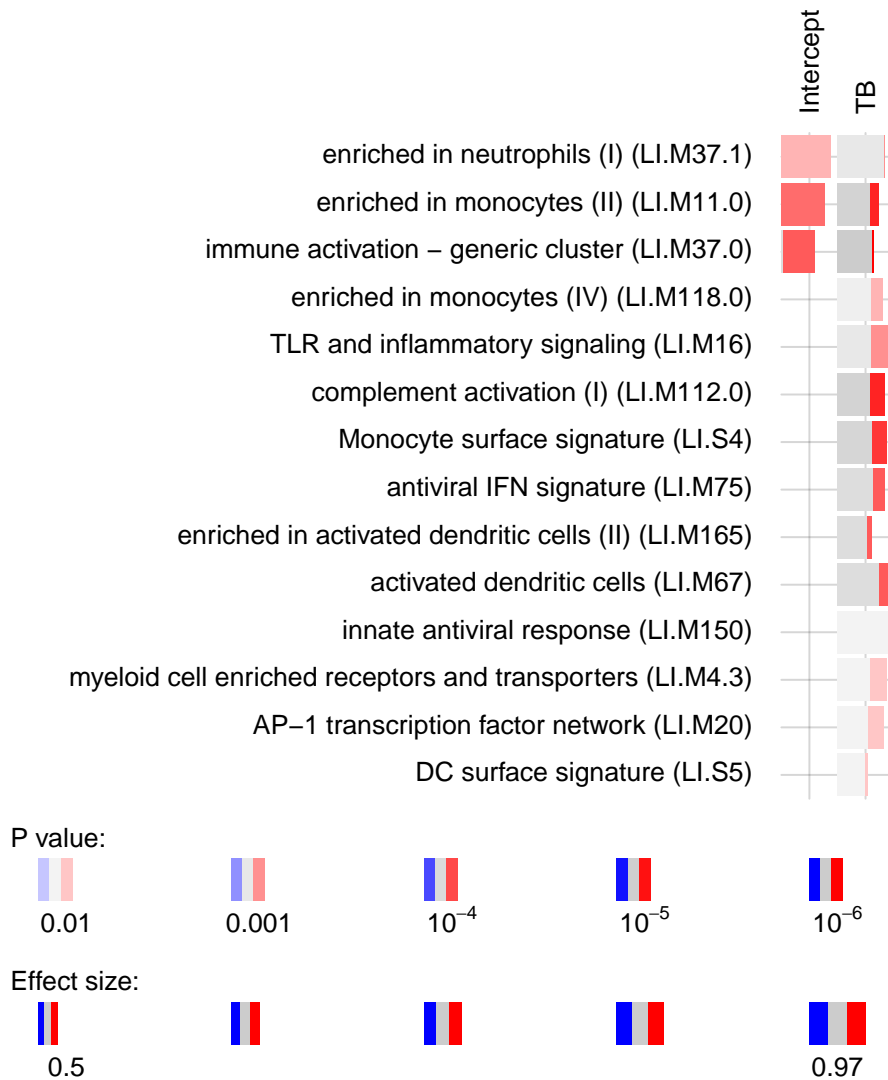
Show a list of tmod results in a heatmap-like visualization which indicates effect sizes as well as p-values:

```
tmodPanelPlot(res.l, text.cex=0.8)
```



Include information about which genes are up- and which are down-regulated:

```
pie <- tmodLimmaDecideTests(fit, genes=Egambia$GENE_SYMBOL)
tmodPanelPlot(res.l, pie=pie, text.cex=0.8, pie.style="rug")
```



Using `tmodDecideTests` for the same purpose:

```
tt.I <-
  topTable(fit, coef="Intercept", number=Inf, sort.by="n")
tt.TB <- topTable(fit, coef="TB", number=Inf, sort.by="n")
pie2 <- tmodDecideTests(Egambia$GENE_SYMBOL,
  lfc=cbind(tt.I$logFC, tt.TB$logFC),
  pval=cbind(tt.I$adj.P.Val, tt.TB$adj.P.Val))
identical(pie[[1]], pie2[[1]])
```

```
## [1] TRUE
```

Using and creating custom sets of modules

A minimal definition of a module set:

```

mymset <- makeTmod(
  modules=data.frame(ID=c("A", "B"),
    Title=c("A title",
      "B title")),
  modules2genes=list(
    A=c("G1", "G2"),
    B=c("G3", "G4"))
)
mymset

```

```

## An object of class "tmod"
## 2 modules, 4 genes

```

MSigDB

Download the MSigDB in XML format. This file can be accessed at the URL http://software.broadinstitute.org/gsea/msigdb/download_file.jsp?filePath=/resources/msigdb/6.1/msigdb_v6.1.xml – follow the link, register and log in, and save the file on your disk (roughly 113 MB).

```

msig <- tmodImportMSigDB("msigdb_v6.1.xml")

```

```

## 8430 modules, 32233 genes

```

Using only Hallmark (H) gene sets from MSigDB :

```

sel <- msig$MODULES$Category == "H"
tmodCERNOtest(tt$GENE_SYMBOL, mset=msig[sel] )

```

```

##           ID           Title
## M5913 M5913      Hallmark interferon gamma response
## M5921 M5921                Hallmark complement
## M5911 M5911      Hallmark interferon alpha response
## M5946 M5946                Hallmark coagulation
## M5890 M5890      Hallmark tnfa signaling via nfkb
## M5930 M5930 Hallmark epithelial mesenchymal transition
## M5932 M5932                Hallmark inflammatory response
## M5953 M5953                Hallmark kras signaling up
## M5892 M5892      Hallmark cholesterol homeostasis
##           cerno N1      AUC      cES      P.Value
## M5913 221.68317 41 0.7786936 2.703453 8.505170e-15
## M5921 217.81028 56 0.6979148 1.944735 8.607634e-09
## M5911 108.39559 20 0.7563566 2.709890 3.192325e-08
## M5946 179.24580 50 0.6779481 1.792458 1.966824e-06
## M5890 148.95123 47 0.6484665 1.584588 2.657694e-04

```



```

## M5930 212.53461 73 0.6371808 1.455717 2.701053e-04
## M5932 184.53035 62 0.6206393 1.488148 3.457724e-04
## M5953 221.76208 82 0.6046637 1.352208 1.790956e-03
## M5892 49.14641 14 0.6138968 1.755229 8.040562e-03
##          adj.P.Val
## M5913 4.252585e-13
## M5921 2.151909e-07
## M5911 5.320542e-07
## M5946 2.458530e-05
## M5890 2.250878e-03
## M5930 2.250878e-03
## M5932 2.469803e-03
## M5953 1.119347e-02
## M5892 4.466979e-02

```

References

Chaussabel, Damien, Charles Quinn, Jing Shen, Pinakeen Patel, Casey Glaser, Nicole Baldwin, Dorothee Stichweh, et al. 2008. “A Modular Analysis Framework for Blood Genomics Studies: Application to Systemic Lupus Erythematosus.” *Immunity* 29 (1). Elsevier: 150–64.

Li, Shuzhao, Nadine Rouphael, Sai Duraisingham, Sandra Romero-Steiner, Scott Presnell, Carl Davis, Daniel S Schmidt, et al. 2014. “Molecular Signatures of Antibody Responses Derived from a Systems Biology Study of Five Human Vaccines.” *Nature Immunology* 15 (2). Nature Publishing Group: 195–204.

Weiner 3rd, January, Shreemanta K Parida, Jeroen Maertzdorf, Gillian F Black, Dirk Repsilber, Anna Telaar, Robert P Mohny, et al. 2012. “Biomarkers of Inflammation, Immunosuppression and Stress with Active Disease Are Revealed by Metabolomic Profiling of Tuberculosis Patients.” *PloS One* 7 (7). Public Library of Science: e40221.