Tutorial for the WGCNA package for R:
III. Using simulated data to evaluate different module detection methods and gene screening approaches

8. Visualization of gene networks

Steve Horvath and Peter Langfelder

December 7, 2011

Contents

0 Setting up the R session

Before starting, the user should choose a working directory, preferably a directory devoted exclusively for this tutorial. After starting an R session, change working directory, load the requisite packages, set standard options, and load the results of previous sections:

```
# Display the current working directory
getwd();
# If necessary, change the path below to the directory where the data files are stored.
# "." means current directory. On Windows use a forward slash / instead of the usual \.
workingDir = ".";
setwd(workingDir);
# Load WGCNA package
library(WGCNA)
library(cluster)
# The following setting is important, do not omit.
options(stringsAsFactors = FALSE);
# Load the previously saved data
load("Simulated-RelatingToExt.RData");
load("Simulated-Screening.RData")
```

8 Network visualization

In this section we provide example methods for visualizing gene co-expression networks.

8.a Multi-dimensional scaling plots

We begin with a multi-dimensional scaling plot. We choose two dimensions. The calculation may take some time.

```
cmd1=cmdscale(as.dist(dissTOM),2)
sizeGrWindow(7, 6)
par(mfrow=c(1,1))
plot(cmd1, col=as.character(colorh1), main="MDS plot",
     xlab="Scaling Dimension 1", ylab="Scaling Dimension 2")
```

The output is shown in Fig. 1. Modules correspond to the fingers in the plot and the intramodular hubs are in the finger tips.
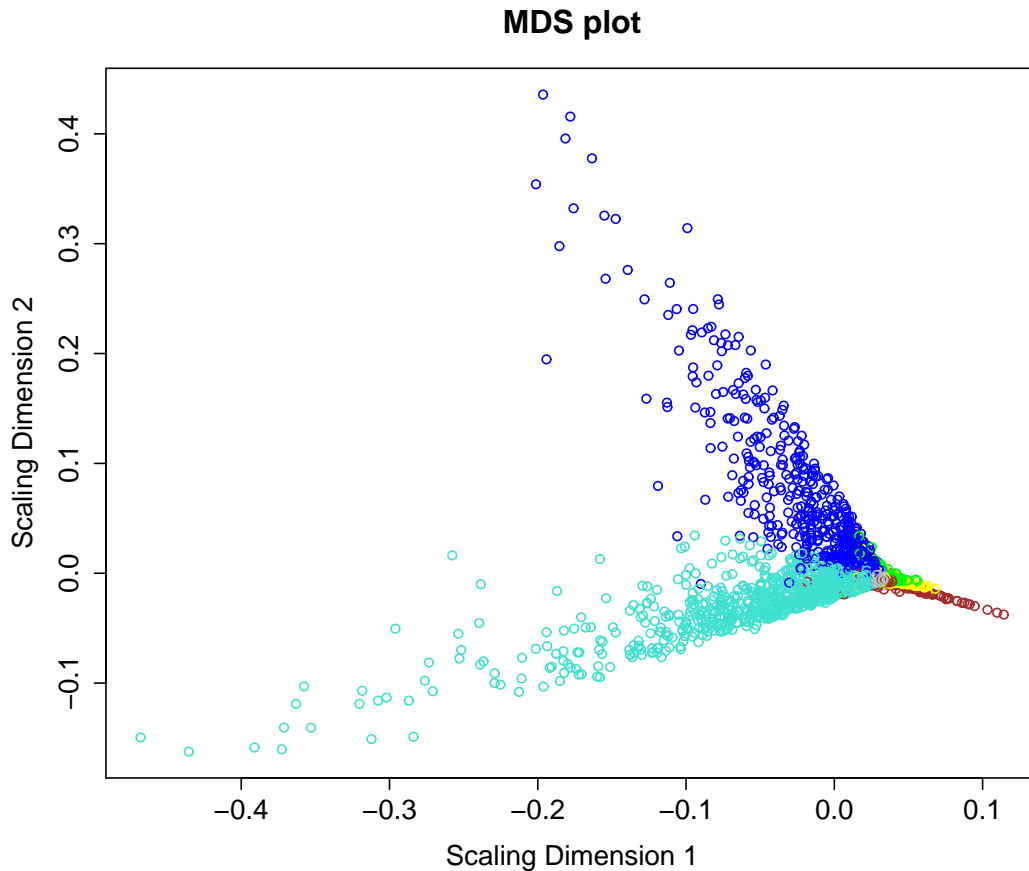

Figure 1: Multi-dimensional scaling plot in which modules tend to correspond to "fingers". Intramodular hubs are in the finger tips.

8.b Topological overlap matrix plot for visualizing the network

We now create a so-called TOM plot, a heatmap plot depicting the topological overlap matrix supplemented by hierarchical clustering dendrograms and the module colors. For large gene sets, say more than 2000 genes this will take a long time. Therefore, we will remove the grey genes from the plot. Even so, generating the plot will likely take a few minutes.

```
power=6
color1=colorDynamicTOM
```

```
restGenes= (color1 != "grey")
diss1=1-TOMsimilarityFromExpr( datExpr[, restGenes], power = 6 )
hier1=flashClust(as.dist(diss1), method="average" )
diag(diss1) = NA;
sizeGrWindow(7,7)
TOMplot(diss1^4, hier1, as.character(color1[restGenes]),
        main = "TOM heatmap plot, module genes" )
```

We have set the diagonal of the dissimilarity to NA and raised it to the power of 4 to bring out the module structure (these changes effectively amount to a change in the color scale of the plot). The resulting plot is shown in Fig. 2. We now create a similar plot from the adjacency matrix:

```
power=6
color1=colorDynamicTOM
restGenes= (color1 != "grey")
diss1=1-adjacency( datExpr[, restGenes], power = 6 )
hier1=flashClust(as.dist(diss1), method="average" )
diag(diss1) = NA;
sizeGrWindow(7,7)
TOMplot(diss1^4, hier1, as.character(color1[restGenes]),
        main = "Adjacency heatmap plot, module genes" )
```

The heatmap plot is shown in Fig. 3.

8.c   Relationships among the top 30 most significant genes

We now restrict the visualization to the 30 most significant genes identified by network screening (Section 7 of this tutorial). We first plot correlation heatmaps for signed network:

```
sizeGrWindow(7,7)
topList=rank(NS1$p.Weighted,ties.method="first")<=30
gene.names= names(datExpr)[topList]
# The following shows the correlations between the top genes
plotNetworkHeatmap(datExpr, plotGenes = gene.names,
                networkType="signed", useTOM=FALSE,
                power=1, main="signed correlations")
```

The next plot will show the correlation in the corresponding unsigned network:

```
sizeGrWindow(7,7)
# The following shows the correlations between the top genes
plotNetworkHeatmap(datExpr, plotGenes = gene.names,
                networkType="unsigned", useTOM=FALSE,
                power=1, main="signed correlations")
```

We now plot the corresponding topological overlap heatmaps:

```
sizeGrWindow(7,7)
# The following shows the TOM heatmap in a signed network
plotNetworkHeatmap(datExpr, plotGenes = gene.names,
                networkType="signed", useTOM=TRUE,
                power=12, main="C. TOM in a signed network")
# The following shows the TOM heatmap in a unsigned network
plotNetworkHeatmap(datExpr, plotGenes = gene.names,
                networkType="unsigned", useTOM=TRUE,
                power=6, main="D. TOM in an unsigned network")
```
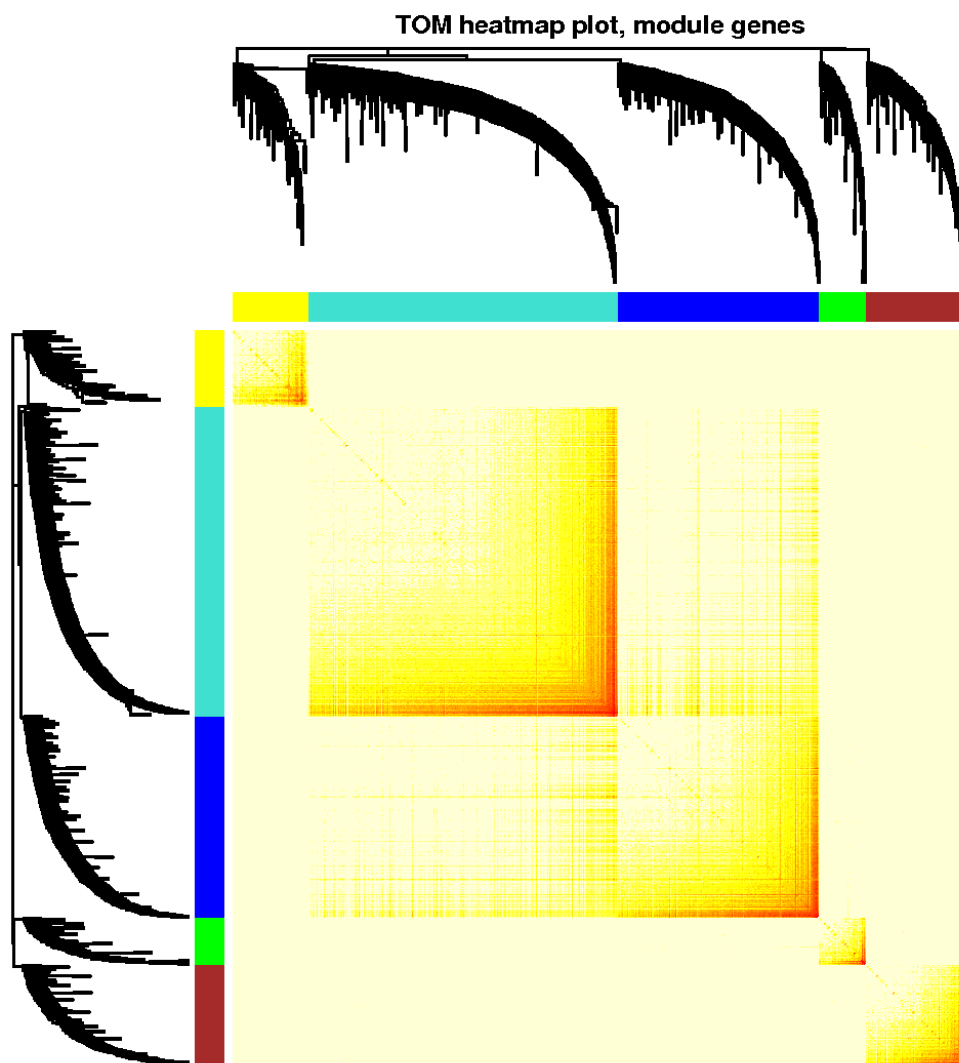
The four plots are shown in Fig. 4.

Figure 2: Heatmap plot of the topological overlap matrix. In the heatmap, rows and columns correspond to single genes, light colors represent low topological overlap, and progressively darker orange and red colors represent higher topological overlap. The corresponding gene dendrograms and module assignment are shown on the left and top.
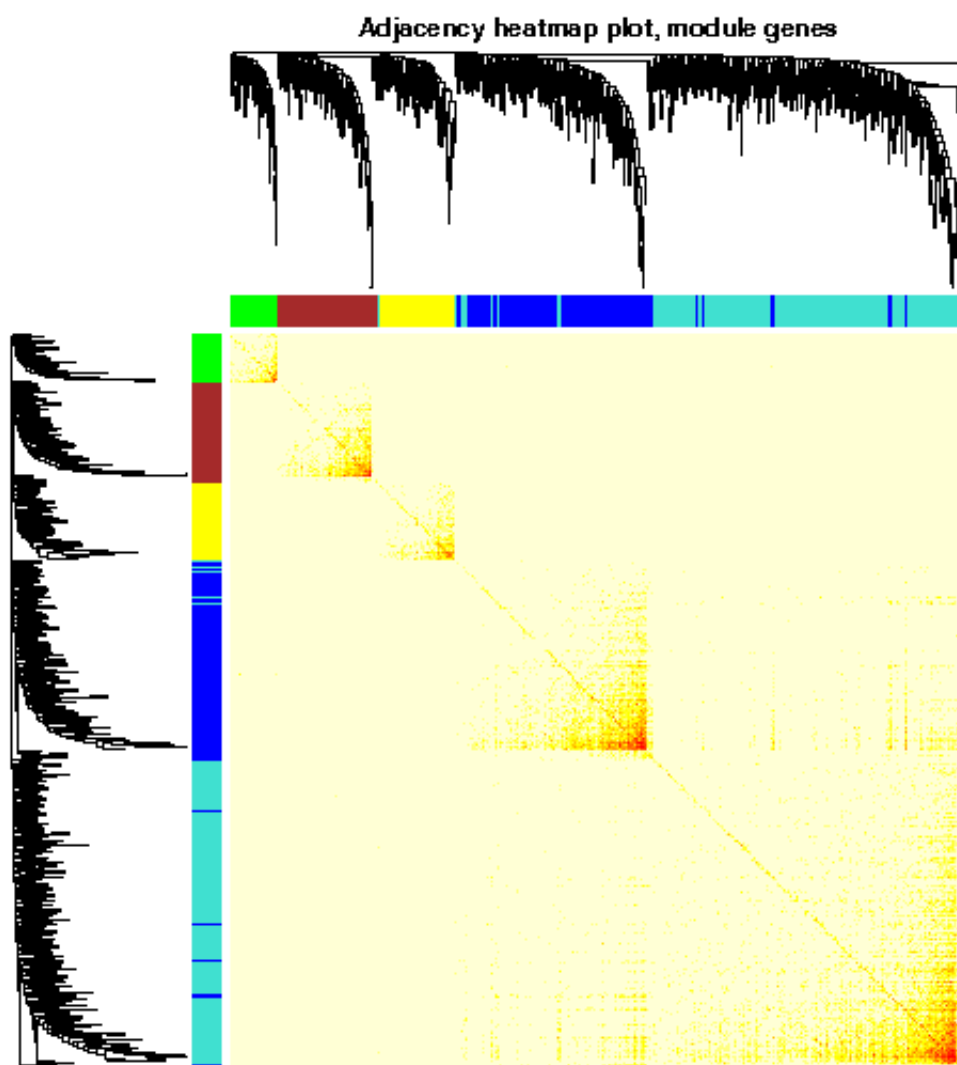
Figure 3: Heatmap plot of the adjacency. In the heatmap, rows and columns correspond to single genes, light colors represent low adjacency, and progressively darker orange and red colors represent higher adjacency. The corresponding gene dendrograms and module assignment are shown on the left and top.
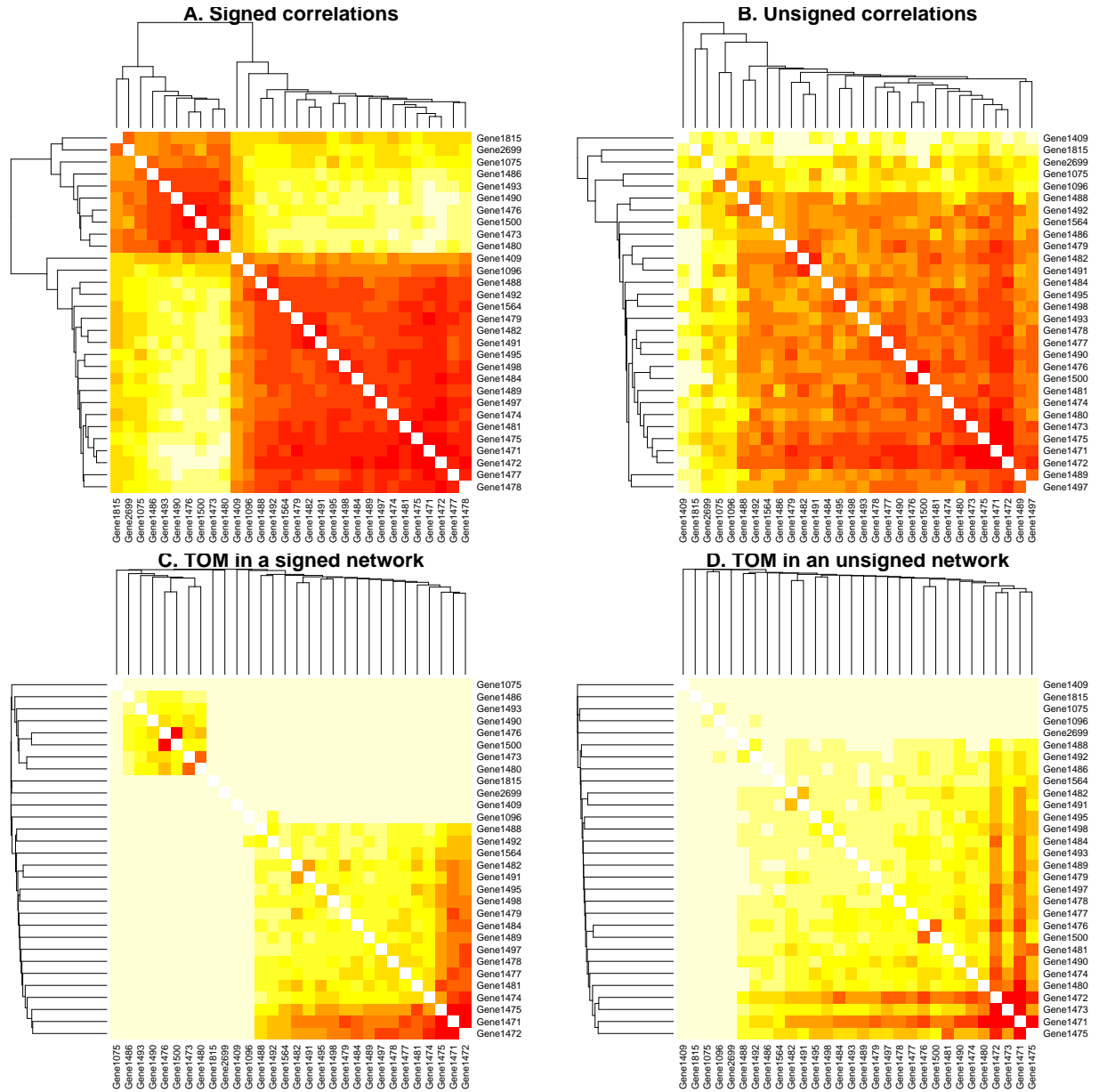
Figure 4: Heatmap plots depicting the relationships among the 30 most significant genes identified by network screening (Section 7). The plots A–D depict, respectively, the signed correlations, the unsigned correlations, the topological overlap in a signed network, and the topological overlap in the unsigned network. Each column and row of the heatmap corresponds to a single gene; light colors mean low correlations or topological overlaps (negative correlation in signed correlation plots); progressively darker colors correspond to higher correlations or topological overlaps.