Networks in microbiomics

Questions

- Are any taxa associated with antibiotic treatment?
- Is there difference in microbiome composition over time (within/between treatments)?
- Is there correlation between abundance of any taxa and metabolic phenotypes?
- *Are there any correlations between the taxa abundance and how it evolves with time and under treatment?*

Correlation networks

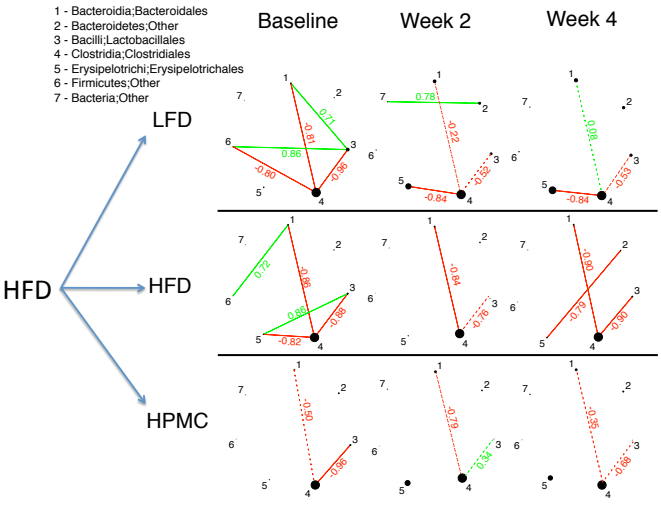
- Relevance (correlation) networks is an approach to understanding the structure of the correlations between multiple (usually homogeneous) variables.
- Based on pairwise univariate tests of association, for example pairwise correlations
- In R we can obtain a correlation matrix for all the variables in a matrix with command: `cor(d, method=method)`

General approach to relevance network construction

- Pick a univariate test appropriate for the data
- Apply the test to all pairs of variables in the dataset and construct a matrix of test statistics
- Use a significance threshold to decide which statistics are significant
- Convert the matrix into a binary matrix, where entry M_{ij} = 1 if the test between variables i and j is significant, and 0 otherwise.
- Treat the binary matrix as indicating adjacencies between nodes (variables) to visualize the result.

A concrete example of a correlation network construction protocol

- Use Pearson correlation test for most abundant taxa
- Perform FDR correction on the pairwise p-values obtained from the test
- Consider only tests that are significant at 10% FDR.

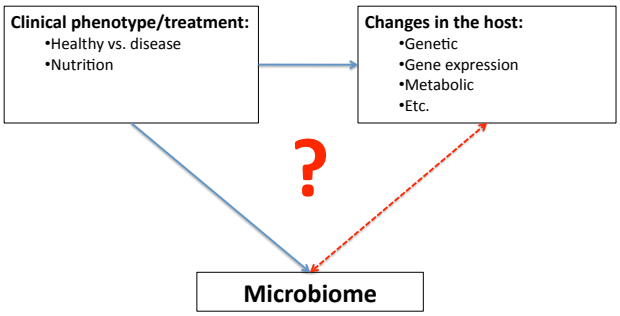


```
library(lattice)
library(network)

plot.cor.network = function(d, method='pearson', fdr.threshold=0.10, main=""){
        par(mar=c(0,0,2,0))
        nvar = dim(d)[2]
        cc = cor(d, method=method)
        varnames = sapply(colnames(d), systematic.name)
        cct = matrix(NA, nrow=nvar, ncol=nvar, dimnames = list(varnames,varnames))
        for(x in 1:(nvar-1)){
          for(y in (x+1):nvar){
                  cct[x,y] = cor.test(d[,x], d[,y])$p.value
          }
        }
        ccp = cct
        cct = p.adjust(cct, method='fdr')
        adjmatrix = matrix(0,nrow=nvar, ncol=nvar)
        adjmatrix[!is.na(cct) & cct<fdr.threshold] = 1
        colsmatrix = matrix(NA, nrow=nvar, ncol=nvar)
        colsmatrix[cc>0 & adjmatrix>0] = 'green'
        colsmatrix[cc<0 & adjmatrix>0] = 'red'
        nn = network(adjmatrix)
        cols = c(apply(as.matrix.network(nn, matrix.type = "edgelist"), 1, function(x) colsmatrix[x[1],x[2]]), NA)
        ll = sapply(colnames(d),systematic.name)
        z=plot(nn, label = as.character(1:nvar), usearrows=F,
                        mode="circle", main=main, label.cex=1.2, pad=0.2,
                        edge.col=cols, vertex.col='black',
                        vertex.cex=colMeans(d)/max(colMeans(d))*3)
        list(cor=cc, pvalue = ccp, fdr=cct, adjmatrix= adjmatrix)
}
```

See: relevance_cor_net.R

Are the microbes drivers or passengers in disease mechanisms?



Examples of non-trivial causal graphs