


WGCNA

Weighted gene co-expression network analysis

Mahsa Ehsanifard

M.S.C. of Molecular genetics

 genecan693

Introduction

A correlation or co-expression weighted network

- Showing correlation between two genes (inter-gene)
- Demonstrating the expression gestures analogy
- Nodes = genes , edges = co-expression link
- Dividing to gene modules

Overview

What is needed in WGCNA analysis for detection of high weighted co-expressed genes ?

(Microarray/RNA-Seq)

Expression
genes matrix

Rows = genes

Columns = samples

(at least 20 samples)

Phenotypic
traits

- Molecular

- Clinical

R programming
language

WGCNA package

Steps for analysis :

1- Filtering & normalization

2- sample clustering

- Detecting the outlier samples

3- Modules & module
eigenegenes

4- TOM matrix

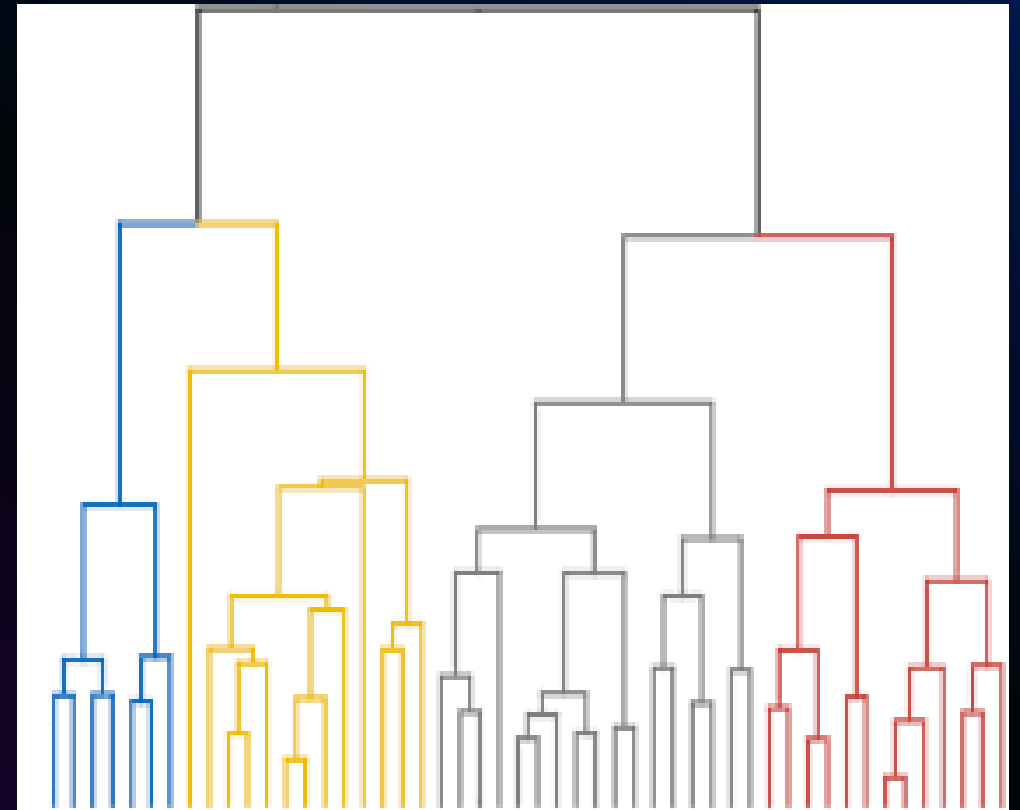
5- co-expressed genes &
trait correlations

Sample clustering

- What is clustering?
 - Brief overview of various clustering methods
 - ✓ Hierarchical clustering
- Cluster identification in hierarchical clustering trees
 - Fixed-height tree cut
 - Adaptive-height tree cut (Dynamic Tree Cut)

Hierarchical clustering

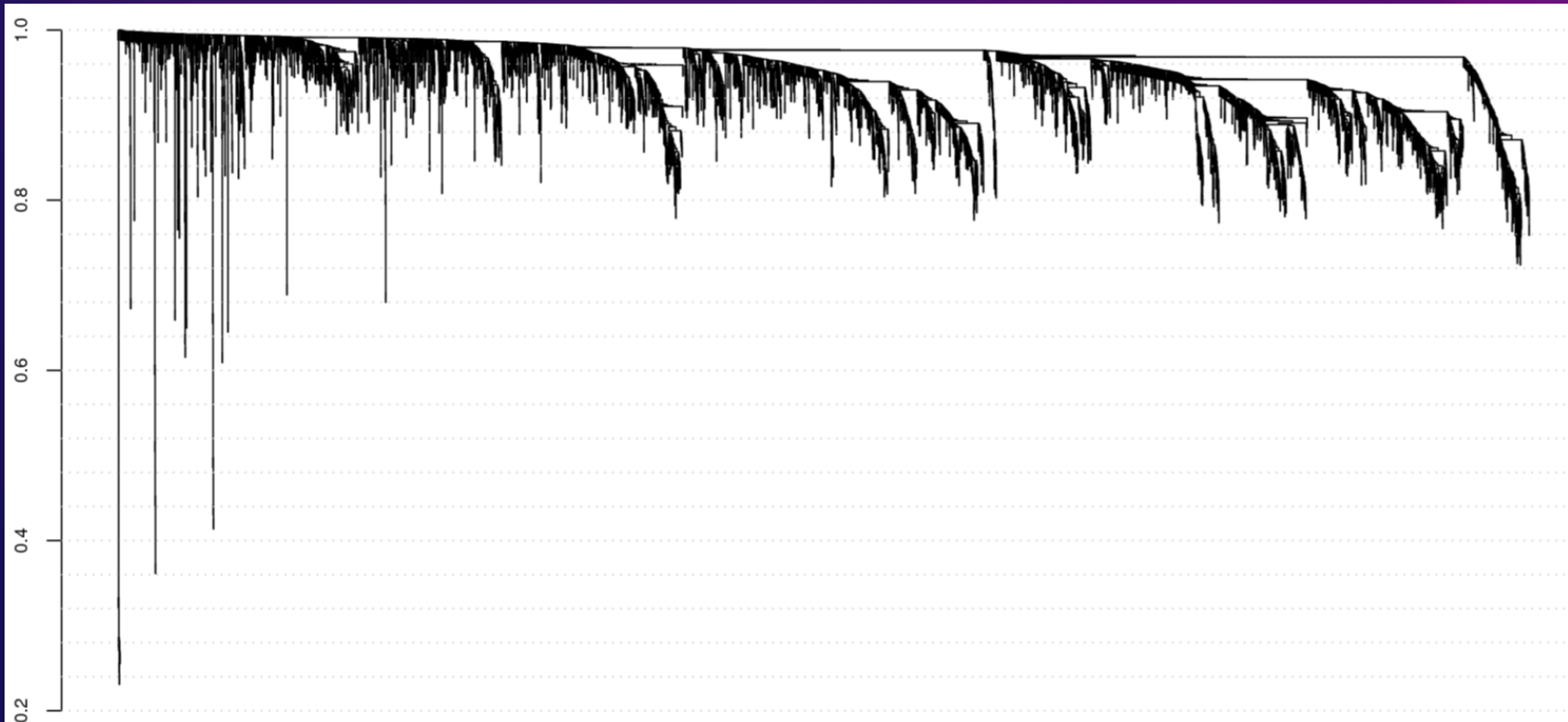
- Good method for exploratory data analysis
- It works well with high-dimensional data
- provides visualization
- Does not require specifying number of clusters beforehand
- For both genes and samples



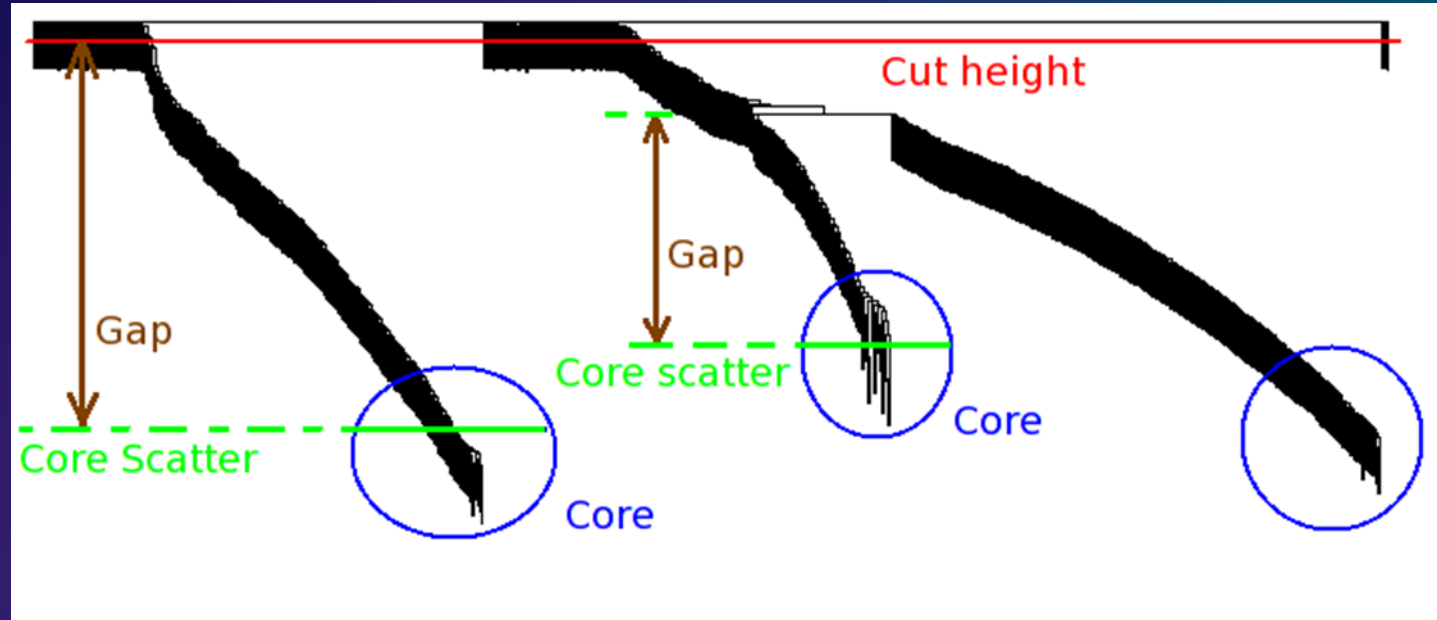
How to deal with objects that are far from clusters? (Outliers)

- A sample is in high distance of other samples
- Cluster cut line for each outlier object
- The outlier sample should be removed

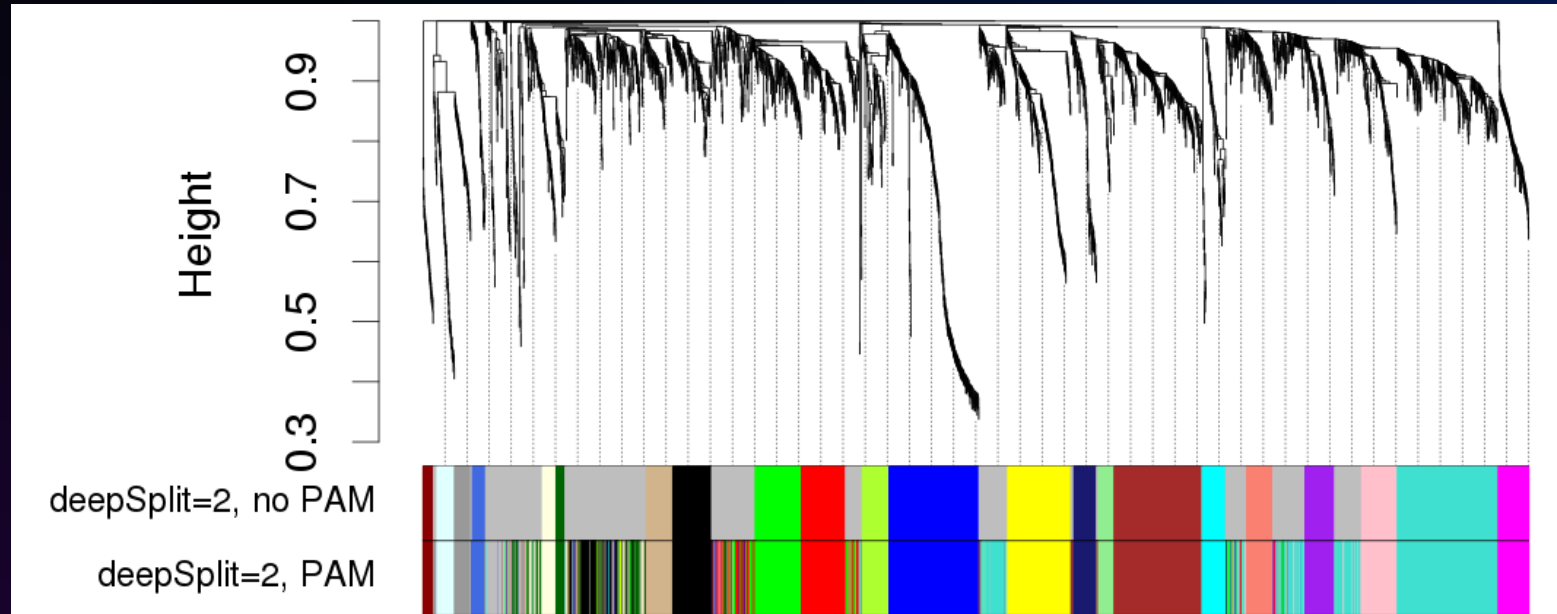
Dynamic Tree Cut



Merging branches



- If the branches meet criteria for being a cluster, they are called separate clusters
- otherwise they are merged into a single cluster



- For genomic data:
- merge modules whose expression profiles are very similar (correlation of eigengenes is too high")

Modules

- Modules group together densely interconnected genes
- High co-expressed and high connectivity
- Consensus modules
 - group together genes densely connected in all conditions

- **Module eigengene**

modules are represented by their eigengenes

Each consensus module has one eigengene in each data set

Module eigengene

- Gene module – trait interaction described by ME
- ME = a representation of a module => **representative module**
- PCA analysis has been done for representing the MEs
- Module has a lot of genes so it is difficult to analysis all the expression values of each gene => ME is the best way

Correlation (co-expression) network

- Positive cor => variable increase/decrease together
- Negative cor => variable increase/decrease in opposing direction
- Pearson = mean correlation
- Biweight = median correlation (**bicor**) => WGCNA package
- Correlation coefficient (R) => $-1 < R < 1$

$R = 1$ => perfect positive

$R = -0.5$ => medium negative

$R = 0.85 \sim 1$ => strong positive

$R = 0.15 \sim 0$ => weak positive

Trait

- Binary trait => usually molecular trait
- Clinical traits => phenotypic information
- Module eigengene - trait correlation

Which driver module is high connected to what trait