



Consiglio Nazionale  
delle Ricerche

**progetti @**



---

**Network analysis integrating miRNA expression  
profiling with MRI biomarkers and clinical data for  
prostate cancer early detection:  
a proof-of-concept study**

---

**FEDERICA CONTE, PHD**

*Institute of Systems Analysis and Computer Science (IASI) "A. Ruberti"*

*CNR - Roma*

## RESEARCH ACTIVITY

Systems Biology  
&  
Modeling Analysis

K-RAS TUMORS

mRNA HALF-LIVES

INSULIN SIGNALING

CANCER RADIOTHERAPY

NEURODEGENERATIVE DISEASES

Bioinformatics  
&  
Network Analysis

VITICULTURE

HUMAN DISORDERS

NETWORK MEDICINE

## COLLABORATIONS



SAPIENZA  
UNIVERSITÀ DI ROMA



UNIVERSITY  
OF TRIESTE



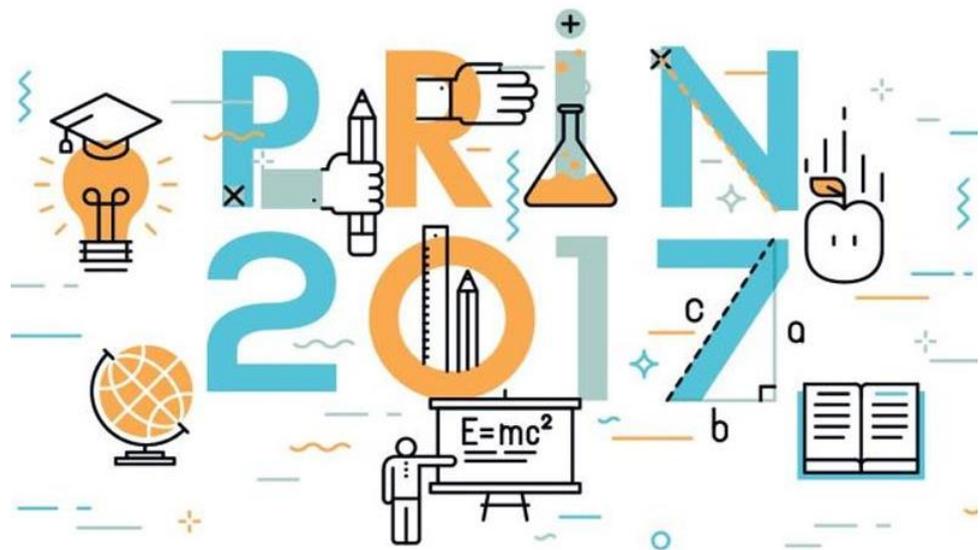
Bambino Gesù  
OSPEDALE PEDIATRICO



SAPIENZA  
UNIVERSITÀ DI ROMA

UMBERTO I  
POLICLINICO DI ROMA





MIUR.AOODGRIC.REGISTRO\_PRIN2017.0003798.29-03-2018

*Ministero dell'Istruzione dell'Università e della Ricerca*

Dipartimento per la formazione superiore e per la Ricerca

Direzione Generale per il Coordinamento, la promozione e la valorizzazione della Ricerca

PRIN: PROGETTI DI RICERCA DI RILEVANTE INTERESSE NAZIONALE – Bando 2017

Prot. 20178L3P38



## 2. Research project title

Regulation of gene expression in grapevine: analysis of genetic and epigenetic determinants

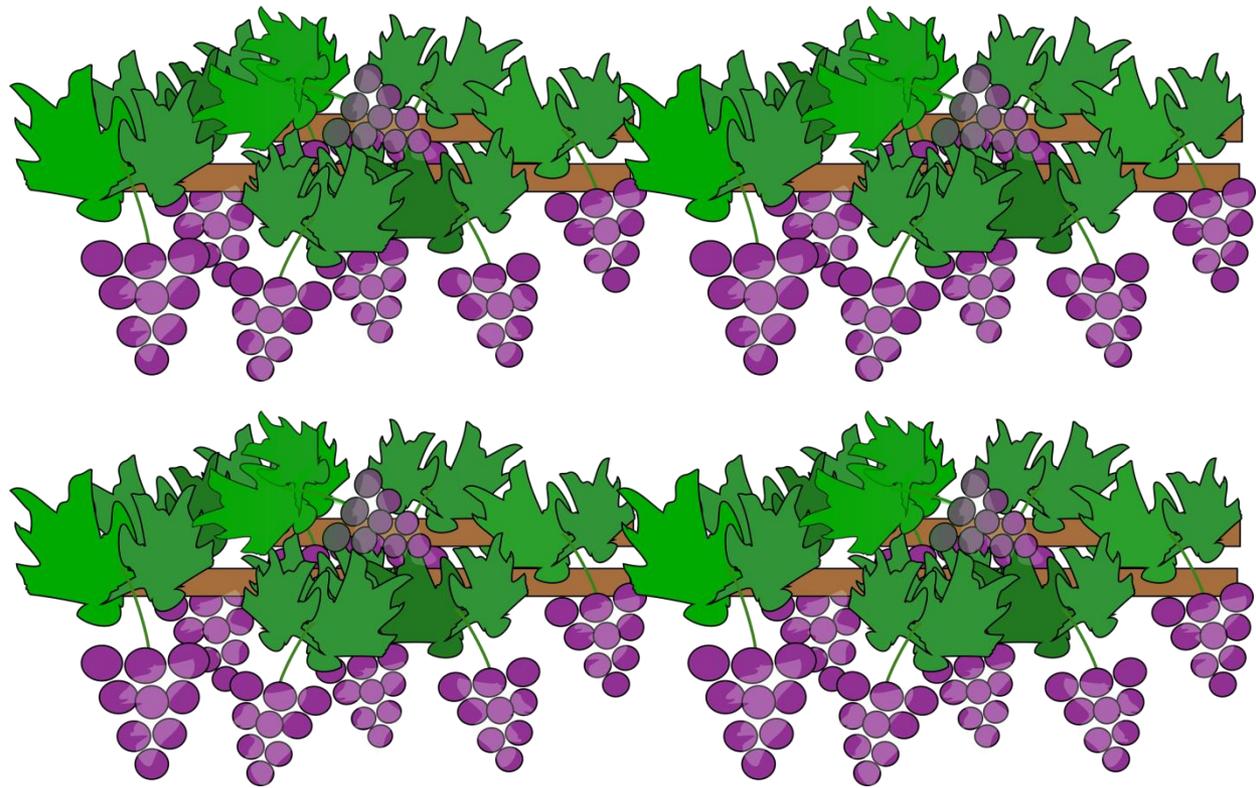
# Personnel of the research unit: Paola Paci

# Research proposal

The aim of this project is to analyse in depth the grapevine genome (~500 Mbp) to search for putative regulatory regions and cis-regulatory variants with a range of wet lab and bioinformatic approaches that will use biochemical, evolutionary and genetic data to look at both genetic and epigenetic factors involved in gene regulation.

## Objective of our research unit

The task of our research unit is the in silico identification of regulatory regions in the grapevine genome and the integration, through bioinformatics approaches, of different types of data (RNASeq for gene expression, ChIPSeq for histone modifications, ATACSeq for chromatin accessibility, BSSeq for DNA methylation, DAPSeq for transcription factor binding site identification). Up to now, our research activity focused on the **development of a pipeline for analyzing RNA-seq data**. This will not only provide basic information on the levels and patterns of gene expression, but will also allow us to evaluate the correlation between the transcriptional activity of genes and epigenetic information when it becomes available during the duration of the project.



**DELAYED**

# **1. Network-based approaches**

# Why we need networks

- **Networks are everywhere...**
  - Molecular networks
  - Cell-cell communication
  - Nervous systems
  - Social networks
- **Networks are powerful tools...**
  - Reduce complexity
  - More efficient than tables
  - Great for data integration
  - Intuitive visualization

Social network



Computer network: internet



Network of airports

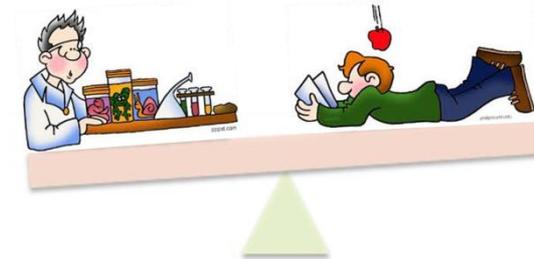
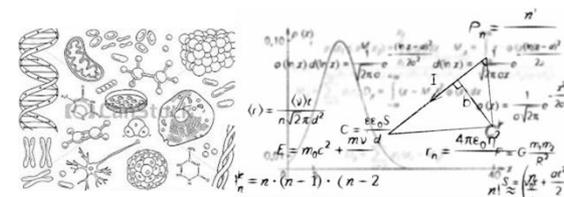
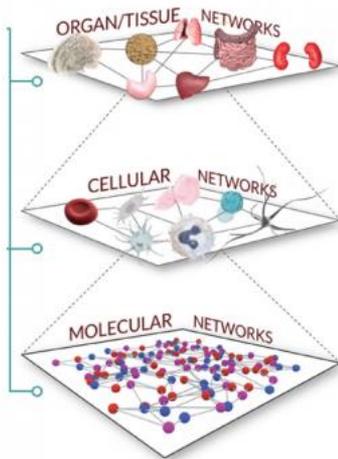


Networks of underground lines: London tube map



# Why we need networks

- **Networks are everywhere... especially in biology!**
- Molecular networks
- Cell-cell communication
- Nervous systems
- Social networks
- **Networks are powerful tools... especially in biology!**
- Reduce complexity
- More efficient than tables
- Great for data integration
- Intuitive visualization
- **Common language**



## **2. Weighted Gene Co-expression Network Analysis (WGCNA)**

# Input Data

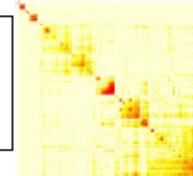
- **Gene expression data:**
  - Microarray
  - RNA-Seq
- **Clinical/phenotypical traits** from the same samples (optional):
  - Weight, race, stage, level of glucose, ecc.

# WGCNA workflow

## Construct a gene co-expression network

**Rationale:** make use of interaction patterns among genes

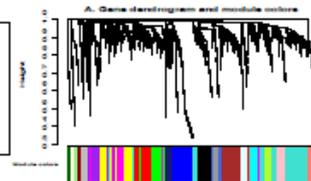
**Tools:** correlation as a measure of co-expression



## Identify modules

**Rationale:** module (pathway) based analysis

**Tools:** hierarchical clustering, Dynamic Tree Cut

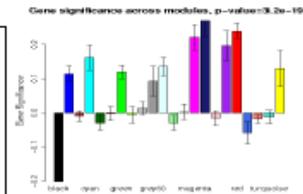


## Relate modules to external information

Array Information: clinical data, SNPs, proteomics

Gene Information: ontology, functional enrichment

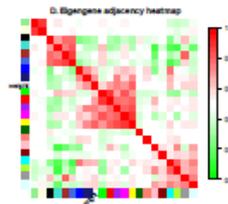
**Rationale:** find biologically interesting modules



## Study module relationships

**Rationale:** biological data reduction, systems-level view

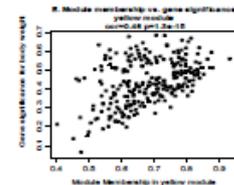
**Tools:** Eigengene Networks



## Find the key drivers in *interesting* modules

**Rationale:** experimental validation, biomarkers

**Tools:** intramodular connectivity, causality testing

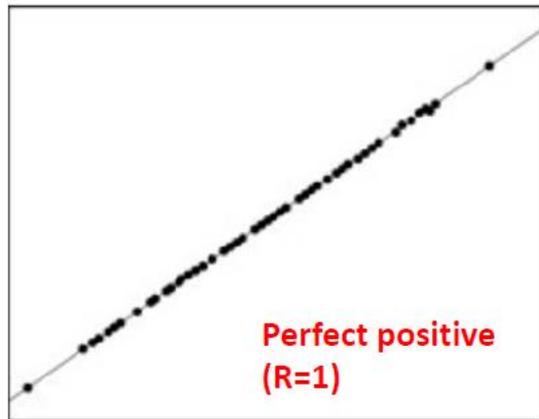


# Gene co-expression network

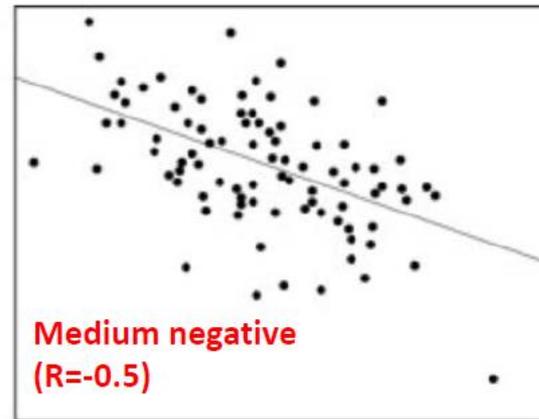
- **Co-expression networks:** undirected, weighted gene networks where nodes correspond to genes and edges between them are determined by the pairwise correlations between their expression profiles.
- **Correlation:** statistical measure for the extent to which two variables fluctuate together
  - Positive correlation: variables increase/decrease together
  - Negative correlation: variables increase/decrease in opposing direction

# Correlation examples

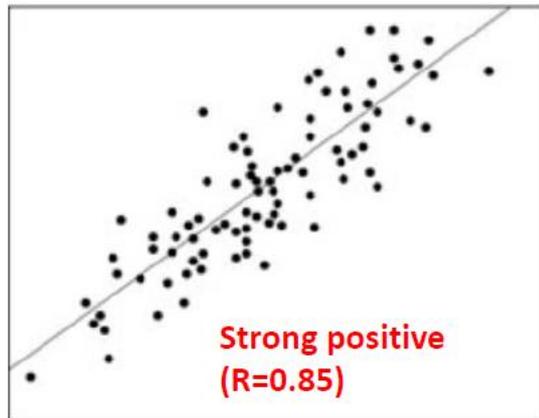
Pearson correlation coefficient R



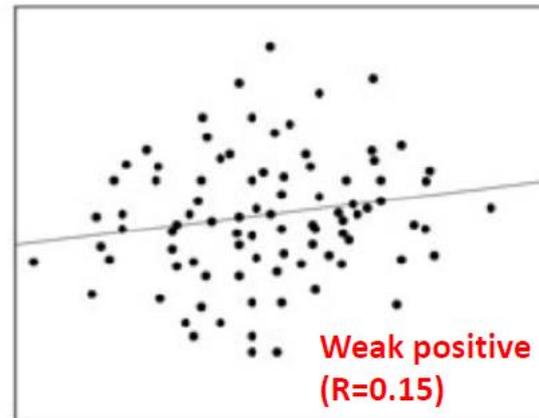
a



b



c



d

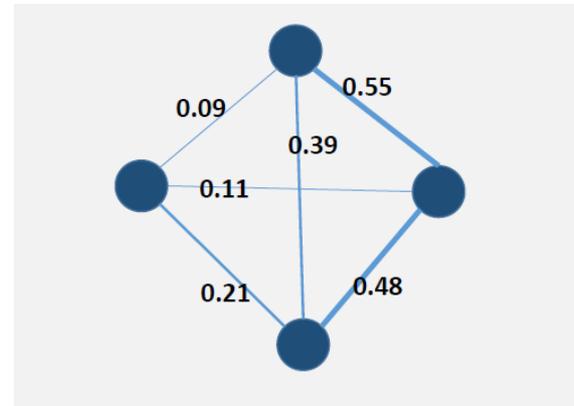
# Adjacency matrix

- The correlation between each pair of genes  $(i, j)$  is raised to a power  $\beta$  (**soft thresholding**):

$$a_{i,j} = |\text{corr}(i, j)|^\beta$$

Adjacency matrix of 4 genes

$a_{i,j}$	gene1	gene2	gene3	gene4
gene1	1	0.55	0.39	0.09
gene2	0.55	1	0.48	0.11
gene3	0.39	0.48	1	0.21
gene4	0.09	0.11	0.21	1



- This amplifies disparity between strong and weak correlations and leads to a fully connected network with genes as nodes and  $a_{i,j}$  as edge weights.

# Example

Power term  $\beta = 4$

## Correlation

$$\text{cor}(i, j) = 0.8$$

$$\text{cor}(k, l) = 0.2$$

$$\frac{0.8}{0.2} = 4 \text{ fold of difference}$$



## Adjacencies

$$|0.8|^4 = 0.4096$$

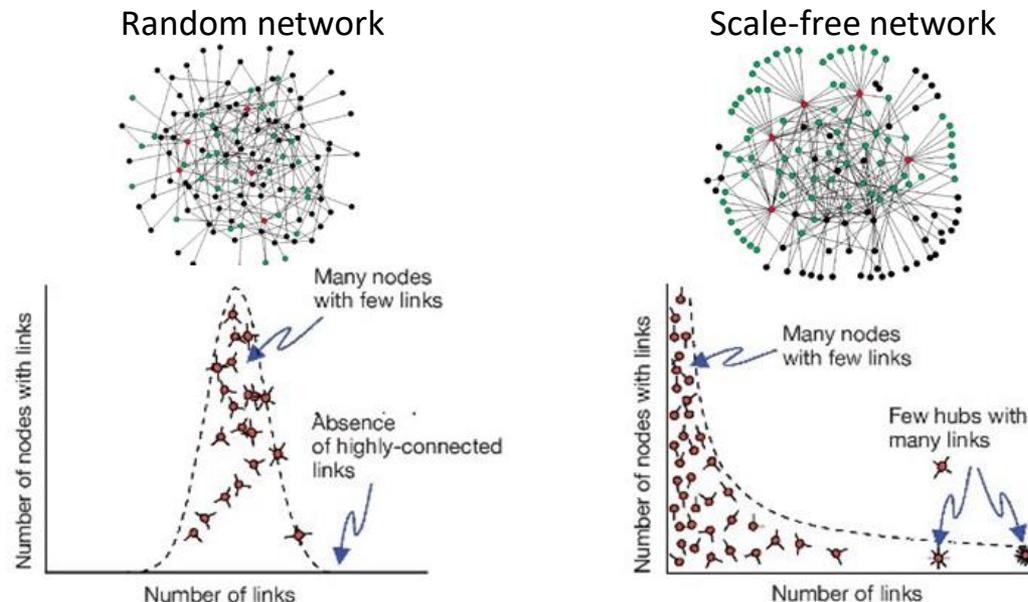
$$|0.2|^4 = 0.0016$$

$$\frac{0.4096}{0.0016} = 256 \text{ fold of difference}$$

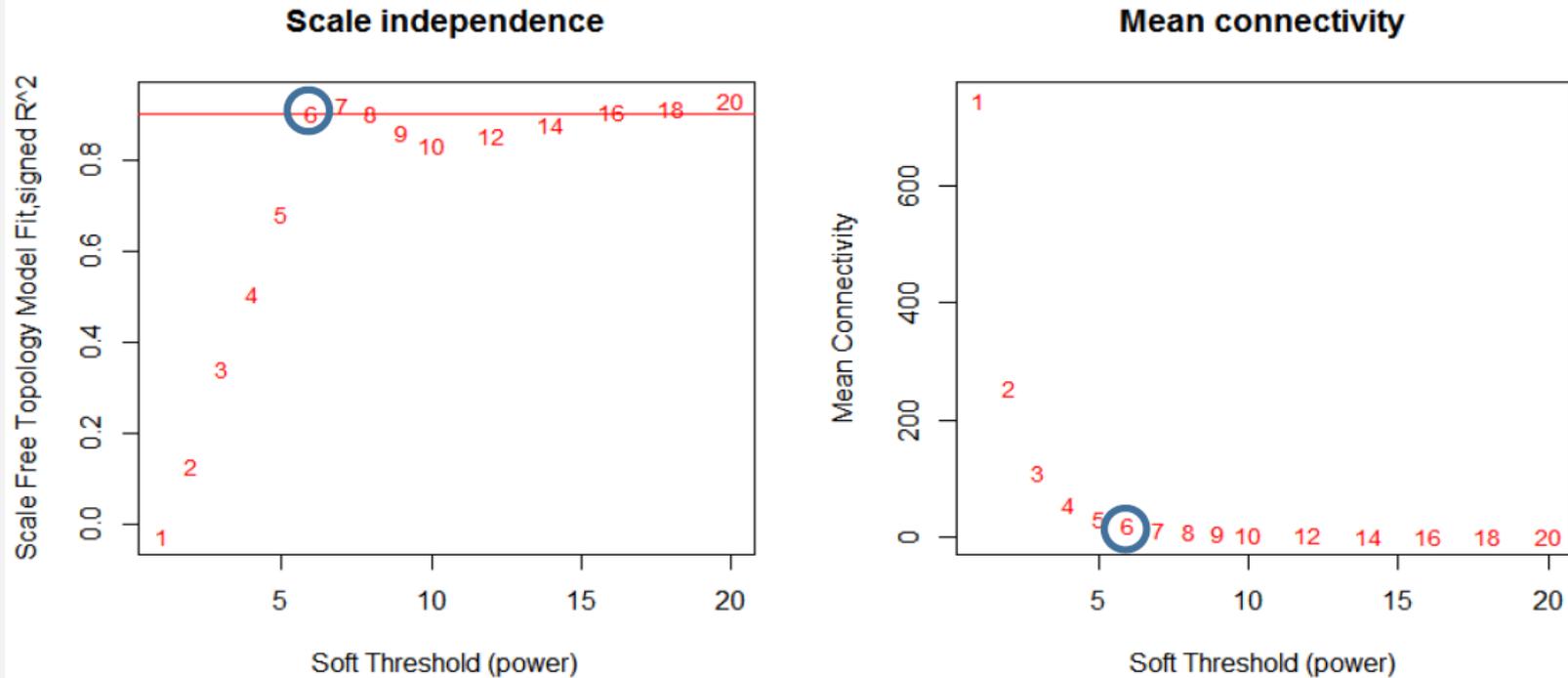
# Selection criterion of power term

Pick the lowest possible  $\beta$  leading to an approximately **scale-free network topology**, that is the typical structure of most of biological networks:

- few nodes with many connections (hub nodes)
- many nodes with few connections (peripheral nodes)
- degree distribution follows a power law, i.e. the probability for a node of having  $k$  links is  $k^{-\gamma}$

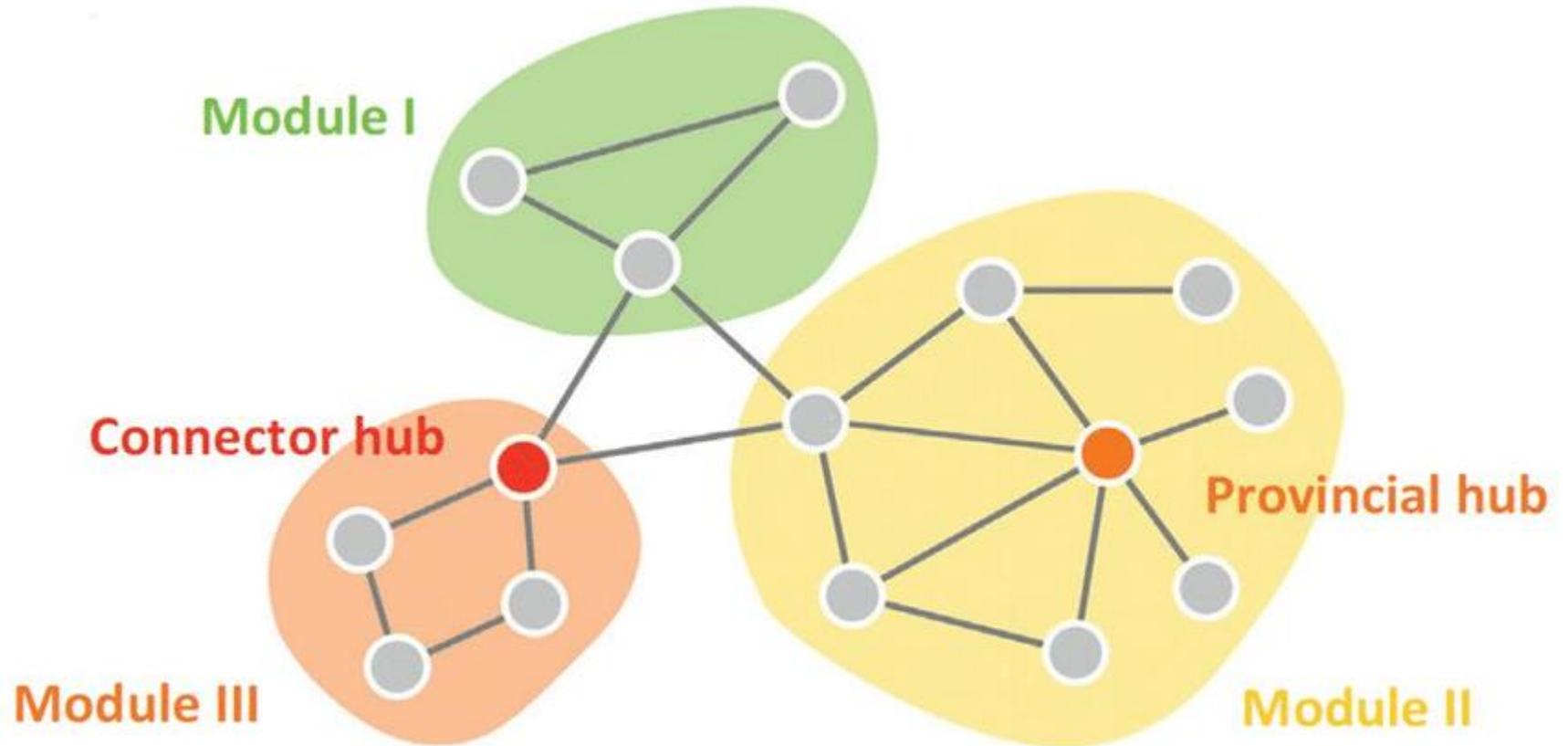


## Pick a power term: Visual Aid in WGCNA



- Left plot: Choose power 6. Lowest possible power term where topology approximately fits a scale free network (on or above red horizontal line).
- Right plot: mean connectivity drops as power goes up. Must not drop too low.

# Modules detection



# Topological Overlap Measure (TOM)

- TOM is a pairwise similarity measure between network nodes (genes) defined as:

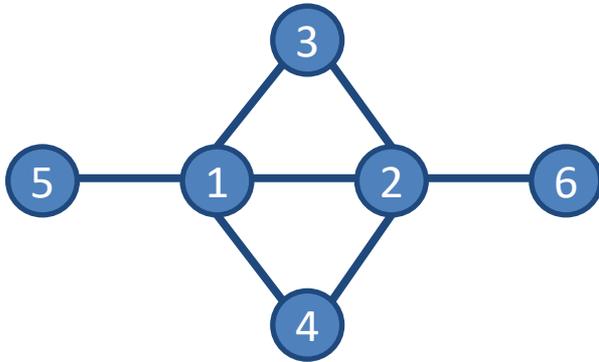
$$TOM_{ij} = \frac{\sum_u a_{iu} a_{uj} + a_{ij}}{\min(k_i, k_j) + 1 - a_{ij}}$$

$$DistTOM_{ij} = 1 - TOM_{ij}$$

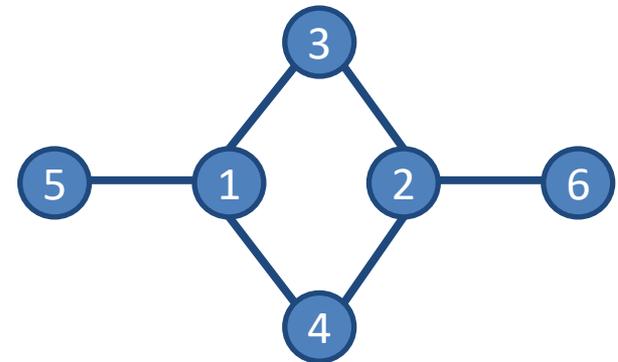
- TOM(i,j) is high if genes i,j have many shared neighbors (overlap of their network neighbors is large)
- High TOM(i,j) implies that genes have similar expression patterns

# Topological Overlap Measure (TOM)

$$TOM_{ij} = \frac{\sum_u a_{iu} a_{uj} + a_{ij}}{\min(k_i, k_j) + 1 - a_{ij}} \quad \rightarrow \quad TOM_{ij} = \frac{\# \text{common neighbors} + \text{link between them}}{\min(\# \text{ link } i, \# \text{ link } j) + 1 - \text{link between them}}$$



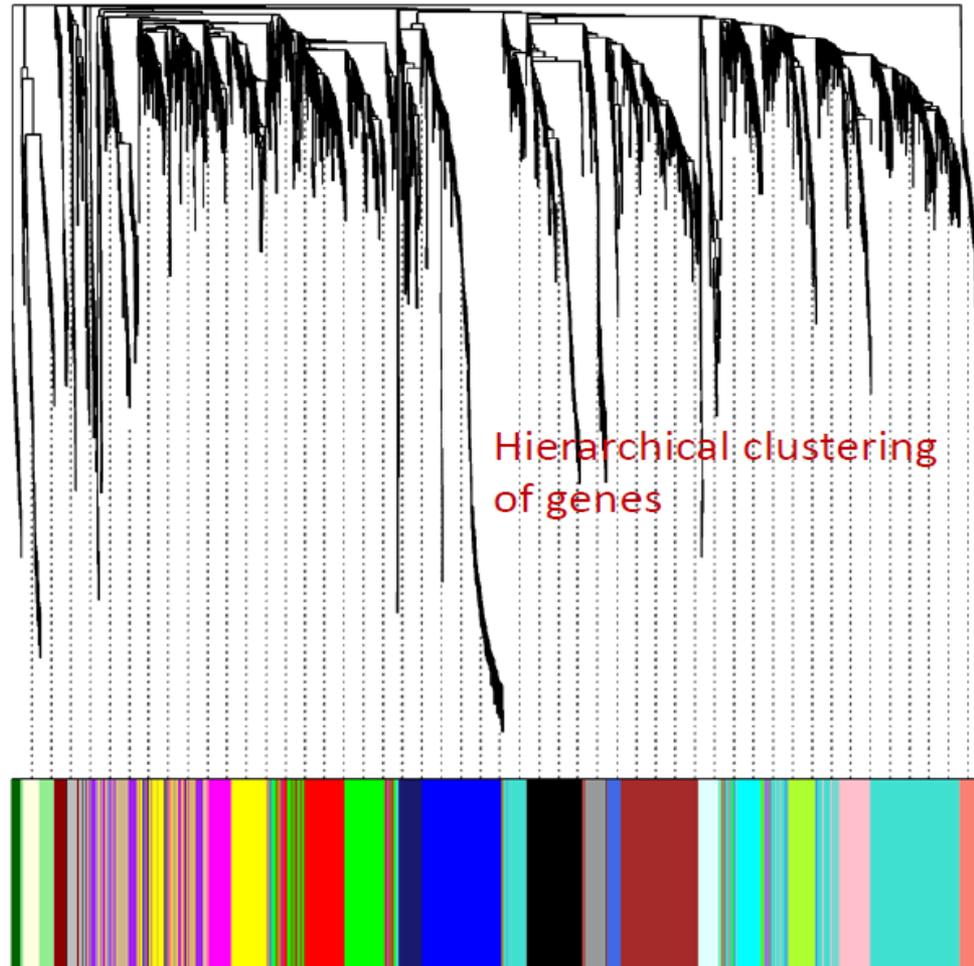
$$TOM_{1,2} = \frac{2+1}{\min(4,4)+1-1} = \frac{3}{4} = 0.75$$



$$TOM_{1,2} = \frac{2+0}{\min(3,3)+1-0} = \frac{2}{4} = 0.5$$

# Hierarchical clustering

Gene dendrogram and module colors



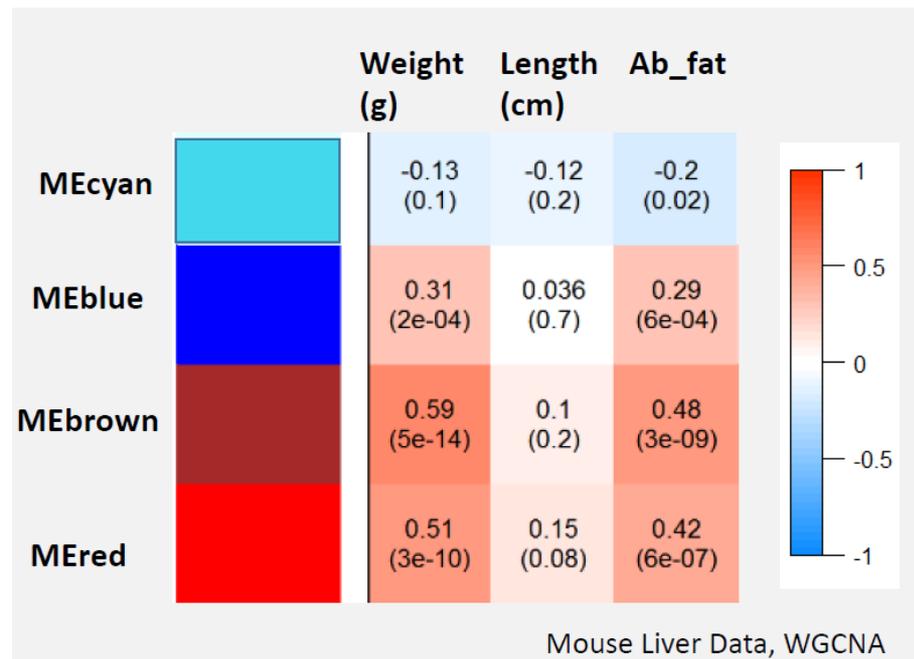
# Module eigengene (ME)

- **What it is:**
  - a 1-dimensional data vector, summarizing the expression data of the genes forming a module
- **How it is computed:**
  - 1st principal component of the expression data of the genes forming a module
- **What it is used for:**
  - modules can be correlated with one another
  - modules can be clustered
  - modules can be correlated with external traits

# Correlate modules to external traits

- Compute correlations between each module eigengene and each trait variable
- Identify modules highly correlated to traits of interest
- Identify traits highly correlated to multiple modules

**Heatmap of module–trait relationships**



# Driver genes identification

- **Aim:**
  - Identify key genes that may influence the expression or function of other genes **OR**
  - may be causal factors for a trait of interest
- **Caveat:**
  - WGCNA cannot show whether gene-gene or gene-trait relationships are causal **BUT**
  - can help find candidate genes for further study
- **Strategy:**
  - Identify those genes within a module that are most representative or highly connected within the module (highest **module membership**) **AND**
  - most strongly correlated with a trait of interest (highest **gene significance**)

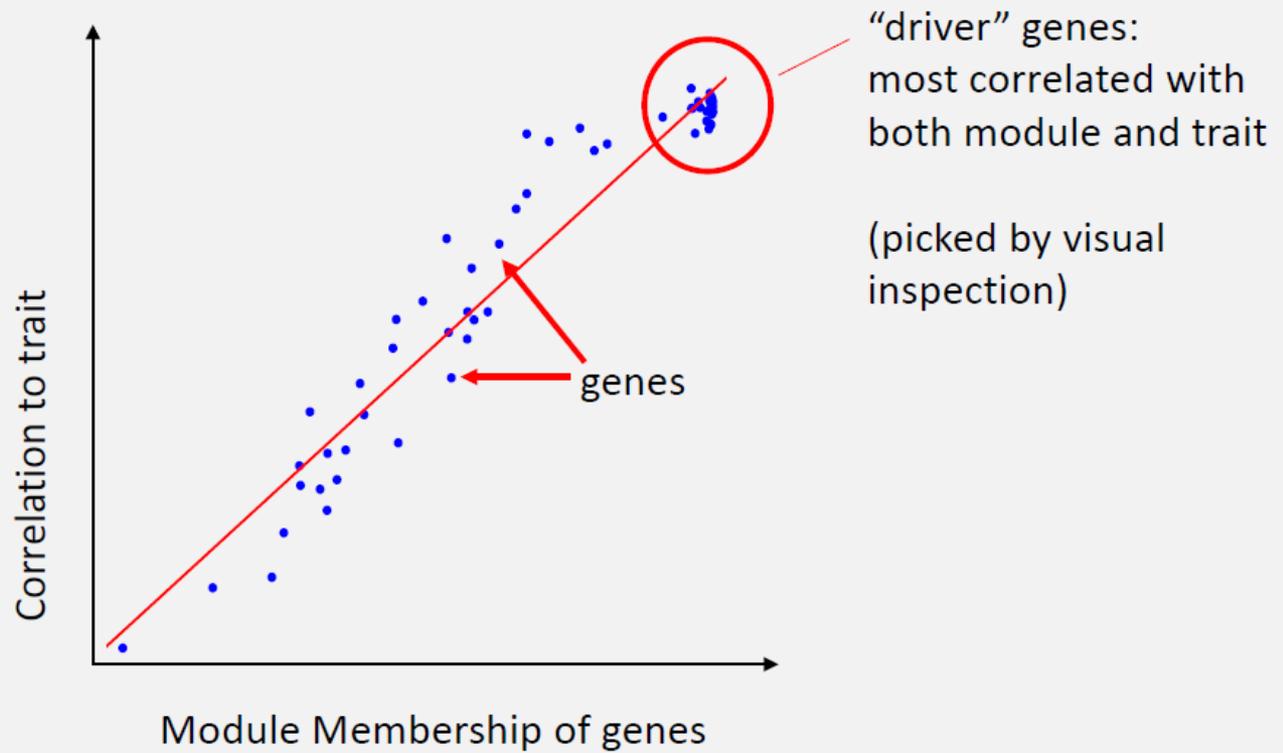
# Module membership (MM)

- MM of a gene is the correlation between its expression profile and the module eigengene of a given module
- Genes with high MM are good representatives of the overall expression profile in the module
- Genes with high MM tend to be “hub” in the module
  - If  $MM \rightarrow 0$  the gene is not part of the module.
  - if  $MM \rightarrow 1$  or  $-1$ , the gene is highly connected to the genes of that module. The sign of MM encodes whether the gene has a positive or a negative relationship with the module eigengene.

# Gene significance (GS)

- GS of a gene is the correlation between its expression profile and external sample traits
- Genes with higher GS are the more biologically significant genes
  - If  $GS \rightarrow 0$  that the gene is not significant with regard to the biological question of interest
  - if  $GS \rightarrow 1$  or  $-1$ , the gene is significant with regard to the biological question of interest. The sign of GS encodes whether the gene has a positive or a negative relationship with the trait of interest.

## Potential driver genes



### **3. WGCNA application for prostate cancer early detection**

# Prostate cancer

- Prostate cancer (PCa) is the carcinoma of prostate gland that may spread to other parts of the body particularly bones and lymph nodes.
- PCa is the fourth most common malignancy worldwide and the second cause of cancer-related death in men.
- PCa can be clinically significant (csPCa) or insignificant (ciPCa):
  - csPCa is often aggressive and potentially metastatic, requiring early detection and possibly multi-modal therapy;
  - ciPCa often never progress and can be safely treated with active surveillance strategies.

# Diagnostic Parameters

- **Prostate-Specific Antigen (PSA) Blood Test:** to look for raised levels of PSA in the blood that may be a sign of cancer in its early stages
- **Digital Rectal Examination:** to rule out prostate enlargement caused by benign prostatic hyperplasia
- **Magnetic Resonance Imaging (MRI):** to assess the extension into the bladder and lymph nodes for staging the cancer and to evaluate bone metastasis.
- **Biopsy:** to aid in the diagnosis and help to determine the Gleason score

**Gleason score** is based upon microscopic appearance and ranges from 2 to 10, with higher numbers indicating higher risks, aggressiveness and mortality ( $\geq 7$  means csPCa)

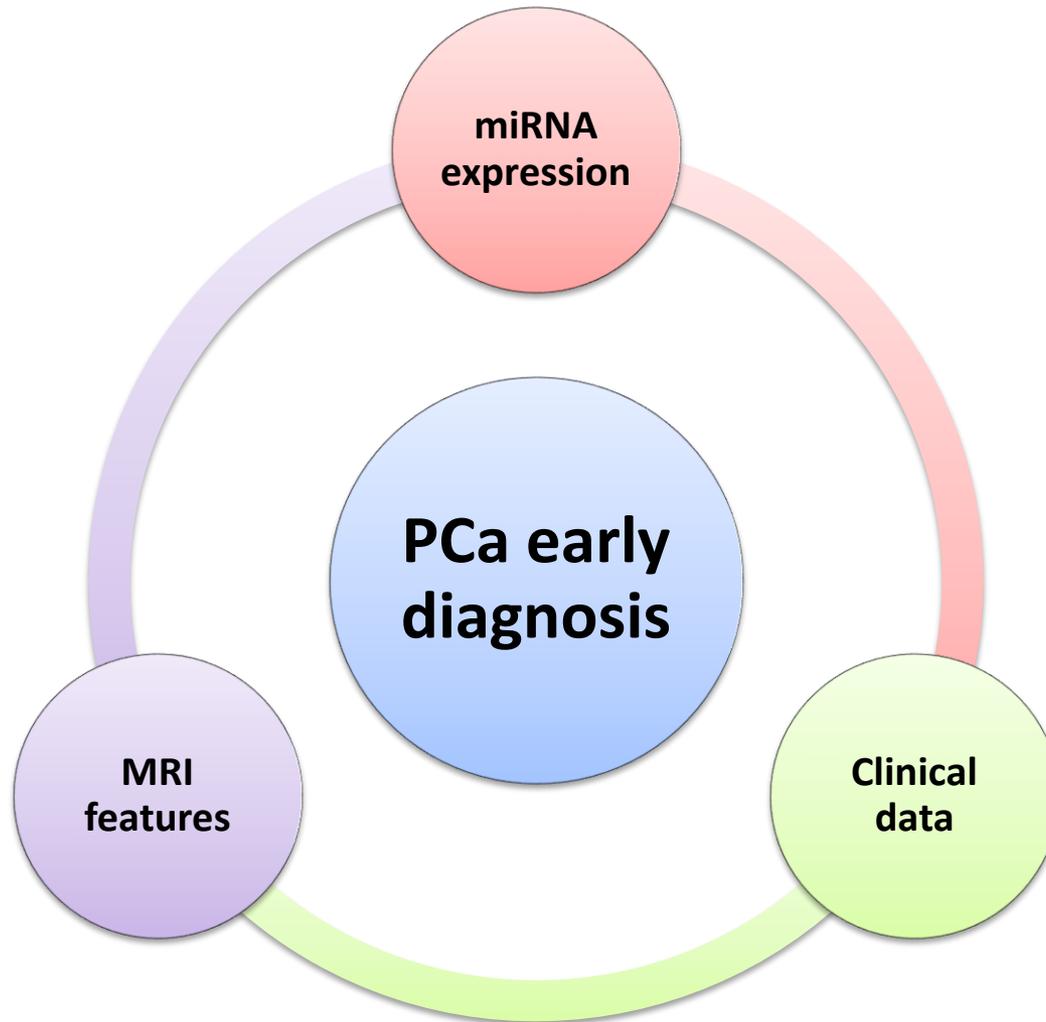
# Magnetic Resonance Imaging (MRI)

- Prostate MRI is recommended by the European association of Urology guidelines as triage test to identify men with csPCa in view of its **high accuracy**
- Nonetheless, MRI still **misses around 11%** csPCa
- Guidelines suggest the use of complementary non-invasive tests (**blood- or urine-based biomarkers**) to implement the accuracy of MRI

# microRNAs as promising biomarkers

- MicroRNAs (miRNAs) are small, non-coding RNA molecules of 21–23 nucleotides
- They are important post-transcriptional regulators of gene expression by translation inhibition or degradation of their target mRNAs
- They represent promising biomarkers for the early detection of PCa, given their stability in tissue and biological fluids

# WGCNA for a non-invasive diagnosis



# Sample collection

Equally sized cohort:

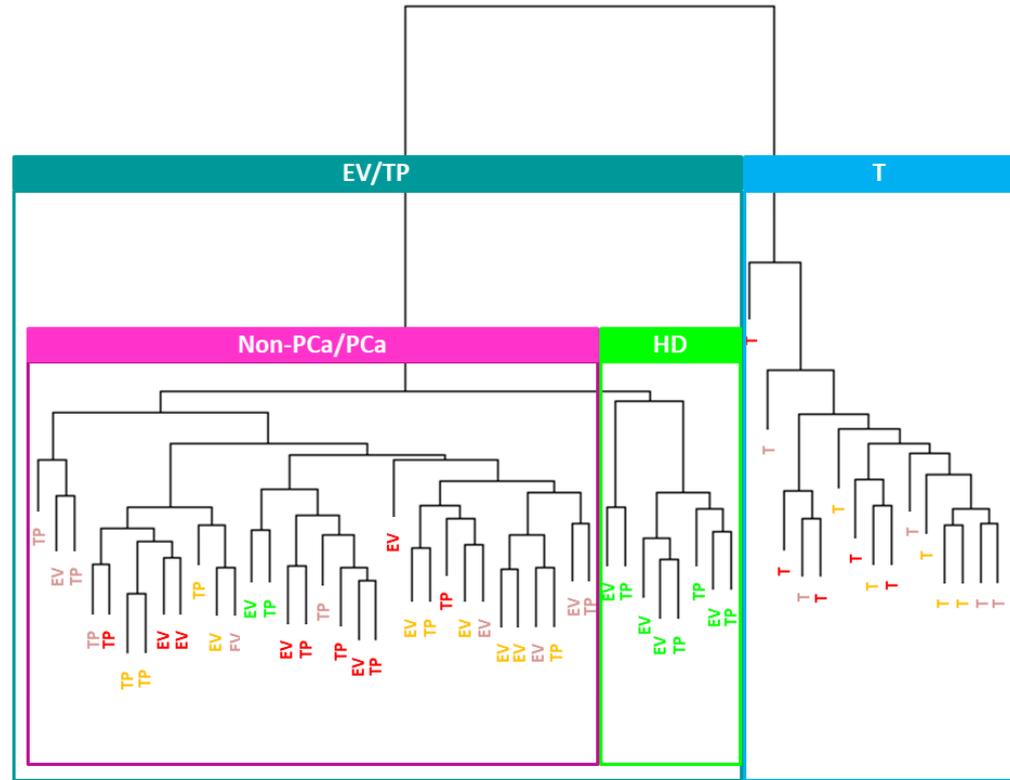
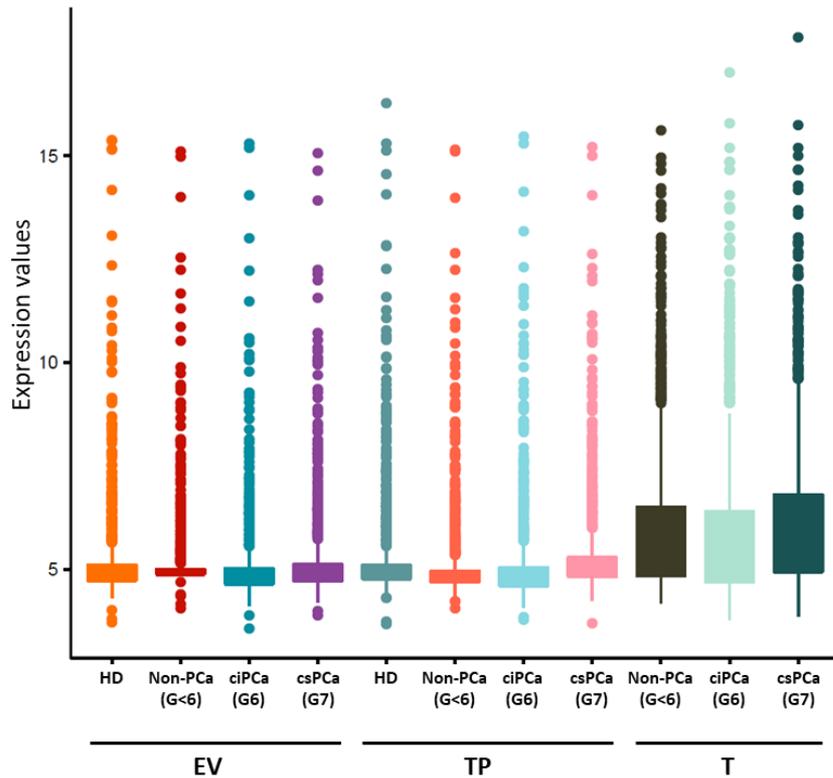
- 5 patients with csPCa (Gleason 7)
- 5 patients with ciPCa (Gleason 6)
- 5 patients with Non-PCa (Gleason < 6)

miRNA from total plasma (TP), extracellular vesicles (EV) and prostate tissue (T)  
+  
mpMRI features  
+  
Clinical data

- 5 healthy donors (HD)

miRNA from TP and EV

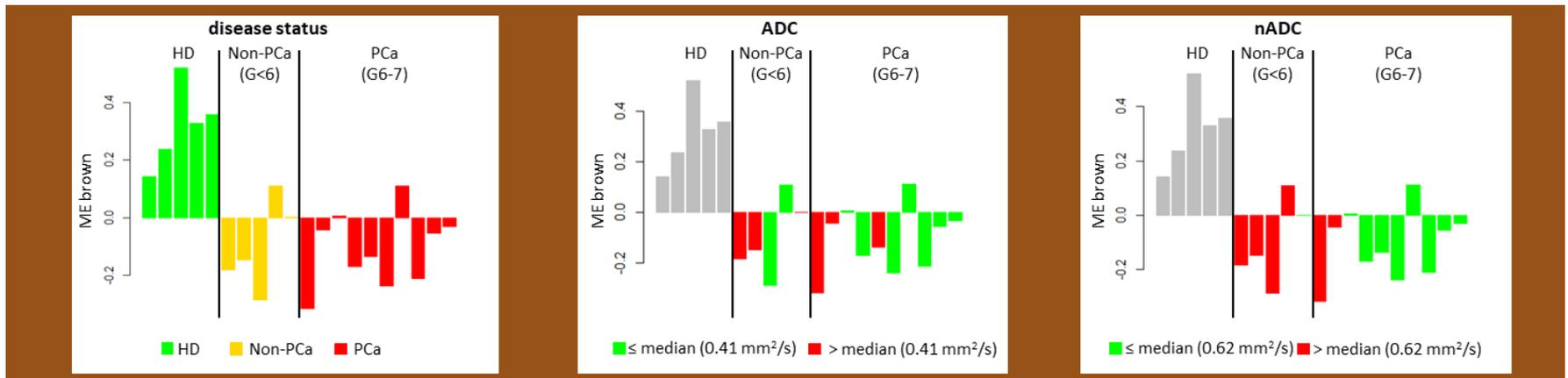
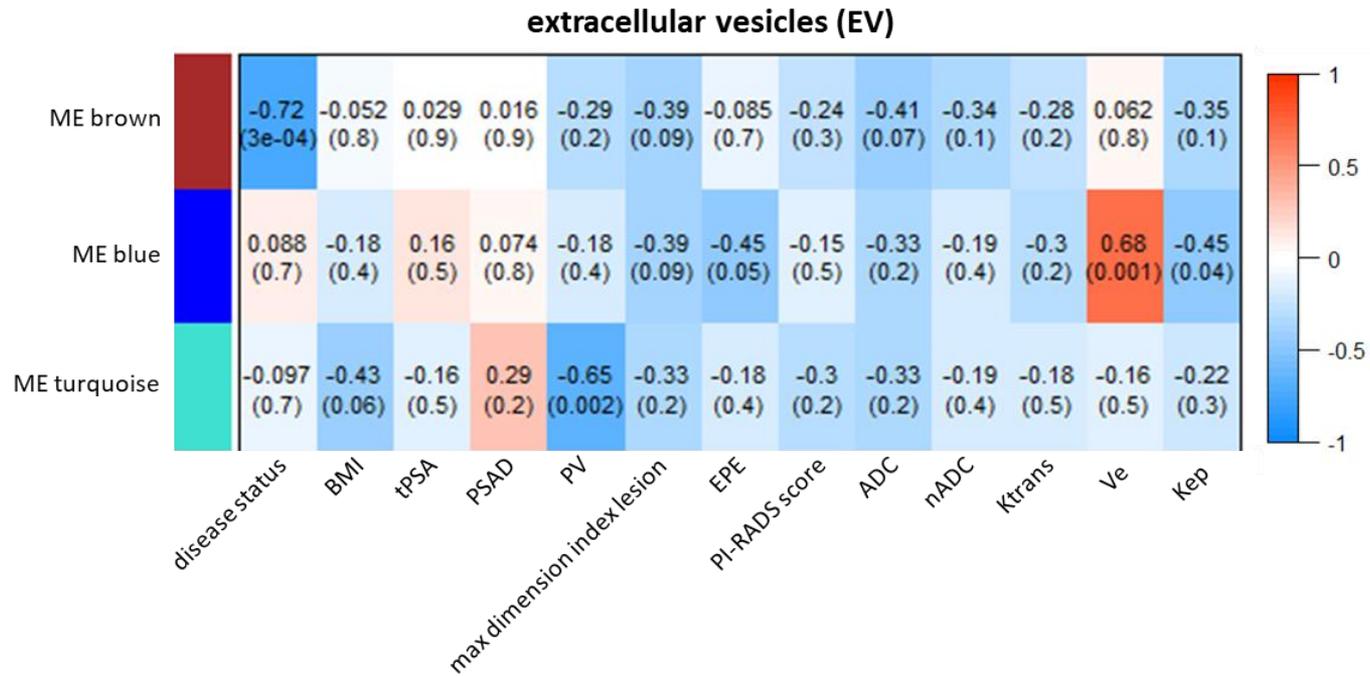
# miRNA expression data



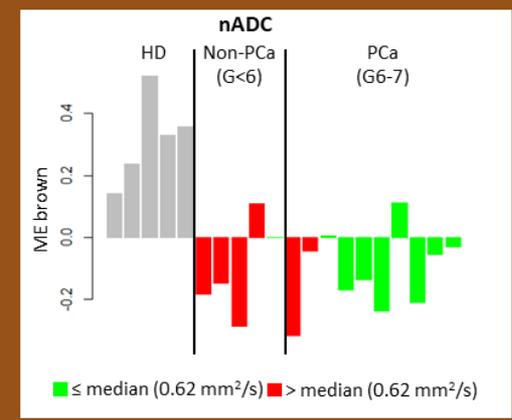
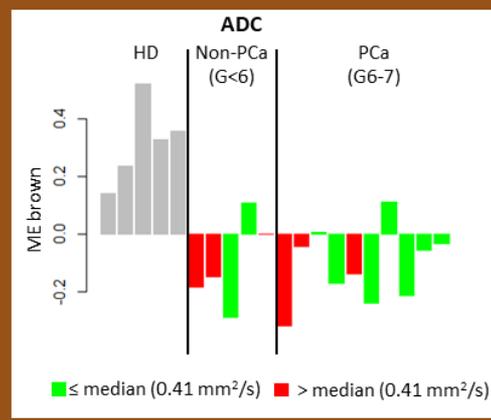
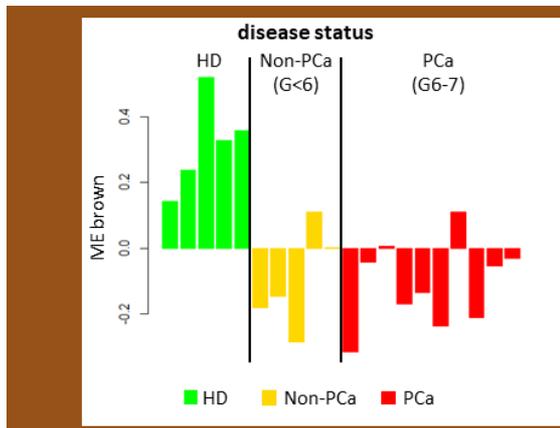
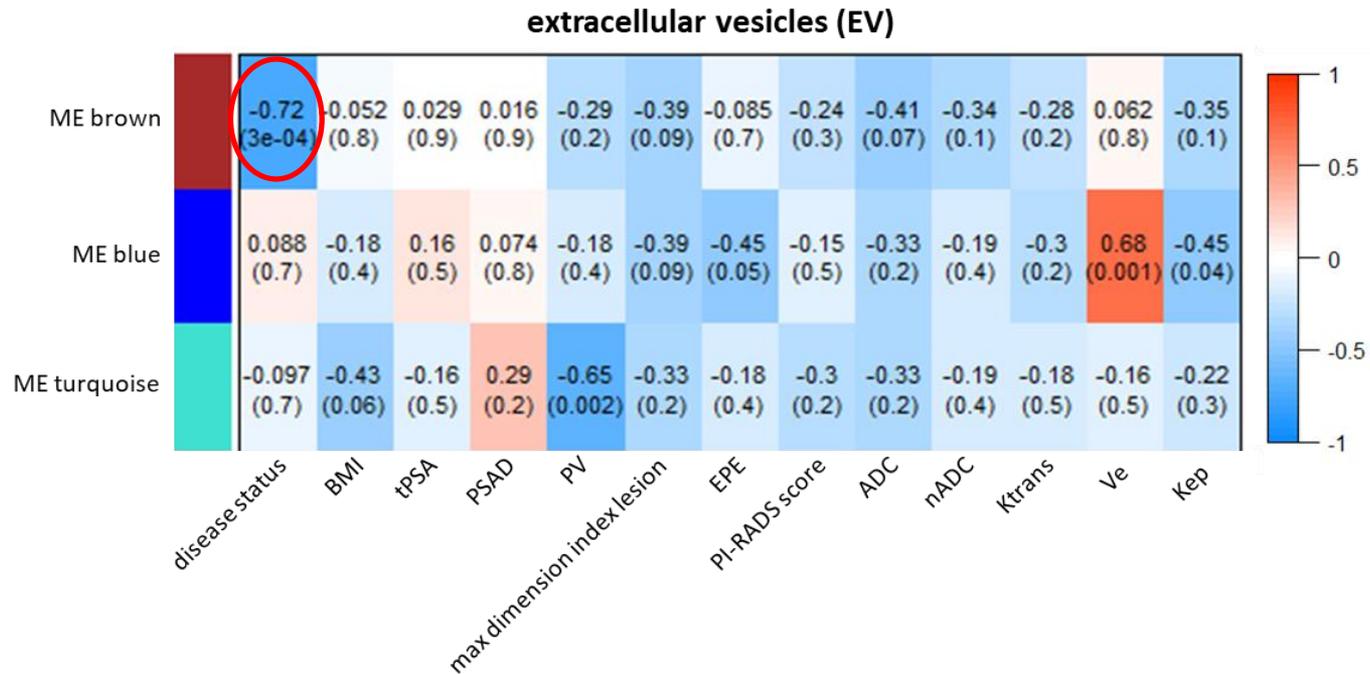
Legend label color:

- Gleason score 7
- Gleason score 6
- Gleason score < 6
- healthy donors

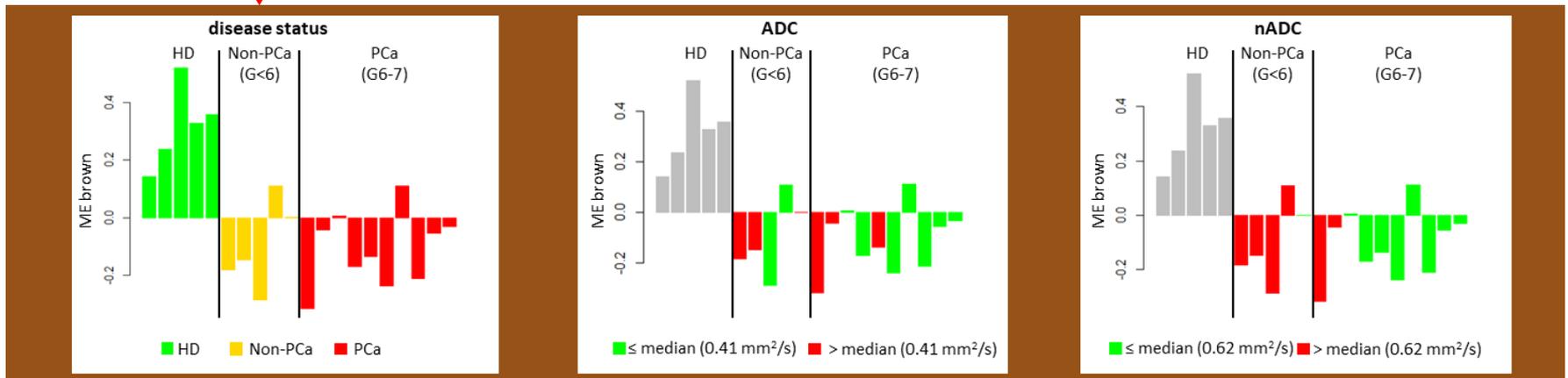
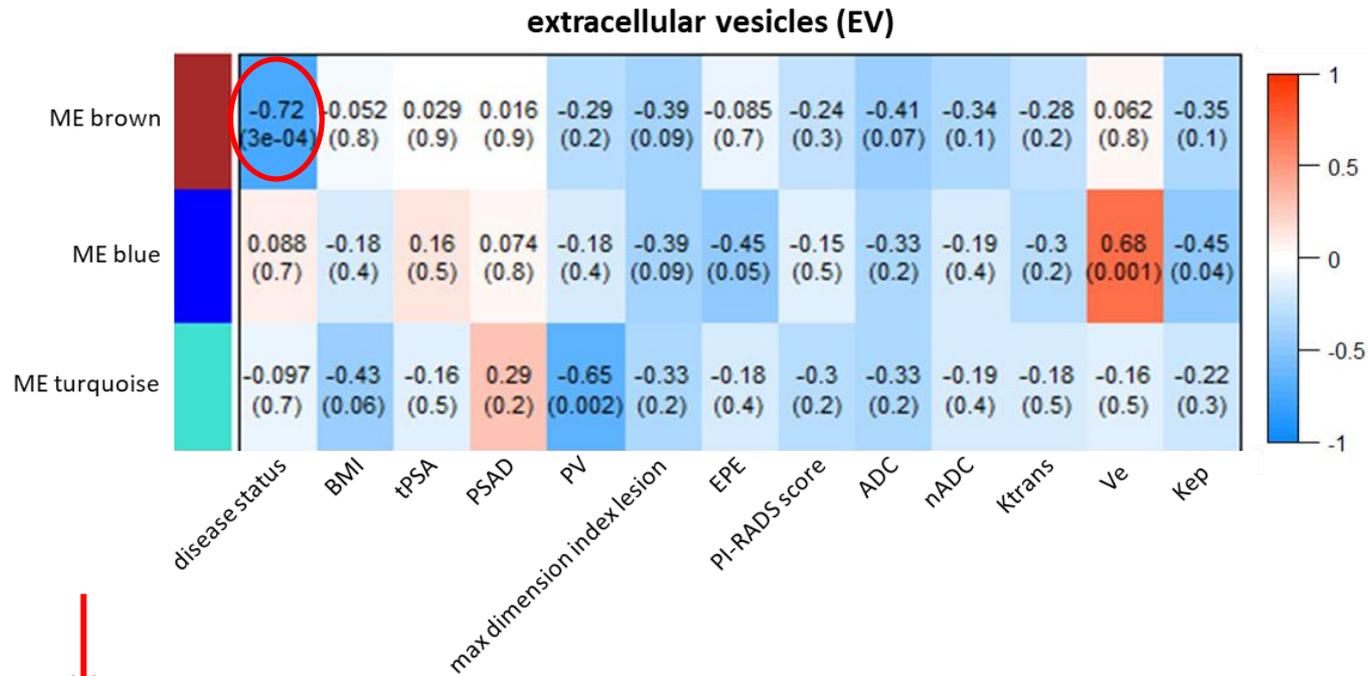
# WGCNA on all cohort of patients



# WGCNA on all cohort of patients

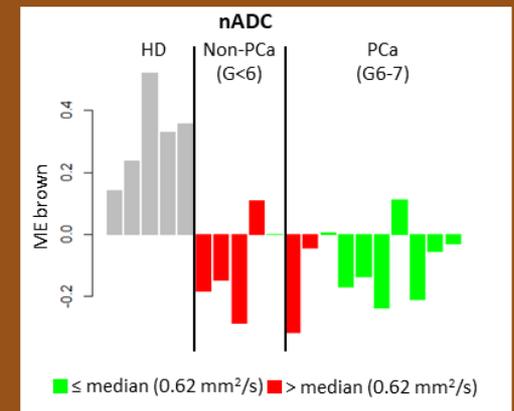
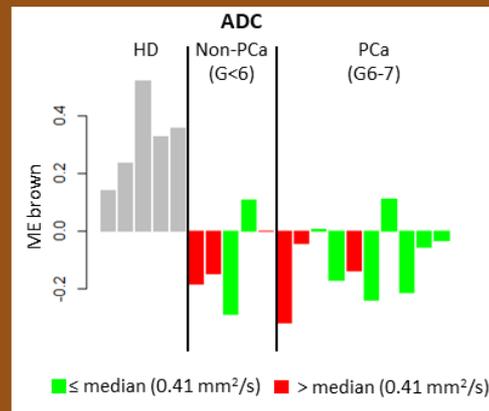
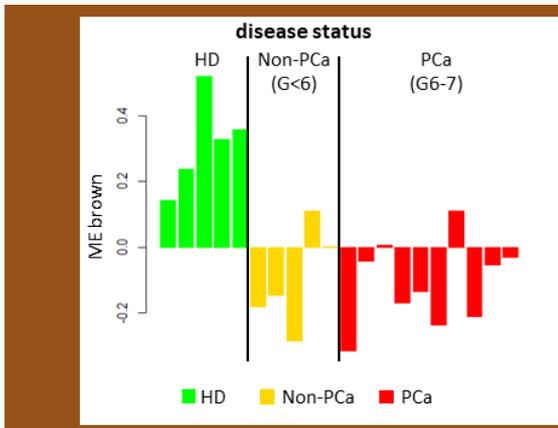
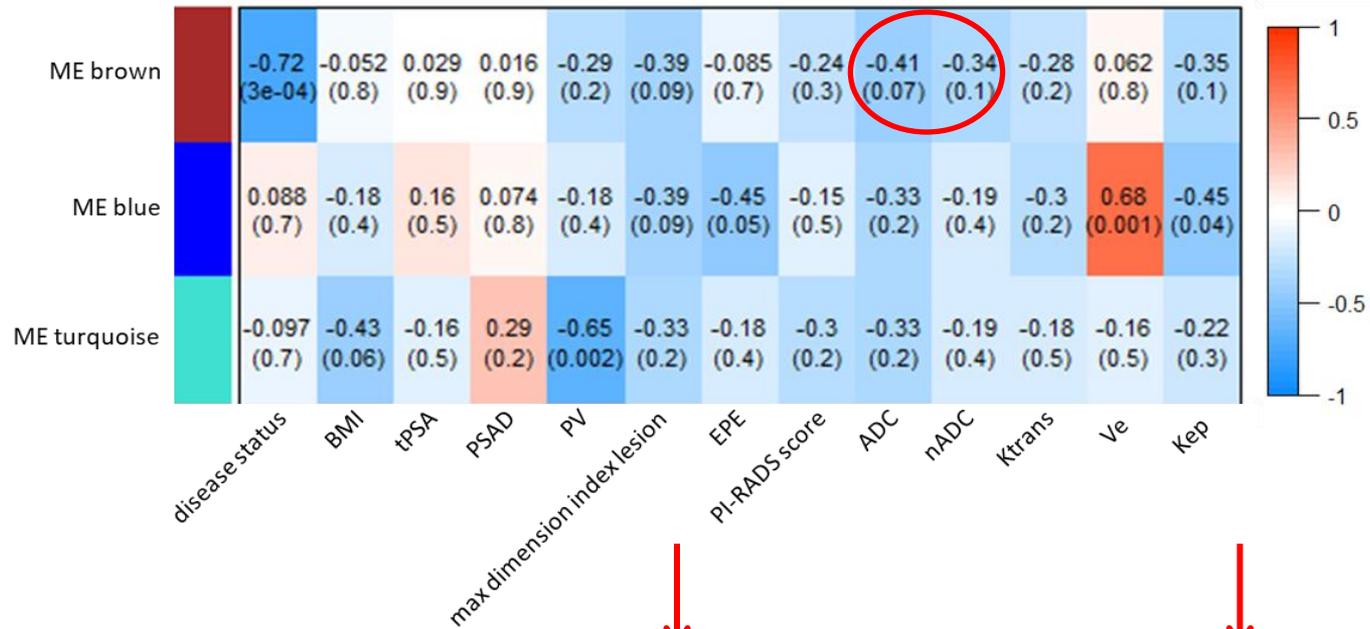


# WGCNA on all cohort of patients

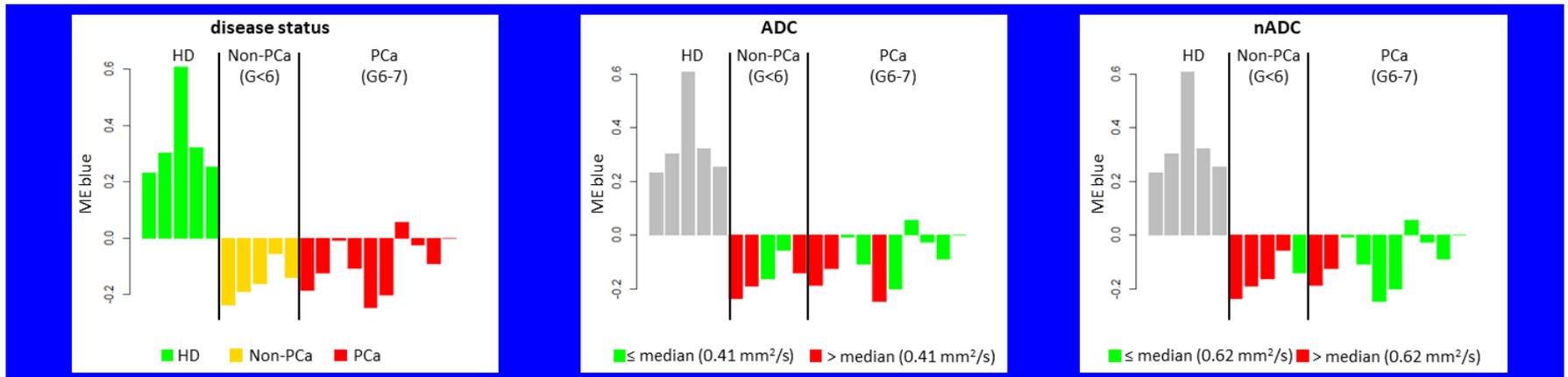
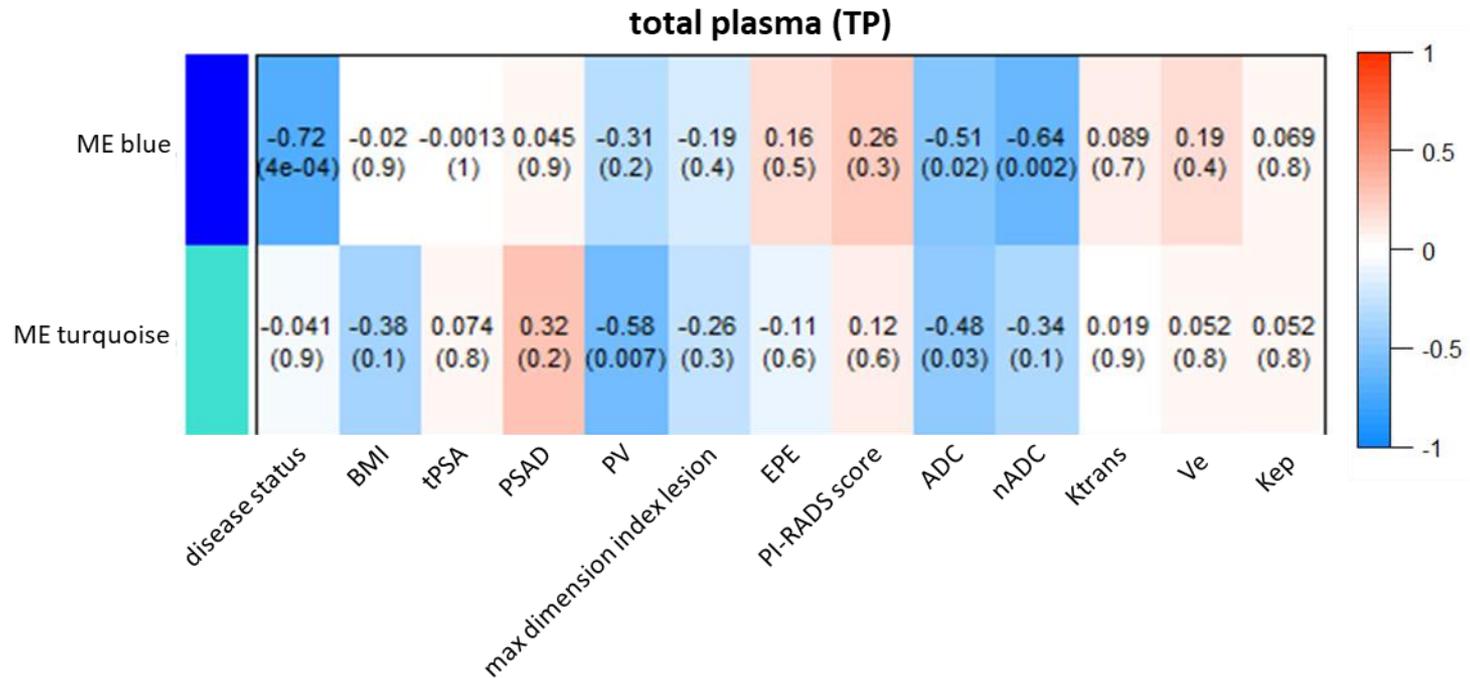


# WGCNA on all cohort of patients

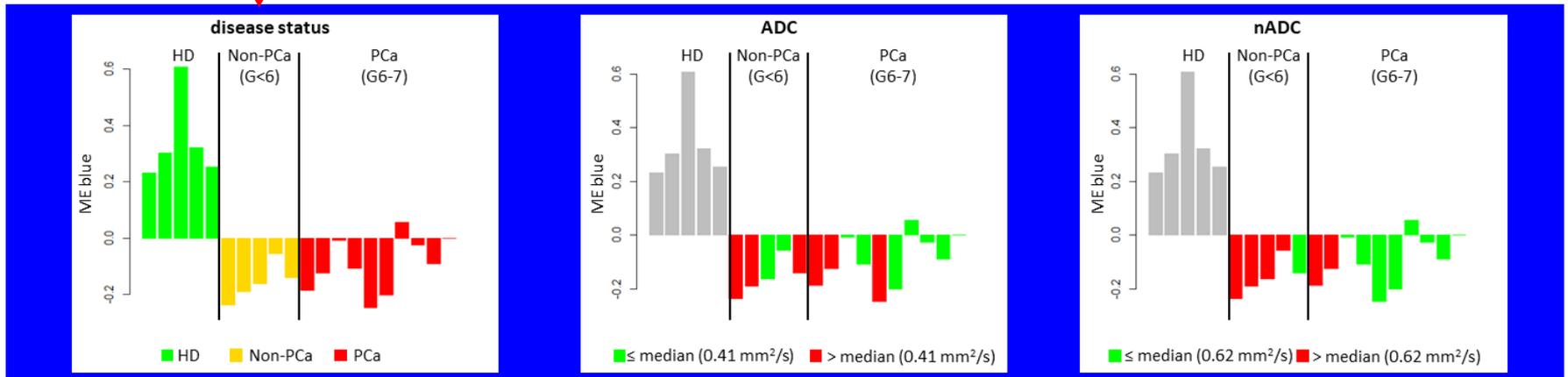
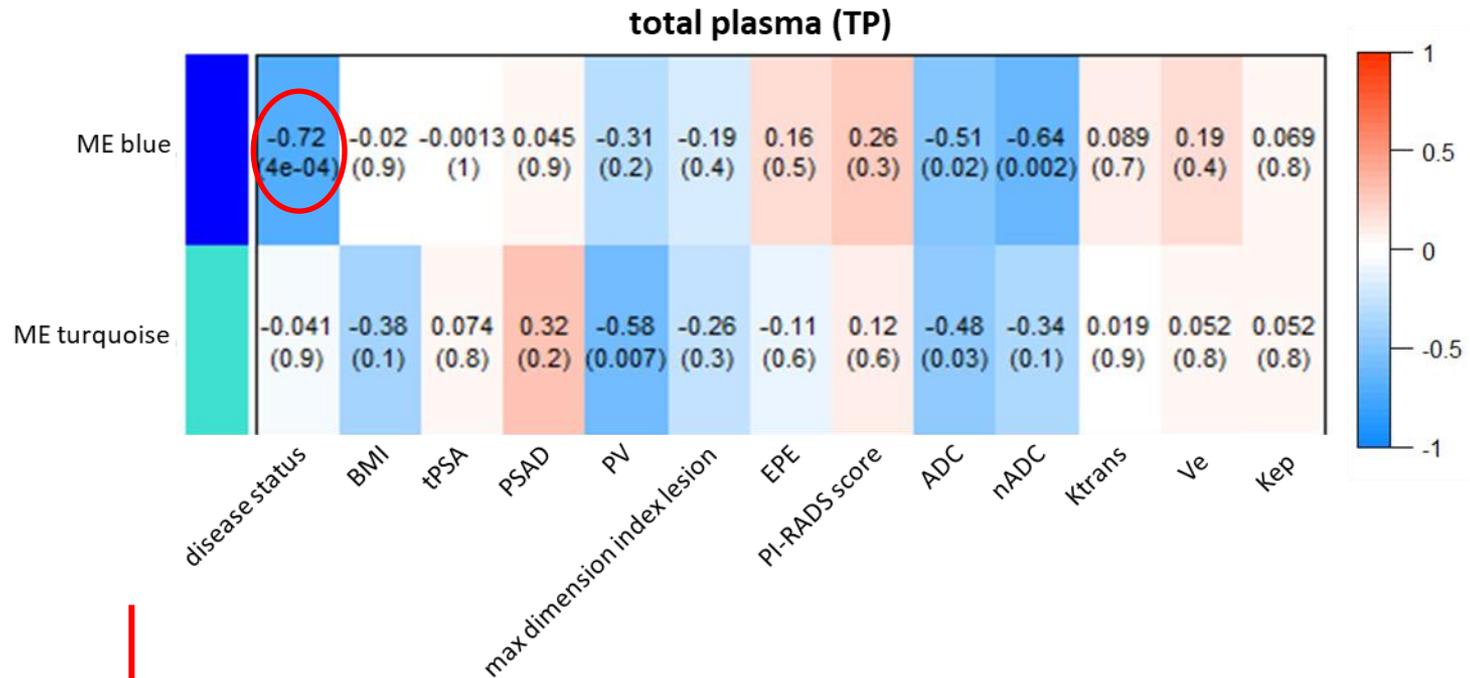
extracellular vesicles (EV)



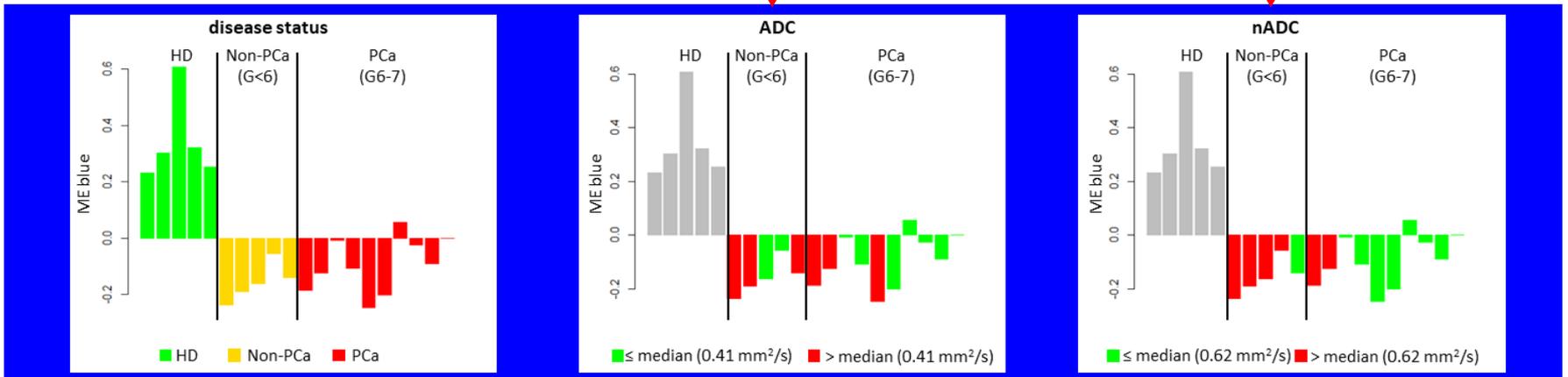
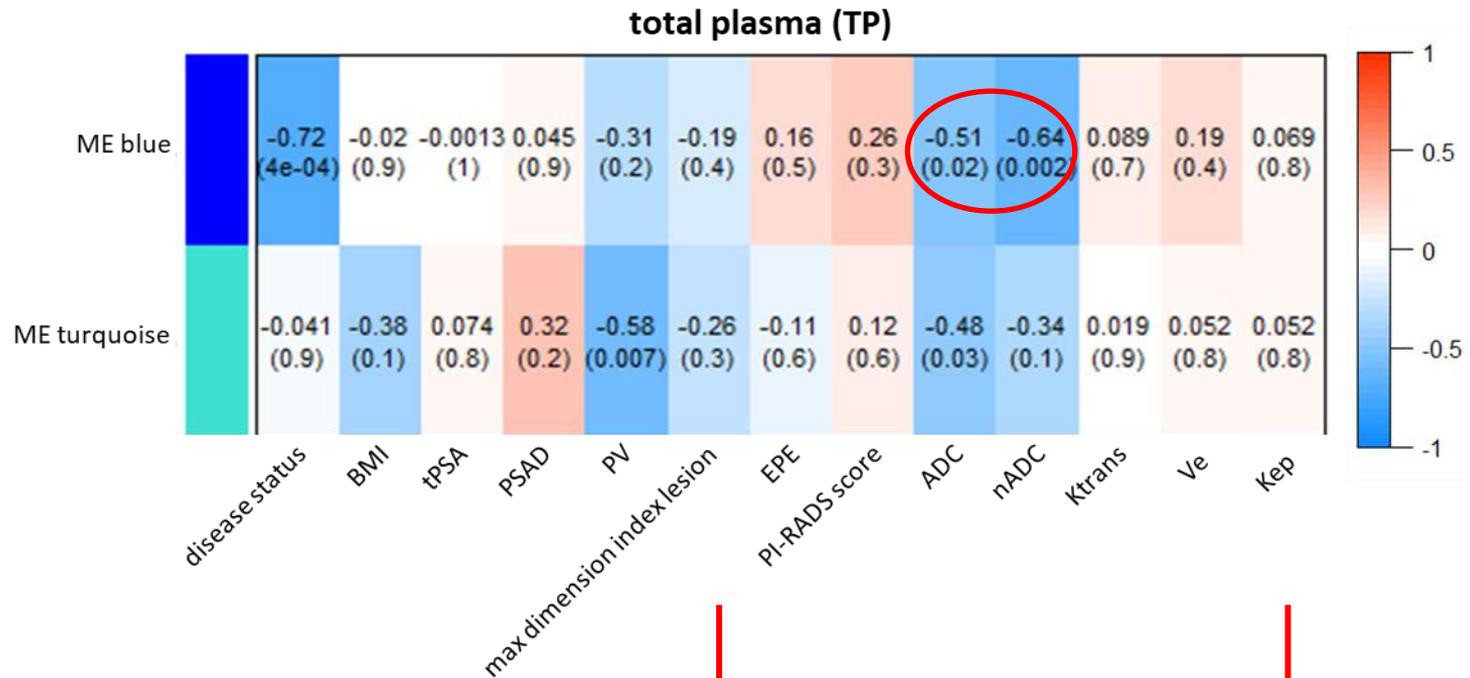
# WGCNA on all cohort of patients



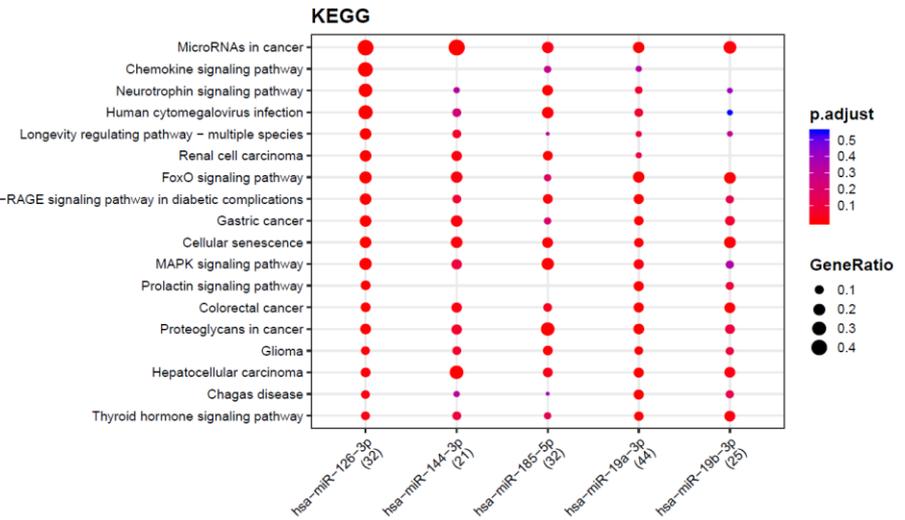
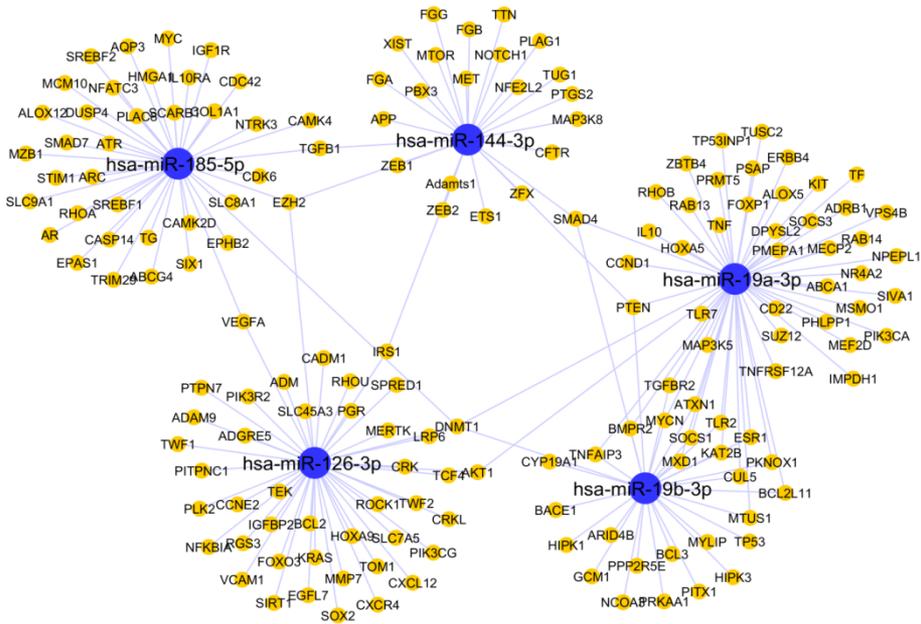
# WGCNA on all cohort of patients



# WGCNA on all cohort of patients

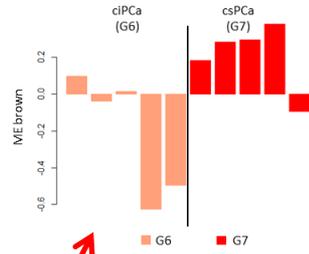


# WGCNA on all cohort of patients



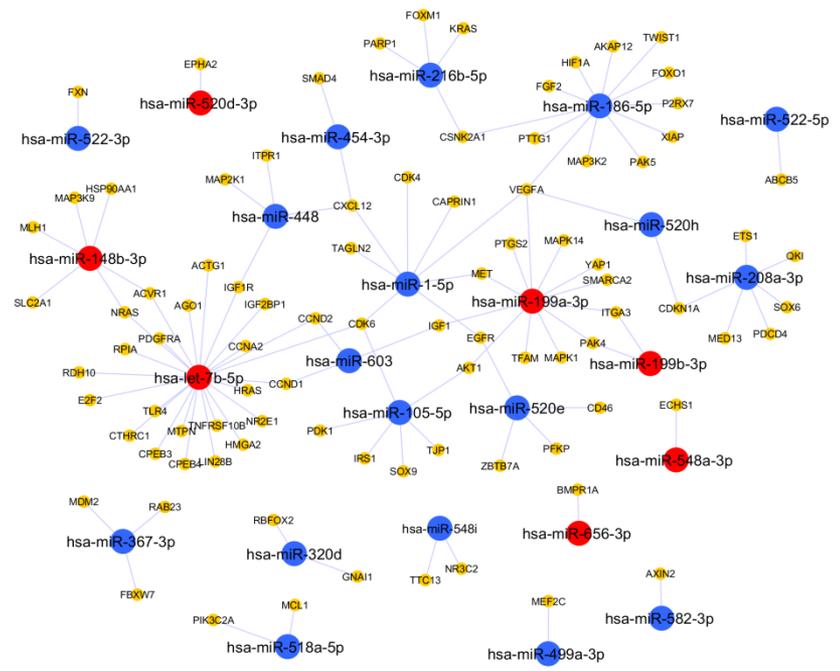
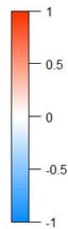


# WGCNA on PCa (G6-7) patients



**extracellular vesicles (EV)**

ME brown	0.66 (0.04)	0.51 (0.1)	0.087 (0.8)	-0.088 (0.8)	0.46 (0.2)	-0.49 (0.1)	0.086 (0.8)	0.42 (0.2)	0.57 (0.08)	-0.16 (0.6)	-0.14 (0.7)	0.37 (0.3)	-0.56 (0.09)	0.49 (0.1)	0.088 (0.8)	-0.064 (0.9)
ME yellow	-0.15 (0.7)	-0.39 (0.3)	-0.17 (0.6)	0.2 (0.6)	0.035 (0.9)	-0.075 (0.8)	-0.47 (0.2)	-0.55 (0.1)	-0.17 (0.6)	-0.55 (0.1)	-0.19 (0.6)	-0.34 (0.3)	0.61 (0.06)	-0.48 (0.2)	-0.22 (0.5)	-0.19 (0.6)
ME blue	0.32 (0.4)	-0.2 (0.6)	-0.46 (0.2)	0.33 (0.4)	0.43 (0.2)	-0.53 (0.1)	-0.46 (0.2)	-0.28 (0.4)	-0.12 (0.8)	-0.52 (0.1)	-0.57 (0.09)	-0.23 (0.5)	-0.4 (0.2)	-0.2 (0.6)	0.093 (0.8)	0.15 (0.7)
ME green	-0.12 (0.7)	-0.069 (0.8)	-0.18 (0.6)	0.18 (0.6)	0.37 (0.3)	-0.14 (0.7)	-0.4 (0.3)	-0.24 (0.5)	-0.13 (0.7)	-0.31 (0.4)	-0.17 (0.6)	-0.11 (0.8)	-0.3 (0.4)	-0.11 (0.8)	-0.17 (0.6)	0.072 (0.8)
ME turquoise	0.18 (0.6)	0.048 (0.9)	-0.48 (0.2)	-0.31 (0.4)	0.17 (0.6)	-0.54 (0.1)	-0.24 (0.5)	-0.038 (0.9)	-0.16 (0.7)	-0.53 (0.1)	-0.62 (0.06)	-0.046 (0.9)	-0.34 (0.3)	-0.05 (0.9)	-0.091 (0.8)	0.2 (0.6)
	grading	age	BMI	tPSA	PSAD	PV	max dimension index lesion	EPE	PI-RADS score	ADC	nADC	Ktrans	Ve	Keap	neoplastic cell percentage	overall mcl

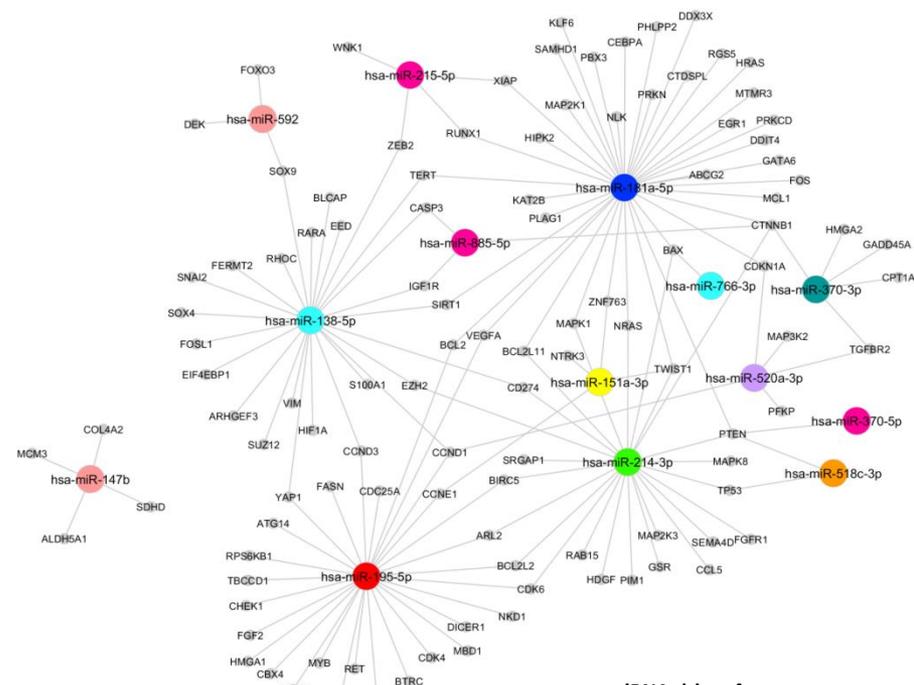
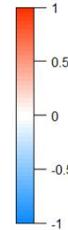
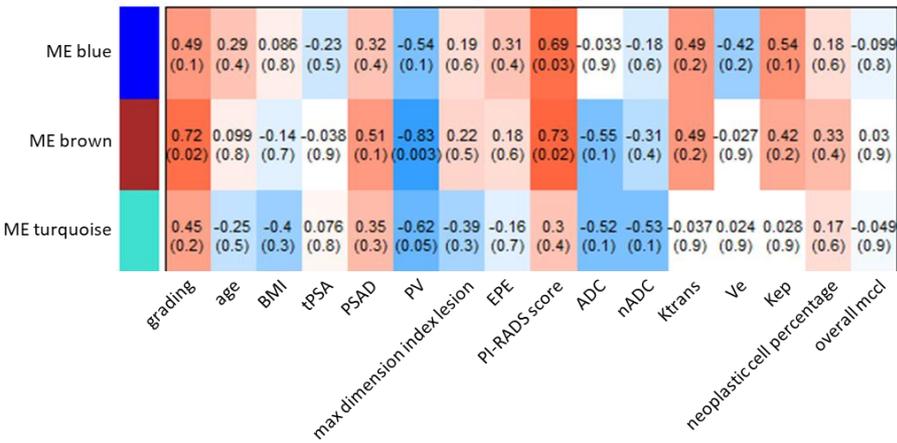


miRNA driver for:  
● grading

# WGCNA on PCa (G6-7) patients



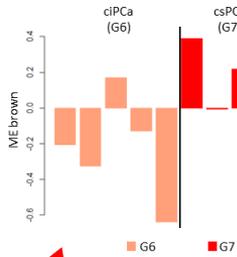
total plasma (TP)



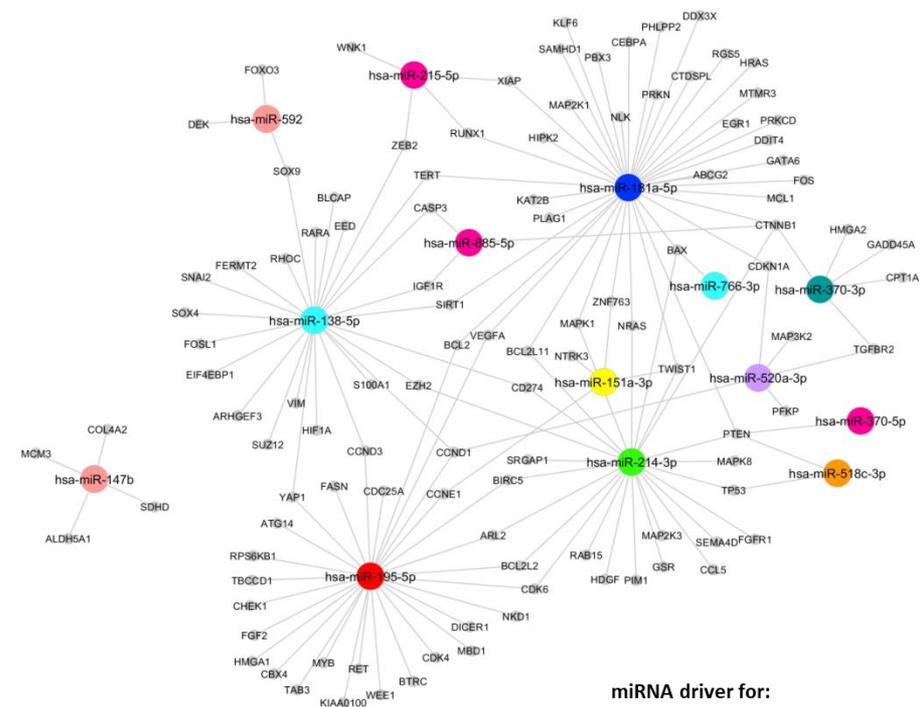
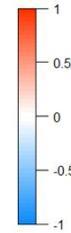
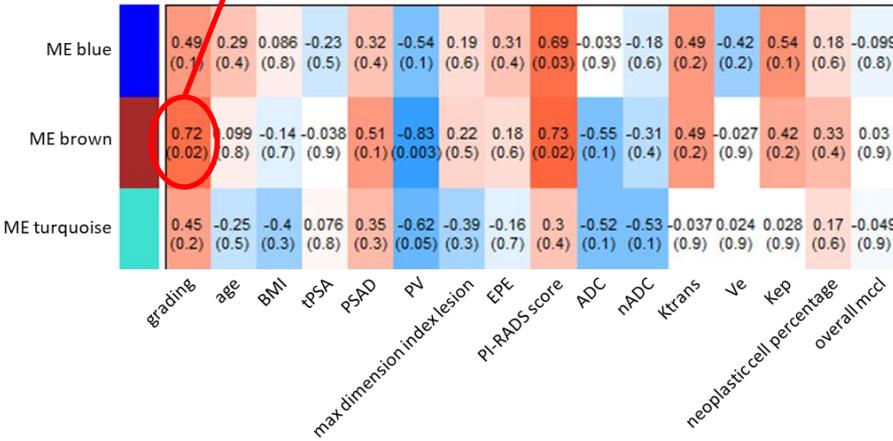
miRNA driver for:

- grading
- PV
- PSAD
- grading, ADC
- PI-RADS, PSAD, Kep
- PI-RADS, Kep, Ktrans
- PSAD, PV, ADC
- PI-RADS, PSAD, PV, Kep, Ktrans
- PI-RADS, PSAD, PV, ADC, Kep
- grading, PI-RADS, ADC, Kep, Ktrans

# WGCNA on PCa (G6-7) patients



total plasma (TP)



- miRNA driver for:**
- grading
  - PV
  - PSAD
  - grading, ADC
  - PI-RADS, PSAD, Kep
  - PI-RADS, Kep, Ktrans
  - PSAD, PV, ADC
  - PI-RADS, PSAD, PV, Kep, Ktrans
  - PI-RADS, PSAD, PV, ADC, Kep
  - grading, PI-RADS, ADC, Kep, Ktrans

# Conclusions

- Our study enabled the identification of several miRNAs with good diagnostic potential when used as a single biomarker or in combination with other imaging/clinical parameters.
- After an experimental validation, they could be suitable for non-invasive diagnosis of PCa using blood as a source of cancer-derived transcripts.
- The clinical implications of this approach would be a dramatic reduction in the number of unnecessary biopsies and a more accurate patient stratification, thus reducing overdiagnosis, overtreatment, and overall related-costs.

## **4. WGCNA application in grapevine**

# Working in progress

young leaf

senescent leaf

veraison soft

young leaf

senescent leaf

veraison soft

	cf.yl.rep1	cf.yl.rep2	cf.yl.rep3	cf.sl.rep1	cf.sl.rep2	cf.sl.rep3	cf.vs.rep1	cf.vs.rep2	cf.vs.rep3	rk.yl.rep1	rk.yl.rep2	rk.yl.rep3	rk.sl.rep1	rk.sl.rep2	rk.sl.rep3	rk.vs.rep1	rk.vs.rep2	rk.vs.rep3
VIT_201s0011g00010	67.64305	84.40926	46.56563	100.756	109.8212	72.31631	167.3562	172.0807	123.0081	105.5222	98.65743	144.002	151.6068	123.1155	134.4458	110.2557	117.2131	120.8792
VIT_201s0011g00030	23.97475	28.96396	72.70773	1.538259	21.96424	10.19845	74.58265	66.92026	57.24139	7.914168	25.36905	39.65272	22.15792	19.52867	12.15329	31.89215	43.66763	56.16609
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

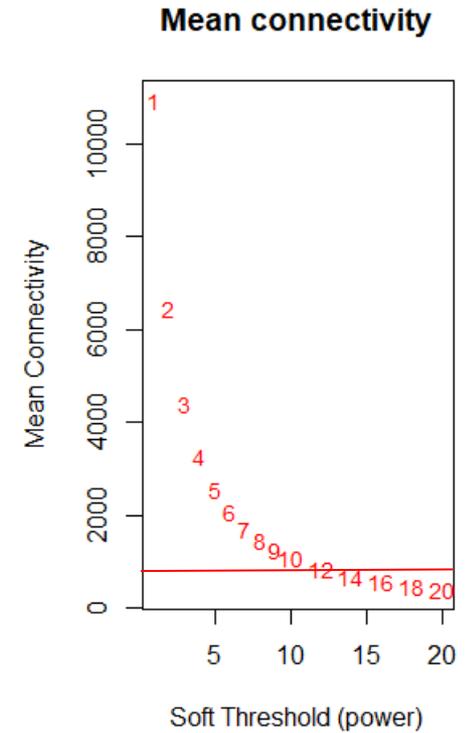
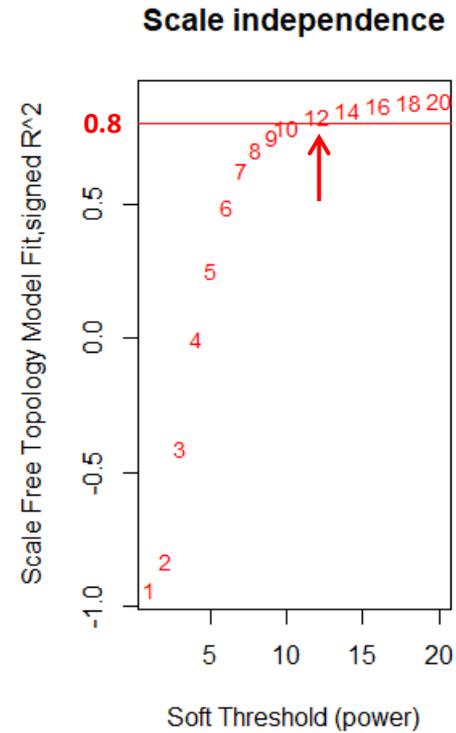
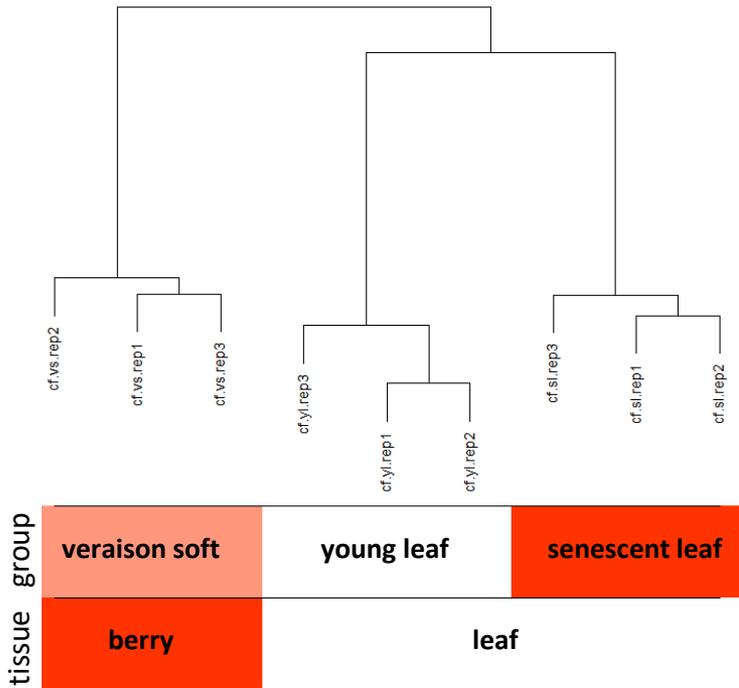
**cabernet**



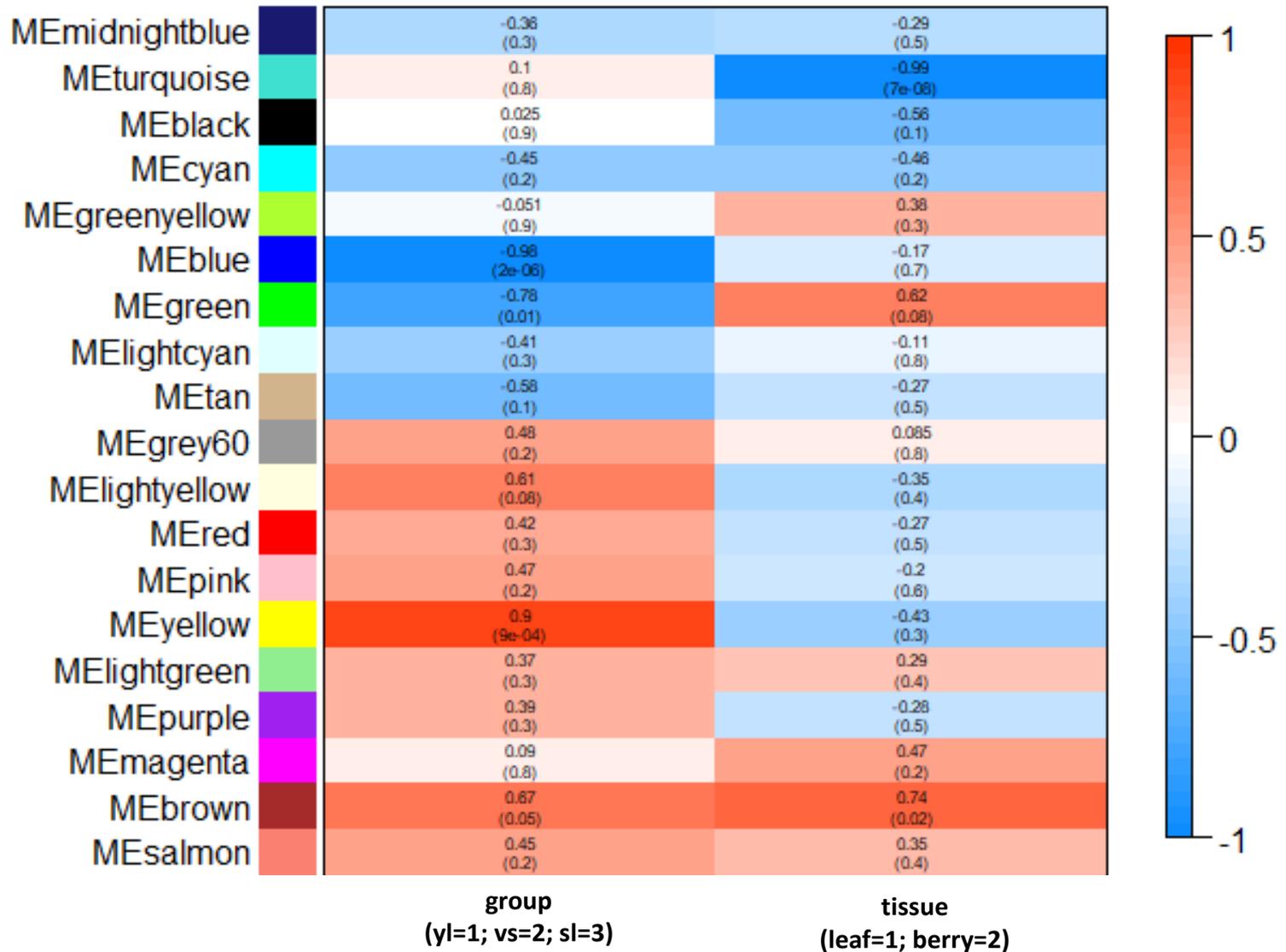
**rkatsiteli**



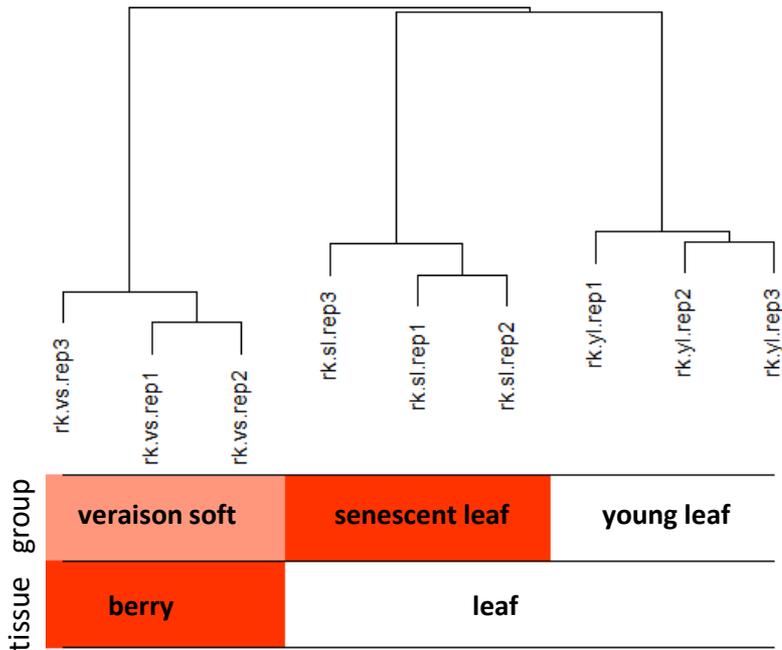
# WGCNA on cabernet



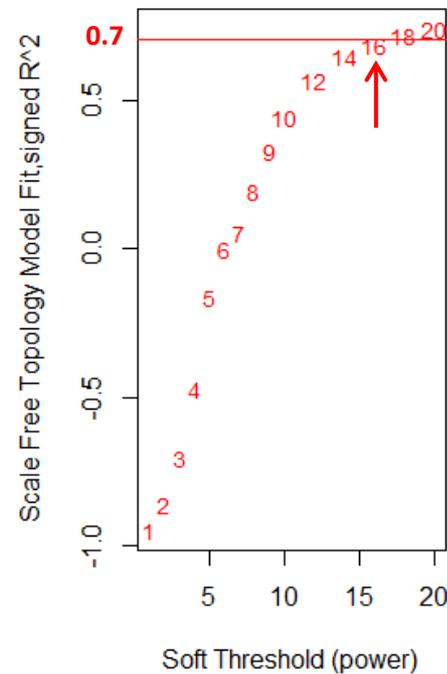
# WGCNA on cabernet



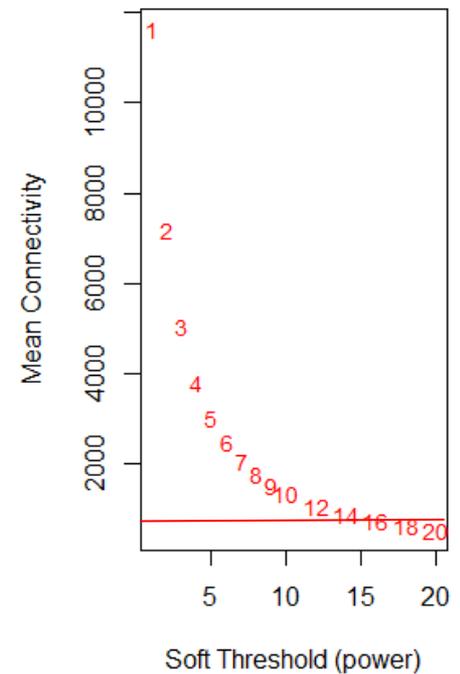
# WGCNA on rkatsiteli



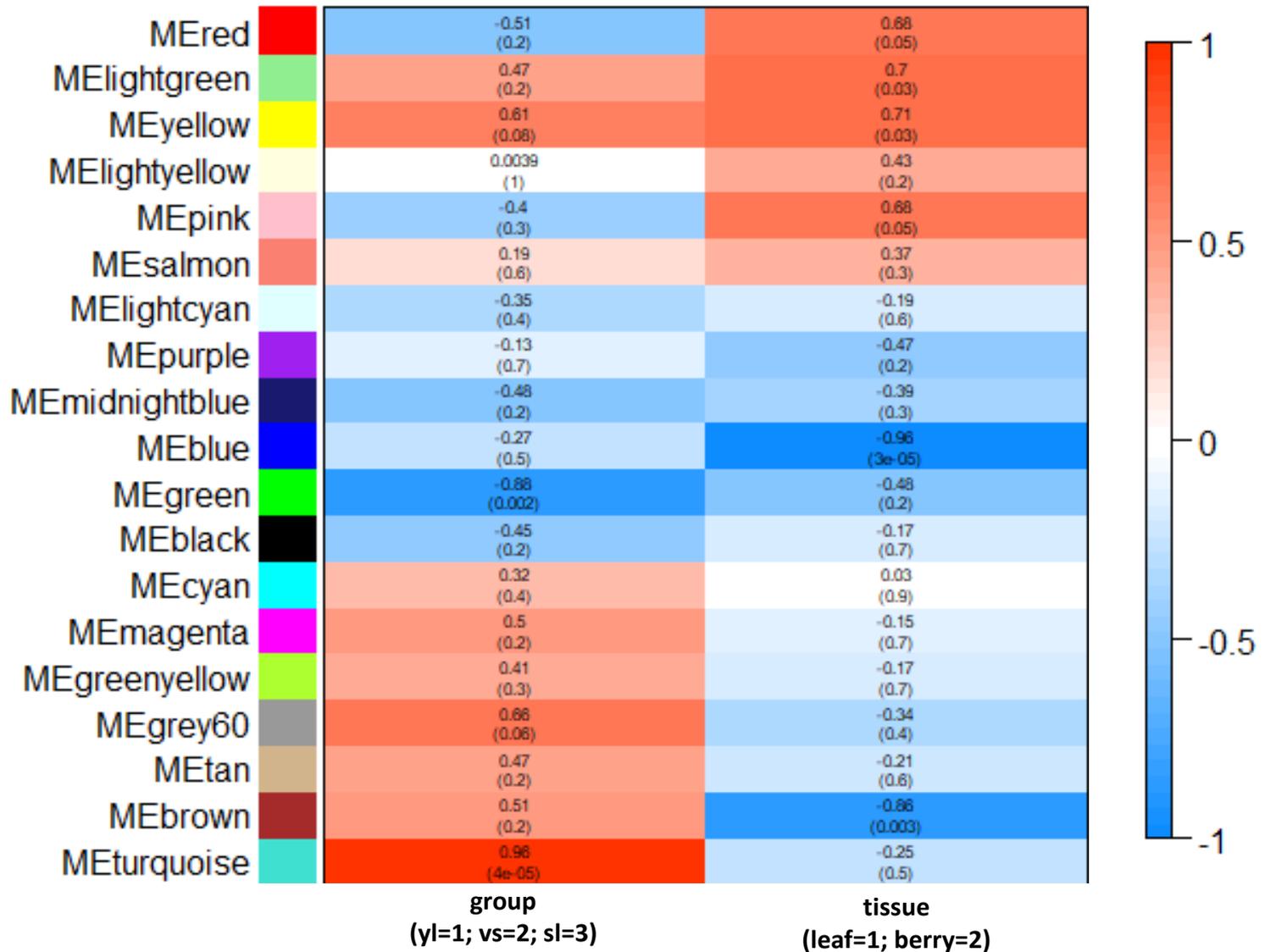
Scale independence



Mean connectivity



# WGCNA on rkatsiteli





THANK  
YOU