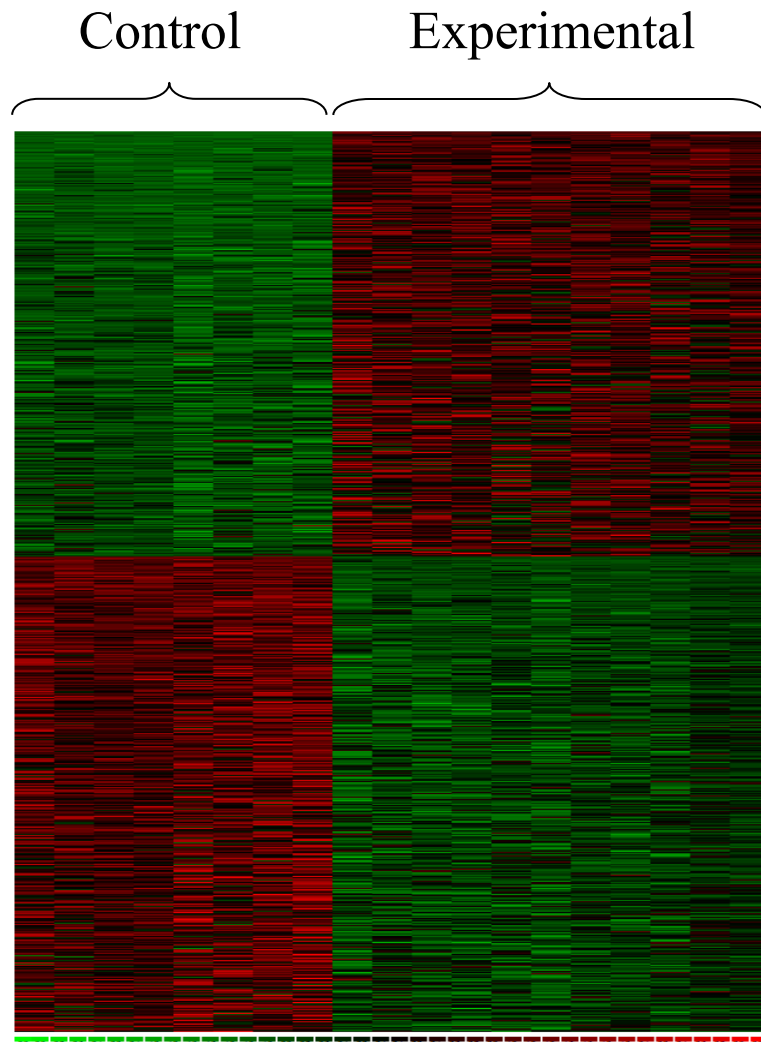# Weighted Correlation Network Analysis and Systems Biologic Applications

Standard differential expression analyses seek to identify individual genes

Control   Experimental

- Each gene is treated as an individual entity

- Often misses the forest for the trees: <u>Fails to recognize that thousands of genes can be organized into relatively few modules</u>

Philosophy of Weighted Gene Co-Expression Network Analysis

- Understand the "system" instead of reporting a list of individual parts
  - Describe the functioning of the engine instead of enumerating individual nuts and bolts
- Focus on modules as opposed to individual genes
  - this greatly alleviates multiple testing problem
- Network terminology is intuitive to biologists
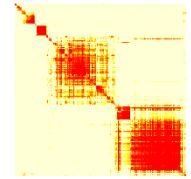
Flight connections and hub airports
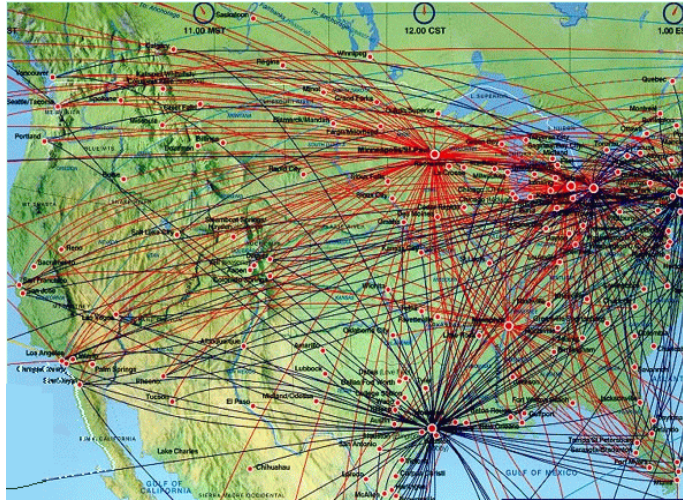
The nodes with the largest number of links (connections) are most important!

What is weighted gene co-expression network analysis?

Construct a network
Rationale: make use of interaction patterns between genes

Identify modules
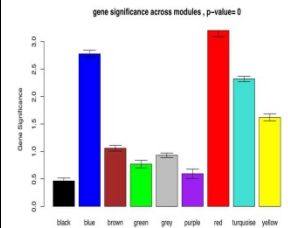Rationale: module (pathway) based analysis

Relate modules to external information
Array Information: Clinical data, SNPs, proteomics
Gene Information: gene ontology, EASE, IPA
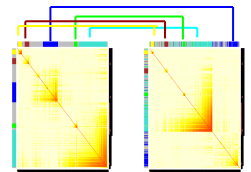Rationale: find biologically interesting modules

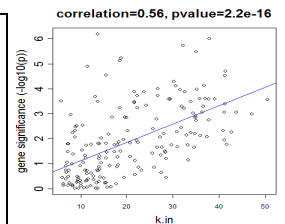## Study Module Preservation across different data
Rationale:
- Same data: to check robustness of module definition
- Different data: to find interesting modules.

Find the key drivers in *interesting* modules
Tools: intramodular connectivity, causality testing
Rationale: experimental validation, therapeutics, biomarkers

Weighted correlation networks are valuable for a biologically meaningful...

- <u>reduction of high dimensional data</u>
  - expression: microarray, RNA-seq
  - gene methylation data, fMRI data, etc.
- <u>integration of multiscale data</u>
  - expression data from multiple tissues
  - SNPs (module QTL analysis)
  - Complex phenotypes

Review of weighted correlation network analysis

Content

Comparison of co-expression measures: mutual information, correlation, and model based indices
- L. Song, P. Langfelder et al 2012 BMC Bioinformatics (2012), 13:328.

Comparing Statistical Methods for Constructing Large Scale Gene Networks (2012)

Allen JD, Xie Y, Chen M, Girard L, Xiao G PLoS ONE 7(1): e29348. doi:10.1371

Aging effects on DNA methylation modules in human brain and blood tissue.
Genome Biol. 2012 Oct 3;13(10):R97. PMID: 23034122

When Is Hub Gene Selection Better than Standard Meta-Analysis?
Langfelder et al (2013) PLoS ONE 8(4): e61505.

Content

Comparison of co-expression measures: mutual information, correlation, and model based indices

    –   L. Song, P. Langfelder et al 2012 BMC Bioinformatics (2012), 13:328.

Network=Adjacency Matrix

- A network can be represented by an adjacency matrix, $A=[a_{ij}]$, that encodes whether/how a pair of nodes is connected.
  - A is a symmetric matrix with entries in [0,1]
  - For unweighted network, entries are 1 or 0 depending on whether or not 2 nodes are adjacent (connected)
  - For weighted networks, the adjacency matrix reports the connection strength between node pairs
  - Our convention: diagonal elements of *A* are all 1.

Two types of weighted correlation networks

Unsigned network, absolute value

$$a_{ij} = | cor(x_i, x_j) |^{\beta}$$

Signed network preserves sign info

$$a_{ij} = | 0.5 + 0.5 \times cor(x_i, x_j) |^{\beta}$$

Default values: β=6 for unsigned and β =12 for signed networks.
We prefer signed networks...

Zhang et al SAGMB Vol. 4: No. 1, Article 17.

Mutual information: an information theory concept

- Measures the mutual dependency of two random variables.
- For categorical variable DX, DY:

$$MI(dx, dy) = \sum_{r=1}^{no.levels.dx} \sum_{c=1}^{no.levels.dy} p(ldx_r, ldy_c) log \left(\frac{p(ldx_r, ldy_c)}{p(ldx_r)p(ldy_c)} \right) \; > =0$$

$$MI(dx, dy) = Entropy(dx) + Entropy(dy) - Entropy(dx, dy)$$

$p(ldx_r)$, $p(ldy_c)$ stands for the frequency of the r-th and c-th level of dx, dy.
Entropy(dx, dy) stands for the joint entropy of variable dx and dy.

- For continuous variable, e.g. gene expression data, it is not straightforward to estimate MI.
  - Need to choose variable discretization method.
  - Need to choose entropy estimation method.
  - Estimate of MI depends on sample size: poor estimate for small sample size
  - Estimation is computational challenging.

Advantages of the correlation coefficient and mutual information when it comes to relating two numeric variables

Correlation

Requires few observations

No hidden parameters

Easy p-value calculation

Sign of relations

Robust to outliers*

* Spearman correlation and biweight midcor are particularly robust.

Mutual Information

Can detect general, non-linear relationship

*Biweight midcorrelation (bicor)

- A robust alternative to the Pearson correlation.

- Definition based on
  - median instead of mean.
  - Assign weights to observations, values close to median receive large weights.

$$covMedianWeighted(x,y) = \frac{sum((x-median(x))w.x(y-median(y))w.y)}{\sqrt{sum(w.x^2)sum(w.y^2)}},$$
(5.16)

where the weights are given by $w.x = weight(robustScale(x))$.

Using these function, the **biweight midcorrelation** between x and y is defined as follows:

$$bicor(x,y) = \frac{covMedianWeighted(x,y)}{\sqrt{covMedianWeighted(x,x)covMedianWeighted(y,y)}}.$$
(5.17)

Reviewed in chapter 5 in book "Weighted Network Analysis"

Well known since the seventies: e.g. book: "Data Analysis and Regression: A Second Course in Statistics", Mosteller and Tukey, Addison-Wesley, 1977, pp. 203-209
Langfelder et al 2012: Fast R Functions For Robust Correlations And Hierarchical Clustering. *J Stat Softw* 2012, **46**(i11):1–17.

Here we address the following questions

1. Is correlation or mutual information a better measurement for expression data?
   - Assess gene-gene relationship
   - Identify co-expression network modules

2. Are there any alternative methods measuring non-linear relationships?
   - Polynomial and spline regression models

Define MI based adjacency

- Mutual information can take on values that are larger than 1. It needs to be transformed to turn it into an adjacency measure.
- Recall adjacency measures association between variables i and j.
  - $0 \leq A_{ij} \leq 1$
  - $A_{ij} = A_{ji}$
  - $A_{ii} = 1$

- For MI, define MI based adjacency AUV2 as:

$$A^{MI, UniversalVersion2} = \frac{MI(dx_i, dx_j)}{max(Entropy(dx_i), Entropy(dx_j))}.$$

- As a transformation of MI, AUV2 is directly comparable to absolute correlation, e.g. |bicor|.
  - AUV2(dx_i, dx_j) = AUV2(dx_j, dx_i)
  - $0 \leq$ AUV2 ≤ 1

Detect gene pairwise relationships

Empirical expression data to compare cor with MI

- 8 data sets from human, mouse and yeast.

Data set	Number of observations
Brain cancer	55
SAFHS (healthy whole blood)	1084
Neurological disease	346
Yeast	44
Mouse adipose	239
Mouse brain	221
Mouse liver	272
Mouse muscle	252

- Tissue-specific mouse data sets were generated by the lab of Dr. Jake Lusis at UCLA.

Ghazalpour A, Doss S, Zhang B, Plaisier C, Wang S, Schadt E, Thomas A, Drake T, Lusis A, Horvath S: **Integrating Genetics and Network Analysis to Characterize Genes Related to MouseWeight.** *PloS Genetics* 2006, **2**(2):8.

# Mutual information vs. biweight midcorrelation in 8 empirical data sets.

- − Circles in the plot below represent gene pairs.
- − Most gene pairs satisfy linear relationships.
- • Circles in the plot below represent gene pairs.
- • Mutual information has strong association with correlation.
- • Most gene pairs satisfy linear relations.
- • There are cases where correlation and MI strongly disagree (blue and red circles).
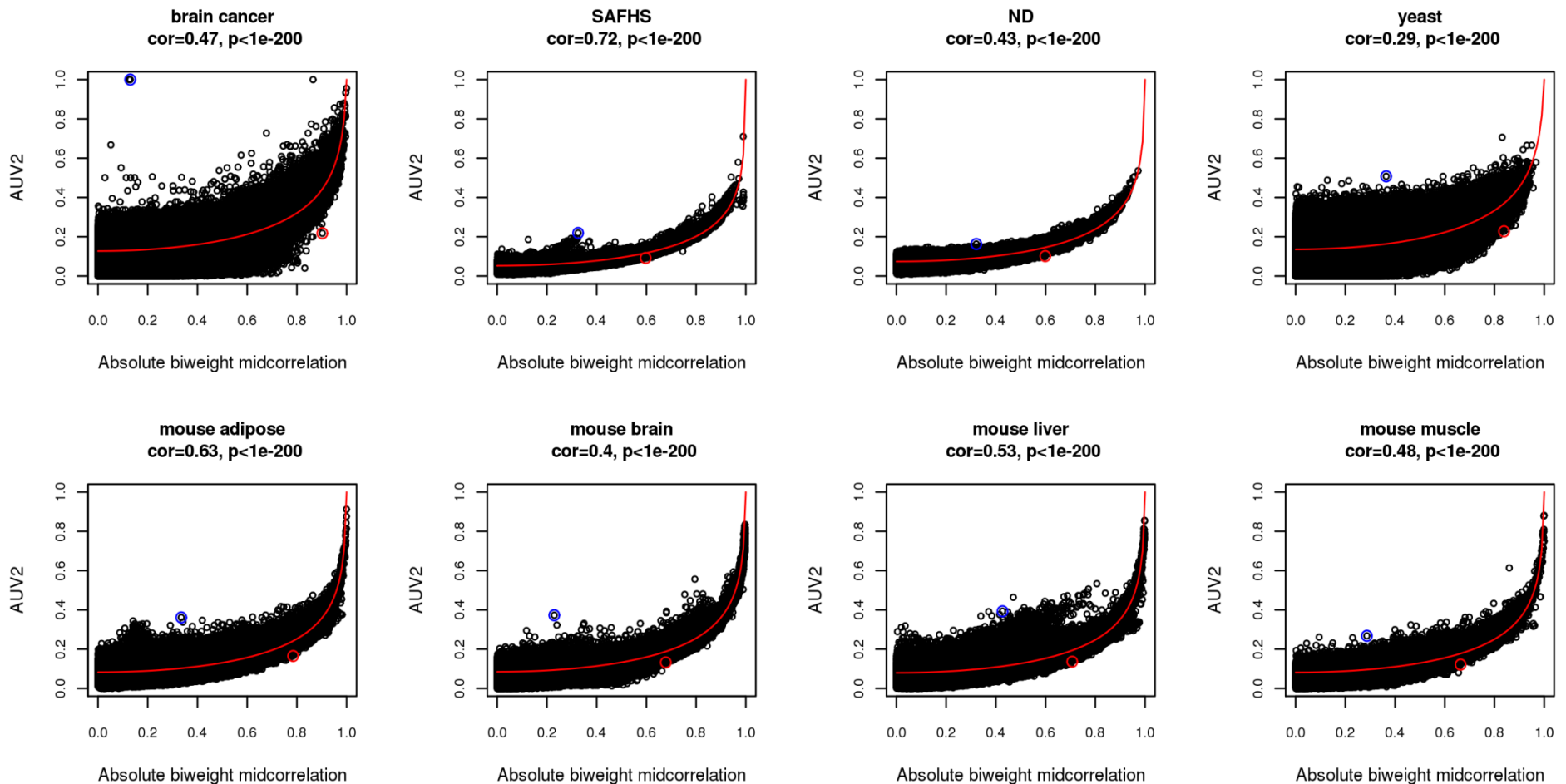
- Expression of gene pairs with most extreme AUV2 and insignificant bicor.
    - Gene pairs are taken from blue circles in the previous plot.
    - Messages: 1) this estimate of MI is susceptible to outliers, 2) MI may detect unusual dependence patterns (which might represent technical artefacts).
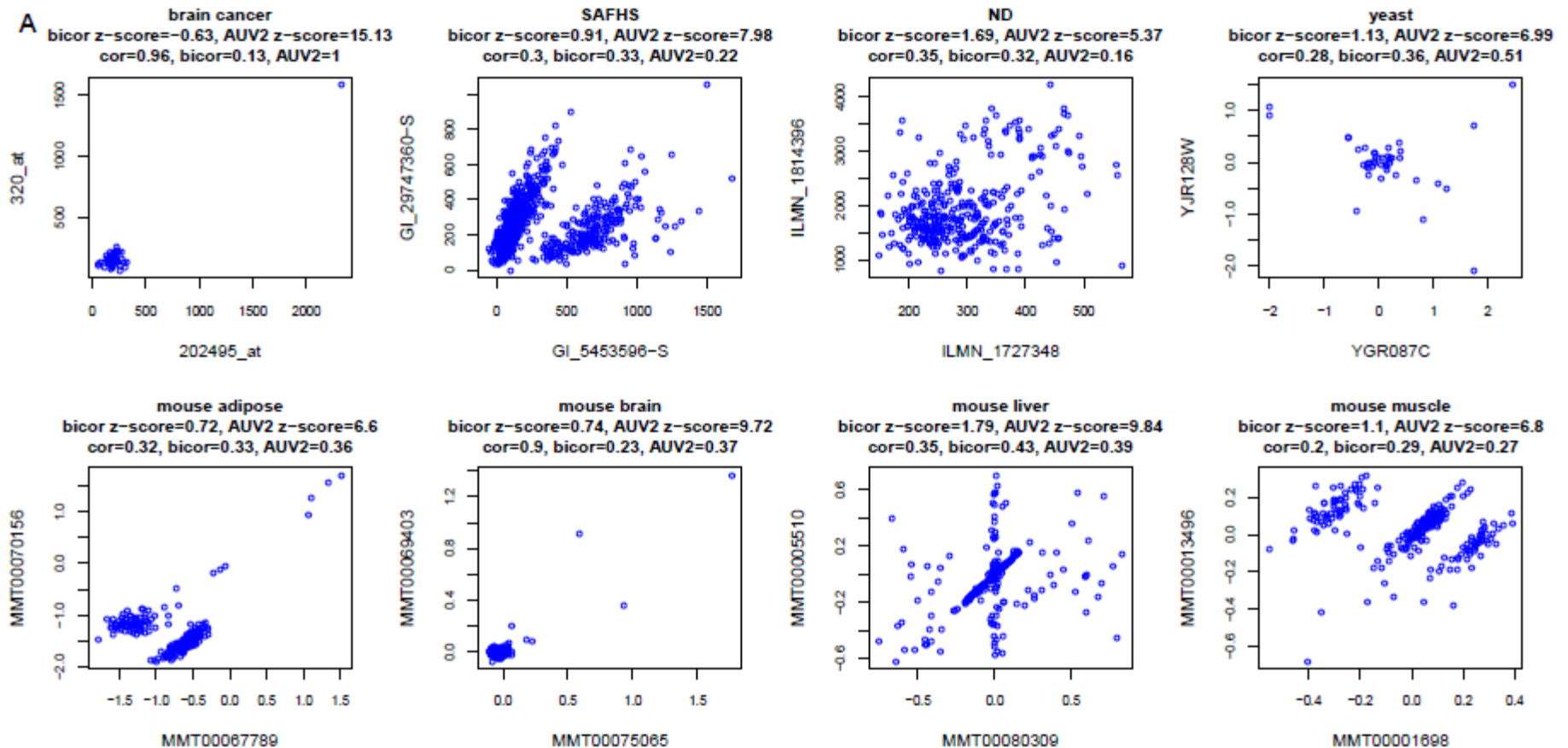
- Expression of gene pairs with most extreme bicor and insignificant AUV2.

  - Gene pairs are taken from red circles in the previous plot.

  - Message: MI misses linear relationships that are detected by bicor.
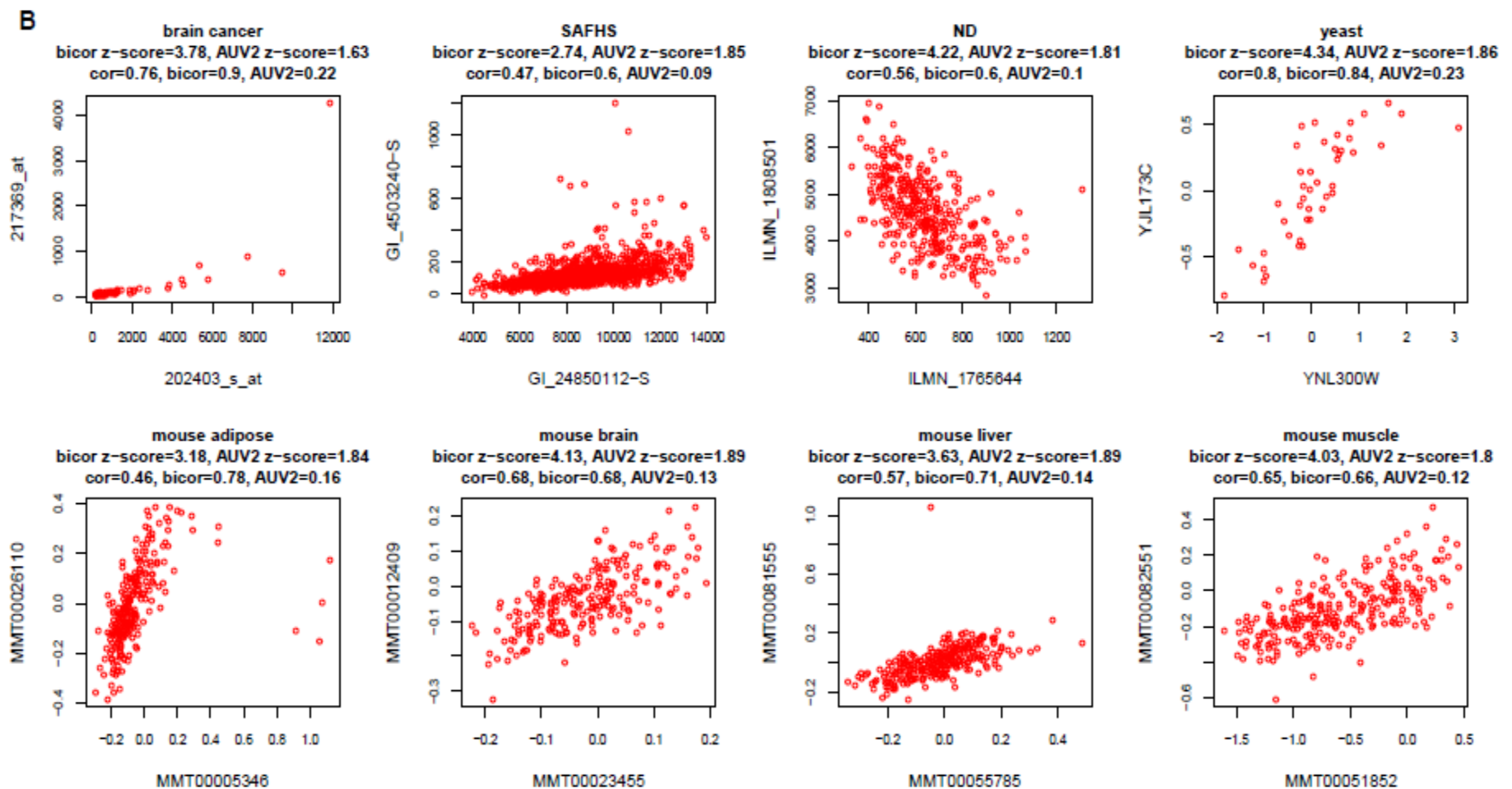
Recall that WGCNA typically uses the topological overlap dissimilarity as input of hierarchical clustering

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

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

- Generalized to the case of weighted networks in Zhang et al (2005, Statistical Applications in Genetics and Molecular Biology: Vol. 4: No. 1, Article 17).
- Generalized to higher order interactions in Yip et al (2006 BMC Bioinformatics 2007, 8:22)

Which measure is better at identifying biologically meaningful modules?

To judge the biological meaning of modules,
we assess their enrichment with respect to
gene ontology categories
as described in step 3 of the following

Construct a network
Rationale: make use of interaction patterns between genes
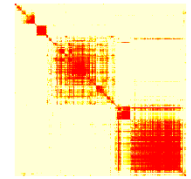
Identify modules
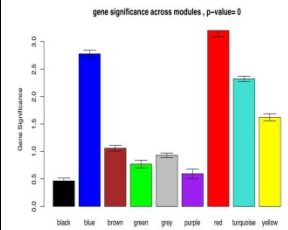Rationale: module (pathway) based analysis

Relate modules to external information
Array Information: Clinical data, SNPs, proteomics
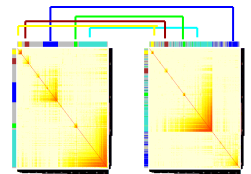Gene Information: gene ontology, EASE, IPA
Rationale: find biologically interesting modules

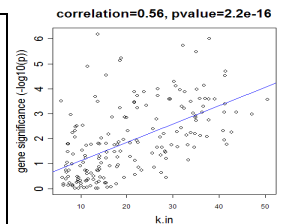## Study Module Preservation across different data
Rationale:
- Same data: to check robustness of module definition
- Different data: to find interesting modules.

Find the key drivers in *interesting* modules
Tools: intramodular connectivity, causality testing
Rationale: experimental validation, therapeutics, biomarkers

We use the same clustering procedure

- Since we want to compare co-expression measures as opposed to different clustering procedures, we use the same clustering procedure for either measure.

- Average linkage hierarchical clustering with dynamic branch cutting
  - Default parameter settings
  - it is a widely used benchmark method in coexpression network analysis.

Defining modules based on a hierarchical cluster tree

*Langfelder P, Zhang B et al (2007) Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut library for R. Bioinformatics 2008 **24**(5):719-720*
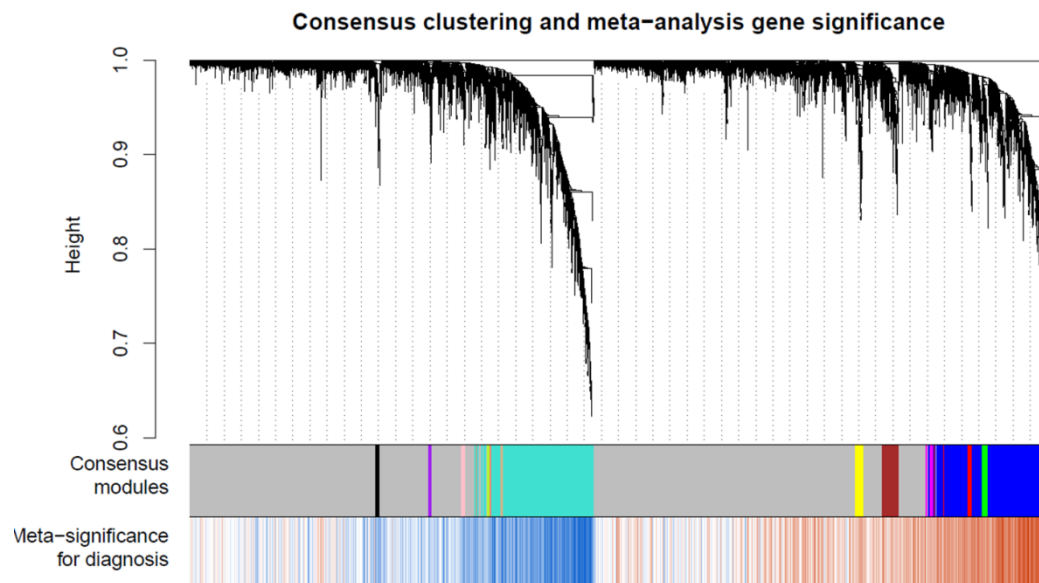
Module=branch of a cluster tree

Dynamic hybrid branch cutting method combines advantages of hierarchical clustering and pam clustering

Cor and MI to define network modules: gene ontology enrichment analysis

- The same 8 empirical data sets are used.
- Correlation or mutual information based adjacency is the only element that can vary in network construction: the same clustering algorithm and the same branch cutting method.
- Enrichment is tested with respect to all GO terms and 5 most significant p-values are retained for each module.
- R function GOenrichmentAnalysis in the WGCNA R package
- P-values from all modules are averaged (details in Song et al 2012)

Modules

5 best p   5 best p   5 best p   5 best p
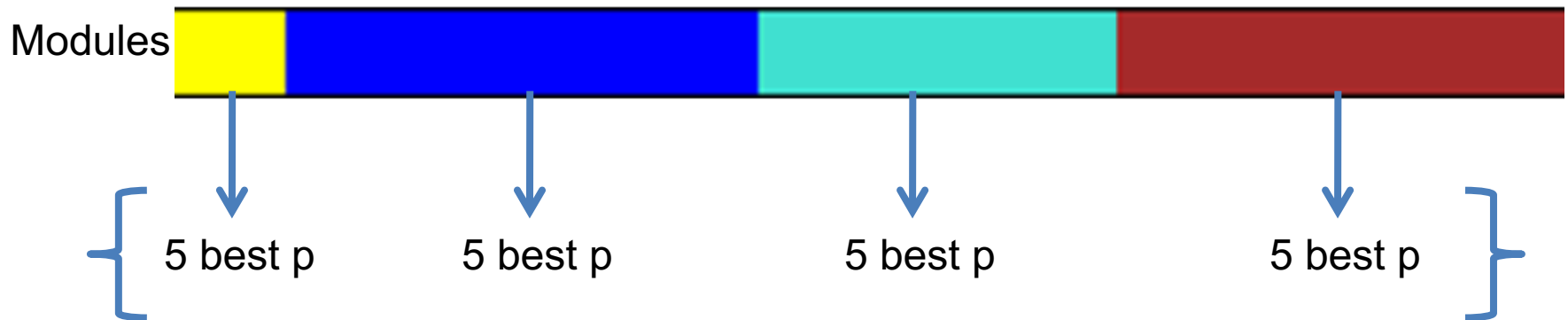
- First, compare AUV2 with 3 bicor based ajacencies: unsigned adjacency, signed adjacency and signed adjacency coupled with Topological Overlap Matrix (TOM).
  - Message: Signed weighted network adjacency coupled with TOM leads to the most significant GO enrichment p-values in all 8 data sets.
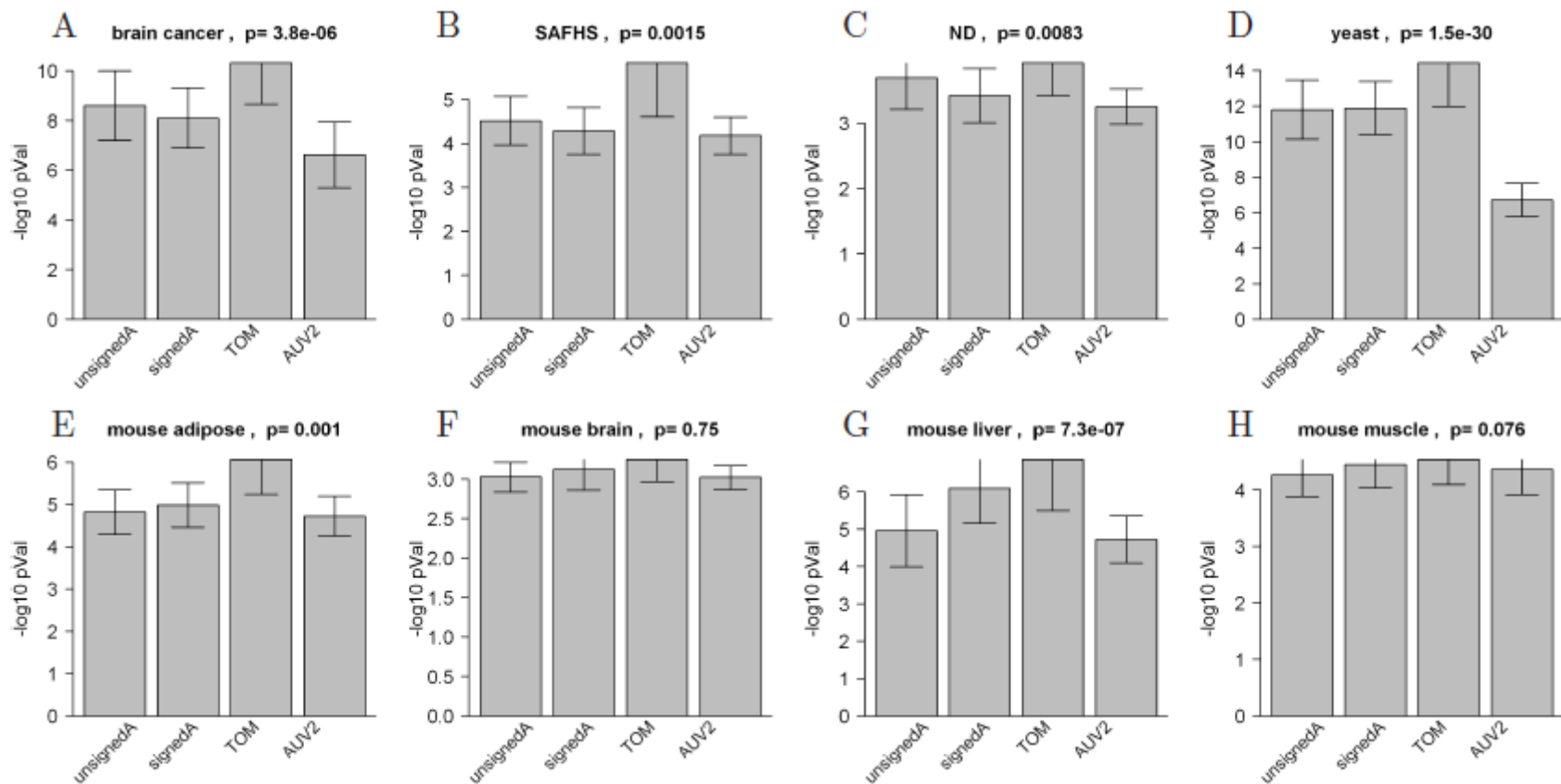
- Second, compare TOM with 6 alternative MI based methods: ARACNE with different tolerance threshold values, CLR, MRNET and RELNET.
  - TOM gives the best p-values in 5 out of 8 data sets. ARACNE is the second best method.
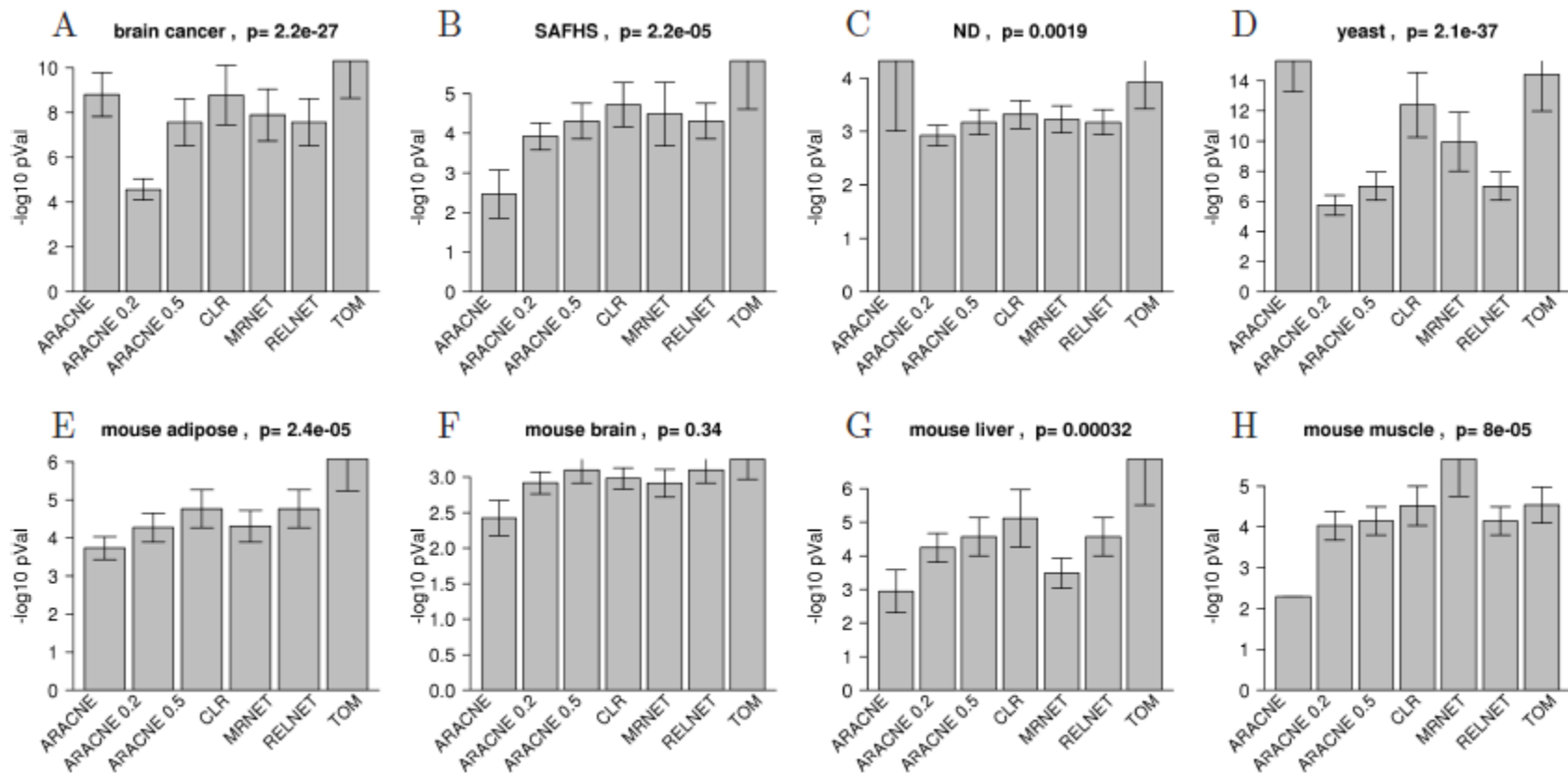
- Third, compare TOM with MI based Maximal Information Coefficient (MIC), the maximal mutual information over many possible grids.
  - TOM outperforms MIC in 6 out of 7 data sets.
  - 8th data set (SAFHS data) is not included due to computation issues of MIC.
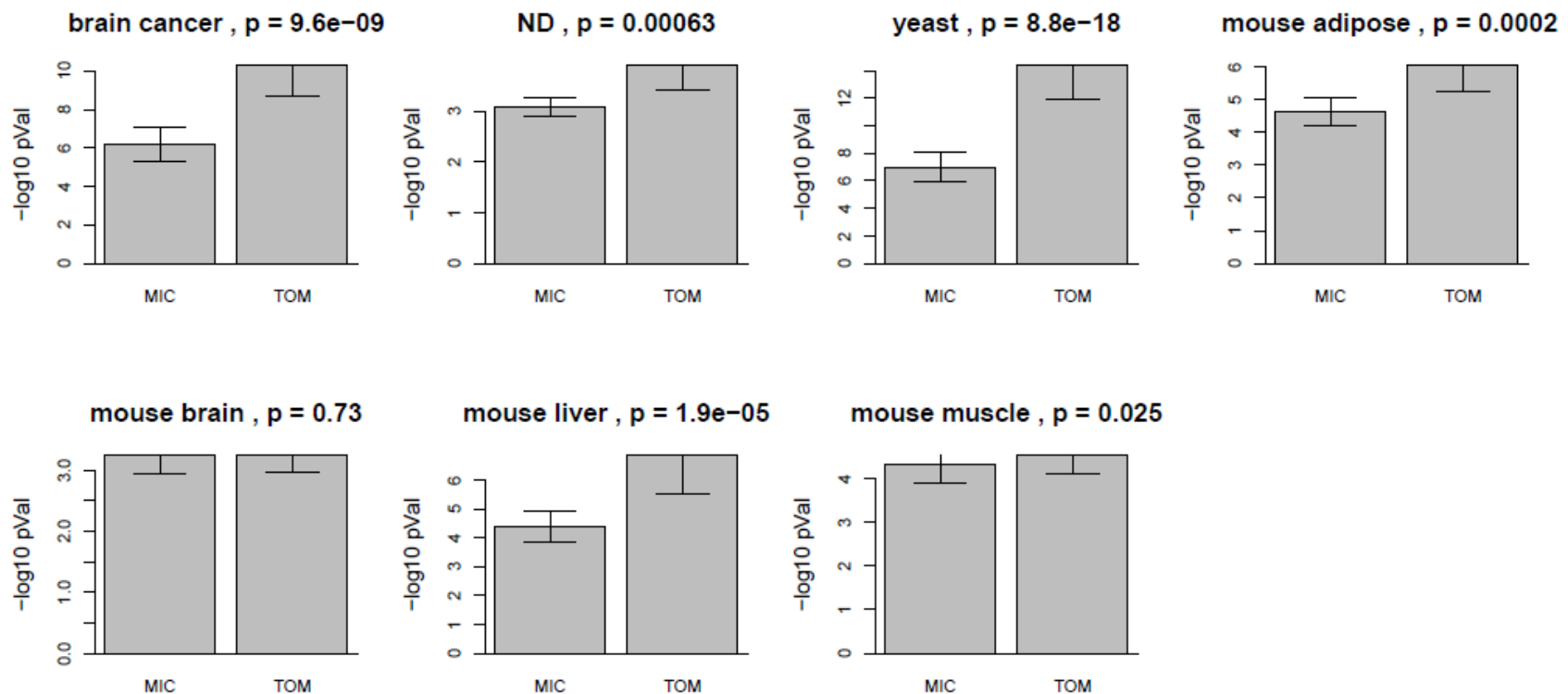
Reshef DN, et al, Sabeti PC: **Detecting Novel Associations in Large Data Sets.** *Science* 2011, **334**(6062):1518–1524.

Polynomial and spline regression models as alternatives to mutual information

Regression models as alternative to MI to estimate non-linear relationships

- Polynomial regression

$$E(y) = \gamma_0 1 + \gamma_1 x + \gamma_2 x^2 \ldots + \gamma_d x^d$$

   - Model fitting index $R^2(x,y)$ can evaluate the relationship between x and y.
   - Reverse the roles of x and y, we get $R^2(y,x)$.
   - Adjacency $= \frac{R^2(x,y)+R^2(y,x)}{2}$
   - Construct networks.

- Local polynomial/spline regression: fit polynomial in sub intervals

- Regression models could detect non-linear relations in both simulation and empirical analyses.

Compare regression models to MI

Pros
- Computational simple and fast
- Standard statistical tests available (F test, chi-square tests, Wald tests etc).
- Model fitting indices are available
- Easy to include covariates

Cons
- Less general than mutual information when it comes to dependence patterns.

WGCNA R functions for calculating an adjacency matrix

- **`adjacency`**
  - Calculate network adjacency based on cor or distance
  - Example:
    A=adjacency(datExpr,type="signed",power=12,
    corFnc = "bicor")
- **`adjacency.fromSimilarity`**
  - Calculate network adjacency from a similarity matrix
- **`adjacency.polyReg`**
  - Adjacency based on polynomial regression
- **`adjacency.splineReg`**
  - Adjacency based on natural cubic spline regression

WGCNA R functions for mutual information

- **mutualInfoAdjacency**
  - Calculate weighted adjacency matrices based on mutual information
  - Ex: A=mutualInfoAdjacency(datExpr, discretizeColumns = TRUE, entropyEstimationMethod = "MM", numberBins = NULL)
- **AFcorMI**
  - Prediction of weighted mutual information adjacency matrix by correlation

Vast literature on mutual info.
Some citations for mutual information

- Horvath S (2011) Weighted Network Analysis. Springer Book, Chapter 14
- Hausser J, Strimmer K (2008) Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. See http://arxiv.org/abs/0811.3579
- Patrick E. Meyer, Frederic Lafitte, and Gianluca Bontempi. minet: A R/Bioconductor Package for Inferring Large Transcriptional Networks Using Mutual Information. BMC Bioinformatics, Vol 9, 2008
- Kraskov A, Stoegbauer H, Andrzejak RG, Grassberger P (2003) Hierarchical Clustering Based on Mutual Information. ArXiv q-bio/0311039

WGCNA R functions for bicor

- **AFcorMI**
  - Predict the weighted mutual information adjacency matrix based on a correlation coefficient (e.g. Pearson or bicor). Details are in book section 14.6 (Weighted Network Analysis)
- **bicor**
  - Biweight Midcorrelation
- **bicorAndPvalue**
  - Biweight Midcorrelation and the associated p-value

Conclusions

- Biweight midcorrelation outperforms mutual information when it comes to

  1) finding biologically meaningful co-expression patterns between pairs of genes

  2) identifying co-expression modules that are enriched with GO categories

- Polynomial and spline regression models are attractive alternatives to mutual information when it comes to detecting non-linear relationships (Song et al 2012)
  - Model based indices enjoy statistical advantages.

- Mutual information can safely be replaced by linear regression based association measures in case of stationary gene expression data.

Content

Comparing Statistical Methods for Constructing Large Scale Gene Networks (2012)

Comparing Statistical Methods for Constructing Large Scale Gene Networks (2012).

Allen JD, Xie Y, Chen M, Girard L, Xiao G

Comments:

- Our group had no involvement in this comparison.

- The study focused on methods for gene regulatory networks.

Abstract

- The goal of this study was to provide a comprehensive evaluation and a practical guide to aid in choosing statistical methods for constructing large scale Gene Regulatory Networks (GRNs)

- Using both simulation studies and a real application in E. coli data, the authors compare different methods

- **Results:**

- (1) GeneNet, WGCNA (Weighted Correlation Network Analysis), and ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks) performed well in constructing the global network structure;

- (2) GeneNet and SPACE (Sparse PArtial Correlation Estimation) performed well in identifying a few connections with high specificity.

Methods

- ## GeneNet (K. Strimmer)
  - Opgen-Rhein, R., and K. Strimmer. 2007. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. BMC Syst. Biol. 1: 37.
- ## ARACNE
  - Margolin, A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R. and Califano, A. (2006). "ARACNE: an Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context". BMC Bioinformatics 7 (Suppl 1): S7.
- ## WGCNA (Zhang et al 2005, Langfelder et al 2008)

The AUCs for 1344-gene network in simulation 1.

WGCNA has consistently good performance across different sample sizes
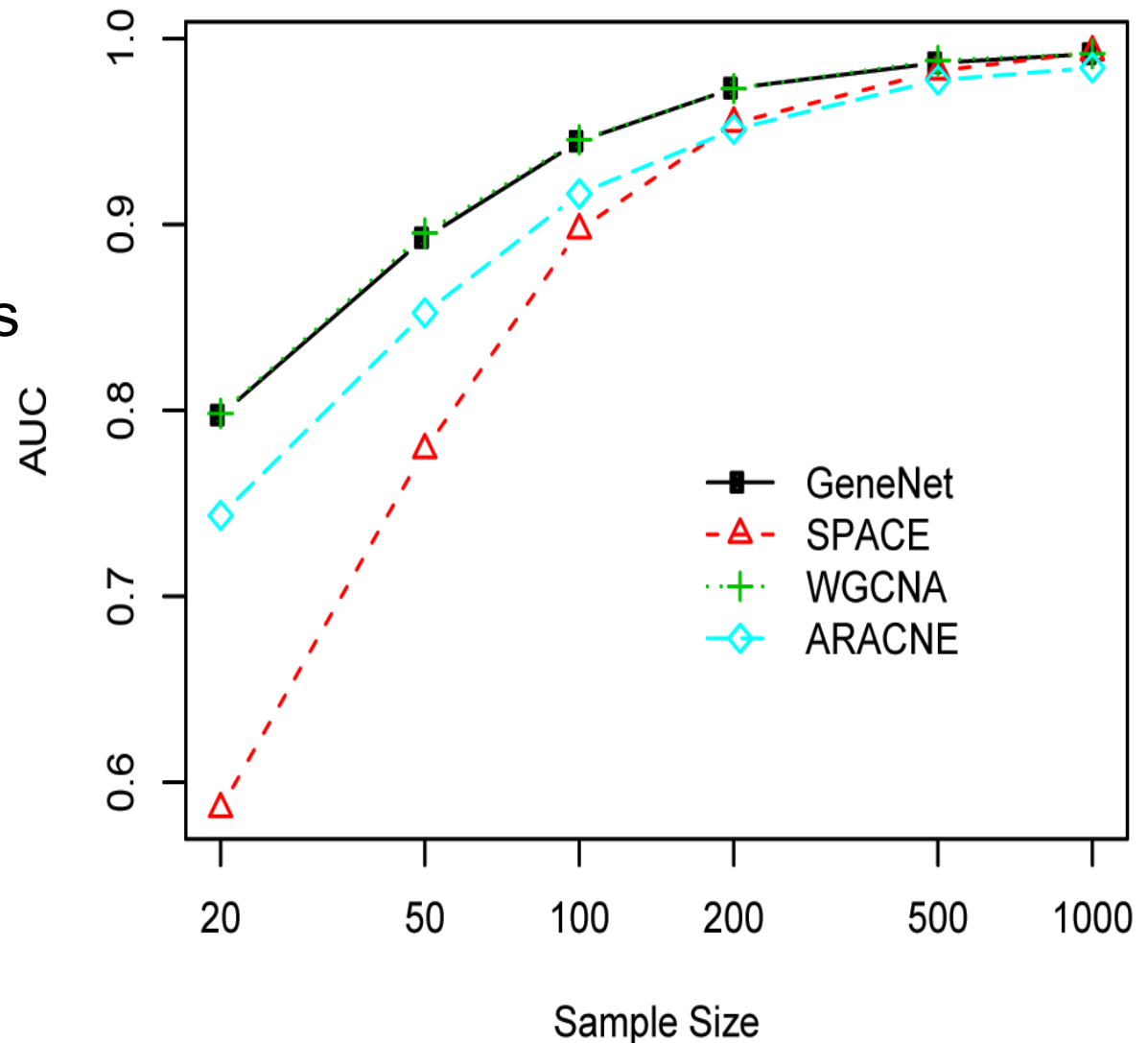
1344 Gene Network AUC

The AUCs for 1344-gene network in simulation 2 with non-normal distribution (Figure 4).

- WGCNA consistently performs well in case of non-normal distributions

Non-Linear 1344 Gene Network AUCs
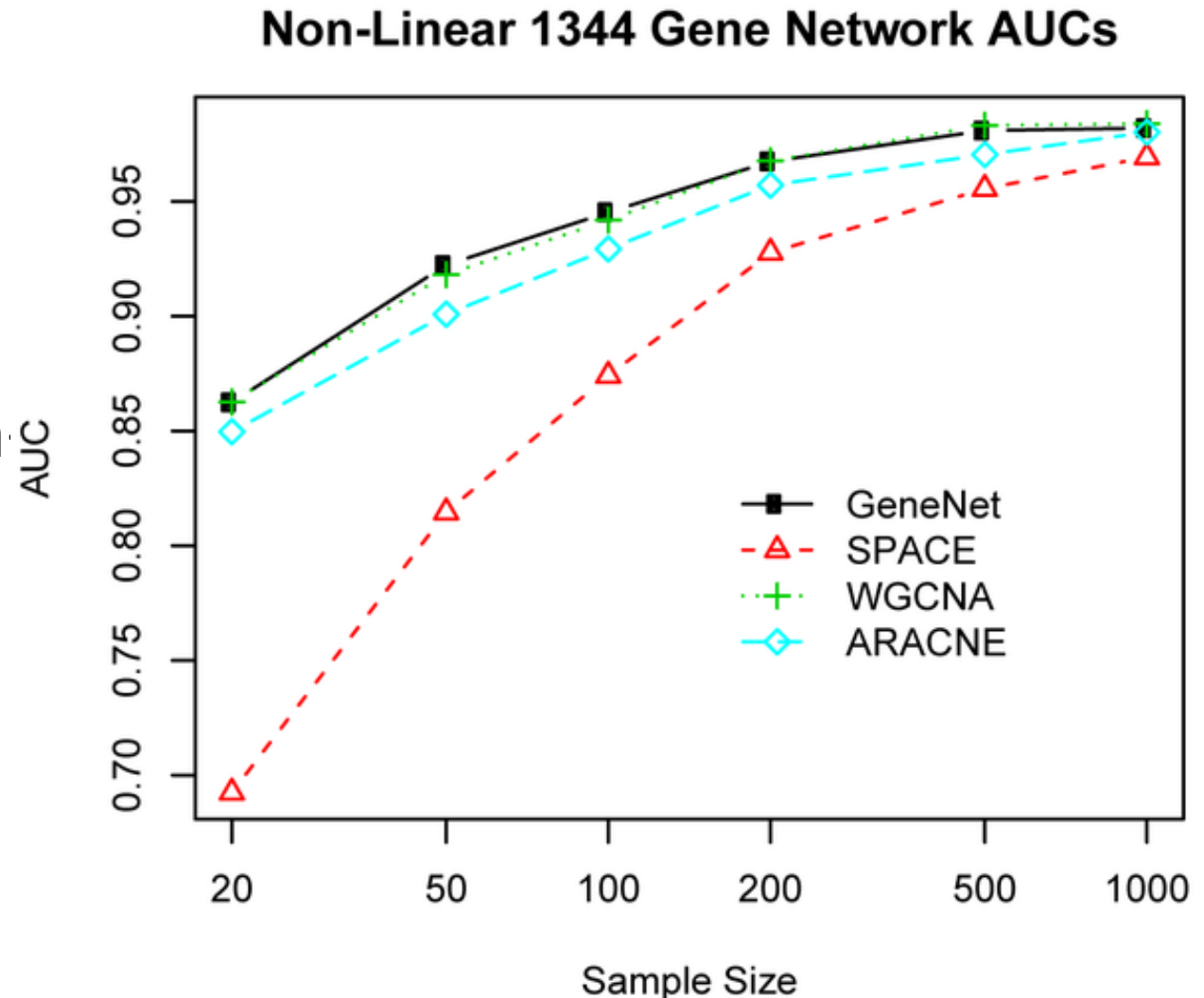
The performance in constructing gene regulatory network in E. coli. Figure 6.

WGCNA performs consistently well across different choices of the specificity
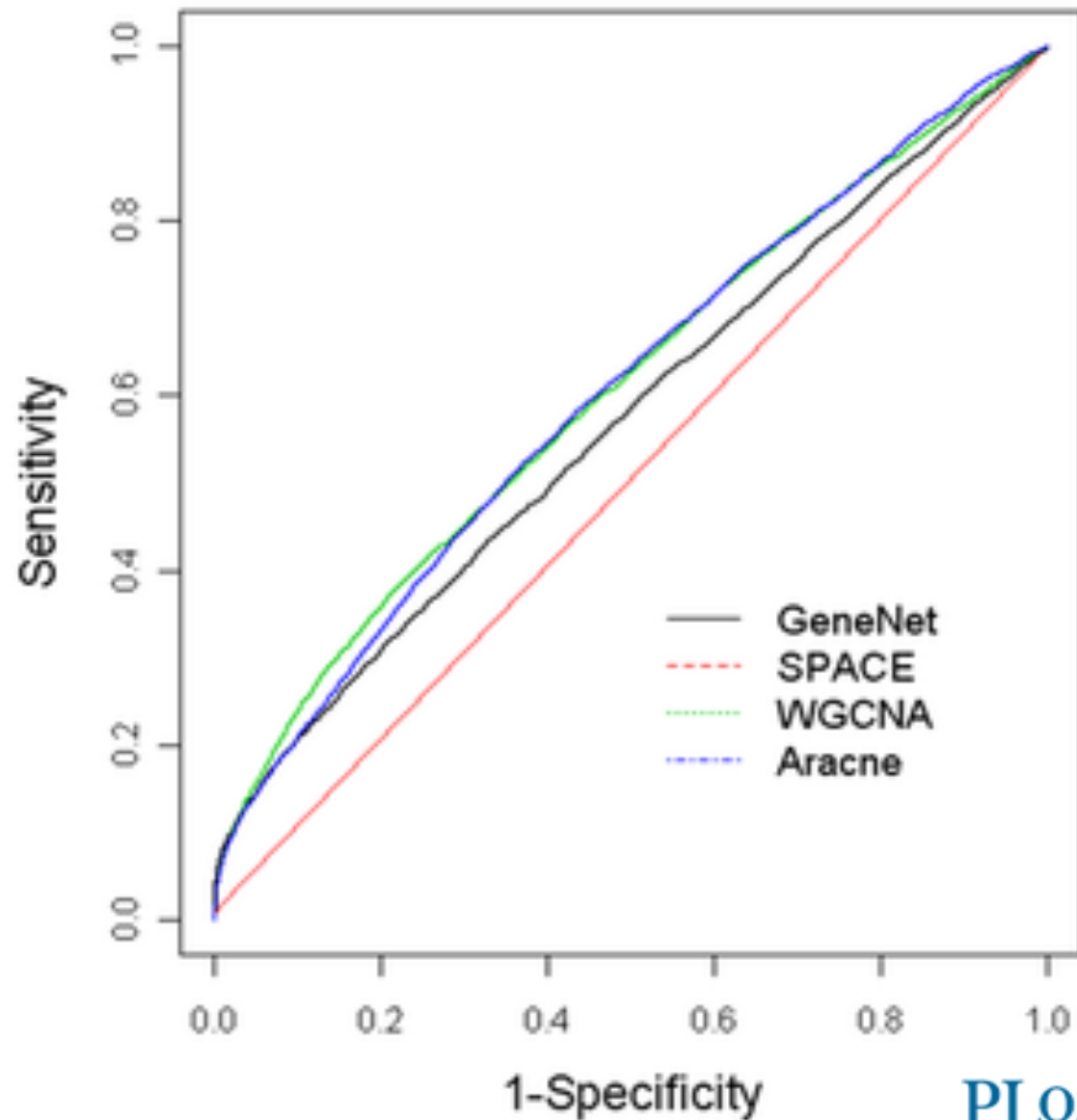
PLoS one

Content

Aging effects on DNA methylation modules in human brain and blood tissue. Genome Biol. 2012 Oct 3;13(10):R97. PMID: 23034122

When Is Hub Gene Selection Better than Standard Meta-Analysis?
Langfelder et al (2013) PLoS ONE 8(4): e61505.

Aging effects on DNA methylation modules in human brain and blood tissue.

Steve Horvath

Senior author: Roel A Ophoff UCLA

Photo courtesy of Luna Productions www.lunaproductions.com

Weighted correlation network analysis (WGCNA) applied to multiple methylation data

Goal: find age related consensus methylation modules in whole blood and brain tissue

Analysis steps of WGCNA

1. Construct a signed weighted correlation network based on multiple, independent data sets
Purpose: keep track of relationships between genes

2. Identify consensus modules
Purpose: find robustly defined and reproducible modules

3. Relate modules to external information
Age
Gene Information: gene ontology, cell marker genes
Purpose: find biologically interesting age related modules

Additional public methylation data used for consensus module analysis

Inf 27k Illumina Array

1) Rakyan et al GSE20236

- 93 different healthy females (31 twin pairs and 31 singletons)
- ranging from 49 to 75 years of age
- Whole blood

2) Healthy individuals from UK ovarian cancer data

- Teschendorff et al 2010, Song et al 2009 GSE19711
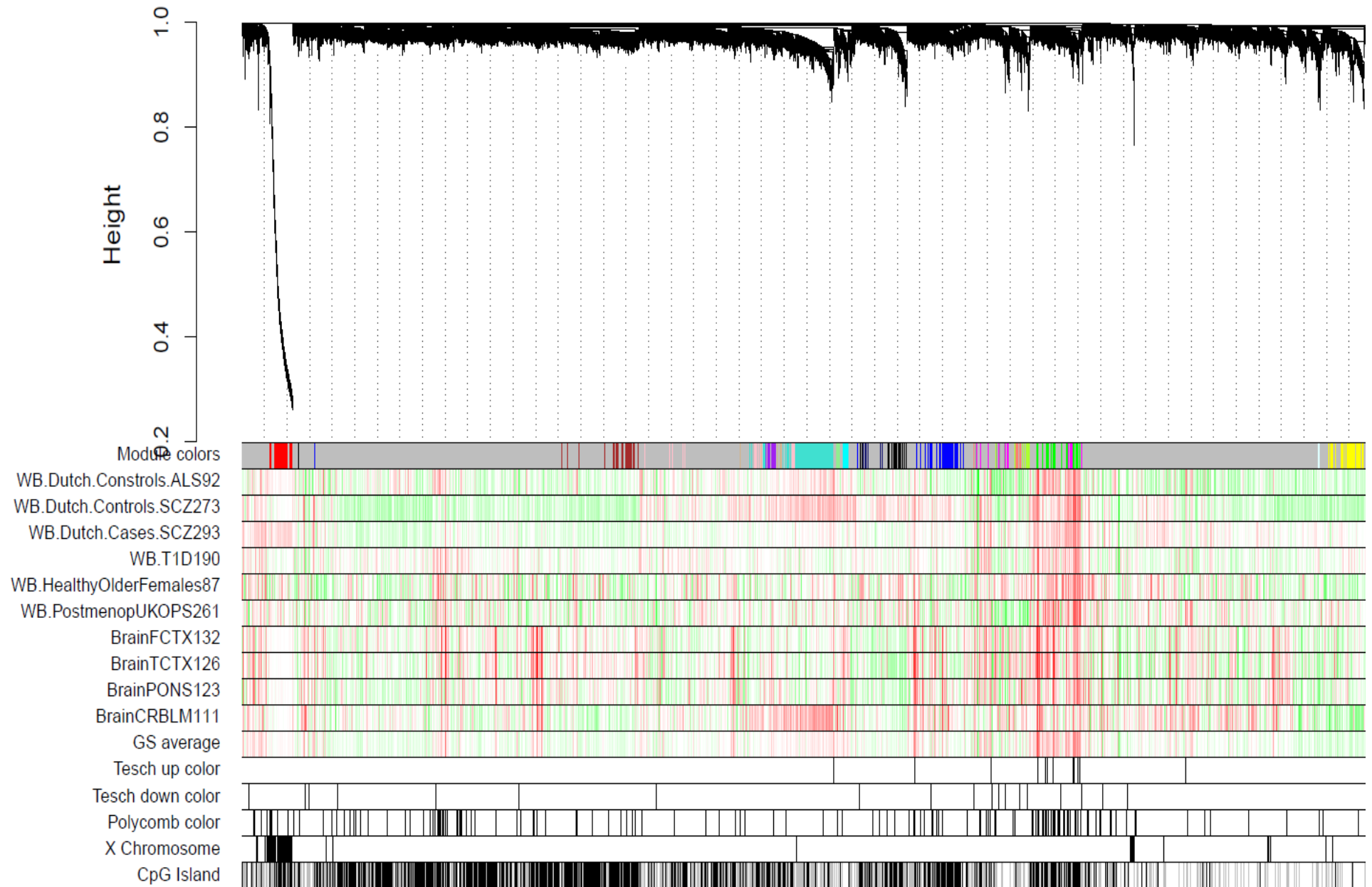- 267 healthy women

3) Type 1 Diabetics (Irish)

- 93 males, 95 females

4) Brain data: about 150 individuals, 4 brain regions

R code

- blockwiseConsensusModules in the WGCNA R package, see the tutorials...

Consensus gene dendrogram and module colors for the 10 methylation dataset recut

Message: green module contains probes positively correlated with age

Consensus module analysis (Figure)

- based on multiple independent DNA methylation data sets.
  - Technical stuff: Average linkage hierarchical clustering tree uses the consensus topological overlap matrix based on calibrated versions of the signed correlation networks in each of the data sets.

- By definition co-methylation modules are present in all of the underlying data sets.

- The first color band underneath the tree visualizes the consensus module colors.

- The other color bands indicate age correlations in multiple data sets
  - red visualizes a high positive correlation

- **The green module contains CpGs that tend to be positively correlated with age.**

Module–age relationships in 10 datasets
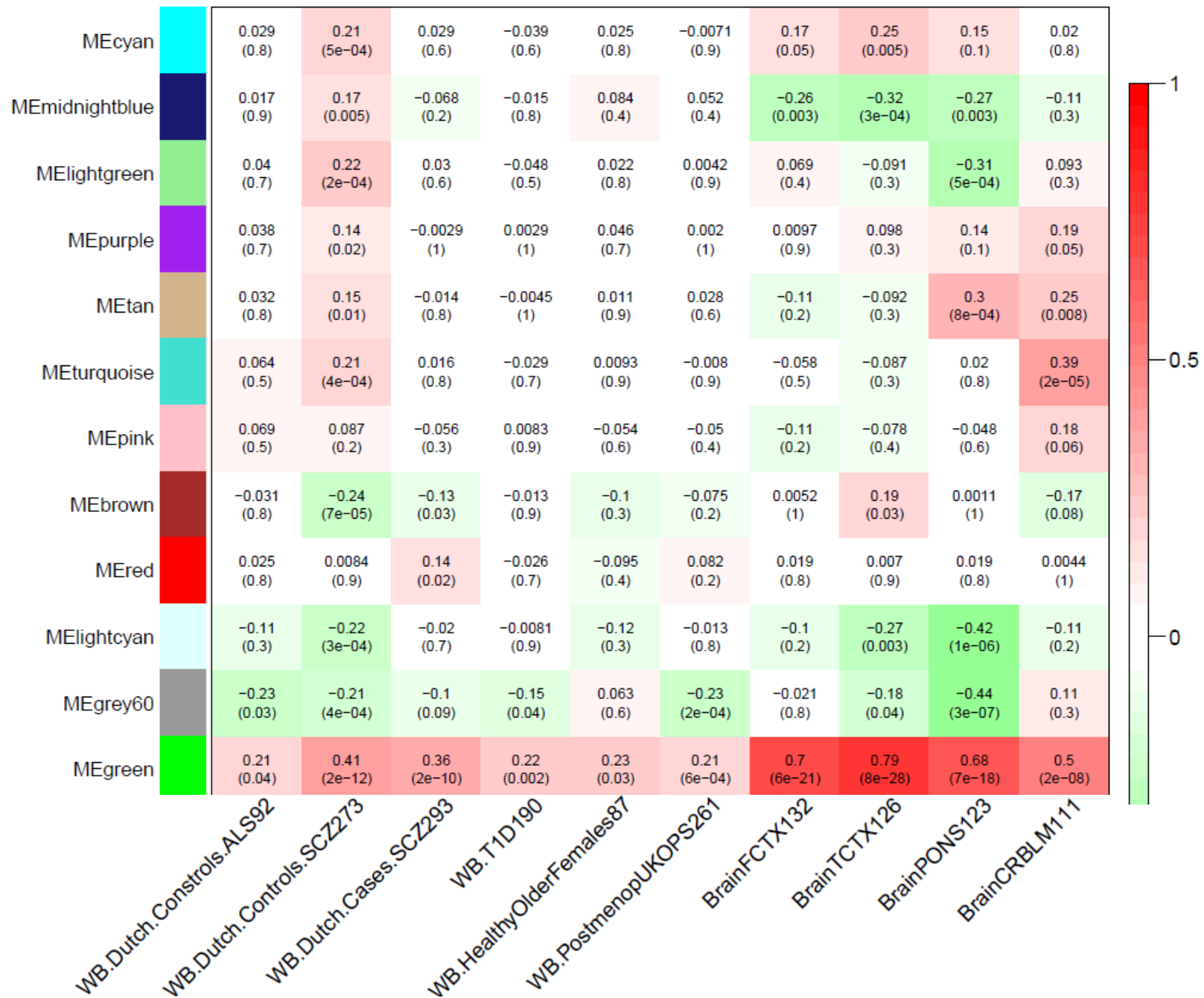
Age relations in brain regions

The green module eigengene is
highly correlated with age in
i) Frontal cortex (cor=.70)
ii) Temporal cortex (cor=.79)
iii) Pons (cor=.68)

But less so in cerebellum (cor=.50).

R code for finding hub genes
`consensusKME` function

consensusKME {WGCNA}   R Documentation

Calculate consensus kME (eigengene-based connectivities) across multiple data sets.

Description

Calculate consensus kME (eigengene-based connectivities) across multiple data sets, typically following a consensus module analysis.

Usage

```
consensusKME(multiExpr, moduleLabels,
          multiEigengenes = NULL,
          consensusQuantile = 0,
          signed = TRUE,
          useModules = NULL,
          metaAnalysisWeights = NULL,
          corAndPvalueFnc = corAndPvalue, corOptions = list(), corComponent = "cor",
          getQvalues = FALSE,
          useRankPvalue = TRUE,
          rankPvalueOptions = list(calculateQvalue = getQvalues, pValueMethod = "scale"),
          setNames = NULL,
          excludeGrey = TRUE, greyLabel = ifelse(is.numeric(moduleLabels), 0, "grey"))
```

Output consensusKME

`Value`

Data frame with the following components (for easier readability the order here is not the same as in the actual output):

`ID`	Gene ID, taken from the column names of the first input data set
`consensus.kME.1, consensus.kME.2, ...`	Consensus kME (that is, the requested quantile of the kMEs in the individual data sets)in each module for each gene across the input data sets. The module labels (here 1, 2, etc.) correspond to those in `moduleLabels`.
`weightedAverage.equalWeights.kME1, weightedAverage.equalWeights.kME2, ...`	Average kME in each module for each gene across the input data sets.
`weightedAverage.RootDoFWeights.kME1, weightedAverage.RootDoFWeights.kME2, ...`	Weighted average kME in each module for each gene across the input data sets. The weight of each data set is proportional to the square root of the number of samples in the set.

Gene ontology enrichment analysis of the green module

- Highly significant enrichment in multiple terms related to cell differentiation, development and brain function
  - e.g. nervous system development (Bonferroni corrected p-value =2.1e-6),
  - neuron differentiation (p=7.6e-5),
  - anatomical structure development (p=0.00013),
  - cell development (p=0.00024),
  - generation of neurons (p=0.00038),
  - neurogenesis (p=0.00052), cell differentiation (p=0.00057).

Green module is enriched with neuronal cell type markers

- Green module genes are enriched for genes that are
  - upregulated in neurons (based on Cahoy et al 2009) (Bonferroni corrected p-values p=5e-9)
  - Related to CA1 area related neurons (based on Newrzella et al 2007) (p=3e-7),
  - downregulated in hippocampus in early Alzheimer's disease (Parachikova et al 2007) (p=3e-6)
- Recall that this module can also be found in blood tissue.
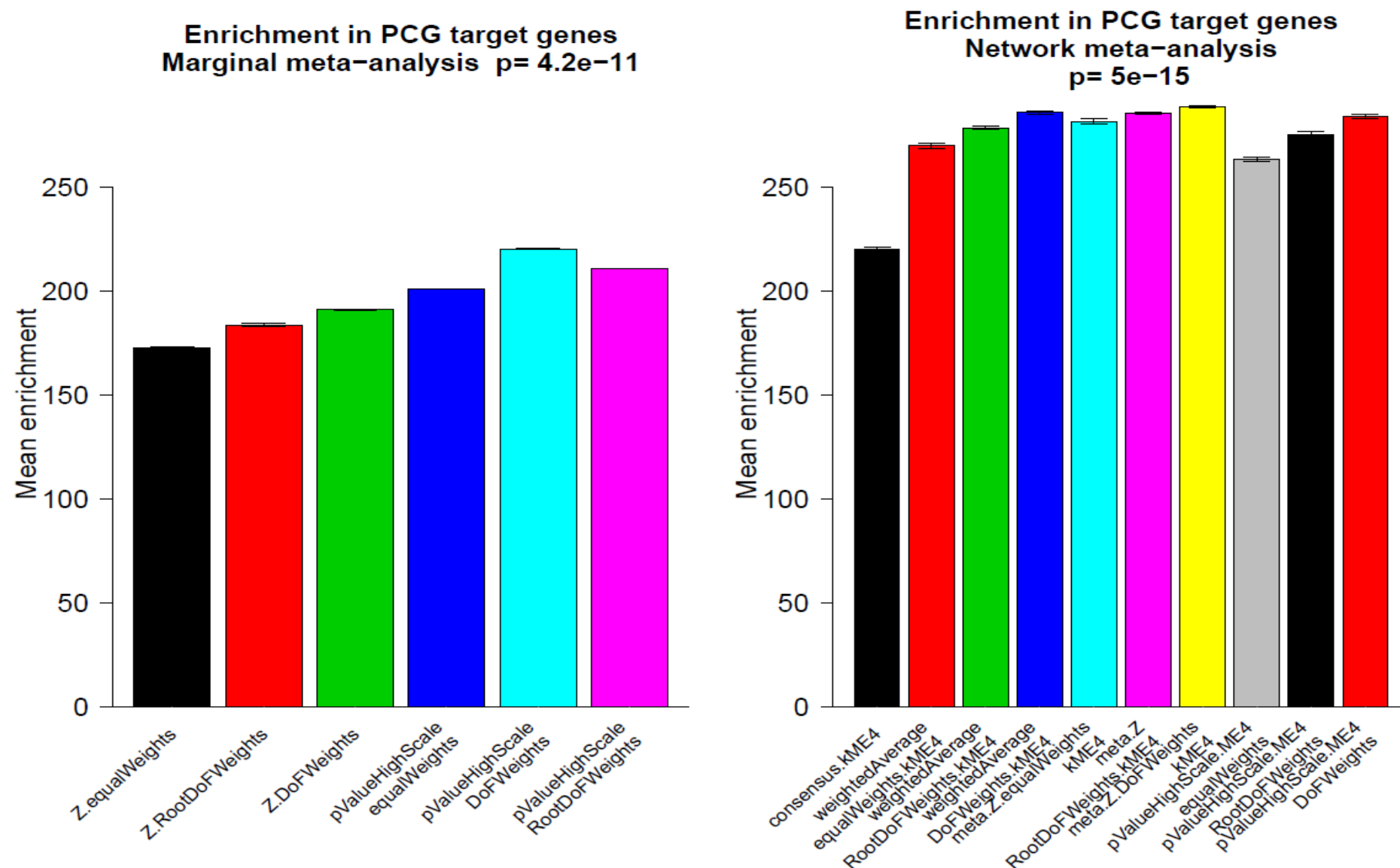
Polycomb-group proteins

Polycomb Group gene expression is important in many aspects of development.

Genes that are hypermethylated with age are significantly enriched with Polycomb group target genes (Teschendorff et al 2010)

This insight allows us to compare different gene selection strategies.

The higher the enrichment with respect to PCGT genes the more signal is in the data.

Comparison of standard screening
vs. consensusKME screening with respect to enrichment for PCG target gene

Enrichment in PCG target genes
Marginal meta−analysis p= 4.2e−11

Enrichment in PCG target genes
Network meta−analysis
p= 5e−15

Relating kME.green and cor.Age
and sequence properties
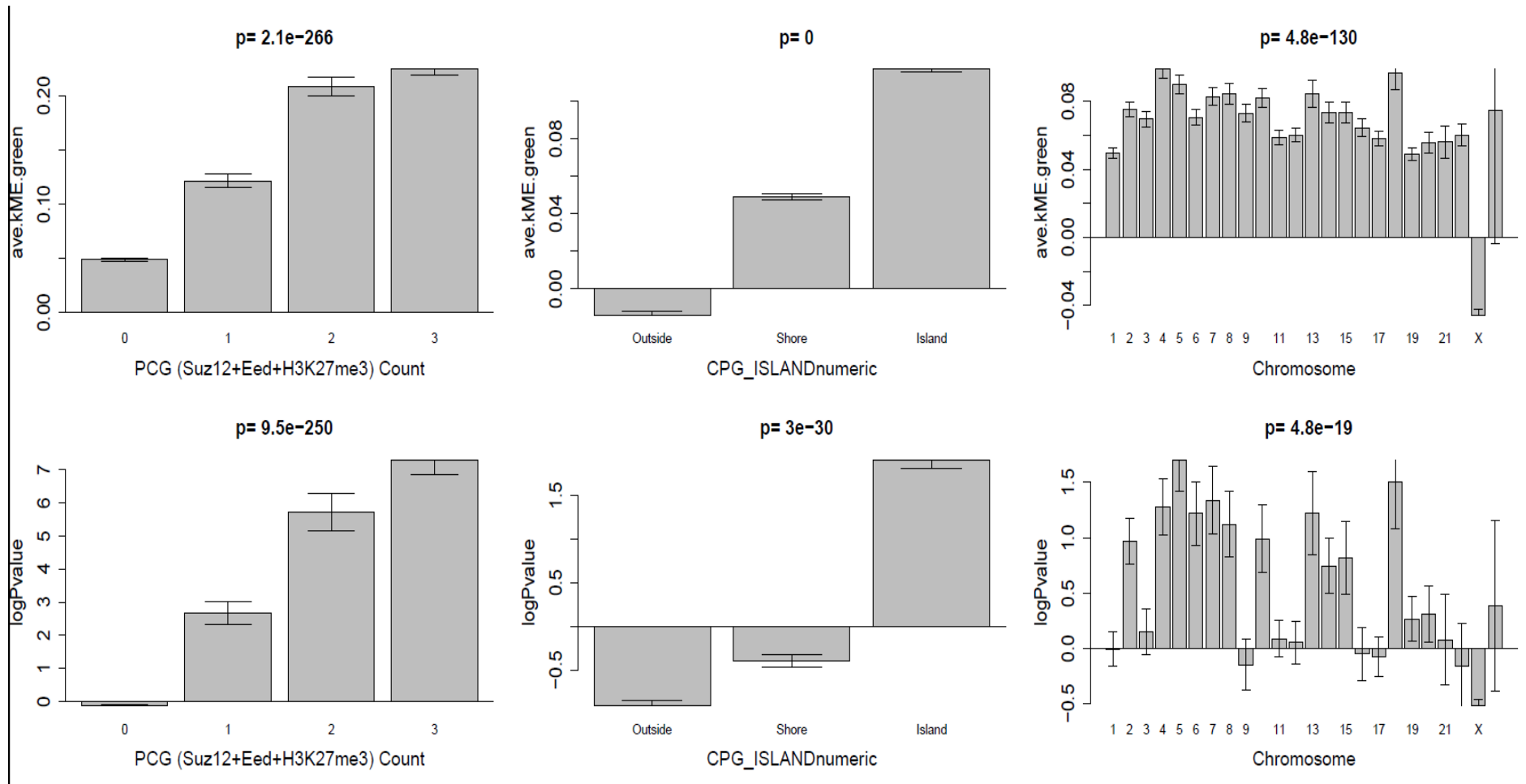
Discussion

- We confirm the findings of many others
  - age has a profound effects on thousands of methylation probes
  - Rough rule: positive/negative correlation for probes inside/outside of CpG islands
- WGCNA reveals the presence of a methylation module that positively correlates with age
  - Enriched with neuronal markers
  - Contains cis regulated genes
- Age related module genes are interesting for studying the age related decline of cognitive functioning

Content

When Is Hub Gene Selection Better than Standard Meta-Analysis?
Langfelder et al (2013) PLoS ONE 8(4): e61505.

When does hub gene selection lead to more meaningful gene lists than a standard statistical analysis based on significance testing?

- Here we address this question for the special case when multiple data sets are available.

- This is of great practical importance since for many research questions multiple gene expression or other -omics data sets are publicly available.

- In this case, the data analyst can decide between a standard statistical approach (e.g., based on meta-analysis) and a co-expression network analysis approach that selects intramodular hubs in consensus modules.

Intramodular hub genes versus whole network hubs

- <u>Intramodular hubs</u> have high intramodular connectivity kME with respect to a given module of interest
- <u>Whole network hubs</u> have high values of whole network connectivity k
  - k= row sum of the adjacency matrix
  - k= number of direct neighbors in case of an unweighted network

Q & A

- **1. Are whole-network hub genes relevant or should one exclusively focus on intramodular hubs?**

- Answer: Focus exclusively on intramodular hubs in trait-related modules.

- **2. Do network-based gene selection strategies lead to gene lists that are biologically more informative than those based on a standard marginal approaches?**

- Answer: Yes, gene selection based on intramodular connectivity leads to biologically more informative gene lists than marginal approaches.

- **3. Do network-based gene selection strategies lead to gene lists that have more reproducible trait associations than those based on a standard marginal approaches?** Answer: Overall no. But in case of a weak signal networks can help.

Criteria for judging gene selection methods

- <u>Criterion 1</u> evaluates the biological insights gained, i.e. it is relevant in basic research.

- <u>Criterion 2</u> evaluates the validation success in independent data sets, i.e. it is relevant when it comes to developing diagnostic or prognostic biomarkers.

Data sets used in the empirical evaluation

- We compare standard meta-analysis with consensus network analysis in three comprehensive and unbiased empirical studies:
- (1) Find genes predictive of lung cancer survival
  - Gold standard=cell proliferation related genes
- (2) Find age related DNA methylation markers
  - Gold standard= Polycomb group target genes
- (3) Find genes related to total cholesterol in mouse liver tissues
 - Gold standard= immune system related genes

R code in the WGCNA package

- For standard screening, we used the <u>metaAnalysis</u> function

- For finding hubs in consensus modules, we used the <u>consensusKME</u> function
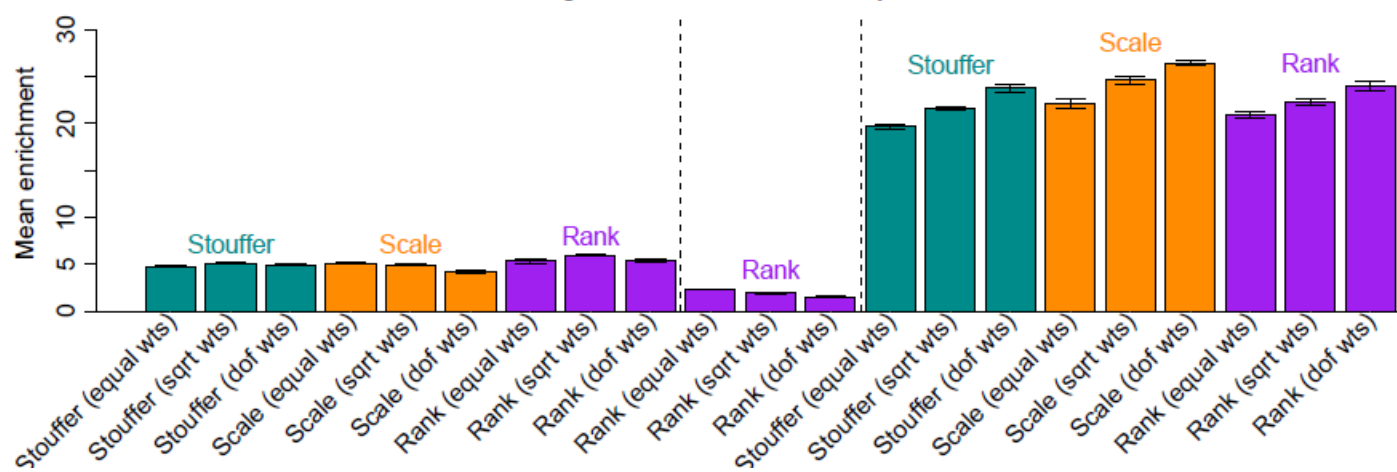
A. Enrichment in "cell cycle", adenocarcinoma expression application
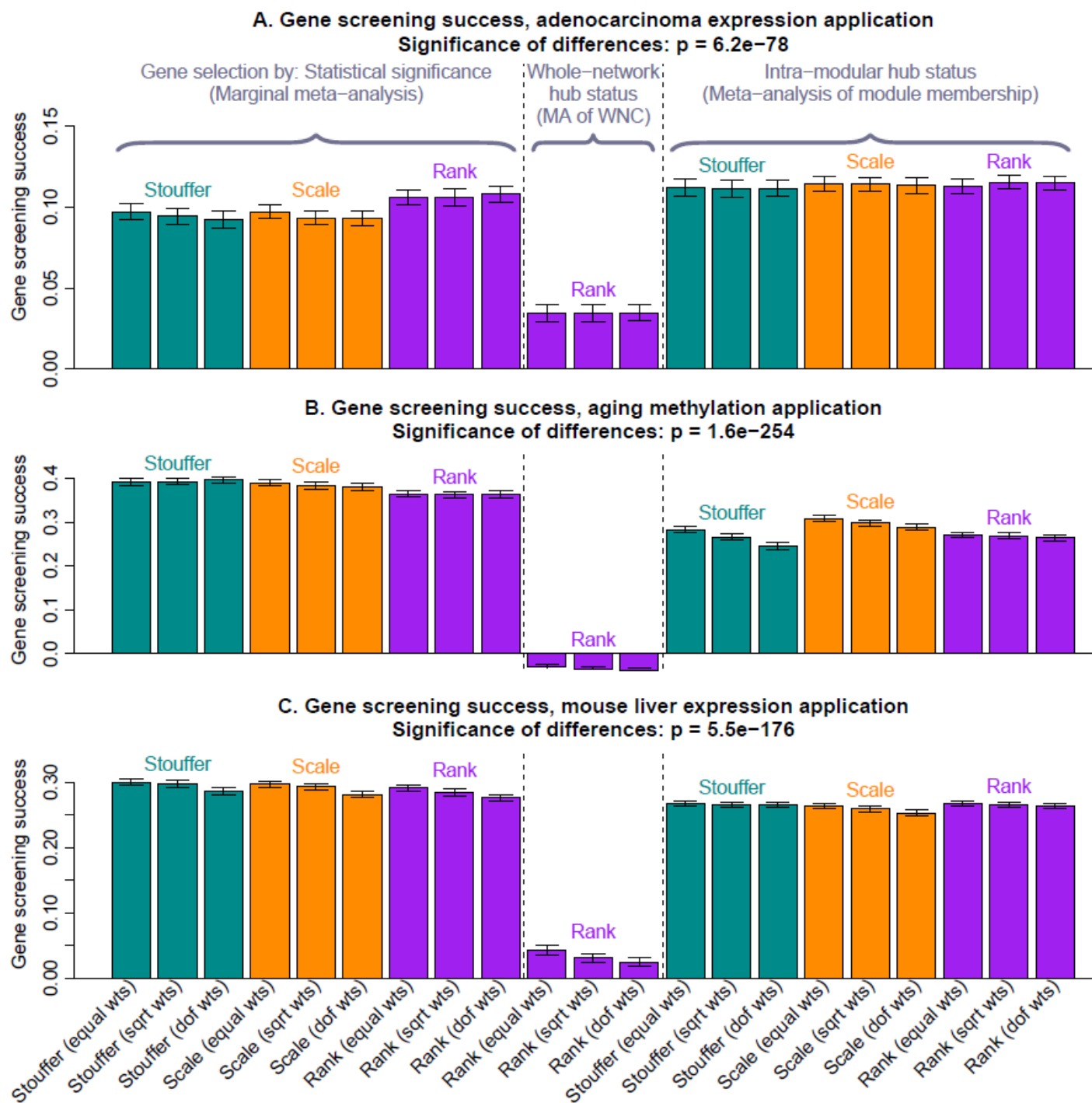Significance of differences: p= 1.5e−31
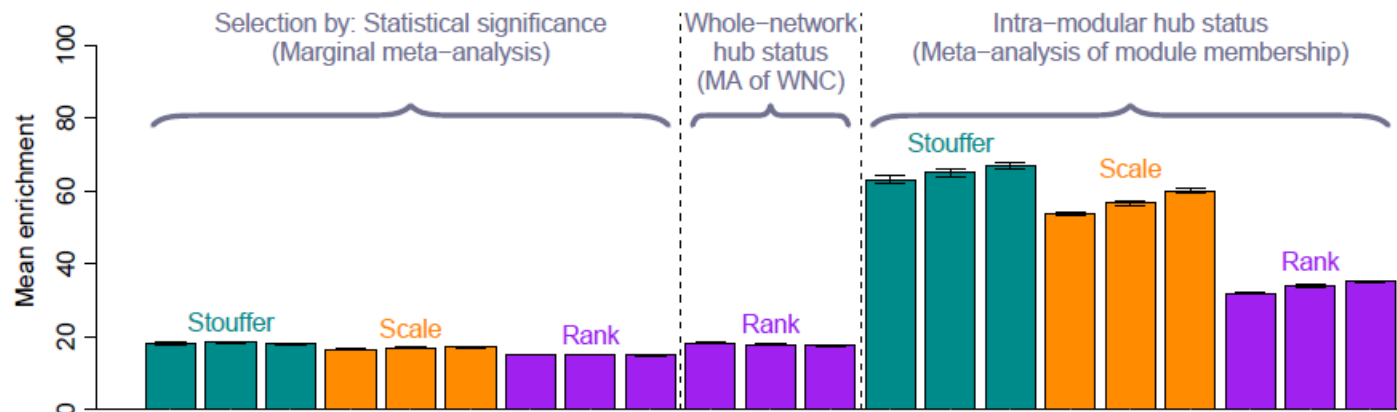
B. Enrichment in PCG targets, aging methylation application
Significance of differences: p= 3.7e−33
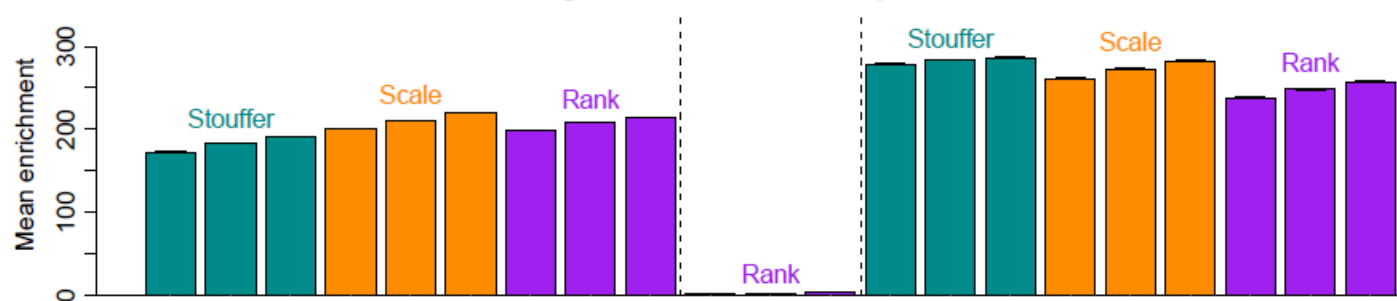
C. Enrichment in "immune system process", mouse liver expression application
Significance of differences: p= 6.5e−31

A. Gene screening success, adenocarcinoma expression application
Significance of differences: p = 6.2e−78

Gene selection by: Statistical significance (Marginal meta−analysis) | Whole−network hub status (MA of WNC) | Intra−modular hub status (Meta−analysis of module membership)

B. Gene screening success, aging methylation application
Significance of differences: p = 1.6e−254

C. Gene screening success, mouse liver expression application
Significance of differences: p = 5.5e−176

Stouffer (equal wts), Stouffer (sqrt wts), Stouffer (dof wts), Scale (equal wts), Scale (sqrt wts), Scale (dof wts), Rank (equal wts), Rank (sqrt wts), Rank (dof wts)

Results

- The results demonstrate that intramodular hub gene status is more useful than a meta-analysis p-value when identifying biologically meaningful gene lists (reflecting criterion 1).

- However, meta-analysis methods perform as good as (if not better) than a co-expression network approach in terms of validation success (criterion 2).