Genome-free mRNA-Seq Data Analysis

BayPass: Outlier testing with SNPs

WGCNA: Gene co-expression network analysis

Genome-free mRNA-Seq Data Analysis

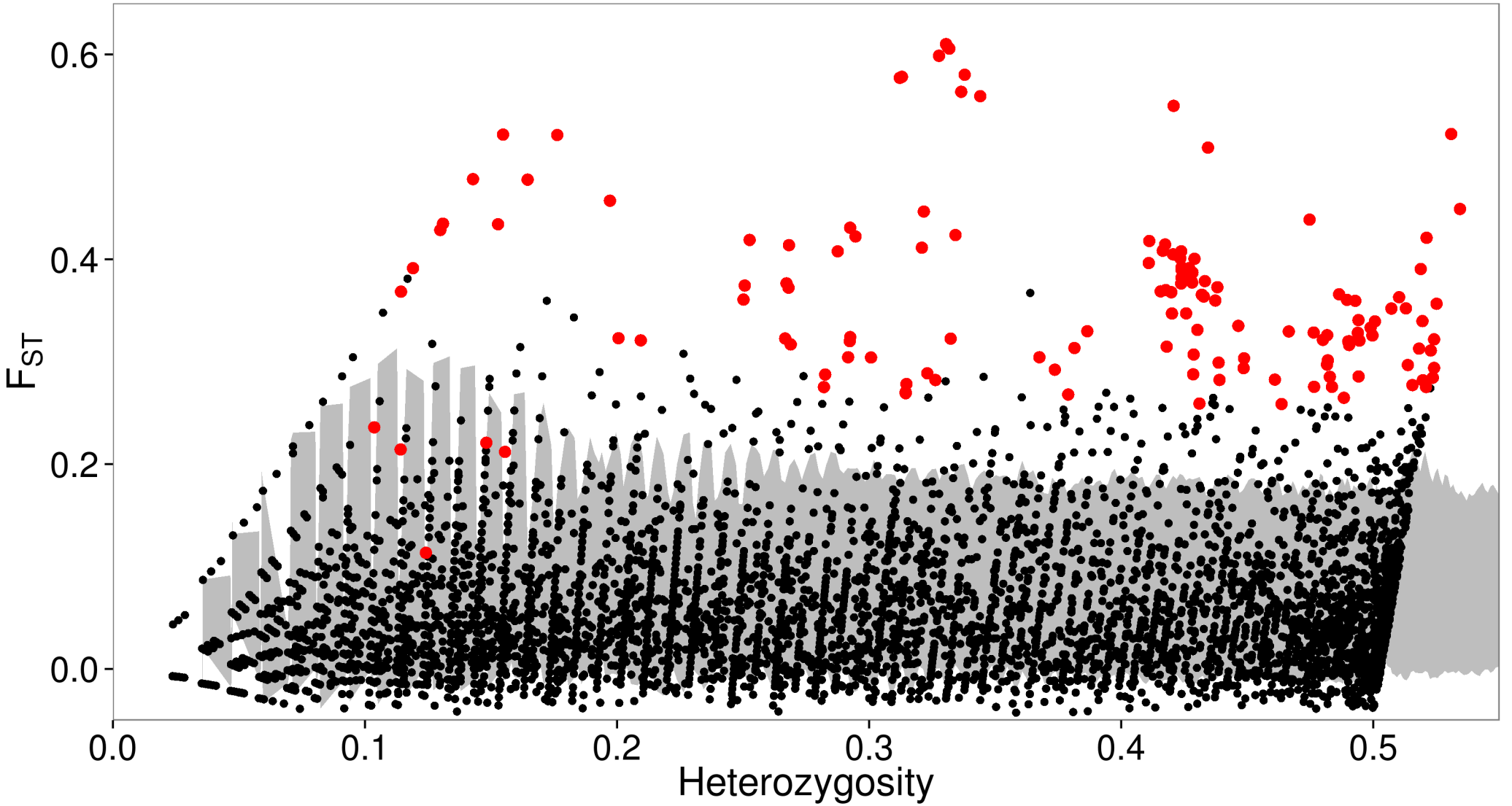
BayPass: Outlier testing with SNPs

One species
Multiple populations
Biallelic SNPs

Which SNPs might be under selection?

F_{ST}-based Outlier Testing:

Test for loci with unusually high F_{ST}



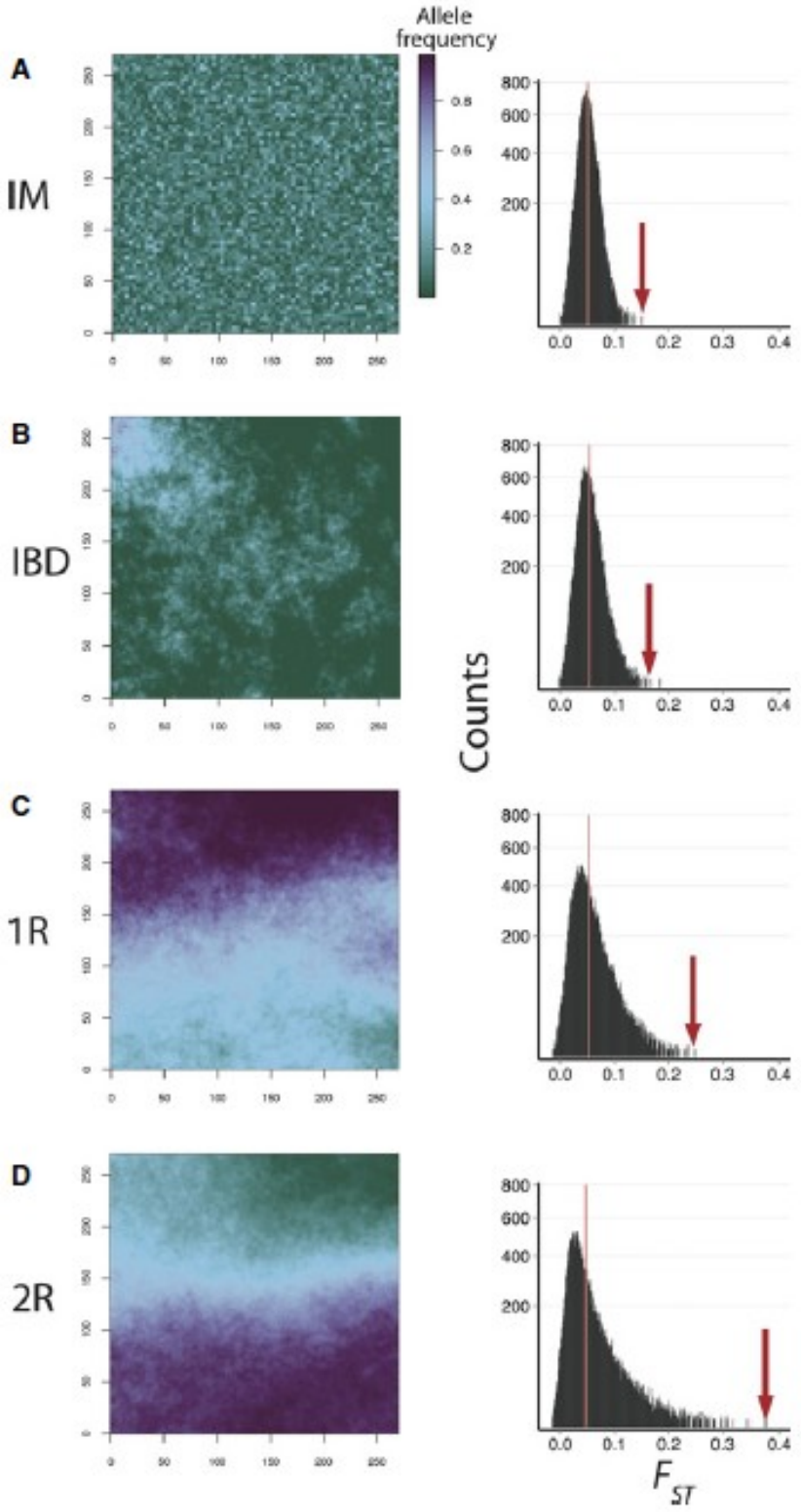
F_{ST}-based Outlier Testing:

Test for loci with unusually high F_{ST}

Assumes that neutral loci have approximately the same F_{ST}

"However, in subdivided populations, by chance the measured F_{ST} can differ substantially from this expectation, causing even neutral genes to vary, sometimes substantially, in their F_{ST}s. F_{ST} outlier tests attempt to account for this neutral variation in F_{ST} and determine which loci have F_{ST} large enough or small enough to show significant evidence of selection. The challenge with outlier tests is to identify how much variation in F_{ST} among loci would be expected (i.e. the null distribution of F_{ST}) in the absence of selection."

Lotterhos & Whitlock 2014



Neutral SNP under different scenarios

Lotterhos & Whitlock 2014

F_{ST}-based Outlier Testing

Older methods: BayeScan, FDIST2

- Assume evolutionary independence
  - Null distribution assumes specific demographic history
  - Independent divergence from a common ancestor
- Perform poorly w/ IDB, expansion, migration, etc.
- May give many false positives

Lotterhos & Whitlock 2014

F_{ST}-based Outlier Testing

Newer methods: BayEnv, BayPass

- Account for population structure
  - Estimate coancestry/covariance among populations
  - X^TX: F_{ST} analog standardized by among-pop covariance
- Perform better in non-equilibrium situations
- Test for covariance with environmental factors

Lotterhos & Whitlock 2014

F_{ST}-based Outlier Testing

Newer methods: BayEnv, **BayPass**

- Account for population structure
  - Estimate coancestry/covariance among populations
  - X^TX: F_{ST} analog standardized by among-pop covariance
- Perform better in non-equilibrium situations
- Test for covariance with environmental factors

Lotterhos & Whitlock 2014

BayPass Input

Fortran

Population-level allele counts (required)

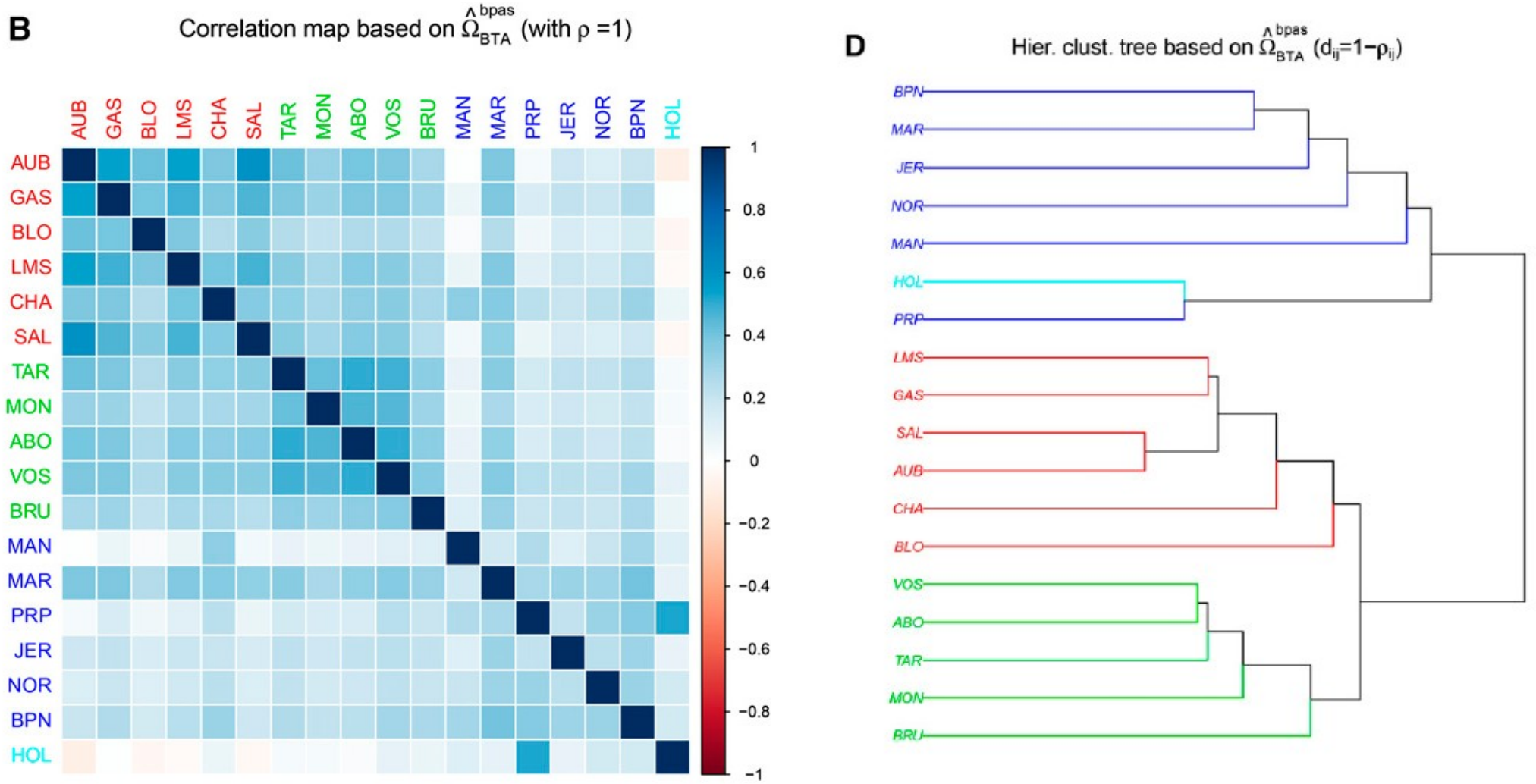
May be in chromosomal order, if known

```
--- file begins here ---
81 19 86 14 2 98 8 92 32 68 23 77
89 11 81 19 9 91 1 99 27 73 27 73
89 11 91 9 0 0 15 85 77 23 80 20
```

Covariate data file (optional)

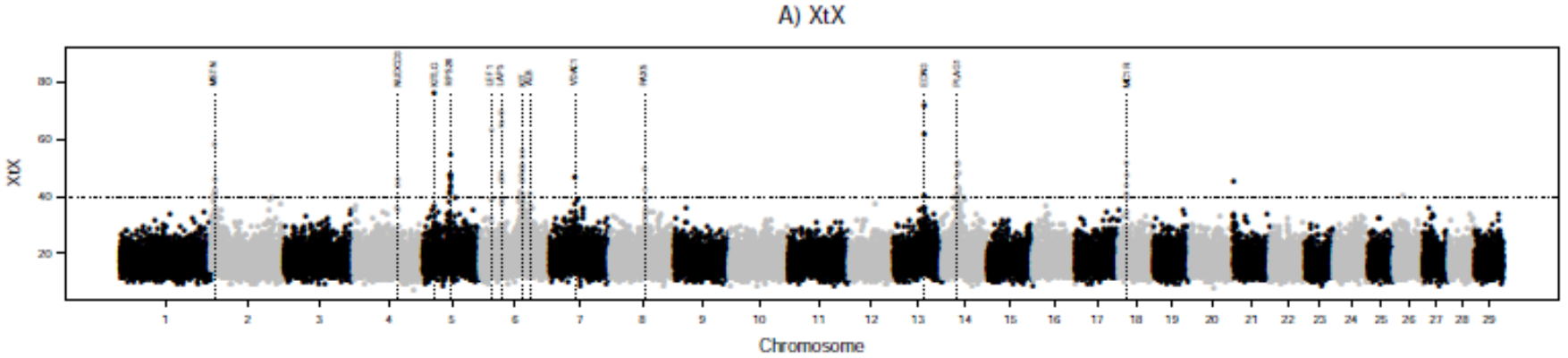
```
--- file begins here ---
150 1500 800 300 200 2500
181.5 172.6 152.3 191.8 154.2 166.8
1 1 0 0 1 1
0.1 0.8 -1.15 1.6 0.02 -0.5
```

Core model: Estimate covariance



Gautier 2015

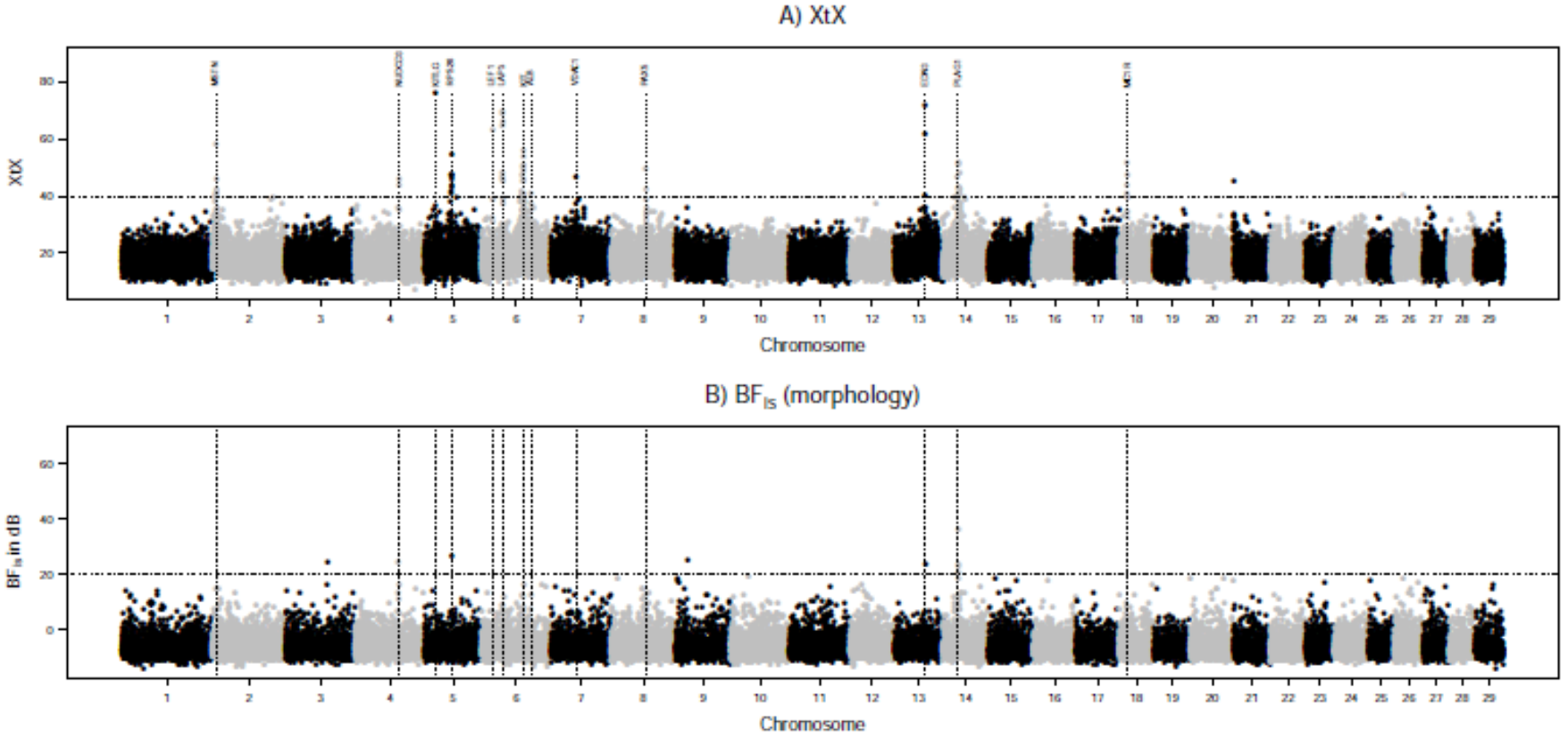
Core model: Calculate X^TX for all SNPs



Calibrate & determine significance via simulated data

Gautier 2015

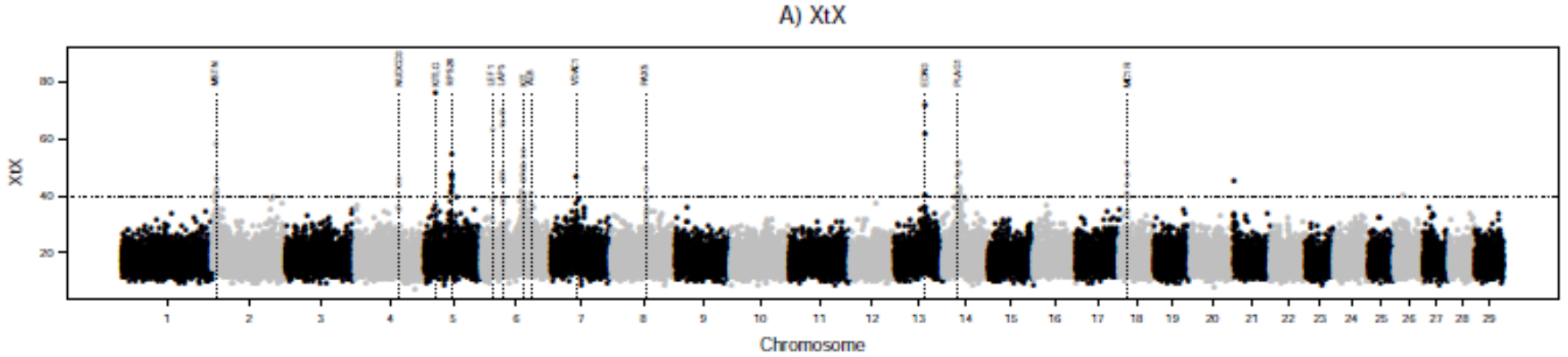
Covariate model: Importance Sampling



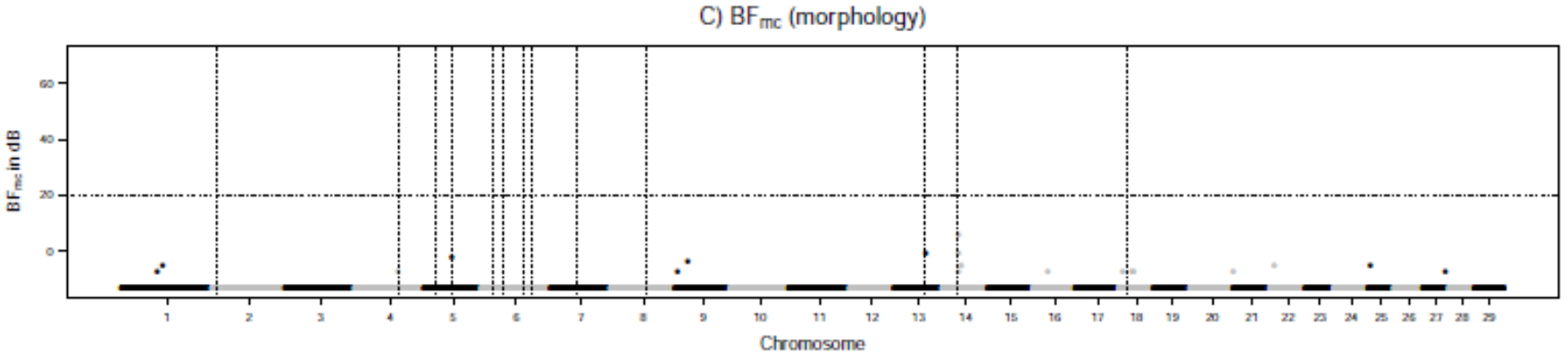
Estimate Bayes Factor for SNP-covariate association

Gautier 2015

Covariate model: Auxiliary Covariate



Estimate Bayes Factor for SNP-covariate association, allowing incorporation of marker position



Gautier 2015

BayPass: What is it good for?

Identifying selection between non-equilibrium populations, with or without marker position information

Genome-free mRNA-Seq Data Analysis

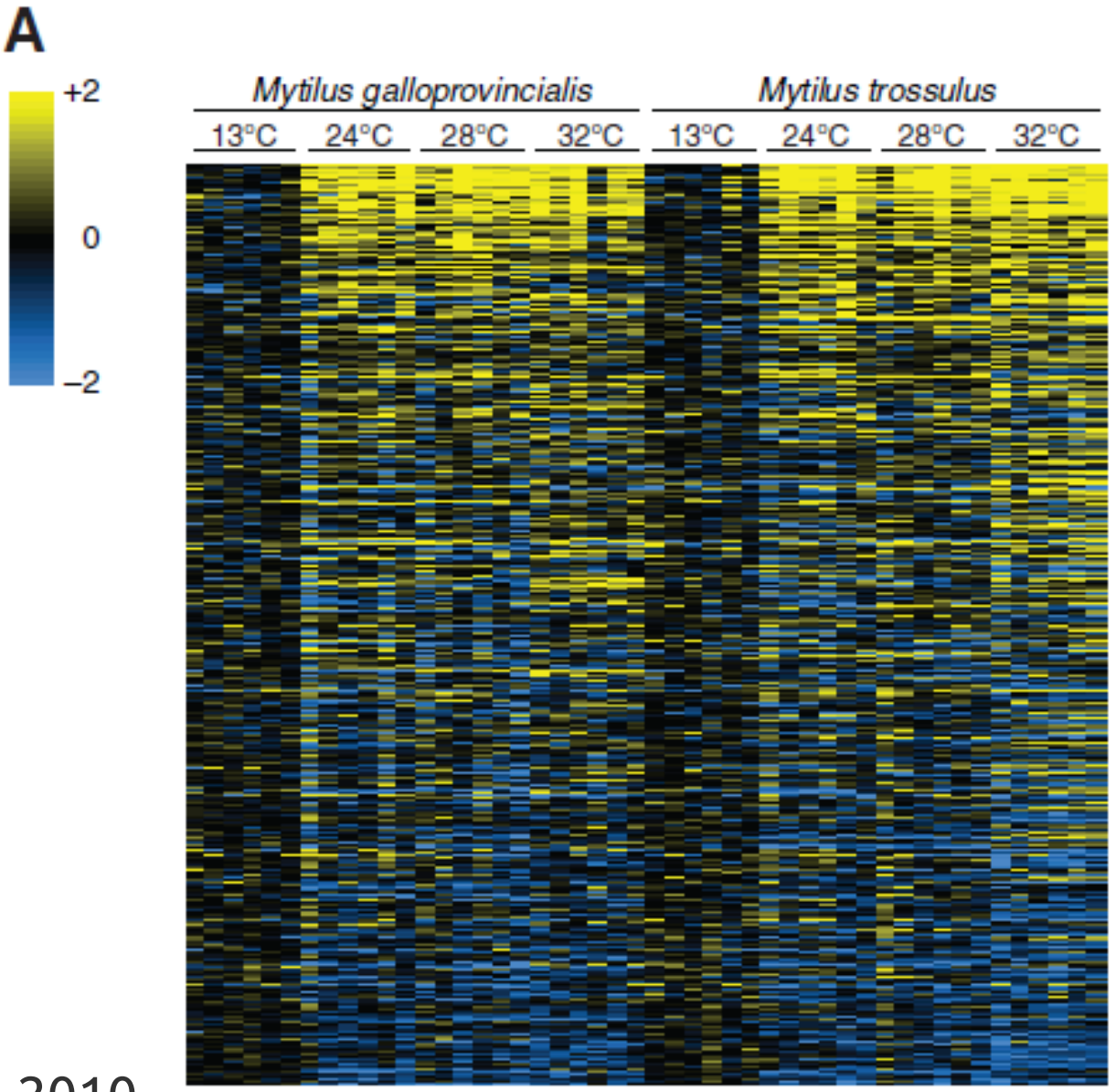
WGCNA: Gene co-expression network analysis

One or more species
Multiple conditions
Normalized expression data

Which genes are changing expression?

Gene Expression Analysis

Which genes are differentially expressed between treatments?



Lockwood *et al.* 2010

Gene Expression Analysis

Gene-by-gene methods: DESeq, EdgeR

- Assess each gene independently
- Enrichment analysis to identify important pathways / processes

Lots of natural variation, 10K+ tests – what's really important?

Gene Expression Analysis

Co-expression network analysis: WGCNA, MEGENA

- Identify modules of co-expressed genes
- Associate modules with traits/conditions of interest
- Enrichment analysis within modules to identify important pathways / processes

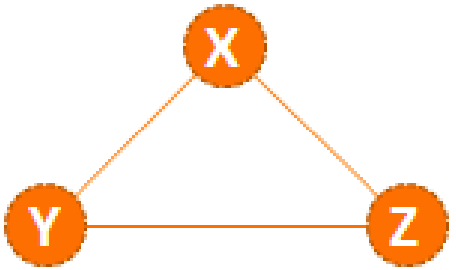
Gene Expression Analysis

Co-expression network analysis: **WGCNA**, MEGENA

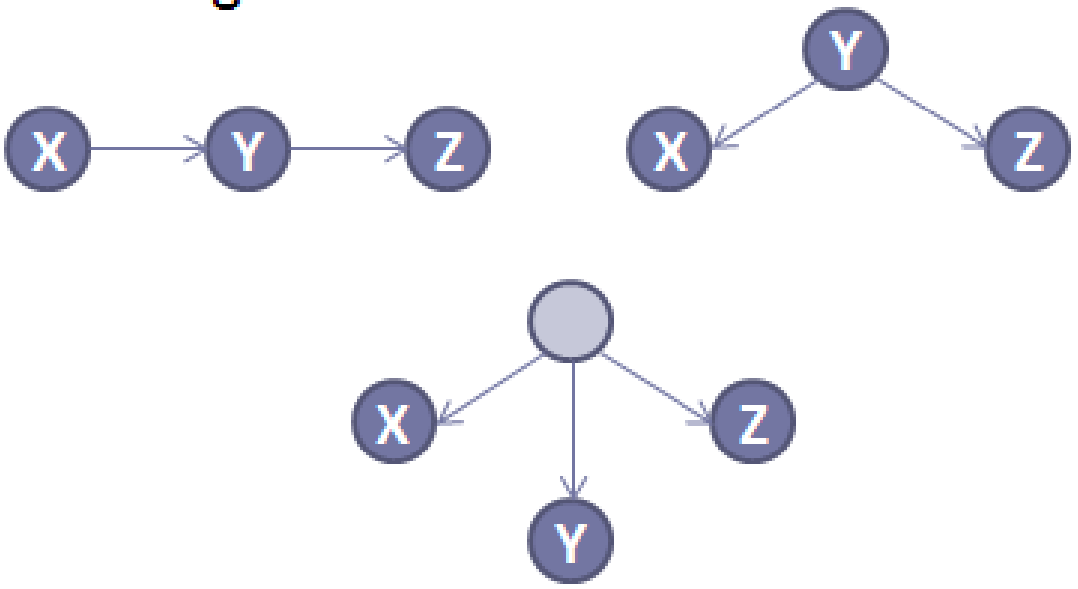
- Identify modules of co-expressed genes
- Associate modules with traits/conditions of interest
- Enrichment analysis within modules to identify important pathways / processes

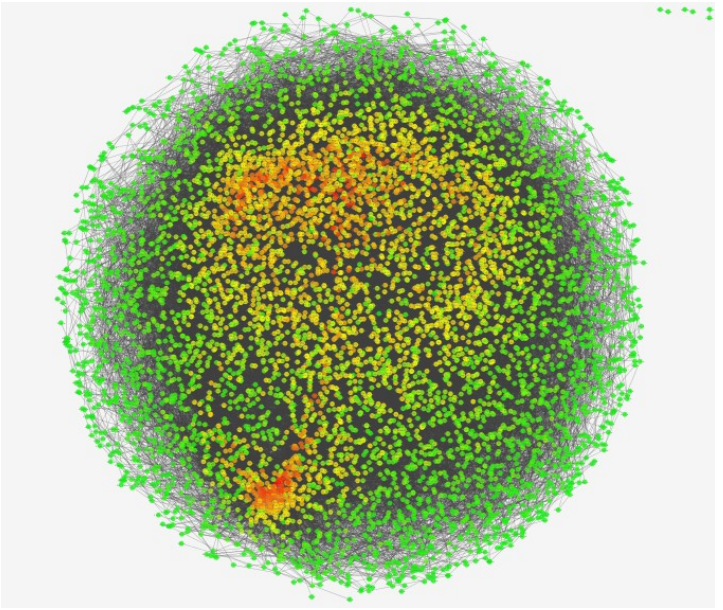
Gene Co-expression Analysis

Gene Co-expression



Gene Regulation





S. Mohammed H. Oloomi

Gene Co-expression Analysis

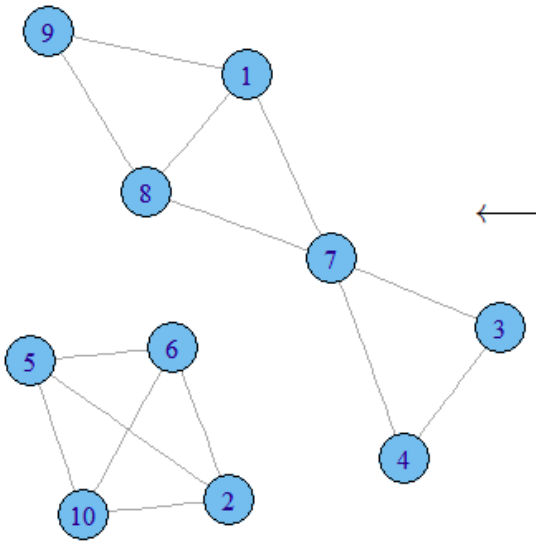
	S_1	S_2	S_3
G_1	43.26	40.89	5.05
G_2	166.6	41.87	136.65
G_3	12.53	39.55	42.09
G_4	28.77	191.92	236.56
G_5	114.7	79.7	99.76
G_6	119.1	80.57	114.59
G_7	118.9	156.69	186.95
G_8	3.76	2.48	136.78
G_9	32.73	11.99	118.8
G_{10}	17.46	56.11	21.41

Gene expression values

$|r(G_i, G_j)|$ → Pearson correlation

	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9	G_{10}
G_1	1.00	0.23	0.61	0.71	0.03	0.35	**0.86**	**1.00**	**0.97**	0.37
G_2	0.23	1.00	0.63	0.52	**0.98**	**0.99**	0.29	0.30	0.46	**0.99**
G_3	0.61	0.63	1.00	**0.99**	0.77	0.53	**0.93**	0.56	0.41	0.51
G_4	0.71	0.52	**0.99**	1.00	0.69	0.41	**0.97**	0.66	0.52	0.40
G_5	0.03	**0.98**	0.77	0.69	1.00	**0.95**	0.48	0.09	0.27	**0.94**
G_6	0.35	**0.99**	0.53	0.41	**0.95**	1.00	0.17	0.41	0.57	**1.00**
G_7	0.86	0.29	**0.93**	**0.97**	0.48	0.17	1.00	**0.83**	0.72	0.16
G_8	**1.00**	0.30	0.56	0.66	0.09	0.41	0.83	1.00	**0.98**	0.42
G_9	**0.97**	0.46	0.41	0.52	0.27	0.57	0.72	**0.98**	1.00	0.58
G_{10}	0.37	**0.99**	0.51	0.40	**0.94**	**1.00**	0.16	0.42	0.58	1.00

Similarity (Co-expression) score



	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9	G_{10}
G_1	0	0	0	0	0	0	1	1	1	0
G_2	0	0	0	0	1	1	0	0	0	1
G_3	0	0	0	1	0	0	1	0	0	0
G_4	0	0	1	0	0	0	1	0	0	0
G_5	0	1	0	0	0	1	0	0	0	1
G_6	0	1	0	0	1	0	0	0	0	1
G_7	1	0	1	1	0	0	0	1	0	0
G_8	1	0	0	0	0	0	1	0	1	0
G_9	1	0	0	0	0	0	0	1	0	0
G_{10}	0	1	0	0	1	1	0	0	0	0

$|r(G_i, G_j)| >= 0.8$ → Significance threshold

Network adjacency matrix

S. Mohammed H. Oloomi

WGCNA Input

R package

Normalized gene expression data (required)

May be transformed, but does not have to be

	F2_2	F2_3	F2_14	F2_15	F2_19	F2_20	F2_23
MMT00000044	-0.01810	0.0642	6.44e-05	-0.05800	0.04830	-0.15197410	-0.00129
MMT00000046	-0.07730	-0.0297	1.12e-01	-0.05890	0.04430	-0.09380000	0.09340
MMT00000051	-0.02260	0.0617	-1.29e-01	0.08710	-0.11500	-0.06502607	0.00249
MMT00000076	-0.00924	-0.1450	2.87e-02	-0.04390	0.00425	-0.23610000	-0.06900
MMT00000080	-0.04870	0.0582	-4.83e-02	-0.03710	0.02510	0.08504274	0.04450
MMT00000102	0.17600	-0.1890	-6.50e-02	-0.00846	-0.00574	-0.01807182	-0.12500
MMT00000149	0.07680	0.1860	2.14e-01	0.12000	0.02100	0.06222751	0.22600

Trait data file (optional)

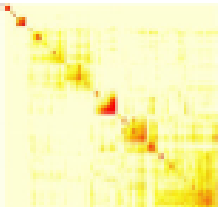
	weight_g	length_cm	ab_fat	other_fat	total_fat	X100xfat_weight	Trigly
F2_2	38.0	10.5	3.81	2.78	6.59	17.342105	14
F2_3	33.5	10.8	1.70	2.05	3.75	11.194030	109
F2_14	33.9	10.0	1.29	1.67	2.96	8.731563	2
F2_15	44.3	10.3	3.62	3.34	6.96	15.711061	71
F2_19	32.9	9.7	2.08	1.85	3.93	11.945289	55
F2_20	44.8	10.3	3.72	3.20	6.92	15.446429	34

Gene annotation information (optional)

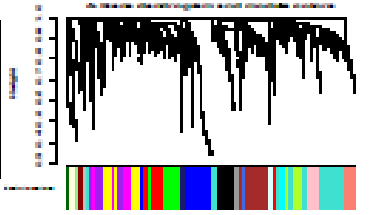
Langfelder & Horvath 2008

WGCNA Overview

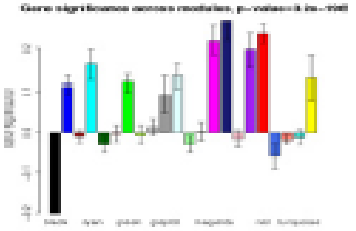
Construct a gene co-expression network
Rationale: make use of interaction patterns among genes
Tools: correlation as a measure of co-expression



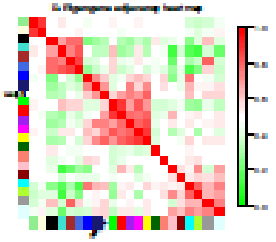
Identify modules
Rationale: module (pathway) based analysis
Tools: hierarchical clustering, Dynamic Tree Cut



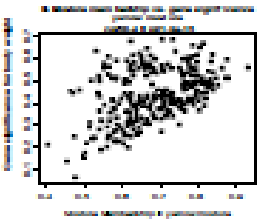
Relate modules to external information
Array Information: clinical data, SNPs, proteomics
Gene Information: ontology, functional enrichement
Rationale: find biologically interesting modules



Study module relationships
Rationale: biological data reduction, systems-level view
Tools: Eigengene Networks

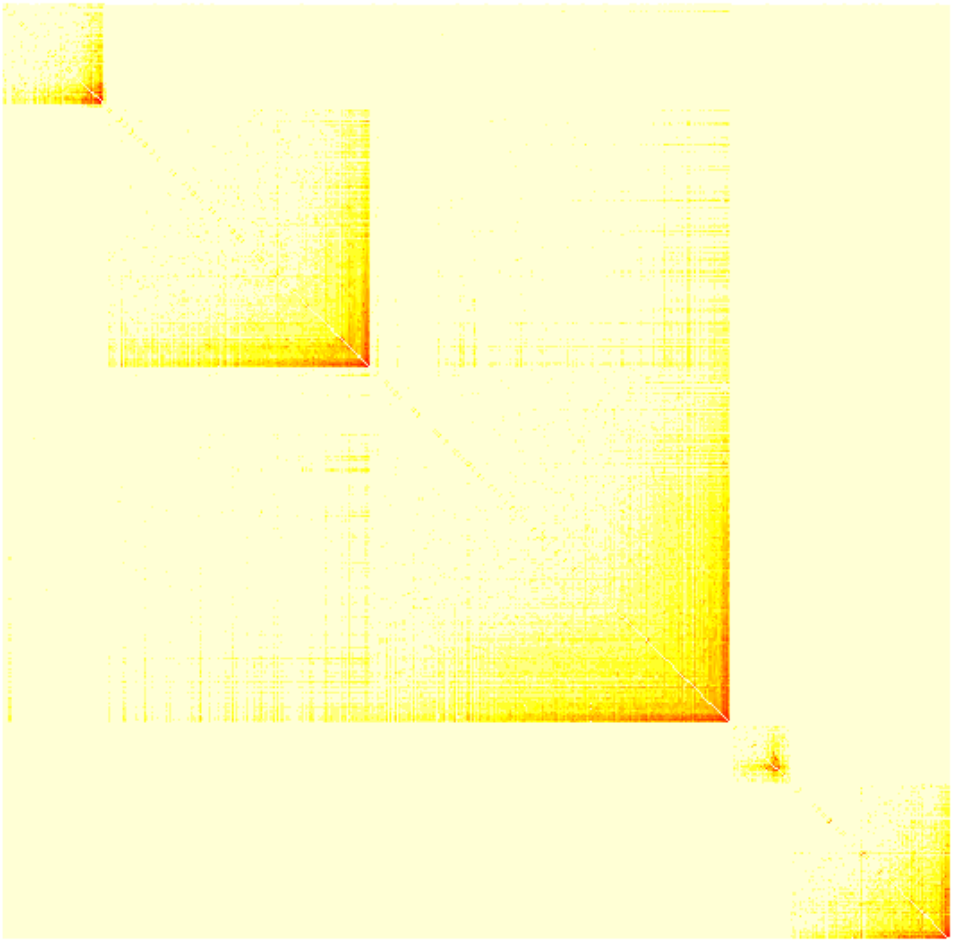


Find the key drivers in *interesting* modules
Rationale: experimental validation, biomarkers
Tools: intramodular connectivity, causality testing



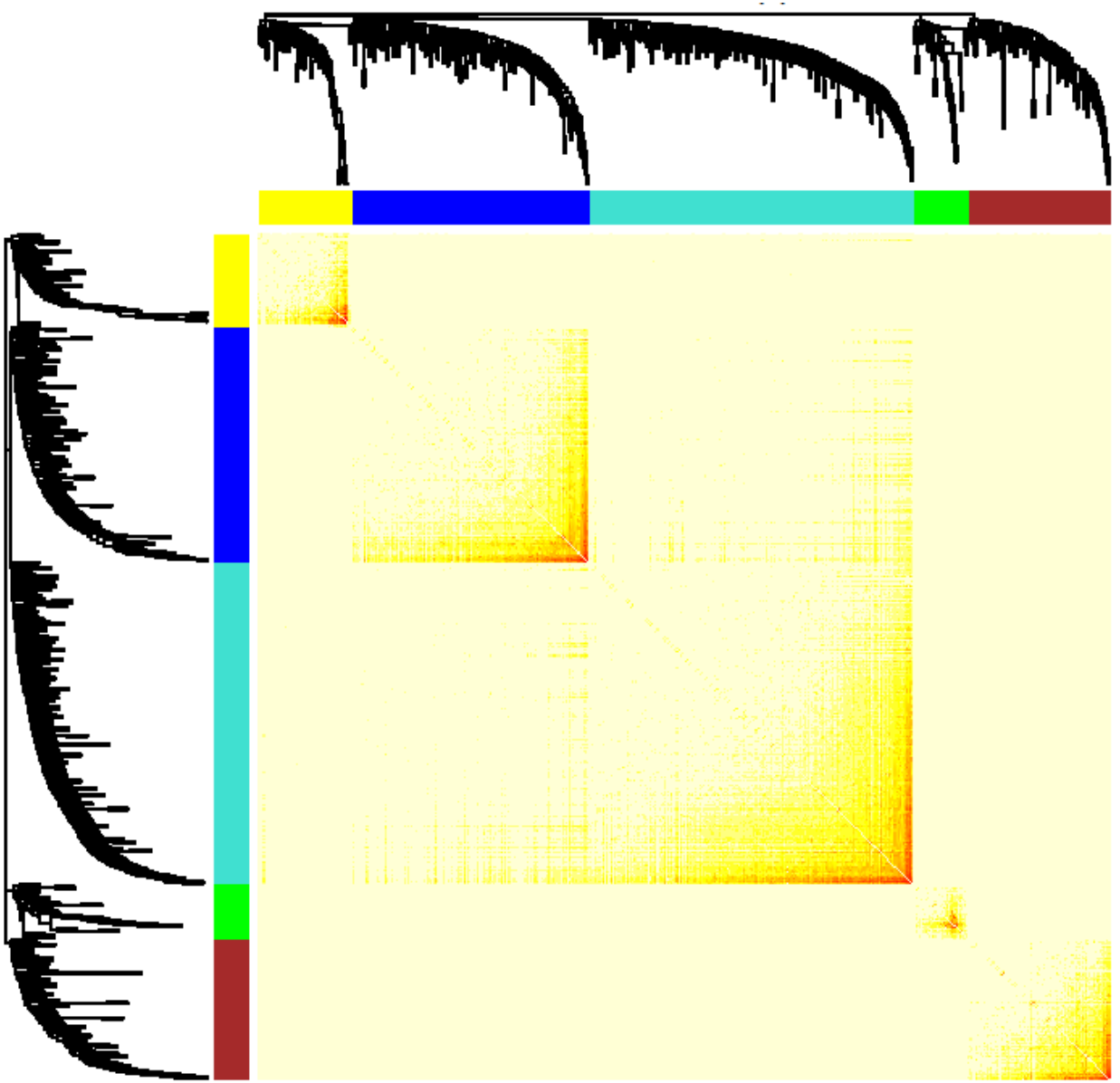
Langfelder & Horvath 2008

Step 1: Construct co-expression network



Langfelder & Horvath 2008

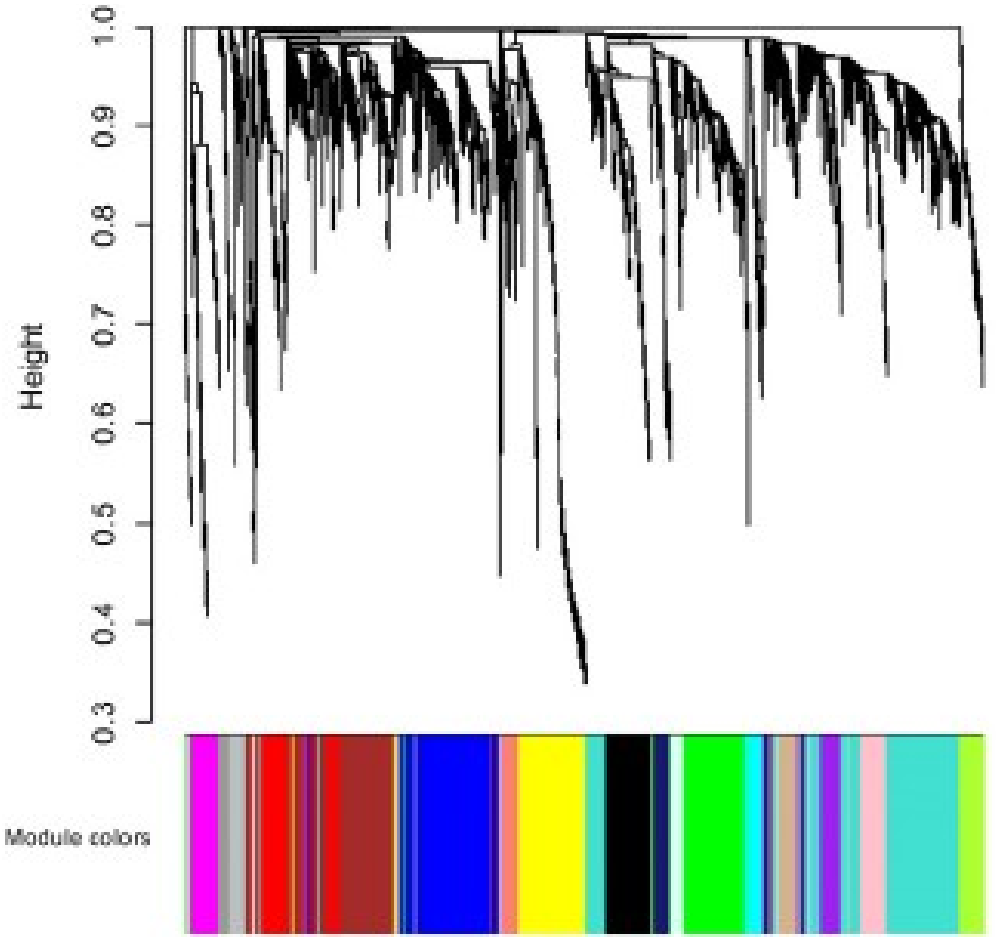
Step 1: Construct co-expression network



Langfelder & Horvath 2008

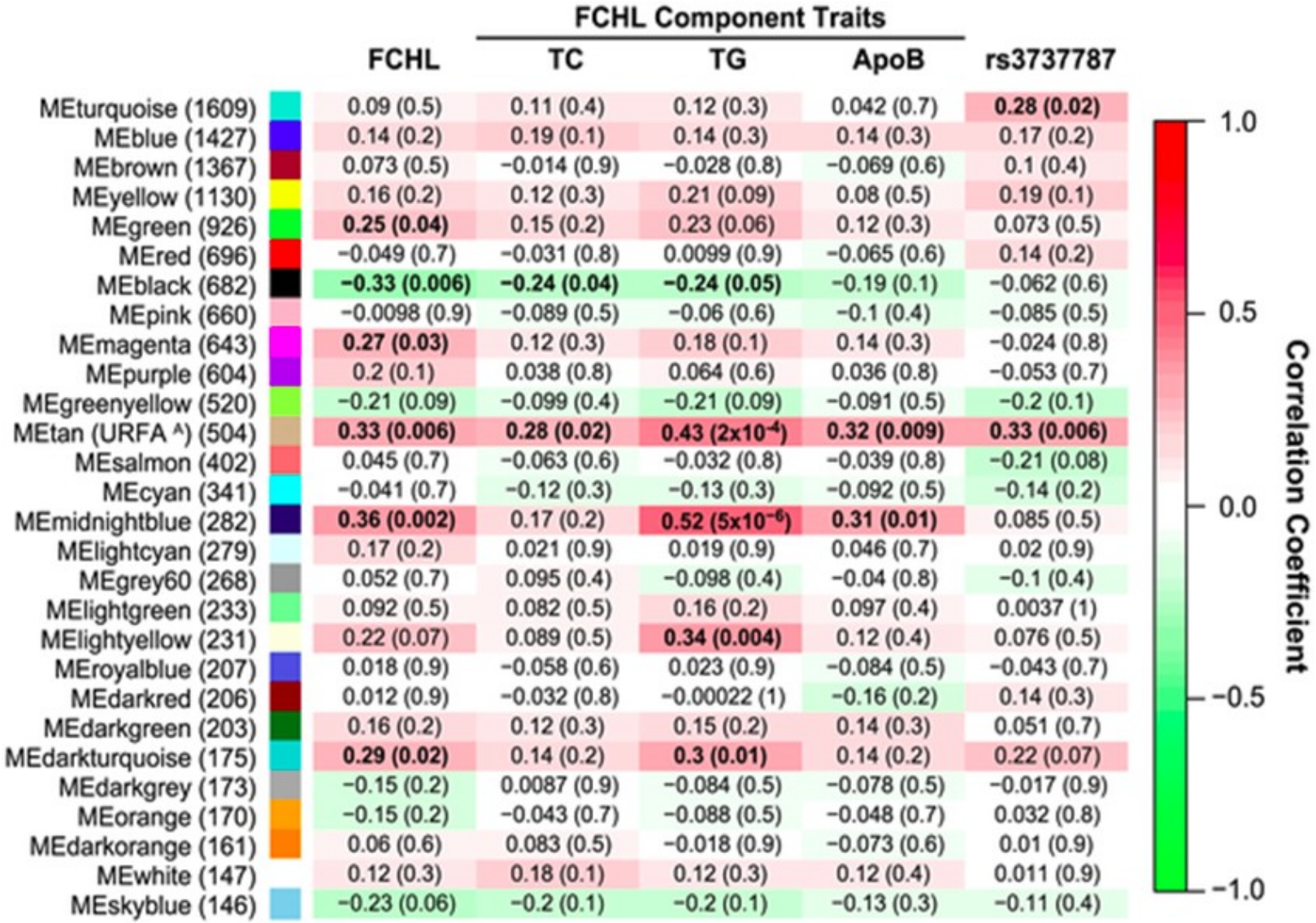
Step 1: Construct co-expression network

Computationally intensive – can perform all at once (ideal), or blockwise with gene subsets



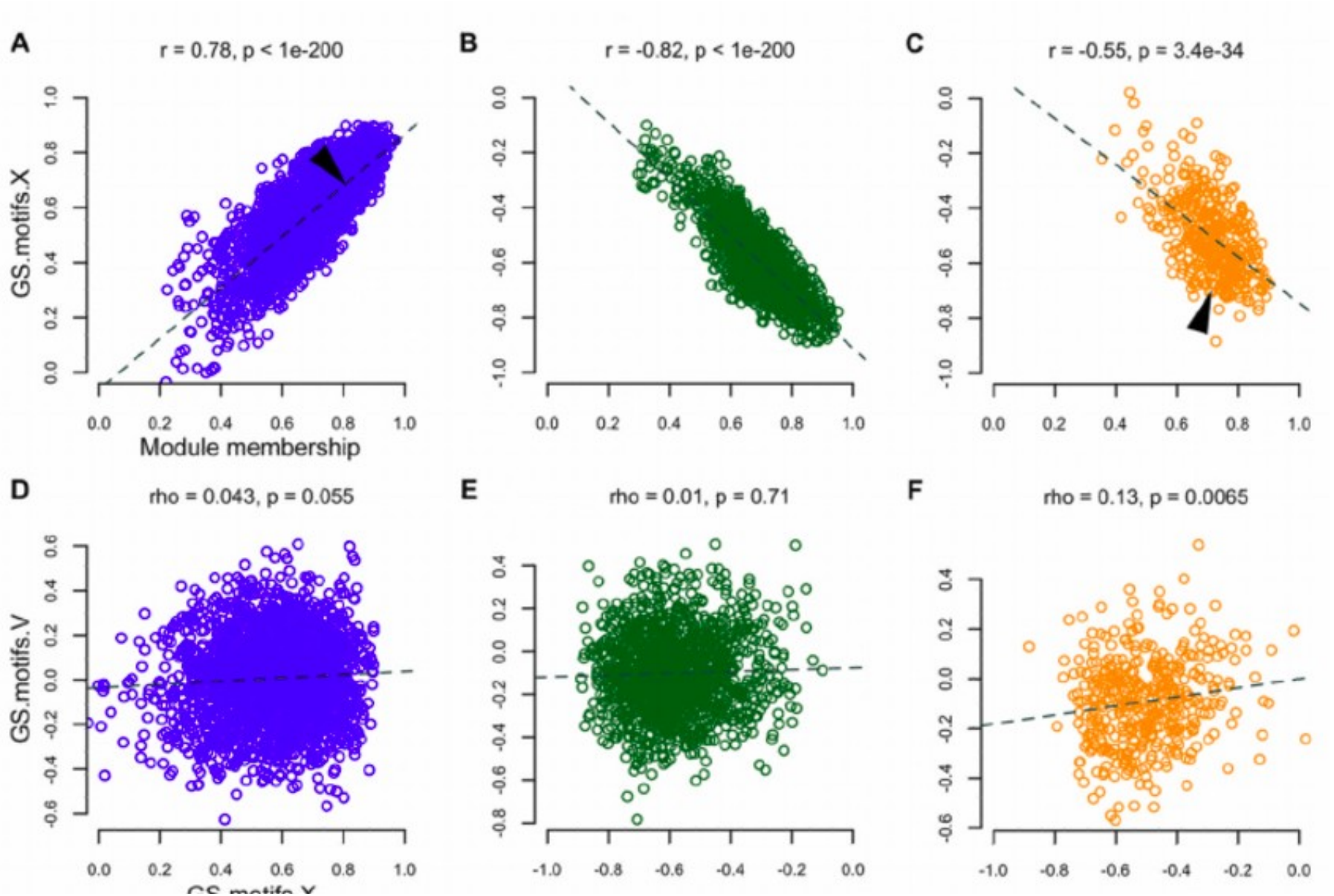
Langfelder & Horvath 2008

Step 2: Relate Modules with Traits



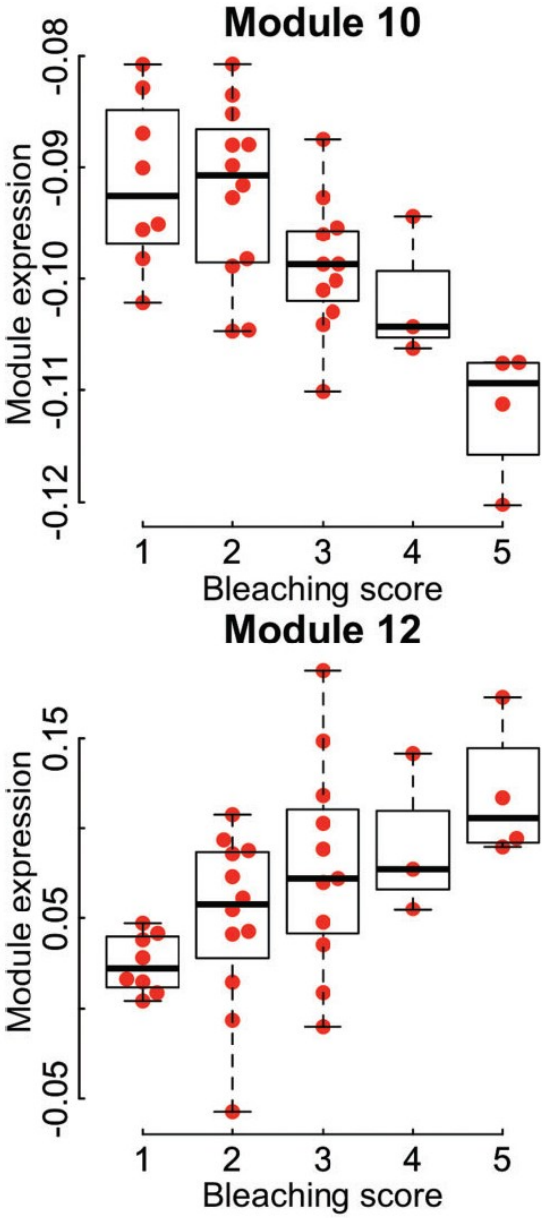
http://www.incodom.kr/Biological_network/WGCNA

Step 2: Relate Modules with Traits

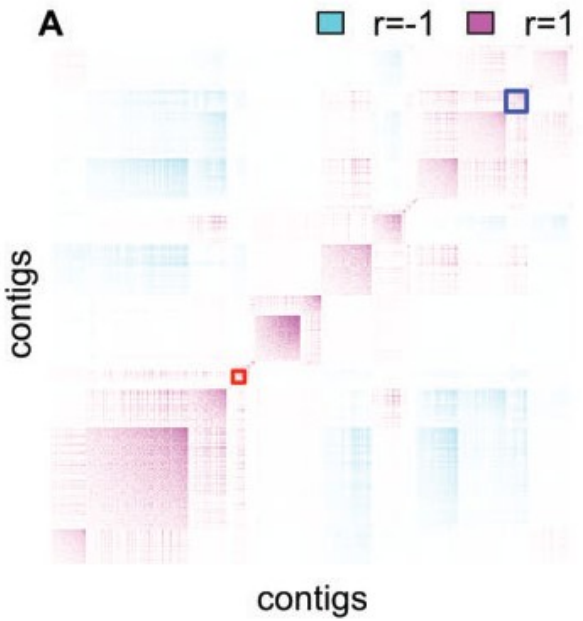


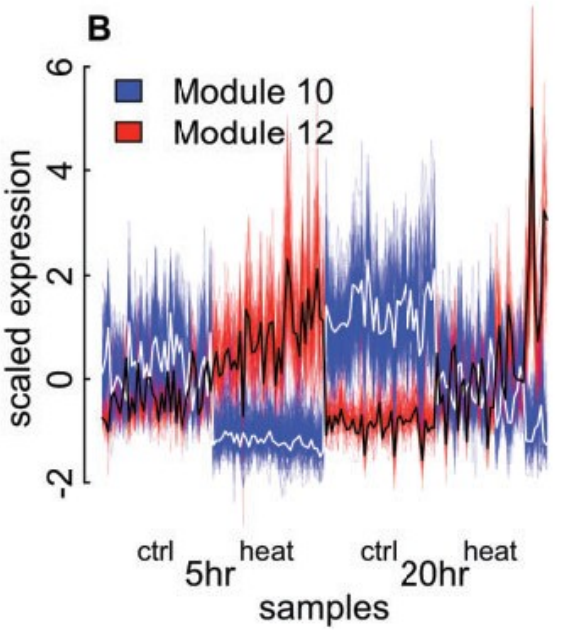
Hilliard et al. 2012

Step 3: Explore Gene Networks



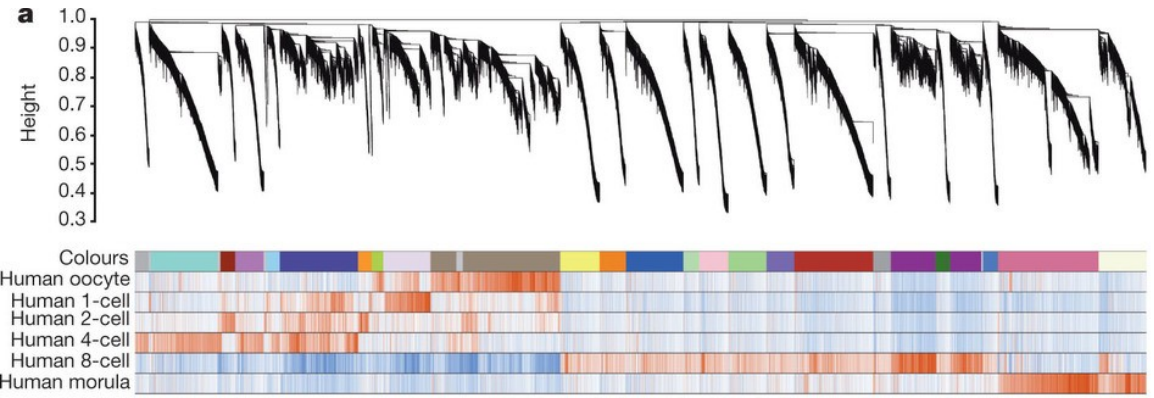
Identify expression patterns associated with a specific biological response.

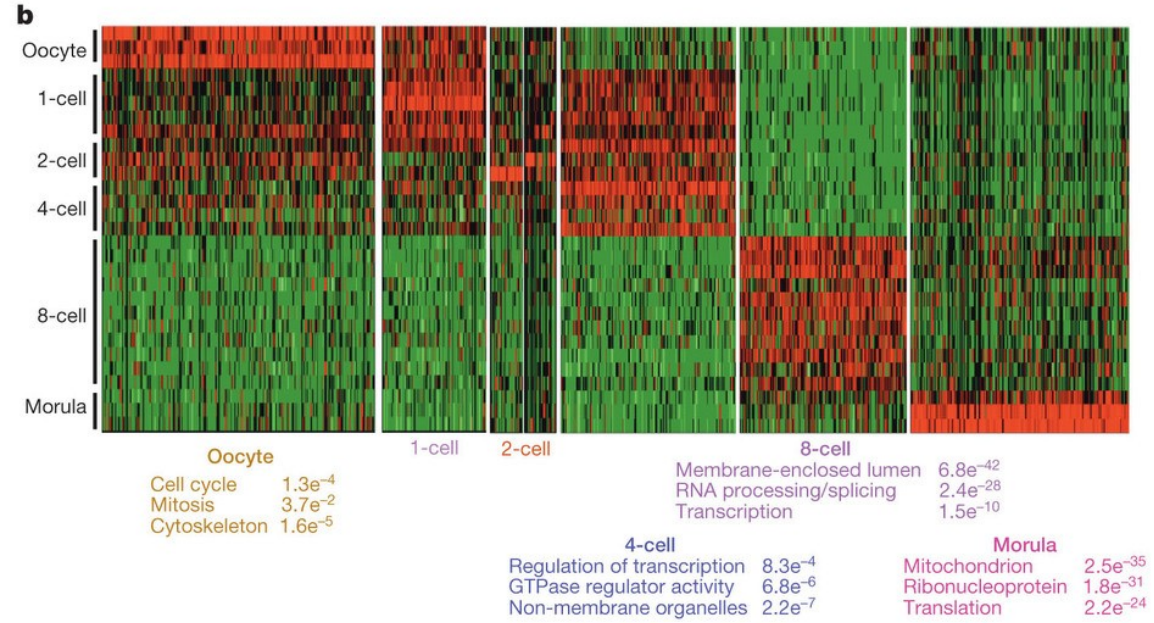


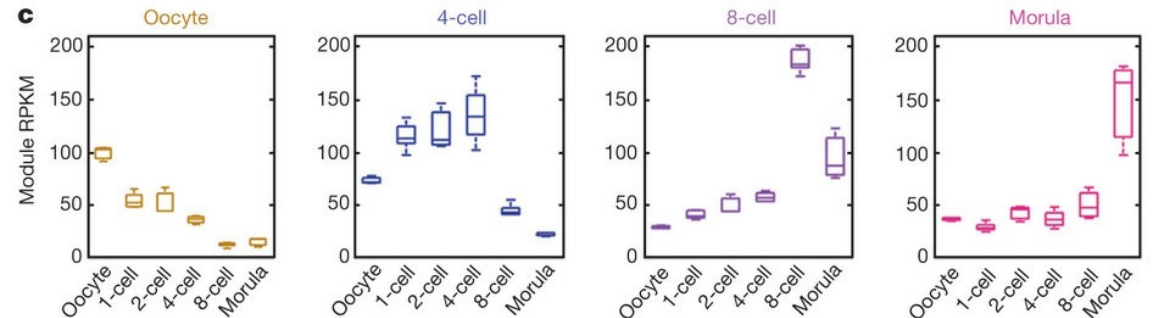


Rose et al. 2015

Step 3: Explore Gene Networks



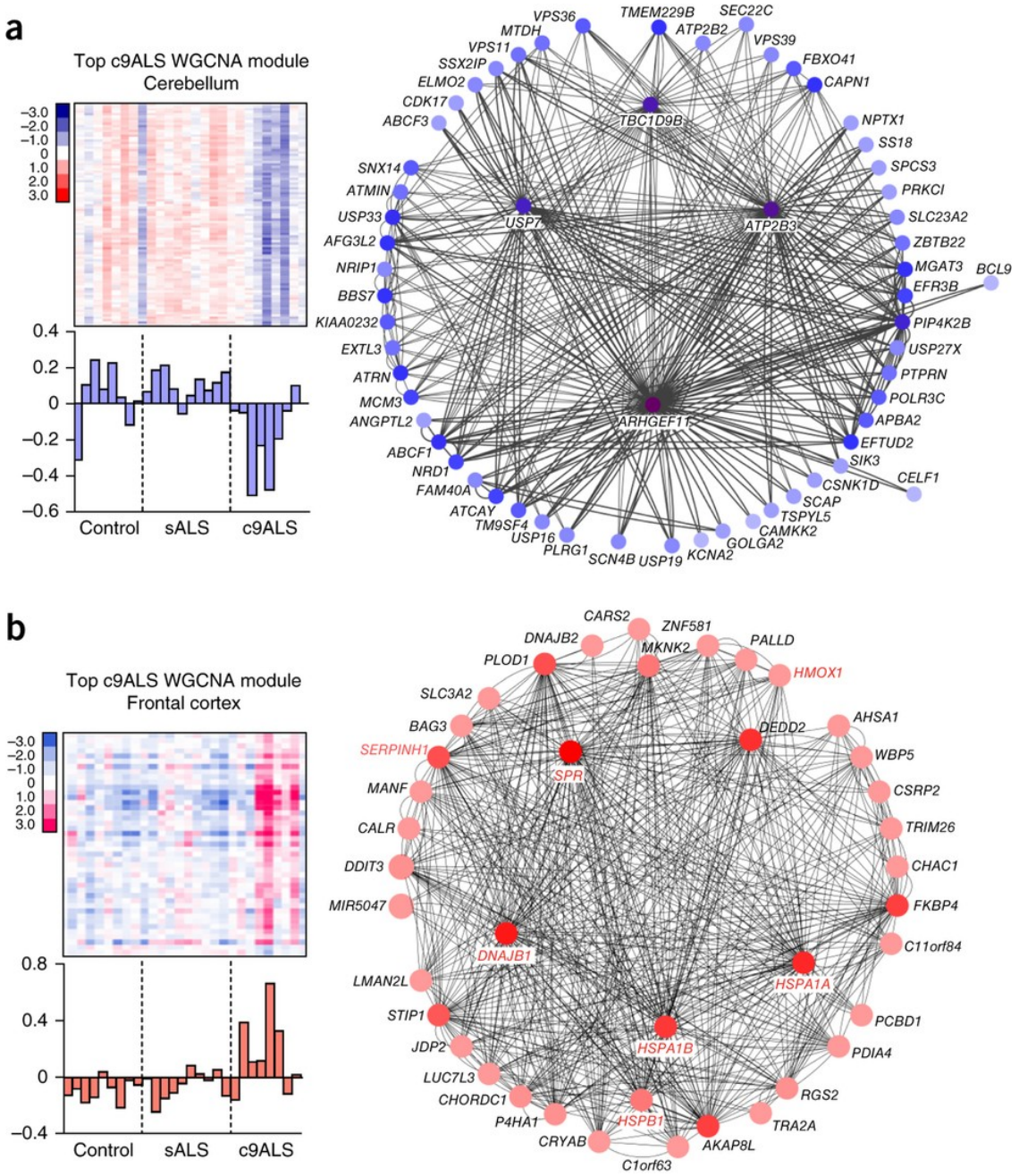




Annotate genes
add GO terms, etc.
Within-module
enrichment analyses.

Xue et al. 2013

Step 3: Explore Gene Networks

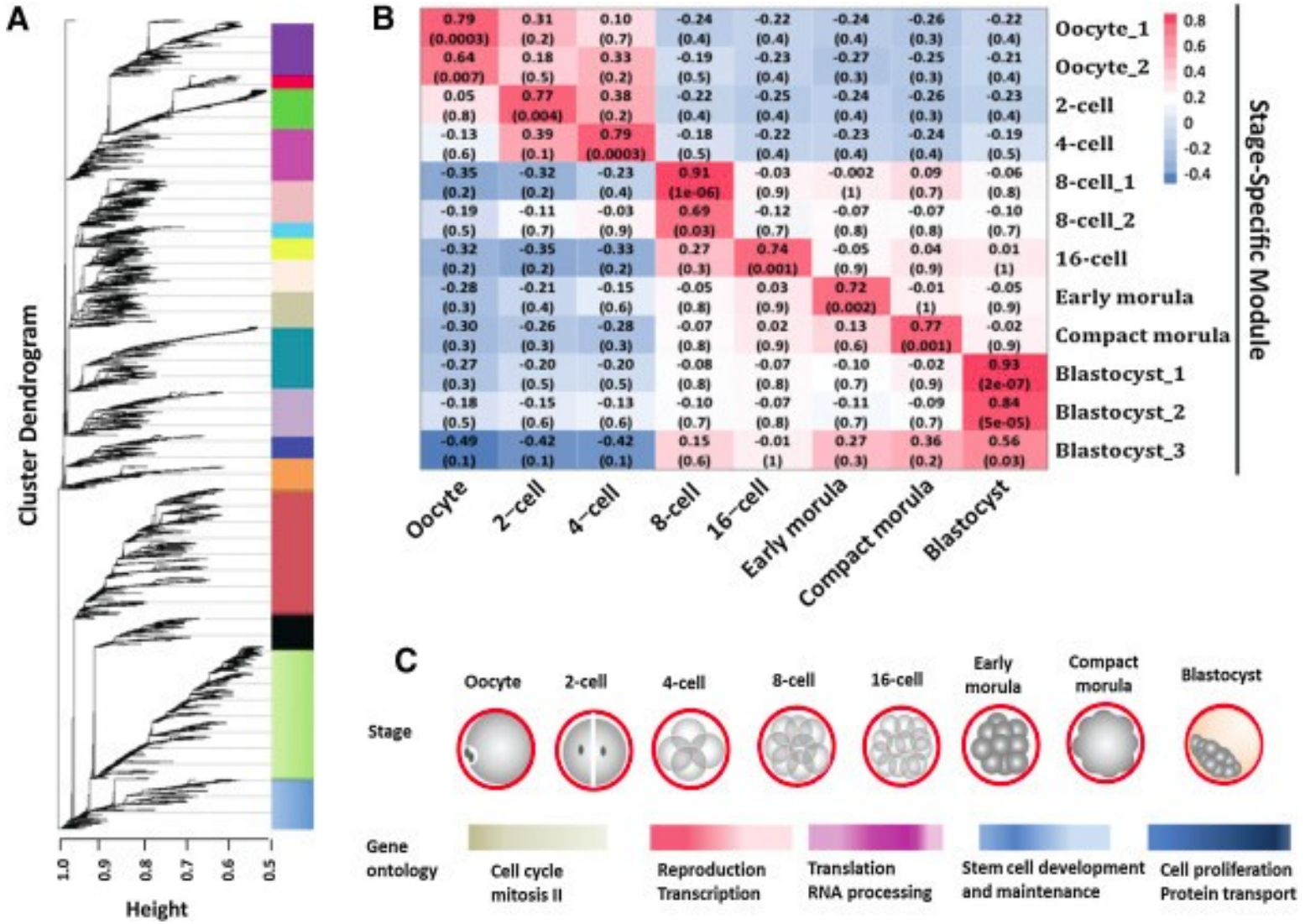


Plot networks & identify key “hub” genes.

Prudencio et al. 2015

Step 3: Explore Gene Networks

Compare expression similarity among different conditions.



Jiang et al. 2014

WGCNA: What is it good for?

Identifying & exploring the gene networks underlying regulatory variation.