Nice R package for making heatmaps
pheatmap

Ca2+/calmodulin-dependent protein kinase, EF-Hand protein superfamily;Ca-dependent protein kinase E=5e-138
Trypsin ; mannan-binding lectin serine protease 1 E=6e-65
Na+/K+ symporter ; solute carrier family 12 (sodium/potassium/chloride transporter), member 2 E=5e-95
Aryl-hydrocarbon receptor nuclear translocator;neuronal PAS domain-containing protein 2 E=1e-56
Transcriptional repressors of the hairy/E(spl) family;hairy and enhancer of split related with YRPW motif E=7e-160
Deoxyribodipyrimidine photolyase/cryptochrome;cryptochrome E=7e-122
Transcriptional repressors of the hairy/E(spl) family;hairy and enhancer of split related with YRPW motif E=9e-29
Deoxyribodipyrimidine photolyase/cryptochrome;cryptochrome E=2e-63

http://cran.r-project.org/
web/packages/pheatmap/
index.html

```
> head(exp)
              green1_dk mega1_dk green2_gr mega2_gr mega3_red green3_dk
isogroup10979  6.296149 6.570837  5.899552 5.356431  6.724022  6.417827
isogroup13244  6.750216 7.029649  5.562750 5.938274  7.087965  6.659479
isogroup1996   7.139482 7.142619  6.171526 6.373726  7.148896  7.033077
isogroup7904   5.356431 5.356431  6.035024 6.498003  5.685713  5.577790
isogroup3676   6.211784 5.871480  6.621497 6.867575  6.154541  5.978316
isogroup29627  5.356431 5.871480  5.356431 5.738803  5.356431  5.356431
              mega4_dk green4_gr mega5_gr green5_red mega6_red   red1_dk
isogroup10979 6.772880  5.916744 5.803765   6.391164  6.483878 6.556468
isogroup13244 7.013566  5.356431 5.615388   6.728181  6.925941 6.689225
isogroup1996  7.263313  5.975031 6.275734   7.300427  7.267887 6.527665
isogroup7904  5.578430  6.457216 6.165769   5.559693  5.645039 5.632151
isogroup3676  6.257922  6.788670 6.936609   5.962260  6.255969 6.216852
isogroup29627 5.356431  5.356431 5.356431   5.356431  5.356431 6.689225
               red2_gr red3_red   red4_dk  red5_gr red6_red
isogroup10979 5.356431 6.378329  6.654834 5.903847 6.371307
isogroup13244 5.677243 6.638230  6.357824 5.650287 6.485595
isogroup1996  6.328771 6.638230  6.935248 6.249401 7.133316
isogroup7904  5.994173 5.356431  5.574034 6.098703 5.567701
isogroup3676  6.818004 6.045460  6.312892 6.660526 6.263160
isogroup29627 6.376728 6.086411  6.461024 6.335759 6.110251
>
```

>library(pheatmap)

>pheatmap(cor(exp),cluster_rows=F,cluster_cols=F)

>pheatmap(cor(exp))

Response to Oxidative Stress

Thioredoxin reductase ; glutathione reductase
Fumarase
Cu2+/Zn2+ superoxide dismutase SOD1
Catalase
Cytochrome c oxidase, subunit Va/COX6
Catalase
Ferritin ; ferritin heavy chain
Cytochrome c oxidase, subunit VIb/COX12
Glutathione peroxidase
Cytochrome oxidase assembly factor COX11

Chromosome Organization

Transcription initiation factor TFIID, subunit TAF
Histone deacetylase complex, catalytic component R
Transcription initiation factor TFIID, subunit TAF
Histone H1/5
Transcription initiation factor TFIID, subunit TAF
Histone acetyltransferase SAGA/ADA, catalytic subu
Structural maintenance of chromosome protein 4
Histones H3 and H4 ; histone H3-like centromeric
chromobox protein
Histone acetyltransferases PCAF/SAGA/ADA, subunit
Histone H1/5
Histone H2B
Histones H3 and H4 ; histone H3-like centromeric
histone-lysine N-methyltransferase MLL3
Chromodomain-helicase DNA-binding protein
Ca2+-binding actin-bundling protein (actinin)
Histone acetyltransferase SAGA/ADA, catalytic
Clathrin coat binding protein/Huntingtin interacting
Histone acetyltransferase (MYST family)
Component of histone deacetylase complex
Histone acetyltransferase (MYST family)
Histone acetyltransferase (MYST family)
Histone H2B ; histone H2B E=1e-57
Histone 2A ; histone H2A E=7e-61

Collagen/Extracellular Matrix Structural Constituent

Collagens (type IV and type XIII), and related
Collagens (type IV and type XIII), and related
Collagens (type IV and type XIII), and related
Collagens (type IV and type XIII), and related
Collagens (type IV and type XIII), and related
Collagens (type IV and type XIII), and related
Collagens (type IV and type XIII), and related
Collagens (type IV and type XIII), and related
Collagens (type IV and type XIII), and related

cAMP-dependent protein kinase complex/regulator activity

Polo-like serine/threonine protein kinase
Ca2+/calmodulin-dependent protein kinase, EF-Hand
FOG: RRM domain
eIF-2alpha kinase GCN2
Protein kinase C substrate, 80 KD protein, heavy chain
Serine/threonine protein kinase ; TP53 regulating
Serine-threonine protein kinase FUSED
Predicted protein kinase KKIAMRE ; cyclin-dependent
Cdk activating kinase (CAK)/RNA polymerase II tran
Cytoskeleton-associated protein and related protein
Serine/threonine protein kinase ; NIMA
Cytoskeleton-associated protein and related protein
Dystonin, GAS (Growth-arrest-specific protein)
Hsp27-ERE-TATA-binding protein/Scaffold attachment
Cytoskeleton-associated protein and related protein
Transforming growth factor beta/activin receptor
cAMP-dependent protein kinase types I and II
Splicing coactivator SRm160/300, subunit SRm300
Ribosomal protein S6 kinase ; p70 ribosomal S6 kinase
Actin binding protein Anillin
p21-activated serine/threonine protein kinase
Serine/threonine protein kinase ; serine/threonine

green1 green2 green3 green4 green5 mega1 mega2 mega3 mega4 mega5 mega6 red1 red2 red3 red4 red5 red6

```
>means=apply(exp,1,mean)
# means of rows

>explc=exp-means
# subtracting them

>pdf
("file_name",height=35,width=9)

>pheatmap(as.matrix(explc),color
= colorRampPalette(rev(c
("#D73027","#D73027","#D73027
","#FEE090","#FFFFBF",
"#91BFDB","#4575B4","#4575B4")
))
(100),cex=0.9,border_color=NA,cl
ustering_distance_rows="correlati
on",clustering_distance_cols="cor
relation")

>dev.off()
```
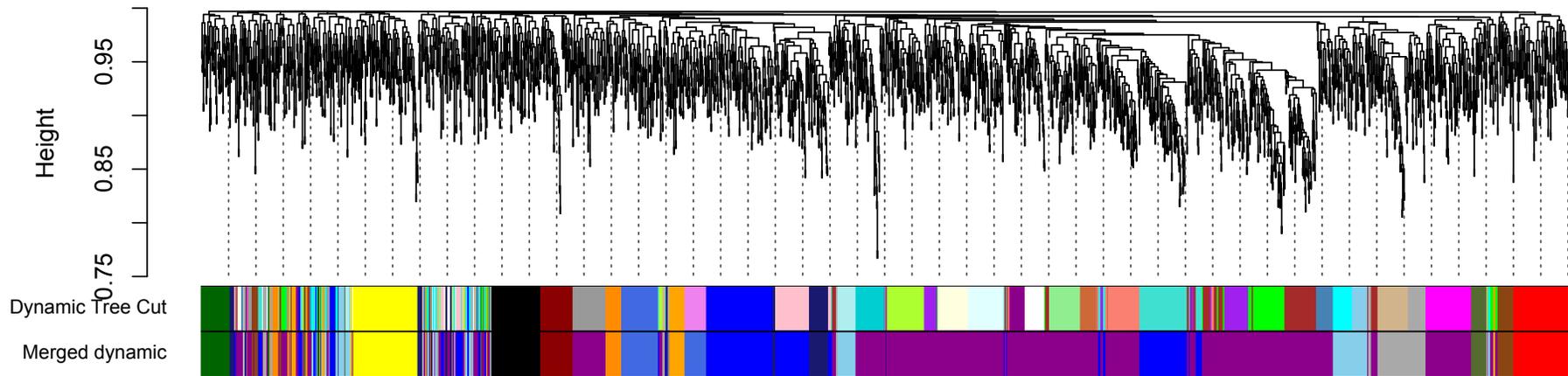
WGCNA

Dept. of Human Genetics, UC Los Angeles (PL, SH), and Dept. of Biostatistics, UC Los Angeles (SH)

Peter (dot) Langfelder (at) gmail (dot) com, SHorvath (at) mednet (dot) ucla (dot) edu

BMC Bioinformatics, 2008 9:559

Special thanks to Rachel Wright for slides

Extensive Tutorials Online

- http://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/
- http://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/

- R package

```
>source("http://bioconductor.org/biocLite.R")
>biocLite("impute")
>install.packages("WGCNA")
```

WGCNA: weighted gene coexpression network analysis

- Identify groups (modules) of co-regulated genes
- Correlate module expression values to trait data
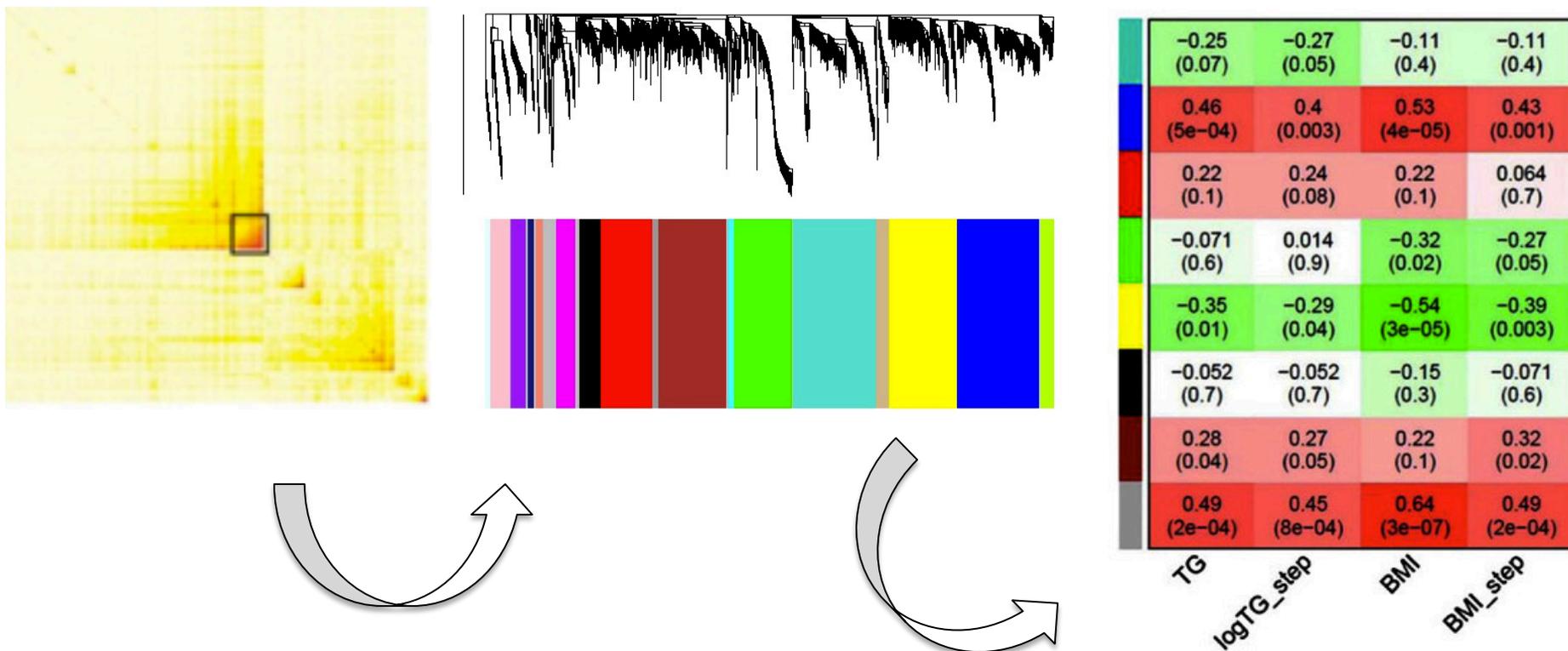
Benefits of WGCNA

- Unsupervised and unbiased
  - Bypasses multiple testing correction
- Finds unique co-expression modules that correspond to clusters of correlated transcripts

What kind of data do you need?

- Normalized counts data files (.csv)
  - VSD file: variance transformed counts data
    - Generated from DESeq
  - The tutorial you use depends on how large your dataset is (18000 genes? Or 2000 genes?)
- Trait Data
  - Made in excel, saved as .csv

VSD file

- \> cdsBlind = estimateDispersions(cds, method="blind")
- \> vsd = varianceStabilizingTransformation (cdsBlind)

```
> head(exp)
            green1_dk mega1_dk green2_gr mega2_gr mega3_red green3_dk
isogroup10979  6.296149 6.570837  5.899552 5.356431  6.724022  6.417827
isogroup13244  6.750216 7.029649  5.562750 5.938274  7.087965  6.659479
isogroup1996   7.139482 7.142619  6.171526 6.373726  7.148896  7.033077
isogroup7904   5.356431 5.356431  6.035024 6.498003  5.685713  5.577790
isogroup3676   6.211784 5.871480  6.621497 6.867575  6.154541  5.978316
isogroup29627  5.356431 5.871480  5.356431 5.738803  5.356431  5.356431
            mega4_dk green4_gr mega5_gr green5_red mega6_red   red1_dk
isogroup10979 6.772880  5.916744 5.803765   6.391164  6.483878 6.556468
isogroup13244 7.013566  5.356431 5.615388   6.728181  6.925941 6.689225
isogroup1996  7.263313  5.975031 6.275734   7.300427  7.267887 6.527665
isogroup7904  5.578430  6.457216 6.165769   5.559693  5.645039 5.632151
isogroup3676  6.257922  6.788670 6.936609   5.962260  6.255969 6.216852
isogroup29627 5.356431  5.356431 5.356431   5.356431  5.356431 6.689225
             red2_gr red3_red  red4_dk  red5_gr red6_red
isogroup10979 5.356431 6.378329 6.654834 5.903847 6.371307
isogroup13244 5.677243 6.638230 6.357824 5.650287 6.485595
isogroup1996  6.328771 6.638230 6.935248 6.249401 7.133316
isogroup7904  5.994173 5.356431 5.574034 6.098703 5.567701
isogroup3676  6.818004 6.045460 6.312892 6.660526 6.263160
isogroup29627 6.376728 6.086411 6.461024 6.335759 6.110251
>
```

Trait Data file (upload in R as .csv)

- Categorical or quantitative traits
- Examples:

Sample	mega	green	red	dark_light	green_light	red_light	weight
green1_dk	0	1	0	1	0	0	8
green2_gr	0	1	0	0	1	0	9
green3_dk	0	1	0	1	0	0	10
green4_gr	0	1	0	0	1	0	11.2
green5_red	0	1	0	0	0	1	7
mega1_dk	1	0	0	1	0	0	9
mega2_gr	1	0	0	0	1	0	12.5
mega3_red	1	0	0	0	0	1	8.7
mega4_dk	1	0	0	1	0	0	9.2
mega5_gr	1	0	0	0	1	0	10.2
mega6_red	1	0	0	0	0	1	13
red1_dk	0	0	1	1	0	0	14
red2_gr	0	0	1	0	1	0	6
red3_red	0	0	1	0	0	1	3
red4_dk	0	0	1	1	0	0	2
red5_gr	0	0	1	0	1	0	9.8
red6_red	0	0	1	0	0	1	10

Construct a gene co-expression network
Rationale: make use of interaction patterns among genes
Tools: correlation as a measure of co-expression
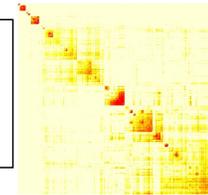
Identify modules
Rationale: module (pathway) based analysis
Tools: hierarchical clustering, Dynamic Tree Cut
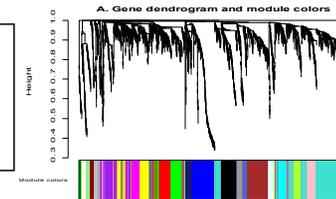
Relate modules to external information
Array Information: clinical data, SNPs, proteomics
Gene Information: ontology, functional enrichment
Rationale: find biologically interesting modules
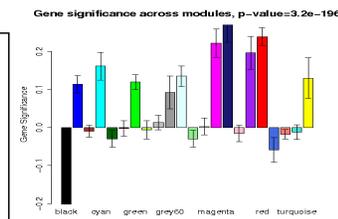
Study module relationships
Rationale: biological data reduction, systems-level view
Tools: Eigengene Networks
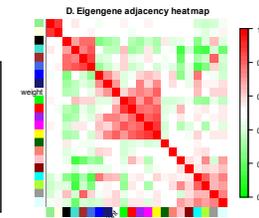
Find the key drivers in *interesting* modules
Rationale: experimental validation, biomarkers
Tools: intramodular connectivity, causality testing
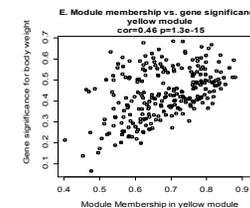
Overview of WGCNA methodology. This flowchart presents a brief overview of the main steps of Weighted Gene Co-expression Network Analysis.
Langfelder and Horvath BMC Bioinformatics 2008 9:559   doi:10.1186/1471-2105-9-559

Step 2: Identifying modules

Dendrogram and gene modules

Clustering dendrogram of top host DEGs, with dissimilarities based on topological overlap shown with assigned module colors (Dynamic tree cut) and upon merging modules whose expression profiles were 75% similar (Merged dynamic).

Cluster Dendrogram
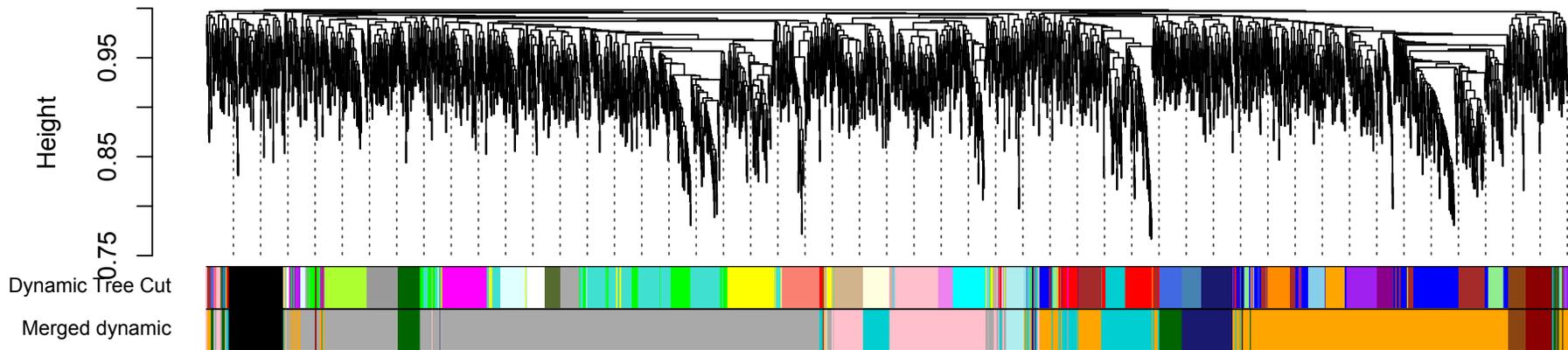
Step 3: Relate Modules to external information

Module-trait relationships
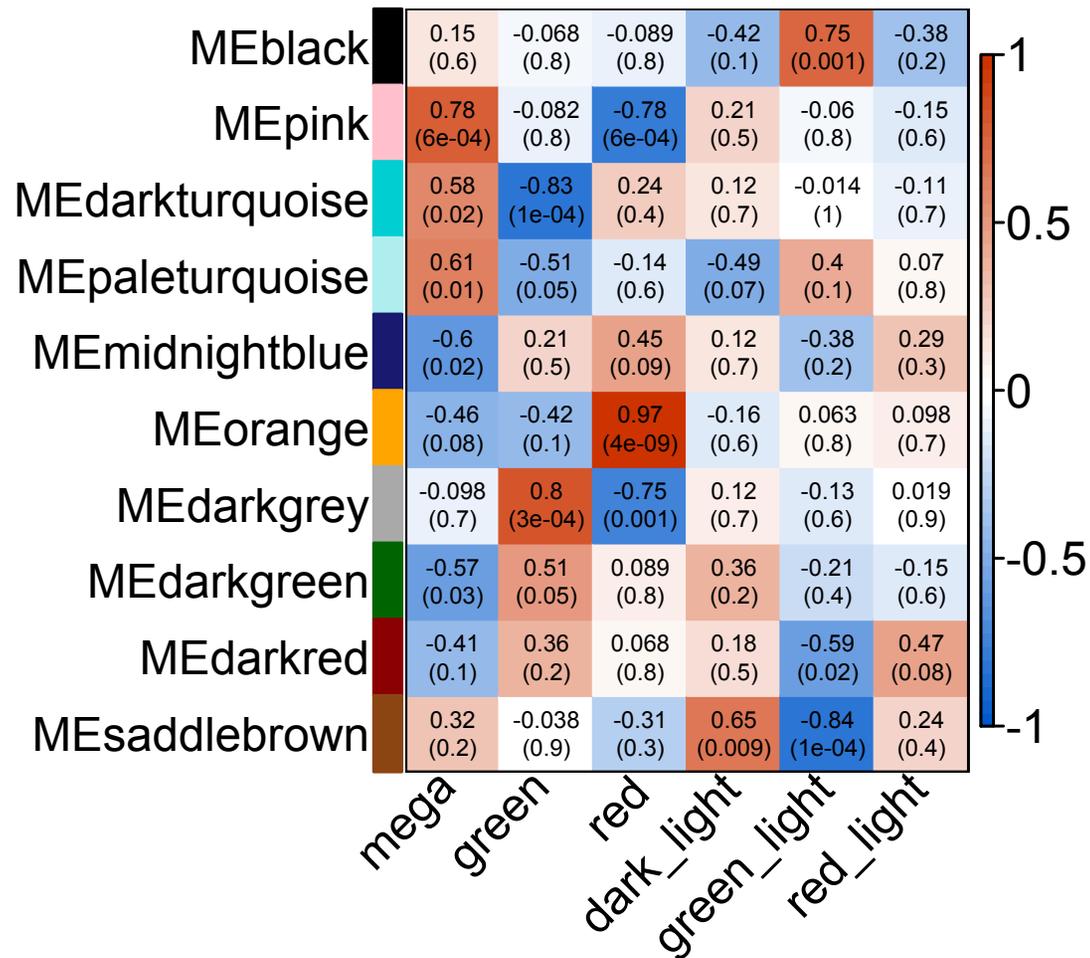
How well do your modules represent your traits?

- Module Eigengene concept
  - Baseline expression in the module
  - "The module eigengene E is defined as the first principal component of a given module. It can be considered a representative of the gene expression profiles in a module."

Module Membership vs. gene significance
cor=0.95, p<1e-200
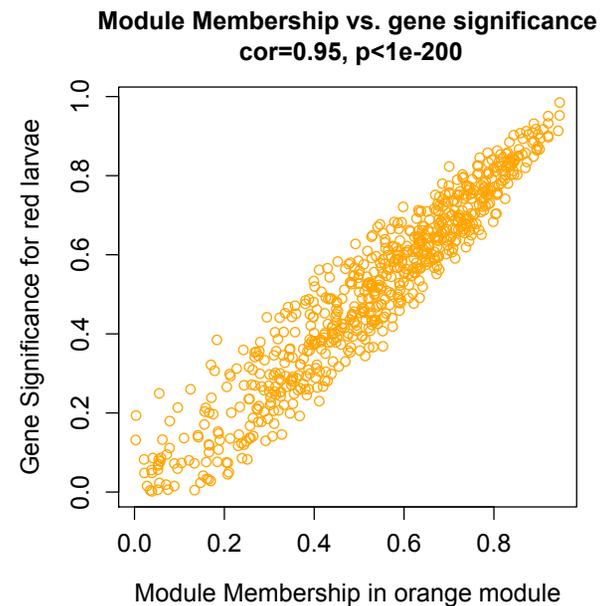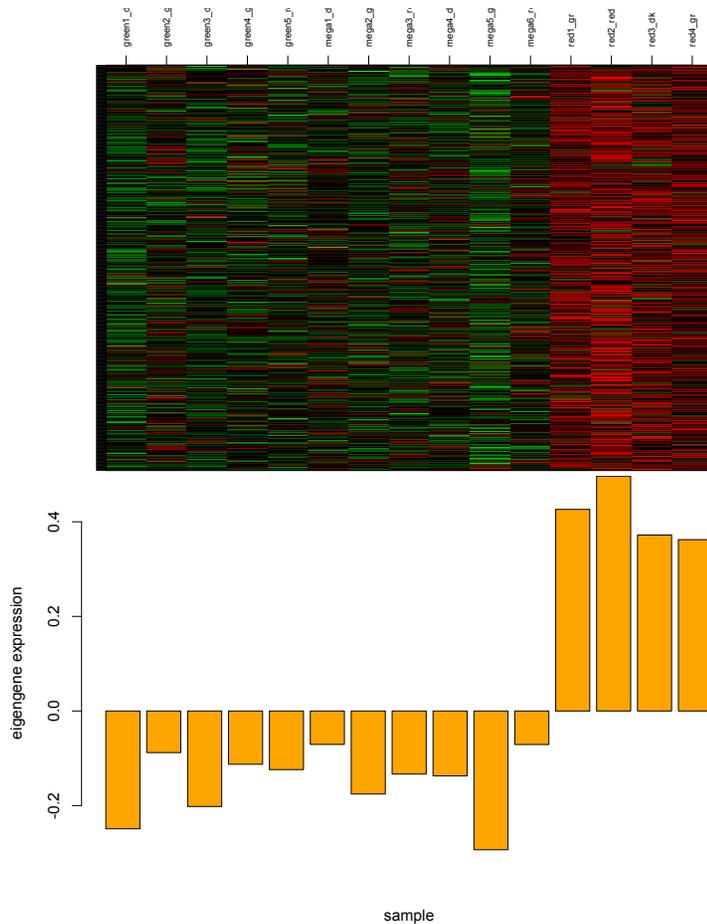
kME concept

- High kME (MM) values = high correlation to the eigengene
- "Highly connected intramodular hub genes tend to have high module membership values to the respective module."
- kME values are useful for selecting the "most important" genes within a module
- "Gene significance of 0 indicates that the gene is not significant with regard to the biological question of interest"

Module Membership vs. gene significance
cor=0.64, p=1.4e-14

Gene Significance for green light
Module Membership in black module

Module Membership vs. gene significance
cor=0.95, p<1e-200

Gene Significance for red larvae
Module Membership in orange module
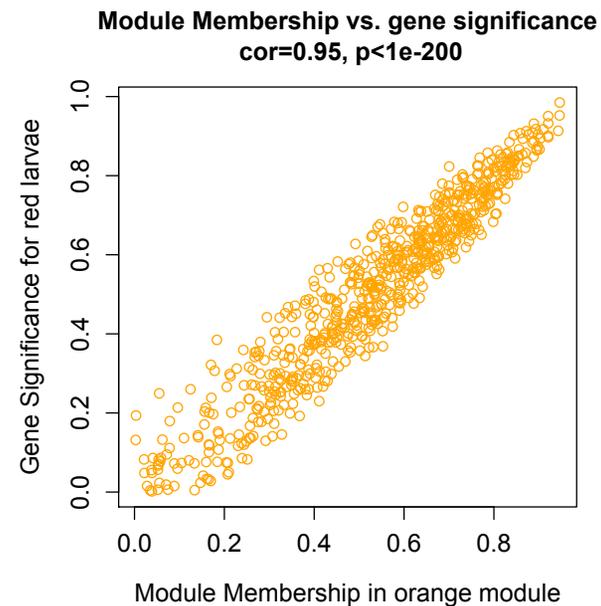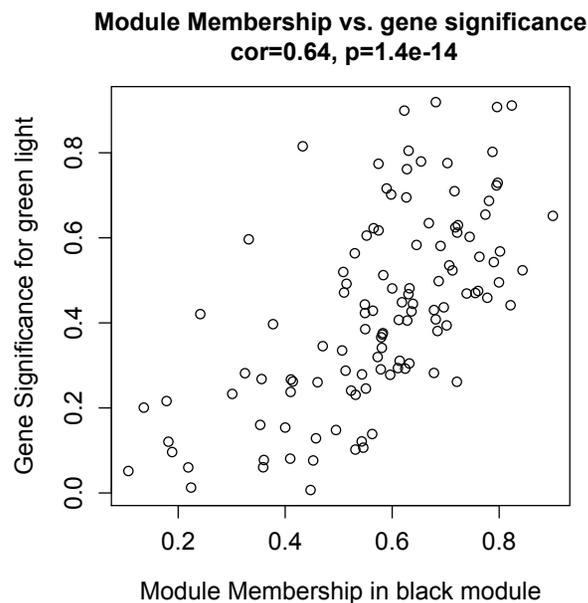
An Example:

- **Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing**
  - Zhigang Xue1*, Kevin Huang2* *et al*
- Analysis of transcriptome dynamics from oocyte to morula in both human and mouse embryos
  - Single-cell RNAseq
- They found that each developmental stage is represented by a small number of functional modules of co-expressed genes
- They also identified conserved key members (or 'hub' genes) of human and mouse networks that likely drive mammalian pre-implantation development

Network analysis of human pre-implantation development.
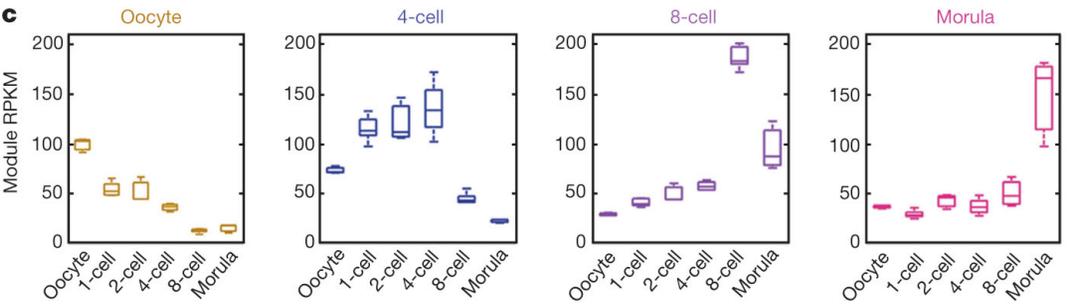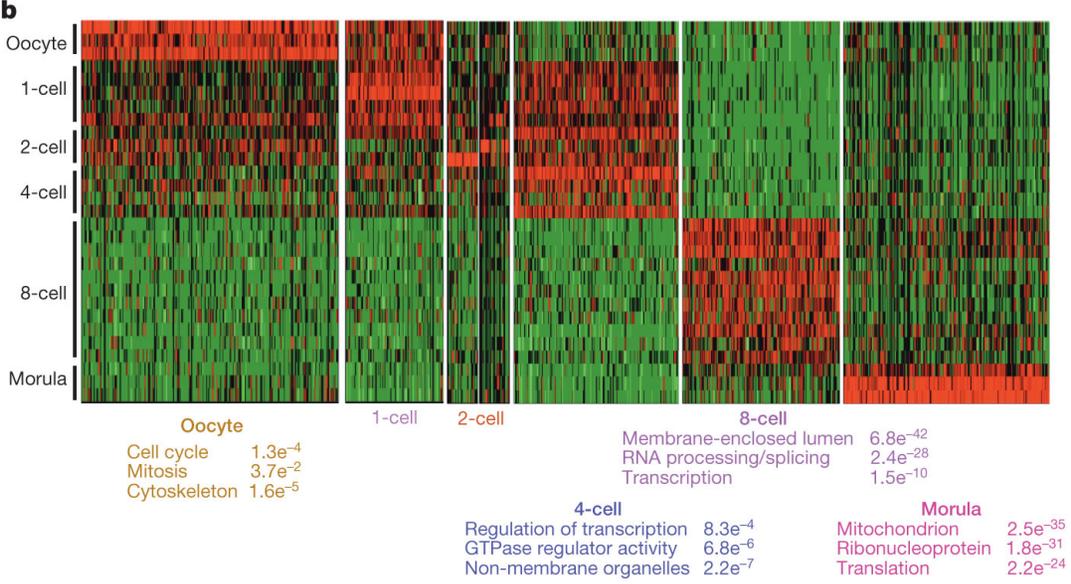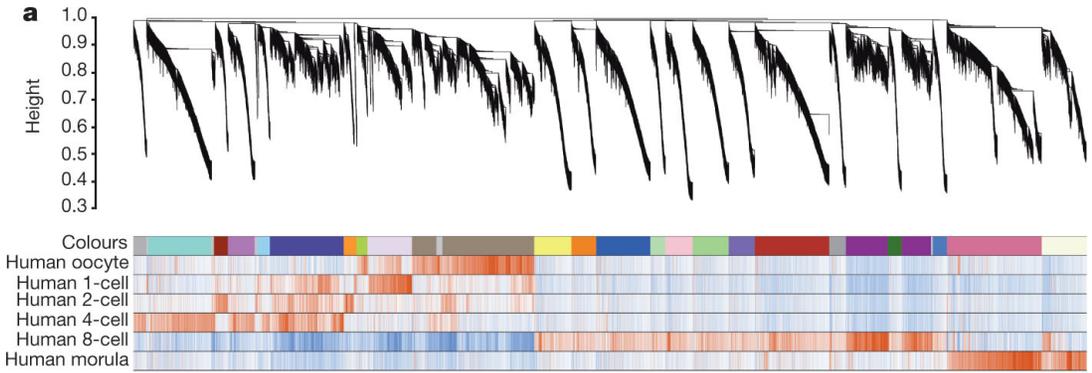
Comparing data sets with module conservation analysis

Experiment A

Experiment B

?
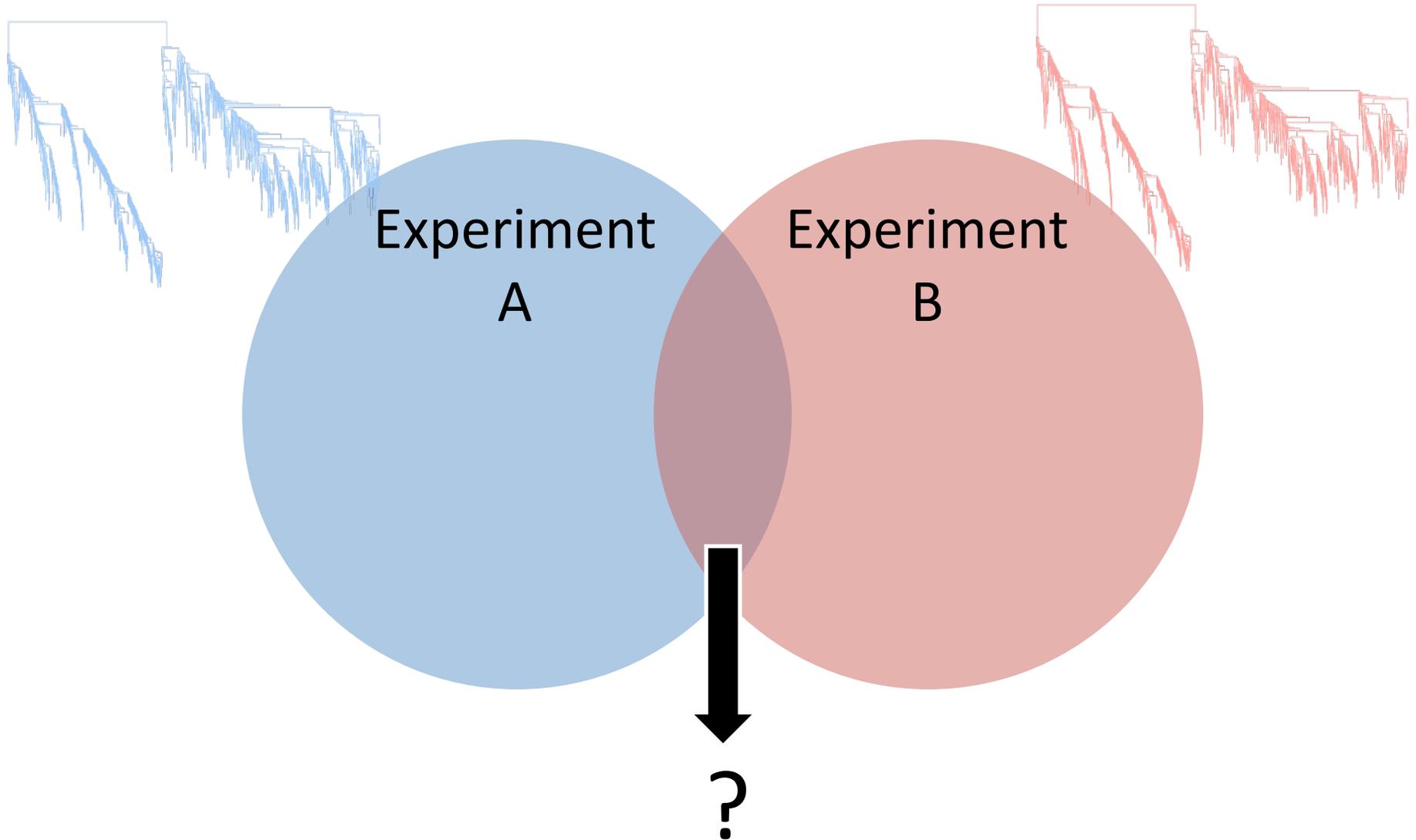
A: Gene expression analysis of *Acropora millepora* larvae challenged with viruses and heat

Photo Credit: Andrew Negri, Australian Institute of Marine Science
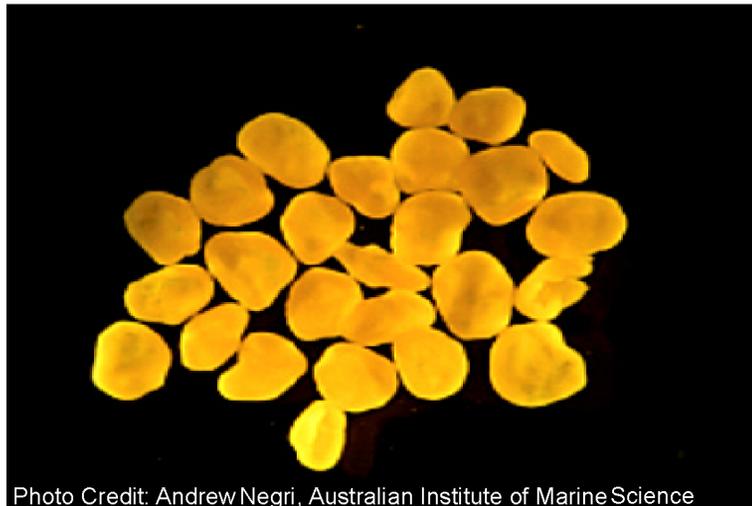
Mix gametes

Add VLPs at two-cell stage
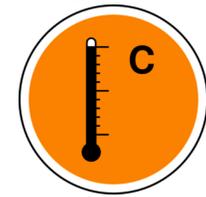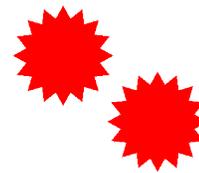
Quantify development through the prawn chip stage

B: Gene expression analysis of adult *Acropora hyacinthus* affected by White Syndrome

Diseased
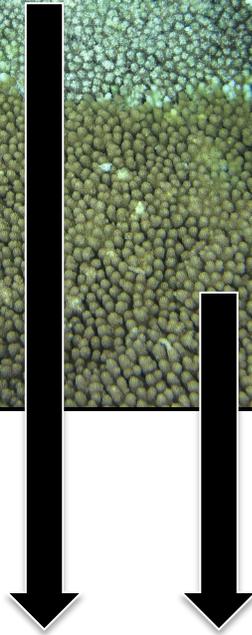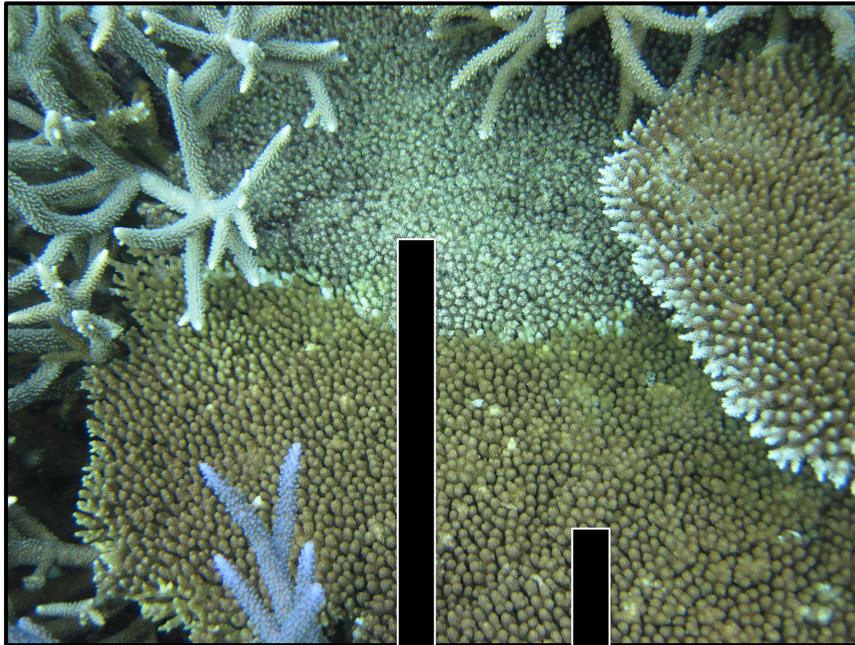
Healthy
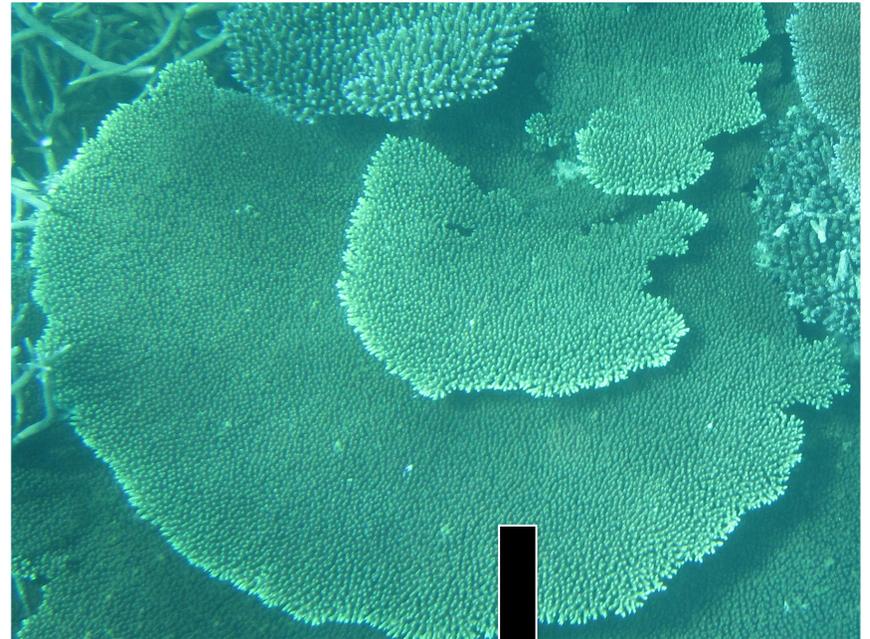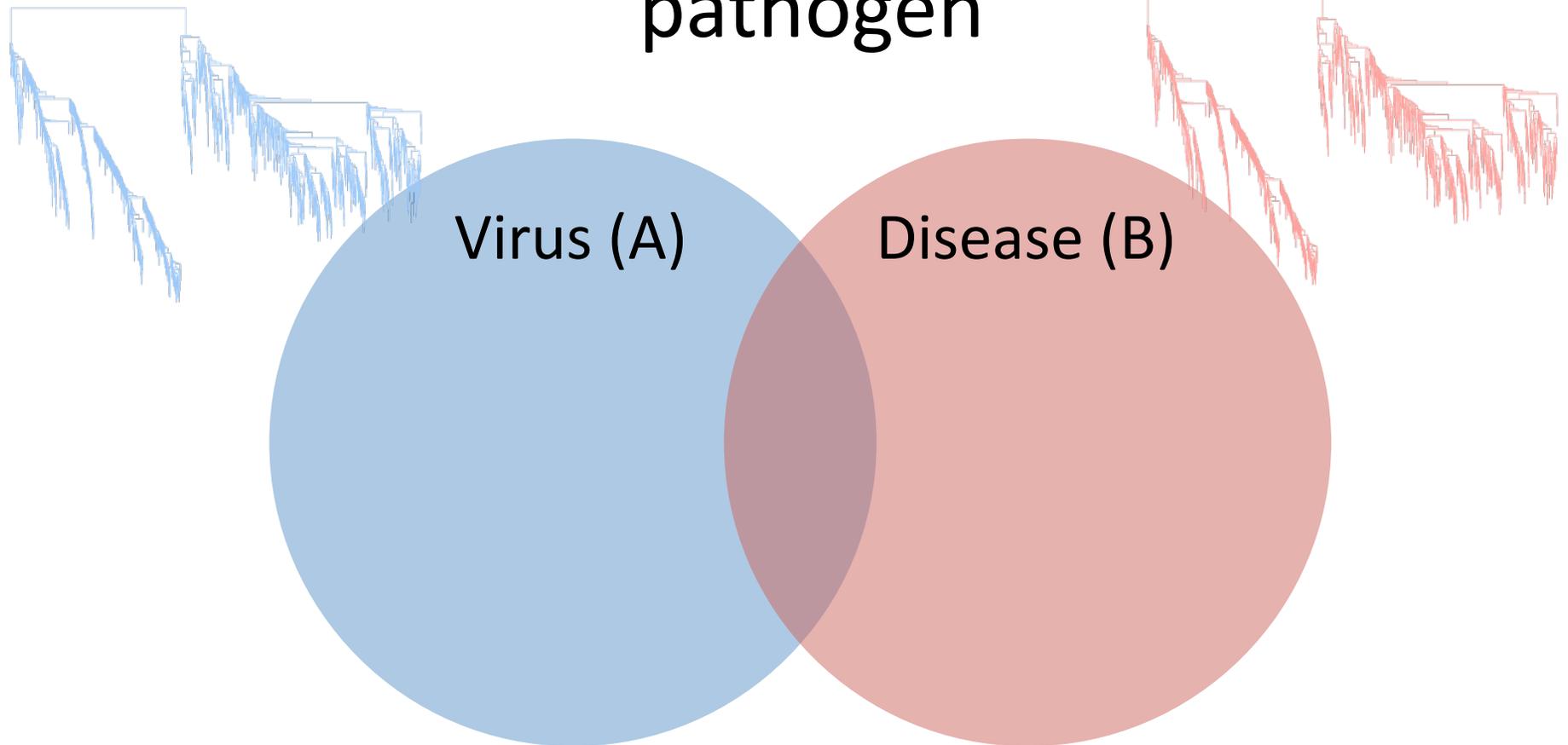
Hypothesis

Common networks may represent vital evolutionarily conserved components of the coral stress response
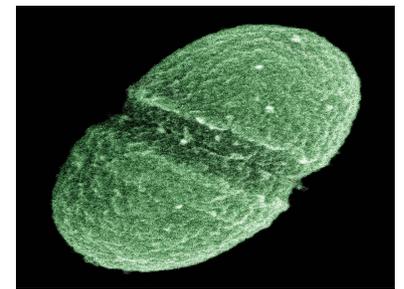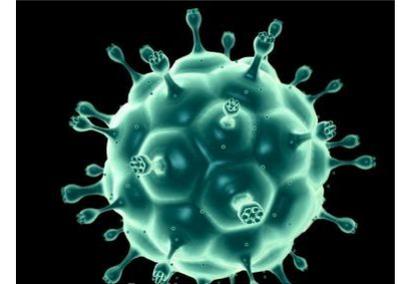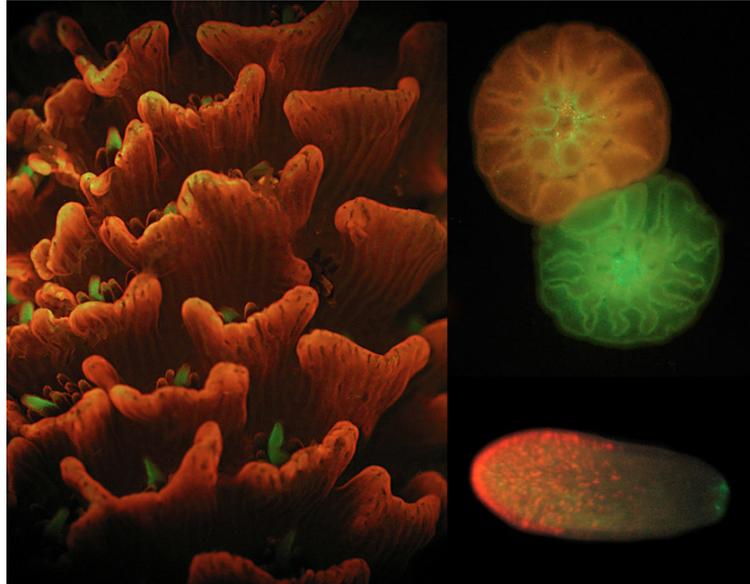
# Identify gene networks in both data sets

Gene clustering on TOM-based dissimilarity (Virus)

Gene clustering on TOM-based dissimilarity (Disease)
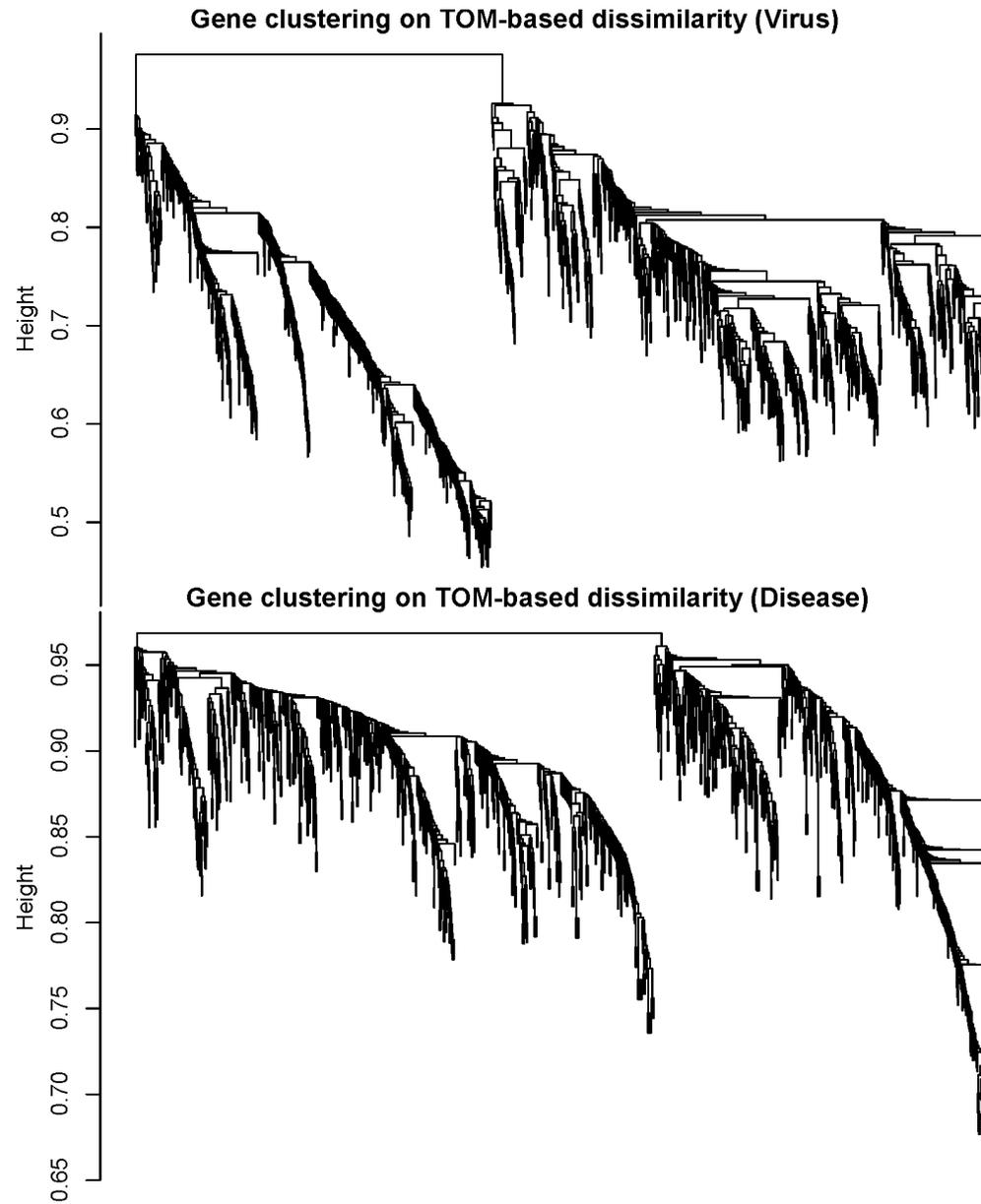
Assign modules to the virus data set
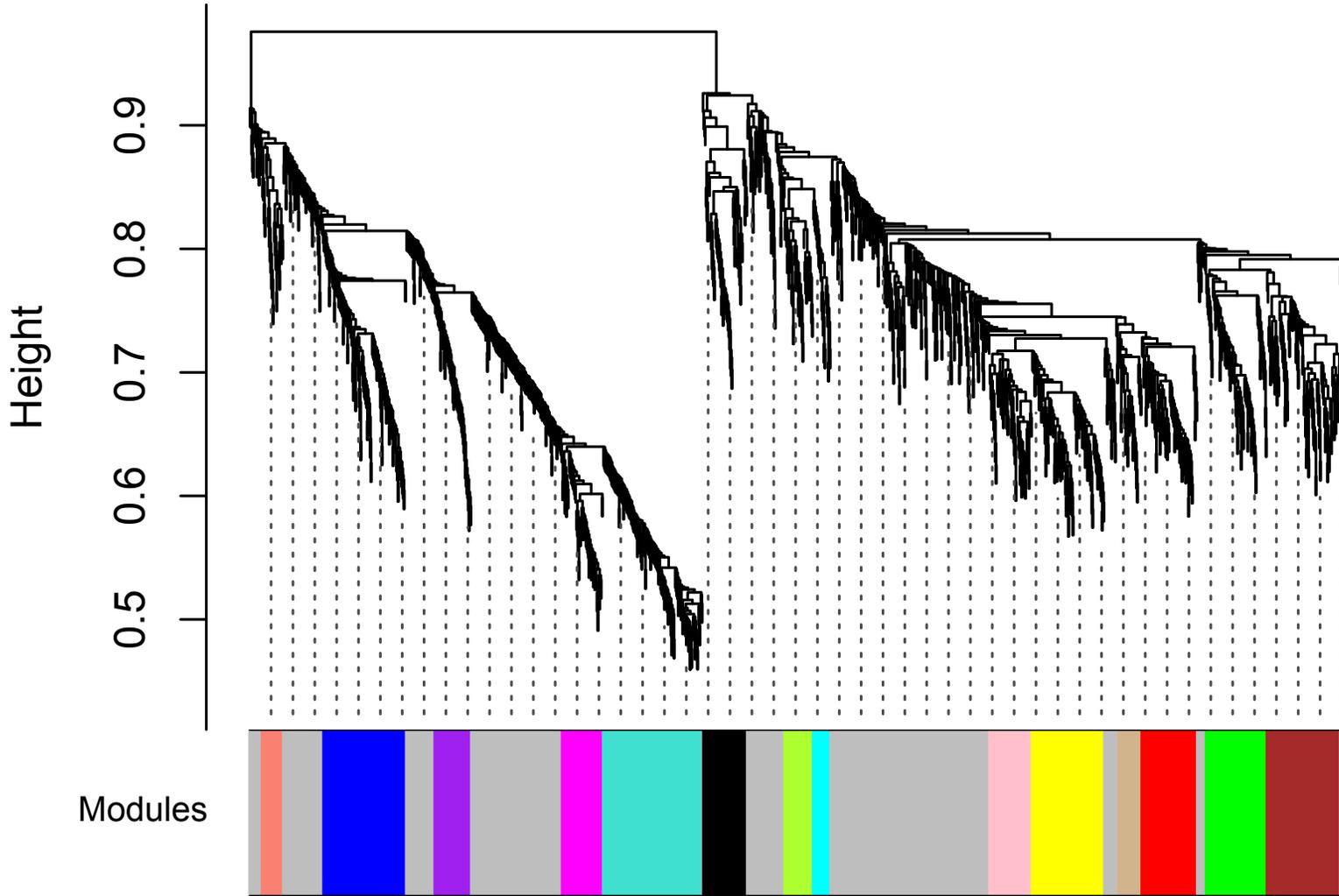
Impose the virus modules on the disease data set
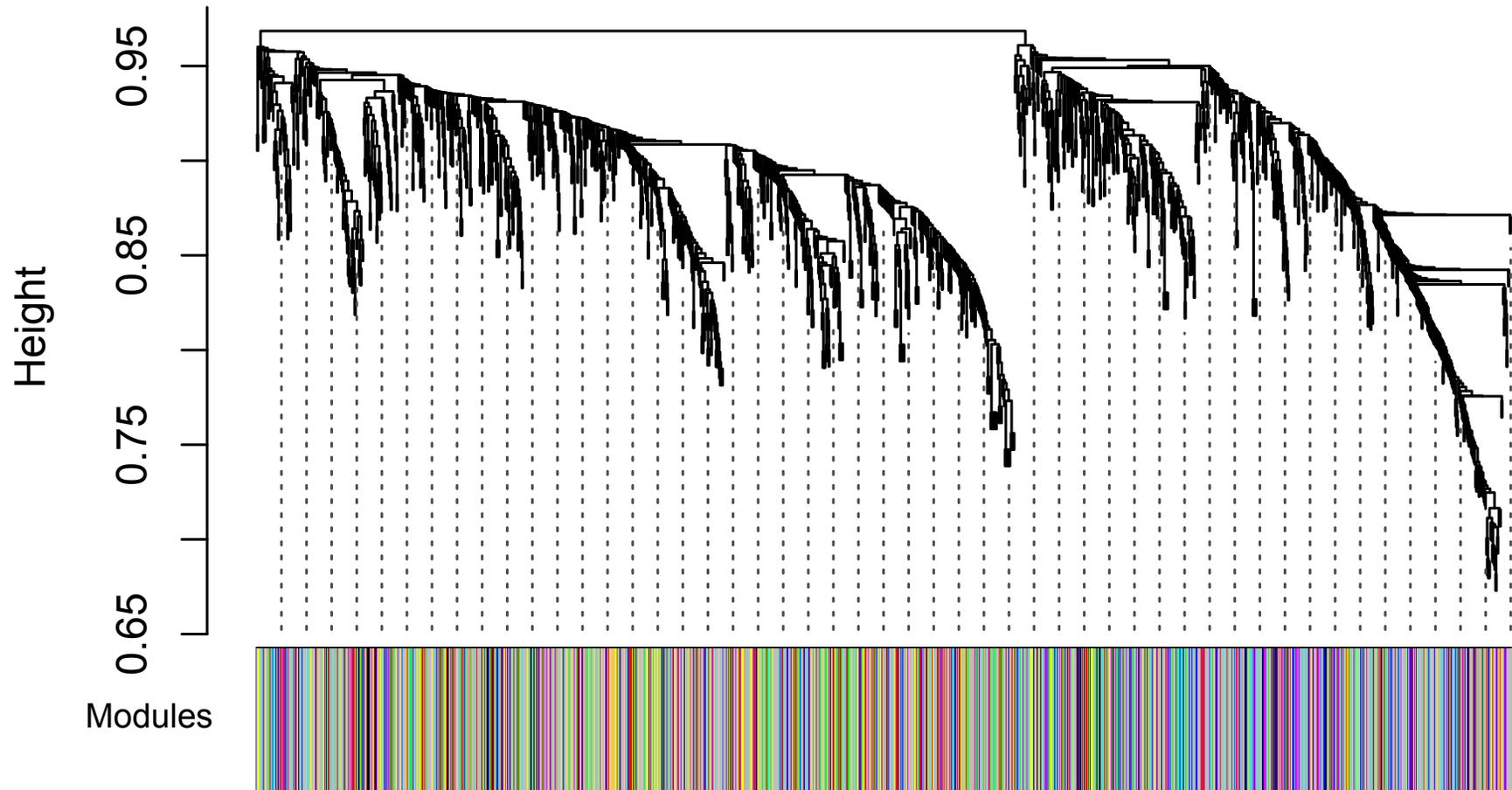
# Three modules were significantly preserved between the data sets
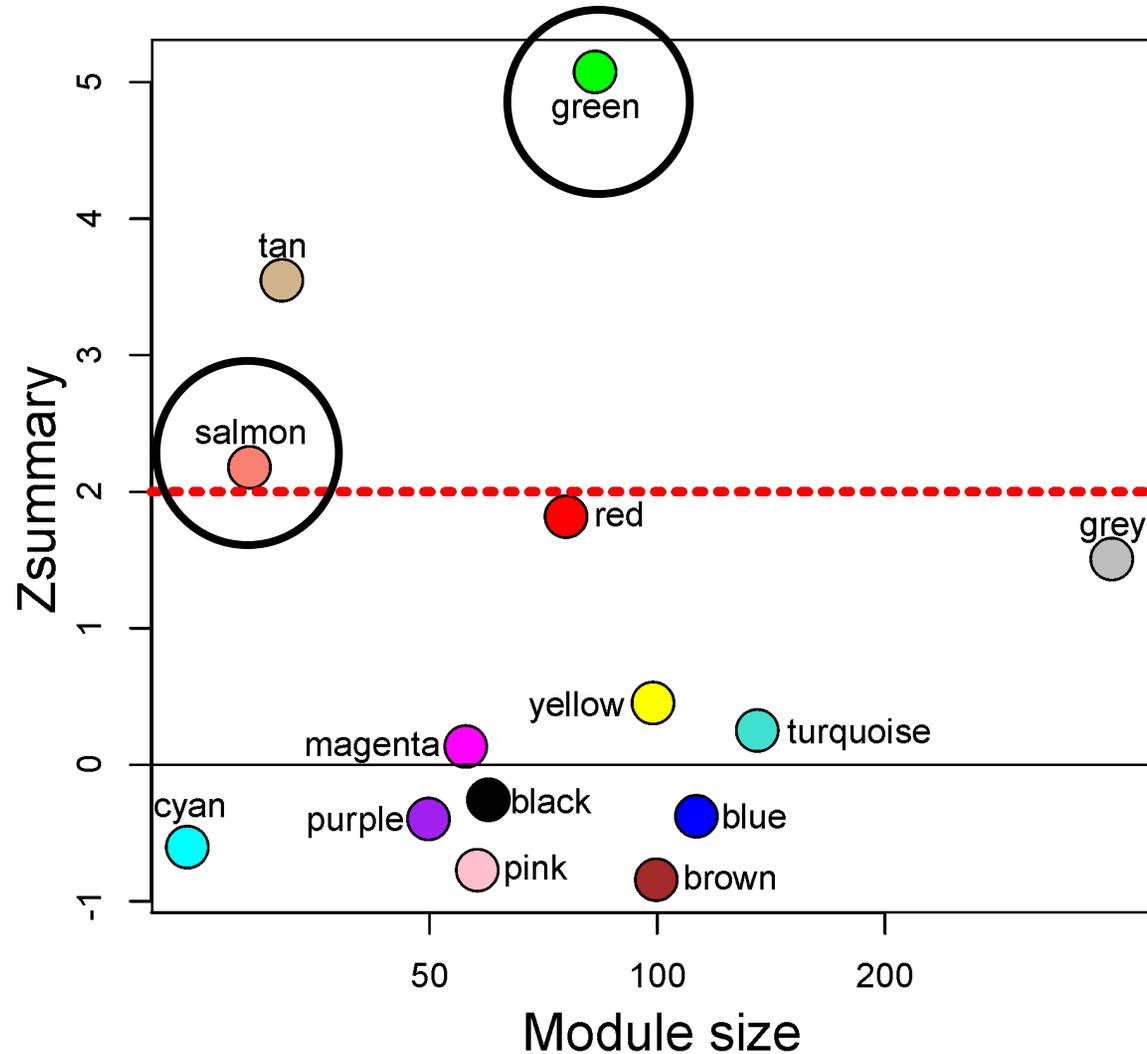
The green module correlates with heat stress
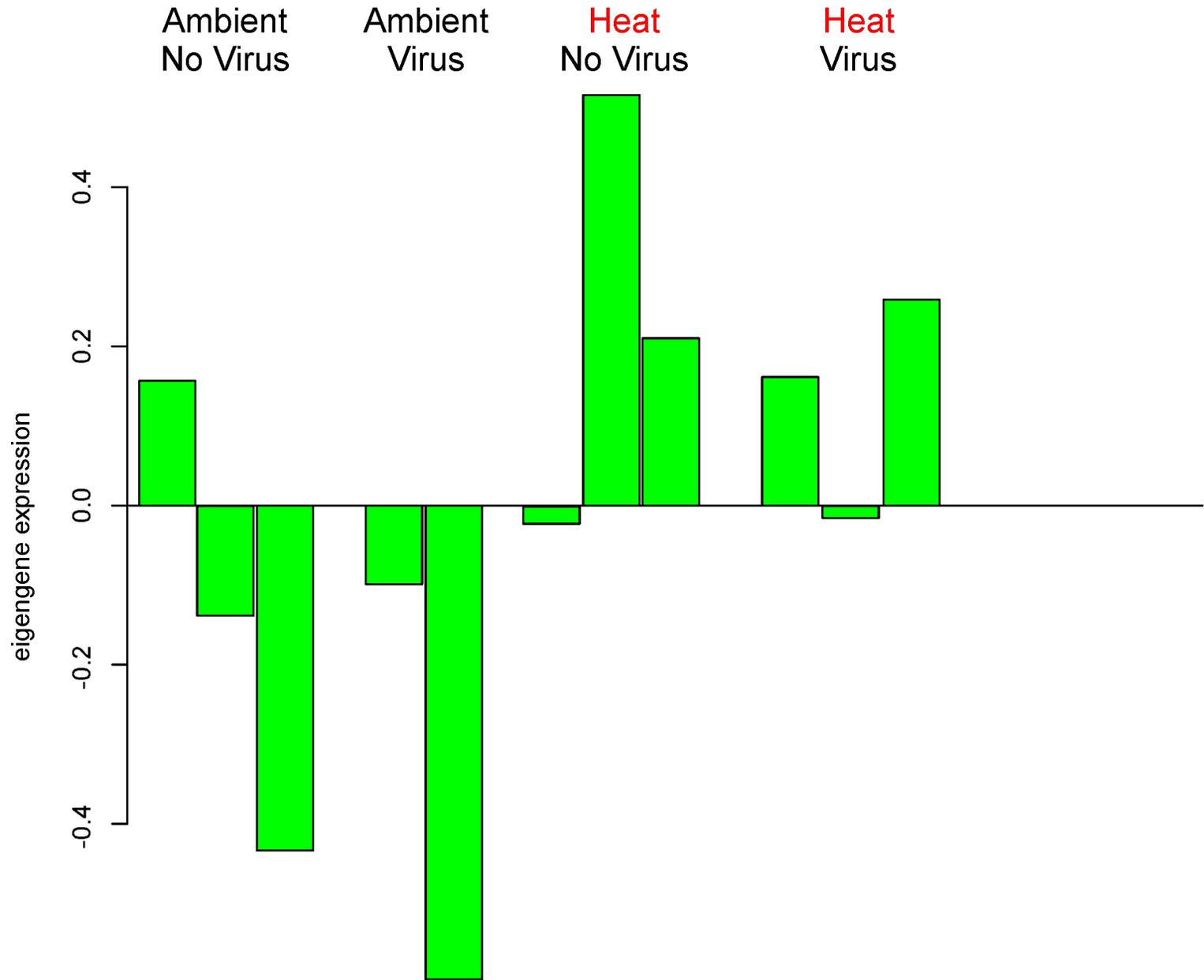
Functions include regulation of cell division, transcription, and protein modification

Annotation	Putative Function
voltage-gated sodium channel alpha, invertebrate	action potential regulation
ATP-binding cassette, subfamily A (ABC1), member 2	iron homeostasis, urate metabolism
WW domain containing adaptor with coiled-coil	transcription regulation
Rho guanine nucleotide exchange factor 2	cell cycle regulation
transitional endoplasmic reticulum ATPase	DNA damage response, protein modification
polycystin 1	cell projection regulation
nuclear factor erythroid 2-related factor 1/3	transcription regulation
alpha-1,3(6)-mannosylglycoprotein beta-1,6-N-acetyl-glucosaminyltransferase	post-translational protein modification
ribosomal RNA-processing protein 1	rRNA processing
mannosyl-oligosaccharide glucosidase	post-translational protein modification

The salmon module correlates with <u>viral challenge</u>
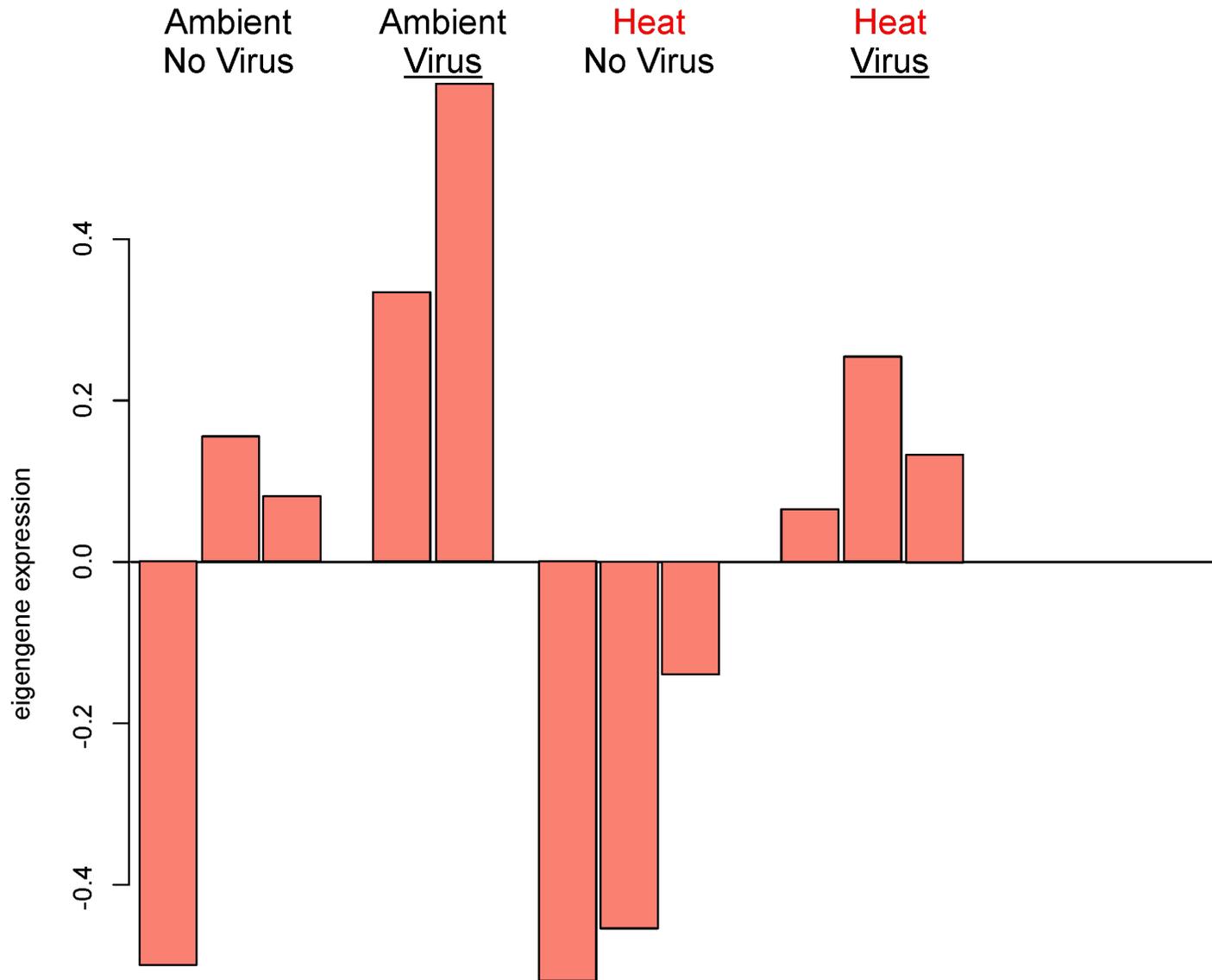
Functions include members of the innate immune response

Annotation	Putative Function
purinergic receptor P2X, ligand-gated ion channel 7	inflammatory response
peptide deformylase	co-translational processing of bacterial proteins
large subunit ribosomal protein L11e	translation
sequestosome 1	apoptosis, autophagy
deoxyhypusine monooxygenase	post-translational protein modification (eIF-5A)
transglutaminase 1	extracellular peptide cross-linking
ubiquinol-cytochrome c reductase subunit 6	oxidative phosphorylation
cofilin	cytokinesis (response to virus)
U3 small nucleolar RNA-associated protein 16	rRNA processing
calbindin D28	cellular response to organic substance