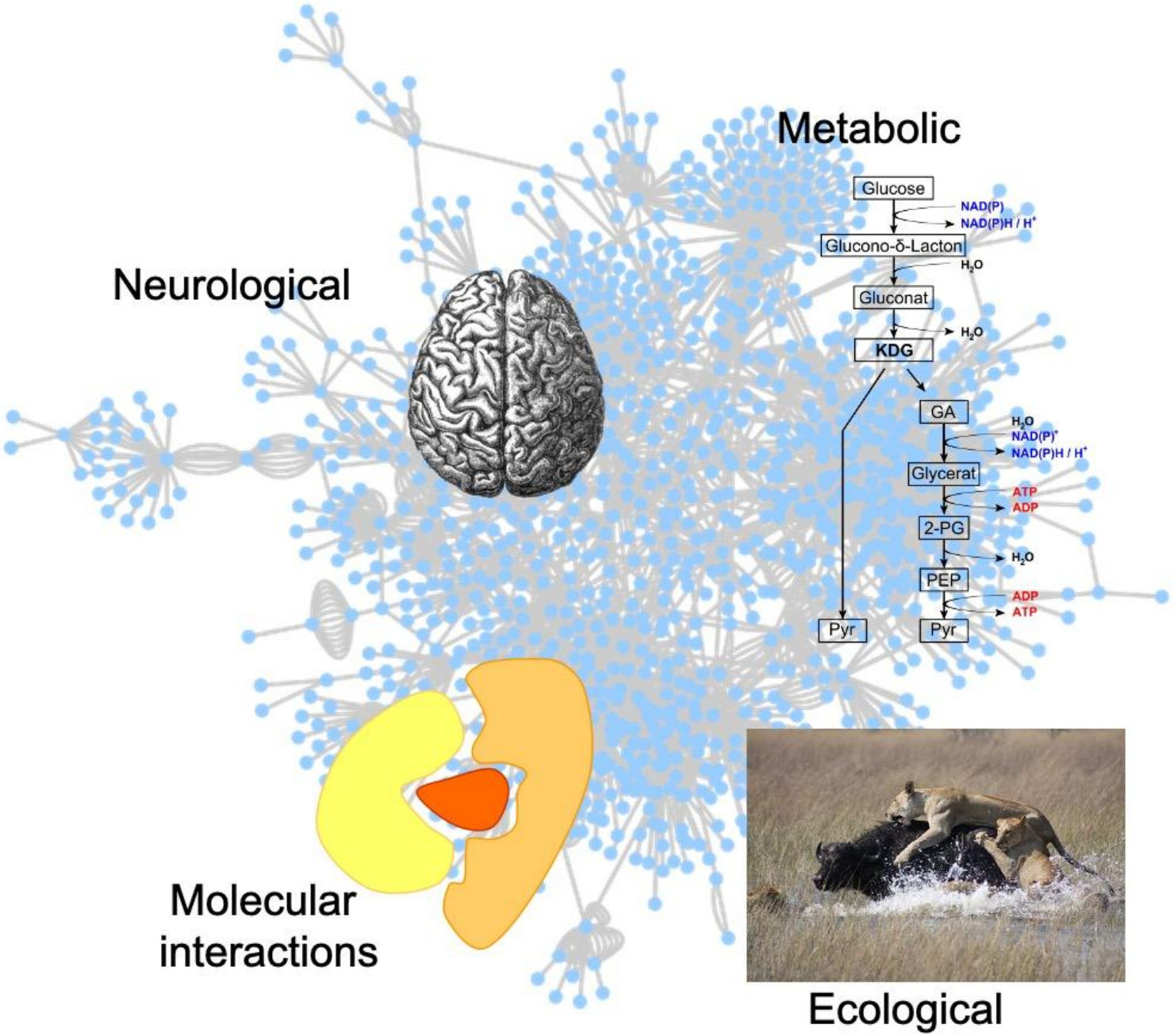
Gene Co–expression Network

Jorge A. Tzec-Interián

June 17, 2024

Network biology

Biological systems are often represented as networks which are complex sets of **binary interactions** or **relations** between different entities.

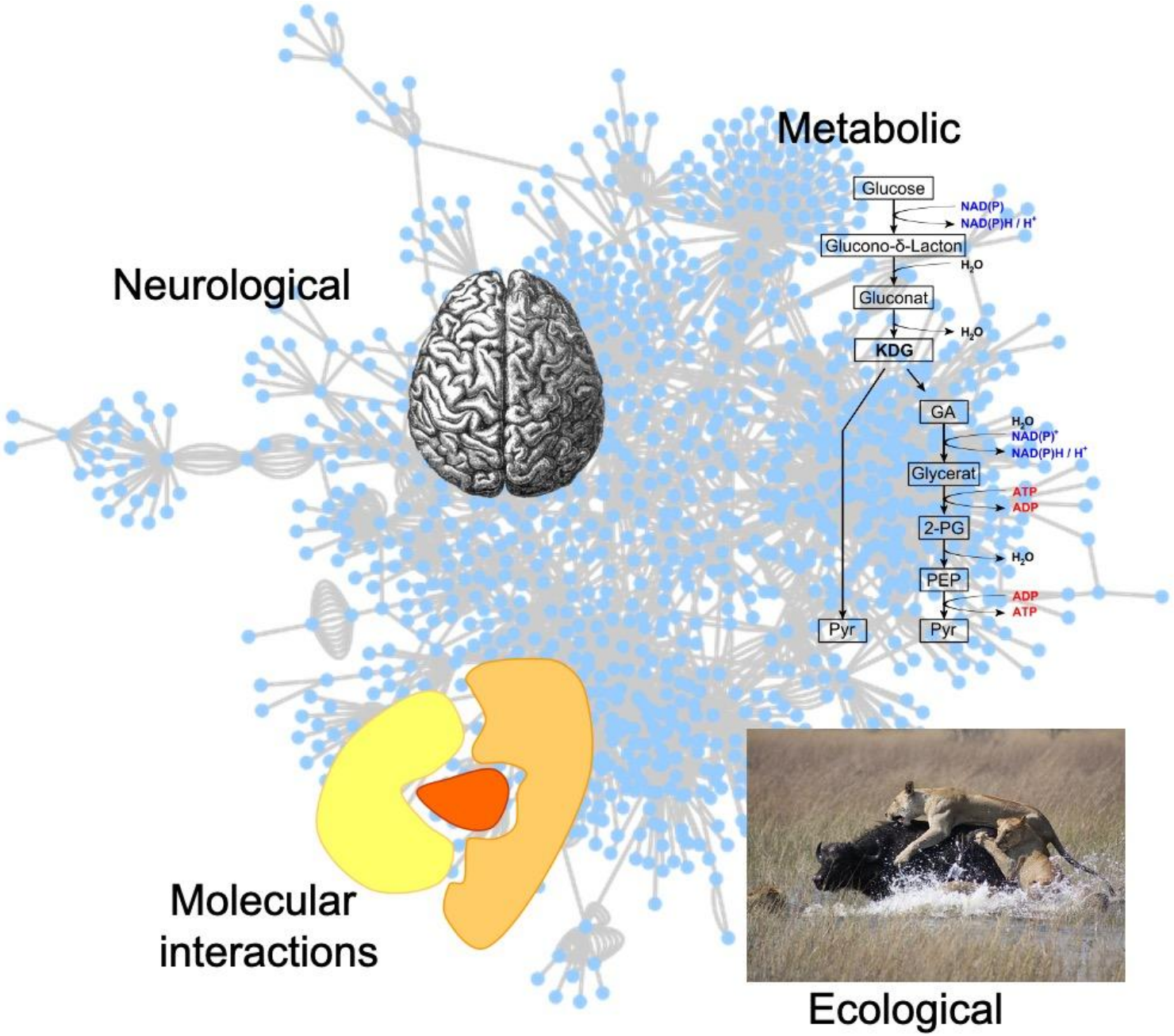


https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/

Different levels of analysis: molecular to the ecosystem level

Approaches: ecological, neurological, metabolic or molecular interaction networks.

Network biology: the representation and analysis of biological systems using tools derived from **graph theory**.



Jorge A. Tzec-Interián | Gene Co-expression Network

https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/

Graph theory and the idea of topology

- Leonard Euler, Swiss mathematician

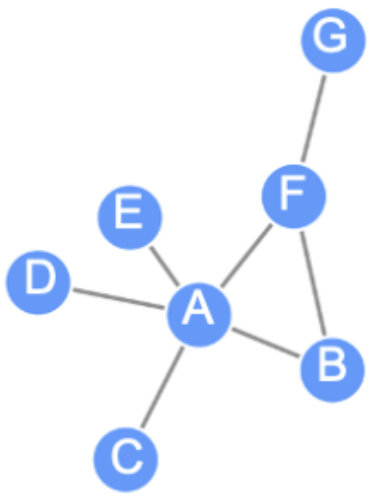
*"[...] the study of graphs, mathematical structures used to model **pairwise relations between objects**. A graph in this context is made up of vertices, nodes, or points which are connected by edges, arcs, or lines".*

https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/

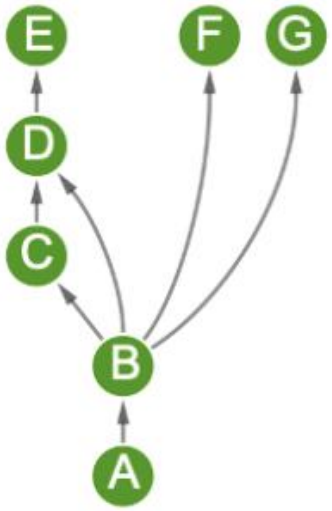
The **nodes** represent different entities (e.g. proteins or genes in biological networks), and **edges** convey information about the links between the nodes

Types of network edges

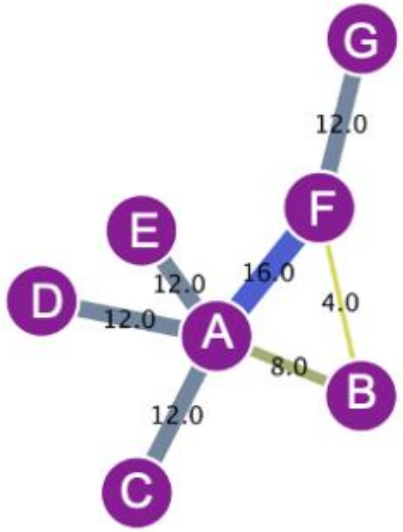
Undirected



Directed



Weighted

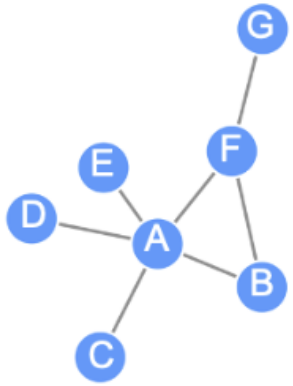


Jorge A. Tzec-Interián | Gene Co-expression Network

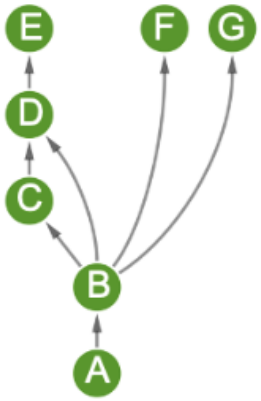
https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/

Adjacency matrices and the edges

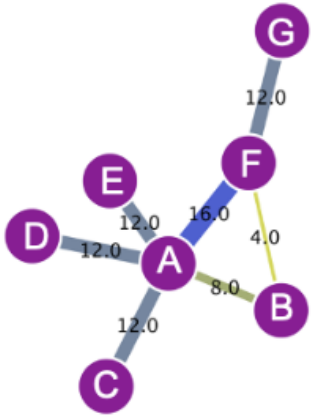
Undirected



Directed



Weighted



Matrices

	A	B	C	D	E	F	G	Degree
A	0	1	1	1	1	1	0	5
B	1	0	0	0	0	1	0	2
C	1	0	0	0	0	0	0	1
D	1	0	0	0	0	0	0	1
E	1	0	0	0	0	0	0	1
F	1	1	0	0	0	0	1	3
G	0	0	0	0	0	1	0	1

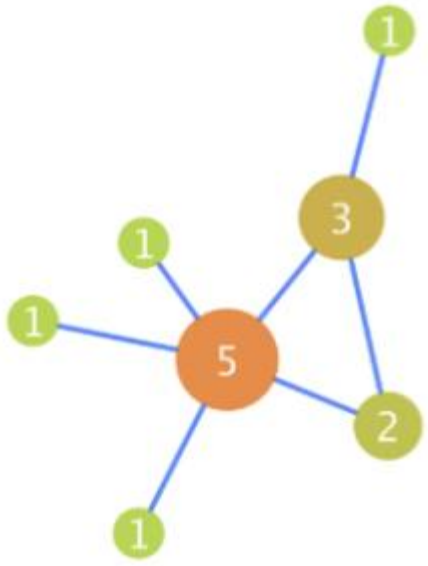
	A	B	C	D	E	F	G	Out-degree
A	0	1	0	0	0	0	0	1
B	0	0	1	1	0	1	1	4
C	0	0	0	1	0	0	0	1
D	0	0	0	0	1	0	0	1
E	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0

	A	B	C	D	E	F	G	Degree
A	0	8	12	12	12	16	12	72
B	8	0	0	0	0	4	0	12
C	12	0	0	0	0	0	0	12
D	12	0	0	0	0	0	0	12
E	12	0	0	0	0	0	0	12
F	16	4	0	0	0	0	12	32
G	12	0	0	0	0	12	0	24

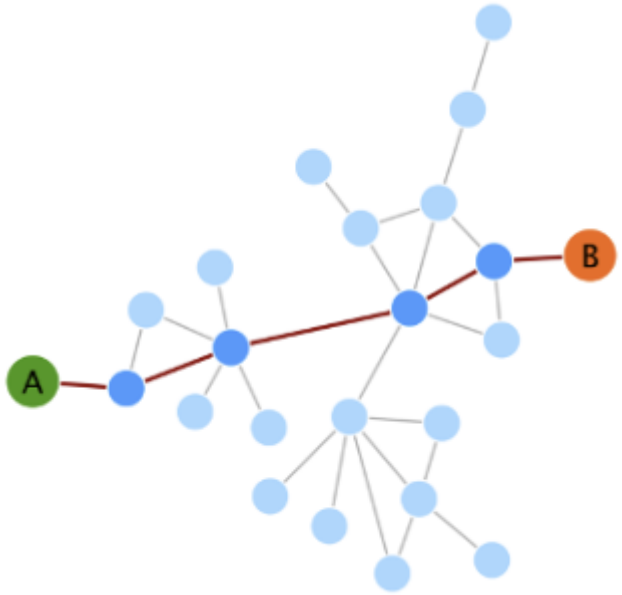
Jorge A. Tzec-Interián | Gene Co-expression Network

Types of network edges

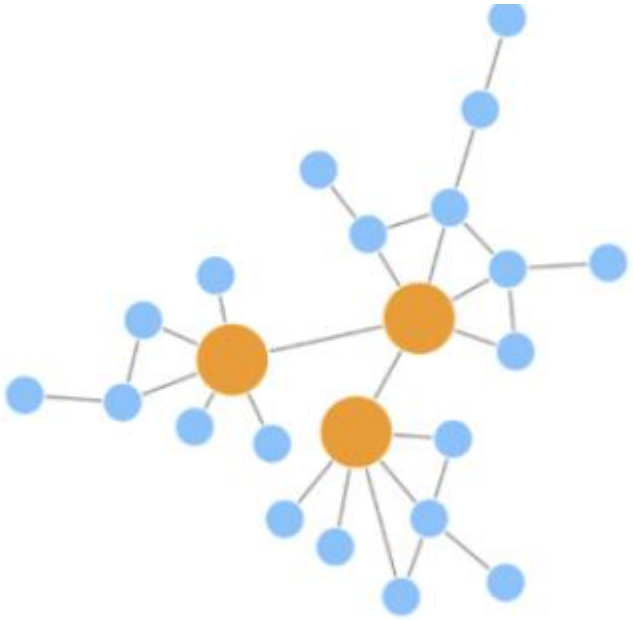
The degree of a network



Shortest paths



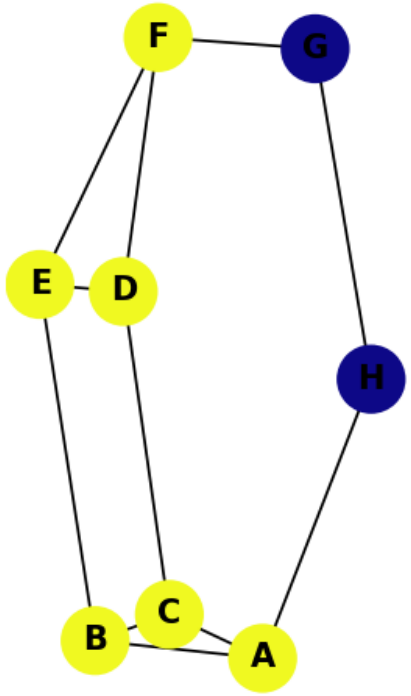
Scale-free networks



https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/

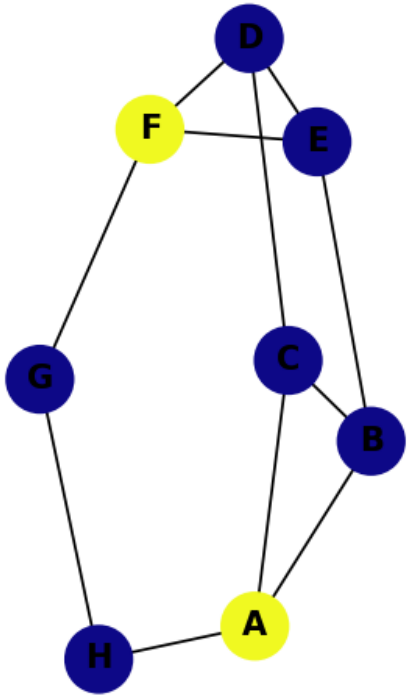
Graph theory: graph types and edge properties

Degree Centrality



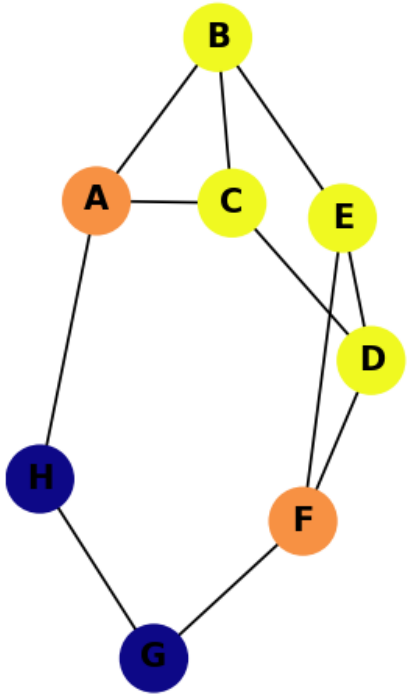
Measures the importance of a node based on the number of direct connections.

Betweenness Centrality



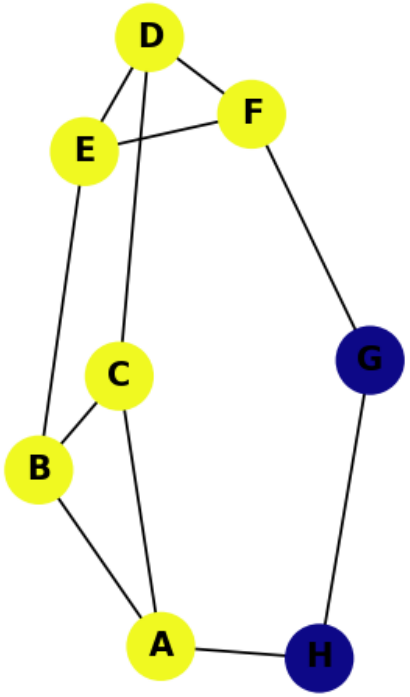
Quantifies the number of times a node acts as a bridge along the shortest path between two other nodes.

Eigenvector Centrality



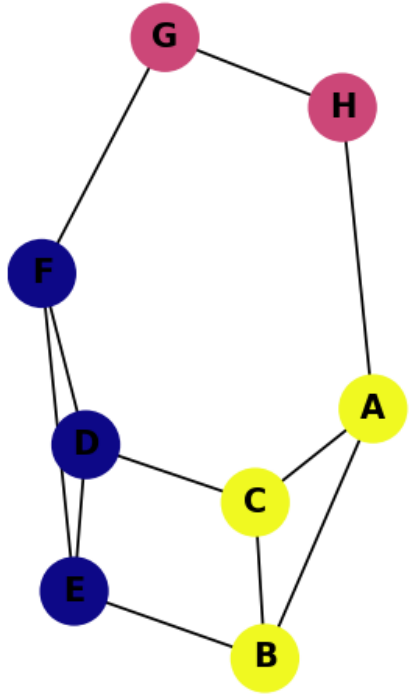
Determines the influence of a node based on the importance of its neighbors

Clustering Coefficient



Measures how likely it is that two neighbors of a node are also connected to each other.

Modularity

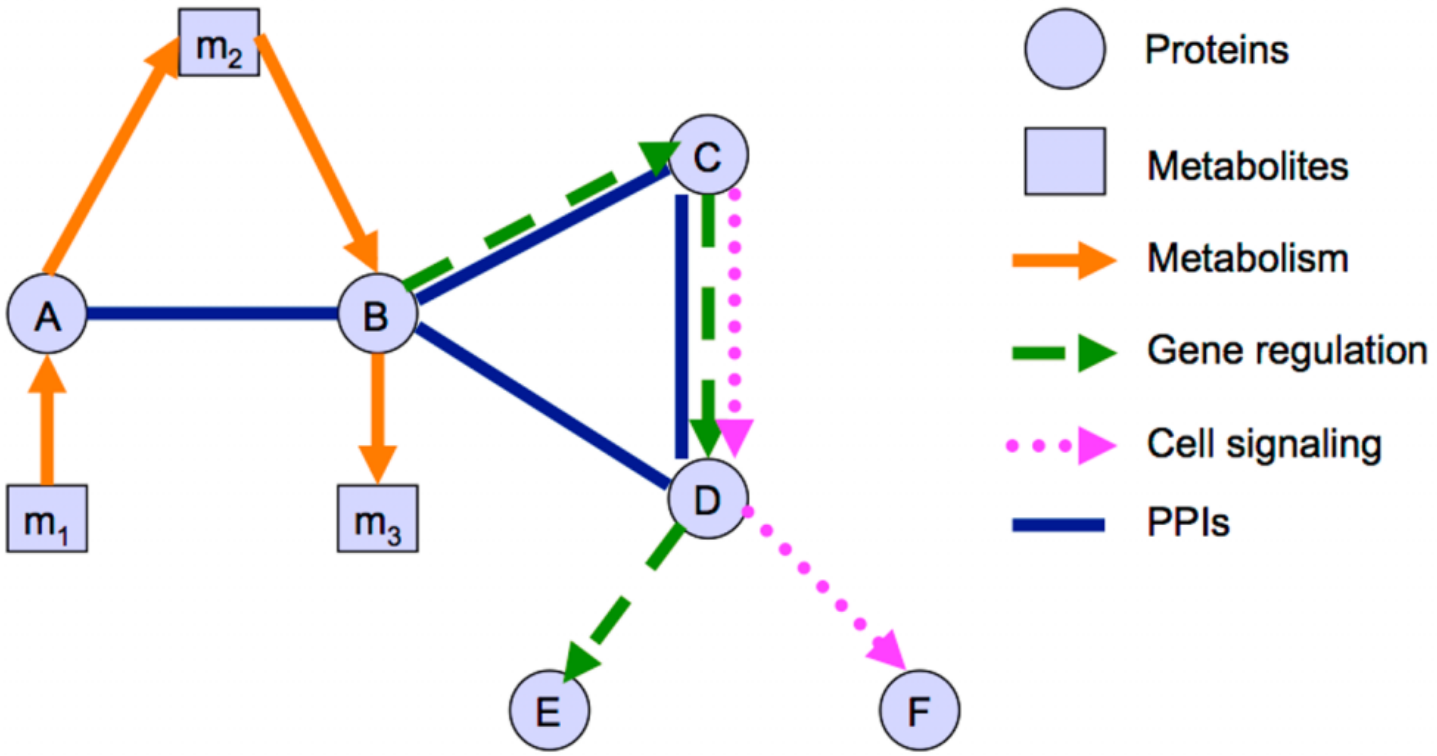


Evaluates the strength of division of a network into clusters or communities.

Jorge A. Tzec-Interián | Gene Co-expression Network

Some of the most common types of biological networks are:

- Protein-protein interaction networks
- Metabolic networks
- Genetic interaction networks
- Gene / transcriptional regulatory networks
- Gene co-expression network
- Cell signaling networks



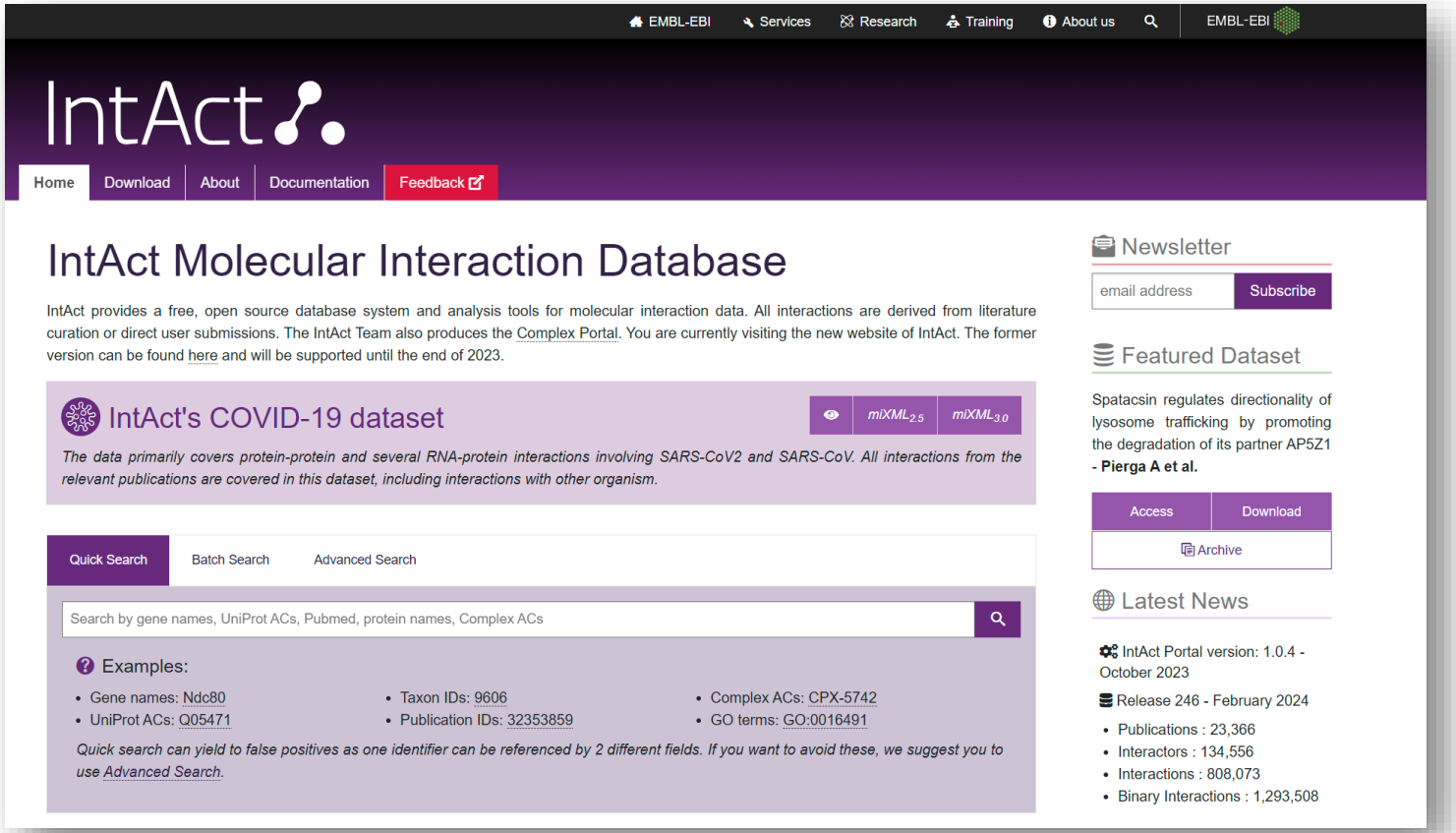
Jorge A. Tzec-Interián | Gene Co-expression Network

https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/

Network Type	Associated Omics	Key Methods
Protein-protein interaction networks	Proteomics	Mass spectrometry, Yeast two-hybrid, Co-immunoprecipitation
Metabolic networks	Metabolomics	LC-MS, GC-MS, NMR spectroscopy
Genetic interaction networks	Genomics	CRISPR screens, GWAS, Whole-genome sequencing
Gene/transcriptional regulatory networks	Epigenomics & Transcriptomics	ChIP-seq, ATAC-seq, RNA-seq
Cell signaling networks	Phosphoproteomics & Proteomics	Mass spectrometry, Kinase activity assays
Gene co-expression networks	Transcriptomics	RNA-seq, Microarrays

The interactome is the totality of Protein-protein interactions (PPIs) that happen in a cell, an organism or a specific biological context.

IntAct provides a free, open source database system and analysis tools for molecular interaction data. All interactions are derived from literature curation or direct user submissions.



Orchard *et al*. 2014; https://www.ebi.ac.uk/intact/search?query=chlamydomonas

The interactome is the totality of Protein-protein interactions (PPIs) that happen in a cell, an organism or a specific biological context.

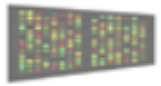
STRING is a database of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations; they stem from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from other (primary) databases.

Data Sources

Interactions in STRING are derived from five main sources:

Genomic Context Predictions

High-throughput Lab Experiments



(Conserved) Co-Expression

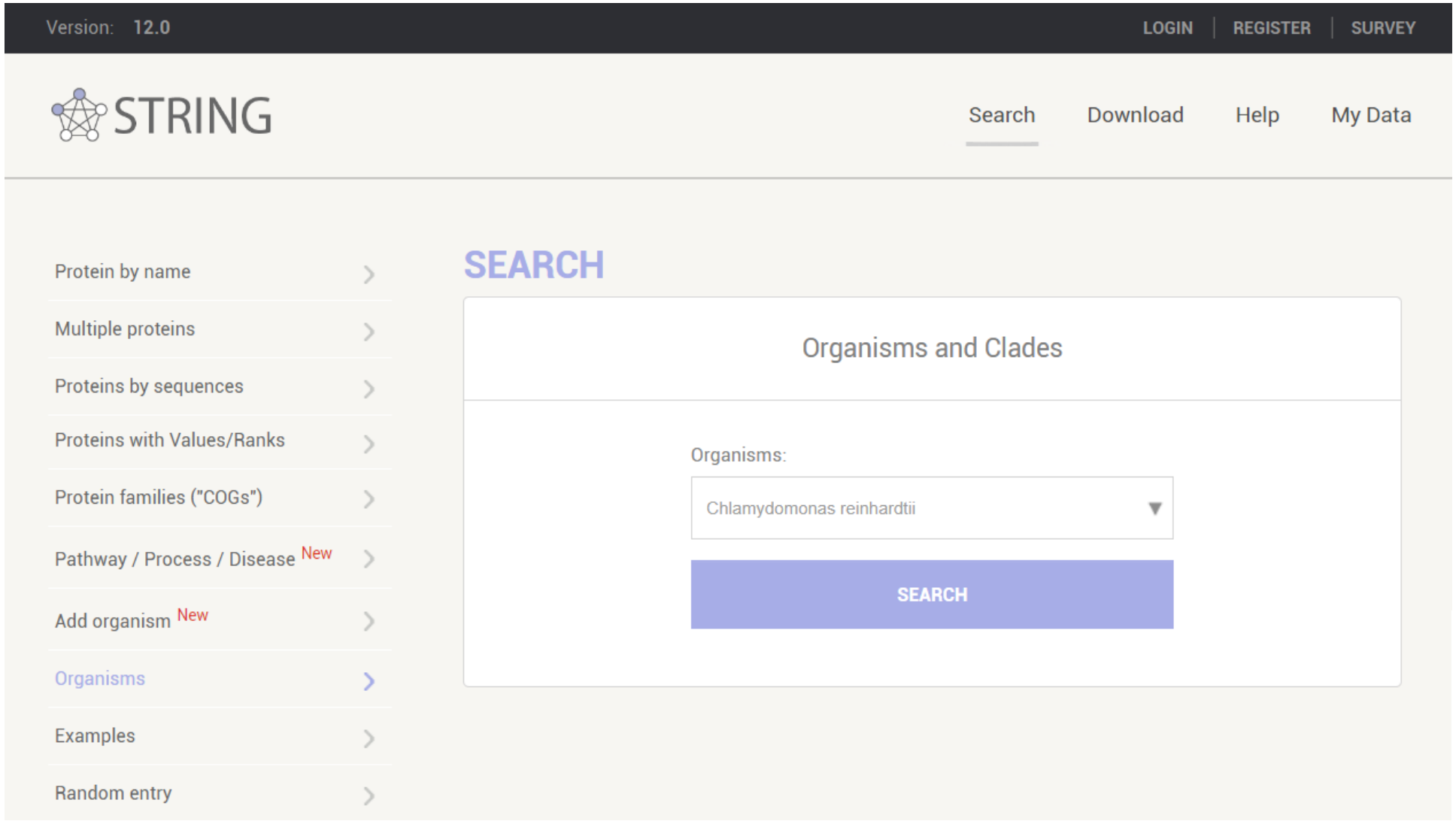
Automated Textmining

Previous Knowledge in Databases

Coverage

The STRING database currently covers 59'309'604 proteins from 12'535 organisms.

Szklarczyk et al. 2023



Cre01.g053000

Jorge A. Tzec-Interián | Gene Co-expression Network

Szklarczyk et al. 2023

Network Type	Associated Omics	Key Methods
Protein-protein interaction networks	Proteomics	Mass spectrometry, Yeast two-hybrid, Co-immunoprecipitation
Metabolic networks	Metabolomics	LC-MS, GC-MS, NMR spectroscopy
Genetic interaction networks	Genomics	CRISPR screens, GWAS, Whole-genome sequencing
Gene/transcriptional regulatory networks	Epigenomics & Transcriptomics	ChIP-seq, ATAC-seq, RNA-seq
Cell signaling networks	Phosphoproteomics & Proteomics	Mass spectrometry, Kinase activity assays
Gene co-expression networks	Transcriptomics	RNA-seq, Microarrays

Transcriptomics: RNAseq

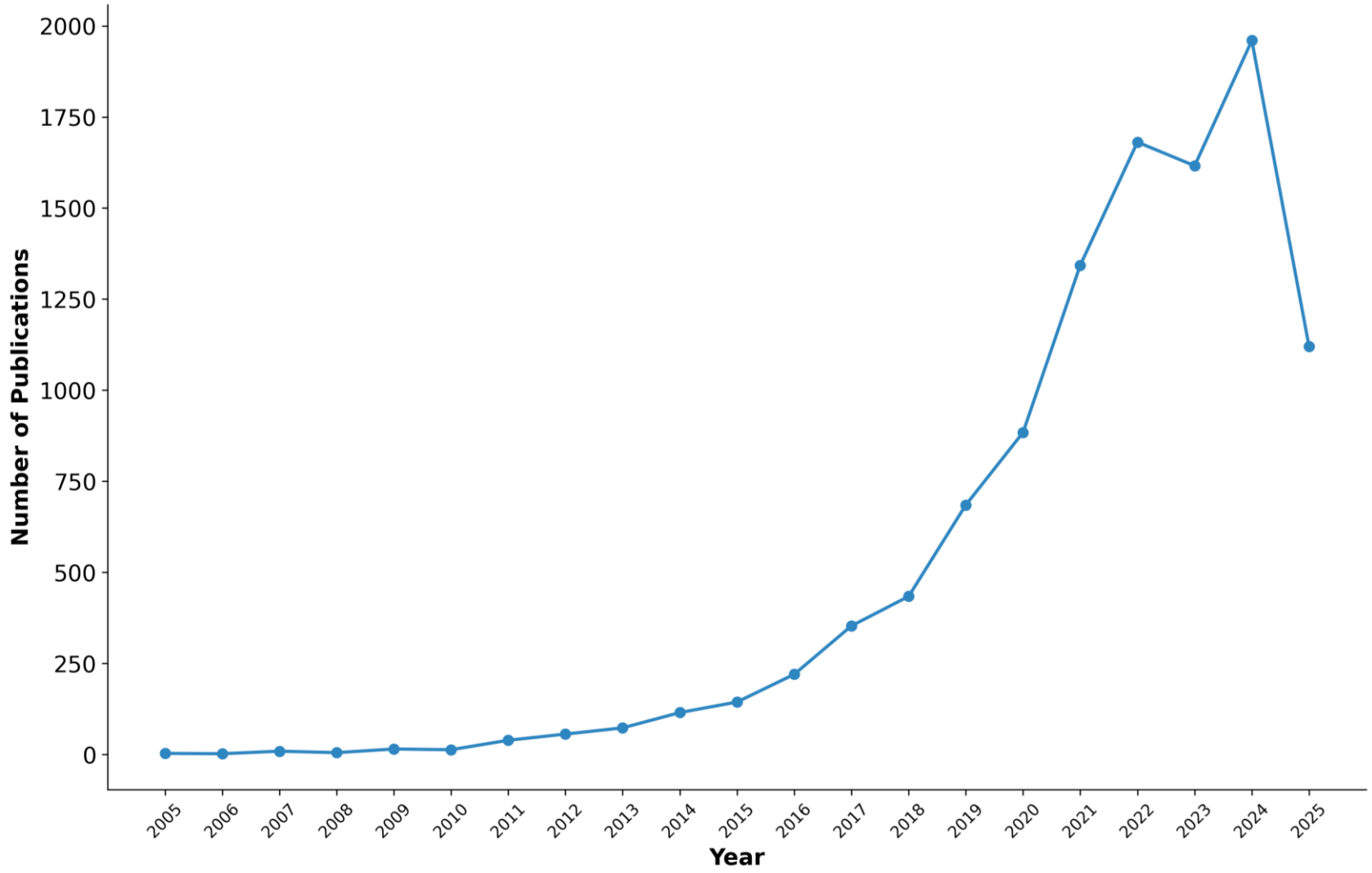
- Quality check
- Trimming (if necessary)
- De novo assembly or Mapping to reference
- Quantification of reads
- Downstream analysis

Gene	S1	S2	S_n
G1	11	5	10
G1	15	80	120
Gn...	40	18	5

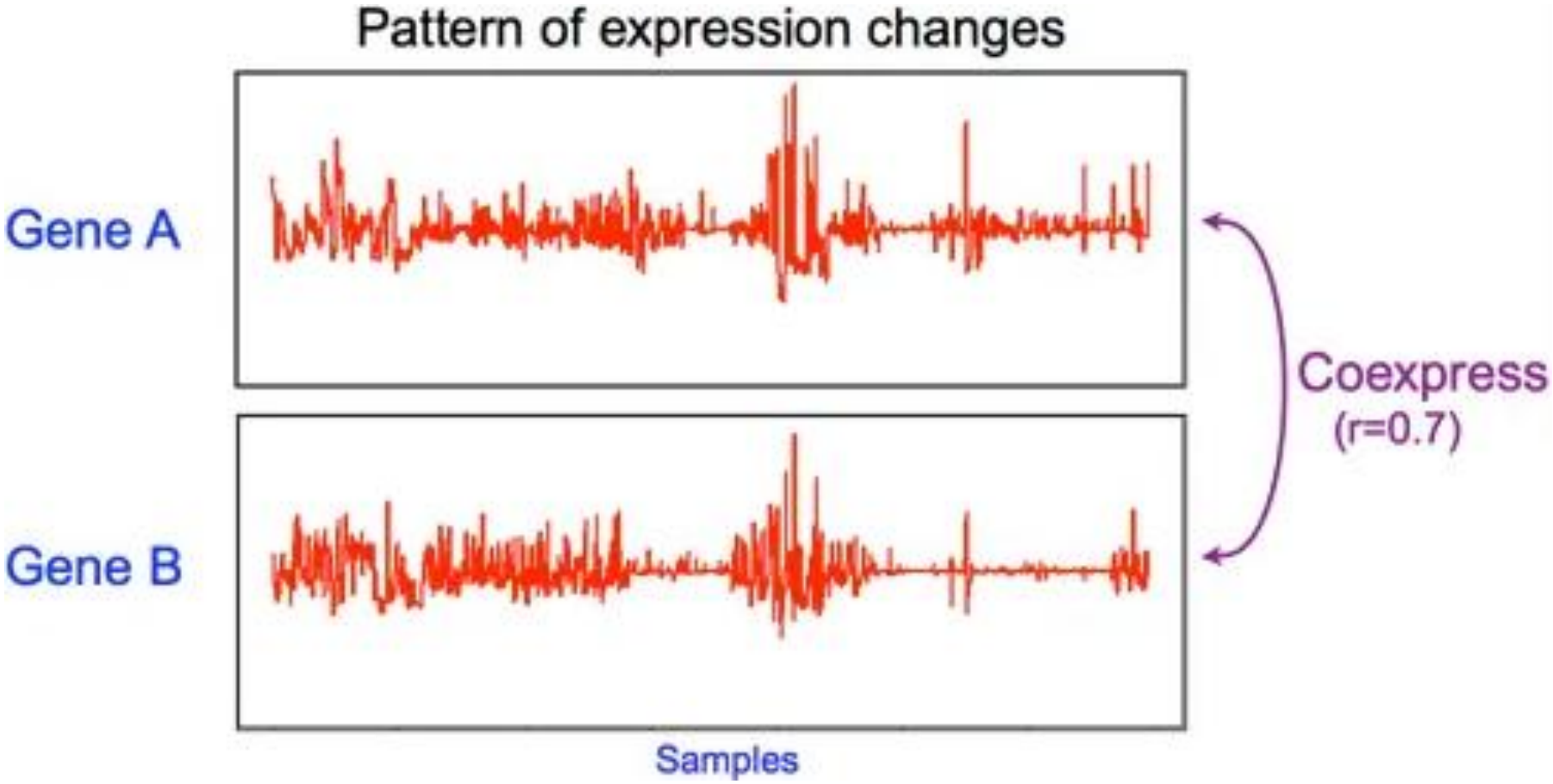
Gene co-expression networks

Gene Co-expresión Network Analysis

Growth of Co-expression Network usage in research (Pubmed Data)



Gene Co-expresión Network Analysis

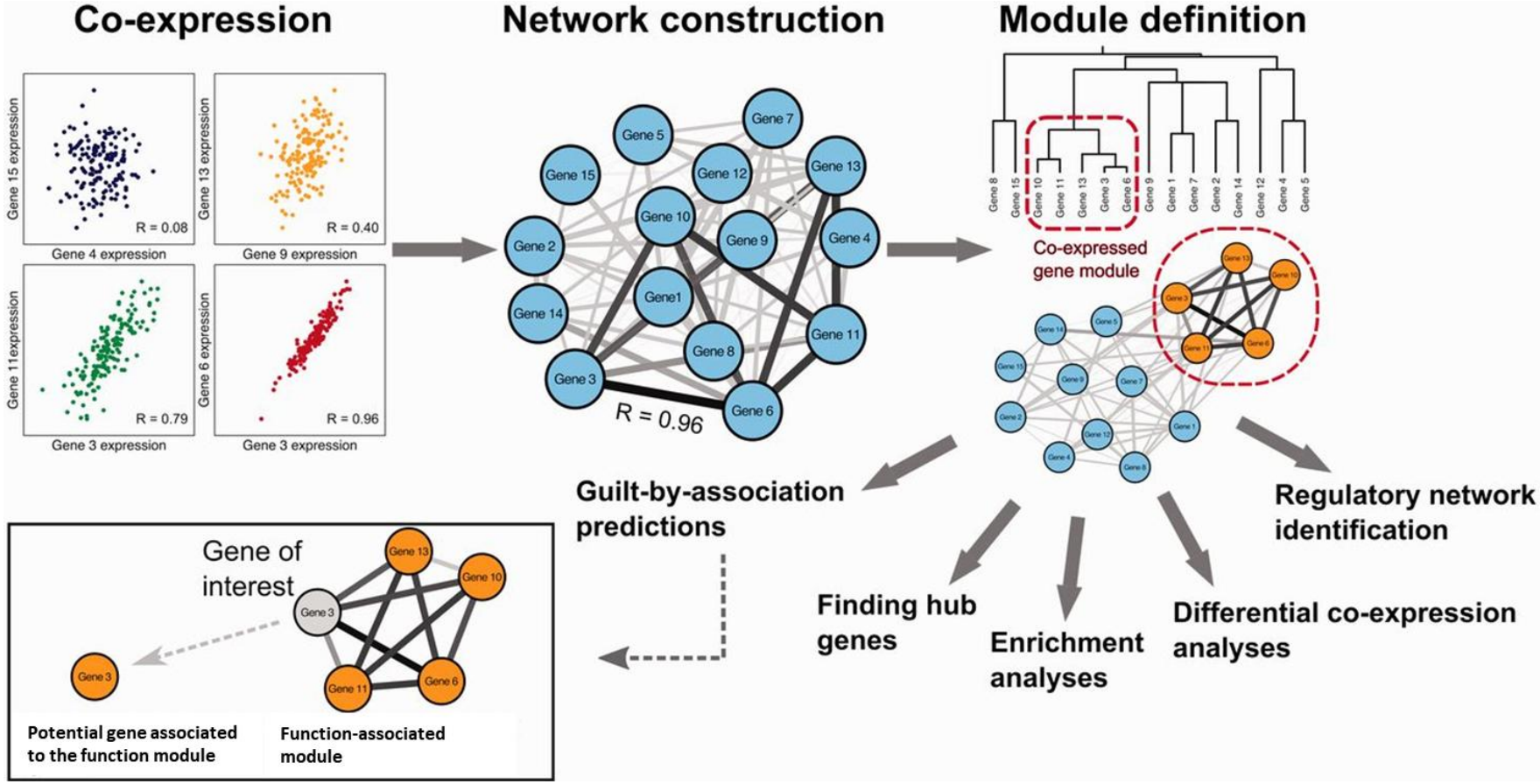


https://coxpresdb.jp/static/overview.shtml

Gene Co-expresión Network Analysis

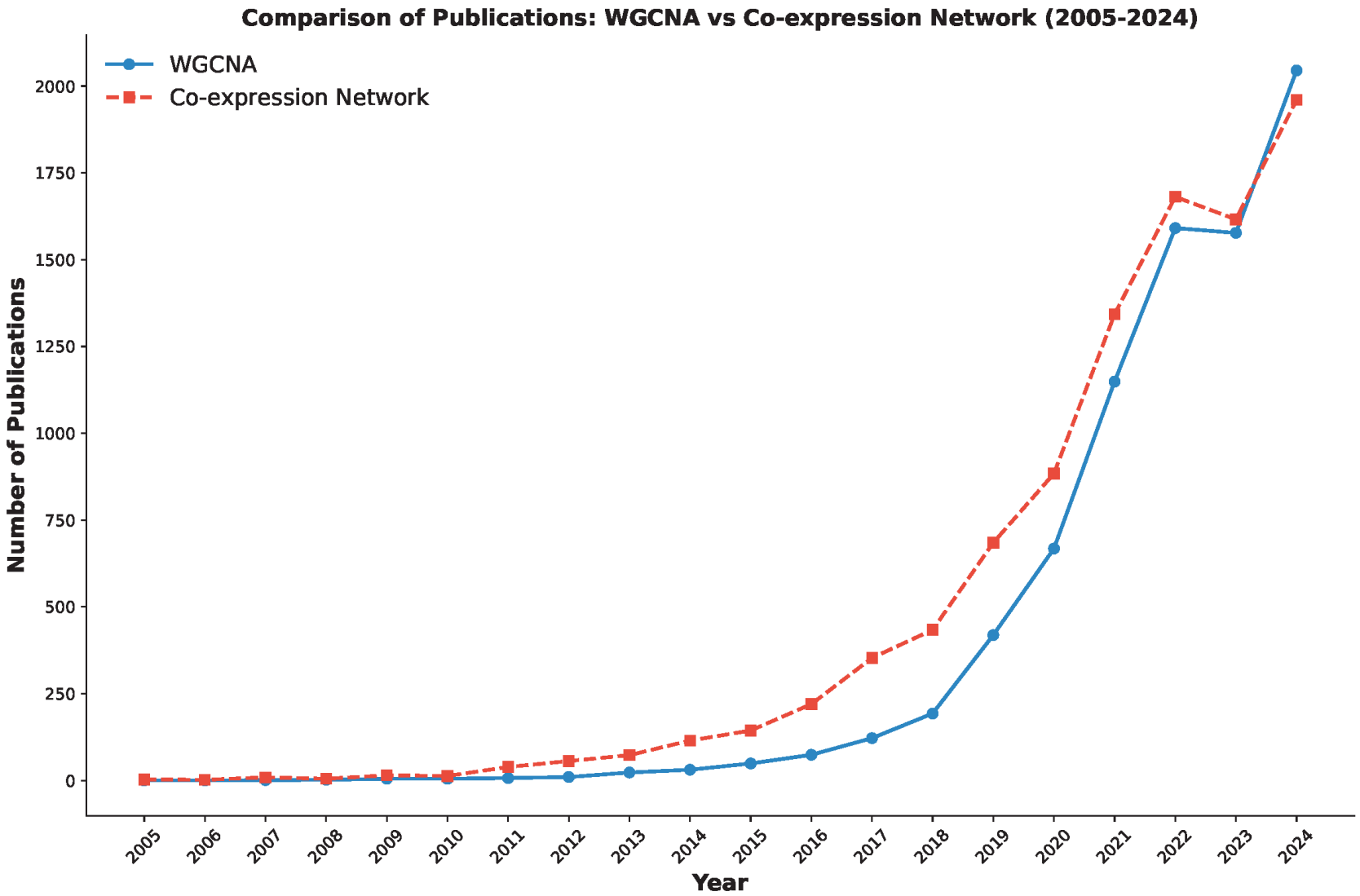
Gene expression data

	S1	S2	S3	Sn
G1	15	20	25	...
G2	30	35	45	...
G3	3	12	20	...
Gn	...	...	...	...



van Dam *et al.*, 2017.

Gene Co-expresión Network Analysis



WGCNA applications

Citation	Scientific Area	Biological System	Number of Samples	Main Findings	WGCNA Application
Wickramasuriya et al., 2021	Animal Science	Bovine mammary glands (lactation)	79 samples	Identified 104, 85, and 26 DEGs across lactation stages. Hub genes (GJA1, AP2A2, NPAS3) regulated lactation. Enriched pathways: ubiquitin-dependent ERAD, chaperone protein refolding.	Constructed co-expression networks for lactation stages, identified modules, hub genes, and enriched pathways using WGCNA combined with meta-analysis.
Yin et al.,2019	Human Health (Oncology)	Human tongue squamous cell carcinoma (TSCC)	57 (TSCC vs. normal)	Identified 9 hub genes (e.g., AIMP2, ANXA2) linked to poor prognosis. Enriched pathways: olfactory transduction, neuroactive ligand-receptor interaction.	Used WGCNA to cluster DEGs into modules, identified key modules correlated with TSCC, and screened hub genes for prognostic analysis.
Yang et al., 2018	Human Health (Oncology)	Human oral squamous cell carcinoma (OSCC, miRNA)	80 (40 OSCC, 40 matched non-tumorous)	Identified two novel miRNA modules associated with OSCC, regulating hub nodes in metabolic networks.	Modified WGCNA pipeline for paired design to reduce batch effects, identified miRNA co-expression modules.
Liu et al., 2017	Human Health (Oncology)	Human osteosarcoma	44 (22 osteosarcoma, 22 normal)	Identified modules enriched in cell cycle, DNA replication. Hub genes (e.g., CDK1, CCNB1) linked to osteosarcoma progression.	Used WGCNA to construct co-expression networks, identify modules, and screen hub genes from DEGs.
Zhang et al., 2023	Human Health (Microbiome)	Human tumor microbiome (TCGA)	1,000+ (multiple tumor types)	Identified oncogenome-associated microbiome modules (OAM). Tissierella linked to tumor suppression.	Built separate WGCNA networks for microbiome and mRNA, correlated modules to identify OAM and prognostic biomarkers.
Wei et al., 2020	Human Health (Oncology)	Human lung adenocarcinoma	517 (TCGA RNA-seq)	Identified modules enriched in immune response, cell adhesion. Hub genes (e.g., CDH1) linked to tumor progression.	Applied WGCNA to RNA-seq data, identified co-expression modules, and correlated with clinical traits.
Eidsaa et al., 2017	Comparative Genomics	Human and mouse tissues	76 human, 66 mouse (multiple tissues)	Enriched transcription factors (e.g., homeobox) in central genes. Disease genes (FOXE3, RHO) highly central.	Compared WGCNA networks using topological overlap and s-core analysis to identify conserved gene roles.
Zhu et al., 2022	Plant Biology (Pathogen Resistance)	Cassava (*Manihot esculenta*) roots infected with Xanthomonas axonopodis	18 (6 time points post-infection, 3 replicates)	Identified MeWRKY IIa genes (e.g., MeWRKY27) as hub genes regulating resistance to cassava bacterial blight. Enriched pathways: plant-pathogen interaction, MAPK signaling.	Used WGCNA to construct co-expression networks from RNA-seq data, identified modules correlated with pathogen response, and screened hub genes.
Liu et al., 2021	Plant Biology (Abiotic Stress)	Maize (*Zea mays*) leaves under drought stress	24 (4 growth stages, 3 replicates, control vs. drought)	Identified 3,451 DEGs and 12 modules. Hub genes (e.g., HSP90, LEA proteins) linked to drought tolerance via ABA and stress defense pathways.	Applied WGCNA to RNA-seq data, identified modules correlated with drought stress at different growth stages, and detected hub genes.

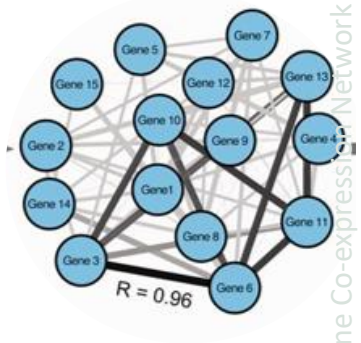
WGCNA: a standard tool for co-expression networks

Langfelder & Horvath (2008), *BMC Bioinformatics*, 9:559
Zhang & Horvath (2005), *Stat Appl Genet Mol Biol*, 4:17.

WGCNA Flow work

Gene expression data

	S1	S2	S3	Sn
G1	15	20	25	...
G2	30	35	45	...
G3	3	12	20	...
Gn	...	...	...	...



Expression Matrix (X) → Correlation Matrix (R) → Adjacency Matrix (A) → TOM (Topological Overlap Matrix) → dissTOM → Clustering → Modules analysis → Network visualization

Jorge A. Tzec-Interián | Gene Co-expression Network

Langfelder & Horvath, 2008; Zhang & Horvath, 2005

Weighted: Continuous edge weights (0 to 1).

Correlation-based: Uses Pearson/Spearman correlations.

Undirected: Symmetric gene relationships, no causality.

Scale-free topology: Power-law degree distribution.

Module detection: Clusters genes via hierarchical clustering, TOM.

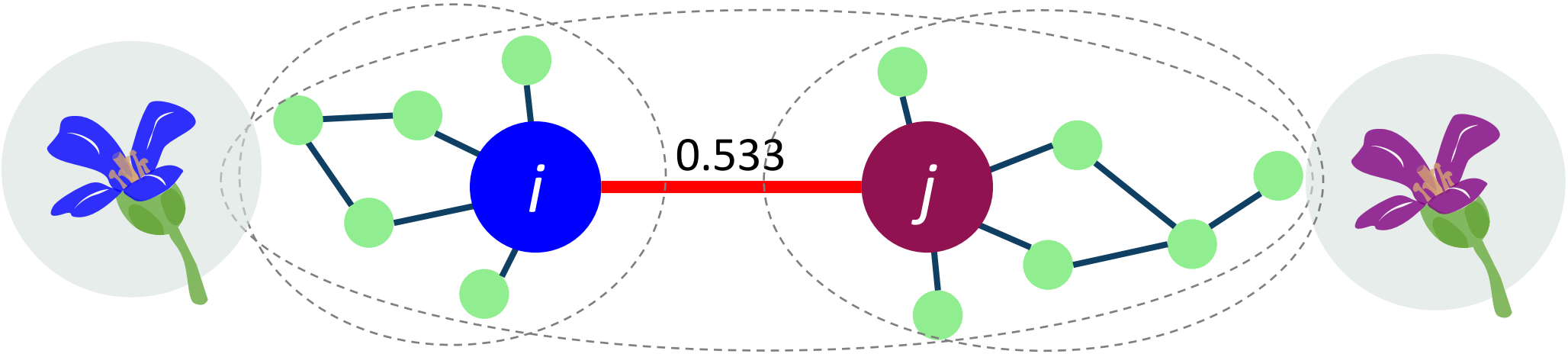
Trait correlation: Links modules to phenotypes via eigengenes.

Soft-thresholding: Adjusts correlations with β for scale-free fit.

Noise-robust: TOM reduces spurious correlations.

Scalable: Handles large datasets (RNA-seq, thousands of genes).

Limitations: No causality, sensitive to batch effects, needs preprocessing.



Langfelder & Horvath, 2008; Zhang & Horvath, 2005 Note: Simulated data

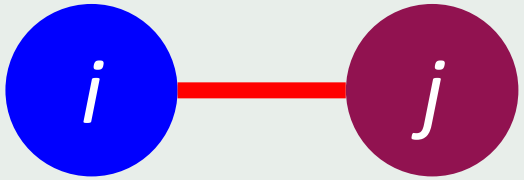
Weighted: Continuous edge weights (0 to 1).

Correlation-based: Uses Pearson/Spearman correlations.

Soft-thresholding: Adjusts correlations with β for scale-free fit.

Highlights strong connections and reduces the impact of weak correlations

$$a_{ij} = |cor(X_i, X_j)|^{\beta}$$



Normalized Expression Matrix

Gene	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Gene A	3.745401	9.507143	7.319939	5.986585	1.560186
Gene B	1.559945	0.580836	8.661761	6.011150	7.080726
Gene C	0.205845	9.699099	8.324426	2.123391	1.818250
Gene D	1.834045	3.042422	5.247564	4.319450	2.912291
Gene E	6.118529	1.394939	2.921446	3.663618	4.560700

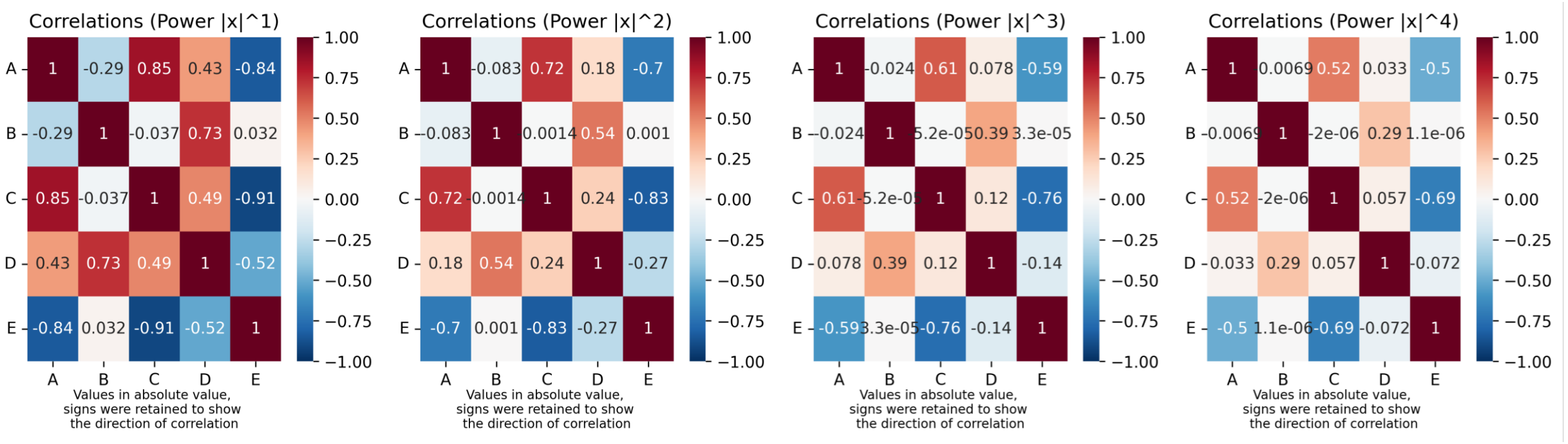
Langfelder & Horvath, 2008; Zhang & Horvath, 2005 Note: Simulated data

Weighted: Continuous edge weights (0 to 1).

Correlation-based: Uses Pearson/Spearman correlations.

Soft-thresholding: Adjusts correlations with β for scale-free fit.

$$a_{ij} = |cor(X_i, X_j)|^{\beta}$$



Jorge A. Tzec-Interián | Gene Co-expression Network

Langfelder & Horvath, 2008; Zhang & Horvath, 2005 Note: Simulated data

WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → Adjacency Matrix (A) → TOM → dissTOM → Clustering

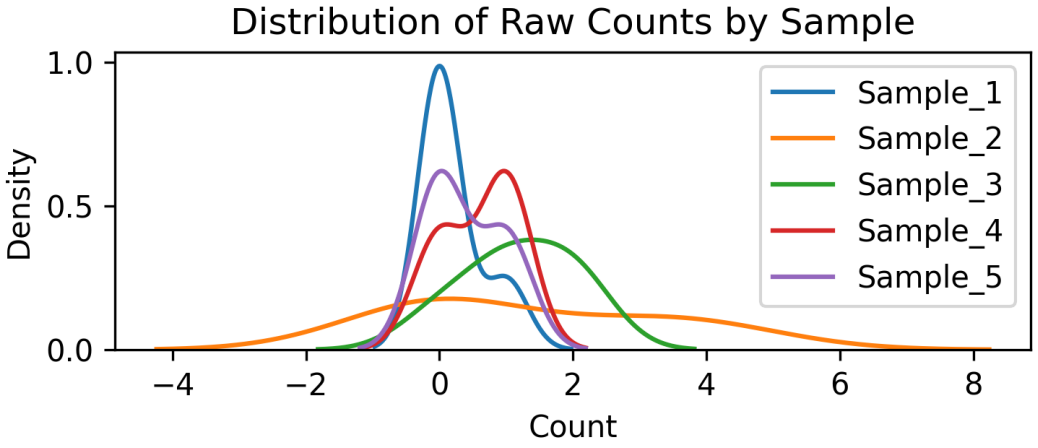
Input: Matrix of gene expression values (row counts)

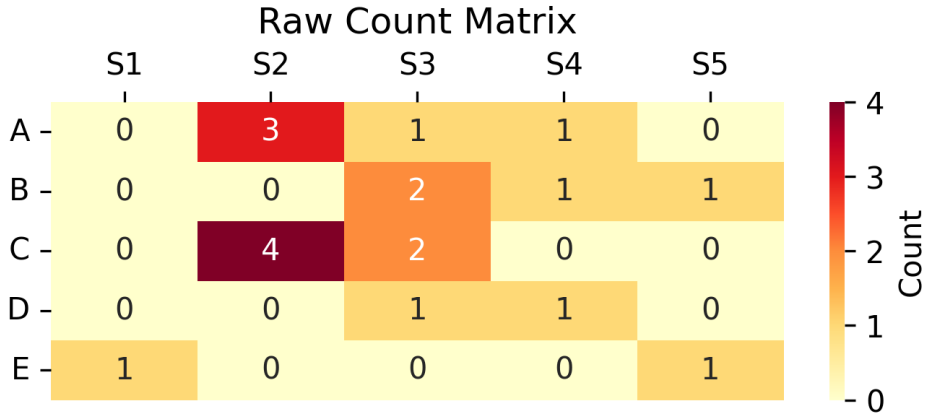
$X = [x_{is}]$, where x_{ij} is the expression of gene i in sample s.

Preprocessing

- Normalization (VST, Rlog, CPM, TMM or log2(TPM+1)
- Filtering (low-variance genes)

Output: A cleaned matrix X (e.g., 5000 genes × 20 samples).



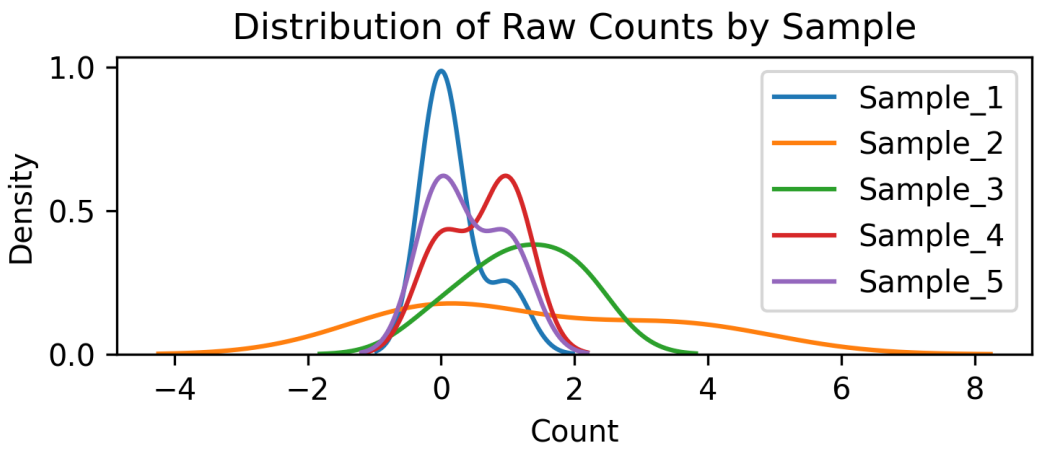


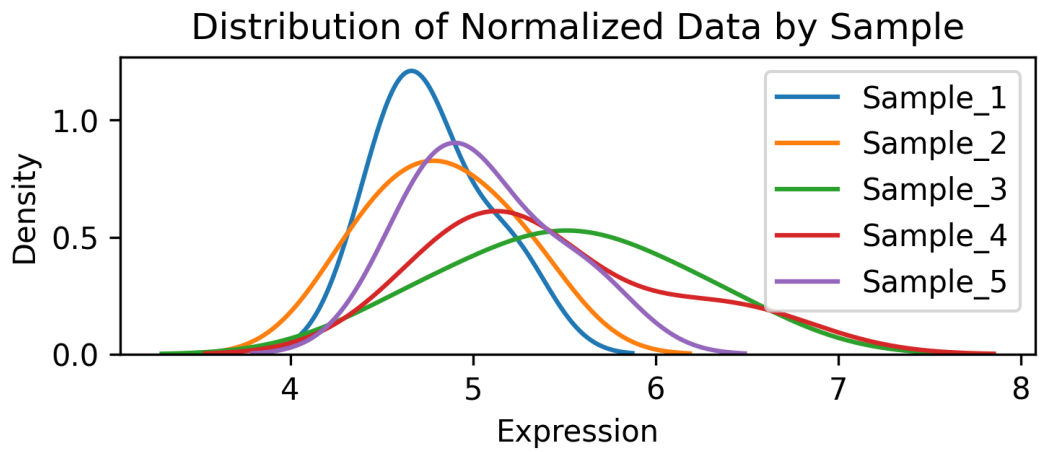
Jorge A. Tzec-Interián | Gene Co-expression Network

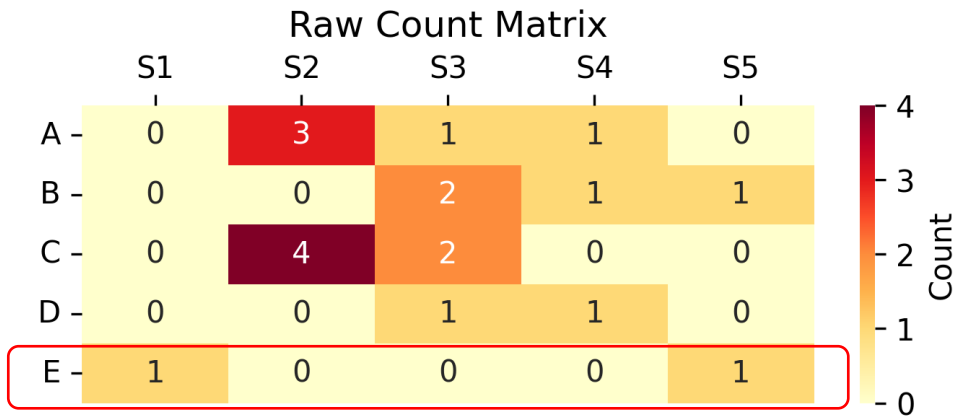
Langfelder & Horvath, 2008; Zhang & Horvath, 2005 Note: Simulated data

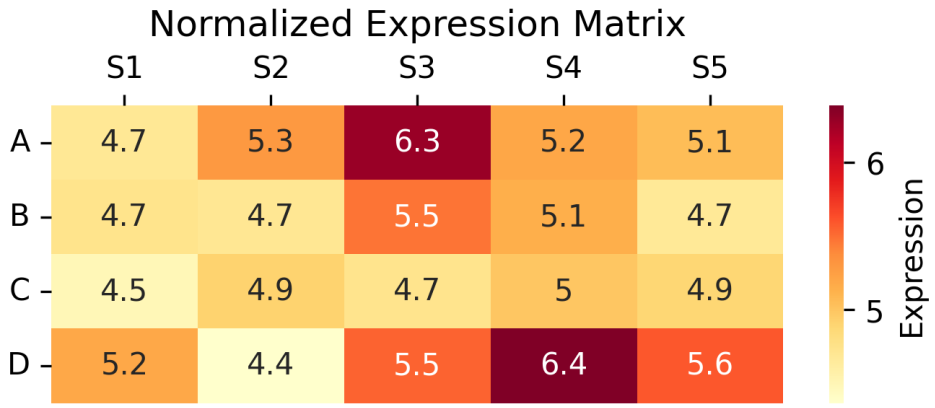
WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → Adjacency Matrix (A) → TOM → dissTOM → Clustering









Langfelder & Horvath, 2008; Zhang & Horvath, 2005

Note: Simulated data

Jorge A. Tzec-Interián | Gene Co-expression Network

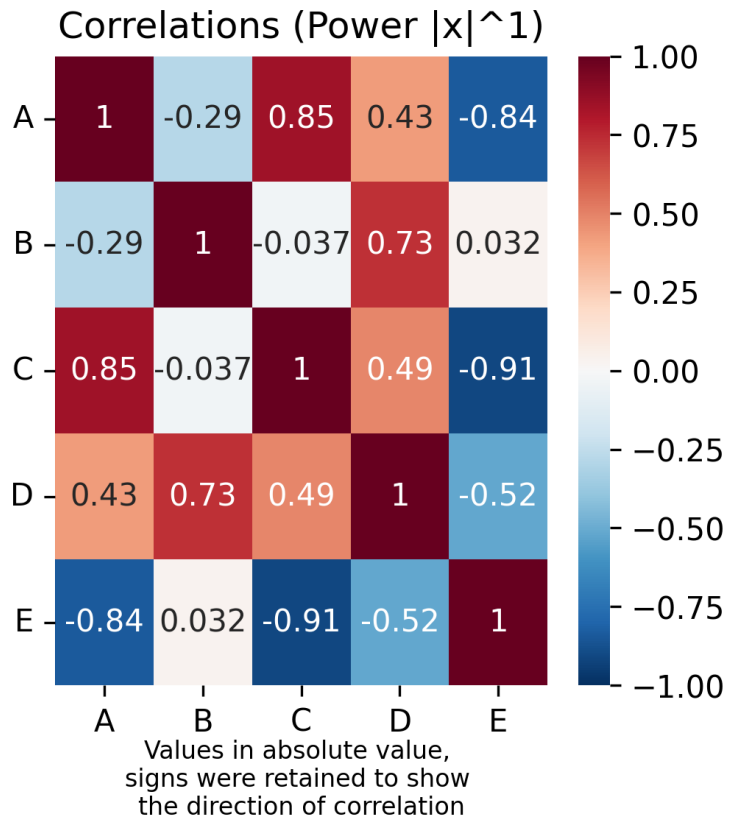
WGCNA
work flow

Expression Matrix (X) → **Correlation Matrix (R)** → Adjacency Matrix (A) → TOM → dissTOM → Clustering

Compute pairwise correlations between genes to measure co-expression

$$r_{ij} = \frac{cov(X_i, X_j)}{\sigma_{x_i}\sigma_{x_i}}$$

Pearson correlation: where X_i and X_j are expression profiles of genes i and j.



Input: A cleaned matrix X .

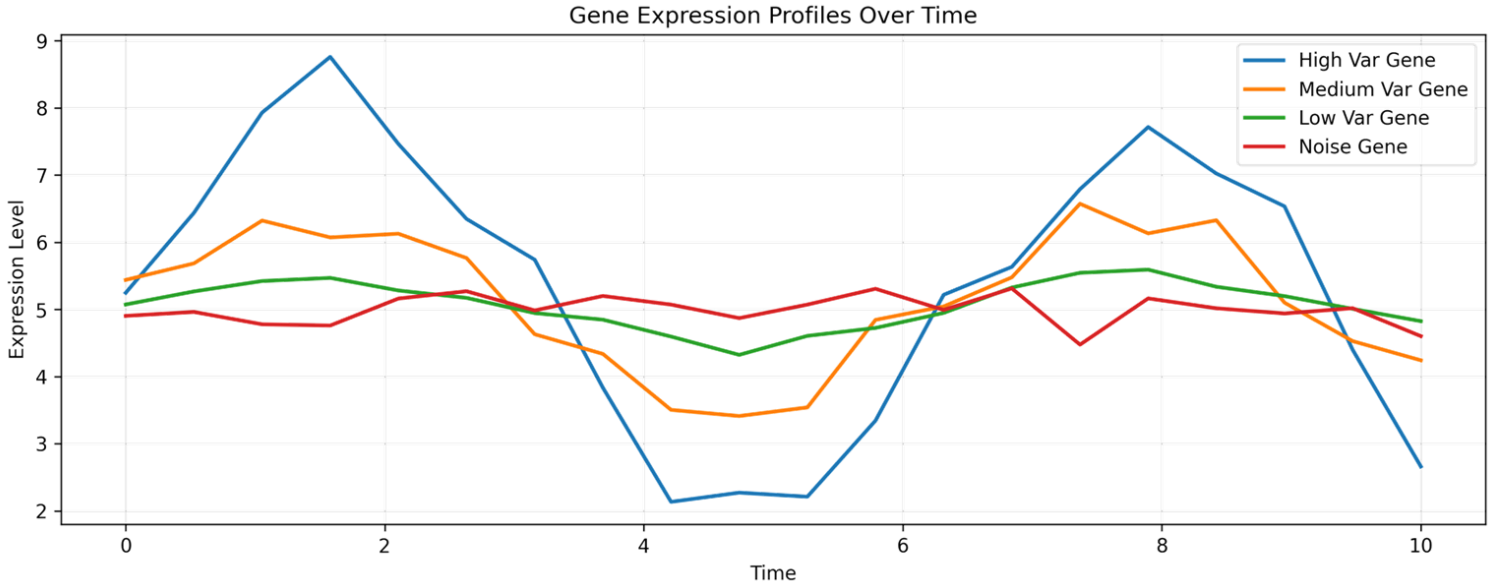
Output: A symmetric $n \times n$ matrix $R = [r_{ij}]$, where $r_{ij} \in [-1,1]$

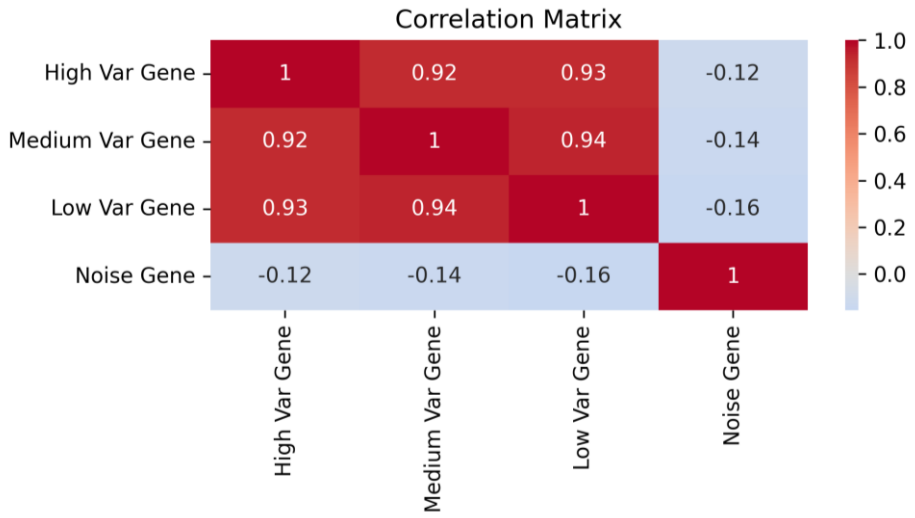
Langfelder & Horvath, 2008; Zhang & Horvath, 2005 Note: Simulated data

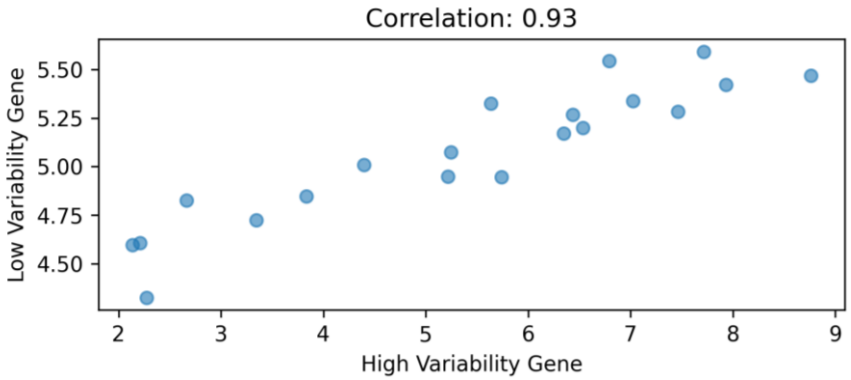
Jorge A. Tzec-Interián | Gene Co-expression Network

WGCNA
work flow

Expression Matrix (X) → **Correlation Matrix (R)** → Adjacency Matrix (A) → TOM → dissTOM → Clustering







Jorge A. Tzec-Interián | Gene Co-expression Network

Langfelder & Horvath, 2008; Zhang & Horvath, 2005 Note: Simulated data

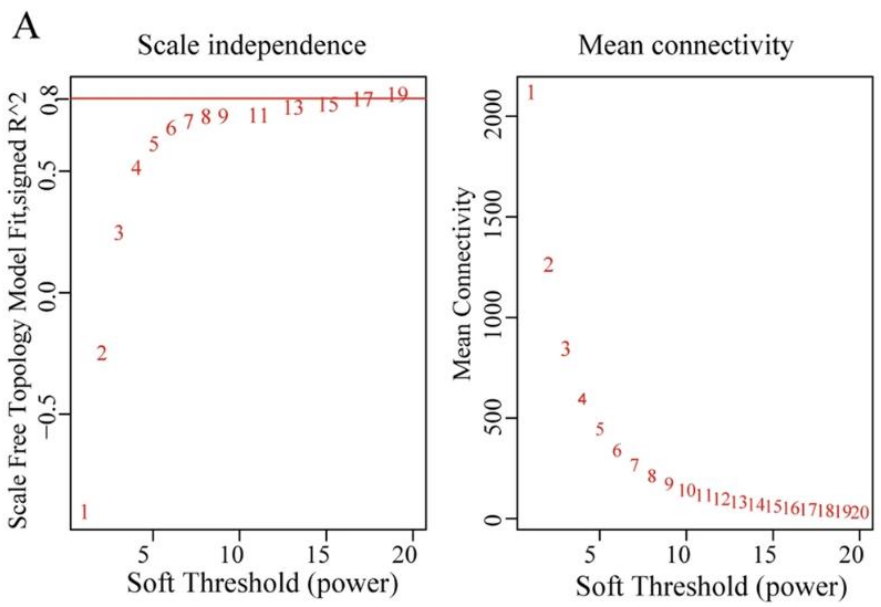
WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → **Adjacency Matrix (A)** → TOM → dissTOM → Clustering

Transform correlations into weighted connections using the power β.

$$a_{ij} = |cor(X_i, X_j)|^{\beta}$$

β is chosen to achieve scale-free topology (from **pickSoftThreshold**)



Zhao et al, 2023

Input:_____

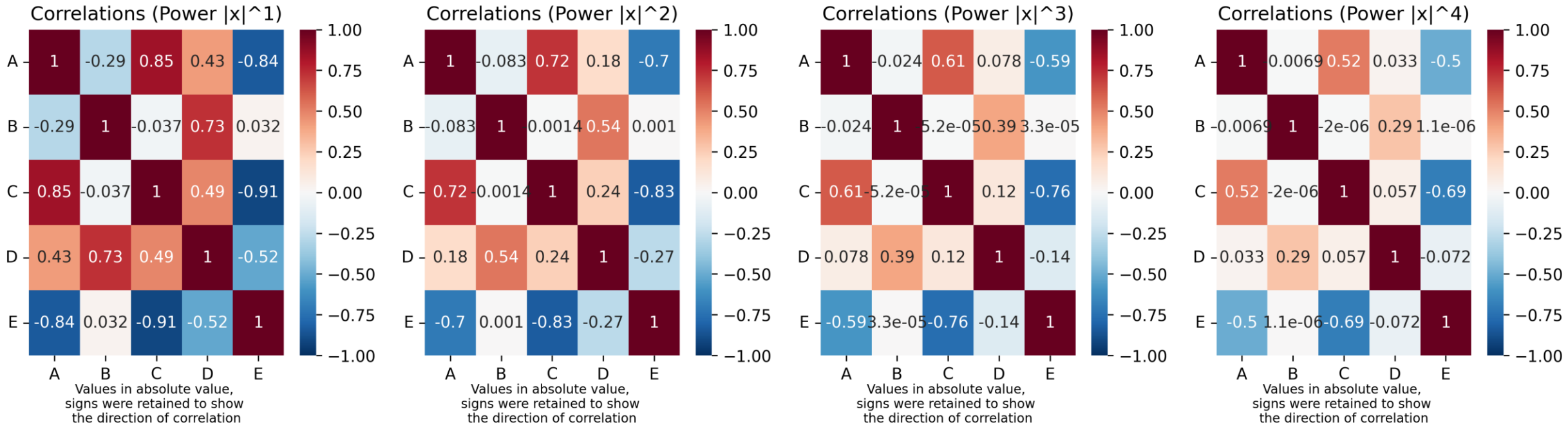
Output: A symmetric $n \times n$ matrix $A = [a_{ij}]$, , where $r_{ij} \in [0,1]$

Langfelder & Horvath, 2008; Zhang & Horvath, 2005

WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → **Adjacency Matrix (A)** → TOM → dissTOM → Clustering

Adjacency matrix at different power values (β).

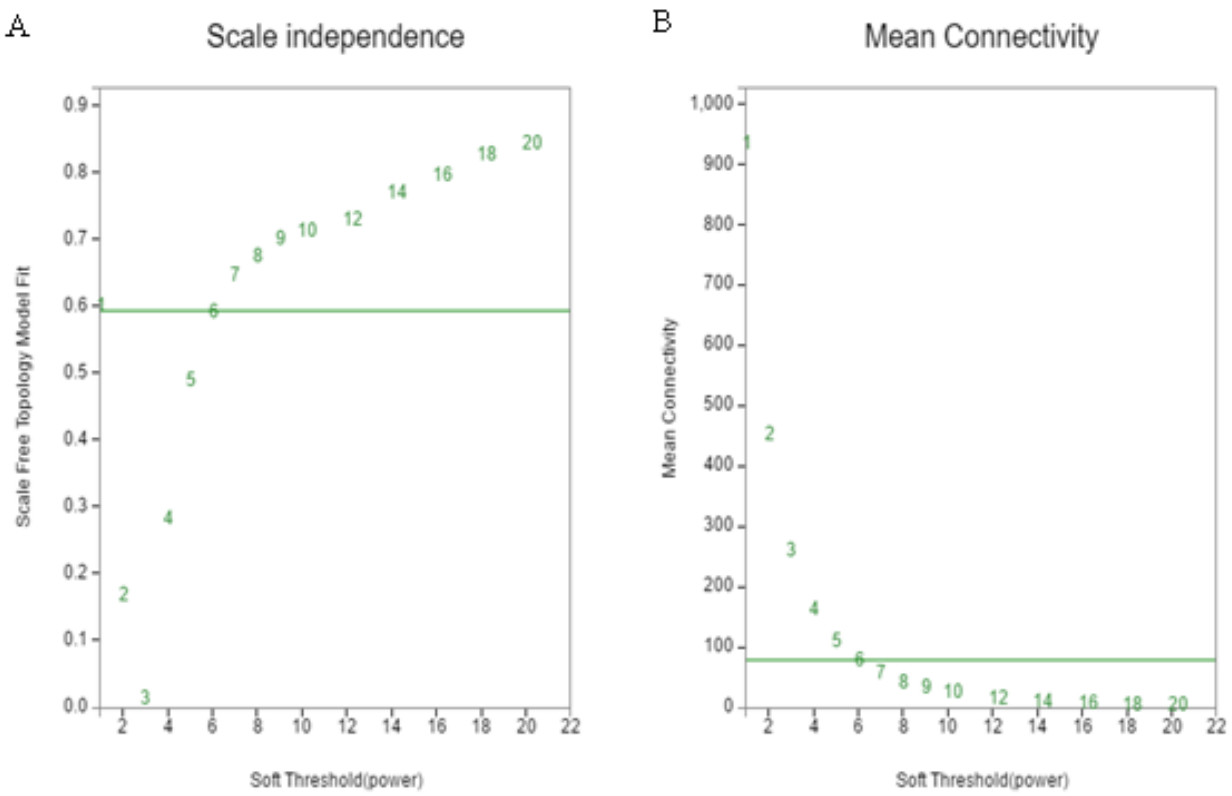


Note: Simulated data

WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → **Adjacency Matrix (A)** → TOM → dissTOM → Clustering

Global Transcriptome and Weighted Gene Co-expression Network Analyses of Growth-Stage-Specific Drought Stress Responses in Maize



Liu et al., 2021	Plant Biology (Abiotic Stress)	Maize (Zea mays) leaves under drought stress	24 (4 growth stages, 3 replicates, control vs. drought)	Identified 3,451 DEGs and 12 modules. Hub genes (e.g., HSP90, LEA proteins) linked to drought tolerance via ABA and stress defense pathways.

WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → **Adjacency Matrix (A)** → TOM → dissTOM → Clustering

Recommended beta values for WGCNA by sample size and network type (From FAQ- Horvarth Lab)

Number of samples	Unsigned and signed hybrid networks	Signed networks
Less than 20	9	18
20-30	8	16
30-40	7	14
more than 40	6	12

Jorge A. Tzec-Interián | Gene Co-expression Network

Langfelder & Horvath, 2008; Zhang & Horvath, 2005; https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/faq.html

WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → **Adjacency Matrix (A)** → TOM → dissTOM → Clustering

More detailed Beta values for WGCNA by sample size and network type

Number of Samples	Network Type	Recommended Beta (Power)	Notes
< 20	Unsigned or Signed Hybrid	9–12	Small sample sizes require higher beta to achieve scale-free topology due to noisy correlations. Verify with pickSoftThreshold.
< 20	Signed	18–24	Signed networks preserve correlation direction, needing higher beta to emphasize strong connections.
20–30	Unsigned or Signed Hybrid	7–9	Moderate sample sizes allow lower beta, balancing connectivity and scale-free fit ($r \geq 0.8$).
20–30	Signed	14–18	Higher beta for signed networks ensures robust module detection.
30–40	Unsigned or Signed Hybrid	6–8	Larger samples improve correlation estimates, reducing beta needed for $r \geq 0.8$.
30–40	Signed	12–16	Signed networks still require higher beta but less than for smaller samples.
> 40	Unsigned or Signed Hybrid	5–7	Large datasets yield stable correlations, allowing lower beta for scale-free topology.
> 40	Signed	10–14	Lowest beta for signed networks in large datasets, maintaining connectivity.

WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → **Adjacency Matrix (A)** → TOM → dissTOM → Clustering

Network types in WGCNA: Unsigned, Signed, and Hybrid

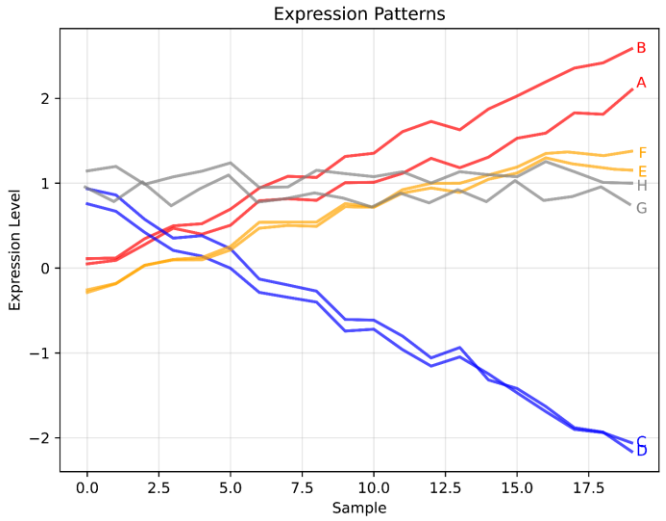
Network type	Formula	Interpretation
Unsigned	$a_{ij} = \lvert cor(X_i, X_j) \rvert^{\beta}$	Emphasizes the magnitude of correlations, ignoring their direction (**positive or negative**). Genes with strong correlations are grouped together regardless of sign.
Signed	$a_{ij} = \left(\frac{1 + cor(X_i, X_j)}{2}\right)^{\beta}$	**Preserves correlation direction**, strengthening positive relationships and weakening negative ones. Useful for distinguishing co-expression patterns.
Hybrid signed	$a_{ij} = \left(\frac{1 + cor(X_i, X_j)}{2}\right)^{\beta}$ If $cor(X_i, X_j) < 0, a_{ij}$=0	Focuses **exclusively on positive correlations**, setting negative correlations to zero. Ideal for studying positive co-expression only.

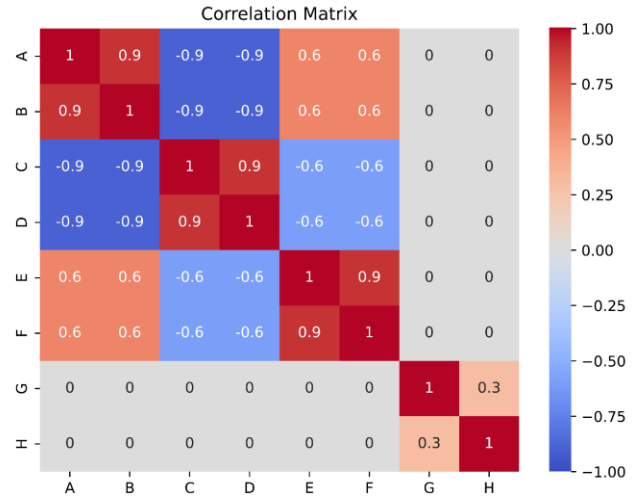
Jorge A. Tzec-Interián | Gene Co-expression Network

Langfelder & Horvath, 2008; Zhang & Horvath, 2005

WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → **Adjacency Matrix (A)** → TOM → dissTOM → Clustering





Correlation Types:

Strong Positive (≈ +0.9): A-B, C-D, E-F

Strong Negative (≈ -0.9): A/B vs C/D

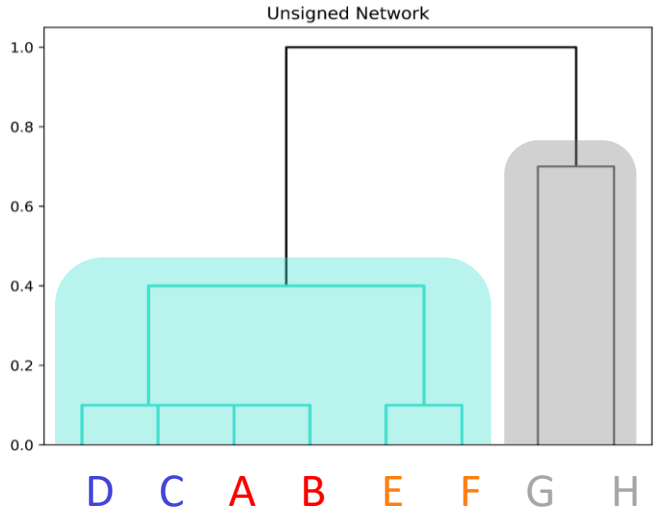
Moderate Positive (≈ +0.6): A/B vs E/F

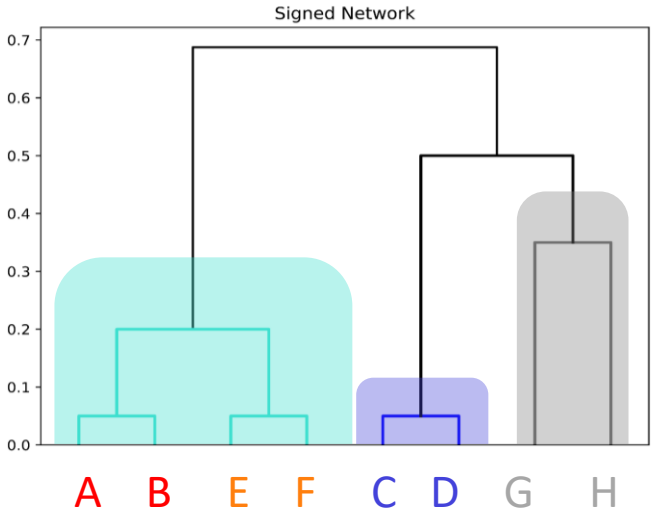
Moderate Negative (≈ -0.6): C/D vs E/F

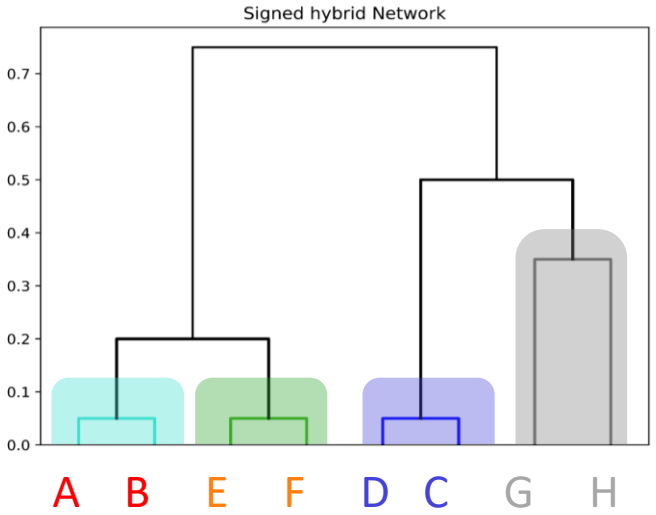
Weak Positive (≈ +0.3): G-H

Near-null (≈ 0): G/H vs others

Expected Modules:
Unsigned: {A, B, C, D, E, F} (turquoise), G-H (gray)
Signed: {A, B, E, F} (turquoise), {C, D} (blue), G-H (gray)
Signed Hybrid: {A, B} (turquoise), {C, D} (blue), {E, F} (green), G-H (gray)







Langfelder & Horvath, 2008; Zhang & Horvath, 2005

Note: Simulated data

WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → Adjacency Matrix (A) → **TOM** → dissTOM → Clustering

Topological Overlap Matrix (TOM): Measure topological similarity between genes, accounting for shared neighbors.

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

Measures shared neighbors.

Scales the numerator to keep TOM [0,1].

Connectivity of gene *i*

$$k_i = \sum_u a_{iu}$$

a_{ij}: Direct connection (adjacency).

$\sum_u a_{iu}a_{uj}$: Shared neighbors.

$\min(k_i k_j)$: Normalizes by connectivity.

Input: A symmetric $n \times n$ matrix $A = [a_{ij}]$, , where $r_{ij} \in [0,1]$

Output: A symmetric $n \times n$ matrix $TOM = [TOM_{ij}]$, where $TOM \in [0,1]$

Langfelder & Horvath, 2008; Zhang & Horvath, 2005

WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → Adjacency Matrix (A) → **TOM** → dissTOM → Clustering

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

Gene	Connected to
i	A, B, C
j	A, C, D

Example:
Understanding in TOM Calculation

$$\sum_u a_{iu} a_{uj}$$

u = A → ($a_{iA} = 1$), ($a_{jA} = 1$) → = 1
u = B → ($a_{iB} = 1$), ($a_{jB} = 0$) → = 0
u = C → ($a_{iC} = 1$), ($a_{jC} = 1$) → = 1
u = D → ($a_{iD} = 0$), ($a_{jD} = 1$) → = 0

Summing up the products:
[1 + 0 + 1 + 0 = 2]

Interpretation:
Genes *i* and *j* share two common neighbors (A and C)

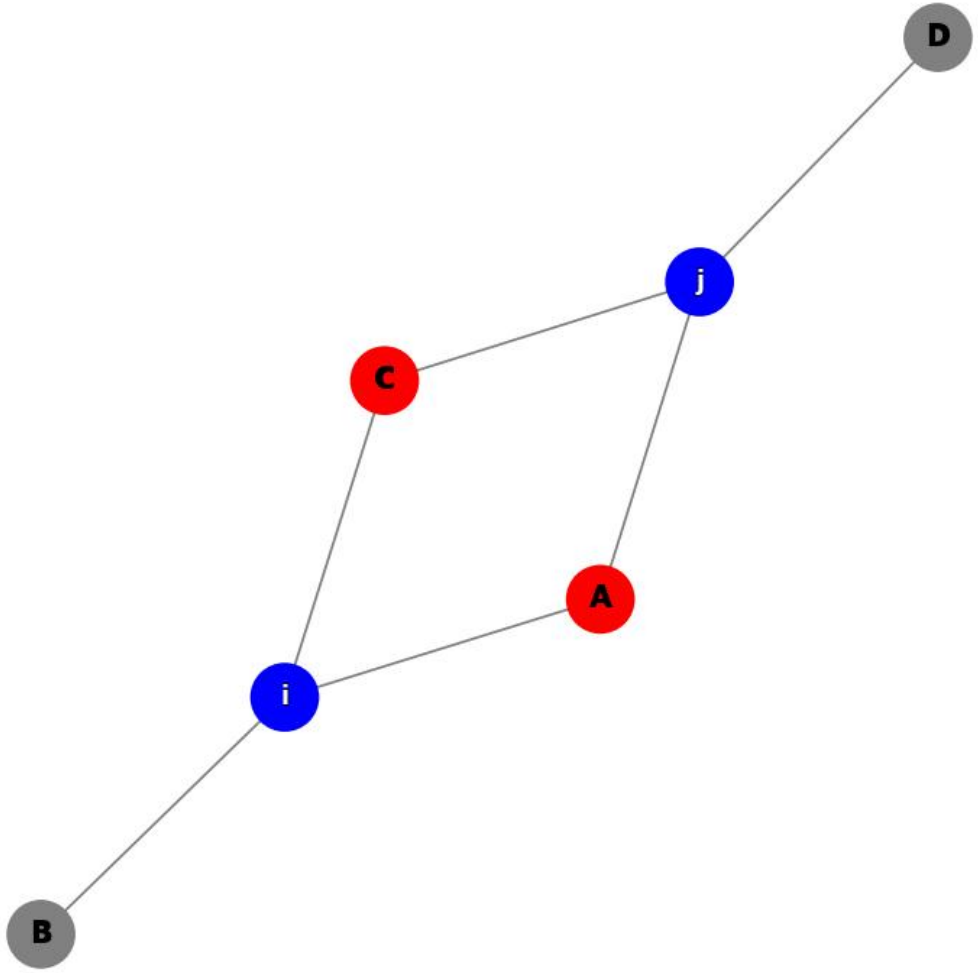
Jorge A. Tzec-Interián | Gene Co-expression Network

WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → Adjacency Matrix (A) → **TOM** → dissTOM → Clustering

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

Gene	Connected to
i	A, B, C
j	A, C, D

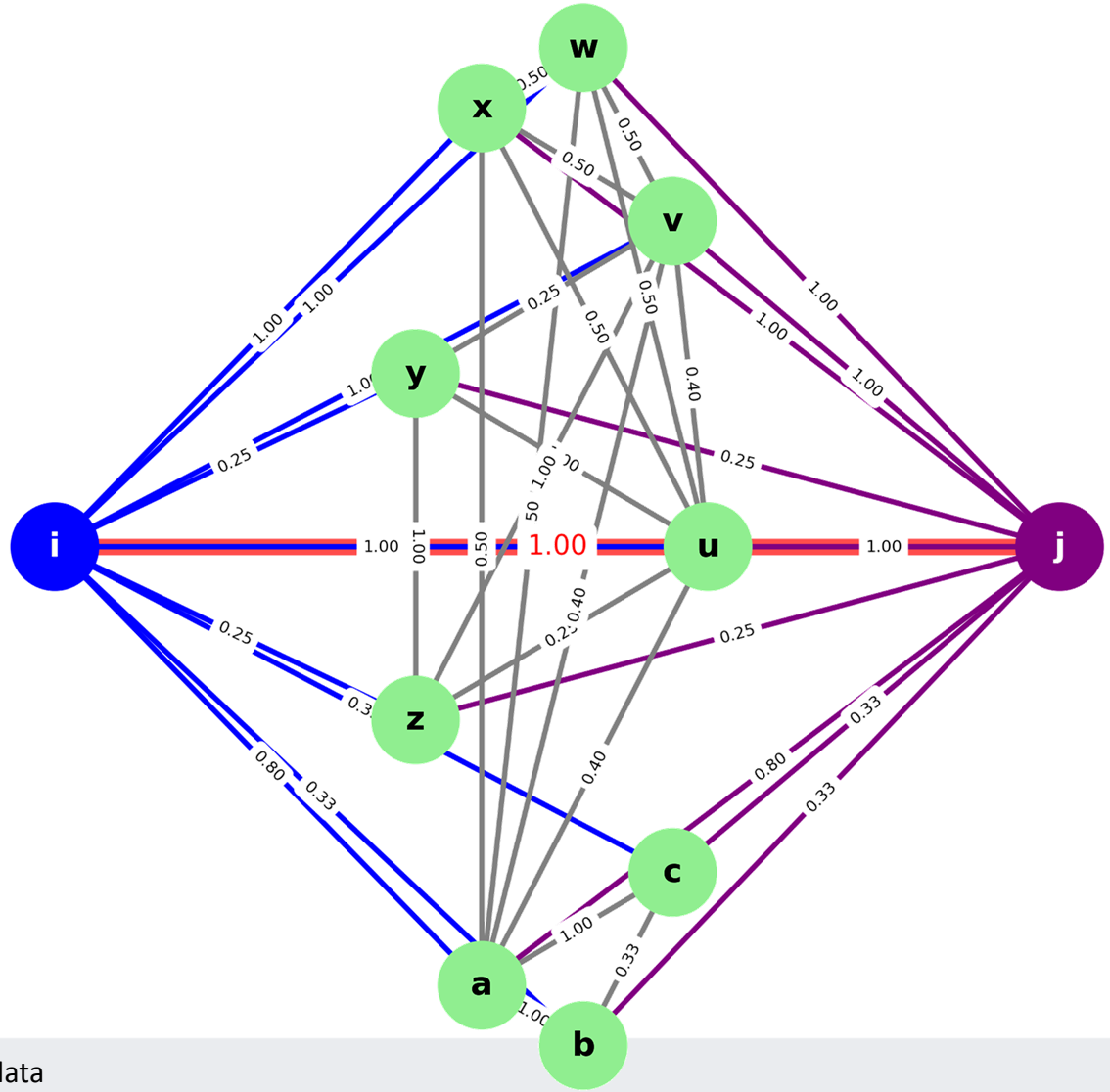


Jorge A. Tzec-Interián | Gene Co-expression Network

WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → Adjacency Matrix (A) → **TOM** → dissTOM → Clustering

Network with Topological Overlap Matrix (TOM). Simulation of *i-j* connections



Genes with similar connection patterns (shared neighbors) have **higher TOM values**, reducing noise.

Jorge A. Tzec-Interián | Gene Co-expression Network

Langfelder & Horvath, 2008; Zhang & Horvath, 2005

Note: Simulated data

WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → Adjacency Matrix (A) → TOM → **dissTOM** → Clustering

Convert TOM into a dissimilarity measure for clustering to group genes into modules.

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

Output: A symmetric $n \times n$ matrix $dissTOM = [dissTOM_{ij}]$, where $dissTOM \in [0,1]$

Jorge A. Tzec-Interián | Gene Co-expression Network

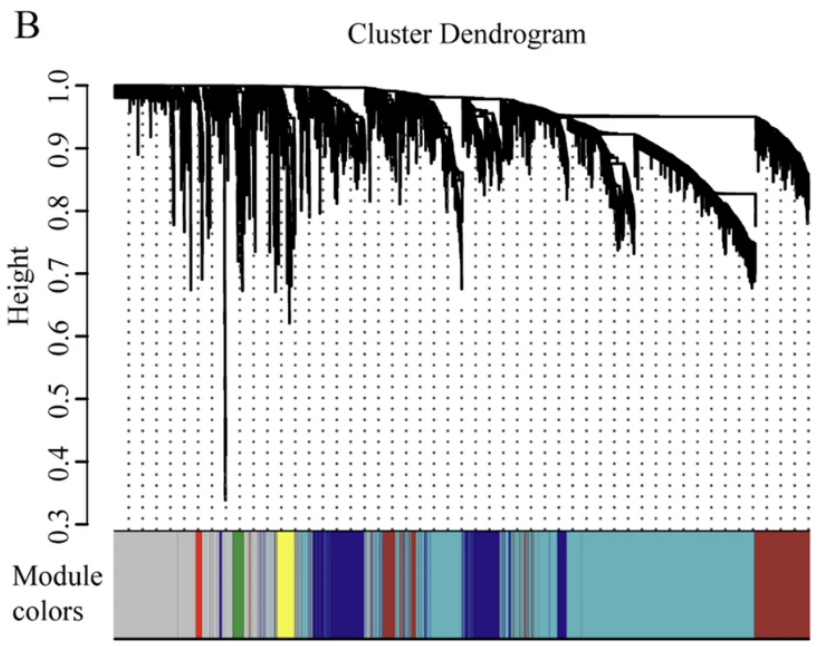
Langfelder & Horvath, 2008; Zhang & Horvath, 2005 Note: Simulated data

WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → Adjacency Matrix (A) → TOM → dissTOM → **Clustering**

STEP 1 Hierarchical Clustering	Use dissTOM ($dissTOM = 1 - TOM$) as distance.
	Method: Average linkage, builds dendrogram.
STEP 2 Dynamic Tree Cutting:	Identifies modules (gene groups with low dissTOM).
	Parameters: minModuleSize (e.g., 30), deepSplit (sensitivity).
STEP 3 Eigengene Calculation:	Eigengene = First principal component of module's expression.
	Summarizes module's expression pattern.
STEP 4 Module Merging	Merge modules if eigengenes correlate highly (e.g., >0.75).

Genes with low dissTOM (high TOM) are clustered together, forming modules.



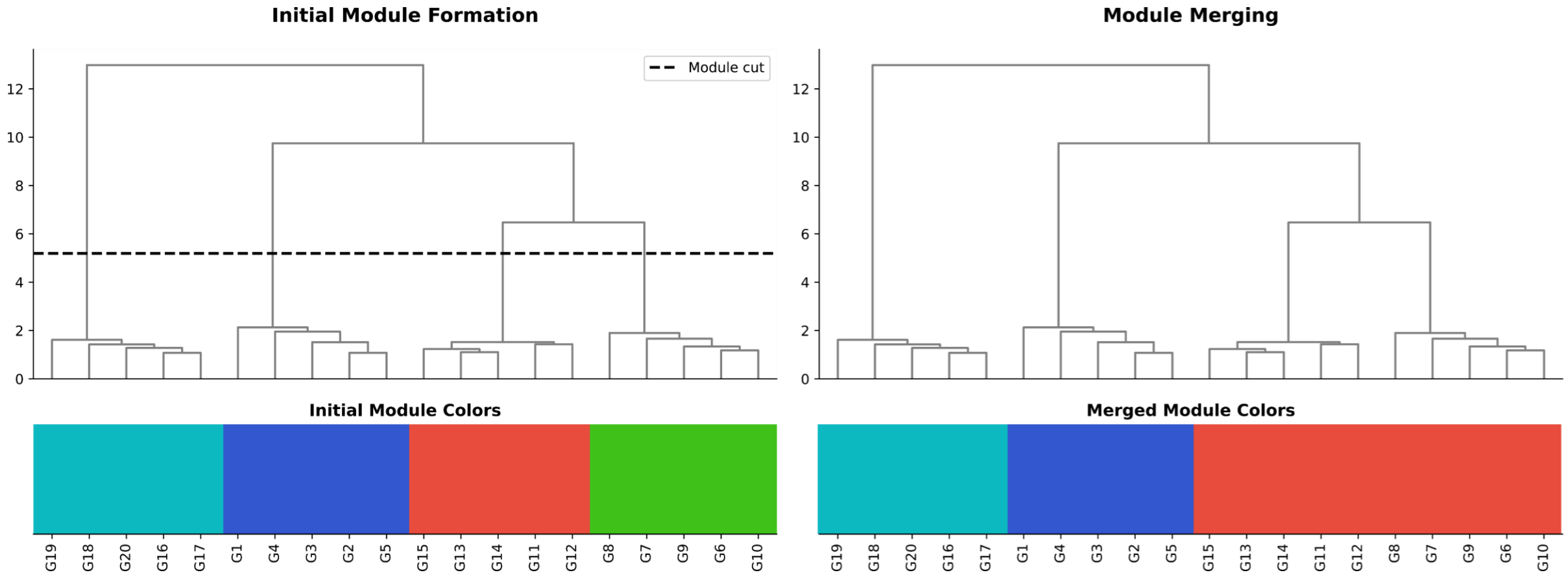
Zhao et al, 2023

Jorge A. Tzec-Interián | Gene Co-expression Network

Langfelder & Horvath, 2008; Zhang & Horvath, 2005 Note: Simulated data

WGCNA
work flow

Expression Matrix (X) → Correlation Matrix (R) → Adjacency Matrix (A) → TOM → dissTOM → **Clustering**



Merge modules if eigengenes correlate highly (e.g., >0.75).

Note: Simulated data

Jorge A. Tzec-Interián | Gene Co-expression Network

WGCNA
work flow

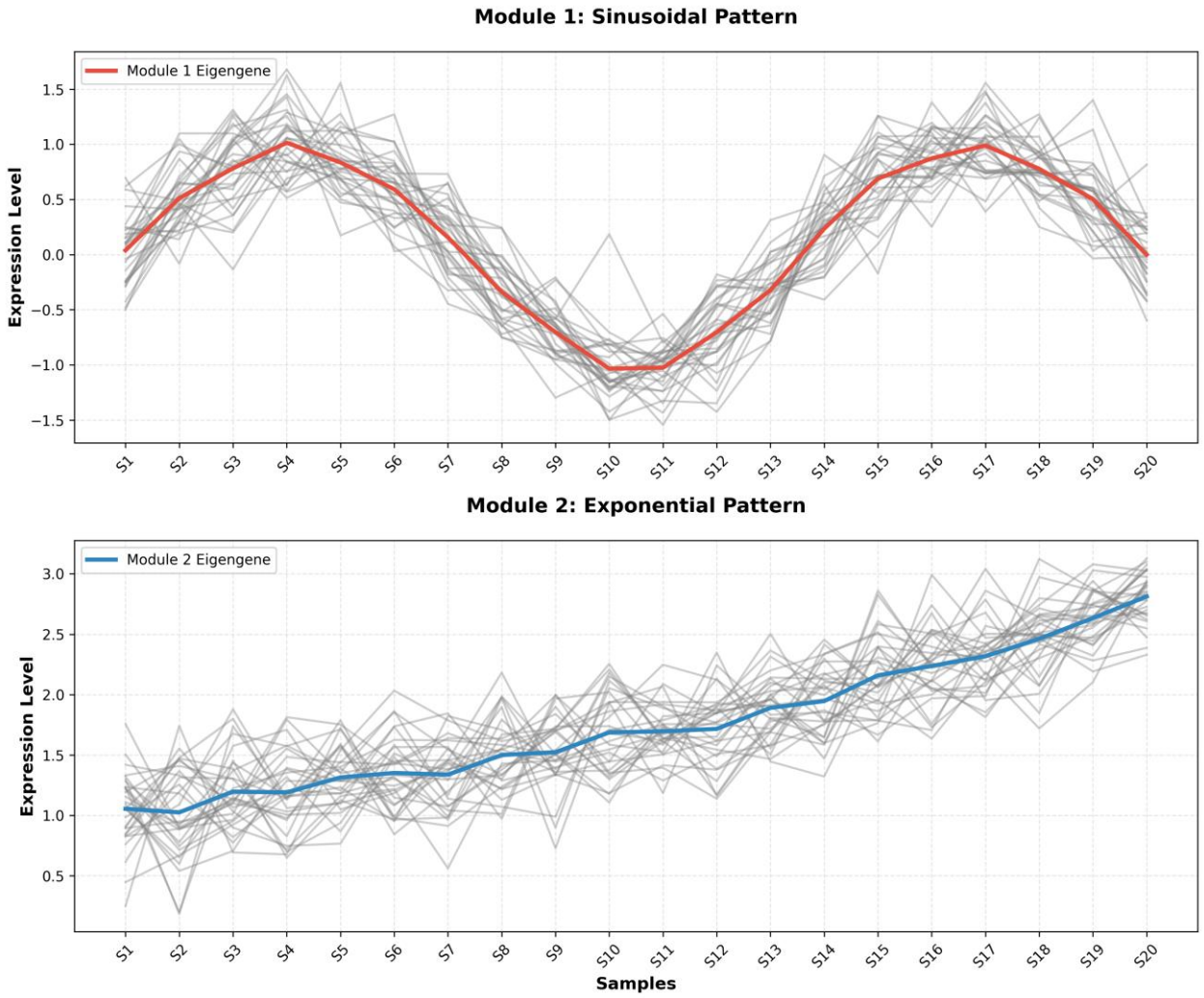
Expression Matrix (X) → Correlation Matrix (R) → Adjacency Matrix (A) → TOM → dissTOM → **Clustering**

The module eigengene is the first principal component (PC1) of the expression profiles of all genes in a module, obtained via PCA.

It summarizes the module's overall expression in a single vector.

It captures the dominant trend of gene expression.

Eigengene's pattern will track the dominant trend of the module's genes, aligning closely with genes that have high module membership (MM)



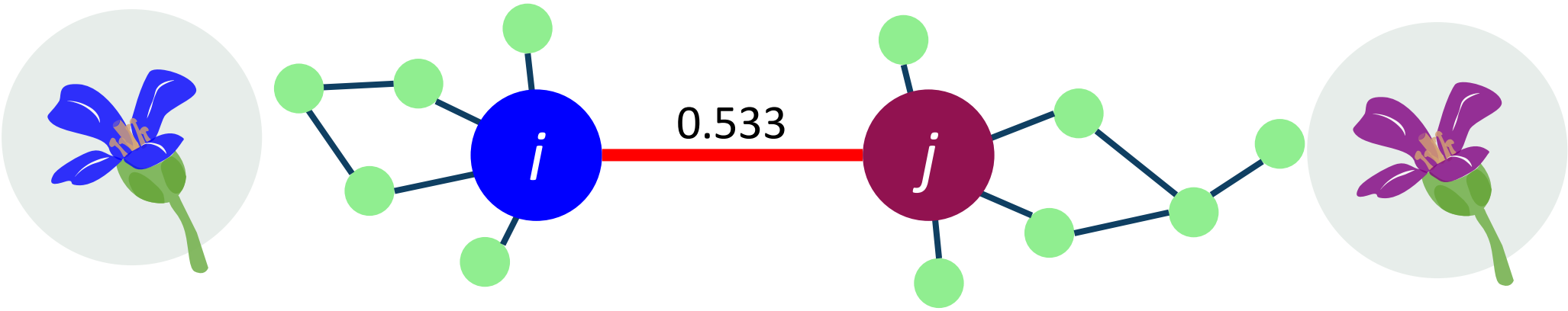
Langfelder & Horvath, 2008; Zhang & Horvath, 2005 Note: Simulated data

WGCNA: Module analysis

Langfelder & Horvath (2008), *BMC Bioinformatics*, 9:559
Zhang & Horvath (2005), *Stat Appl Genet Mol Biol*, 4:17.

Correlate **module eigengenes with external traits** (e.g., disease status, treatment) to identify biologically relevant modules.

METHOD: **Pearson correlation, p-values for significance**



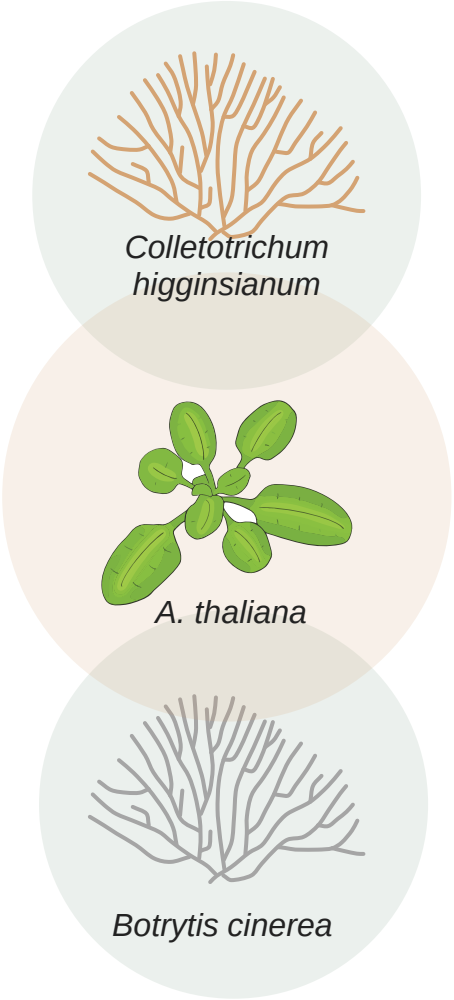
Result: Identifies modules linked to biological traits.

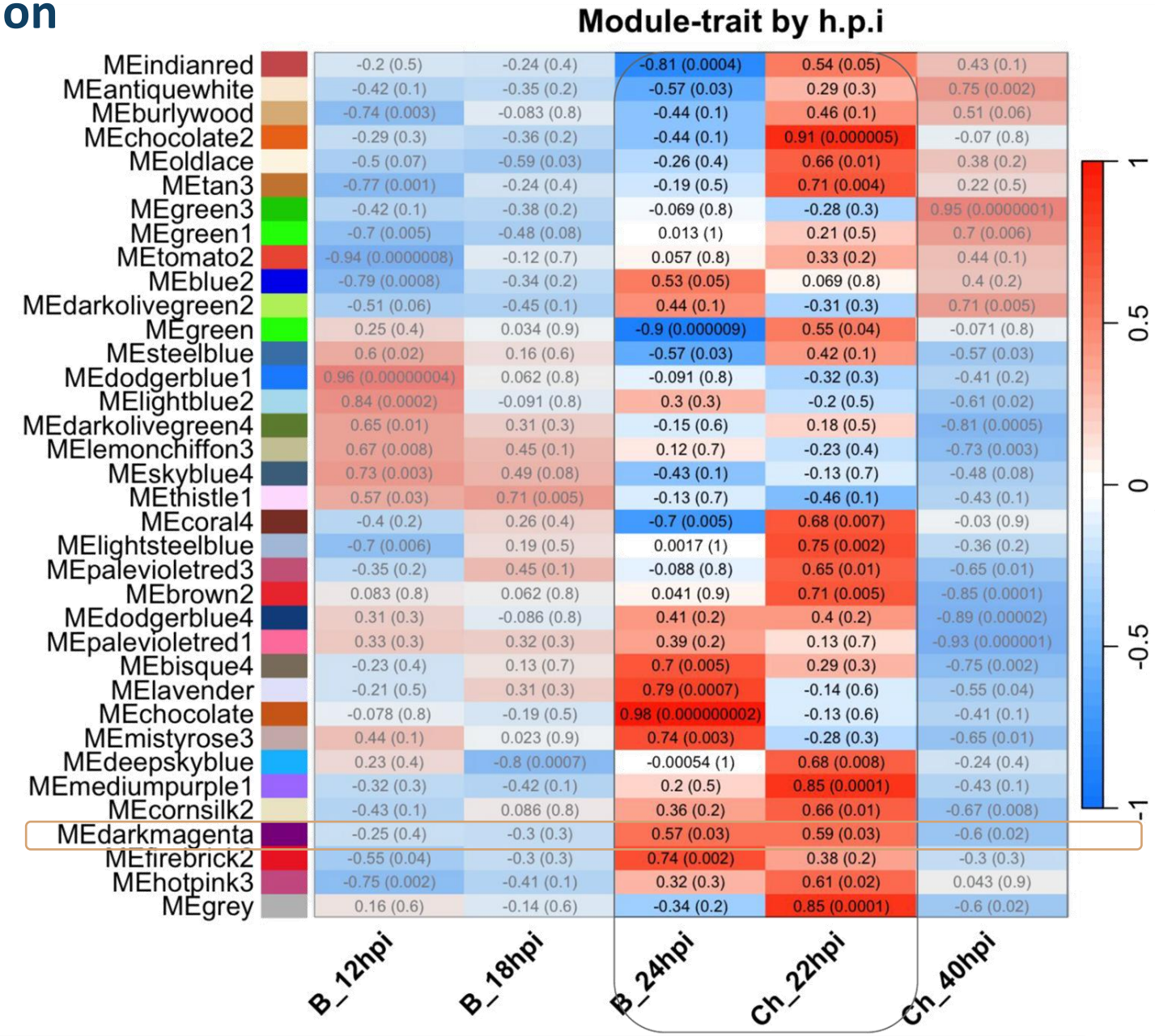
Modules with high correlation to a trait (e.g., cancer vs. healthy) are likely involved in that trait's biology.

Jorge A. Tzec-Interián | Gene Co-expression Network

Langfelder & Horvath, 2008; Zhang & Horvath, 2005 Note: Simulated data

Identification of core genes-modules in fungal infection





Jorge A. Tzec-Interián | Gene Co-expression Network

Soto-Cardinault *et al.*, 2023

Interpretation: Genes with high GS and high MM in a trait-correlated module are key candidates for biological roles.

Gene Significance (GS): Correlation between a gene's expression and a trait.

$$GS_i = |cor(X_i, trait)|$$

Module Membership (MM): Correlation between a gene's expression and its module's eigengene.

$$MM_i = |cor(X_i, ME_{module})|$$

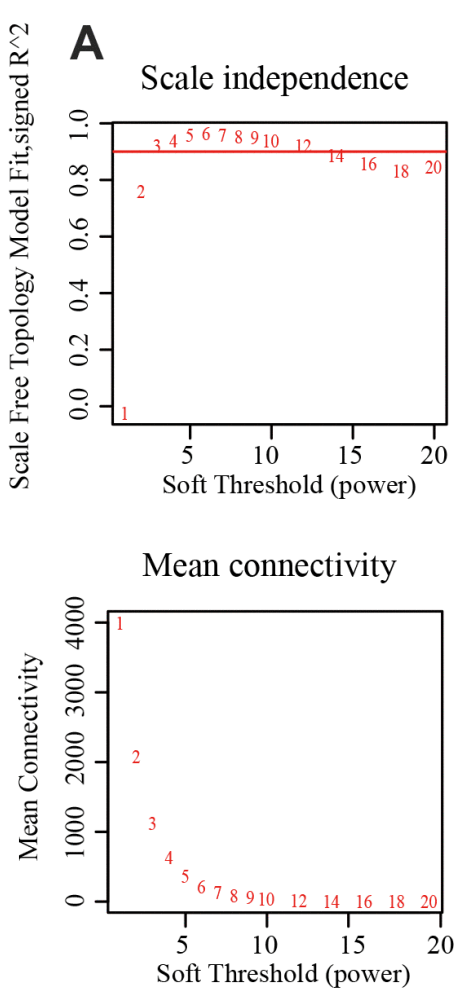
GS: How much a gene correlates with a trait.

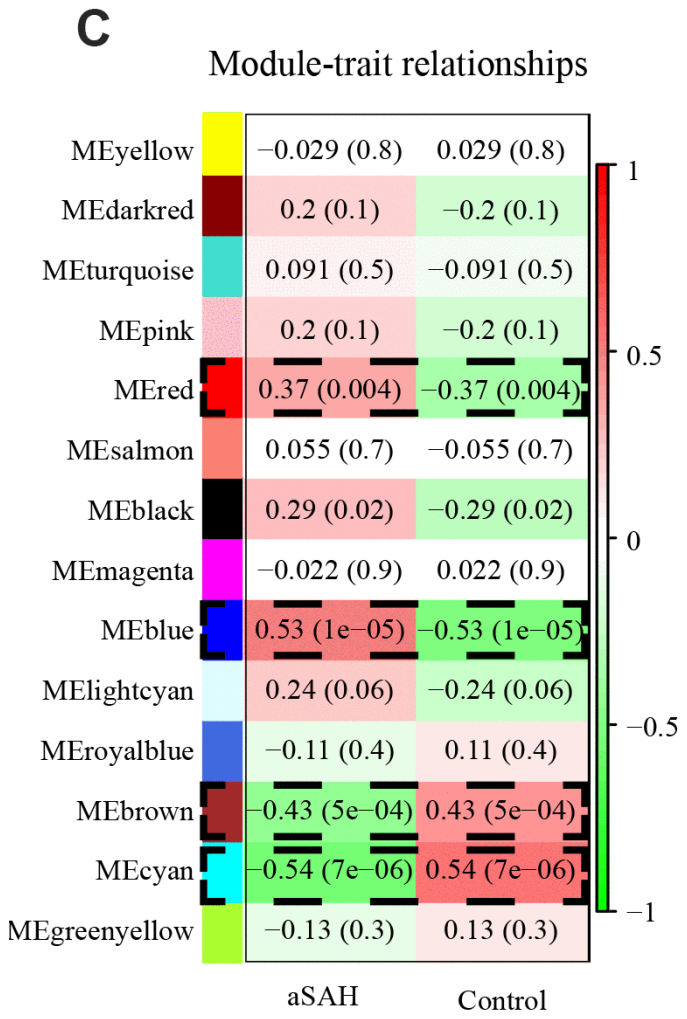
+

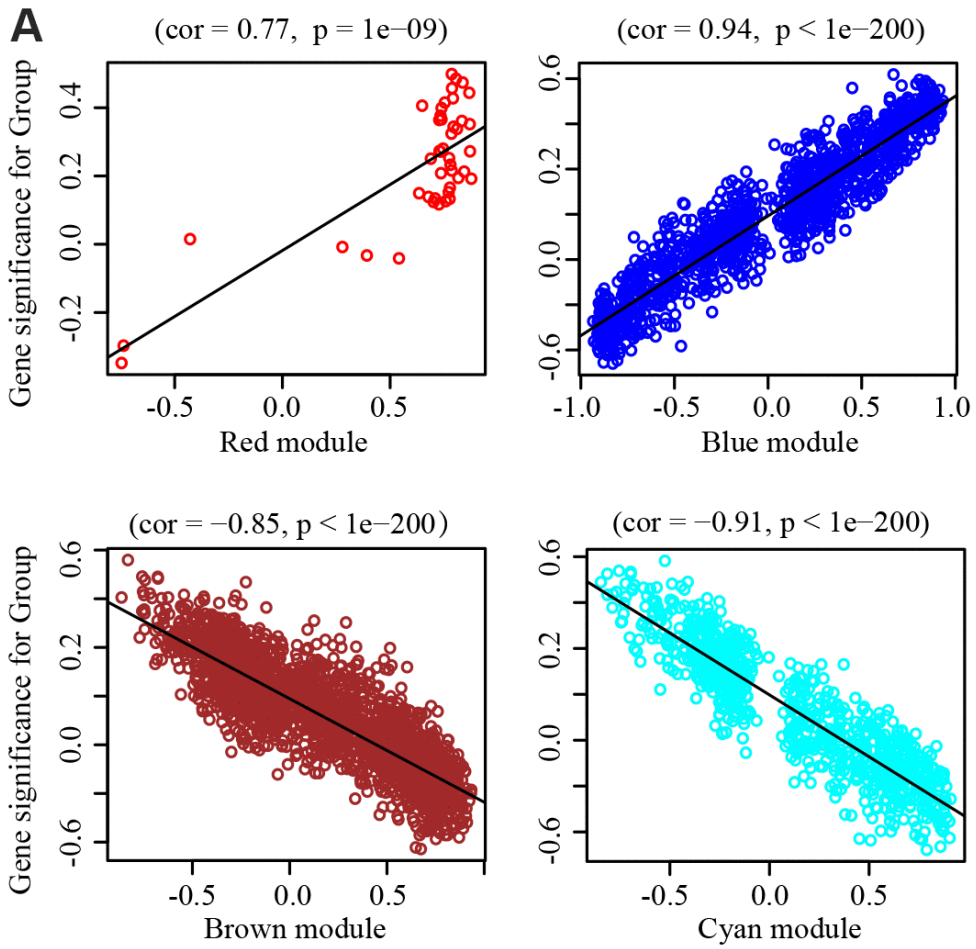
MM: How central a gene is to its module.

High GS + high MM = Important gene in trait-related module.

Examples: Identifying critical genes associated with aneurysmal subarachnoid hemorrhage by weighted gene co-expression network analysis



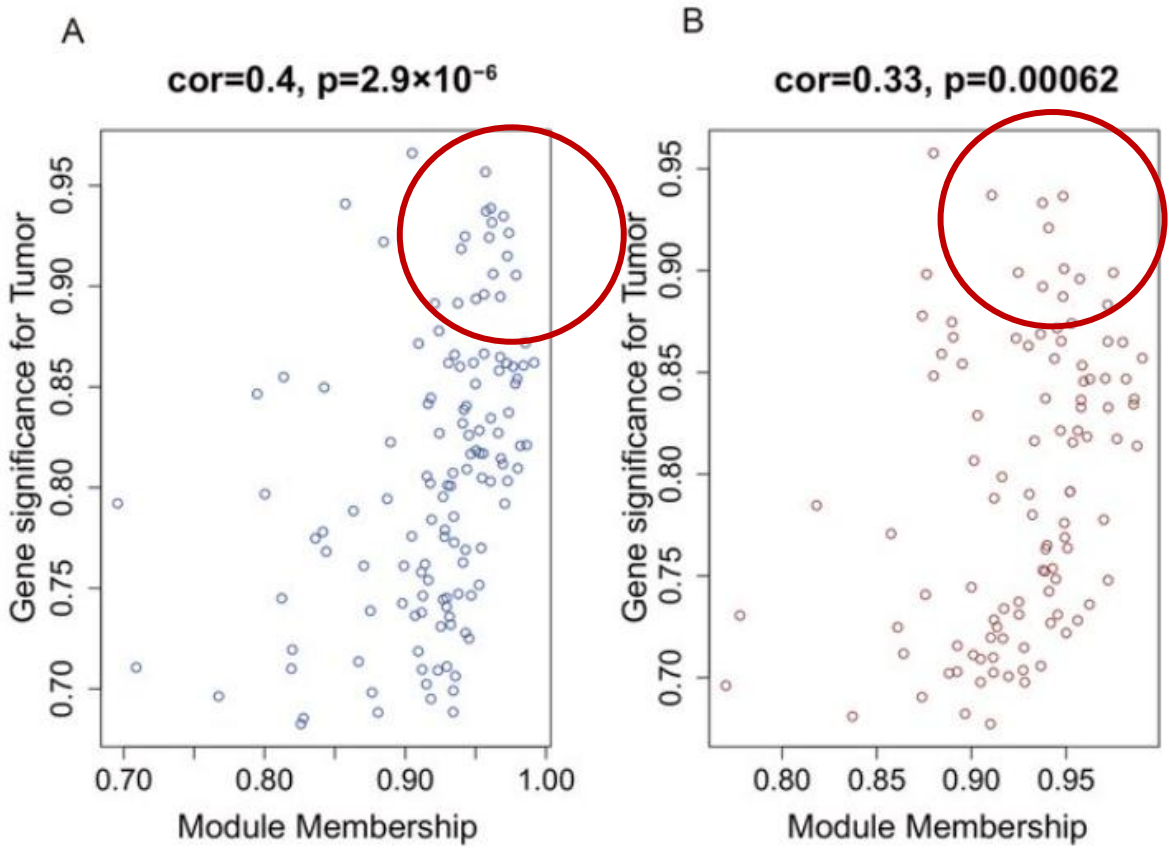


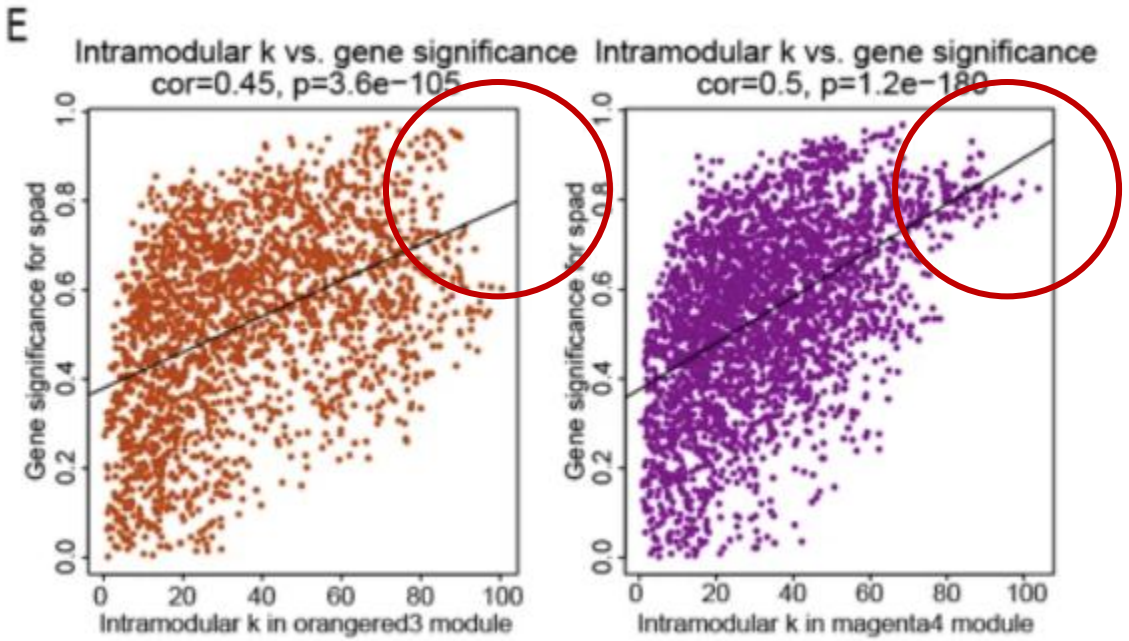


Jorge A. Tzec-Interián | Gene Co-expression Network

Yan et al., 2021 (10.18632/aging.203542)

More examples





Jorge A. Tzec-Interián | Gene Co-expression Network

Ma et al., 2025 (10.3389/fcell.2025.1554972); Qin et al., 2024 (10.1186/s12870-024-05450-5)

Intramodular Connectivity: Measures how strongly a gene connects to others in its module.

$$k_{within,i} = \sum_{j \in module, j \neq i} a_{ij}$$

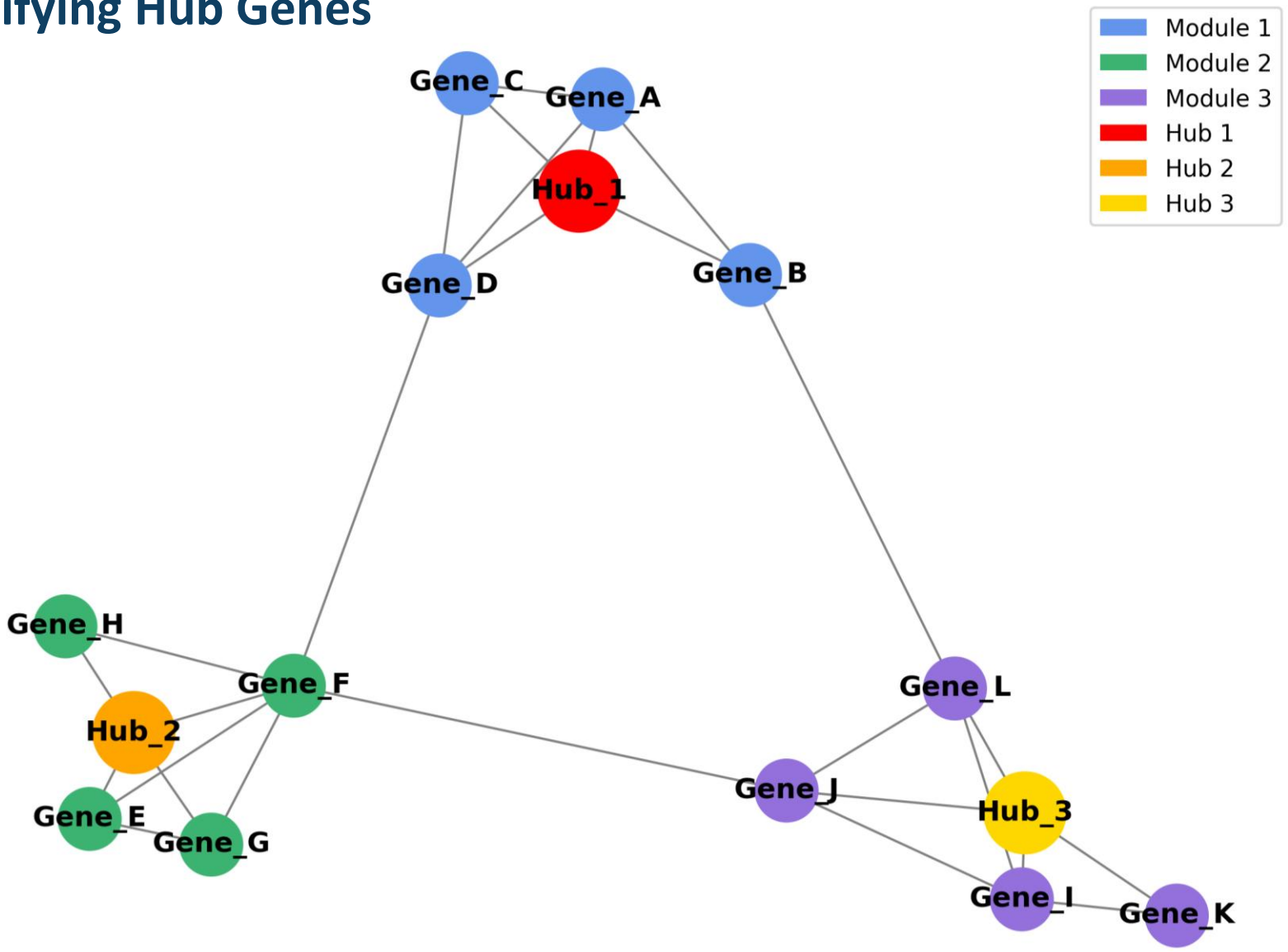
Role: Identifies **hub genes** central to a module's biology.

Hub Genes

- Select genes with high MM (e.g., top 10%) or high intramodular connectivity.
- Highly connected within a module.
- Top 10% MM or connectivity

Biological Insight

- Hubs are often functionally important (e.g., transcription factors in pathways).



Jorge A. Tzec-Interián | Gene Co-expression Network

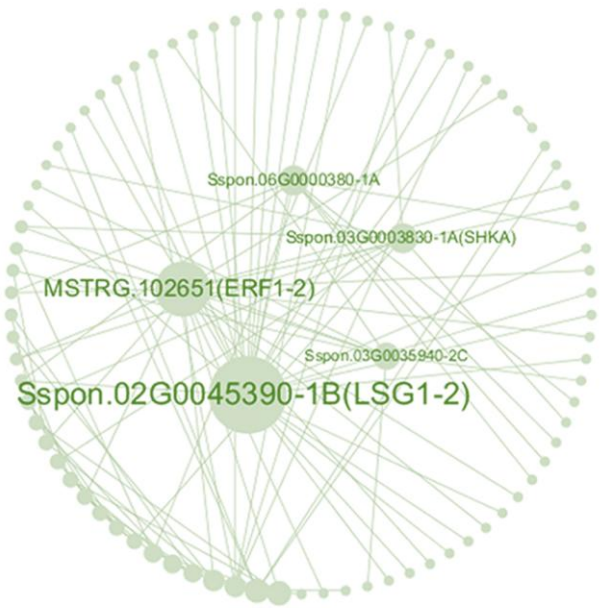
Note: Simulated data

Transcriptome and WGCNA reveal hub genes in sugarcane tiller seedlings in response to drought stress

A

Module–Trait relationships

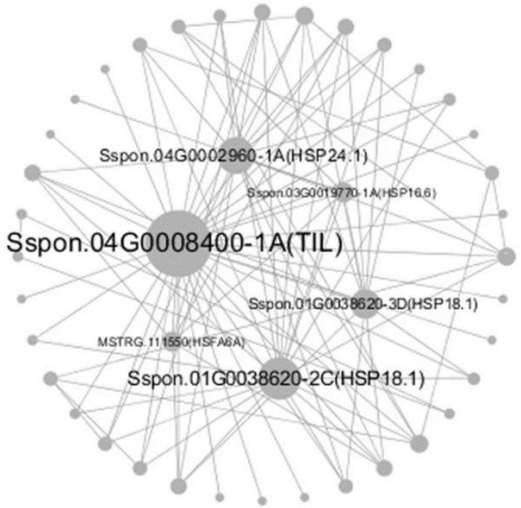
	IAA	ZT	GA3	ABA	SLs	SOD	POD	MDA	Pro	sucrose	protein	starch	height	diameter
MM.darkgrey	0.055 (0.8)	0.021 (0.9)	−0.31 (0.2)	0.0052 (1)	0.1 (0.7)	0.41 (0.09)	0.46 (0.05)	0.57 (0.01)	0.42 (0.08)	0.47 (0.05)	0.48 (0.04)	0.44 (0.07)	−0.45 (0.06)	−0.38 (0.1)
MM.white	0.061 (0.8)	−0.15 (0.5)	−0.26 (0.3)	−0.09 (0.7)	0.048 (0.9)	0.45 (0.06)	0.57 (0.01)	0.56 (0.02)	0.21 (0.4)	0.28 (0.3)	0.5 (0.04)	0.29 (0.2)	−0.47 (0.05)	−0.46 (0.05)
MM.darkseagreen4	−0.48 (0.05)	−0.42 (0.08)	−0.5 (0.03)	−0.39 (0.1)	−0.24 (0.3)	−0.29 (0.2)	0.16 (0.5)	0.11 (0.7)	−0.27 (0.3)	−0.25 (0.3)	−0.27 (0.3)	−0.19 (0.4)	−0.54 (0.02)	−0.014 (1)
MM.lightpink4	−0.37 (0.1)	−0.56 (0.02)	−0.41 (0.09)	−0.43 (0.07)	−0.16 (0.5)	−0.22 (0.4)	−0.11 (0.7)	0.18 (0.5)	−0.24 (0.3)	−0.21 (0.4)	−0.068 (0.8)	−0.18 (0.5)	−0.61 (0.008)	−0.029 (0.9)
MM.darkmagenta	0.25 (0.3)	0.25 (0.3)	−0.14 (0.6)	0.1 (0.7)	−0.061 (0.8)	0.41 (0.09)	0.42 (0.08)	0.072 (0.8)	0.39 (0.1)	0.34 (0.2)	0.28 (0.3)	0.33 (0.2)	0.14 (0.6)	−0.39 (0.1)
MM.ivory	0.19 (0.4)	0.16 (0.5)	−0.35 (0.2)	0.024 (0.9)	−0.028 (0.9)	0.5 (0.04)	0.63 (0.005)	0.36 (0.1)	0.47 (0.05)	0.48 (0.04)	0.42 (0.08)	0.44 (0.07)	−0.2 (0.4)	−0.49 (0.04)
MM.darkgreen	0.48 (0.05)	0.42 (0.08)	−0.018 (0.9)	0.39 (0.1)	0.28 (0.3)	0.82 (4e−05)	0.65 (0.004)	0.65 (0.003)	0.82 (3e−05)	0.82 (3e−05)	0.78 (1e−04)	0.83 (2e−05)	−0.085 (0.7)	−0.5 (0.03)
MM.grey60	0.55 (0.02)	0.32 (0.2)	0.31 (0.2)	0.44 (0.07)	0.31 (0.2)	0.72 (8e−04)	0.59 (0.01)	0.53 (0.02)	0.47 (0.05)	0.54 (0.02)	0.62 (0.007)	0.57 (0.01)	0.1 (0.7)	−0.26 (0.3)
MM.mediumorchid	−0.13 (0.6)	−0.54 (0.02)	−0.16 (0.5)	−0.3 (0.2)	0.029 (0.9)	0.036 (0.9)	0.19 (0.4)	0.22 (0.4)	−0.22 (0.4)	−0.2 (0.4)	0.095 (0.7)	−0.21 (0.4)	−0.48 (0.05)	−0.39 (0.1)
MM.thistle1	0.015 (1)	−0.41 (0.09)	0.16 (0.5)	−0.18 (0.5)	0.12 (0.6)	−0.19 (0.5)	−0.089 (0.7)	−0.21 (0.4)	−0.37 (0.1)	−0.4 (0.1)	−0.23 (0.4)	−0.44 (0.07)	−0.034 (0.9)	−0.12 (0.6)
MM.cyan	−0.66 (0.003)	−0.81 (4e−05)	−0.37 (0.1)	−0.61 (0.007)	−0.31 (0.2)	−0.73 (6e−04)	−0.54 (0.02)	−0.33 (0.2)	−0.7 (0.001)	−0.79 (8e−05)	−0.63 (0.005)	−0.69 (0.002)	−0.46 (0.06)	0.15 (0.6)
MM.honeydew1	−0.4 (0.1)	−0.28 (0.3)	−0.016 (1)	−0.23 (0.4)	−0.22 (0.4)	−0.61 (0.008)	−0.51 (0.03)	−0.57 (0.01)	−0.53 (0.02)	−0.64 (0.004)	−0.61 (0.007)	−0.58 (0.01)	0.085 (0.7)	0.2 (0.4)
MM.darkred	0.012 (1)	0.19 (0.5)	0.081 (0.7)	−0.09 (0.7)	−0.25 (0.3)	−0.16 (0.5)	−0.21 (0.4)	−0.24 (0.3)	−0.27 (0.3)	−0.068 (0.8)	−0.1 (0.7)	−0.18 (0.5)	0.27 (0.3)	0.54 (0.02)
MM.greenyellow	−0.032 (0.9)	0.33 (0.2)	0.21 (0.4)	−0.072 (0.8)	−0.35 (0.2)	−0.41 (0.09)	−0.38 (0.1)	−0.61 (0.007)	−0.4 (0.1)	−0.25 (0.3)	−0.42 (0.09)	−0.36 (0.1)	0.64 (0.004)	0.74 (5e−04)
MM.skyblue	0.44 (0.07)	0.39 (0.1)	0.13 (0.6)	0.3 (0.2)	0.32 (0.2)	0.42 (0.08)	0.13 (0.6)	0.25 (0.3)	0.74 (4e−04)	0.72 (7e−04)	0.48 (0.04)	0.6 (0.009)	0.16 (0.5)	−0.34 (0.2)
MM.coral2	0.21 (0.4)	0.3 (0.2)	0.27 (0.3)	0.22 (0.4)	0.2 (0.4)	−0.12 (0.6)	−0.35 (0.2)	−0.22 (0.4)	0.17 (0.5)	0.11 (0.7)	−0.16 (0.5)	0.062 (0.8)	0.44 (0.06)	0.35 (0.2)
MM.brown4	0.25 (0.3)	−0.13 (0.6)	0.15 (0.5)	−0.14 (0.6)	0.24 (0.3)	−0.096 (0.7)	−0.23 (0.4)	−0.16 (0.5)	0.0085 (1)	0.0052 (1)	−0.12 (0.6)	−0.15 (0.5)	0.24 (0.3)	0.13 (0.6)
MM.darkslateblue	0.12 (0.6)	0.022 (0.9)	0.32 (0.2)	−0.081 (0.7)	0.043 (0.9)	−0.43 (0.07)	−0.51 (0.03)	−0.59 (0.01)	−0.34 (0.2)	−0.33 (0.2)	−0.54 (0.02)	−0.45 (0.06)	0.59 (0.01)	0.56 (0.01)
MM.grey	−0.24 (0.3)	0.038 (0.9)	−0.34 (0.2)	−0.25 (0.3)	−0.44 (0.07)	−0.31 (0.2)	−0.33 (0.2)	−0.28 (0.3)	−0.083 (0.7)	−0.048 (0.8)	−0.28 (0.3)	−0.11 (0.7)	−0.034 (0.9)	−0.016 (1)



Positively correlated

Darkgreen

- LSG1-2
- ERF1-2
- SHKA



Grey60

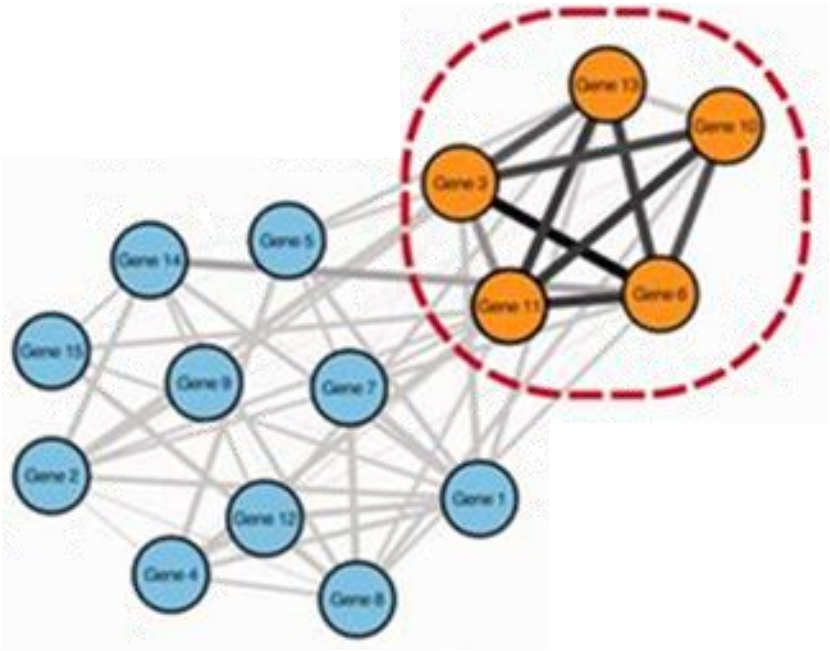
- HSP18.1
- HSP24.1
- HSFA6A

Jorge A. Tzec-Interián | Gene Co-expression Network

Tang et al., 2023 (10.1038/s41598-023-40006-x)

Term	Definition	Formula	Synonyms	Role	Scope
Total Connectivity (k_i)	Sum of a gene's connection strengths to all other genes.	$k_j = \sum_{j \neq i} a_{ij}$	None	Identifies global hub genes.	Whole network
Intramodule Connectivity($k_{within,i}$)	Sum of a gene's connection strengths within its module.	$k_{within,i} = \sum_{j \in module, j \neq i} a_{ij}$	k_{MI}	Identifies module-specific hub genes.	Within module
Module Membership ($k_{ME.i}$)	Correlation of a gene's expression with its module eigengene.	$k_{ME,i} = \lvert cor(X_i, ME_{module}) \rvert$	MM, k_{ME}	How central a gene is to its module.	Per gene
Module Eigengene (ME)	First principal component summarizing a module's expression.	PCA-based	None	Represents module expression for trait correlation and merging.	Per module
Gene Significance (GS)	Correlation of a gene's expression with a trait.	$GS_i = \lvert cor(X_i, trait) \rvert$	None	How much a gene correlates with a trait.	Per gene

Langfelder & Horvath, 2008; Zhang & Horvath, 2005



What biological functions or processes are enriched in the module?	Gene Ontology enrichment analysis
What transcription factors are present?	Transcription factors identification
Do the genes share any metabolic pathways?	KEGG pathways overrepresentation
Are all the genes known?	Functional annotation (BLAST)
What domains are common among the genes?	Domains and motifs identification

Network visualization

Edges

fromNode	toNode	weight	direction	fromAltName	toAltName
Cre01.g002500	Cre01.g004300	0.0472115097377545	undirected	Cre01.g002500	Cre01.g004300
Cre01.g002500	Cre01.g007200	0.075540941421398	undirected	Cre01.g002500	Cre01.g007200
Cre01.g002500	Cre01.g012450	0.077157660339242	undirected	Cre01.g002500	Cre01.g012450
Cre01.g002500	Cre01.g014050	0.0298752721040778	undirected	Cre01.g002500	Cre01.g014050
Cre01.g002500	Cre01.g014900	0.049277455778783	undirected	Cre01.g002500	Cre01.g014900
Cre01.g002500	Cre01.g024050	0.0652546097539393	undirected	Cre01.g002500	Cre01.g024050
Cre01.g002500	Cre01.g024750	0.0305700084946545	undirected	Cre01.g002500	Cre01.g024750
Cre01.g002500	Cre01.g033700	0.0474965532771704	undi		
Cre01.g002500	Cre01.g034150	0.021413152373439	undi		
Cre01.g002500	Cre01.g034950	0.10183455802104	undi		
Cre01.g002500	Cre01.g039900	0.0294571612332167	undi		
Cre01.g002500	Cre01.g040100	0.0336744459548217	undi		
Cre01.g002500	Cre01.g044850	0.0827187184591748	undi		
Cre01.g002500	Cre01.g046237	0.0602551595748371	undi		
Cre01.g002500	Cre01.g048400	0.0268996366558188	undi		
Cre01.g002500	Cre01.g053000	0.110454190957158	undi		

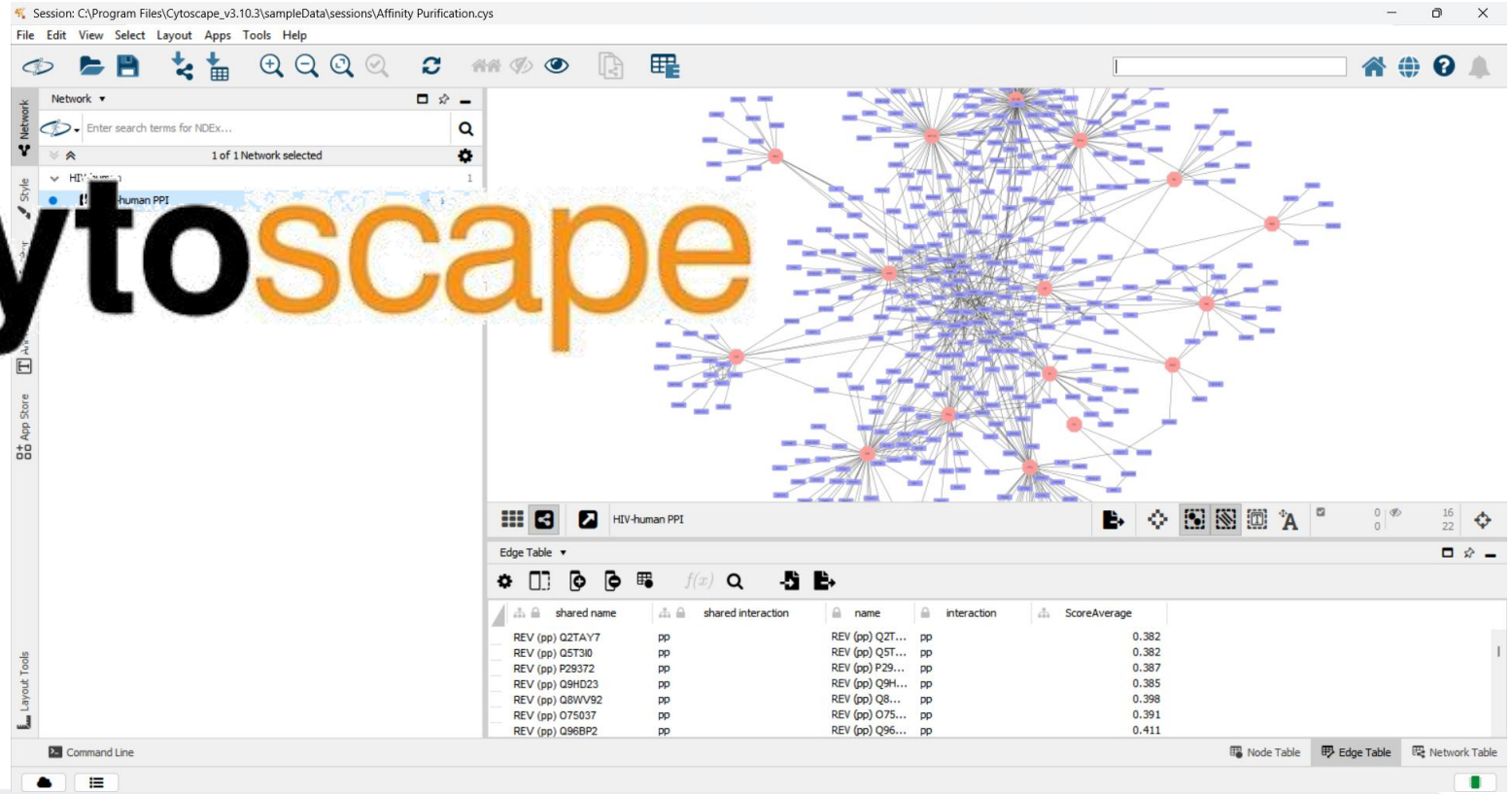
Nodes

"fromNode"	"toNode"	"weight"
"Cre01.g002500"	"Cre01.g004300"	0.0472115097377545
"Cre01.g002500"	"Cre01.g007200"	0.075540941421398
"Cre01.g002500"	"Cre01.g012450"	0.077157660339242
"Cre01.g002500"	"Cre01.g014050"	0.0298752721040778
"Cre01.g002500"	"Cre01.g014900"	0.049277455778783
"Cre01.g002500"	"Cre01.g024050"	0.0652546097539393
"Cre01.g002500"	"Cre01.g024750"	0.0305700084946545
"Cre01.g002500"	"Cre01.g033700"	0.0474965532771704
"Cre01.g002500"	"Cre01.g034150"	0.021413152373439
"Cre01.g002500"	"Cre01.g034950"	0.10183455802104
"Cre01.g002500"	"Cre01.g039900"	0.0294571612332167
"Cre01.g002500"	"Cre01.g040100"	0.0336744459548217
"Cre01.g002500"	"Cre01.g044850"	0.0827187184591748

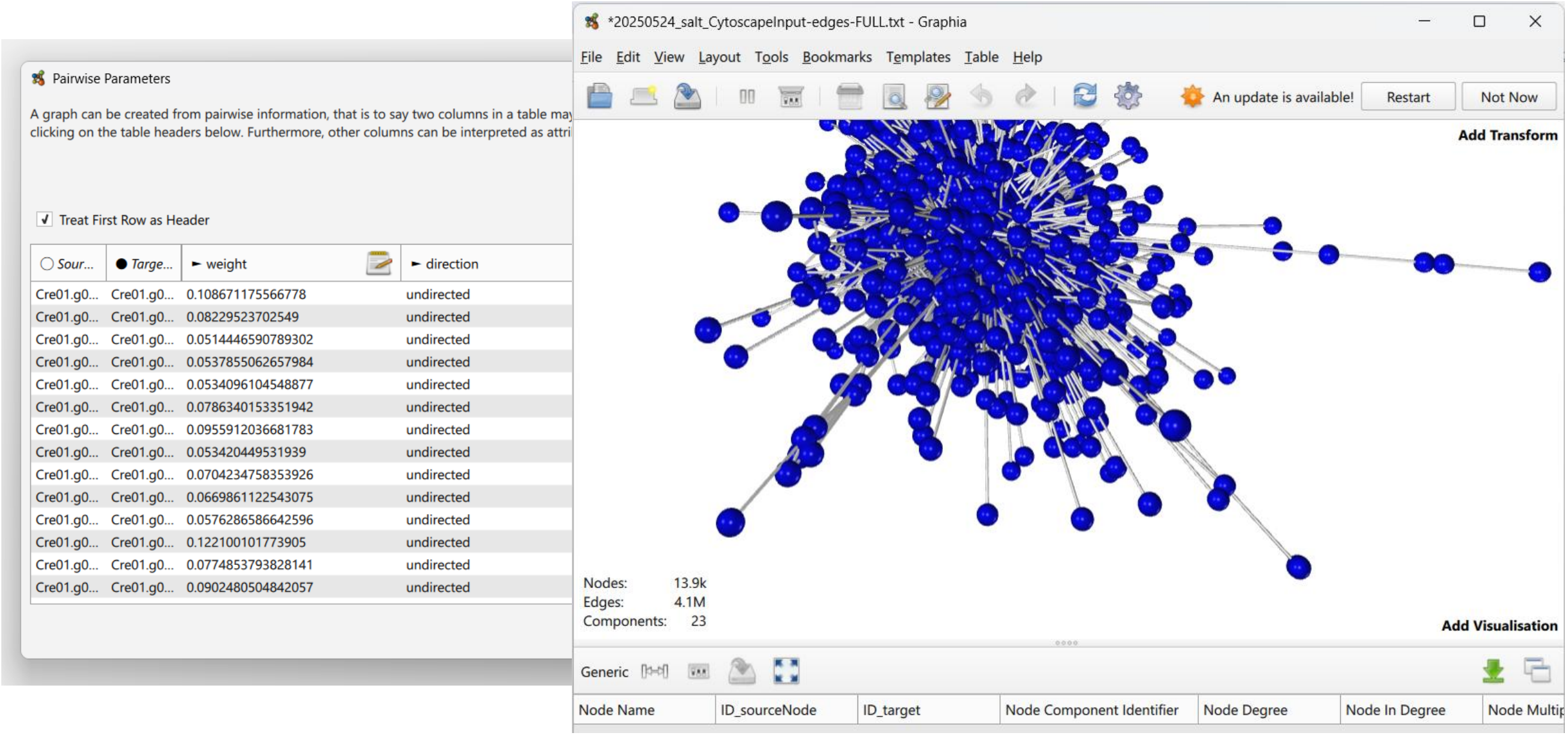
1. Export: Node (genes, MM, GS) and edge (adjacency/TOM) files (exportNetworkToCytoscape).

2. Visualize: Color by module or connectivity, use layouts.

3. Annotate: Add GO/KEGG terms for function.



https://web.cytoscape.org/


Pairwise Parameters
A graph can be created from pairwise information, that is to say two columns in a table may
clicking on the table headers below. Furthermore, other columns can be interpreted as attri
Treat First Row as Header
Sour... Targe... weight direction
Cre01.g0... Cre01.g0... 0.108671175566778 undirected
Cre01.g0... Cre01.g0... 0.08229523702549 undirected
Cre01.g0... Cre01.g0... 0.0514446590789302 undirected
Cre01.g0... Cre01.g0... 0.0537855062657984 undirected
Cre01.g0... Cre01.g0... 0.0534096104548877 undirected
Cre01.g0... Cre01.g0... 0.0786340153351942 undirected
Cre01.g0... Cre01.g0... 0.0955912036681783 undirected
Cre01.g0... Cre01.g0... 0.053420449531939 undirected
Cre01.g0... Cre01.g0... 0.0704234758353926 undirected
Cre01.g0... Cre01.g0... 0.0669861122543075 undirected
Cre01.g0... Cre01.g0... 0.0576286586642596 undirected
Cre01.g0... Cre01.g0... 0.122100101773905 undirected
Cre01.g0... Cre01.g0... 0.0774853793828141 undirected
Cre01.g0... Cre01.g0... 0.0902480504842057 undirected
*20250524_salt_CytoscapeInput-edges-FULL.txt - Graphia
File Edit View Layout Tools Bookmarks Templates Table Help
An update is available!
Restart
Not Now
Add Transform
Nodes: 13.9k
Edges: 4.1M
Components: 23
Add Visualisation
Generic
Node Name
ID_sourceNode
ID_target
Node Component Identifier
Node Degree
Node In Degree
Node Multip

R/igraph

Program	Interface	Language/Platform	Online/Installed	WGCNA Input Compatibility	Role & Features	Official Link
WGCNA (R)	CLI	R	Installed	Adjacency matrix, TOM, edge lists, module assignments, node attributes (kME, GS)	Built-in visualization (e.g., TOMplot, plotDendroAndColors) for heatmaps, dendrograms. Exports to Cytoscape/VisANT. Limited static plots.	CRAN WGCNA
igraph	CLI	R, Python, C/C++	Installed	Edge lists, adjacency matrices, node attributes (e.g., module colors, kME)	Versatile for network analysis/visualization. Supports custom attributes. High-performance for large networks.	igraph R, igraph Python
NetworkX	CLI	Python	Installed	Edge lists, adjacency matrices, node attributes	Easy-to-use Python library for network creation/analysis. Supports WGCNA edge lists. Slower for large networks.	NetworkX PyPI
Graph-tool	CLI	Python, C++	Installed	Edge lists, adjacency matrices	High-performance C++ library for statistical network analysis. Visualizes WGCNA networks via edge weights. Steeper learning curve.	Graph-tool
Tabnetviz	CLI	Python, Graphviz	Installed	Edge lists, node/edge tables	Lightweight CLI tool using Graphviz for SVG visualizations. Maps WGCNA node attributes to visual properties.	Tabnetviz, Python version
Cytoscape	GUI	Java	Installed	Edge lists, node tables (e.g., module assignments, kME, GS)	Open-source platform for complex network visualization. Imports WGCNA via edge lists or R. Supports GO/KEGG.	Cytoscape
Graphia	GUI	Java	Installed	Edge lists, node tables	Open-source tool for interactive network exploration. Visualizes WGCNA networks with custom layouts. Handles large and complex datasets.	Graphia
Gephi	GUI	Jav	Installed	Edge lists, node tables	Open-source tool for interactive network exploration. Visualizes WGCNA networks with custom layouts. Suitable for large networks.	Gephi
Omics Playground (WGCNA Module)	GUI	Web-based	Online	Expression data (processes WGCNA internally)	User-friendly platform for WGCNA analysis/visualization. Limited customization. Requires registration.	Omics Playground
PyWGCNA	CLI	Python	Installed	Edge lists, module assignments, node attributes	Python WGCNA implementation with visualization (e.g., module-trait heatmaps, PPI networks). Supports functional enrichment.	PyWGCNA PyPI, GitHub

WGCNA applications

Citation	Scientific Area	Biological System	Number of Samples	Main Findings	WGCNA Application
Wickramasuriya et al., 2021	Animal Science	Bovine mammary glands (lactation)	79 samples	Identified 104, 85, and 26 DEGs across lactation stages. Hub genes (GJA1, AP2A2, NPAS3) regulated lactation. Enriched pathways: ubiquitin-dependent ERAD, chaperone protein refolding.	Constructed co-expression networks for lactation stages, identified modules, hub genes, and enriched pathways using WGCNA combined with meta-analysis.
Yin et al.,2019	Human Health (Oncology)	Human tongue squamous cell carcinoma (TSCC)	57 (TSCC vs. normal)	Identified 9 hub genes (e.g., AIMP2, ANXA2) linked to poor prognosis. Enriched pathways: olfactory transduction, neuroactive ligand-receptor interaction.	Used WGCNA to cluster DEGs into modules, identified key modules correlated with TSCC, and screened hub genes for prognostic analysis.
Yang et al., 2018	Human Health (Oncology)	Human oral squamous cell carcinoma (OSCC, miRNA)	80 (40 OSCC, 40 matched non-tumorous)	Identified two novel miRNA modules associated with OSCC, regulating hub nodes in metabolic networks.	Modified WGCNA pipeline for paired design to reduce batch effects, identified miRNA co-expression modules.
Liu et al., 2017	Human Health (Oncology)	Human osteosarcoma	44 (22 osteosarcoma, 22 normal)	Identified modules enriched in cell cycle, DNA replication. Hub genes (e.g., CDK1, CCNB1) linked to osteosarcoma progression.	Used WGCNA to construct co-expression networks, identify modules, and screen hub genes from DEGs.
Zhang et al., 2023	Human Health (Microbiome)	Human tumor microbiome (TCGA)	1,000+ (multiple tumor types)	Identified oncogenome-associated microbiome modules (OAM). Tissierella linked to tumor suppression.	Built separate WGCNA networks for microbiome and mRNA, correlated modules to identify OAM and prognostic biomarkers.
Wei et al., 2020	Human Health (Oncology)	Human lung adenocarcinoma	517 (TCGA RNA-seq)	Identified modules enriched in immune response, cell adhesion. Hub genes (e.g., CDH1) linked to tumor progression.	Applied WGCNA to RNA-seq data, identified co-expression modules, and correlated with clinical traits.
Eidsaa et al., 2017	Comparative Genomics	Human and mouse tissues	76 human, 66 mouse (multiple tissues)	Enriched transcription factors (e.g., homeobox) in central genes. Disease genes (FOXE3, RHO) highly central.	Compared WGCNA networks using topological overlap and s-core analysis to identify conserved gene roles.
Zhu et al., 2022	Plant Biology (Pathogen Resistance)	Cassava (*Manihot esculenta*) roots infected with Xanthomonas axonopodis	18 (6 time points post-infection, 3 replicates)	Identified MeWRKY IIa genes (e.g., MeWRKY27) as hub genes regulating resistance to cassava bacterial blight. Enriched pathways: plant-pathogen interaction, MAPK signaling.	Used WGCNA to construct co-expression networks from RNA-seq data, identified modules correlated with pathogen response, and screened hub genes.
Liu et al., 2021	Plant Biology (Abiotic Stress)	Maize (Zea mays) leaves under drought stress	24 (4 growth stages, 3 replicates, control vs. drought)	Identified 3,451 DEGs and 12 modules. Hub genes (e.g., HSP90, LEA proteins) linked to drought tolerance via ABA and stress defense pathways.	Applied WGCNA to RNA-seq data, identified modules correlated with drought stress at different growth stages, and detected hub genes.

WGCNA applications

Citation	Scientific Area	Biological System	Number of Samples	Main Findings	WGCNA Application
Wickramasuriya et al., 2021	Animal Science	Bovine mammary glands (lactation)	79 samples	Identified 104, 85, and 26 DEGs across lactation stages. Hub genes (GJA1, AP2A2, NPAS3) regulated lactation. Enriched pathways: ubiquitin-dependent ERAD, chaperone protein refolding.	Constructed co-expression networks for lactation stages, identified modules, hub genes, and enriched pathways using WGCNA combined with meta-analysis.
Yin et al.,2019	Human Health (Oncology)	Human tongue squamous cell carcinoma (TSCC)	57 (TSCC vs. normal)	Identified 9 hub genes (e.g., AIMP2, ANXA2) linked to poor prognosis. Enriched pathways: olfactory transduction, neuroactive ligand-receptor interaction.	Used WGCNA to cluster DEGs into modules, identified key modules correlated with TSCC, and screened hub genes for prognostic analysis.
Yang et al., 2018	Human Health (Oncology)	Human oral squamous cell carcinoma (OSCC, miRNA)	80 (40 OSCC, 40 matched non-tumorous)	Identified two novel miRNA modules associated with OSCC, regulating hub nodes in metabolic networks.	Modified WGCNA pipeline for paired design to reduce batch effects, identified miRNA co-expression modules.
Liu et al., 2017	Human Health (Oncology)	Human osteosarcoma	44 (22 osteosarcoma, 22 normal)	Identified modules enriched in cell cycle, DNA replication. Hub genes (e.g., CDK1, CCNB1) linked to osteosarcoma progression.	Used WGCNA to construct co-expression networks, identify modules, and screen hub genes from DEGs.
Zhang et al., 2023	Human Health (Microbiome)	Human tumor microbiome (TCGA)	1,000+ (multiple tumor types)	Identified oncogenome-associated microbiome modules (OAM). Tissierella linked to tumor suppression.	Built separate WGCNA networks for microbiome and mRNA, correlated modules to identify OAM and prognostic biomarkers.
Wei et al., 2020	Human Health (Oncology)	Human lung adenocarcinoma	517 (TCGA RNA-seq)	Identified modules enriched in immune response, cell adhesion. Hub genes (e.g., CDH1) linked to tumor progression.	Applied WGCNA to RNA-seq data, identified co-expression modules, and correlated with clinical traits.
Eidsaa et al., 2017	Comparative Genomics	Human and mouse tissues	76 human, 66 mouse (multiple tissues)	Enriched transcription factors (e.g., homeobox) in central genes. Disease genes (FOXE3, RHO) highly central.	Compared WGCNA networks using topological overlap and s-core analysis to identify conserved gene roles.
Zhu et al., 2022	Plant Biology (Pathogen Resistance)	Cassava (*Manihot esculenta*) roots infected with Xanthomonas axonopodis	37 transcriptomic datasets	Identified MeWRKY IIa genes (e.g., MeWRKY27) as hub genes regulating resistance to cassava bacterial blight. Enriched pathways: plant-pathogen interaction, MAPK signaling.	Used WGCNA to construct co-expression networks from RNA-seq data, identified modules correlated with pathogen response, and screened hub genes.
Liu et al., 2021	Plant Biology (Abiotic Stress)	Maize (Zea mays) leaves under drought stress	24 (4 growth stages, 3 replicates, control vs. drought)	Identified 3,451 DEGs and 12 modules. Hub genes (e.g., HSP90, LEA proteins) linked to drought tolerance via ABA and stress defense pathways.	Applied WGCNA to RNA-seq data, identified modules correlated with drought stress at different growth stages, and detected hub genes.

Tools Similar to WGCNA or Based on WGCNA

Tool	Interface	Language/Platform	Main Features	Official Download Link
WGCNA	CLI	R	Constructs co-expression networks using weighted correlations.	CRAN WGCNA
BioNERO	CLI	R	All-in-one package for network inference. Uses WGCNA for co-expression networks (GCNs). Infers gene regulatory networks (GRNs) via CLR, GENIE3, ARACNe. Includes preprocessing, PCA, heatmaps, and topological analysis. User-friendly for non-experts.	Bioconductor BioNERO
multiWGCNA	CLI	R	Extends WGCNA for multidimensional datasets (e.g., disease + time/space). Constructs networks at three levels: combined, disease/wildtype, secondary trait. Maps modules across levels. Analyzes differential module expression and preservation.	CRAN multiWGCNA
PyWGCNA	CLI	Python	Python implementation of WGCNA. Faster than R WGCNA for large datasets. Includes module-trait correlation, hub gene detection, and functional enrichment (GO, KEGG, REACTOME). Compares modules across datasets (e.g., AD models).	PyPI PyWGCNA, GitHub
CEMiTool	CLI	R	Infers co-expression modules using WGCNA-like correlation-based methods. Integrates functional enrichment (GO, KEGG) and interaction data (e.g., PPI). Generates interactive HTML reports. Simpler interface than WGCNA for beginners.	Bioconductor CEMiTool
iterativeWGCNA	CLI	Python, R	Python-wrapped extension of WGCNA. Improves module robustness by iterative refinement. Increases gene inclusion in modules. Supports custom WGCNA parameters (e.g., power, merge cut height). Requires R WGCNA.	PyPI iterativeWGCNA, GitHub
GEN (WGCNA Module)	GUI	Web-based	Online platform for bulk RNA-seq analysis. Includes WGCNA for co-expression networks. Offers differential expression, enrichment, and GRN inference. No coding required. Limited customization.	GEN Tools
Exp Omics	GUI	Web-based	ExpOmics is a suite of six user-friendly web apps with 28 advanced functions for multi-omics expression analysis, including WGCNA transcriptome analysis. Limited customization.	ExpOmics
Gene Expression Nebulas	GUI	Web-based	Tools in aid of further downstream analysis of both bulk and single-cell RNA-seq data, including WGCNA. Limited customization.	Gene Expression Nebulas

Gene co-expression network vs gene regulation network

Similarities

- **Network Structure**: Both represent biological relationships as graphs with nodes (genes) and edges (interactions).
- **Biological Insight**: Identify functional modules or pathways (e.g., co-expressed genes in a disease or regulatory cascades).
- **Data-Driven**: Rely on high-throughput data (e.g., RNA-seq, ChIP-seq) to infer relationships.
- **Visualization**: Use tools like Cytoscape or igraph to visualize networks.

Network type	Basis of Edges	Directionality	Data Input	Biological Meaning	Inference Method
Co-expression	Edges based on correlation of expression profiles (e.g., Pearson correlation raised to power beta). No causality implied.	Undirected (symmetric relationships)	Gene expression data (e.g., RNA-seq counts).	Groups genes with similar expression patterns (e.g., modules in WGCNA).	Statistical (e.g., correlation, TOM in WGCNA).
Regulatory	Edges represent causal interactions (e.g., transcription factor binds to gene promoter), often inferred from binding data or motifs	Directed (e.g., transcription factor → target gene).	Binding data (e.g., ChIP-seq), motif databases, or expression + prior knowledge.	Models mechanistic control (e.g., gene regulation in a pathway).	Integrative (e.g., Bayesian networks, motif scanning, or machine learning).

Tool	Interface	Language/Platform	Input Data	Key Features	Official Link
ARACNe	CLI	R, Python	Gene expression (e.g., RNA-seq)	Infers regulatory networks using mutual information. Removes indirect interactions via DPI. Suitable for large datasets.	ARACNe GitHub
GENIE3	CLI	R, Python	Gene expression	Machine learning (random forests) to infer directed regulatory interactions. High accuracy for transcription factor targets. Integrates with WGCNA outputs.	GENIE3 GitHub Bioconductor
SCENIC	CLI	Python, R	Expression + motif/CisBP data	Combines expression with motif analysis to infer regulons (transcription factor + targets). Visualizes via heatmaps or t-SNE. Single-cell focus.	SCENIC
iRegulon	GUI, CLI	Cytoscape plugin	Expression, motif databases	Predicts transcription factors via motif enrichment. Integrates with Cytoscape for visualization. User-friendly GUI.	iRegulon
TRRUST	GUI	Web-based	Curated regulatory data	Database of human/mouse transcription factor-target interactions. Visualizes networks online. Limited to known interactions.	TRRUST
GeNeCK	GUI	Web-based	Gene expression, transcription factors	A comprehensive online toolkit for constructing gene networks using expression data and optional hub gene input.	GeNeCK

Gracias por su atención

Csardi G, Nepusz T (2006). The igraph software package for complex network research. *InterJournal*, Complex Systems, 1695. https://igraph.org

EMBL-EBI (n.d.). *Network analysis of protein interaction data – An introduction*. https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/

Horvath S (2011). Weighted Network Analysis: Applications in Genomics and Systems Biology. Springer Science & Business Media. https://doi.org/10.1007/978-1-4419-8819-5

Langfelder P, Horvath S (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559. https://doi.org/10.1186/1471-2105-9-559

Ma X, Zhou H, Wang R (2025). Integration of machine learning and bulk sequencing reveals FOSB as an exosome-related gene in abdominal aortic aneurysm. *Front Cell Dev Biol*, **13**, 1554972.

Orchard S, *et al.* (2014). The MIntAct project—IntAct as a common curation platform. *Nucleic Acids Res*, **42**, D358–D363. https://doi.org/10.1093/nar/gkt1115

Qin H, Guo J, Jin Y, *et al.* (2024). Integrative analysis of transcriptome and metabolome provides insights into leaf variegation in *Heliopsis helianthoides*. *BMC Plant Biol*, **24**, 731. https://doi.org/10.1186/s12870-024-05450-5

Soto-Cardinault C, Childs KL, Góngora-Castillo E (2023). Network analysis of public RNA-seq data reveals plant defense mechanisms in *Arabidopsis*. *Genes*, **14**(12), 2223. https://doi.org/10.3390/genes14122223

Szklarczyk D, *et al.* (2023). The STRING database in 2023: functional enrichment of protein networks. *Nucleic Acids Res*, **51**, D638–D646. https://doi.org/10.1093/nar/gkac1034

Tang Y, Li J, Song Q, *et al.* (2023). Transcriptome and WGCNA reveal hub genes in sugarcane tiller seedlings in response to drought stress. *Sci Rep*, **13**, 12823. https://doi.org/10.1038/s41598-023-40006-x

Van Dam S, Vosa U, van der Graaf A, Franke L, de Magalhaes JP (2018). Gene co-expression analysis for functional classification and disease prediction. *Brief Bioinform*, **19**(4), 575–592. https://doi.org/10.1093/bib/bbw139

Yan Z, Wu Q, Cai W, *et al.* (2021). Identifying critical genes in aneurysmal subarachnoid hemorrhage via WGCNA. *Aging (Albany NY)*, **13**(18), 22345.

Zhang B, Horvath S (2005). A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*, **4**, Article 17. https://doi.org/10.2202/1544-6115.1128