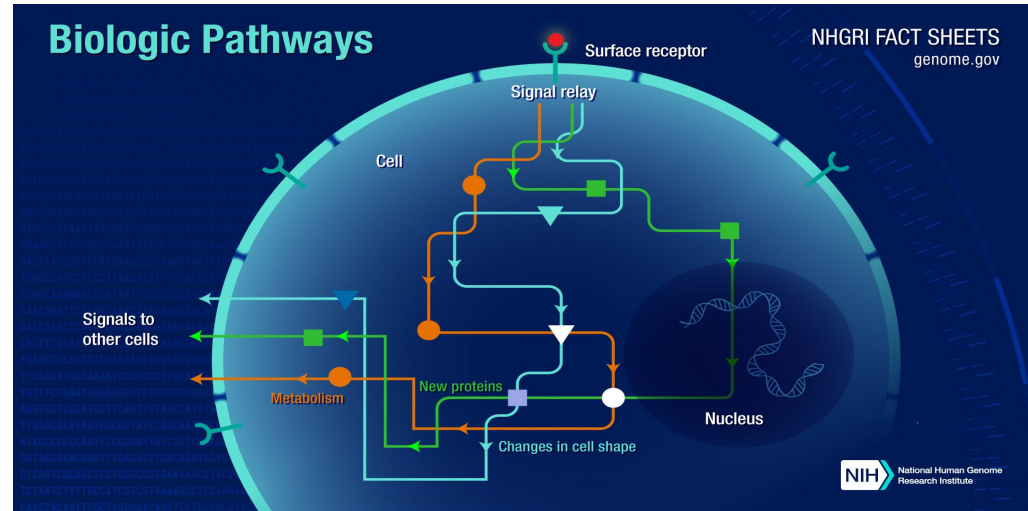


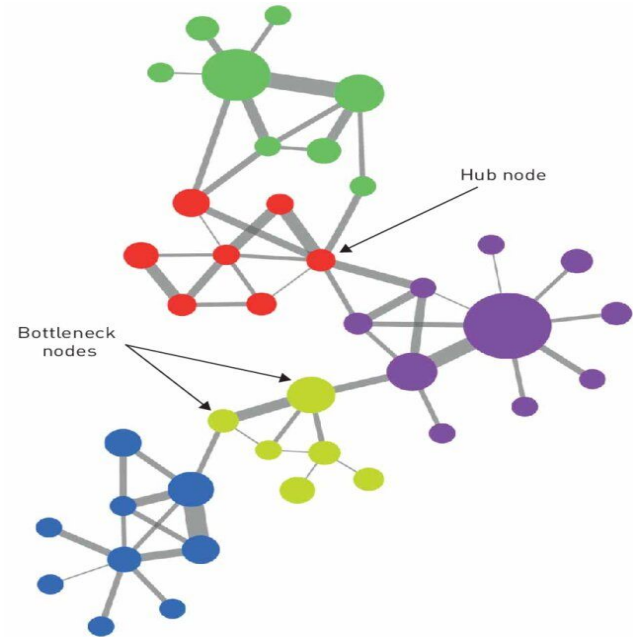
Gene Co-Expression Networks (GCEs): From Clustering to Insights Part 1

As. univ. Bozdog Alexandru



What Are Gene Co-Expression Networks?

- GCEs explore **interactions** between **genes**, beyond static clustering.
- **Nodes**: Genes; **Edges**: Correlations in expression.
- **Importance**: Reveal hidden relationships, key players in pathways, and potential therapeutic targets.
- **Applications**: Cancer research, liver function, and drug discovery.



Understanding Gene Expression

	gene1	gene2	gene3	gene4	gene5
exp 1	0.1	0.3	0.6	0.7	0.9
exp 2	0.1	0.2	0.5	0.7	0.5
exp 3	0.1	0.2	0.1	0.9	0.6
exp 4	0.1	0.2	0.5	0.6	0.6

$N_e = 4$

$N_g = 5$

- Gene expression: How much a gene is "on" or "off" in a condition.
- Measured via RNA levels; stored in a gene expression matrix.
- Rows: Genes; Columns: Samples; Values: Expression levels.
- Challenge: High-dimensional datasets require advanced analysis.

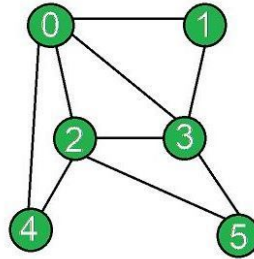
Microarray data

	mRNA samples				
	sample1	sample2	sample3	sample4	sample5 ...
1	0.46	0.30	0.80	1.51	0.90
2	-0.10	0.49	0.24	0.06	0.46
3	0.15	0.74	0.04	0.10	0.20
4	-0.45	-1.03	-0.79	-0.56	-0.32
5	-0.06	1.06	1.35	1.09	-1.09

Gene expression level of gene i in mRNA sample j

Building a Gene Co-Expression Network

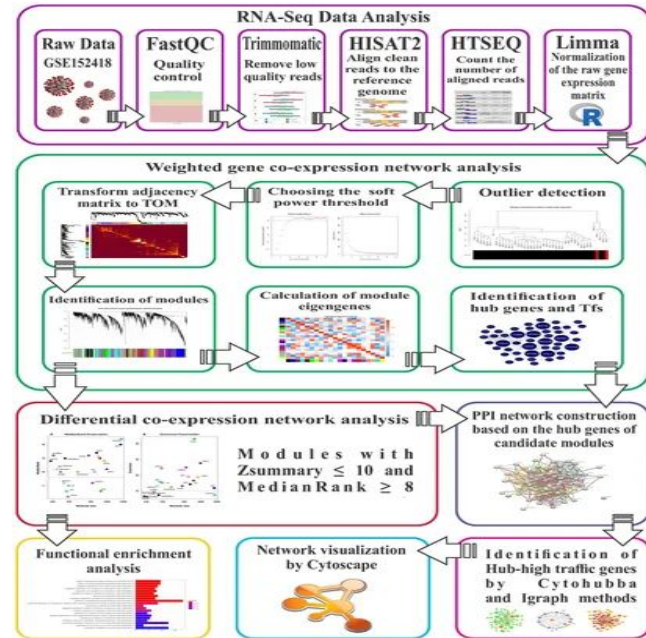
- Preprocessing: Normalize data and filter low-variance genes.
- Calculate correlations (Pearson, Spearman, or mutual information).
- Threshold correlations to create an adjacency matrix.
- Represent the network as a graph.
- Modules can be detected via community detection algorithms such as Louvain or Spectral clustering



	0	1	2	3	4	5
0	0	1	1	1	1	0
1	1	0	0	1	0	0
2	1	0	0	1	1	1
3	1	1	1	0	0	1
4	1	0	1	0	0	0
5	0	0	1	1	0	0

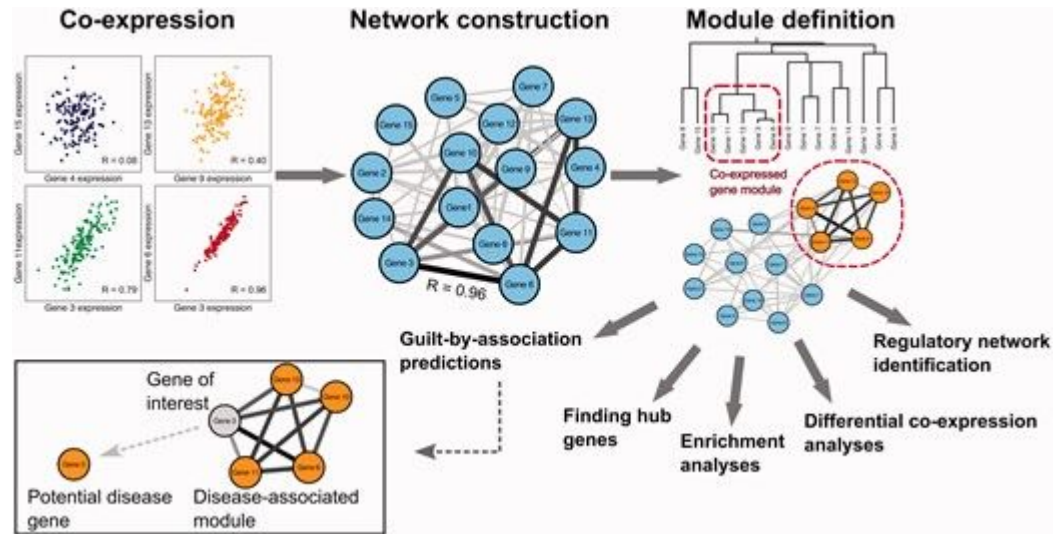
Weighted Gene Co-Expression Network Analysis (WGCNA)

- Soft Thresholding: Converts correlations into connection strengths.
- Topological Overlap Matrix (TOM): Measures overlap in neighbors.
- Hierarchical Clustering: Groups genes into robust modules.
- Example: Glioblastoma research pinpointed immunotherapy targets.



Interpretation Pitfalls in Gene Co-Expression Networks

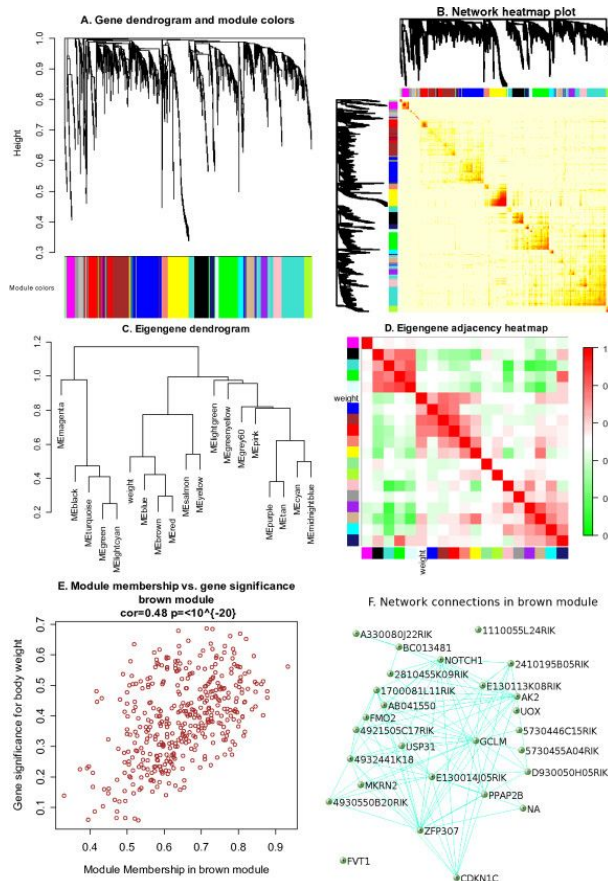
- **Correlation \neq causation** — co-expression does not imply direct regulation.
- **Indirect edges**: gene pairs may correlate because of a third hidden factor.
- **Confounders**: batch effects, tissue heterogeneity, library size \rightarrow false modules.
- **Threshold sensitivity**: module boundaries shift with correlation cutoffs.
- **Sample size**: small n inflates correlation estimates.
- **Hub bias**: highly variable/expressed genes often appear as hubs.



From: van Dam, S., Vösa, U., van der Graaf, A. et al. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform* 19, 575–592 (2018). doi:10.1093/bib/bbw139

Thresholding &

- Hard cutoffs on $|r|$ (e.g. ≥ 0.6).
- Target network density (1–5%).



Sparsification

- k-Nearest Neighbors for strongest edges per node.
- Soft thresholding ($|r|^\beta$) — foundation of WGCNA.

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

Key Takeaways and What's Next?

- GCEs reveal complex gene interactions, aiding in health and disease research.
- WGCNA and community detection algorithms uncover robust modules.
- Applications in cancer biology, liver function, and drug discovery.
- Next Lab: Visualization + validation & the Diseasesome.

