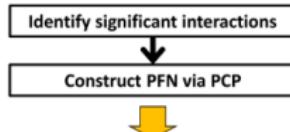


# Multiscale Embedded Gene Co-expression Network Analysis (MEGENA)

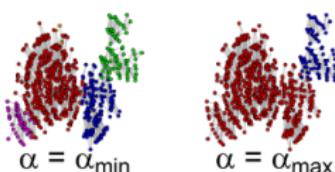
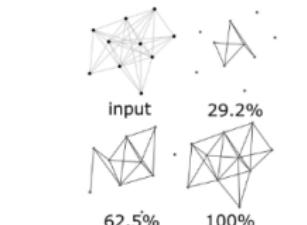
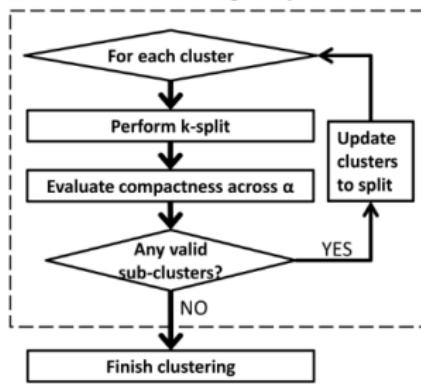
Brayan Gutierrez

# MEGENA Flow Chart

## A Fast PFN Construction



## B Multi-scale Clustering Analysis



## C Downstream Analyses

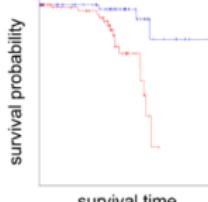
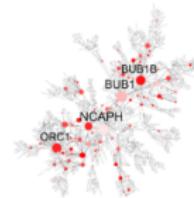
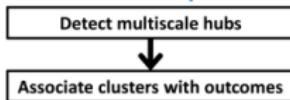


Fig 1. Flow chart of MEGENA. A) Fast planar filtered network construction. Significant interactions are first identified and then embedded on topological surface via a parallelized screening procedure described in the text. On the right, a toy example is illustrated to show construction of PFN from a thresholded network by FDR (top left), and gradual construction of PFN with number of included links and screened pairs shown on the top of each. B) Multi-scale clustering: Beginning from connected components of the initial PFN as the parent clusters, clustering is performed for each parent cluster and compactness of the sub-clusters are evaluated. These steps are described in the dotted box. The clustering is performed iteratively until there remains no further parent

# MEGENA vs. Other Methods

- Existing methods:
  - Require predefined parameters
  - Inability to reproduce scale-free properties
  - Lack support for multiscale organization
- MEGENA:
  - Developed to address issues above
  - Parallelized network construction
  - Efficiently handles large-scale genomic data
  - Novel clustering technique that can uncover hierarchy of co-expressed gene clusters

# The MEGENA Framework

- Four-step methodology for inferring and analyzing gene co-expression networks
- Overcome computational and conceptual shortcomings of previous methods (Planar Maximally Filtered Graph (PMFG), Weighted Gene Co-expression Network Analysis (WGCNA), etc.)

# Fast Planar Network Construction (FPFNC)

- First major step
- Takes in raw gene expression data → Compute pairwise similarities between genes
- Variety of similarity measures including:
  - Pearson's Correlation Coefficient (PCC)
  - Mutual Information (MI)
  - Euclidean Distance
- Includes optional step to remove insignificant interactions (False Discovery Rate)

## FPFNC Cont.

- The core of FPFNC is the ability to construct a Planar Filtered Network (PFN)
  - Embeds gene pairs onto a topological sphere, ensuring no links cross each other
- To make the process scalable, parallelized screening procedure (PCP) is introduced
  - Efficiently identifies subset of gene pairs that are highly likely to be embedded
  - Reduces computational time required compared to existing serial algorithms

# Multiscale Clustering Analysis (MCA)

- Can identify gene co-expression modules (aka “communities” or “clusters”)
- **Modules:** groups of genes whose expression profiles are tightly intertwined, suggesting they are involved in similar biological processes or disease states
- Clustering performed by the novel MCA procedure developed within MEGENA framework

## MCA Cont. 1

- Uses hierarchical divisive approach to dissect complex interactions within PFN
- Turns this into coherent, multi-scale clusters
- Incorporates three distinct criteria to identify clusters:
  - **Shortest Path Distances (SPD):** Optimizes for within-cluster compactness
  - **Local Path Index (LPI):** Optimizes for local clustering structure, which is particularly effective in PFNs due to their abundance of 3- and 4-cliques
  - **Overall Modularity (Q):** A measure used to identify an optimal partition of the network

## MCA Cont. 2

- Uses a compactness measure,  $v(\alpha)$ , a function of resolution parameter  $\alpha$
- Used to identify clusters at different scales
- Smaller  $\alpha$  value, the more compact the cluster
- Multiscale capability allows the framework to identify both coarse-grained and compact clusters that can coexist within a single network
- Addresses a key limitation of many other clustering methods

# Hub Genes and Significance

- MEGENA provides a formal method for identifying most influential genes within modules: the hub genes
- **Hub Genes:** “highly connected members” of the co-expression network
- Identification of these hubs is performed by a dedicated procedure called Multiscale Hub Analysis (MHA)
- MHA works by:

Grouping the different scales that show similar connectivity patterns → Identify significant hubs within each scale → Identify multiscale hubs by combining significance scores for each node across all different scales

# Innovations

Method	Key Innovations	Purpose
FPFNC	Parallelized screening procedure (PCP); quality control of co-expression similarities	Construct a robust, scalable network efficiently
MCA	Hierarchical divisive clustering; uses SPD, LPI, and Q; compactness measure with a resolution parameter ( $\alpha$ )	Identify multiscale clusters with varying degrees of compactness
MHA	Grouping of scales; combining significance scores across scales	Identify robust, highly connected hub genes at single and multiple scales

# Our Data

## References

Song, W.-M., & Zhang, B. (2015). *Multiscale Embedded Gene Co-expression Network Analysis*. PLoS Computational Biology, 11(11), e1004574.