Notes from *WGCNA: an R package for weighted correlation network analysis* (Langfelder & Horvath, 2008)

1. Why WGCNA was developed

- Traditional clustering methods (like hierarchical clustering or k-means) group genes based on similarity but may miss higher-level *network structure*.
- WGCNA (Weighted Gene Co-expression Network Analysis) extends clustering by building a **network** where:
  - Nodes = genes
  - Edges = weighted correlations between expression profiles
- This makes it possible to find **modules** of tightly co-expressed genes rather than just isolated clusters.

---

2. Core concepts

- **Soft-thresholding:**
  - Instead of hard-cutting correlation values (e.g., 0.8 cutoff), WGCNA uses a power function to emphasize strong correlations while keeping weaker ones.
  - Leads to networks that approximate a **scale-free topology** (a property seen in many biological systems).
- **Modules:**
  - Defined as groups of genes with high topological overlap (not just correlation).
  - Identified via hierarchical clustering of the **Topological Overlap Measure (TOM)**.
- **Module eigengenes:**
  - Each module is summarized by its **first principal component** (the "eigengene"), capturing the main expression pattern of the module.
  - Modules can be correlated with external traits (e.g., AMD vs. control, severity stages, age).

---

3. Advantages of WGCNA vs. classical clustering

- Goes beyond simple grouping — **captures relationships between clusters/modules**.
- Allows testing how modules relate to biological or clinical variables (e.g., AMD phenotype).
- Identifies **hub genes** within modules (genes with the highest intramodular connectivity, potential key drivers).
- Applicable to **any high-dimensional data** (not just gene expression).

---

4. Relevance to your AMD project

- If you use **different clustering methods** (k-means, spectral, hierarchical), WGCNA is a **network-based complement**:
  - Clustering → finds groups of genes with similar expression.
  - WGCNA → finds **network modules** with biological relevance and highlights candidate **hub genes** that might drive AMD pathways.

- WGCNA is especially useful for diseases like AMD where **complex pathways and interactions** are involved (oxidative stress, complement system, lipid metabolism, angiogenesis).
- You can test whether WGCNA modules associate with:
    - **AMD case/control status**
    - **Disease severity/stages**
    - **Age** (a key covariate)

---

5. Limitations / Considerations

- Requires relatively large sample sizes for stable module detection.
- Sensitive to preprocessing (normalization, batch effects).
- Module definitions can vary between datasets — often needs **consensus networks** across cohorts.
- WGCNA is exploratory: modules suggest hypotheses but need biological validation (pathway analysis, wet lab experiments).

---

6. Presentation-ready framing

When comparing clustering approaches for AMD gene expression data:

- **K-means/hierarchical:** find groups based on direct similarity.
- **Spectral clustering:** uses graph Laplacians to capture complex structure.
- **WGCNA:** builds a weighted co-expression network → identifies biologically meaningful **modules** and **hub genes** linked to AMD.
- This can help **prioritize candidate genes** for further validation in AMD pathogenesis.

Notes from *hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data* (Morabito et al., 2023)

1. Why hdWGCNA?

- Traditional **WGCNA** works well for bulk RNA-seq, but struggles with:
  - **Single-cell & spatial transcriptomics** (sparse, noisy, high-dimensional).
  - Datasets with **hundreds of thousands to millions of cells**.
- **hdWGCNA** adapts WGCNA for these data types, scaling network analysis to modern high-throughput transcriptomics.

---

2. Key innovations

- **High-dimensional scalability:** Efficient enough to handle ~1 million single cells.
- **Seurat integration:** Built to work seamlessly with the widely used single-cell analysis framework.
- **Module detection in single-cell/spatial data:** Allows discovery of cell-type-specific or spatially restricted **co-expression modules**.
- **Downstream tools included:** Gene set enrichment, statistical testing, visualization.

---

3. How hdWGCNA relates to clustering

- Like WGCNA, hdWGCNA constructs **gene co-expression networks** and identifies **modules** (clusters of genes with highly correlated expression).
- Unlike **k-means, hierarchical, or spectral clustering**, which directly group genes based on similarity, hdWGCNA:
  - Builds a **weighted network** first.
  - Clusters genes into modules based on **topological overlap** (shared network connections).
  - Summarizes each module with a **module eigengene** (main expression pattern).
- This approach helps reveal **biological structure** that standard clustering might miss, especially in complex diseases.

---

4. Applications shown in the paper

- Applied to **single-cell datasets** from neurological disorders (ASD, Alzheimer's).
- Identified **disease-relevant gene modules** linked to pathology.
- Demonstrated that **hub genes** (genes with highest connectivity within modules) can highlight potential **drivers of disease pathways**.

---

5. Relevance to AMD project

- **AMD is complex**: involves oxidative stress, inflammation, lipid metabolism, angiogenesis — likely requiring multi-gene modules rather than isolated genes.
- hdWGCNA could help:
  - Detect **gene co-expression modules** specific to retinal cell types (e.g., RPE cells, photoreceptors, microglia).
  - Correlate modules with AMD traits (case/control, severity, age).
  - Highlight **hub genes** that may play key roles in AMD progression.

- Especially powerful if your AMD data includes **single-cell or spatial transcriptomics**.

---

6. Strengths vs. other clustering methods

- **K-means / hierarchical clustering:** simple, interpretable, but may not capture higher-order relationships.
- **Spectral clustering:** good at finding non-linear structures, but doesn't integrate external traits.
- **hdWGCNA:** network-based, scalable, trait-integrated → ideal for identifying biologically relevant **modules** and **hub genes** in AMD.

---

7. Limitations / caveats

- Requires careful preprocessing (normalization, batch correction) in single-cell data.
- Modules may differ across datasets → need replication or consensus analysis.
- As with WGCNA, results are **hypothesis-generating** and need biological validation.

Notes from *MEGENA: Multiscale Embedded Gene Co-expression Network Analysis* (Song & Zhang, 2015)

1. Why MEGENA was developed

- Traditional clustering and network methods often:
  - Miss **hierarchical/multiscale structures** in gene networks.
  - Identify only *flat* modules (one level of grouping).
- MEGENA is designed to capture **multiscale modular organization** of gene co-expression networks.
- Relevant for complex diseases (like AMD) where gene interactions may occur at multiple scales (small functional clusters within larger pathways).

---

2. Core methodology

- **Planar Filtered Network (PFN):**
  - Starts by constructing a co-expression network filtered into a planar graph to reduce noise and redundancy.
- **Multiscale clustering:**
  - Detects **modules within modules** (hierarchical communities).
  - Allows analysis at different scales of network resolution.
- **Hub gene detection:**
  - Identifies genes with high connectivity inside each module (candidate key drivers).
- **Functional enrichment:**
  - Tests modules for enrichment in biological pathways and traits.

---

3. Comparison to other clustering approaches

- **K-means / hierarchical clustering:** group genes but usually at one scale.
- **WGCNA:** finds modules, but tends to emphasize one resolution (module size depends on chosen parameters).
- **MEGENA:** explicitly designed to uncover *nested and overlapping modules*, providing a richer view of gene regulation.

---

4. Applications demonstrated in the paper

- Applied MEGENA to gene expression datasets from **human tissues** and **mouse models**.
- Showed that modules are **biologically meaningful** (enriched for known pathways).
- Identified **hub genes** as potential regulators of disease-relevant networks.

---

5. Relevance to AMD research

- AMD involves multiple biological layers:
  - Inflammation, angiogenesis, oxidative stress, complement system, lipid metabolism.
- MEGENA could:
  - Detect **small, specialized gene modules** (e.g., complement activation genes).

    - Reveal how these nest inside **larger functional modules** (e.g., immune regulation networks).
    - Prioritize **hub genes** within AMD modules as candidate biomarkers or therapeutic targets.
- Particularly useful if you want to highlight **multi-level gene interactions** instead of just flat clusters.

---

6. Strengths

- Captures **hierarchical network structure** (multiscale).
- Identifies both global and local patterns in co-expression.
- Good for uncovering **complex disease pathways** with multiple interacting modules.

---

7. Limitations

- Computationally more intensive than simple clustering.
- Requires careful preprocessing (normalization, batch correction).
- As with WGCNA/hdWGCNA, results are exploratory → need external validation.

---

8. Framing for your presentation

- You can present MEGENA alongside **WGCNA and hdWGCNA** as:
    - **WGCNA:** flat modules (bulk data).
    - **hdWGCNA:** scalable modules (single-cell/spatial).
    - **MEGENA:** multiscale hierarchical modules (nested clusters).
- For AMD:
    - Classical clustering → broad groups.
    - Network-based clustering (WGCNA/hdWGCNA) → co-expression modules.
    - MEGENA → **nested gene networks** that could reflect the multi-layered biology of AMD