Network Science applied to Epilepsy

BioBytes Group

MO413 · 1º semester of 2024 · Institute of Computing · UNICAMP

Carnot Luiz B. G. Filho ra255785 · **Isabella C. Galvão** ra199200 · **João Paulo F. da Silva** ra263792

Laíz Luiza M. de Oliveira ra200744 · **Ana Carolina da Hora** ra272499 · **Vitoria D. M. Pinho** ra188511

23/05/2024

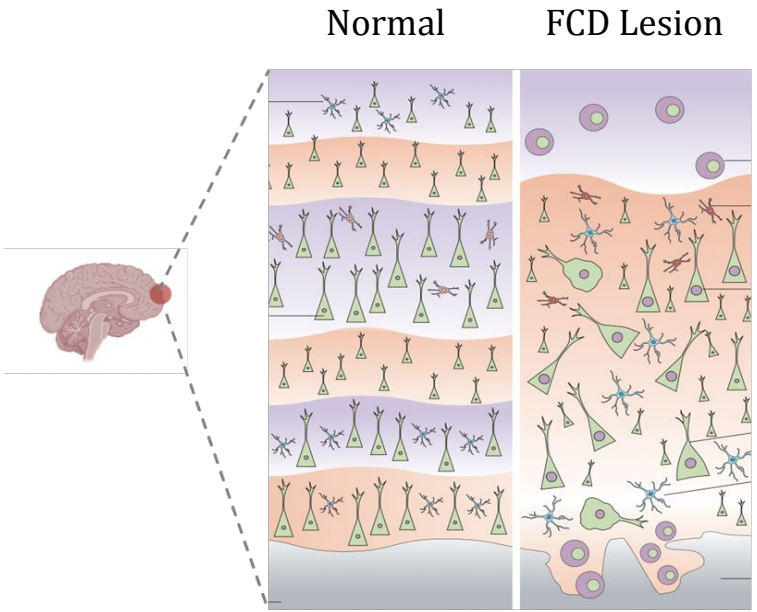
Introduction

Epilepsy

- Epilepsy is a disease caused by the **synchronous**, **excessive** or **abnormal** activity of neuronal cells in the brain.

Focal Cortical Dysplasia

- Pediatric epilepsy
- Malformation of cortical development
- Cortical and cellular abnormalities



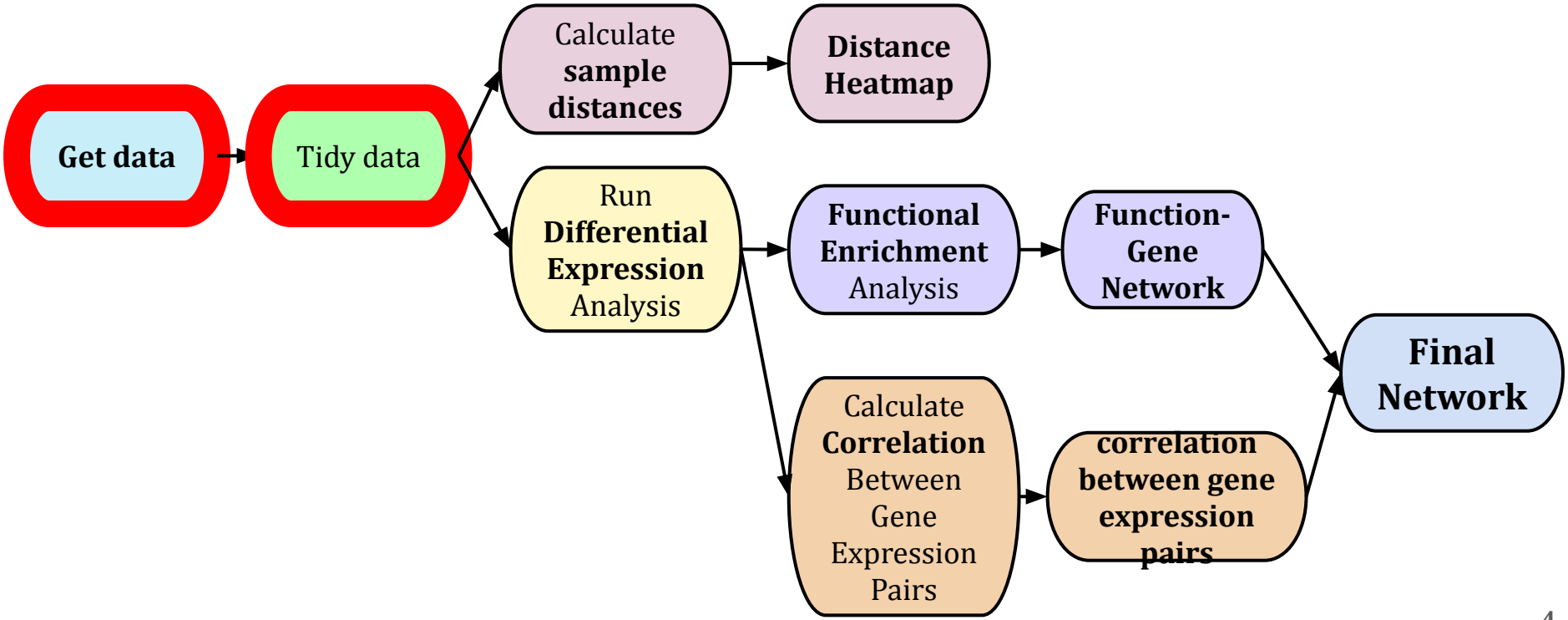
Sisodiya et al., Lancet Neur., 2009

Proposal and Research Questions

Analyze the correlation between **genes** and **biological functions** in the analysis of differentially expressed genes in patients diagnosed with **epilepsy** caused by Focal Cortical Dysplasia type 2b (FCD 2b)

- What are the differentially expressed genes between **disease** and **control group**?
- Which **biological functions** are most represented in the disease? → **functional networks**
- How are these genes being regulated? → **regulation networks**

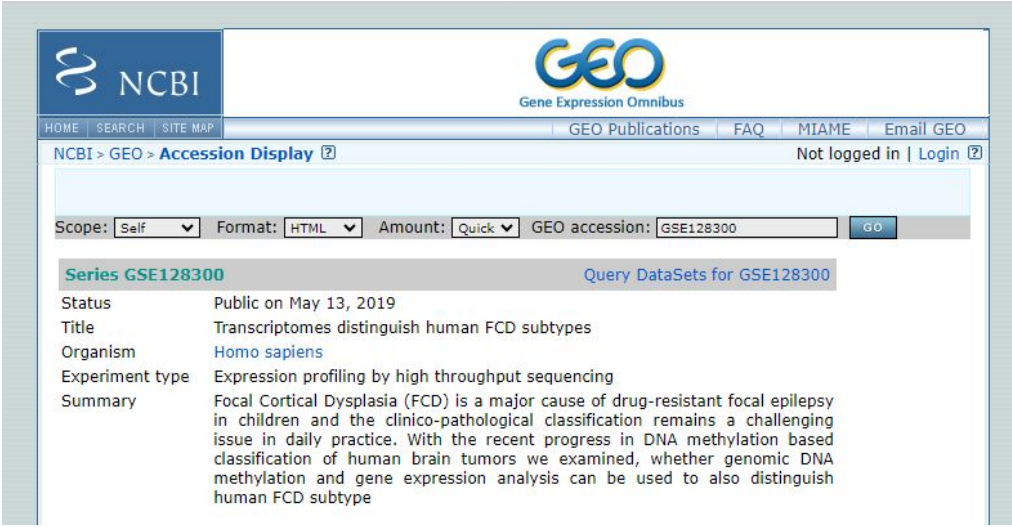
Workflow


Get data
Tidy data
Calculate sample distances
Distance Heatmap
Run Differential Expression Analysis
Functional Enrichment Analysis
Function-Gene Network
Calculate Correlation Between Gene Expression Pairs
correlation between gene expression pairs
Final Network

Get Data

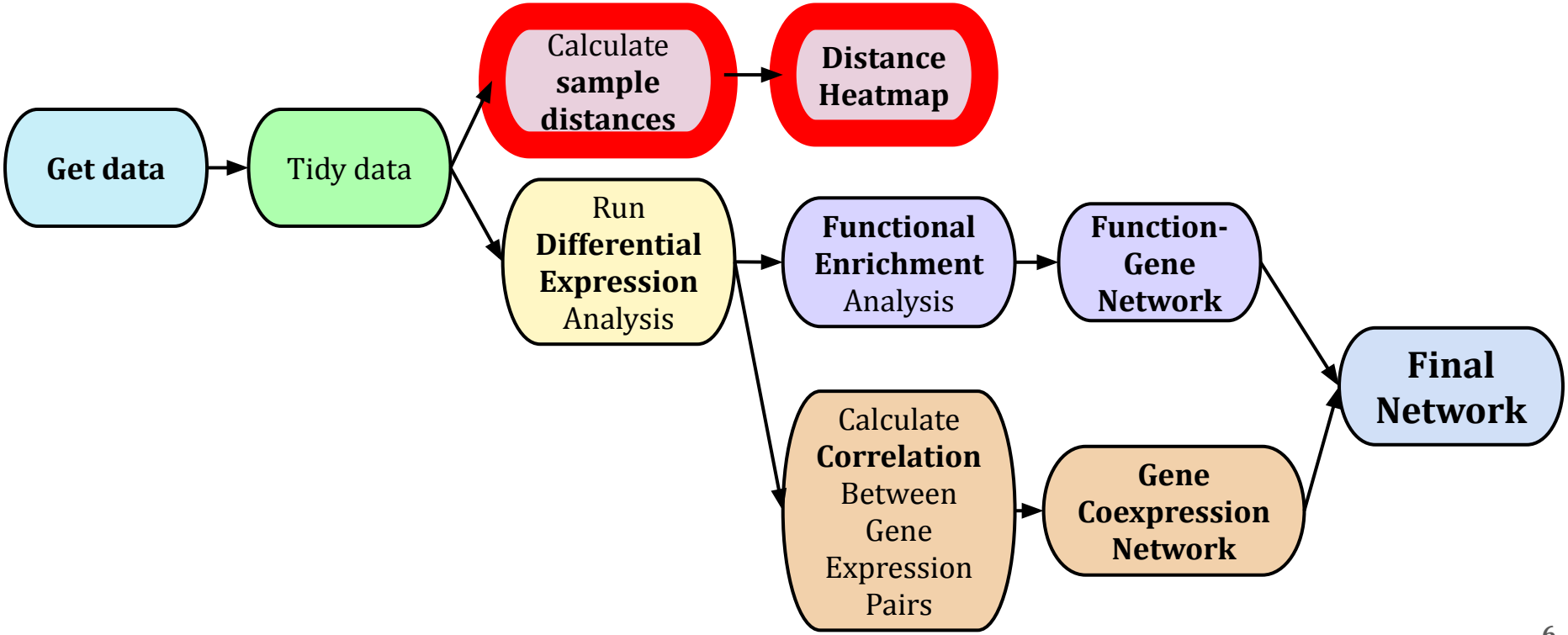
GSE128300

- RNA-seq data
- Expression matrix
- n = 5 FCD IIb
- n = 8 control
- Metadata obtained from the authors (.xlsx)



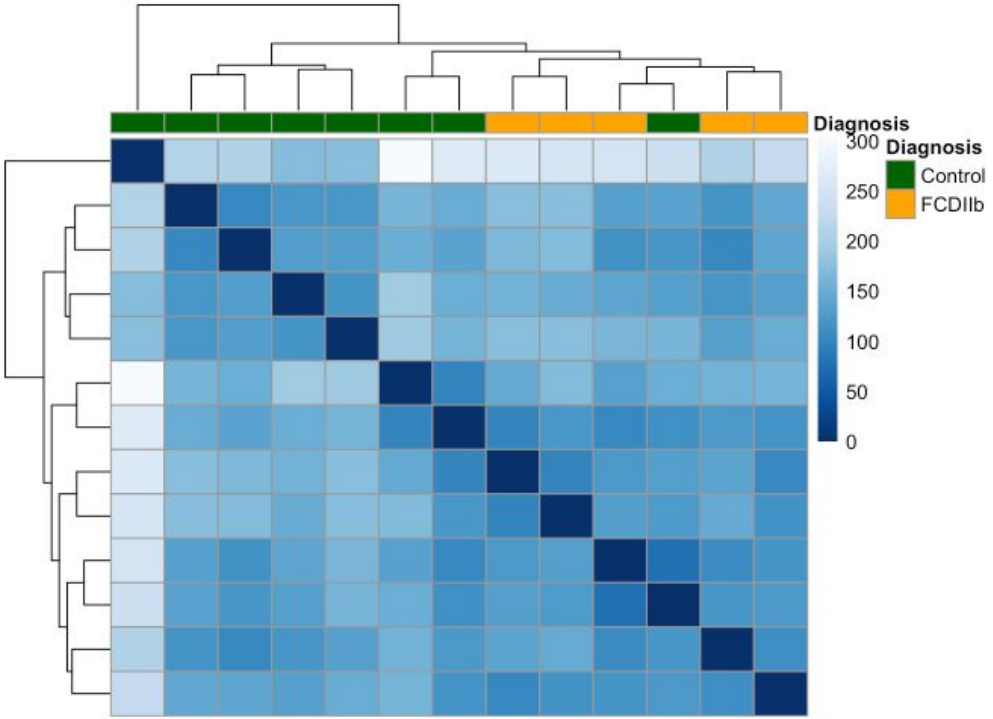
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128300

Workflow

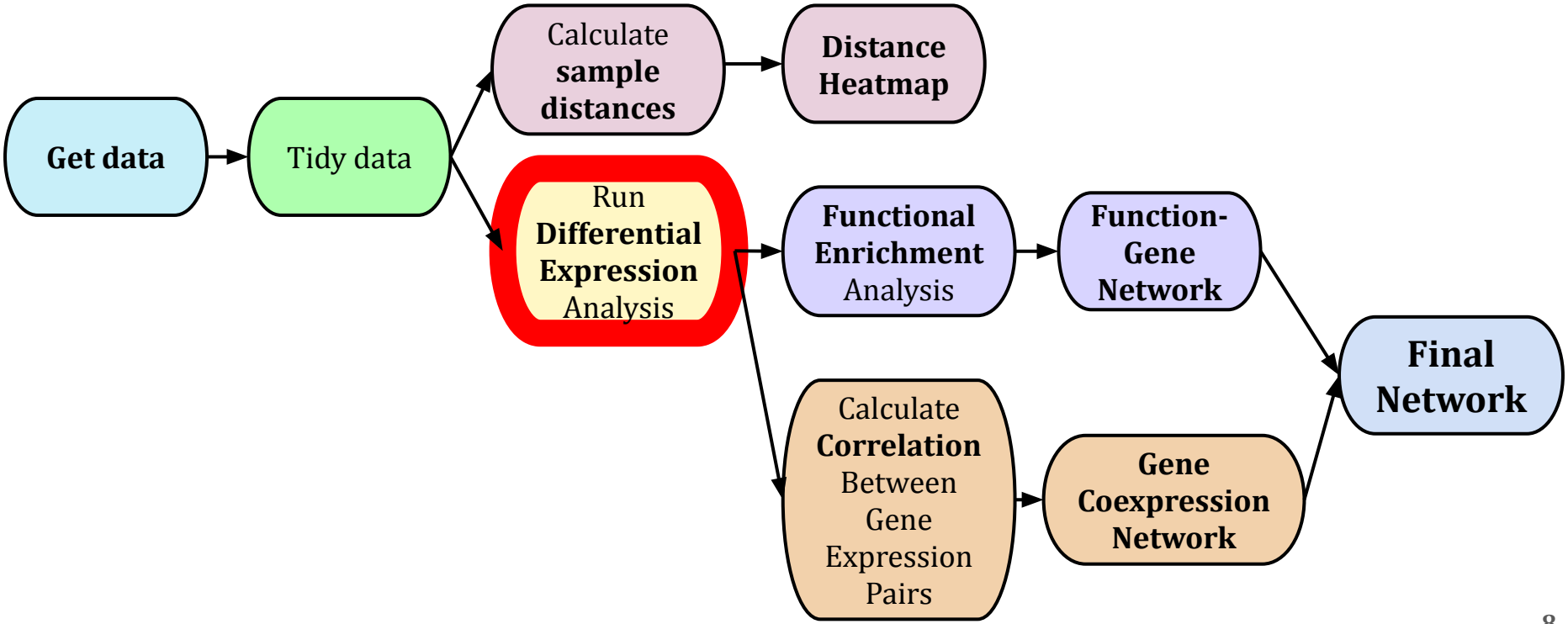

Get data
Tidy data
Calculate sample distances
Distance Heatmap
Run Differential Expression Analysis
Functional Enrichment Analysis
Function-Gene Network
Calculate Correlation Between Gene Expression Pairs
Gene Coexpression Network
Final Network

Sample Distances

- FCD IIb and controls are transcriptomically distinct



Workflow


Get data
Tidy data
Calculate sample distances
Distance Heatmap
Run Differential Expression Analysis
Functional Enrichment Analysis
Function-Gene Network
Calculate Correlation Between Gene Expression Pairs
Gene Coexpression Network
Final Network

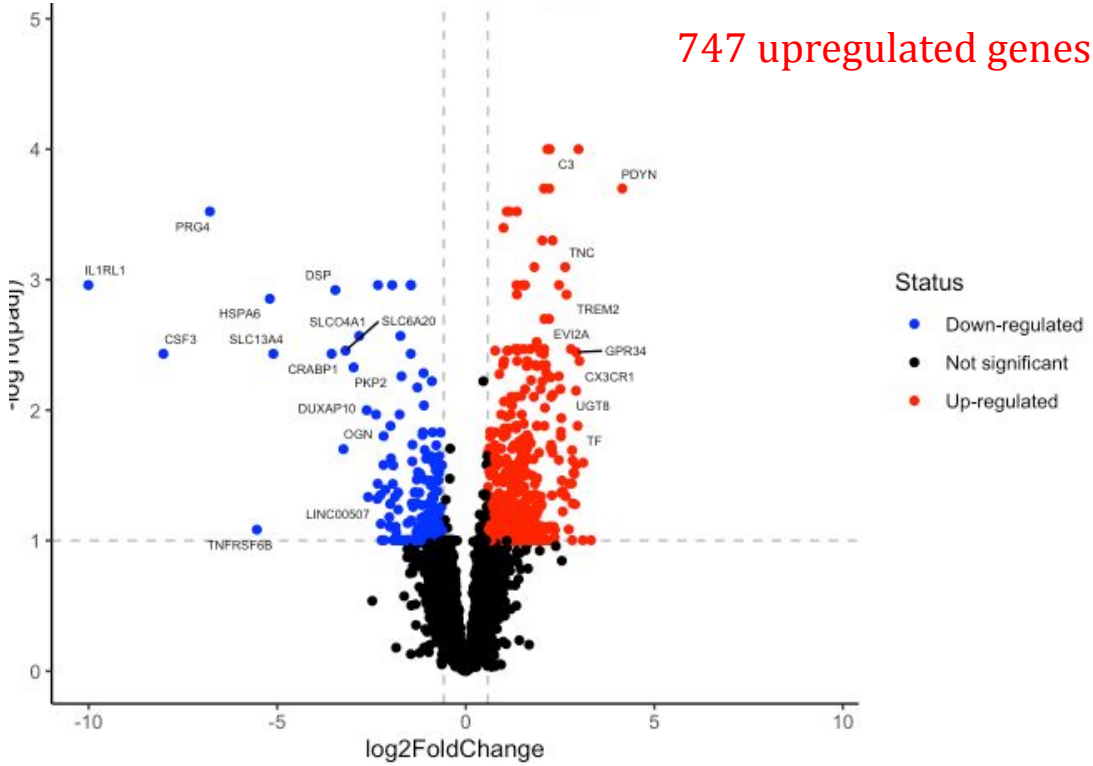
Differential Expression Analysis

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

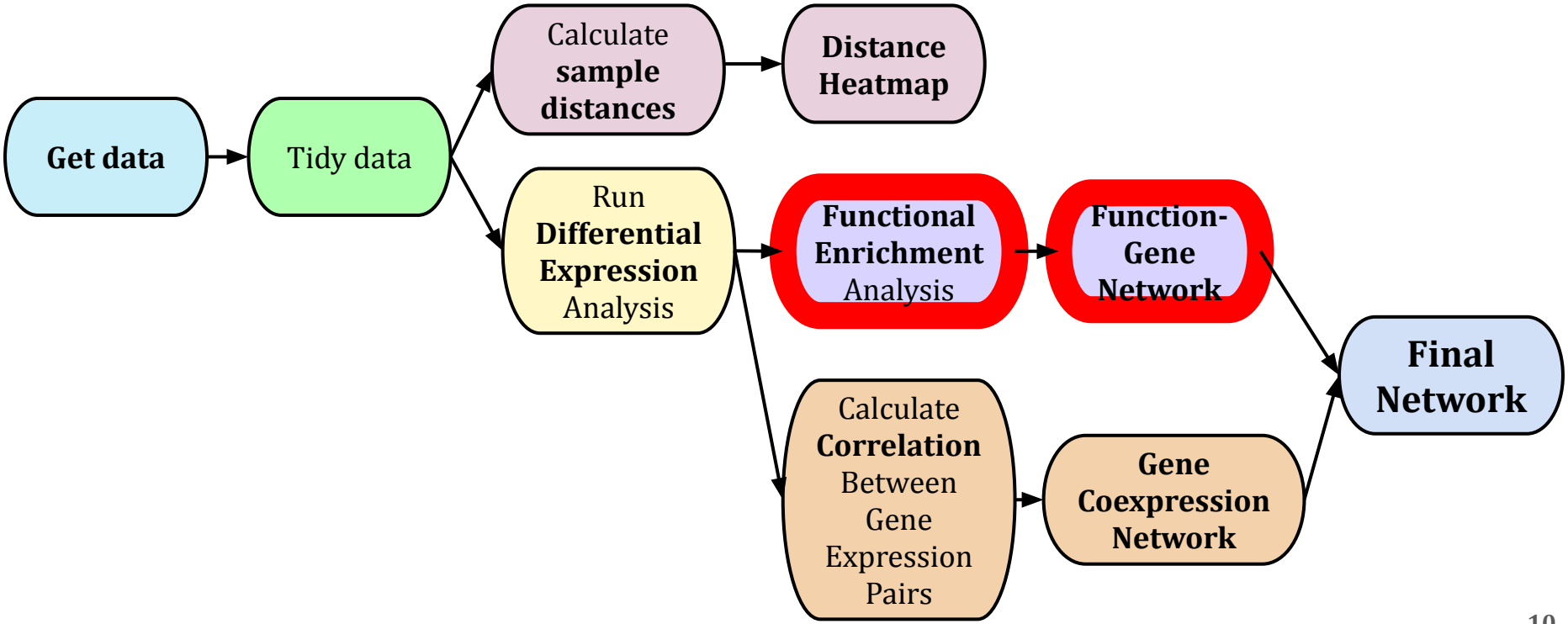
Criteria:

- FC > 1.5
- padj < 0.1

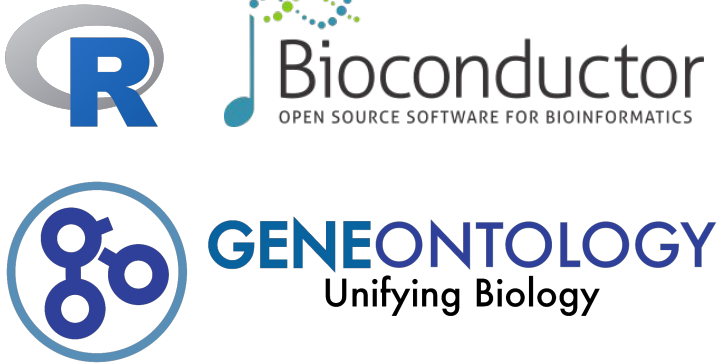
747 upregulated genes



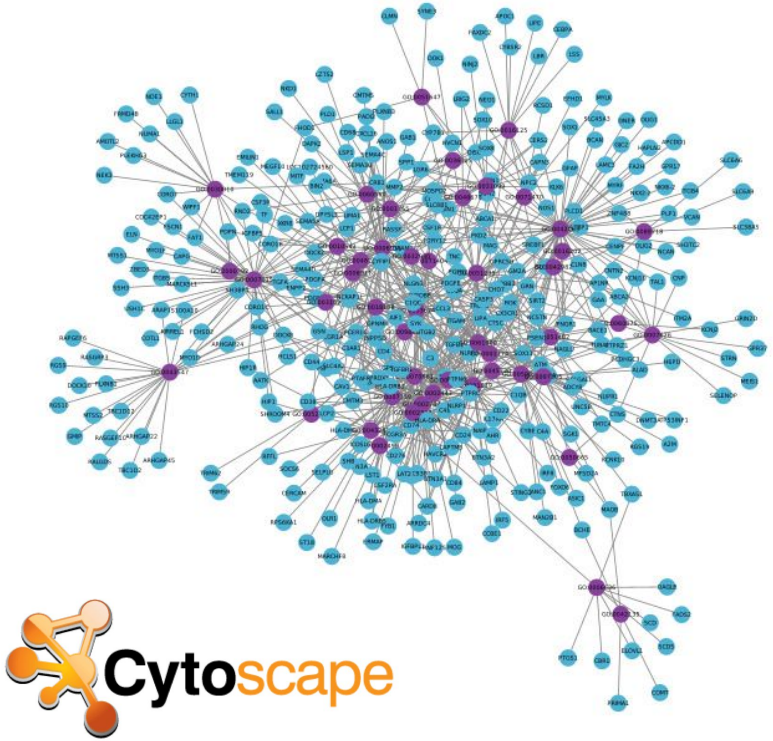
Workflow


Get data
Tidy data
Calculate sample distances
Distance Heatmap
Run Differential Expression Analysis
Functional Enrichment Analysis
Function-Gene Network
Calculate Correlation Between Gene Expression Pairs
Gene Coexpression Network
Final Network

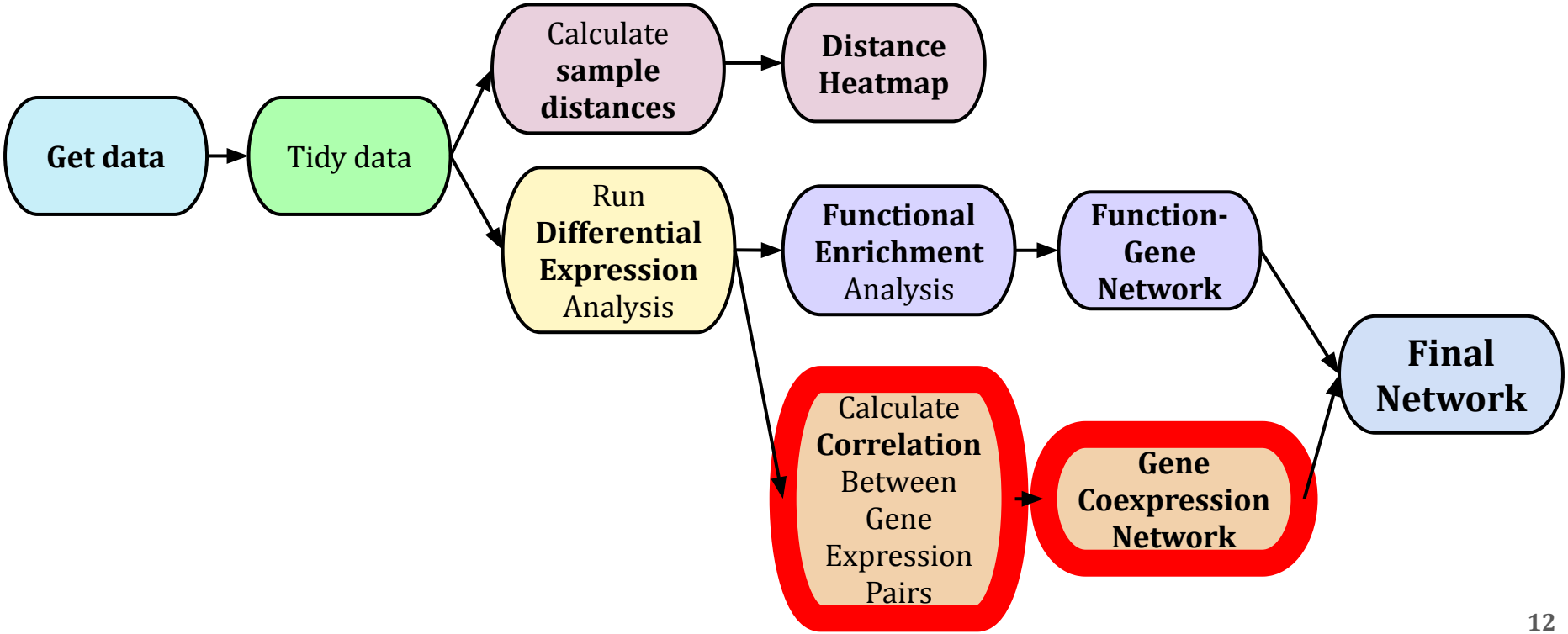
Functional Enrichment Analysis



- ClusterProfiler
  - EnrichGO
    - ont = "BP"

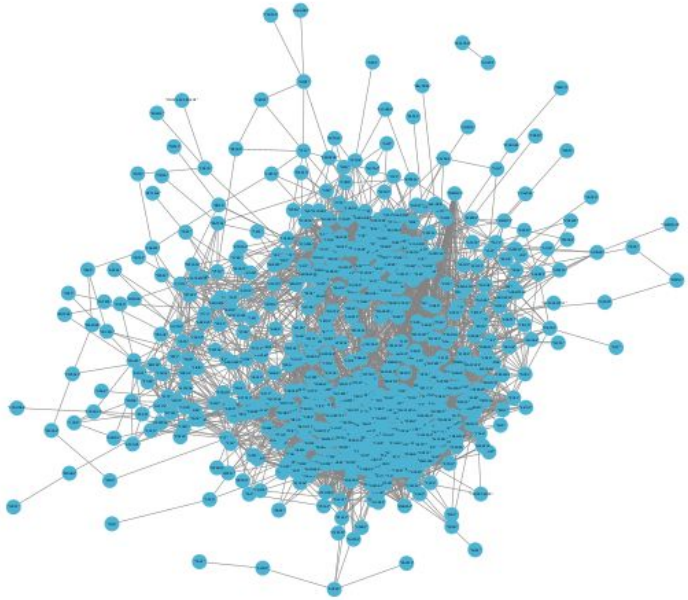


Workflow

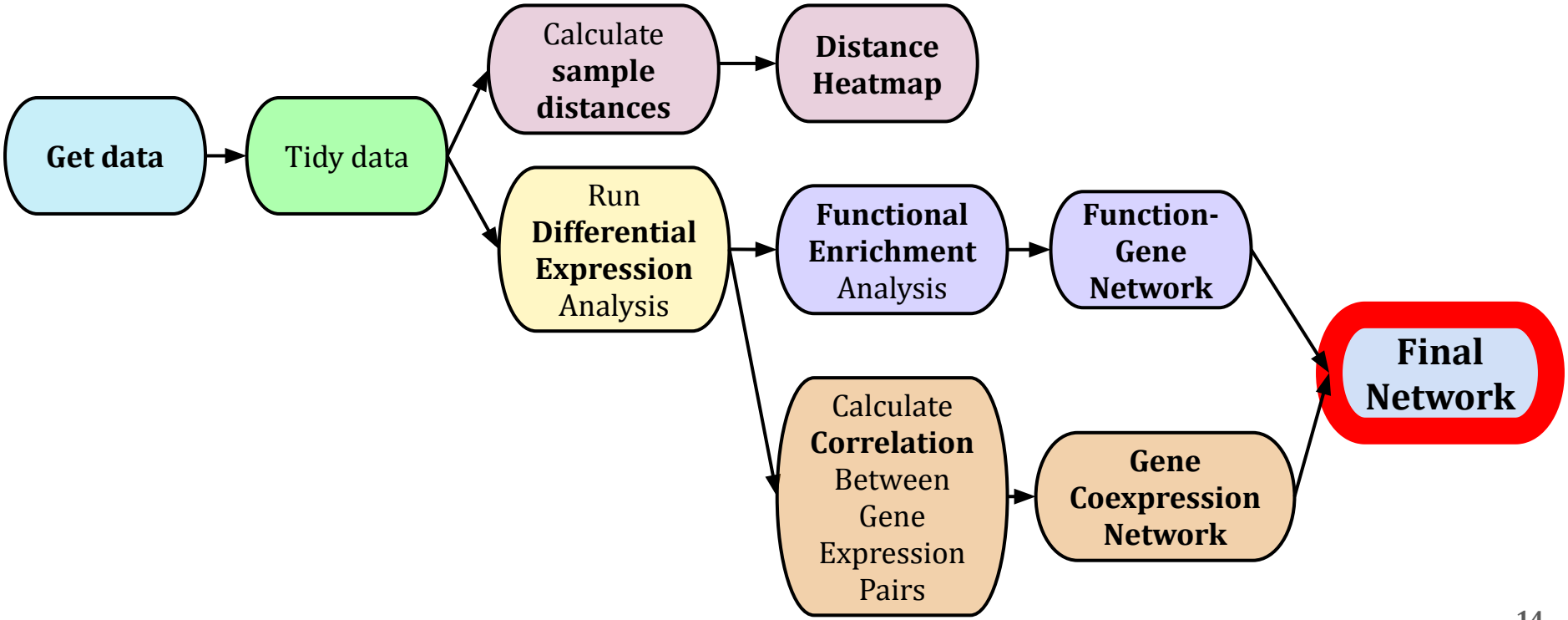

Get data
Tidy data
Calculate sample distances
Distance Heatmap
Run Differential Expression Analysis
Functional Enrichment Analysis
Function-Gene Network
Calculate Correlation Between Gene Expression Pairs
Gene Coexpression Network
Final Network

Functional Enrichment Analysis

- Pearson Correlation
  - P < 0.05 and statistic > 0.85

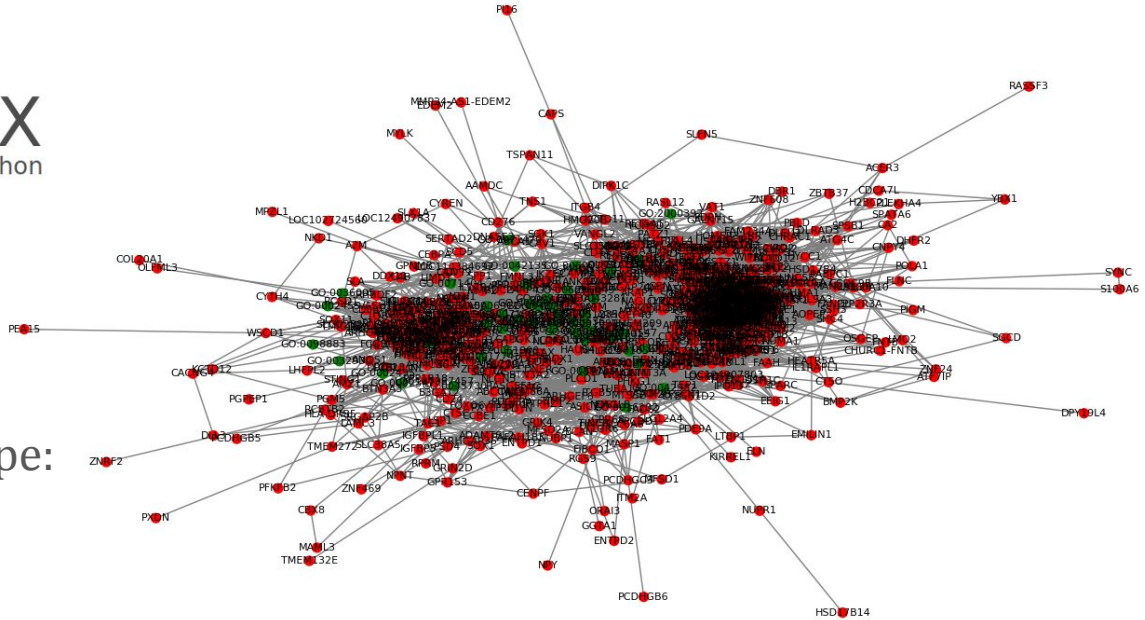


Workflow

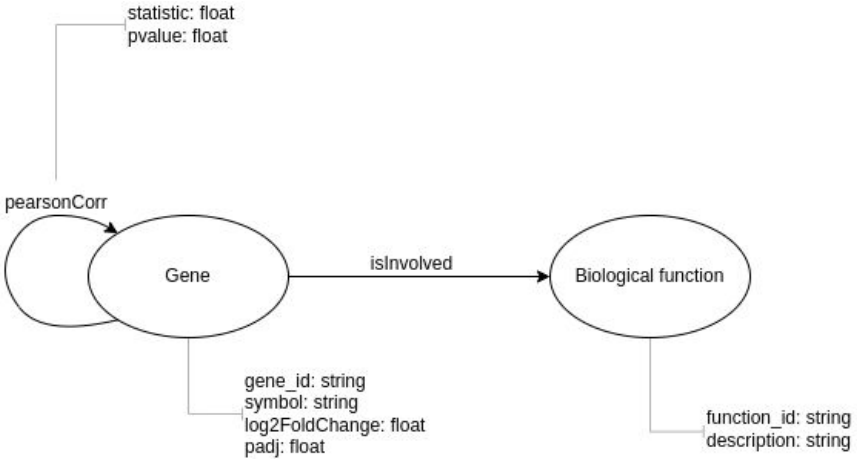

Get data
Tidy data
Calculate sample distances
Distance Heatmap
Run Differential Expression Analysis
Functional Enrichment Analysis
Function-Gene Network
Calculate Correlation Between Gene Expression Pairs
Gene Coexpression Network
Final Network

Final Network

- NetworkX instead Cytoscape:
  - Need to use heterogeneous edges and nodes to perform analyzes

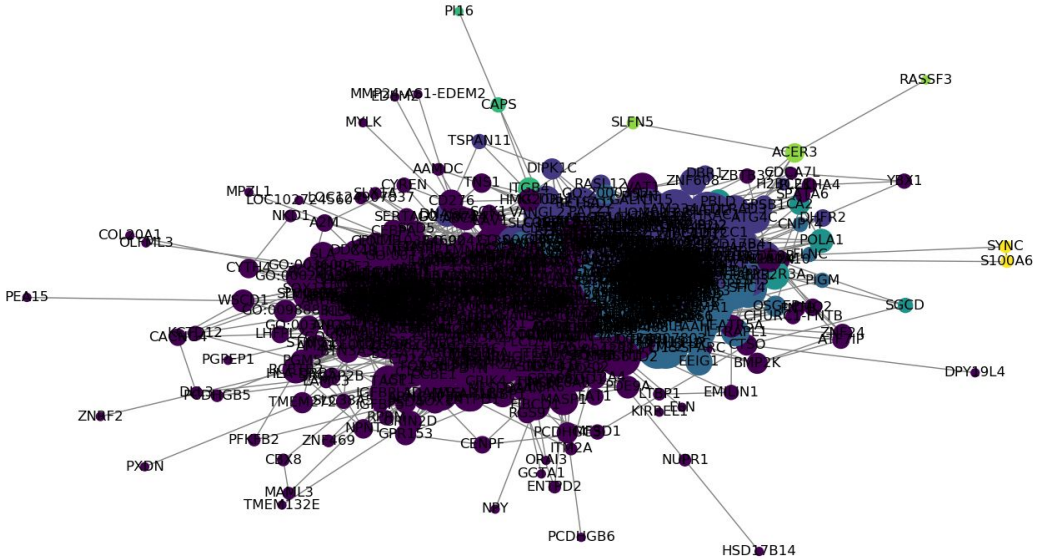


Logic Model


statistic: float
pvalue: float
pearsonCorr
Gene
isInvolved
Biological function
gene_id: string
symbol: string
log2FoldChange: float
padj: float
function_id: string
description: string

Preliminary analysis

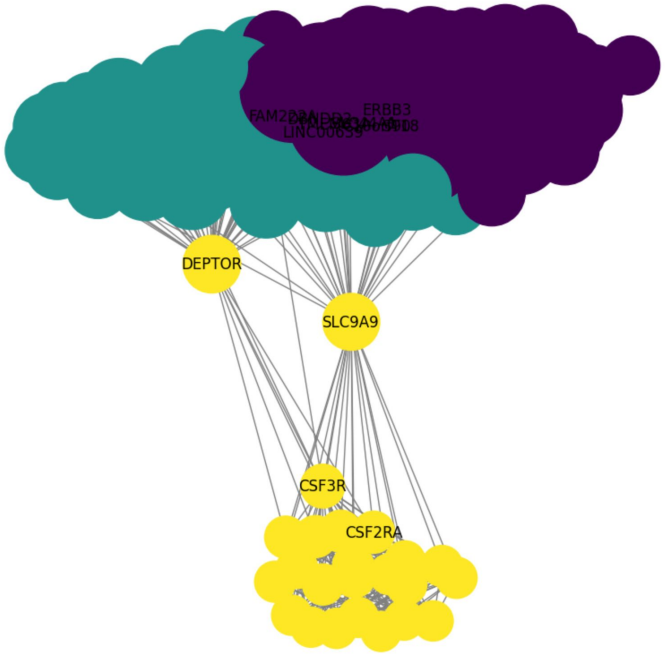
- Community detection through the **greedy modularity communities** algorithm:
  - 6 communities



Preliminary analysis

- Node filter with **degree** greater than 70 applied:
  - 3 largest communities
  - Highest degree genes in each community:

Gene	Degree	Community
ERBB3	181	blue
DBNDD2	190	purple
DEPTOR	43	yellow



Future analysis and Problems

Future analysis:

- Use other algorithms to detect **communities**
- Identify which **biological functions** are most relevant
- Find most relevant genes/functions using **betweenness centrality**
- **Perturbation and robustness analysis**: assess how the removal or perturbation of specific genes affects network structure and function
- Improve visualization of analyzes by exporting graphs to **Cytoscape**

Problems:

- Difficulty using node and edge properties in Cytoscape and Neo4j
- Difficulty in visualizing the most important genes/functions, due to the large graph

Thank You!
Questions?

References

[1] - Patel P, Moshé SL. The evolution of the concepts of seizures and epilepsy: What's in a name? Epilepsia Open. 2020 Jan 10;5(1):22-35. doi: 10.1002/epi4.12375. PMID: 32140641; PMCID: PMC7049807.

[3] - Drugs.com: https://www.drugs.com/health-guide/seizure.html

[4] - Associations of B-Type Natriuretic Peptide and Its Coding Gene Promoter Methylation With Functional Outcome of Acute Ischemic Stroke: A Mediation Analysis - DOI:10.1161/JAHA.120.017499

[5] - DESeq2 (Bioconductor Software Package): https://bioconductor.org/packages/release/bioc/html/DESeq2.html

References

[6] - Gene Ontology: https://geneontology.org/

[7] - Cytoscape: https://cytoscape.org/

[8] - Transcriptomes distinguish human FCD subtypes: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128300

[9] - Brain image theme: Getty Images (Yuichiro Chino)

[10] - DNA image theme: iStock (Shutter2U)