BINF 5003: Data Mining, Modeling, and Biostatistics

Week 14 - Module 6

Review

Gene Set Enrichment Analysis (GSEA)

What are gene sets?

- A **gene set** is a collection of genes grouped together based on shared characteristics, such as:
  - Biological pathways (e.g., glycolysis, immune response).
  - Chromosomal location.
  - Functional annotations (e.g., transcription factors, kinase activity).
  - Gene expression patterns under certain conditions.

- Examples of widely used gene sets:
  - **KEGG (Kyoto Encyclopedia of Genes and Genomes):** Pathway-based groupings.
  - **Gene Ontology (GO):** Grouped by molecular function, cellular component, and biological process.
  - **Hallmark gene sets:** Broad biological themes (e.g., hypoxia, apoptosis).

Why Gene Sets Matter?

- Biological systems often function through groups of genes working together in pathways or networks

- Analyzing individual genes may miss the bigger picture, but focusing on gene sets can reveal system-level changes

Gene Set Enrichment Analysis (GSEA) – example

Gene
ATM
MAPK3
MTOR
RAF1
PIK3CA
AKT1
MAPK1

Sort the genes **based on fold change** or **statistical significance** (e.g., p-value)

Gene	Rank
MAPK1	1
MAPK3	2
RAF1	3
…	…
PIK3CA	18
AKT1	19
MTOR	20

Gene Set Enrichment Analysis (GSEA) – example

Gene
ATM
MAPK3
MTOR
RAF1
PIK3CA
AKT1
MAPK1

Sort the genes **based on fold change** or **statistical significance** (e.g., p-value)

Gene	Rank	
MAPK1	1	**MAPK Pathway**
MAPK3	2	
RAF1	3	
…	…	
PIK3CA	18	**PI3K-Akt Pathway**
AKT1	19	
MTOR	20	

Gene Set Enrichment Analysis (GSEA) – example

Our genes of interest

Gene	Rank*
MAPK1	1
MAPK3	2
RAF1	3
…	…
PIK3CA	18
AKT1	19
MTOR	20

MAPK is pathway is **upregulated** because its genes are at the top of the ranking

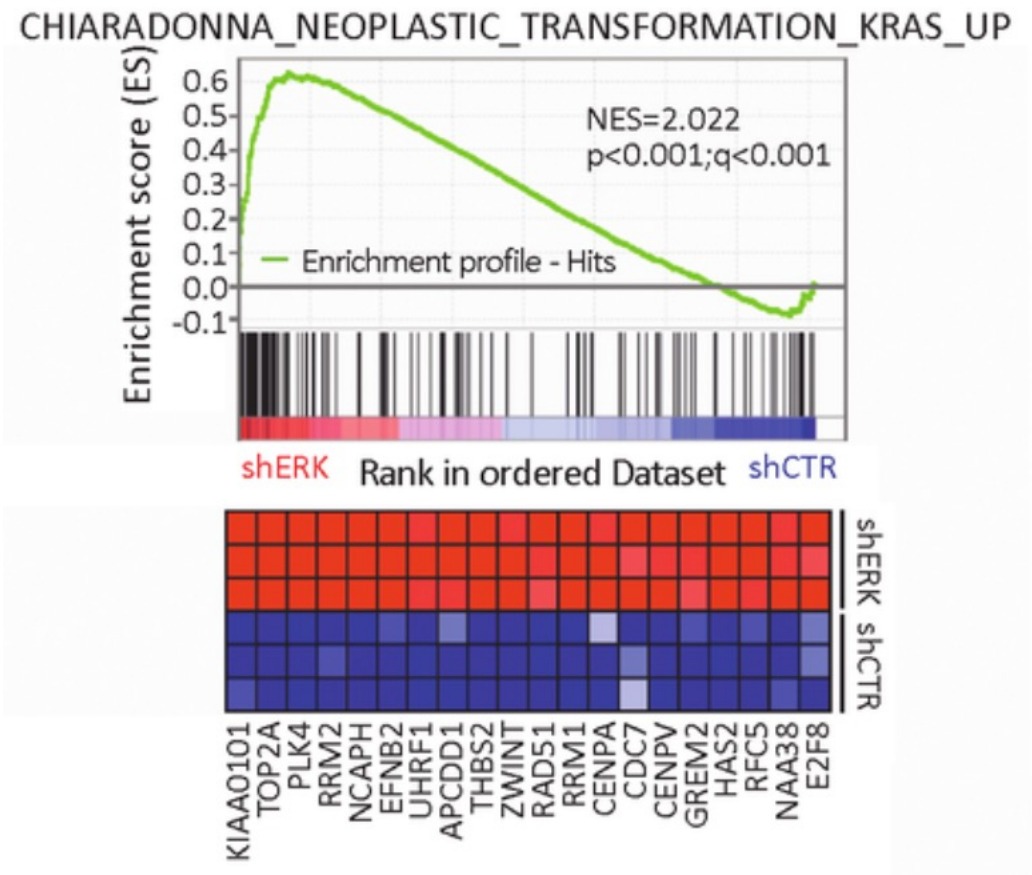
PI3K-Akt is pathway is **down regulated** because its gene are at the bottom of the ranking

MAPK Pathway Genes :
MAPK1 , MAPK3, RAF1 ,BRAF, MEK1, KRAS,…

PI3K-Akt Pathway Genes:
PIK3CA, AKT1, MTOR, MYC, FOXO1, HIF1A, …

Visualizing GSEA Results - Enrichment Plots

- Show the distribution of a gene set across a ranked list
- Peaks indicate where gene sets are enriched; steep peaks suggest strong enrichment



https://www.researchgate.net/figure/Gene-set-enrichment-analysis-GSEA-of-genes-highly-expressed-in-cells-that-escaped-from_fig3_306034299

Let’s finish the GSEA tutorial