

T02 - Integrando Redes

Dado o a lista de 2.000 genes diferencialmente expressos extraídos do artigo:

Radich, J. P., Dai, H., Mao, M., Oehler, V., Schelter, J., Druker, B., Sawyers, C., Shah, N., Stock, W., Willman, C. L., Friend, S., & Linsley, P. S. (2006). Gene expression changes associated with progression and response in chronic myeloid leukemia. *Proceedings of the National Academy of Sciences of the United States of America*, 103(8), 2794–2799.

<https://doi.org/10.1073/PNAS.0510423103>

Os genes estão disponíveis no arquivo `most-differentially-expressed-genes.csv`, que além do código único do gene (Gene.ID) seu símbolo oficial (Gene.symbol) e o Log do Fold Change (logFC).

A partir deste arquivo, construa um grafo apresentando as interações físicas entre estes genes, associado ao grafo correspondente ao Wnt signaling pathway, conforme detalhado a seguir:

1. Encontre as relações entre os genes usando o STRING (<https://string-db.org/>). Configure o STRING (em Settings) para usar somente physical subnetwork e active interaction sources de: Experiments e Databases. Exporte este grafo.
2. Gere um no STRING o grafo do Wnt signaling pathway do KEGG (use a busca do Pathway no KEGG). Exporte este grafo.
3. Combine todos os dados no Neo4j. Importe os genes e sua expressão diferencial de `most-differentially-expressed-genes.csv`, bem como as redes geradas por (1) e (2), tendo o cuidado de colocar no mesmo nó todos os dados do mesmo gene.
4. Exporte o grafo integrado para ser lido no Cytoscape.
5. Monte no Cytoscape uma apresentação visual do grafo integrado. Diferencie os genes que fazem parte do pathway de algum modo (exemplos: forma, tamanho, etc.). Para os nós que tem valor de expressão diferencial, apresente-os em cores de uma escala de cores (por exemplo, de vermelho a azul) conforme a expressão.

A tarefa deve ser realizada por trio. Deve ser submetido apenas o arquivo final do Cytoscape e uma imagem do grafo final.



Disciplina interdisciplinar baseada em problemas, que envolve aplicação de ciência de dados em saúde. Este ano tem enfoque em saúde no contexto de biologia de sistemas e ciência de redes.