

T03 - Análise de Enriquecimento de Conjuntos de Genes (GSEA) e Análise de Rede das vias enriquecidas

Enunciado do Problema:

Experimentos destinados a quantificar a expressão genética em larga escala frequentemente resultam em centenas de genes que exibem diferenças estatisticamente significativas entre os grupos comparados (por exemplo, um grupo de pacientes comparado com um grupo controle). Após a identificação dos genes diferencialmente expressos, métodos de análise de enriquecimento de vias podem ser utilizados para identificar os mecanismos biológicos subjacentes ao fenômeno estudado. Um exemplo desse método é a Análise de Enriquecimento de Conjuntos de Genes usando o teste hipergeométrico. Usaremos dados extraídos de uma análise de expressão diferencial que comparou dados de expressão gênica entre dois grupos de pacientes com artrite reumatóide (pacientes com erosão óssea vs pacientes sem erosão). Somente os genes com redução do nível de expressão em pacientes com erosão óssea serão analisados (down-regulated).

Dados Fornecidos:

- Lista de DEGs com a magnitude do nível de expressão (Sup-Table-4-List_of_Down_regulated_Erosion_vs_Without-Erosion.csv)
- Conjuntos de genes com as vias biológicas (gene_sets_GOBP.gmt)

Tarefas:

1. Realizar a análise de Enriquecimento usando o método GSEA:
 - Ler o arquivo Sup-Table-4-List_of_Down_regulated_Erosion_vs_Without-Erosion.csv contendo IDs de genes ("Gene_name") que serão usados.
 - Realizar a análise de enriquecimento usando a lista DEG fornecida e os conjuntos de genes (gene_sets_GOBP.gmt), e calcular os escores de enriquecimento (combined score) e valores de p corrigido para testes múltiplos (por exemplo, correção FDR). Essa etapa pode ser realizada com a biblioteca gseapy no python usando o algoritmo do Enrichr ou na plataforma web do Enrichr (<https://maayanlab.cloud/Enrichr/>). (Referência: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-128#Sec2>)
 - Identificar conjuntos de genes significativamente enriquecidos com base em valores de p ajustados (p ajustado $< 0,05$).
2. Usando o resultado da análise de enriquecimento, construir uma rede das vias enriquecidas. A rede terá dois tipos de nós: (i) nós representando as vias enriquecidas; e (ii) nós representando os genes associados a estas vias. As vias e genes serão conectadas com arestas. Desta forma um gene pode ser associado a uma ou várias vias.
 - Calcular uma medida de centralidade para cada via (pode ser o *Degree* ou o *betweenness centrality*, ou outra medida) e representar o tamanho dos nós das vias com essa medida.
 - Quais são as duas vias mais importantes com base na medida de centralidade?
 - Quais são os 3 genes hub associados às vias?
3. Detetar as comunidades presentes nessa rede. Usar pelo menos três algoritmos de detecção de comunidades.

4. Qual é a interpretação biológica dos achados?

Entrega:

Deve ser entregue um documento PDF com o texto da análise e um arquivo do Cytoscape com a rede e as respectivas métricas aplicadas.