Análise de Expressão Diferencial e Enriquecimento Gênico (RNA-seq)

📌 Sobre o Projeto

Este projeto foi desenvolvido como parte do meu aprendizado em bioinformática, seguindo como referência o tutorial [RNA-seq Analysis in Python](https://www.youtube.com/watch?v=wIvxFEMQVwg) por [Nome do Criador, se disponível]. Todos os direitos sobre a metodologia original pertencem ao criador do tutorial. Este repositório contém minha implementação pessoal com adaptações e extensões para fins educacionais.

📌 Visão Geral

O workflow inclui:

Pré-processamento de dados de expressão gênica

Análise diferencial (DESeq2/pyDESeq2)

Gene Set Enrichment Analysis (GSEA)

Visualização dos resultados (clustermaps, volcano plots, enrichment plots)

🔧 Dependências

python

Copy

# Pacotes principais

Python >= 3.8

pandas

numpy

scipy

matplotlib

seaborn

# Bioinformática

gseapy >= 1.0.0

pydeseq2

scanpy (opcional para análise single-cell)

📊 Principais Análises

1. Expressão Diferencial

python

Copy

# Exemplo com pyDESeq2

dds = DeseqDataSet(

counts=count\_matrix,

metadata=sample\_metadata,

design\_factors="Condition"

)

dds.deseq2()

results = DeseqStats(dds).summary()

2. Análise de Enriquecimento (GSEA)

python

Copy

gsea\_results = gp.prerank(

rnk=gene\_ranking,

gene\_sets=['GO\_Biological\_Process\_2023', custom\_genes],

permutation\_num=1000

)

3. Visualização

Clustermap de expressão:

python

Copy

sns.clustermap(

expression\_matrix,

z\_score=0,

cmap='RdYlBu\_r',

figsize=(10,12)

)

GSEA Plot:

python

Copy

gseaplot(

rank\_metric=ranking,

term='GO\_term\_of\_interest',

\*\*gsea\_results['GO\_term\_of\_interest']

)

💡 Insights Biológicos

Os resultados permitem identificar:

Vias biológicas significativamente enriquecidas

Padrões de co-expressão gênica

Mecanismos moleculares ativados/repirimidos entre condições

Nota Legal:  
Este projeto foi desenvolvido exclusivamente para fins educacionais. A metodologia básica foi adaptada do tutorial vinculado, com modificações para acomodar versões atualizadas de pacotes e necessidades específicas do meu dataset. O código neste repositório não tem fins comerciais.