

Módulo 1: Download e Processamento com Galaxy

Dataset sugerido do GEO:

Exemplo: [GSE117795 - abrir link e explorar os dados](#)

Organismo: Homo sapiens

Plataforma: Illumina

Condições: Knockdown vs Controle

Passos no Galaxy:

1. Acesse o Galaxy: <https://usegalaxy.org>
2. Crie uma conta
3. Use a ferramenta "**Get Data > UCSC Main table browser**" ou "**Upload**" arquivos do GEO.
4. Use **FastQC** para checar qualidade.
5. Use **Trimmomatic** (opcional) para cortar adaptadores.
6. Faça alinhamento com **HISAT2**.
7. Faça contagem com **featureCounts**.
8. Exporte a **matriz de contagem** (TSV) para análise no R.

Módulo 2: Migrando para o R

1. Instale os pacotes e siga o código
2. Para entender os contrastes, consulte a apostila

Módulo 3: Visualizações e Enriquecimento

1. Seguir no código

Módulo 4: Comparação Galaxy vs R + Relatório

Comparando resultados:

- Compare os arquivos **DEGs.csv** gerado via R com os outputs do Galaxy (**DESeq2** wrapper).
- Verifique a consistência de genes expressos diferencialmente.
- Aplique filtros idênticos ($p_{adj} < 0.05$ e $|\log_2FC| > 1$).

Extra!



1. Introdução ao formato e lógica de arquivos RNA-seq

- Explicação dos formatos FASTQ, BAM/SAM, GTF/GFF, Count Matrix.
- Qualidade dos dados e o que é medido (contagens de transcritos, isoformas).
- Como interpretar um arquivo de contagens.



2. Controle de Qualidade Profundo

- Comparar QC com FastQC e MultiQC.
- Interpretar gráficos de qualidade (per base quality, GC content, adapter content).
- Identificar quando fazer trimming.



3. Trimming e pré-processamento avançado (opcional)

- Quando remover adaptadores e qual ferramenta usar.
- Comparar Trimmomatic vs fastp (Galaxy e CLI).

4. Comparação entre pipelines de mapeamento

- HISAT2 vs STAR vs Salmon (quasi-mapeamento).
- Quando usar alinhamento direto versus pseudoalignment (quantificação rápida).
- Galaxy: Salmon wrapper.

5. Análise de expressão gênica em nível de transcrito

- Introdução a `tximport` no R para usar com quantificações do Salmon.
- DESeq2 com nível de transcritos agregados.

6. Comparações múltiplas / contrastes personalizados

- Como usar `contrast = c(...)` para diferentes grupos.
- Comparar controle x tratado, tratado A vs tratado B, etc.
- Comparações complexas: tratamento ao longo do tempo, interação entre genótipos.

7. Análise de enriquecimento mais sofisticada

- Visualizações com `cnetplot`, `emapplot`, e `gseaplot` (clusterProfiler).
- GSEA (Gene Set Enrichment Analysis) com `fgsea` ou `clusterProfiler::GSEA`.
- Importar conjuntos de genes personalizados (.gmt).

8. Exportação e reprodutibilidade

- Exportar todos os resultados em CSV ou Excel.

- Gerar scripts reprodutíveis com RMarkdown.
 - Salvar ambiente R e script `.R` passo a passo.
-

9. Visualizações interativas

- Volcano plot com `EnhancedVolcano`.
 - PCA com `plotly::ggplotly` para exploração interativa.
 - App Shiny básico para explorar os resultados.
-

10. Introdução à integração com outras camadas ômicas

- Como combinar expressão gênica com:
 - Metilação de DNA
 - ATAC-seq
 - Proteômica

AVALIAÇÃO FINAL

Analise um dataset real do GEO, gere a matriz de contagem, realize o DESeq2, produza os gráficos e interprete os 10 genes mais significativos.