

Exercício 6: Análise Detalhada de um Dataset de Expressão Gênica

Objetivo:

- Consolidar os conhecimentos em análise de dados de expressão gênica utilizando o R.
- Explorar diversas ferramentas e técnicas para uma análise mais profunda.
- Desenvolver habilidades na interpretação e visualização de resultados.

Dataset:

Utilizaremos o dataset GSE10072 do GEO como ponto de partida. No entanto, para este exercício, sugerimos que o aluno escolha um dataset de seu interesse, relacionado à sua área de pesquisa.

Carregamento dos dados e pré-processamento:

Code snippet

```
library(GEOquery)
library(tidyverse)

# Baixando os dados do GEO
gse <- getGEO("GSE10072")
exprSet <- gse[[1]]

# Transformando os dados em um data frame
exprData <- exprs(exprSet)
```

Análise exploratória:

Code snippet

```
# Boxplot da expressão gênica
boxplot(exprData, main = "Distribuição da expressão gênica")

# PCA
pca <- prcomp(t(exprData))
plot(pca$x[,1:2], col = factor(pData(exprSet)$group),
     xlab = "PC1", ylab = "PC2")
```

Identificação de genes diferencialmente expressos (DEGs):

Code snippet

```
library(limma)
```

```
# Criando um design experimental
design <- model.matrix(~0 + group, data = pData(exprSet))
colnames(design) <- levels(pData(exprSet)$group)

# Ajustando o modelo linear
fit <- lmFit(exprSet, design)
cont.matrix <- makeContrasts(groupB-groupA, levels=colnames(design))
fit2 <- contrasts.fit(fit, contrasts=cont.matrix)
fit2 <- eBayes(fit2)

# Top DEGs
topTable(fit2)
```

Enriquecimento de vias bioquímicas:

Code snippet

```
library(clusterProfiler)

# Vetor com os genes DEGs
deg <- topTable(fit2, number=Inf)$Gene.symbol

# Enriquecimento KEGG
keggResult <- enrichKEGG(gene = deg, organism = 'hsa', pAdjustMethod =
"BH")
dotplot(keggResult, showCategory=10)
```

Integração com outras fontes de dados:

- Utilizar pacotes como sva para corrigir efeitos de lote.
- Integrar dados de metilação usando pacotes como minfi.

Análise mais profunda:

- **WGCNA:** Identificar módulos de co-expressão de genes.
- **GSEA:** Analisar enriquecimento de conjuntos de genes pré-definidos.
- **Modelos de rede:** Construir modelos de rede para simular a dinâmica dos sistemas biológicos.

Observações:

- **Escolha do método:** A escolha do método de análise depende do tipo de dados, do design experimental e das perguntas de pesquisa.

- **Interpretação dos resultados:** É fundamental interpretar os resultados no contexto biológico e validar os achados com a literatura existente.
- **Visualização:** A visualização dos dados é essencial para a compreensão dos resultados. Utilize diferentes tipos de gráficos para explorar os dados de diferentes perspectivas.

Próximos passos:

- **Explorar outros datasets:** Aplicar as técnicas aprendidas em outros datasets para consolidar o conhecimento.
- **Desenvolver projetos independentes:** Propor projetos de análise de dados de expressão gênica relacionados à sua área de interesse.
- **Aprofundar o conhecimento:** Continuar estudando novas ferramentas e técnicas para análise de dados biológicos.

Lembre-se: Este é apenas um exemplo de como realizar uma análise completa de dados de expressão gênica. A complexidade da análise pode variar dependendo do dataset e das perguntas de pesquisa.