

WGCNA共表达网络分析

1. 加载R包

- 1.1 读取R包: `library()`
- 1.2 启用多线程: `enableWGCNAThreads()`

2. 数据导入与清洗

- 2.1 读取表达矩阵: `read.csv()`
- 2.2 检查样本数: `stopifnot()`
- 2.3 过滤低表达基因: `rowSums()`、`expr_data[]`
- 2.4 离群样本检测: `hclust()`、`plot()`

3. 网络构建与模块识别

- 3.1 选择软阈值 β : “`powers <- c(5:20)`”、`pickSoftThreshold()`、`abline()`、“`power_use <- 14`”
- 3.2 构建网络: “`cor <- WGCNA::cor`”、`blockwiseModules()`
- 3.3 模块可视化: `labels2colors()`、`plotDendroAndColors()`

4. 模块-性状关联分析

- 4.1 计算模块特征值 (ME): `MEs <- net$MEs`、`MEs <- orderMEs(MEs)`
- 4.2 读取性状数据: `read.delim()`
- 4.3 分类变量数值化: `Gender`、`Disease_Stage`、`Treatment`
- 4.4 连续变量标准化: `Age`、`Survival_days`、`Glucose_mmolL`
- 4.5 查看预处理后数据: `head(trait_data)`
- 4.6 偏相关控制混杂因素: `matrix()`
- 4.7 绘制热图: `sizeGrWindow(10,6)`、`labeledHeatmap()`

5. 功能富集与关键基因挖掘

- 5.1 提取目标模块 (如蓝色模块): `expr_filt[blue_module,]`
- 5.2 GO富集分析 (放宽阈值): `bitr()`、`enrichGO()`
- 5.3 可视化 (Top10通路): `dotplot()`
- 5.4 识别枢纽基因 (模块内连接度Top10): `TOMsimilarity(adjacency)`、`intramodularConnectivity()`、`print(top_hubs)`

常见问题解析

1. 分析失败或结果异常

①软阈值无法满足 $R^2 > 0.8$ ，原因是样本异质性高（如批次效应或离群样本），可检查样本聚类树，剔除离群样本，或改用经验Power值；②模块数量过多/过少，需调整参数：增大minModuleSize减少模块数或降低deepSplit增加模块大小。

2. 模块-性状关联不显著

①性状数据未数值化（如分类变量未转换），需重新检查性状矩阵格式；②样本量不足或性状变异度低，可增加样本量或选择变异更大的性状。

参考文献

- [1] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis [J]. BMC Bioinformatics, 2008, 9: 559.
- [2] Rao X, Zhou Y, Xue J, Zhou Z. Identification of DNA damage repair-related diagnostic markers for idiopathic pulmonary fibrosis via WGCNA [J]. Comput Biol Chem, 2025, 119: 108563.