

BIOF3 组学数据分析

04 WGCNA 与共表达模块

导出日期：2026年5月12日

04 WGCNA 与共表达模块

WGCNA (Weighted Gene Co-expression Network Analysis) 通过构建基因共表达网络，把上万个基因压缩成几十个模块，每个模块代表一组协同变化的基因。在多组学整合中，WGCNA 的模块可以作为"特征压缩"的手段，也可以用来检验某个模块在不同数据集或不同组学层之间是否保守。

基本流程

```
library(WGCNA)
allowWGCNAThreads()

# 输入：行是样本，列是基因（注意转置）
datExpr <- t(rna_final)

# 1. 选择软阈值 (soft threshold)
powers <- c(1:20)
sft <- pickSoftThreshold(datExpr, powerVector = powers, verbose = 3)

# 画 scale-free topology fit 图
plot(sft$fitIndices[, 1], -sign(sft$fitIndices[, 3]) * sft$fitIndices[, 2],
     xlab = "Soft Threshold (power)",
     ylab = "Scale Free Topology Model Fit (signed R^2)",
     type = "n", main = "Scale independence")
text(sft$fitIndices[, 1], -sign(sft$fitIndices[, 3]) * sft$fitIndices[, 2],
     labels = powers, col = "red")
abline(h = 0.85, col = "red")
```

选择 R^2 首次超过 0.85 的 power 值。通常 RNA-seq 数据在 6-12 之间。

构建网络与模块识别

```
# 2. 一步法构建网络
net <- blockwiseModules(datExpr,
  power = 8,
  TOMType = "unsigned",
  minModuleSize = 30,
  reassignThreshold = 0,
  mergeCutHeight = 0.25,
  numericLabels = TRUE,
  pamRespectsDendro = FALSE,
  verbose = 3)

# 模块颜色
moduleColors <- labels2colors(net$colors)
table(moduleColors)

# 画聚类树 + 模块颜色
plotDendroAndColors(net$dendrograms[[1]], moduleColors[net$blockGenes[[1]],
  "Module colors", dendroLabels = FALSE,
  hang = 0.03, addGuide = TRUE, guideHang = 0.05)
```

模块与表型的关联

每个模块用 module eigengene (ME, 即模块内基因表达的第一主成分) 来代表。然后计算 ME 与临床表型的相关:

```
MEs <- net$MEs
# 计算 ME 与表型的相关
trait_data <- data.frame(
  subtype = as.numeric(factor(col_data$subtype)),
  stage = as.numeric(factor(col_data$stage))
)

cor_ME_trait <- cor(MEs, trait_data, use = "p")
pval_ME_trait <- corPvalueStudent(cor_ME_trait, nrow(datExpr))

# 热图展示
library(ComplexHeatmap)
Heatmap(cor_ME_trait, name = "Correlation",
  cell_fun = function(j, i, x, y, w, h, fill) {
    if (pval_ME_trait[i, j] < 0.05)
      grid.text("*", x, y)
  })
```

模块保守性分析

如果你有两个独立数据集 (比如 TCGA 和自己的队列), 可以检验某个模块在第二个数据集中是否保守。这用 `modulePreservation` 函数:

```
# 准备第二个数据集
datExpr2 <- t(rna_validation)

# 设置多集数据
multiExpr <- list(
  Discovery = list(data = datExpr),
  Validation = list(data = datExpr2)
)
multiColor <- list(Discovery = moduleColors)

# 计算保守性
mp <- modulePreservation(multiExpr, multiColor,
  referenceNetworks = 1,
  nPermutations = 200,
  randomSeed = 1,
  verbose = 3)

# Zsummary > 10 表示高度保守, 2-10 中等, < 2 不保守
stats <- mp$preservation$Z$ref.Discovery$inColumnsAlsoPresentIn.Validation
print(stats[, c("moduleSize", "Zsummary.pres")])
```

在多组学整合中的角色

WGCNA 模块在整合分析中有两个用途:

1. **特征降维**: 用 ME 代替上千个基因, 减少后续整合模型的输入维度。比如把 20 个模块的 ME 和蛋白组数据一起送进 MOFA2。
2. **跨层验证**: 在 RNA 层发现的模块, 检查对应基因在蛋白层或甲基化层是否也有一致的模式。

参考资源

- [WGCNA 官方教程](#)
- [WGCNA FAQ](#)
- [Langfelder & Horvath 2008 原始论文](#)
- [modulePreservation 说明](#)



扫码关注微信公众号【生信F3】

获取文章完整内容, 分享生物信息学最新知识。