

## BIOF3 组学数据分析

# 03 跨层相关性探索

导出日期：2026年5月12日

## 03 跨层相关性探索

在跑复杂的整合模型之前，先做一轮简单的相关性分析。它能告诉你：哪些基因在不同组学层之间有一致的信号，哪些层之间整体相关性高，以及数据里有没有明显的异常样本。

### 全局相关性：RV 系数

RV 系数衡量两个矩阵之间的整体相似度，取值 0-1，类似于矩阵层面的 Pearson 相关。

```
library(FactoMineR)

# 计算两两 RV 系数
rv_rna_meth <- coeffRV(t(rna_final), t(meth_final))$rv
rv_rna_prot <- coeffRV(t(rna_final), t(prot_final))$rv
rv_meth_prot <- coeffRV(t(meth_final), t(prot_final))$rv

cat("RNA vs Methylation RV:", round(rv_rna_meth, 3), "\n")
cat("RNA vs Protein RV:", round(rv_rna_prot, 3), "\n")
cat("Methylation vs Protein RV:", round(rv_meth_prot, 3), "\n")
```

通常 RNA 和蛋白的 RV 系数在 0.3-0.6 之间，甲基化和表达的相关性更低（因为甲基化主要是负调控，且只有 promoter 区域的甲基化与表达强相关）。

### Cis 相关分析

Cis 相关指的是同一个基因在不同组学层之间的相关性。最经典的例子是：某个基因的 promoter 甲基化水平与它自身的 mRNA 表达之间的 Pearson 相关。

```
# 找到两层都有的基因
common_genes <- intersect(rownames(rna_final), rownames(meth_final))

# 逐基因计算相关
cis_cor <- sapply(common_genes, function(g) {
  cor(as.numeric(rna_final[g, ]),
      as.numeric(meth_final[g, ]),
      method = "pearson", use = "complete.obs")
})

# 分布
summary(cis_cor)
hist(cis_cor, breaks = 50, main = "Cis correlation: RNA vs Methylation",
     xlab = "Pearson r", col = "steelblue")
```

预期结果：大部分基因的 cis 相关接近 0，少部分呈现强负相关（promoter 甲基化抑制表达）。如果你看到大量正相关，可能是映射到了 gene body 而不是 promoter。

## 跨层相关热图

用 ComplexHeatmap 画一张跨层相关矩阵，展示 top 变异基因在不同组学层之间的关系：

```
library(ComplexHeatmap)
library(circlize)

# 选 top 500 变异基因 (在 RNA 层)
rv <- apply(rna_final, 1, var)
top_genes <- names(sort(rv, decreasing = TRUE))[1:500]

# 只保留三层都有的
top_common <- Reduce(intersect, list(top_genes,
                                     rownames(meth_final),
                                     rownames(prot_final)))

# 计算 RNA vs Protein 的逐基因相关
rna_prot_cor <- sapply(top_common, function(g) {
  cor(as.numeric(rna_final[g, ]),
      as.numeric(prot_final[g, ]),
      use = "complete.obs")
})

# 热图：行是基因，列是相关类型
cor_mat <- cbind(
  RNA_Meth = cis_cor[top_common],
  RNA_Prot = rna_prot_cor[top_common]
)

col_fun <- colorRamp2(c(-1, 0, 1), c("blue", "white", "red"))
Heatmap(cor_mat, name = "Pearson r",
        col = col_fun,
        cluster_columns = FALSE,
        show_row_names = FALSE,
        column_title = "Cross-layer cis correlation")
```

## Trans 相关的注意事项

Trans 相关（一个基因的甲基化与另一个基因的表达）在计算上是  $O(n^2)$  的，特征数多时不现实。常见策略：

- 只算已知调控关系的 pair（比如 TF 和 target）
- 先用 WGCNA 找模块，再算模块间相关
- 用 MOFA2 的因子来间接捕捉 trans 效应

## 相关性不等于因果

高相关不代表一层驱动了另一层。甲基化和表达的负相关可能是因果（甲基化沉默基因），也可能是共同受上游调控。如果需要因果推断，考虑 Mendelian randomization 或 mediation analysis。

## 参考资源

- [ComplexHeatmap 完整手册](#)

- [FactoMineR RV 系数](#)
- [circlize 颜色映射](#)
- [Methylation-expression cis 分析综述 \(Bock 2012\)](#)



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。

