科研绘图 | 一文解决差异表达基因聚类热图分析

Hierarchical clustering and heatmap analysis

通过 “**DESeq2 差异分析优化版**” 该文得到差异表达基因后，进一步对差异基因进行聚类分析有助于我们更加充分的理解基因表达的调控模式。具体案例见以下几篇本实验室文献：

Reference:

Zeng et al., Nature communications. 2021 (Fig. 5b)

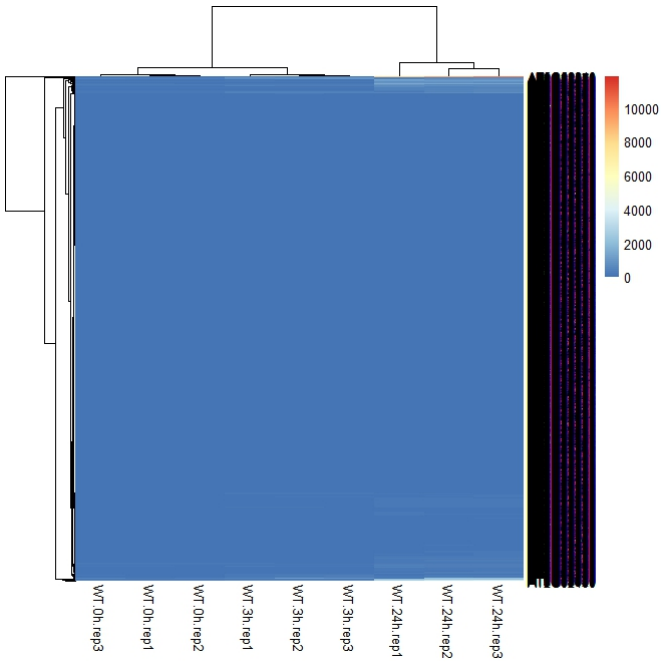
Li et al., Plant cell. 2022 (Fig. 4c)

Jiang et al., Nature plants. 2022 (Fig. 2c)

本文将使用 R 包-**pheatmap** 进行差异表达基因的可视化及聚类分析，示例数据文件名为 **Gene_expresion.txt**，包含低温处理 0h，3h 和 24h 后 2790 个基因在野生型 (WT) 样品中的基因表达 (**TPM**) 定量数据，每个样品各三个生物学重复。

1. 读取数据并简单绘图

```
setwd("E:/生信总结/Figure/Heatmap") #设置工作路径
library(pheatmap) #读取R包
data<-read.table("Gene_expresion.txt",header=T,row.names=1,sep="\t") #读取基因表达文件
range(data) #查看数据最小值与最大值
pheatmap(data) #默认参数绘图
```

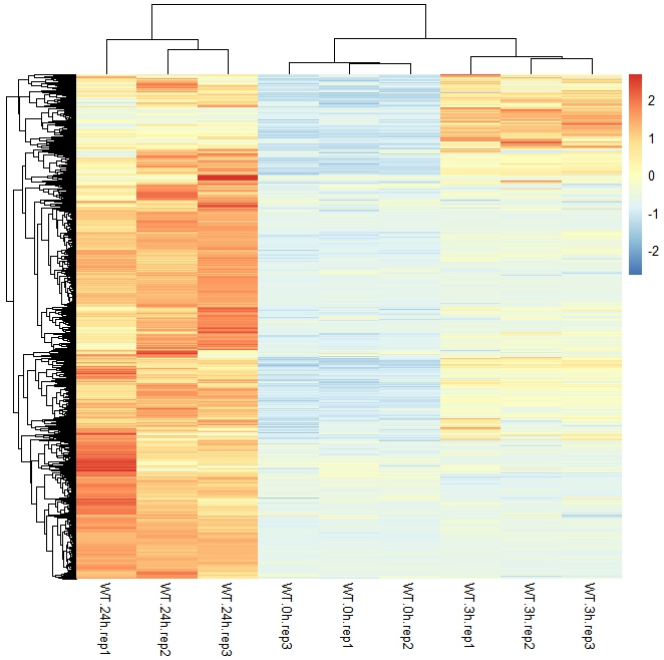


2. 通过 z-score 进行归一化

2.1 输入数据集最大值与最小值相差过大，使用 **scale = "row"** 进行归一化，并用 **show_row=F** 去掉行名；

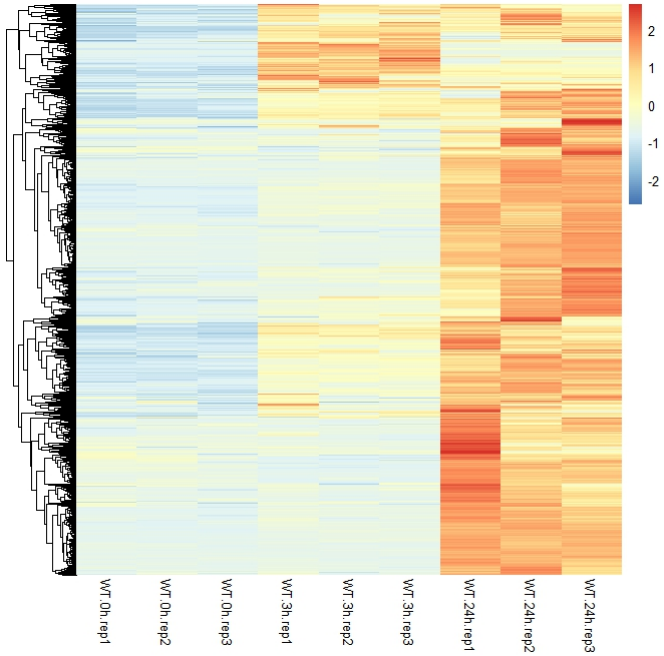
此外需注意，也可对 TPM 进行对数转换，比如 **log10 (TPM+1)**，转换后若最大值与最小值差异小可不使用 scale 参数进行归一化

```
pheatmap(data,scale = "row",show_row=F)
```



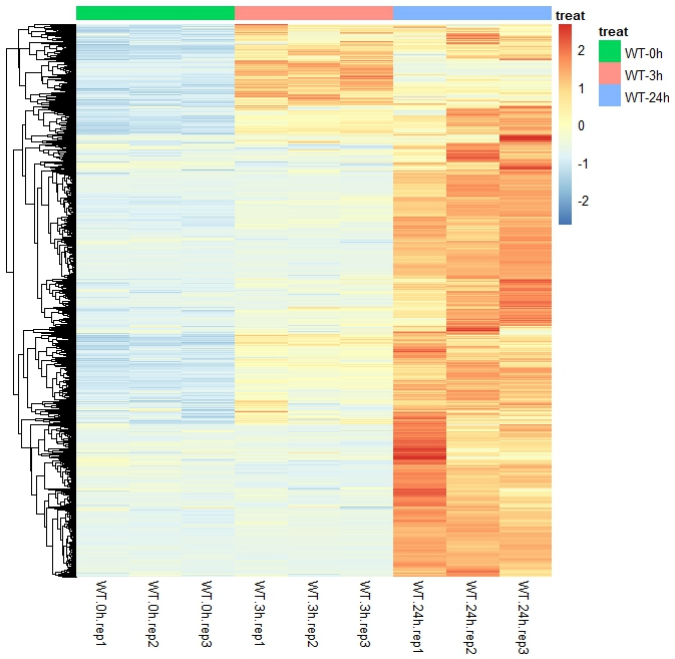
2.2 pheatmap 默认对行列均进行聚类，可修改 **cluster_col=** 和 **cluster_row=** 参数进行修改

```
pheatmap(data,scale = "row",show_row=F,cluster_col=FALSE,cluster_row=TRUE)
```



2.3 使用 **annotation_col** 参数构建列注释信息

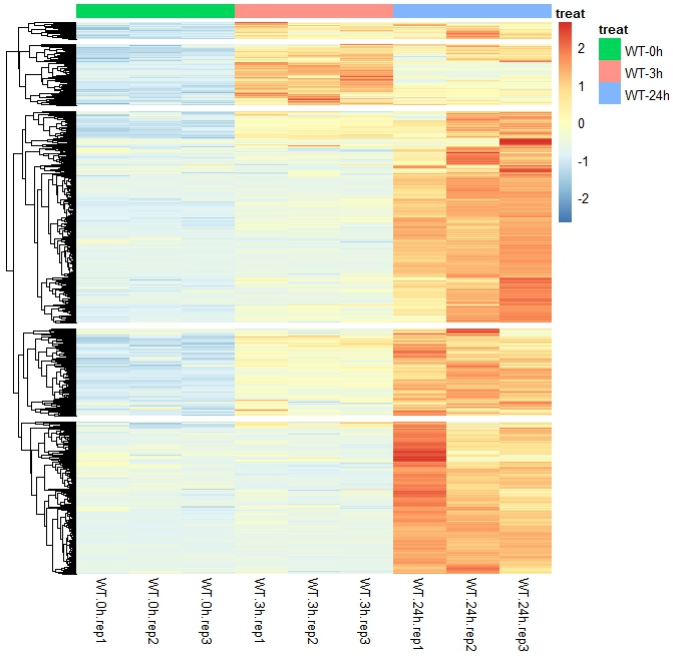
```
ann_col = data.frame(
  treat=c(rep("WT-0h",3),rep("WT-3h",3),rep("WT-24h",3)
)) #根据样品数据构建一个数据集
row.names(ann_col) <- colnames(data) #更改行名使其与输入数据保持一致
ann_col$treat = factor(ann_col$treat , levels=c('WT-0h','WT-3h','WT-24h')) #对注释信息进行排序
pheatmap(data,scale =
"row",show_row=F,cluster_col=FALSE,cluster_row=TRUE,annotation_col = ann_col)
```



3. 对聚类后的数据根据表达量变化模式分 cluster

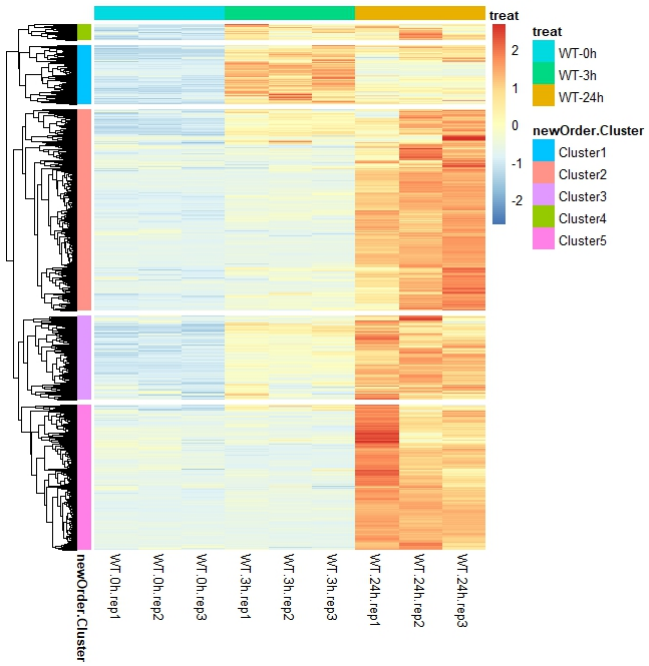
3.1 通过 cutree_rows 参数分簇，并提取分簇后的数据集

```
list=pheatmap(data,scale = "row")
row_cluster=cutree(list$tree_row,k=5) #通过k分五个cluster，可修改
newOrder=data[list$tree_row$order,] #将聚类及分簇后的数据集提取为新的data
newOrder[,ncol(newOrder)+1]=row_cluster[match(rownames(newOrder),names(row_clust
er))]
colnames(newOrder)[ncol(newOrder)]="Cluster"
write.table(newOrder,file ="cluster.csv",sep =",",quote=FALSE)
pheatmap(data,scale =
"row",show_row=F,cluster_col=FALSE,cluster_row=TRUE,annotation_col =
ann_col,cutree_rows = 5)
```



3.2 根据 cluster 构建行的注释信息

```
ann_row = data.frame(
  newOrder$Cluster
  ) #提取分簇信息
ann_row$newOrder.Cluster<-paste('Cluster',ann_row$newOrder.Cluster,sep="")
row.names(ann_row) <- rownames(newOrder) #修改行名
pheatmap(data,scale =
"row",show_row=F,cluster_col=FALSE,cluster_row=TRUE,annotation_col =
ann_col,cutree_rows = 5,annotation_row = ann_row)
```



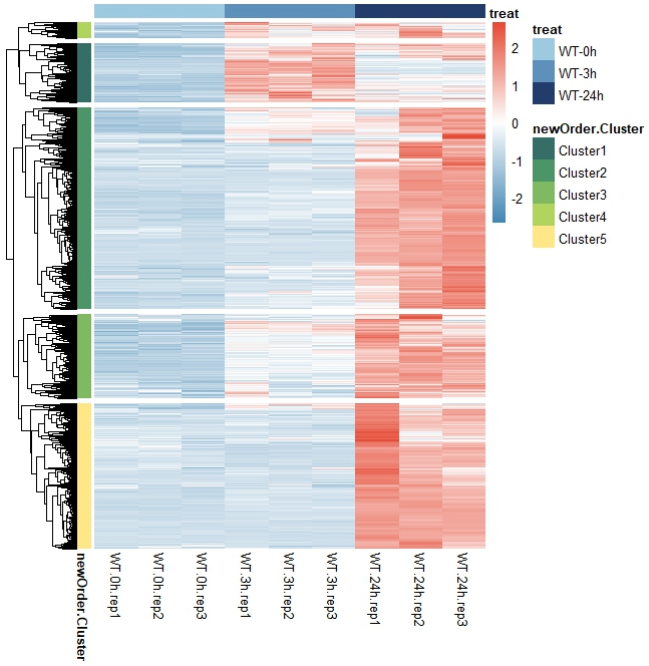
3.3 修改颜色

```
#参数 annotation_colors 修改注释颜色
treat_color <- c("#9ecae1","#5D90BA","#223D6C")
names(treat_color) <- c("WT-0h","WT-3h","WT-24h")

cluster_color <- c("#356d67","#4c9568","#7fb961","#b0d45d","#ffe788")
names(cluster_color) <-
c("Cluster1","Cluster2","Cluster3","Cluster4","Cluster5")

ann_colors <- list(treat=treat_color,newOrder.Cluster=cluster_color)

#参数 color=colorRampPalette 对应单元格颜色
pheatmap(data,scale =
"row",show_row=F,cluster_col=FALSE,cluster_row=TRUE,annotation_col =
ann_col,cutree_rows = 5,annotation_row =
ann_row,color=colorRampPalette(c("#4387B5","white","#E64B35B2"))
(100),annotation_colors = ann_colors)
```



以上为常用参数信息，其它修改信息自行查看官方文档!