**“Linux 生物信息技术基础”课程小组总结报告**

组：G12 次：11 组长：马小凯 执笔：陈晨

1. **时间**

2023年 5 月 21 日

1. **方式**

线上腾讯会议

1. **主题**

RNA-seq数据的研究分析

1. **内容**
2. 关于RNA-seq的分析工具

通过同学上课时候的分享以及Pereira在2017年文章的记述，我们找到了以上功能的不同分析工具，并对其中部分进行了研究学习。

1. 单细胞测序数据的差异表达分析方法Differential expression

首先是对单细胞测序数据的差异表达分析方法Differential expression进行了学习。

无论是传统的多细胞转录组测序（bulk RNA-seq）还是单细胞转录组测序（scRNA-seq），差异表达分析（differential expression analysis）是比较两组不同样本基因表达异同的基本方法，可获得一组样本相对于另一组样本表达显著上调（up-regulated）和下调的基因（down-regulated），从而可进一步研究这些差异表达基因的功能，包括富集的通路（pathway）或生物学过程（biological process）。由于单细胞测序技术的局限性，单细胞测序数据通常具有高噪音，有较高的dropout问题，即很多低表达或中度表达的基因无法有效检测到。所以，以前针对传统多细胞转录组测序数据开发的差异表达检测方法或软件不一定完全适用于单细胞测序数据。若想比较不同细胞亚型或不同条件下的细胞表达差异时，为了能得到可靠的结果，需要选定一个好的差异表达分析方法。

近年来，有不少专门针对单细胞转录组测序数据的差异表达分析方法相继被开发出来，如MAST (Finak et al., 2015)、SCDE (Kharchenko et al., 2014)、 DEsingle (Miao et al., 2018)、 Census (Qiu et al., 2017)、 BCseq (Chen and Zheng, 2018)等。

根据易同学分享的PPT，我们使用了其中的DESeq2工具进行了相关内容的学习分析。

DESeq2介绍：

DESeq2是一个为高维计量数据的归一化、可视化和差异表达分析而设计的一个R语言包。它通过经验贝叶斯方法(empirical Bayes techniques)来估计对数倍数变化(log2foldchange）和离差的先验值，并计算这些统计量的后验值。它由美国北卡罗莱纳大学教授Michael Love(michaelisaiahlove@gmail.com)于2014年发布，目前仍在更新与维护中，是目前差异表达分析方面最常用的R包。

DESeq2安装：

在R语言命令行或Rstudio运行以下命令(version "4.2")：

if (!require("BiocManager", quietly = TRUE))

install.packages("BiocManager")

BiocManager::install("DESeq2")

DESeq2的使用：

1. 读入和处理数据，加载表达量矩阵gene\_count\_matrix.csv

# 加载包

library(DESeq2)

library(pheatmap) # 用于作热图的包

library(ggplot2) # 用于作图的包

# 读入数据，注意设置工作路径

countData <- as.matrix(read.csv("gene\_count\_matrix.csv",row.names="gene\_id"))

1. 提前准备信息：包括将分组信息构建成向量，告诉差异分析函数是要分析哪些变量间的差异。
2. 开始进行差异表达分析

dds <- DESeqDataSetFromMatrix(countData = countData, colData = colData, design = ~ condition)

查看结果(利用head)

1. 可视化

#可以直接用DESeq2的plotCounts

dds <- makeExampleDESeqDataSet()

plotCounts(dds1, gene = "AT4G38770",intgroup = "condition") # 指定某个基因

#可用ggplot对图片样式进行修改，并用ggrepel进行标注

d=data.frame(t(subset(countData,rownames(countData)=="AT4G38770")))

ggplot(d, aes(x = condition, y = AT4G38770, color = condition))+

geom\_point(position=position\_jitter(w=0.2,h=0))+

geom\_text\_repel(aes(label=rownames(d)))+

theme\_bw()+

ggtitle("AT4G38770")+

theme(plot.title=element\_text(hjust=0.5))