## 2.2 Seurat 单细胞数据分析 (Dataset: SEPSIS)

* Content: 0 个基因
* Anonymous Comment: 后续这块还是卡严点

Reply: 好的

## 2.6 GEO 数据获取 (Dataset: PB\_GSE63042)

* Content: 缺失样本计数数据的原因包括运行未通过 50% 的对齐率或由于技术原因处理失败)
* Anonymous Comment: 现在GEO数据库有Download RNA-seq counts，用这个也不行吗 , 其次，对齐率低（<50%）的原因一般实在下机数据质控中过滤样本，现在我们下载数据如果方差不齐，可以矫正，如果通过PCA或者热图能明显存在离群样本再进行过滤。

Reply: 系官方文件缺失。已直接沟通。

## 2.7 Limma 差异分析 (Dataset: PB\_GSE63042)

* Content: 过滤 count
* Anonymous Comment: 后续如果是count数据，尽量还是用DESeq2做差异

Reply: …

## 2.9 Limma 差异分析 (Dataset: PB\_GSE74224)

* Content: 创建设计矩阵，对比矩阵，差异分析：Sepsis vs SIRS。使用 limma::lmFit, limma::contrasts.fit, limma::eBayes 拟合线形模型。以 limma::topTable 提取所有结果，并过滤得到 P.Value 小于 0.05，|Log2(FC)| 大于 0.5 的统计结果。
* Anonymous Comment: 这些内容都属于方法，通篇汇总只需要提及一次就行

Reply: 有些数据集的分析会存在不同。

## 2.11 Limma 差异分析 (Dataset: BL\_GSE243217)

* Content: adj.P.Val 小于 0.05，|Log2(FC)| 大于 1 的统计结果
* Anonymous Comment: 同一个项目中，相同分析的阈值要保持一致

Reply: 不同的数据特点不一样。Sepsis vs Healthy 与 Sepsis vs SIRS 采用了不同的阈值。

## 2.14 GEO 数据获取

* Content: 2.14 GEO 数据获取 (Dataset: PB\_GSE69063)
* Anonymous Comment: 最终选择了几个数据集，并进行汇总，内容包含样本数量，使用平台，组织、等临床信息基线表

Reply: 后面汇总的部分，有附加的表格。

## 3.1 Seurat 单细胞数据分析 (SEPSIS)

* Content: 线粒体基因的比例小于 10%
* Anonymous Comment: 过滤之后数据包含多少基因多少细胞，得对结果进行简单描述

Reply: 已调整

* Content:
* Anonymous Comment: 后续尽量展示样本间或者组间

Reply: 好的

* Content:
* Anonymous Comment: 添加一张不分组的umap图

Reply: 已调整

## 3.2 ClusterProfiler 富集分析 (BULK)

* Content:
* Anonymous Comment: 可以添加一份不同细胞间差异基因的富集分析图

Reply: 未添加。

## 3.3 Bulk RNA-seq (Sepsis vs Non-infectious)

* Content: SIRS
* Anonymous Comment: 所有首次缩写都需要全称

Reply: 已添加。

## 3.6 Lasso 诊断模型建立 (BULK)

* Content: 以 GSE134347 (sepsis - noninfectious) 创建诊断模型
* Anonymous Comment: 前面以不同疾病类型的数据集做差异取交集，现在只在sepsis - noninfectious组中进行模型，这中间需要依据，或者说衔接说明

Reply: 其它几个做验证数据集而已。以 GSE134347 做训练集。

* Content: 对应的特征数 (基因数) 分别为 15, 12。使用数据集 GSE134347 (SI), GSE232404 (SH), GSE243217 (SH), GSE63042 (SI), GSE69063 (SH), GSE74224 (SI) 验证。
* Anonymous Comment: 这个感觉意义不大，因为你不存选出的是那几个基因，都会通过验证，原因在于你的训练集和验证集其实没有本质区别

Reply: 你这里理解错了，基因集的选择与这些数据集无关。训练集和验证集相互之间一点关系都没有。

* Content: **Fig.** **23** BULK lambda 1se ROC
* Anonymous Comment: 如果说同一个基因集在不同数据集中的表达趋势相反，而我们只做了ROC，那么所谓的验证会不会存在偷换概念，因为你这么多基因都是在所有数据集中都是差异的，会直接导致你的ROC曲线的结果都会很好

Reply: 同上。这些基因集不是来自这些数据集的差异表达。没有过拟合的问题。

# 4 总结

* Content: 本研究从 T 细胞出发
* Anonymous Comment: 文中除了自己在scFEA出选定的T细胞外，无任何证据证明这个出发点是合适的

Reply: 这个分析的前期思路上，已经提到：“脓毒症会改变能量代谢，导致免疫细胞代谢重编程，从而破坏先天和适应性免疫反应，引发过度炎症和免疫抑制，代谢重编程有望成为未来脓毒症治疗的新靶点。”所以这里，选择 T 细胞出发。此外，文中补充了富集分析的比较，可以作为佐证，联系 T 细胞。

* Content: 探究脓毒症和 SIRS 或 noninfectious 诊断标志物
* Anonymous Comment: 这里需要确认到底是哪个，都到结论了，我们不能不知道我们最终探究的是啥

Reply: 这里探究的是脓毒症和非脓毒症 (Sepsis Negative) 之间的诊断问题。SIRS 与 noninfectious，是同一类问题，相比与 Sepsis 和 Healthy，更难以诊断。文中用了多个数据集做了验证，效果良好。同时也能区分脓毒症和健康贡献者。(SIRS, systemic inflammatory response syndrome，文中的 SIRS 指的是 Infection-Negative Systemic Inflammation，因为这些 GSE 中是标注为 SIRS 的，我这里没有改动。数据集中的 SIRS 都是这一类)