

# 虚拟敲除 + 机器学习 + 单细胞

## 四月新文复现课程目录

### 第1节：AI + 公共数据库发文范式解读

1. 公共数据库 bulk 转录组数据集的筛选与下载
2. 训练集、验证集和单细胞数据整体分工设计
3. 样本分组信息整理与分析对象确定
4. 整篇文章技术路线图与主图逻辑梳理

### 第2节：差异分析

1. 整合疾病与正常样本表达矩阵，完成分组信息整理与分析准备
2. 使用 limma 进行差异表达分析，筛选疾病相关 DEGs
3. 设置差异基因筛选阈值，提取显著上调和下调基因
4. 绘制火山图和差异基因表达图，展示 OA 与正常组转录组差异

### 第3节：WGCNA 筛选候选基因

1. 基于预处理后的表达矩阵构建 WGCNA 共表达网络
2. 选择合适软阈值，识别具有生物学意义的共表达模块
3. 分析模块与疾病表型之间的相关性，筛选关键功能模块
4. 提取关键模块基因，与差异基因结果整合得到候选基因集

### 第4节：交集基因功能富集与通路机制解析

1. 差异基因与共表达模块基因取交集
2. GO 富集分析解析生物学功能
3. KEGG 富集分析定位关键信号通路
4. STRING 构建交集基因 PPI 网络
5. CytoHubba 计算网络关键节点

### 第5节：PPI 网络与机器学习筛选 Hub 基因

1. 构建 Random Forest 模型，对交集基因进行特征重要性排序
2. 提取 Random Forest 排名前列基因，筛选具有稳定判别能力的候选基因
3. 构建 XGBoost 模型，对候选基因进行再次筛选与重要性评估
4. 整合 PPI、Random Forest 和 XGBoost 三部分结果，最终锁定 hub genes

### 第6节：核心基因交叉筛选与外部数据验证

1. PPI、RF、XGBoost 三套结果取交集
2. 锁定 PTPRC、ITGB2、CX3CR1 等核心基因
3. 外部 GEO 数据集验证核心基因表达差异
4. 评估核心基因的稳健性与可重复性

### 第7节：孟德尔随机化验证核心基因因果关系

1. 以核心基因为暴露构建 MR 分析框架
2. 利用 SNP 工具变量评估基因与疾病的因果关系
3. 进行异质性、多效性和留一法敏感性检验
4. 从“相关”推进到“方向性因果证据”

### 第8节：免疫浸润分析与核心基因免疫关联

1. GSVA 评估疾病与正常样本的免疫浸润差异
2. 比较不同免疫细胞在疾病中的变化趋势
3. 分析核心基因与免疫细胞浸润的相关性
4. 建立“核心基因—免疫微环境”联系

### 第9节：单细胞分析——细胞注释与核心基因定位

1. 单细胞数据预处理与降维聚类
2. 主要细胞群注释与亚群识别
3. 观察核心基因在不同细胞类型中的表达分布
4. 明确关键基因主要落在哪类细胞上

### 第10节：虚拟敲除

1. 利用 scTenifoldKnn 进行 PTPRC 虚拟敲除
2. 观察敲除后基因调控网络的变化
3. 推断 PTPRC 对 NK 细胞和免疫环境的影响
4. 完成整篇文章“筛选—因果—单细胞—机制验证”的闭环总结