# 从生信小白快速培训成生信大神的培训课程内容如下

## 常用生信分析

### 数据目录内多样本整合

RNAseq测序数据目录下多样本的合并处理

CEL芯片数据目录下多样本的合并处理

非CEL芯片数据目录下多gsm样本的合并处理

RCC芯片数据目录下多样本的合并处理

### 数据清洗

提取出符合条件的特定行列子数据集

从gpl注释文件中提取出基因注释

将提取的基因注释信息跟表达矩阵合并

从表达矩阵中提取出分组信息

从样本注释文件中提取出分组数据

从样本注释文件中提取出分组和子分组数据

从样本注释文件中提取出分组列和批次效应列

构建ExpressionSet分析对象

按照某些特征的高低阈值对样本进行分组

同一物种不同类型基因id间相互转换

不同物种间同源基因转换

### 质控和标准化

基因表达芯片的质控

表达芯片进行归一化处理

### 样本分群聚类和剔除异常样本

聚类分群分析

剔除异常样本

### 多数据集整合和批次效应去除

多个数据集合并

去除批次效应

### 差异分析

通用差异分析

通用批量差异分析

按照某个基因或特征的高低分组差异表达分析

TCGA的肿瘤-正常差异表达分析

### 富集分析

go和kegg通路富集分析

分析像铁死亡等某些通路marker基因在go,kegg富集的哪些通路里

可视化go和kegg富集到的指定生物学通路

使用go和kegg导入指定通路数据集来富集指定的生物学通路

查看一组基因或某个通路的一组marker基因哪些有差异表达

GSEA富集分析

将多个GSEA富集结果绘制在一张图中

ssGSEA/GSVA单样本通路得分的富集分析

### WGCNA加权共表达分析

1.WGCNA表达数据的预处理和查看是否有离群样本

2.剔除离群样本

3.wgcna的软阈值计算

4.wgcna的网络构建

5.wgcna的性状最相关的模块的散点图可视化

6.wgcna的TOM拓扑矩阵计算和导出到cytoscape

### 甲基化分析

读取数据目录下的idat文件的甲基化全流程一键分析

读取甲基化beta矩阵文件的甲基化全流程一键分析

差异甲基化位点结果的棒棒糖图可视化

DMRs差异甲基化区域的基因组circos图

相关性和互作调控分析

获取某个基因的核心ppi互作蛋白列表

一组基因涉及的两个数值矩阵文件的相关性分析

### 多种结果进行整合

多种结果取交集基因

## 单细胞测序分析

### 读取单细胞数据构建seurat处理对象

读取多个h5文件构建seurat对象

读取表达矩阵文件构建seurat对象

读取10X的matrix,barcodes,features文件构建seurat对象

### 单细胞测序数据读取和处理

1.单细胞测序数据质控

2.多样本整合

3.单细胞数据缩放和PCA降维

### 细胞注释

1.使用SingleR自动进行细胞注释

2.细胞cluster分群聚类,为手动注释细胞做准备

3.挑选最好的分辨率resi,进行分群结果的可视化

4.单细胞聚类后的所有cluster差异分析

使用ucell来根据每个细胞的marker基因集来计算cluster的细胞注释信息

将整理的每个cluster对应的细胞名称文件注释到seurat对象中并对注释好的结果进行可视化

1.3把细胞名称整合到差异分析结果中

提取部分分组的seurat子集进行后续分析

单细胞的差异分析和富集分析

细胞比例在不同分组中的差异比较

单细胞的msigdb的gmt基因集通路富集分析

单细胞自定义基因集的通路富集分析

所有类型细胞中基因在不同分组(如肿瘤vs正常)的差异表达分析

### 目的基因在单细胞中的可视化作图

单独绘制每个基因的feature plot

单独绘制每个基因的violin plot

单独绘制每个基因的box plot

一组基因的feature plot拼图绘制

一组基因的vln plot的拼图绘制

一组基因的dot plot的拼图绘制

一组基因列表的heatmap plot的拼图绘制

### 单细胞测序的细胞通讯分析

对每个分组单独进行所有类型细胞的细胞通讯分析计算

部分感兴趣的细胞的细胞通讯计算

各个分组总细胞通讯贡献度计算和可视化

筛选细胞通讯贡献度lr有显著差异的通路

对贡献度差异通路可视化

### 绘图可视化

### 箱式图绘制

基本箱式图绘制

绘制一个表达矩阵中每个样本的箱式图,展示质控标准化结果

一张图中绘制一组基因基因表达水平

对一组基因每个基因单独绘制一张箱式图

一张图中绘制多个基因的试验vs对照的差异箱式图

单个基因的多类别的试验vs对照差异箱式图

### 分布密度图绘制

常规密度图绘制

使用表达矩阵进行密度图绘制

### 条形图绘制

绘制正负水平对比条形图

没有分组信息,对数据进行melt，样本id映射到x轴，特征变量映射为填充颜色的百分比条形图

对数据进行melt,把分组信息+样本id作为x轴类别名称,特征变量作为填充颜色的百分比条形图

有分组信息，特征变量映射为x轴,用分组信息填充颜色的百分比条形图

绘制分类占比簇状条形图,不加上总类别

绘制分类占比簇状条形图,加上总类别

### 热图绘制

根据所有差异基因绘制基因表达的聚类热图

绘制top差异基因表达的聚类热图

绘制感兴趣基因的聚类热图

### 散点图

火山图

批量火山图

### 可视化多种结果的交集

韦恩图

## 数据库挖掘

### GEO数据下载处理

使用GSE编号来下载GEO数据

使用GSE编号自动批量下载GEO数据

使用GSE编号加文件名批量下载

下载GEO的series\_matrix表达矩阵

### TCGA数据下载

TCGA转录组RNAseq数据下载

TCGA临床数据下载

TCGA突变数据下载

TCGA拷贝数变异数据下载

TCGA蛋白芯片RPPA数据下载

TCGA的mirna测序数据下载

TCGA的甲基化芯片数据下载

### TCGA数据提取

从TCGA的转录组rda格式数据中提取出counts,fpkm,tpm数据

从TCGA的表达矩阵中提取出分组信息

批量提取出单基因的泛癌组学数据

提取TCGA泛癌mrna转录组的tpm数据

提取TCGA单癌肿的mrna转录组的tpm数据

提取TCGA泛癌mrna转录组配对样本的tpm数据

提取TCGA单癌肿的mrna转录组配对样本的tpm数据

提取TCGA泛癌的临床病理信息

提取TCGA单癌肿的临床病理信息

提取TCGA泛癌的TMB数据

提取TCGA泛癌的MSI数据

提取TCGA泛癌的免疫细胞比例数据

提取TCGA泛癌的免疫评分数据

TCGA泛癌基因差异可视化

TCGA任意组学的单基因泛癌箱式图

单基因肿瘤组织中的转录水平泛癌箱式图

单基因肿瘤和正常组织中的转录水平泛癌箱式图

单基因配对样本的转录水平泛癌箱式图

单基因单癌肿配对样本的转录水平箱式图

单基因单癌肿年龄分组的转录水平箱式图

单基因单癌肿性别分组的转录水平箱式图

单基因单癌肿临床分期的转录水平箱式图

单癌肿按单基因高低分组的转录组差异热图

### TCGA基因集泛癌差异分析

单癌症按单基因高低分组差异分析后的GSEA富集分析

单癌症按单基因高低分组差异后的GSEA-KEGG富集分析

基因集在肿瘤组织中的转录水平泛癌箱式图

基因集肿瘤和正常组织中的转录水平泛癌箱式图

基因集配对样本的转录水平泛癌箱式图

基因集单癌肿配对样本的转录水平箱式图

基因集单癌肿年龄分组的转录水平箱式图

基因集单癌肿性别分组的转录水平箱式图

基因集单癌肿临床分期的转录水平箱式图

### TCGA泛癌基因相关性分析

单基因跟TMB的泛癌相关性

单基因跟MSI的泛癌相关性

单基因跟免疫检查点的泛癌相关性

单基因跟趋化因子的泛癌相关性

单基因跟免疫抑制因子的泛癌相关性

单基因跟免疫刺激因子的泛癌相关性

单基因跟免疫细胞的泛癌相关性

单基因跟免疫评分的泛癌相关性heatmap

单基因跟免疫评分的泛癌相关性triangle

基因-基因间相关性散点图

跟目的基因共表达的基因筛选和通路富集

### TCGA泛癌基因集相关性分析

基因集跟TMB的泛癌相关性

基因集跟MSI的泛癌相关性

基因集跟免疫检查点的泛癌相关性

基因集跟趋化因子的泛癌相关性

基因集跟免疫抑制因子的泛癌相关性

基因集跟免疫刺激因子的泛癌相关性

基因集跟免疫细胞的泛癌相关性

基因集跟免疫评分的泛癌相关性heatmap

基因集跟免疫评分的泛癌相关性triangle

### TCGA的预后生存分析

单基因转录水平高低分组的生存曲线

单基因转录水平的泛癌cox回归森林图

基因集转录水平的泛癌cox回归森林图

单基因来区分癌症的肿瘤和正常组织的ROC曲线