

# 课程目录

## 第一节课：AI + 公共数据库发文范式解读：从“相关性堆结果”到“MR×单细胞的因果机制证据链”

### 基于 Deepseek 高效精读“公共数据库挖掘型论文”

用 Deepseek 快速拆解这类文章最核心的发文逻辑：为什么只靠公开数据也能写出“机制感”、怎样把多队列/多数据库拼成一条可信的证据链、创新点通常落在哪里（因果、机制落点、外部验证、可解释性）。

### Deepseek 拆解论文叙事框架：一条可复用的“公共数据因果链”

从“提出临床/生物学问题”开始 → 公共转录组多队列发现与稳健验证 → 富集/网络锁定候选 → eQTL – GWAS MR 建立方向性因果（不是相关）→ 敏感性分析自证可靠 → 单细胞定位到细胞类型/状态 → 轨迹/通讯补齐“谁在驱动、如何互动”的机制解释 → 形成可投稿的主图叙事闭环。

### ChatGPT：把这套范式迁移到你的疾病方向（选题到主图规划）

现场示范如何用 ChatGPT 做“可发文选题设计”：

- 1) 如何选合适的公共数据组合（表达队列/结局 GWAS/eQTL/scRNA）
- 2) 如何定义暴露与结局、提前规划 MR 的可行性（IV 数量、组织匹配、方向）
- 3) 如何把“MR 结果”落到单细胞的验证任务（要看哪些细胞、哪些状态、哪些互动）

## 第二节课：公共数据处理和差异分析

- 1.从 GEO 数据库下载芯片数据，整理为训练集
- 2.下载 RNAseq 数据进行临床分组信息，整理为验证集表达矩阵
- 3.数据预处理：探针注释、重复基因合并、缺失处理
- 4.每个样本 QC、数据标准化和批次效应处理
- 5.limma 包进行差异分析：得到表达差异 DEGs，作为候选基因

## 第三节课：功能富集与蛋白互作网络：PPI/关键节点筛选（“把故事讲成机制”）

- 1.多队列一致性验证：方向一致、显著性一致、可视化一致
- 2.GO、KEGG 和 Reactome 富集：显著通路筛选与可视化
- 3.GSEA 通路富集分析：从“基因列表”到“通路级解释”的转换

4.STRING 构建 PPI 网络：导出网络文件 + Cytoscape 可视化

5.网络关键节点 (hub) 与模块 (cluster) 筛选：构建“机制候基因”

形成“基因—通路—网络”三层证据材料

**第四节课：eQTLGen 数据库构建孟德尔随机化的工具变量 (IVs)** 1.从

eQTLGen 数据库下载 Significant cis-eQTLs2.阈值筛选与暴露显著相关的

SNP3.进行连锁不平衡去相关，确保保留的 SNP 彼此独立 4.过滤弱工具变

量，满足与暴露因子显著相关且与结局变量相互独立

**第五节课：基于 eQTL 暴露与疾病 GWAS 多基因孟德尔随机化分析** 1.

将暴露数据与疾病结局数据统一转换为 outcome/exposure 标准格式 2.

用 SNP 取暴露与结局的交集，生成 SNP 的结局数据表 3.每个基因对应的

暴露子集做等位基因方向一致化 4.MR 分析并做异质性、多效性、单 SNP、

留一法等敏感性检验**第六节课：筛选有意义的暴露因素和基因单独做孟德**

**尔分析** 1.读取 MR 的全量结果，只保留 IVW 方法 p 值显著的结果 2.显著

暴露再做 OR 方向一致性筛选，确保同一暴露在不同方法下效应方向一致 3.

多效性 (pleiotropy) 与 异质性 (heterogeneity) 结果按照 p 值进行筛

选 4.用最终 IVW 筛出的暴露 (基因) 列表反向过滤 eQTL 数据，得到“高

可信 IV 集合”

5.筛选出有意义的基因后，单独对每个基因做孟德尔随机化分析

**第七节课：诊断与预测能力验证：ROC/AUC + 多指标对比**

1.单基因 ROC 与多基因联合 ROC：训练集 + 验证集 AUC 比较

2.交叉验证/重复抽样：避免偶然高 AUC 的伪结果

3.校准曲线与分层分析：不同亚组/不同批次下稳定性评估

4.Nomogram 构建与可解释输出：让临床读者看得懂、用得上

5.DCA 决策曲线：补齐临床净获益证据（文章结构更完整）

### **第八节课：机制挖掘一：单基因/核心基因的通路机制（GSEA / GSVA）**

1.按核心基因表达分组构建比较框架

2.单基因 GSEA (KEGG/Hallmark)：输出标准化 NES、p.adjust、核心富集基因

3.GSVA 打分：把通路活性转成样本级矩阵，便于画图与统计

4.与临床分组/表型关联：把通路变化写成 机制解释段落

5.输出 核心基因→关键通路 的主图级图组（GSEA 曲线 + 通路热图等）

### **第九节课：机制挖掘二：免疫浸润与炎症环境（ssGSEA/多算法交叉验证）**

1.免疫细胞基因集准备与 ssGSEA 打分 (GSVA)

2.免疫细胞差异比较：疾病组 vs Control 分层比较

3.核心基因与免疫细胞相关：Spearman 相关 + 热图/散点回归

4.多算法交叉验证：xCell/EPIC/CIBERSORT 思路与结果一致性检查

5.输出 基因—免疫—通路 三联证据图，补齐审稿人常问的免疫解释

### **第十节课：调控与关联：ceRNA 网络 + 疾病/药物关联（公共数据库可完成）**

1. miRNA 靶向预测（如 miRTarBase/TargetScan 思路）得到 miRNA—mRNA 对

2. lncRNA—miRNA 关联（如 starBase 思路）补齐上游调控
- 3.构建 lncRNA—miRNA—mRNA (ceRNA) 网络：导出边表/节点表并可视化
- 4.疾病关联/证据补齐（如 CTD 思路）：核心基因与疾病关联排名展示
- 5.输出 调控网络图 + 疾病关联图 ，让文章从 结果 变成 更像机制论文

### **第十一节课：单细胞验证一：把 MR 基因落到具体的细胞类型上**

- 1.下载单细胞数据与注释信息：建立可复用对象（Seurat 对象）
- 2.QC、归一化、高变基因、降维聚类：标准化流程与关键参数
- 3.细胞类型注释：marker 规则 + 自动注释（SingleR 和 ScType 思路）+ 人工校正
- 4.核心基因在细胞类型中的表达：FeaturePlot/VlnPlot/DotPlot
- 5.输出 关键基因在哪些细胞表达、差异在哪些细胞最明显 的主图证据

### **第十二节课：单细胞验证二：轨迹推断 + 细胞通讯（把机制写成 谁驱动、如何互作）**

- 1.选定关键细胞亚群：基于标记基因与表型差异进行聚焦
- 2.轨迹分析（Monocle2 和 3/scVelo）：状态转变与关键节点解释
- 3.核心基因沿轨迹变化：把因果基因接到状态变化上
- 4.细胞通讯分析（CellChat）：差异互作通路 with 配体/受体细胞识别
- 5.文章化交付：把 MR + 单细胞结果整理为 可投稿的主图叙事顺序 + 图注要点 + 结果段落模板