

有助于精准识别个体化的疾病进展模式，为精准医疗提供数据支撑。

2. 拓展疾病早期预测与干预的新策略

传统的疾病研究主要依赖静态基因表达分析，难以捕捉疾病发展的关键转折点。本研究基于单细胞组学数据，结合因果推断方法，能够动态监测疾病状态的变化，并识别疾病恶化的临界点，为早期预测和个体化干预提供新思路。

3. 推动生物网络建模和因果推断技术的发展

本研究结合因果推断、动态网络分析及多组学数据融合，构建个体特异性分子因果网络，不仅拓展了网络生物学在复杂疾病研究中的应用，也为系统生物学、网络医学等领域提供新的方法论支持。

4. 发现关键动态生物标志物，优化个体化治疗方案

通过挖掘疾病恶化过程中的关键基因和分子网络，本研究能够筛选出具有潜在临床价值的动态生物标志物，为靶向治疗、药物开发及个性化诊疗方案设计提供科学依据，推动疾病精准防治的发展。

三 研究内容与方法

本课题围绕“复杂疾病个体特异性因果网络构建及临界点预测”展开，结合单细胞转录组学 (scRNA-seq) 数据，蛋白质-蛋白质相互作用 (PPI) 网络、统计因果推断方法和动态网络分析，探索个体化疾病进展的分子机制，并识别疾病恶化的临界点。本研究预采用一个模拟数据集与五个真实数据集：利用微分方程生成基因调控网络，主要评估统计方法的有效性与稳定性；以流感感染单细胞数据 (GEO: GSE30550)，结直肠癌单细胞数据 (GEO: GSE161277)，CPTAC 数据库中胶质母细胞瘤单细胞数据作为癌症与神经系统疾病的建模数据集；以 TCGA 数据库中结直肠癌 (COAD/READ) 和胶质母细胞瘤 (GBM) 转录组数据作为模型验证数据集。具体研究方法与技术路线如下：

step 1. 获取上述单细胞数据集，数据来源于 GEO 数据库 (<https://www.ncbi.nlm.nih.gov/geo/>) 与 TCGA 数据库 (<https://portal.gdc.cancer.gov/>)；

step 2. 数据预处理

- 质量控制：去除低质量细胞（如基因检测数过低或者线粒体基因表达比例过高的细胞）；
- 归一化与标准化：采用 log-normalization 或 SCTransform 进行数据标准化处理；
- 降维分析：采用 PCA（主成分分析）或者 UMAP/t-SNE（非线性降维）进行数据