

欧易生物空间转录组基础分析方法

I 生物信息分析流程

1 测序数据质量控制及基因定量

建库测序及数据分析部分由上海欧易生物医学科技有限公司完成。高通量测序中产生的原始数据 (raw reads) 为 fastq 格式序列。采用 10x genomics 官方软件 Space Ranger(version 2.0.1) 对 Visium 空间转录组测序数据和明场显微镜切片图像进行处理, 检测组织在芯片中的捕获区域, 与参考基因组(人: GRCh38, 小鼠: mm10)进行比对分析, 并根据 Spatial barcode 信息, 将每个 spot 的 reads 区分开来, 并对总 spot 数、每个 spot 中的 reads 数、检测到的基因数以及 UMIs 数目进行统计, 从而对样品的质量进行评估。

2 基因定量质控及数据预处理

在 Space Ranger 初步质控结果的基础上, 使用 Seurat^[1](version 4.3.0)软件包对数据进行进一步质控及处理。利用 SCTransform^[2] 函数对数据进行归一化处理, 检测高方差特征, 并将数据存储在 SCT 矩阵中。

3 降维与聚类分析

使用 Seurat 包中的 FindVariableGenes 函数筛选 Top 3000 个高变基因 (HVGs, highly variable genes), 利用高变基因的表达谱进行 PCA (主成分) 降维分析, 通过 UMAP(非线性降维)将结果在二维空间上进行可视化。

批注 [A1]: 如果样本间存在批次, 则将 PCA 降维分析替换为 harmony (version 0.1.0)包中的 RunHarmony 分析, 以此矫正批次效应。

Korsunsky, I., Millard, N., Fan, J. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 16, 1289–1296 (2019)

4 空间特征基因鉴定

使用 Seurat 包中的 FindAllMarkers 函数(test.use = bimod)进行 marker 基因鉴定, 即找到每种 Spot 群相对于其他 Spot 群差异上调表达的基因, 这些基因就是每种 Spot 群潜在的 marker 基因。通过 VlnPlot 和 FeaturePlot 函数对鉴定得到的 Marker 基因进行可视化。

5 空间细胞类型注释

RCTD^[3] (version 1.1.0)是一种稳健的细胞类型分解方法, 能够利用从单细胞 RNA-seq 中获悉的细胞类型概况来分解细胞类型混合物, 同时校正跨测序技术的差异。在运行 RCTD 时, creat.RCTD 函数使用默认参数, 但每种细胞类型至少有 1 个细胞且每个 spot 至少有 1 个 UMI (唯一分子标识符), run.RCTD 函数 doublet_mode 设置为 FALSE, 从而推断每个 Spot 位点中细胞类型的组成情况。

批注 [A2]: 根据实际情况选择是否放入文章里

6 差异基因及富集分析

使用 Seurat 包中的 FindMarkers 函数进行差异基因筛选，根据 pvalue 小于 0.05 以及差异倍数大于 1.5 倍的条件筛选出差异显著基因，并通过超几何分布检验进行差异显著基因的 GO 和 KEGG 富集分析。

批注 [A3]: 或 1.2 倍，根据实际结果修改。

如果您的研究课题使用了欧易的测序和分析服务，我们期望您在文章发表时，在文章方法部分或致谢部分引用或提及欧易生物：The sequencing and bioinformatics analysis were provided by OE Biotech Co., Ltd. (Shanghai, China).

II 参考文献

- [1] Stuart, Tim, et al. "Comprehensive integration of single-cell data." Cell 177.7 (2019): 1888-1902.
- [2] Hafemeister, Christoph, and Rahul Satija. "Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression." Genome biology 20.1 (2019): 296.
- [3] Cable, Dylan M., et al. "Robust decomposition of cell type mixtures in spatial transcriptomics." Nature biotechnology 40.4 (2022): 517-526.