**生物医药合作项目开发**

**研究方向： 小鼠结肠炎的单细胞测序数据 ;**

**委托人： 公司内部 ;**

**受托人： 杭州铂赛生物科技有限公司 .**

# 1 研究背景

## 1.1 思路

# 2 可行性

# 3 创新性

# 4 GEO 检索方法

## 4.1 数据分析平台

在 Linux pop-os x86\_64 (6.9.3-76060903-generic) 上，使用 R version 4.4.2 (2024-10-31) (<https://www.r-project.org/>) 对数据统计分析与整合分析。

## 4.2 GSE 数据搜索 (Dataset: COLITIS\_SINGLE\_CELL)

使用 Entrez Direct (EDirect) <https://www.ncbi.nlm.nih.gov/books/NBK3837/> 搜索 GEO 数据库 (esearch -db gds)，查询信息为: ((colitis[Description] AND (single cell[Description]) AND ((3:20[Number of Samples]) AND (GSE[Entry Type]) AND (Mus musculus[Organism]))。 排除 summary 中包含 ‘knock|deficien|SuperSeries|transgenic|regulat|CD[0-9]+|cancer|treatment|single.nuclei|interferons|microbiota’ 的数据例。 以 GEOquery 获取 GSE 数据集 (n=11)。

## 4.3 GSE 数据搜索 (Dataset: COLITIS)

使用 Entrez Direct (EDirect) <https://www.ncbi.nlm.nih.gov/books/NBK3837/> 搜索 GEO 数据库 (esearch -db gds)，查询信息为: ((colitis[Description]) AND ((6:1000[Number of Samples]) AND (GSE[Entry Type]) AND (Mus musculus[Organism]))。 以正则匹配，滤除包含 ‘single cell’ 或 ‘scRNA’ 的数据例。仅获取类型包含 ‘Expression profiling by high throughput sequencing’ 或 ‘Expression profiling by array’ 的数据例。排除 summary 中包含 ‘knock|deficien|SuperSeries|transgenic|regulat|CD[0-9]+|cancer|treatment|single.nuclei|single.cell|interferons|microbiota|treated|Inhibit’ 的数据例。 以 GEOquery 获取 GSE 数据集 (n=31)。 从元数据中匹配并排除包含关键词的数据：‘knock|deficien|SuperSeries|transgenic|regulat|CD[0-9]+|cancer|treatment|single.nuclei|single.cell|interferons|microbiota|treated|Inhibit’，共得到 8 个数据集。

# 5 参考文献和数据集

## 5.1 GSE 数据搜索 (COLITIS\_SINGLE\_CELL)

以 Entrez Direct (EDirect) 搜索 GEO 数据库 (检索条件见方法章节) 。匹配 gdsType 中包含“high throughput sequencing”的描述，未过滤掉任何数据 (n = 11)。这些数据为：GSE249582, GSE234592, GSE168806, GSE194411, GSE210415, GSE207178, GSE201723, GSE168033, GSE178749, GSE114374, GSE117216等。 可用数据，及其组别为：

* **GSE249582**, **Type**: RNA-seq
  + control (n = 3)
  + DSS\_induced\_colitis (n = 3)
* **GSE234592**, **Type**: (scRNA-seq) RNA-seq
  + Unknown (n = 2)
* **GSE168806**, **Type**: RNA-seq
  + ADT\_allsamples (n = 1)
  + IL (n = 1)
  + Naïve\_gene\_expression (n = 1)
  + PBS\_gene\_expression (n = 1)
  + Untreated\_gene\_expression (n = 1)
* **GSE194411**, **Type**: (scRNA-seq) RNA-seq
  + Propyzamide\_induced\_colitis (n = 1)
  + Propyzamide\_Naive (n = 1)
  + Vehicle\_induced\_colitis (n = 1)
  + Vehicle\_Naive (n = 1)
* **GSE210415**, **Type**: (scRNA-seq) RNA-seq
  + control (n = 2)
  + Dss (n = 2)
* **GSE207178**, **Type**: (scRNA-seq) RNA-seq
  + control (n = 2)
  + diphtheria\_toxin (n = 2)
* **GSE201723**, **Type**: (scRNA-seq) RNA-seq
  + anti\_GFP\_24hr (n = 3)
  + anti\_GFP\_48hr (n = 3)
  + SZN1326p\_24hr (n = 3)
  + SZN1326p\_48hr (n = 3)
  + Uninjured\_24hr (n = 2)
  + Uninjured\_48hr (n = 2)
* **GSE168033**, **Type**: (scRNA-seq) RNA-seq
  + Normal (n = 1)
  + Recovered (n = 1)
  + Severe\_colitis (n = 1)
* **GSE178749**, **Type**: (scRNA-seq) RNA-seq
  + baseline (n = 1)
  + DSS (n = 2)
* **GSE114374**, **Type**: (scRNA-seq) RNA-seq
  + Healthy (n = 2)
  + Ulcerative\_Colitis (n = 2)
* **GSE117216**, **Type**: RNA-seq
  + Control\_18345 (n = 1)
  + Control\_18346 (n = 1)
  + Control\_19083 (n = 1)
  + DSS\_17954 (n = 1)
  + DSS\_19083 (n = 2)
  + DSS\_19424 (n = 1)

## 5.2 GSE 数据搜索 (COLITIS)

以 Entrez Direct (EDirect) 搜索 GEO 数据库 (检索条件见方法章节) 。 可用数据，及其组别为：

* **GSE280748**, **Type**: RNA-seq
  + ceAF (n = 5)
  + DSS (n = 5)
* **GSE239820**, **Type**: RNA-seq
  + cortex (n = 69)
  + distal\_colon (n = 66)
  + hippocampus (n = 65)
  + midbrain (n = 64)
  + proximal\_colon (n = 64)
  + striatum (n = 67)
* **GSE214600**, **Type**: RNA-seq
  + C (n = 199)
  + D (n = 204)
* **GSE210405**, **Type**: RNA-seq
  + day0\_\_replicate (n = 3)
  + day2\_\_replicate (n = 3)
  + day4\_\_replicate (n = 3)
  + day6\_\_replicate (n = 3)
* **GSE166277**, **Type**: RNA-seq
  + C5KO\_rep (n = 2)
  + C5WT\_rep (n = 2)
  + C9KO\_rep (n = 2)
  + C9WT\_rep (n = 2)
  + P5KO\_rep (n = 3)
  + P5WT\_rep (n = 3)
  + P9KO\_rep (n = 2)
  + P9WT\_rep (n = 2)
* **GSE42768**, **Type**: Microarray
* **GSE35609**, **Type**: Microarray
  + Healthy\_Control (n = 4)
  + TNBS\_colitis (n = 30)
* **GSE5864**, **Type**: Microarray
  + Colon\_Saline\_w (n = 1)
  + Colon\_saline\_w6 (n = 2)
  + Colon\_TNBS\_w (n = 4)