**生物医药合作项目开发**

**研究方向： SCG2 与神经母细胞瘤 ;**

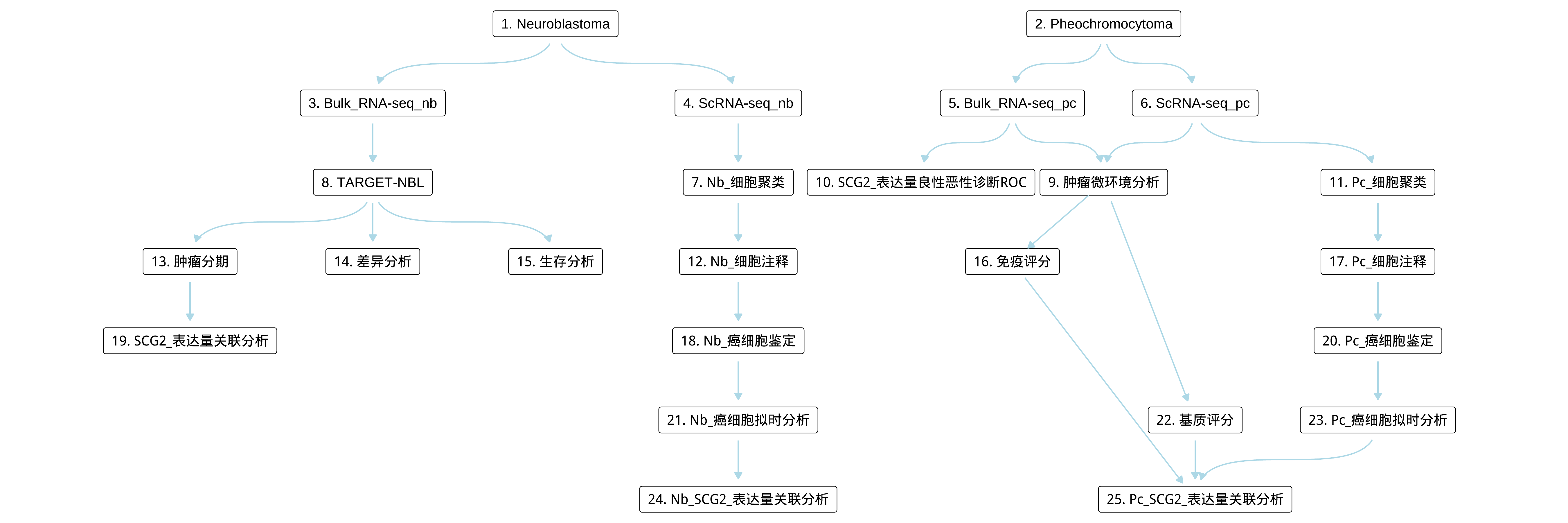
**委托人： 孙念政 ;**

**受托人： 杭州铂赛生物科技有限公司 .**

# 1 研究背景

## 1.1 思路

* SCG2表达量与神经母细胞瘤分期呈线性负相关，其上调表达可以诱导神经母细胞瘤细胞进一步分化。
* SCG2可以作为神经母细胞瘤危险度评估和预后预测的生物学标志。
* SCG2可以作为鉴别嗜铬细胞瘤良恶性的生物学标志。
* SCG2通过改变肿瘤微环境，从而诱导神经母细胞瘤细胞及嗜铬细胞瘤细胞进一步分化



**Fig.** Route

# 2 GEO 检索方法

## 2.1 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。

## 2.2 GSE 数据搜索 (Dataset: SCRNA\_PG)

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: ((Paraganglioma[Description] AND (single cell[Description]) AND ((3:20[Number of Samples]) AND (GSE[Entry Type]) AND (Homo sapiens[Organism]))，转化为数据表格。

## 2.3 GSE 数据搜索 (Dataset: SCRNA\_PC)

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: ((Pheochromocytoma[Description] AND (single cell[Description]) AND ((3:20[Number of Samples]) AND (GSE[Entry Type]) AND (Homo sapiens[Organism]))，转化为数据表格。

## 2.4 GSE 数据搜索 (Dataset: SCRNA\_NB)

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: ((Neuroblastoma[Description] AND (single cell[Description]) AND ((3:20[Number of Samples]) AND (GSE[Entry Type]) AND (Homo sapiens[Organism]))，转化为数据表格。 以正则匹配，仅保留 ‘summary’ 或 ‘title’ 中包含 ‘single cell’ 或 ‘scRNA’ 的数据例。仅查询临床数据，因此滤除匹配到关键词 in vitro, cell line, CD[0-9]+, vehicle, vector, DMSO, /ml, nm 的数据例。 (注：以上仅为查找合适的 GEO 数据所做的数据筛选，与实际分析无关) 。仅获取类型包含 ‘Expression profiling by high throughput sequencing’ 或 ‘Expression profiling by array’ 的数据例。此外，排除 summary 或 title 中匹配到字符集 EX (KO, WT, *WT*, *KO*, wildtype, mutant, knock, deficien, absen, SuperSeries, transgenic, CD[0-9]+) 的数据例。上述得到 19 个 GSE 数据集。以 R 抓取网页 (例如，RCurl::getURL 抓取 )，解析 ‘Overall design’ 和 ‘Samples’ 模块，匹配字符集 EX，排除匹配到的数据例。排除 ‘Overall design’ 中包含 in vitro, cell line, CD[0-9]+, vehicle, vector, DMSO, /ml, nm 的数据例。为了获得包含对照样本的数据集，(描述字符过少的例外) 以正则匹配 (‘Samples’, ‘Summary’, ‘Overall design’) 包含 normal, control, healthy, ctrl, adjacent, N[0-9]+, C[0-9]+ 的数据例。共 7 个。仅获取包含 ‘protein coding’ 测序的数据集，排除 ‘Samples’ 和 ‘Overall design’ 中包含 siRNA, miRNA, miR, lncRNA 字符的数据例。余下共 7 个。 以 GEOquery 获取 GSE 数据集 (n=7)。

# 3 参考文献和数据集

## 3.1 PCPG

### 3.1.1 PCPG scRNA-seq

* PMID: 38407266, <https://datadryad.org/dataset/doi:10.5061/dryad.rjdfn2zkg#citations>
* PMID: 40190189, <https://ngdc.cncb.ac.cn/search/specific?db=hra&q=HRA005846> (不可获取)
* …

### 3.1.2 GSE 数据搜索 (SCRNA\_PG)

以 Entrez Direct (EDirect) 搜索 GEO 数据库 (检索条件见方法章节) 。未发现合适数据集。

未检索到数据。

### 3.1.3 GSE 数据搜索 (SCRNA\_PC)

以 Entrez Direct (EDirect) 搜索 GEO 数据库 (检索条件见方法章节) 。未发现合适数据集。

未检索到数据。

## 3.2 NB

### 3.2.1 GSE 数据搜索 (SCRNA\_NB)

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：

* **GSE161975**, **Type**: (scRNA-seq) RNA-seq
  + ethanol (n = 2)
  + Nutlin\_3\_treated (n = 2)
* **GSE216155**, **Type**: (scRNA-seq) RNA-seq
  + ATRX\_deleted\_sample (n = 2)
  + Control\_sample (n = 5)
  + MYCN\_amplified\_sample (n = 5)
  + Sporadic\_sample (n = 5)
* **GSE218450**, **Type**: RNA-seq
  + left\_adrenal\_gland (n = 7)
  + left\_cervical\_primary\_tumor (n = 4)
  + lymph\_node\_left\_inguinal (n = 5)
* **GSE220607**, **Type**: (scRNA-seq) RNA-seq
  + marrow (n = 12)
* **GSE220651**, **Type**: (scRNA-seq) RNA-seq
  + AML\_single\_cell\_multiplexed\_feature (n = 2)
  + AML\_single\_cell\_multiplexed\_GEX (n = 2)
* **GSE261828**, **Type**: (scRNA-seq) RNA-seq; Other
  + Biopsy (n = 2)
  + Surgical\_specimen (n = 10)
* **GSE288099**, **Type**: RNA-seq
  + CTRL (n = 2)
  + NEN (n = 2)
  + RA (n = 2)
  + RA+NEN (n = 2)