

## BIOF3 组学数据分析

# Jupyter 与交互式分析环境

导出日期：2026年5月11日

## Jupyter 与交互式分析环境

组学分析的第一个问题是：在哪里写下可以重复运行的分析过程？

Jupyter Notebook 把代码、说明、图表和运行结果放在一起，适合探索性分析和教学演示。本章介绍如何使用本地 Jupyter、JupyterLab、Google Colab 和服务器 Notebook，并说明 Notebook 与普通脚本的分工。

### 学习目标

完成本章后，你应该能够：

- 理解 Notebook 在可重复分析中的作用
- 区分本地 Jupyter、JupyterLab、Google Colab 和服务器环境
- 掌握 Notebook 的基本结构
- 知道什么时候该把 Notebook 稳定下来改为脚本

### Notebook 适合做什么

Jupyter Notebook 是一种把代码、说明文字、图表和运行结果放在一起的交互式文档。它适合探索性分析和教学演示。

适合：

- 记录分析思路
- 逐步运行代码
- 快速查看表格和图
- 写下参数解释
- 分享小型分析示例

不适合：

- 长时间无人值守的大流程
- 大规模批处理任务
- 保存敏感数据输出
- 没有版本控制的正式生产流程

推荐做法：Notebook 用来探索和记录，稳定后把关键步骤整理成 `.R`、`.py` 或工作流脚本。

### 常见交互式环境

#### Jupyter Notebook

经典 Notebook 界面，适合初学者。文件通常是 `.ipynb`，内部保存代码单元、Markdown 文本、输出结果和元数据。

启动方式：

```
jupyter notebook
```

## JupyterLab

JupyterLab 是更完整的工作界面，支持 Notebook、终端、文本编辑器和文件浏览器。做真实项目时更推荐 JupyterLab。

启动方式：



## Google Colab

Google Colab 是云端 Notebook 环境，打开浏览器即可运行 Python。它适合教学和轻量实验，不适合长期保存重要环境或处理隐私数据。

适合：

- 快速试代码
- 课堂演示
- 共享小型 Notebook
- 临时使用 GPU

注意：

- 运行环境可能变化
- 会话可能断开
- 文件需要显式保存
- 不要上传未脱敏的人类样本数据

## 服务器 Notebook

在真实组学项目中，数据常放在服务器上。可以在服务器启动 JupyterLab，再通过浏览器访问。

一般流程：



如果要从本地浏览器访问远程 Notebook，通常会用 SSH 端口转发：

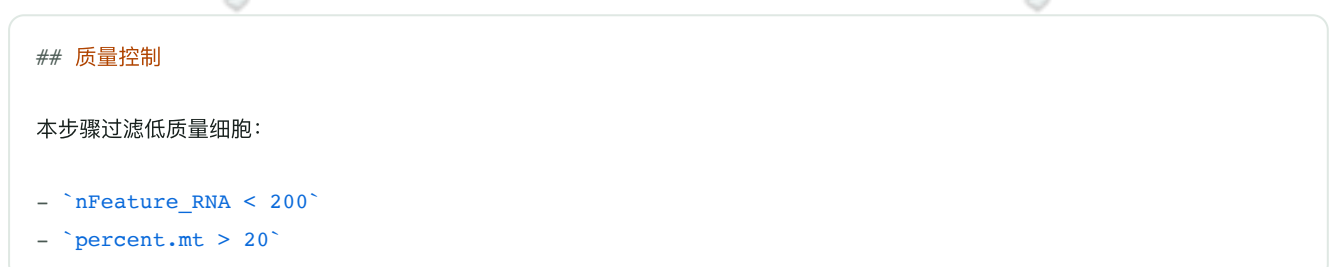


服务器环境要特别注意权限、数据位置和端口安全。

## Notebook 的基本结构

### Markdown 单元

用来写说明、记录参数和解释结果。



## Code 单元

用来运行代码。

```
import pandas as pd

metadata = pd.read_csv("data/metadata.csv")
metadata.head()
```

## 输出结果

输出可以是表格、图像、日志或错误信息。正式分析时，不建议只依赖 Notebook 里的输出；重要结果应该保存到 results/。

```
metadata.to_csv("results/metadata_checked.csv", index=False)
```

## 记录运行环境

一个能重跑的 Notebook 至少要把运行环境记下来。

Python 中可以记录版本：

```
import sys
import pandas as pd

print(sys.version)
print(pd.__version__)
```

R 中可以记录环境：

```
sessionInfo()
```

这些信息对后续复现结果、排查问题都有用。

## 从 Notebook 走向脚本

Notebook 的价值不是"把所有分析塞在一个文件里"，而是帮你逐步检查数据结构、记录思路。稳定下来的步骤应该整理到 .R 或 .py 脚本，便于批量运行、版本控制和团队协作。

结构建议：

- notebooks/ 放探索性分析和教学 Notebook
- scripts/ 放稳定的可复现脚本
- results/ 放最终图表、表格和报告

## 下一步

继续学习：

- [公共数据库与数据检索](#)
- [R 数据整理与 ggplot2 可视化](#)
- [单细胞实践 01](#)

## 参考资源

- Jupyter 官方文档: <https://docs.jupyter.org/>
- JupyterLab 文档: <https://jupyterlab.readthedocs.io/>
- Google Colab: <https://colab.research.google.com/>



扫码关注微信公众号【生信F3】

获取文章完整内容，分享生物信息学最新知识。

