为scATAC项目制定工作盘、清理和归档计划

完整操作教程 · 参考固定 · 往返恢复验收

用ArchR专用保存、重新载入及依赖清单完成归档验收，再按任务回收工作盘。

运行边界：示例代码需按安装版本和真实项目输入复核；未提供真实生物数据，语法或小样本检查不代表科学结果和性能验收。

scATAC项目的保存计划应覆盖输入片段、Arrow、项目结构、QC和下游结果。单独保存一个R对象文件，不能自动带走全部Arrow数据；反过来只留Arrow，也可能丢失项目状态与分析记录。下面用专用保存和重新载入完成一次可验证归档，再决定回收哪些工作文件。

按数据生命周期划分位置

内容	日常用途	回收前的条件
原始片段/BAM及索引	追溯和重新构建	有来源、校验与明确保留政策
Arrow与项目结构	后续分析和项目恢复	完整归档能重新载入
临时文件与失败半成品	当前构建/故障定位	写入结束且已确认可丢弃
QC、代码与参数	解释筛选和复现	与对应输入和项目一起保留
导出矩阵与图	交付和展示	不把它们当完整项目替身

先记录每类当前占用、保留周期和真实挂载点。容量按同时存在计算：复制归档时旧项目和新项目可能共存，下一批构建也可能已经启动。不能只按最终图表或压缩片段大小选择工作盘；独立备份的容量和恢复地点还要另列。

用专用接口创建新的归档目录

保存前让相关Arrow写入结束，并确认项目状态一致。以下需要兼容ArchR，源目录应是已经saveArchRProject保存过的项目。两个参数为源项目和新的归档目录；overwrite与dropCells显式关闭，避免覆盖已有归档或在归档中顺便改变细胞集合。

```
library(ArchR)
args <- commandArgs(trailingOnly=TRUE)
stopifnot(length(args)==2L, !dir.exists(args[2]), !file.exists(args[2]))
proj <- loadArchRProject(path=args[1], force=FALSE, showLogo=FALSE)
arrows <- getArrowFiles(proj)
stopifnot(length(arrows)>0L, all(file.exists(arrows)))
baseline <- list(cells=getCellNames(proj), arrow_count=length(arrows),
                 version=sessionInfo())
saveArchRProject(ArchRProj=proj, outputDirectory=args[2],
                 overwrite=FALSE, load=FALSE, dropCells=FALSE)
saveRDS(baseline, file.path(args[2], "archive-baseline.rds"))
```

保存会组织项目输出和Arrow，仍需足够空间与权限；函数返回不代表远端备份已完成。load=FALSE避免立即载入另一份项目，但不等于零内存或零I/O。源项目、归档和输入之间的关联要写入清单，部分写入失败的目录不能标记为归档通过。

在新R会话中执行恢复验收

将下例中的归档路径换成实际位置。loadArchRProject会按保存结构重新规范路径，因此用它恢复，不只readRDS一个内部文件。先验收细胞标识与Arrow可访问，再根据项目需要抽查矩阵和已有降维、峰集等内容。

```
library(ArchR)
archive <- "/absolute/path/to/new-archive"
baseline <- readRDS(file.path(archive,"archive-baseline.rds"))
check <- loadArchRProject(path=archive, force=FALSE, showLogo=FALSE)
```

```
files <- getArrowFiles(check)
stopifnot(length(files)==baseline$arrow_count, all(file.exists(files)),
          identical(getCellNames(check), baseline$cells))
print(check)
writeLines(capture.output(sessionInfo()), file.path(archive,"restore-session.txt"))
```

如果归档迁到另一台机器，还要在目标账号验证全部文件权限和实际读取，而非只在原机器打开成功。force=TRUE可忽略某些可选组成缺失，不应用来掩盖不完整交接。校验值证明字节一致，不能替代项目结构、细胞与所需分析结果的验收。

清理写成明确的任务清单

清理记录应列绝对路径、内容类别、是否仍被项目引用、对应成功归档和处理人。仅空目录可用rmdir作保守回收；本篇不提供清空共享目录的递归命令。Arrow仍在getArrowFiles返回清单中，就不能仅因很大而视为临时文件。

下一批构建按剩余真实空间和已测新增峰值排期，防止待归档项目积压。若需要通过dropCells缩减Arrow，那属于另一次数据修改，应独立命名并重新验收，不能与原样归档混淆。只有恢复测试和保留安排完成，才据记录回收旧工作副本。

配套练习包

本PDF对应的ZIP包含本文代码块、主题练习模板或合成数据、README和SHA256校验文件。请从同一教程页面下载练习包；先核对校验值，再在独立目录复现。

参考资料

ArchR saveArchRProject · 来源资料

ArchR loadArchRProject · 来源资料

ArchR getArrowFiles · 来源资料

操作建议：先在合成小样本上验证流程，再依据项目参考、标签合同、存储策略与获配资源验收真实数据。