检查合并前的数据层、对象冗余和处理顺序

完整操作教程 · 参考固定 · 往返恢复验收

从assay与layer盘点开始，在明确重新分析的前提下建立counts-only副本，验证样本、细胞数和保留范围后再决定是否扩容。

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。

多个Seurat对象可以合并，但“合成一个对象”不代表数据层、归一化结果、降维和图都会按原样保留。动手前先说明用途：只是汇总原始counts、继续整合，还是直接下游比较；再核对对象的assay、layer、细胞名和参考版本。

合并前列出必须保留的内容

检查项	核对方式	常见风险
软件版本	记录Seurat与SeuratObject版本	不同版本的数据层行为不同
assay/layer	列出RNA、SCT等及counts/data	把不同分析阶段的层混在一起
细胞标识	检查样本来源和条码重复	不同样本同名细胞被误认为同一细胞
元数据	检查样本、批次、条件列	合并后无法追溯来源
降维和图	明确是否需要重新计算	旧图或整合结果并非通用合并产物

给样本加来源标识后合并

以下示例假设每个对象都包含RNA counts和sample列。先在副本上操作，避免覆盖原始RDS。Seurat的merge可为细胞名添加样本前缀；合并后按预期比较输入/输出细胞数、重复条码和样本分布。

```
library(Seurat)
objs <- list(S01 = obj1, S02 = obj2)
stopifnot(all(vapply(objs, inherits, logical(1), "Seurat")))
stopifnot(all(vapply(objs, function(x) "RNA" %in% Assays(x), logical(1))))
merged <- merge(
  x = objs[[1]], y = objs[-1],
  add.cell.ids = names(objs), merge.data = FALSE
)
stopifnot(ncol(merged) == sum(vapply(objs, ncol, integer(1))))
table(merged$orig.ident, useNA = "ifany")
Layers(merged[["RNA"]])
```

如果输入对象的处理步骤或归一化方法不一致，合并原始counts并重新按计划处理通常更容易解释；若确实要保留data层，应先确认各对象使用相同方法和参数。merge不会替代批次整合，也不会保留所有图、命令历史和feature级元数据。

何时需要 JoinLayers

Seurat v5合并后可能保留按样本拆分的counts/data层，这对后续IntegrateLayers工作流可能是有意设计。若后续函数要求单一层，且研究流程确实不再需要分层，才在副本上执行JoinLayers并重新检查layer名称、细胞数和下游输入。不要为了让某个函数“能跑”就盲目连接，也不要把counts-only对象当成已完成归一化的对象。

验收并保留可恢复版本

保存新文件前核对合并前后细胞总数、每个样本的细胞数、细胞名唯一性、关键metadata和counts维度；抽查若干细胞的原始计数是否来自预期样本。将原始对象保留为只读，记录脚本、版本和是否JoinLayers。只有确认结果满足后续分析目标，才把新对象作为项目输入。

用样本表避免合并错配

合并前准备一张样本表，每个对象一行，至少包含对象路径、样本名、实验条件、批次、建库类型、参考版本和预期细胞数。读入后将样本名写入metadata，再检查每个对象的细胞名是否唯一。对于不同物种、不同参考或不同计数流程得到的对象，先确认feature ID和基因注释可以比较；不能因为对象结构相似就直接合并。示例中的对象列表名称要与样本表一致，真实项目可用循环读取并按清单逐一核对，避免文件顺序变化造成样本标签错位。

出现细胞数不一致时先暂停保存，分别比较每个输入对象的ncol、细胞条码和metadata行名；不要用强制改名或丢弃行来掩盖问题。合并后抽取每个样本若干细胞，核对RNA counts是否非空且来源标记正确。保存新的RDS后重新载入一次，确认对象仍可用，再把脚本和sessionInfo一并归档。

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

SeuratObject merge.Seurat · 来源资料

Seurat Essential Commands: merge and JoinLayers · 来源资料

Seurat DietSeurat · 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。