按比较数量与文件规模安排并发和工作盘

将rMATS的prep与post拆开管理，核对BAM路径一致性和中间文件集合，再按阶段资源控制批量并发。

rMATS 分阶段运行后，`--task inte` 报错或 post 找不到 BAM，常见原因不是算力不够，而是 prep 与 post 对输入路径的记录对不上，或者临时目录里混入旧结果。先把问题拆为三段：BAM 是否可读、每个 BAM 是否完成 prep、post 使用的分组清单能否与 prep 记录一一对应。这样比直接重复整个分析更容易定位。

理解 prep、inte 和 post 的职责

任务	作用	应检查
prep	逐个处理 BAM 并在临时目录写入 .rmats	日志、退出码、输入路径和文件数
inte	检查 b1/b2 中的 BAM 是否与临时记录对应	完整路径和每个样本是否一对一匹配
post	读取 prep 结果、检测事件并进行统计	临时文件来源、GTF 和分析参数一致

官方说明特别要求 prep 与 post 中 BAM 的完整路径保持一致；prep 使用 `/data/a.bam`，post 不能改成 `a.bam` 或另一个挂载路径。并行 prep 的 `.rmats` 文件还可能同名，汇总时需用官方 `cp_with_prefix.py` 添加不同前缀，防止覆盖或漏样本。

按原路径做完整性检查

下面命令假定 b1、b2 清单中的路径与 prep 时完全相同，TMP_POST 是一个只含本次运行 `.rmats` 文件的新目录。先运行 integrity 检查；检查未通过时不要继续 post。清单中的重复项、漏项和路径变更都应先修正。

```
set -euo pipefail
: "${RMATS_PY:?set rmats.py path}"
: "${B1:?set group one BAM list}"
: "${B2:?set group two BAM list}"
: "${GTF:?set the same annotation GTF}"
: "${TMP_POST:?set clean merged temp directory}"
: "${OUT_POST:?set new post output directory}"
: "${READ_LENGTH:?set read length}"
: "${THREADS:?set allocated threads}"
test -r "$B1" && test -r "$B2" && test -r "$GTF"
test -d "$TMP_POST" && test ! -e "$OUT_POST"
mkdir -p "$OUT_POST"
common=(--b1 "$B1" --b2 "$B2" --gtf "$GTF" -t paired
  --readLength "$READ_LENGTH" --nthread "$THREADS"
  --od "$OUT_POST" --tmp "$TMP_POST")
python "$RMATS_PY" "${common[@]}" --task inte
python "$RMATS_PY" "${common[@]}" --task post
```

按报错处理，避免混用旧结果

如果 integrity 提示缺少匹配项，逐条比较清单中的完整路径、文件名和 `.rmats` 来源；如果 post 报临时数据缺失，确认每个 BAM 的 prep 已成功且所有批次文件都已复制。若输入、GTF 或关键参数改过，应为本次分析重新建立 prep 和 post 目录，不能把新旧结果混合。对 BAM 可先运行 quickcheck，但它只检查文件头和结尾，不会扫描文件中间所有数据；文件仍需靠完整读取或下游运行发现内部损坏。

结束后归档版本、命令、b1/b2 清单、GTF 校验值、日志和输出文件列表。统计结果还需结合实验设计、重复结构、事件数及原始 reads 质控解释。一次完整性检查通过只能说明输入对应关系正确，不代表生物学结论已经成立。

常见提示怎样逐项排查

若 integrity 提示某个文件不匹配，先打印 b1、b2 中每个路径，与 prep 日志和 `.rmats` 文件记录逐条核对，特别留意软链接解析后路径、大小写和共享存储挂载点。若 post 未找到准备结果，检查临时目录是否指向汇总后的目录，文件是否在复制时被覆盖，是否漏掉某批 prep。若任务能完成但结果数量异常，再检查输入 BAM 记录、GTF 染色体命名、读长和文库方向；先确认这些前提，避免把输入问题误认为是计算资源不足。

建立一份交接清单：样本名、分组、BAM 完整路径、prep 批次、`.rmats` 来源目录、GTF 版本、rMATS 版本、读长、文库类型和任务状态。批次文件复制后记录前缀与文件数量，运行 post 前留存清单快照。发生失败时只重做受影响的 prep 批次，并在新的汇总目录重新整理文件；不要在原目录反复覆盖，避免新旧 `.rmats` 被同一次 post 混读。

rMATS 官方 integrity 任务只核对 BAM 路径与临时记录的一对一关系，它不能证明 GTF 正确，也不能评价统计模型是否符合实验设计。quickcheck 可快速发现缺少头部或文件结尾的 BAM，但不会读取文件中间所有数据。因此仍要结合完整运行日志、样本表与结果文件做最终验收。

参考资料

rMATS-turbo 官方 README：分阶段 prep/post 与完整性检查 · 官方资料

rMATS-turbo 官方参数说明 · 官方资料

samtools quickcheck 官方手册 · 官方资料