# 检查FASTA、GTF、版本与索引构建日志

先核对参考序列与注释是否属于同一版本，再区分格式、权限与索引资源问题，保留能定位失败阶段的证据。

Cell Ranger自建参考构建失败时，先区分输入和注释问题、STAR索引资源不足，以及运行目录或权限问题。FASTA和GTF序列名称不一致，不会因为增加内存而修好；输入已核对但在STAR索引阶段退出，才需要结合内存和日志评估资源。本文只讨论Gene Expression的cellranger mkref。

## 先固定文件来源和版本

记录Cell Ranger版本、物种、组装版本、FASTA与GTF来源和日期。尽可能从同一来源取得同一组装的两个文件；一个版本的FASTA不能直接与另一组装的GTF混用。把原始文件单独归档并记录SHA256，所有过滤和构建都在新目录生成新文件，便于失败后回到原始输入。

## 比较FASTA片段名和GTF序列名

FASTA片段名取标题行第一个空格前字段；GTF第一列也必须指向该片段。以下命令列出两边名称并查找仅出现在一边的值。名称集合匹配不代表组装生物学正确；还要确认GTF含第三列为exon的记录，并且第九列含gene_id和transcript_id属性。

```
grep '^>' reference.fa | sed 's/^>//' | awk '{print $1}' | sort -u > fasta_names.txt
grep -v '^#' annotation.gtf | cut -f1 | sort -u > gtf_names.txt
comm -23 fasta_names.txt gtf_names.txt > fasta_only.txt
comm -13 fasta_names.txt gtf_names.txt > gtf_only.txt
grep -v '^#' annotation.gtf | awk -F '\t' '$3=="exon"{n++} END{print "exon_rows",n+0}'
grep -v '^#' annotation.gtf | grep -m 3 'gene_id'
```

若名称不匹配，先核对下载包和染色体命名规则。不要全局替换chr前缀来掩盖组装差异，因为这可能造成名称碰撞。若GTF属性需转换，先参照10x格式要求和来源注释说明，并保留过滤规则和记录数。mkgtf筛选会改变保留基因类型，属于分析输入定义，应在项目记录里说明。

## 根据失败阶段判断是否资源不足

使用独立输出目录运行并保存完整日志。若在输入预检阶段停止，先修复路径、格式、FASTA/GTF对应关系或权限。若日志显示STAR genome index阶段，并且作业被内存限制终止，才核对实际作业额度和对应版本的--memgb。当前v10.1文档指出，memgb必须大于输入FASTA的基因组大小；典型3 Gb人基因组构建示例约需32 GB内存和8 core hours，但具体注释、版本和环境会影响结果，不能当作所有任务的固定规格。

| 现象 | 先核对 | 处理方向 |
| --- | --- | --- |
| 找不到文件或参数错误 | 路径、权限、命令帮助 | 修正命令，不扩容 |
| 序列名称不匹配 | 组装来源与名称集合 | 换配套文件或有记录地修正 |
| GTF注释报错 | exon、gene_id、transcript_id | 检查来源与格式要求 |
| STAR阶段OOM | 作业内存、memgb、并发 | 核实额度后调整资源 |

## 构建后做最小验收

日志出现参考成功创建提示后，检查reference.json、fasta、genes和star子目录，记录输入及输出清单。再选与文库类型匹配的小样本运行对应pipeline，确认Cell Ranger能读取参考并生成预期summary。索引目录存在不等于注释适合研究问题。保存软件版本、输入校验值、过滤规则、资源记录、日志和验收样本，后续复用时才能知道索引由什么数据生成。

此外，FASTA体积不是mkref内存的直接换算值。先用fasta文件记录和版本号判断组装规模，再按官方参数说明设置memgb，并确认作业调度器实际给了足够内存；参数申请超过队列上限时，作业仍可能启动失败。构建时间较长时保留

独立日志和运行目录，不要在其他成员正在使用参考时覆盖旧路径。需要在多个项目间复用时，发布新版本目录并同步提供输入来源、校验值和验证结果。

## 参考资料

Building a Custom Reference for 3′ Gene Expression（10x官方） · 官方资料

Cell Ranger v10.1命令行参数（10x官方） · 官方资料

Troubleshooting Cell Ranger Errors（10x官方） · 官方资料