Baysor大规模分割实操：先做输入检查，再验证区域与资源

用Baysor C++ 0.9.0核对分子表、坐标单位和参数；先预览代表区域，再运行完整独立区域，并按输出文件和资源日志验收。

空间转录组分子表很大时，直接全量运行会让输入错误、参数不合适和资源不足混在一起。下面以当前原生 C++ 版 Baysor 0.9.0 为例，按“核对输入—小范围预览—完整区域试跑—检查结果”的顺序处理。旧 Julia 版 0.7.1 的命令和部分输出定义不同，先用 baysor --version 记录实际版本；新版本用 --threads 控制线程，不要把 JULIA_NUM_THREADS 当成等价设置。

先检查输入与坐标单位

每行应代表一个检测到的分子。基础表需要数值型 x、y 和非空 gene 列；三维数据还要确认 z 有效且不是全列常数。若平台列名不同，可用 -x、-y、-g 映射；不要只改表头而漏掉对应参数。检查坐标是否落在预期视野范围、有无 NA、异常单位或整批重复记录。处于相同坐标的真实分子不应仅因坐标相同就删除。transcript_id 是可选列，若下游流程依赖它，再按目标格式确认是否保留。

参数	怎么理解	如何确定
-s / scale	近似细胞半径	与坐标同单位，不是直径；先查平台标定并看预览
-m	真实细胞预期分子数	按检测平台和组织稀疏度设定；它不是最终小细胞过滤阈值
--threads	工作线程数	不超过任务配额；记录实际使用值
prior	可选的既有分割	确认坐标、标签及先验置信度含义

先预览，再选代表区域

对大型数据先用坐标范围生成预览，例如 baysor preview -m 30 --x-min 0 --x-max 2000 --y-min 0 --y-max 2000 -o preview.html molecules.csv。这里的 30 和坐标范围只是命令示意，不是适用于所有平台的建议值；预览也需要正整数 -m。打开 HTML 检查低质量分子、局部基因组成和空间分布，再决定是否调整参数。坐标范围须使用输入表的单位。

分割试跑优先选无需跨边界合并的真实完整视野或组织区域，并记录样本、区域ID、坐标范围、输入分子数和文件校验值。小裁块适合比较参数、估算资源，却不能把各块的细胞编号直接拼成全组织结果：边缘细胞、跨边界分子和局部上下文都会改变分割。若必须分块，需要单独验证重叠宽度、边界归属、跨块去重及细胞身份合并规则；在验证前，应把分块结果当作探索性结果。

运行与结果验收

先选一组有依据的 -m、-s，用独立输出目录运行一个代表区域。确认日志中的线程数和输入路径，再扩大到其他完整区域；不要先把多个样本拼成一个分割任务。示例中的参数仅演示命令结构：

```
baysor --version
baysor preview -m 30 --x-min 0 --x-max 2000 \
  --y-min 0 --y-max 2000 -o preview.html molecules.csv
baysor run -m 30 -s 8 --threads 8 -p \
  -o run_region_01 molecules.csv
find run_region_01 -maxdepth 1 -type f -printf '%f\n' | sort
```

把示例中的文件名、坐标范围、半径、分子数和线程数换成实际值。scale=8 只有在坐标单位及细胞半径与之相符时才成立。每个区域用独立目录，避免覆盖前次结果。默认输出包含分子分配 segmentation.csv、细胞统计 segmentation_cell_stats.csv、多边形和计数矩阵；加 -p 会生成 diagnostic_report.html 与 segmentation_plot.html。先确认这些文件存在且非空，再打开报告检查噪声分子比例、异常大或过小的细胞、边界形态，以及每细胞分子数分布。输出中的 cell 0 表示噪声/未分配，不应计为普通细胞。

用实测值规划资源

用同一代表区域和参数测一次墙钟时间、峰值内存、线程数及输入输出大小，再按区域分开估算后续批次。不能只按分子数线性外推：组织密度、基因组成、聚类和报告生成也会影响耗时与内存；因此至少比较稀疏和密集区域。若试跑内存接近可用上限或磁盘写入明显增长，先降低并发、缩小单次区域或检查输出需求，再决定整批运行。把 Baysor 版本、命令、输入校验值、坐标单位、参数、日志和输出目录写进运行记录，另存原始输入及参数文件，便于复查和复现。本文没有声称在特定硬件上完成基准测试，资源计划应以本实验室试跑记录为准。

参考资料

Baysor 官方分割参数与输出检查 官方文档
Baysor 官方输入格式与坐标过滤 官方文档
Baysor 官方预览数据集教程 官方文档
Baysor 官方输出文件说明 官方文档
从 Julia 版迁移到 C++ 版 官方文档