NicheNet多细胞对实操：复用参考、保存输入并控制并发

完整操作教程 · 参考固定 · 往返恢复验收

把各发送—接收细胞对的输入独立保存，顺序复用同一NicheNet模型，防止并发复制参考矩阵和混淆背景基因。

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。

NicheNet分析多个细胞对，适合复用的是固定版本的先验模型，不是把一个细胞对的背景和差异基因复制给所有组合。本文从已经确认的发送细胞表达基因、接收细胞表达基因及响应基因集开始，批量计算配体活动。差异分析和表达比例标准应在上游方案中确定，并随每个细胞对保存。

准备可以独立审查的细胞对清单

文件或字段	内容	判断依据
ligand_target_matrix.rds	靶基因行、配体列的先验矩阵	物种、基因体系及模型版本
lr_network.rds	含from、to列的配体—受体网络	与先验模型及物种兼容
pairs.rds	唯一名称的列表，每项四个字符向量	以下四项分别有上游来源
geneset、background	接收细胞响应基因、表达背景	响应基因属于背景；方向和对照明确
sender、receptors	发送者表达基因、接收者表达受体	表达标准预先确定而非按结果调参

可以用saveRDS(list(pair_A=list(geneset=响应向量,background=背景向量,sender=发送者向量,receptors=受体向量)),"pairs.rds")保存真实上游向量。这里不填虚构标记基因代替分析。背景应代表接收细胞中有机会被选为响应的基因，不能只用显著基因；不同接收者与比较一般需要分别计算。

载入一次模型，按细胞对顺序计算

```
library(nichenetr)
M <- readRDS("ligand_target_matrix.rds")
lr <- readRDS("lr_network.rds")
pairs <- readRDS("pairs.rds")
stopifnot(is.matrix(M),!is.null(rownames(M)),!is.null(colnames(M)),
 !anyDuplicated(rownames(M)),!anyDuplicated(colnames(M)),
 all(c("from","to") %in% names(lr)),!anyNA(lr[,c("from","to")]),
 is.list(pairs),length(pairs)>0,!is.null(names(pairs)),
 !anyNA(names(pairs)),all(nzchar(names(pairs))),!anyDuplicated(names(pairs)),
 !file.exists("nichenet-output"),dir.create("nichenet-output"))
check <- function(x) is.character(x) && length(x)>0 && !anyNA(x) &&
 all(nzchar(x)) && !anyDuplicated(x)
for (i in seq_along(pairs)) {
 x <- pairs[[i]]
 stopifnot(all(c("geneset","background","sender","receptors") %in% names(x)),
   all(vapply(x[c("geneset","background","sender","receptors")],check,logical(1))),
   all(x$geneset %in% x$background))
 bg <- intersect(x$background,rownames(M))
 gs <- intersect(x$geneset,bg)
 lig <- intersect(unique(lr$from[lr$from %in% x$sender & lr$to %in% x$receptors]),
                  colnames(M))
 out <- sprintf("nichenet-output/pair-%03d",i); stopifnot(dir.create(out))
 saveRDS(list(name=names(pairs)[i],original=x,background=bg,geneset=gs,
   potential_ligands=lig),file.path(out,"inputs.rds"))
 writeLines(c(paste("background",length(bg),"of",length(x$background)),
   paste("geneset",length(gs),"of",length(x$geneset)),
   paste("ligands",length(lig))),file.path(out,"coverage.txt"))
 stopifnot(length(gs)>0,length(bg)>length(gs),length(lig)>0)
 ans <- predict_ligand_activities(geneset=gs,background_expressed_genes=bg,
   ligand_target_matrix=M,potential_ligands=lig,single=TRUE)
 stopifnot(setequal(ans$test_ligand,lig),!anyDuplicated(ans$test_ligand))
 saveRDS(ans,file.path(out,"activities.rds"))
 stopifnot(all(vapply(ans[,c("auroc","aupr","aupr_corrected","pearson")],
   function(v) all(is.finite(v)),logical(1))))
 write.csv(ans,file.path(out,"activities.csv"),row.names=FALSE)
}
writeLines(capture.output(sessionInfo()),"nichenet-output/session.txt")
```

```
/usr/bin/time -v -o nichenet.time.txt Rscript pairs.R > nichenet.log 2>&1
```

保存为pairs.R后运行，矩阵仅在进程开始载入。每个编号目录保存实际使用的背景、响应和候选配体，可从name字段映射回细胞对。输入不存在模型中的基因被交集排除，必须查看coverage和原始列表，解释缺失，而不是自动接受覆盖率下降；这里没有给出所有研究通用的合格百分比。

结果如何验收

官方函数返回配体及AUROC、AUPR、校正AUPR、Pearson等活动指标。检查候选配体是否完整、得分是否有限及有无警告，再按选定指标排序。非有限值不能默认为零参与排名；先查响应与背景是否退化、候选是否缺少信息，并保留失败细胞对的输入。

结果是对响应基因的解释能力与候选通讯假设，不等同于已经证明配体结合、受体激活或因果作用。若两个发送者共享同一候选配体，活动值相似也不直接说明发送者贡献相同；需结合表达、空间和实验信息。只看高分配体名单无法判断输入是否合理。

在有效细胞对上测试并发

记录先验矩阵大小、细胞对数量、每对候选数和完整峰值RAM。顺序计算可避免每个worker各载一份大矩阵；若要并发，用相同模型和输入试少量独立任务，观察任务组总内存。更多CPU适用于有效任务吞吐不足的情形，不能修复响应定义、物种或参考匹配问题。

保留失败细胞对而非填充分数

如果候选配体为空，先查发送者表达、接收者受体、网络边与模型列名的各步交集。若响应集等于背景，活动评估也缺少区分条件。此类任务应单列为输入或方法失败，不补零分数混入其他细胞对的排名。

先把发送者、接收者和响应基因集定义清楚

NicheNet的ligand activity分析需要明确哪类细胞作为sender、哪类作为receiver，以及receiver中哪组响应基因被解释。通常geneset来自receiver中与问题相关的差异或上调基因，background来自receiver中实际检测且模型矩阵覆盖的表达基因；sender侧用于筛选表达配体，receiver侧用于筛选受体。若角色方向反了，分析仍可能输出排名，但回答的是错误问题。把条件比较方向、细胞注释版本、差异分析阈值和每一组使用的细胞数写入pair配置。

基因ID映射和模型覆盖率先于运行

表达矩阵、ligand-target矩阵和LR网络必须采用兼容的物种和基因符号体系。先统一Ensembl ID、旧符号、同义名与大小写，再报告geneset、background、sender ligand、receiver receptor分别与模型覆盖的交集数量。背景集应表示在接收细胞中可检测或可能响应的基因，并限制在模型矩阵可评估范围内；geneset需是背景的子集。若交集过小，先查物种、ID映射和过滤条件，不要用任意补入的基因扩大背景。

activity分数是模型优先级而非因果证明

AUROC、AUPR、corrected AUPR、Pearson等指标描述模型的ligand-target先验能否区分观察到的基因集，不表示配体在实验中真实分泌，也不能证明受体结合导致该响应。排名会受geneset大小、background构造、矩阵版本和potential ligand集合影响。建议报告多个指标、候选配体的表达与受体关系、模型覆盖率和敏感性分析；若基因集小，可通过合适的置换或bootstrap评估排名稳定性，但要说明抽样设定。不要把排名第一写成已确认的上游调控因子。

细胞类型注释与伪重复问题

若receiver或sender由单细胞聚类得到，先核对每个样本的细胞数、样本来源和注释可靠性。把多个供体合并后直接做细胞级表达或差异检验，可能把细胞当成独立生物学重复；差异基因和表达筛选应按研究设计选择合适的样本级或伪bulk策略。不同供体的组成差异也会改变配体和受体检出。把样本、批次、条件和细胞类型的映射保留在输入表中，对低细胞数的细胞群标记不稳定性。

按pair拆分任务并控制内存

每个sender-receiver对的geneset、background、potential ligands可分开保存，便于发现哪对输入交集为空或模型覆盖不足。先用一个代表pair测运行时间与内存，再逐步增加并行；矩阵本身可能较大，多个R进程各自加载矩阵会复制内存占用。可在计算节点预加载共享对象或按平台推荐方式并行，但需要验证序列化和worker内存。失败时保存pair ID、输入交集统计和R session信息，不要因为一个pair失败就重新运行全部条件。

先检查输入特征数再解释分数

每个pair运行前记录geneset总数、背景数、与模型矩阵交集数、potential ligand数以及sender/receiver的表达比例。覆盖率很低时，活动排名容易由少量基因驱动；背景集过窄或过宽也会改变AUPR等指标。可以绘制覆盖漏斗和每一步删除原因，先确认ID映射而非不断调小阈值。所有组别使用一致、可解释的筛选规则，若因物种模型覆盖不同而调整，需明确标注。

表达配体和受体并不等于真实信号传递

LR网络中的配体—受体边来自先验数据库，不代表本样本中蛋白已分泌、受体已位于细胞表面或相互作用发生。表达筛选阈值会影响候选集，低表达但有功能的配体可能被剔除，而环境RNA或双细胞也可能造成假阳性表达。结合细胞质量控制、样本来源、蛋白/空间证据和已知生物学背景审查优先候选。把表达支持、网络先验和activity评分分列报告，避免合成一个“通信证据”标签。

结果展示应保留排名稳定性

对重点pair可改变合理范围内的差异基因阈值、表达比例或背景定义做敏感性对照，观察top ligand是否稳定。小基因集可采用随机基因集或bootstrap方案构建参考分布，但需要匹配背景和基因集长度，并报告迭代次数和随机种子。不要把活动分数差距很小的候选硬排成确定次序；可以展示排名区间或多个指标并列。若不同设定下候选完全改变，结果应表述为依赖参数、需进一步验证。

给activity排名加上可验证的证据链

为排名靠前的配体逐项整理：sender是否表达该配体、receiver是否表达候选受体、LR数据库中是否有对应关系、响应geneset中哪些基因是该配体的预测靶标，以及这些基因在实际数据中的表达变化。把每一层证据和来源分开记录。可以挑选多种方法或独立数据进行交叉检查，但不要只根据NicheNet自身的先验网络循环验证自身结论。最终候选仍应通过实验或独立队列进一步检验。

条件比较要保持方向一致

若geneset来自两个条件间差异表达，必须明确哪一个条件作为receiver的响应方向以及geneset取上调、下调还是全部差异基因。把方向相反的基因混在一个响应集合，可能使ligand activity难以解释。分开分析上调和下调、或按预先定义的响应程序建立多个gene set，并保持背景与模型参数一致；同一细胞类型的不同条件比较则分别保存输入清单，避免复制时沿用错误的sender或receiver。

如何把结果整理给实验人员

交接表可包含配体、发送者表达比例、受体、接收者表达比例、activity指标、模型覆盖的响应靶基因、差异方向、不同样本支持和优先验证理由。将低覆盖、仅单样本支持或对阈值敏感的候选单独标记。实验团队可据此选择阻断/刺激条件或测量指标，但分析结果本身不保证某个配体是唯一调控者。

复现材料清单

归档R与NicheNet版本、LR网络和ligand-target矩阵来源及校验和、细胞注释、表达矩阵、差异结果、geneset/background构建脚本、pair配置、随机种子、运行日志和每对输出。若使用了预计算的人或鼠模型，报告其物种和版本。不要只保存最后一张热图；排名依赖输入基因集合与模型版本，完整中间信息才便于复核。

按真实规模做试运行

如果使用多个receiver细胞群，应为每个群分别建立表达背景和响应geneset，避免用一个大背景混合不同细胞类型。保存每个pair的细胞数和样本数，确保活动排名比较的是同一分析口径。跨群体排名差异可能来自表达覆盖和基因集差异，不一定代表配体作用强弱真的不同。

建议在输出表中保留未通过输入检查的pair及原因，而不是从结果文件中静默删除。这样团队能区分“没有候选配体”“没有足够模型覆盖”和“分析没有执行”，后续补数据或调整实验设计时也知道从哪一步继续。

结论与复核记录

如多个ligand-target矩阵版本并存，确认它们来自同一物种与相同建模流程。矩阵规模和数据库更新会影响交集及排名，不能只因文件名相同就认为参数一致。保存模型来源、版本、下载日期及哈希，之后才能准确复跑。

对最终候选做生物学解释时，记录受体表达、通路背景和目标基因方向，并指出哪些证据来自模型先验。若不同证据冲突，保留冲突和不确定性，不要通过删除不支持的候选让排名看起来更整齐。

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

NicheNet官方分步教程 · 来源资料

配体活动函数与输出列 · 来源资料

NicheNet官方说明 · 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。