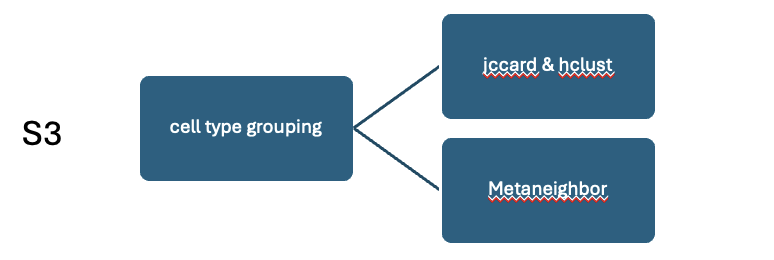
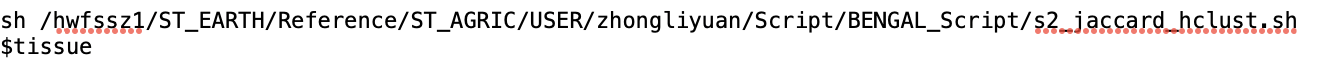
1第一个workflow





输入：/Files/ResultData/Workflow/W202408290000623/glob-3f996b450b66fda16ba957c868415e8d/\*.rds

2第二个workflow

merge.sh

SCTransform的seurat整合方法——在merge的前面的处理

SCTransform的seurat整合方法：<https://github.com/satijalab/seurat/issues/7542#issuecomment-1631534467>

merge 需要包含几种integration方法，RLIGER，scvi，harmony必须

输入：水蕨茎用来做integration的输入：任务号，W202408270004060；W202408290001968； W202409020000024；W202408270004059

https://github.com/satijalab/seurat/issues/7542#issuecomment-1631534467

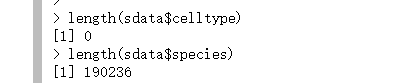
解决环境问题

cell\_type\_grouping 使用环境s2\_jaccard\_R，后面统一名称为c-type-group\_r\_01

开始测试脚本jccard

注意脚本的意义所在

many\_higher\_expr.rds的celltype和species数量不对等



Integration的资源准备完成，先把integration的部分测试起来。

第一个的两个R脚本已经测试ok后面就是使用work.sh测试一下。

对于scvi和harmony的参数进行修改

明天将harmony和scvi进行封装，另外对于rliger进行测试。学习SCT

2024/9/6

1Jaccard\_meta的测试完成，输出的可视化的结果，里面使用了SCT，问题是SCT的作用是什么，与一般seurat处理的区别是什么呢？

2Harmony和scvi的work.sh都已经完成，但是测试结果不佳，对于机器学习部分，如何提高学习效率和学习的质量，在参数的设置如何达到优化的问题？

3Rliger的部分已经用convert进行的封装，另外是否可以和rliger一起去解决，所以install rliger。然后将rliger封装在一个环境中。

4另外对于钟博说的SCT的可选项操作，我们还是要去看github对于SCT的解释和说明。SCT的意义是什么！

SCTransform的位置问题，它是在第二个wdl之前进行处理，且是可选项？我发现SCT是R语言，输入也是SeuratObject，SCT代替了以往的Seurat的normalization，Finder Variable Features，scale data。生成了一个SCT assay，那么我们回到哪个问题，SCT处理的数据应该是RNA的counts部分？，SCT copy一份经过SCT处理的RNA的counts部分生成的新的SCT assay，那么对于integration而言，它的处理对象就应该是SCT assay。所以我们应该设置active assay 为SCT assay， DefaultAssay(object = scRNA) <-"SCT" ， 对于rliger而言无伤大雅，将处理后的SCT输出为rds，而对于将要做scvi和harmony处理来说，应该是要输出为h5ad，即anndata格式的数据，那么如何定义h5ad数据结构的结构层数也是SCT，或者换个方式我们可不可以只输出SCT 的assay，对于counts不输出。这是一个关于SCT处理之后的输出问题，当然也是对于Seurat和Anndata数据结构的重新认识。

我似乎明白了，harmony本身是一个python脚本，通过SCTransform可以实现再R里面integration。同时scvi是没办法实现的在R里面。SCTransform的实质是改变了normalization的质量，为什么不考虑对rliger的前处理进行替换。

结尾：scvi和harmony封装好了，scttransform和rliger都封装好了，后面就是对其进行加判断，什么时候用harmony，什么时候用SCTransform。Ok，下周再做个，回去要好好看书。

对于SCTransform\_harmony还可以优化CellType为celltype，这样可以增强其通用性

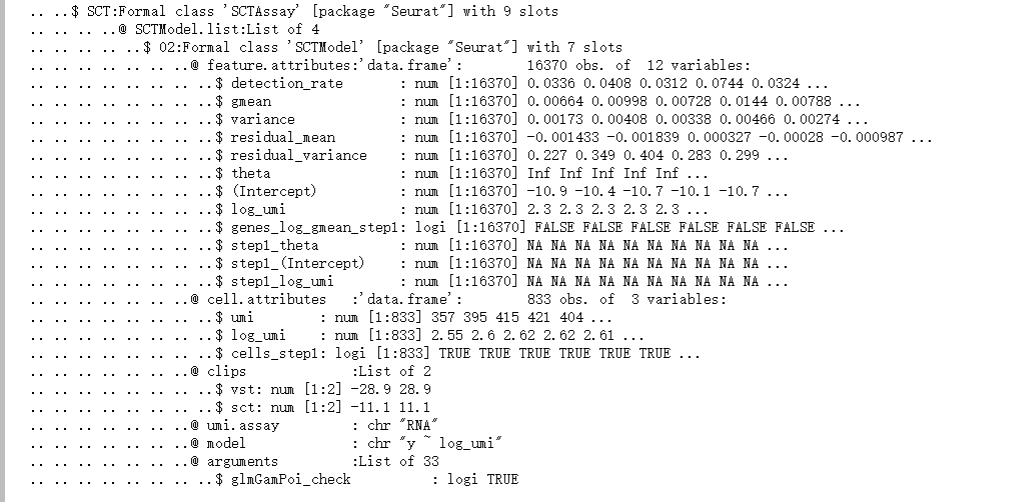
关于图的内容作为修改，设置biosample用于设置整合。最终输出输出整合后的。

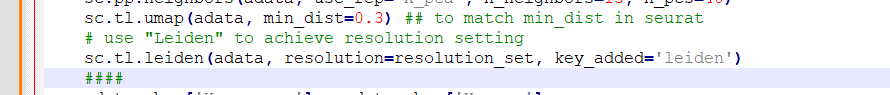
输出的图像三个图

另外设计一个python脚本用于concat两个生物样本的整合，并对其进行直接的处理，比如biosample，celltype，species的问题，输出处理好的h5ad。

另外对于各个integration的resolution的参数进行传参！输出的图像改变，biosample,species,celltype.还有三个小提琴图的认识。就输出四个图吧，多比少好。

输出小提琴图





之前在dataget部分已经使用leiden算法进行设置了resolution，再次使用leiden算法会不会影响效果呢？

To achieve the future defaults please pass: flavor="igraph" and n\_iterations=2. directed must also be False to work with igraph's implementation.

sc.tl.leiden(adata, resolution=resolution\_set, key\_added='leiden')

果然报错了！！！

关于harmony部分的图像输出等问题

scvi的测试

Liger和SCT的测试及其输出

Wdl

Scvi与leiden不兼容，使用这个算话似乎不可靠，Louvain算法的代码，暂且使用这个吧

sc.tl.louvain(adata, resolution=0.5)

2024/9/11

scvi的环境问题后面再去解决，先把当下能跑出来的跑出来

Scvi的conda安装一直有问题？

Harmony已经封装好了，希望不要报错。

另外其它脚本应该问题不大，跑一下就指导了，sct和liger在离线测试都不错。

关注一下输入和输出的正常性

另外对于sct和harmony的选择控制后面在设定吧

