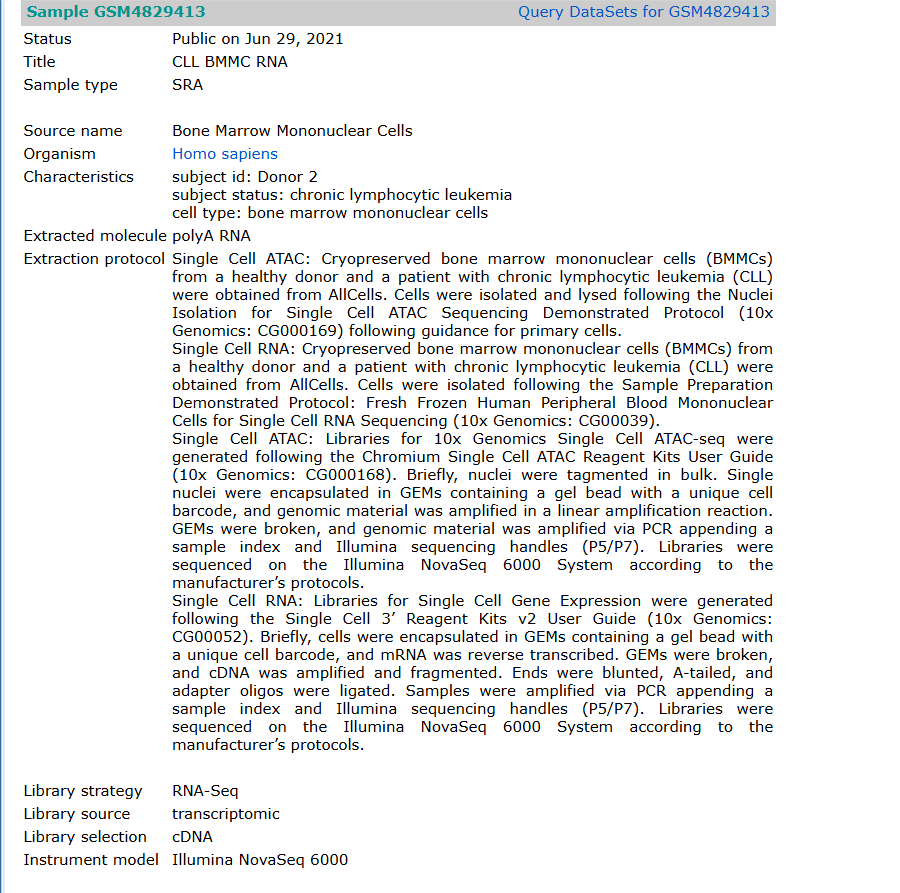
下一步：在GEO里面找一个数据集用Scanpy处理出scRNA和scATAC 并构建出对象 进行分析HLF

一开始搜索的是B-lineage acute lymphoid leukemias（GEPIA里面说的HLF直接相关）但是发现并没有这个单细胞数据集

最后选择了一个这个

这个的搜索条件为Search (((leukemia[Title]) AND single cell[Title])) AND (b lymphocytic leukemia AND single cell)

选了一个这个只有一个样本的 有scrna & scatac



看了一下celloracle提供的scrna & scatac处理过程

分别用的是scanpy & cicero

* 看一下作为scanpy的输入合不合适

可以的：10x的三个标准文件：h5，txt，csv

* Cicero呢？

应该可以 先试试这个

Scanpy处理

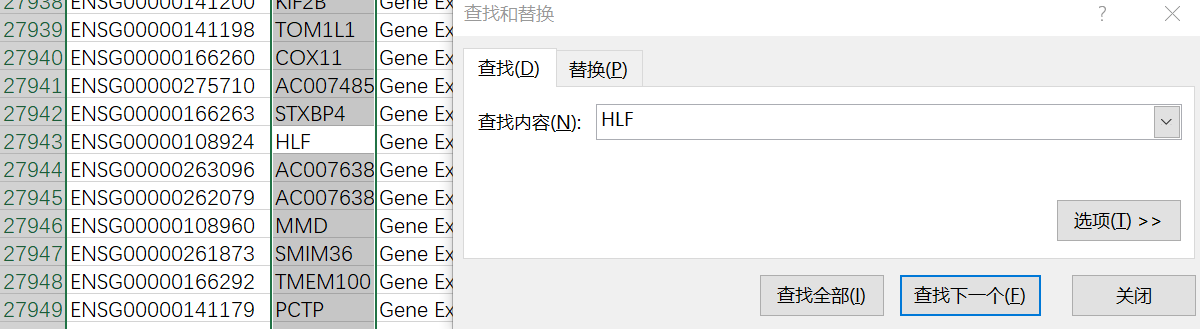
看了一下之前的两篇教程没有说输入标准的 新在网上搜一下

暂时跑完了：总结一下：输入前一定看一下表格

结果：感觉数据质量不太好 很多marker都没出来

重点：HLF也不在前2000

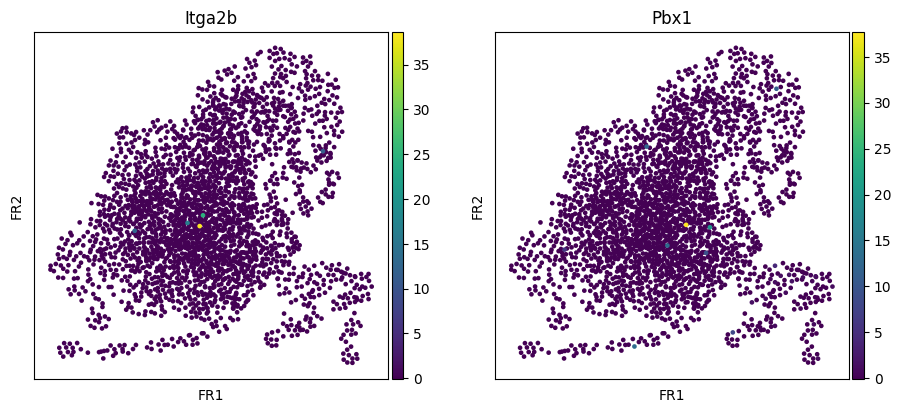
在feature里面本来是有的 在feature\_makers之后就没有了 悲~

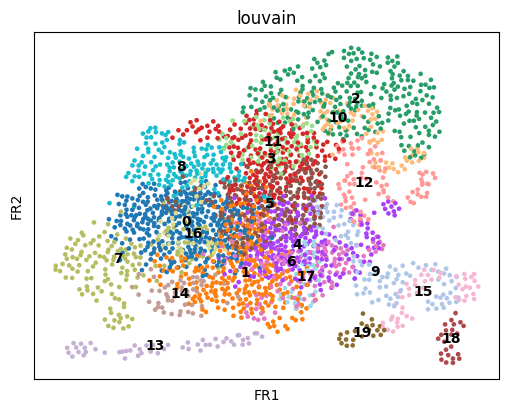


感觉这个数据集还是不行】

Scanpy: not HLF as marker genes







要不要用TOSICA注释一下？

Cicero处理

我之前是用的R包跑的

但是我现在重新看celloracle的教程发现也是用的R包

用的也是10X的数据 可以的

首先是要在中服务器配置cicero 参考大服务器的时候吧

按照celloracle官方的教程 需要用到monocle3的函数

但是在安装units包的时候遇到了问题

Cicero: reducing dimension using tSNE

在使用样本的ATAC数据构建GRN的时候 花的时间实在是比较多

下面这几行代码跑了两天还是不行

%%time

# Scan motifs. !!CAUTION!! This step may take several hours if you have many peaks!

tfi.scan(fpr=0.02,

motifs=None, # If you enter None, default motifs will be loaded.会自动加载上面所说的gimme.vertebrate.v5.0

verbose=True

)

直接放弃