RNAseq pipeline更新汇总

特点：4-6样本在3小时内可以跑完， 不再消耗大量内存。

代码地址：<https://github.com/adamtongji/rainbowRNA>

标准部分：

1. 可执行名为”rnapipe\_test”, 运行方式依旧是“rnapipe\_test config.txt”。原来的pipeline不变动。
2. 现在mapping 软件为hisat2，一般一个文件1-2小时，4CPU一个样本，我们服务器满负荷是8个样本同时运行（多进程 ：8），一般设为6 或者4。 内存暂时没有瓶颈。 表达量软件为featureCount，不再需求samtools sort。 Mapping rate文件现只有单独文件的mapping情况，同star的mapping rate文件不太一样。
3. genome index文件为hisat2的index, 常见物种在hisat2官网都可以下载，大小为几个G，由于内存原因，现不支持重新build基因组index文件。 Hg19在/home/Public/hisat\_index/hg19\_index/，前缀为genome, 因此我们设置参数需要为： /home/Public/hisat\_index/hg19\_index/genome 。
4. 新增Expr\_dir参数，默认为输出文件夹下expr文件夹。 该参数是为了支持从自定义文件夹（本地） 单独跑差异表达及下游分析，输入文件需要是同count表达量文件格式相同。

CircRNA

1. 现targetscan不再扫描exon以外的区域。现支持人和小鼠， rat还未测试。

**RNAseq测试需求：**

1. Single end数据是否能跑通（pairend第一个）

2. 在本地从自己expr\_dir开始用de\_full是否稳定

3. 不同类型hisat2 index对运行时间和mapping率是否有影响（只跑mapping那步）

**CircRNA测试需求：**

1. 以前客户数据测试是否能找到circRNA，是否找到合理数量的miRNA target。

**主要反馈内容：**

1. 是否能跑完整，是否报错，以及差异表达富集和以前结果差别如何？
2. 样本数，文件大小， 运行总时间（以log中中文时间为准，计时器多线程暂时有Bug）。
3. 其他常见或特殊的需求。