

# MSAM: MBD-Seq Analysis Method

## 中文使用手册

该软件集成了 Bowtie、MACS 以及课题组自己编写的处理 MBD-seq 的软件包。可以从测序数据开始分析，直至差异甲基化区域的筛选及功能注释。

适用操作系统：Linux(For All Applications);

Windows(For Applications without Bowtie and MACS)

需安装的软件：Java 虚拟机（最新版本）、Bowtie-0.12.7、MACS-1.4。

需提供的数据文件：fq 格式的原始测序数据(测序生成);

Bowtie 需要的物种序列索引 (Bowtie 网站下载);

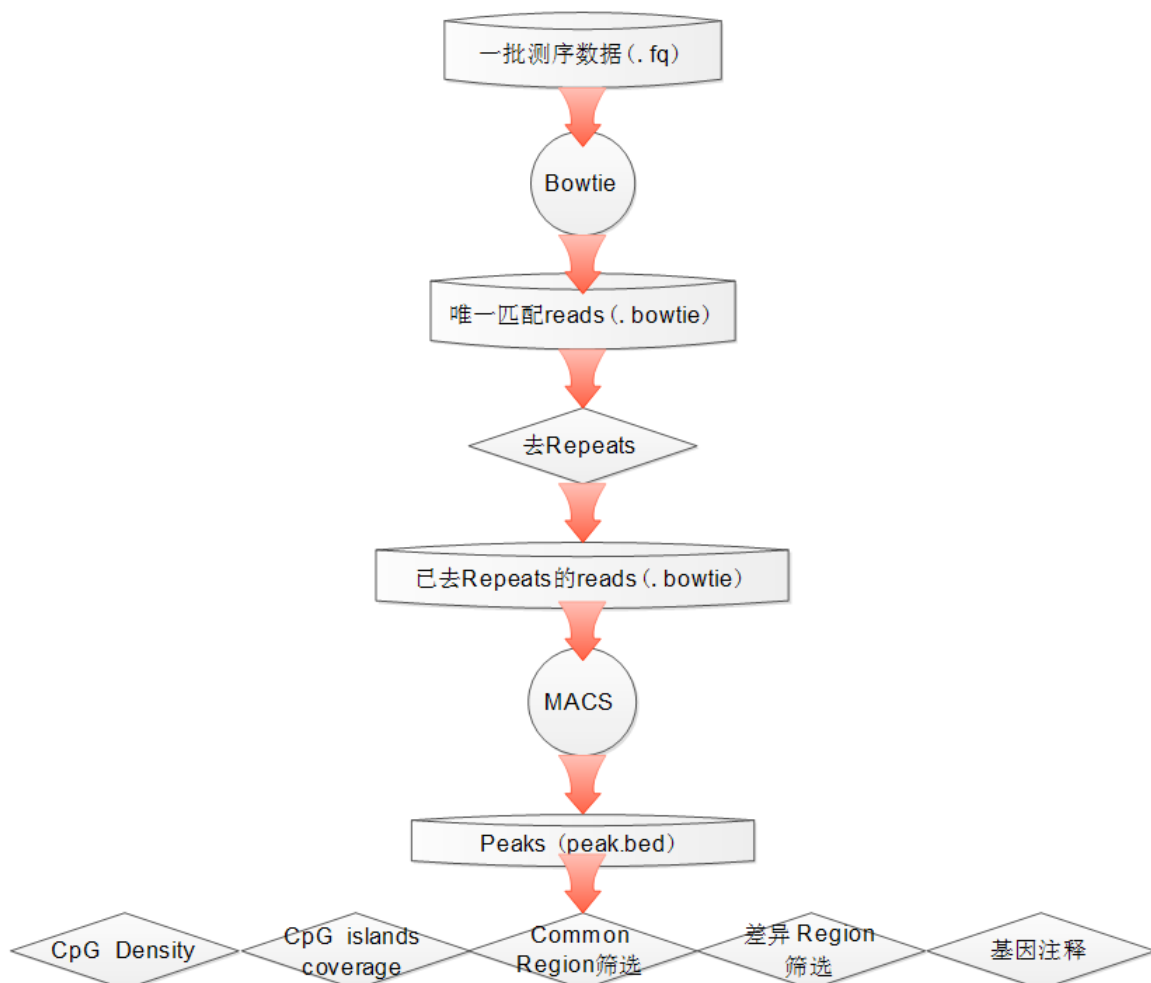
物种的重复序列数据 (UCSC 下载，含重复序列类型);

物种 CpG 位点位置文件 (Perl 程序计算);

物种 CpG 岛文件 (UCSC 下载);

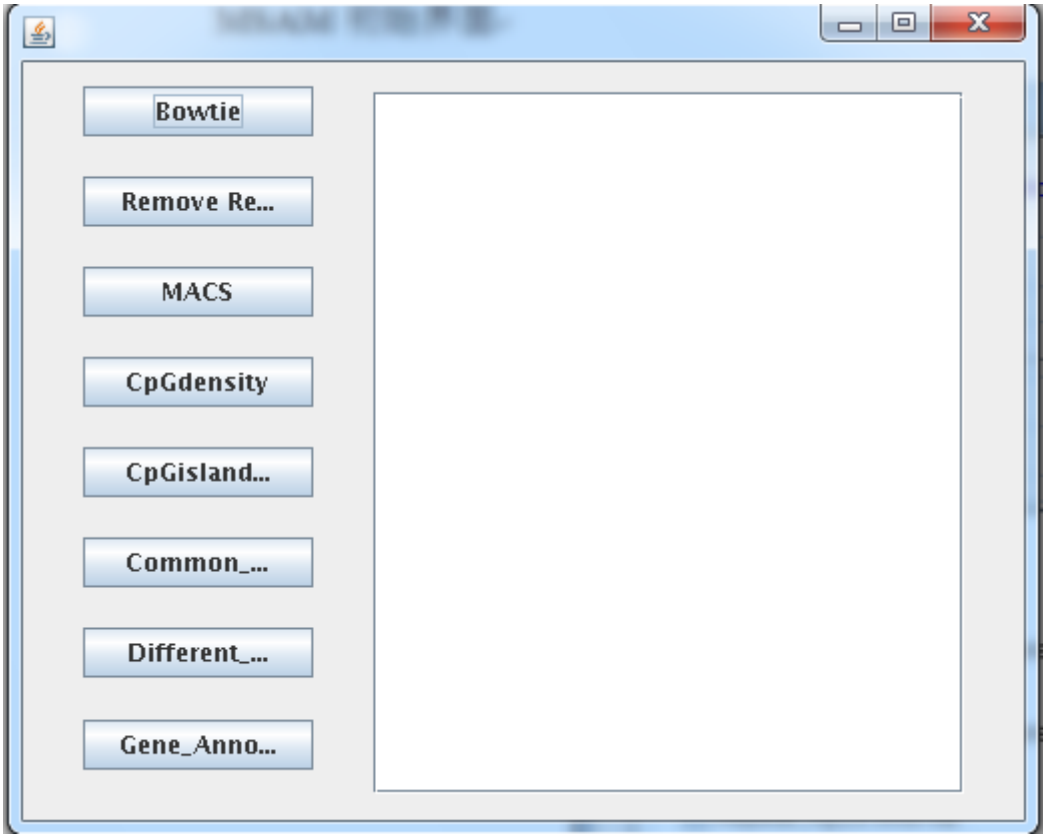
Refseq 基因文件 (UCSC 下载)

使用流程：

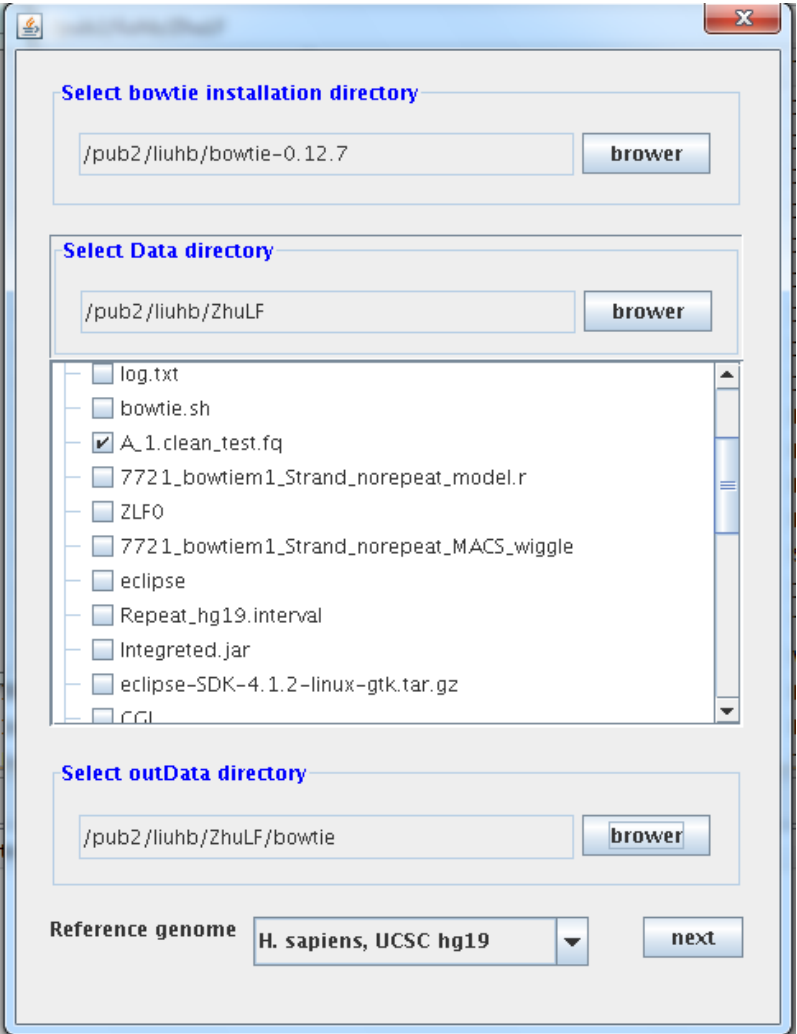


使用说明：

MSAM 初始界面



点击 **bowtie** 出现



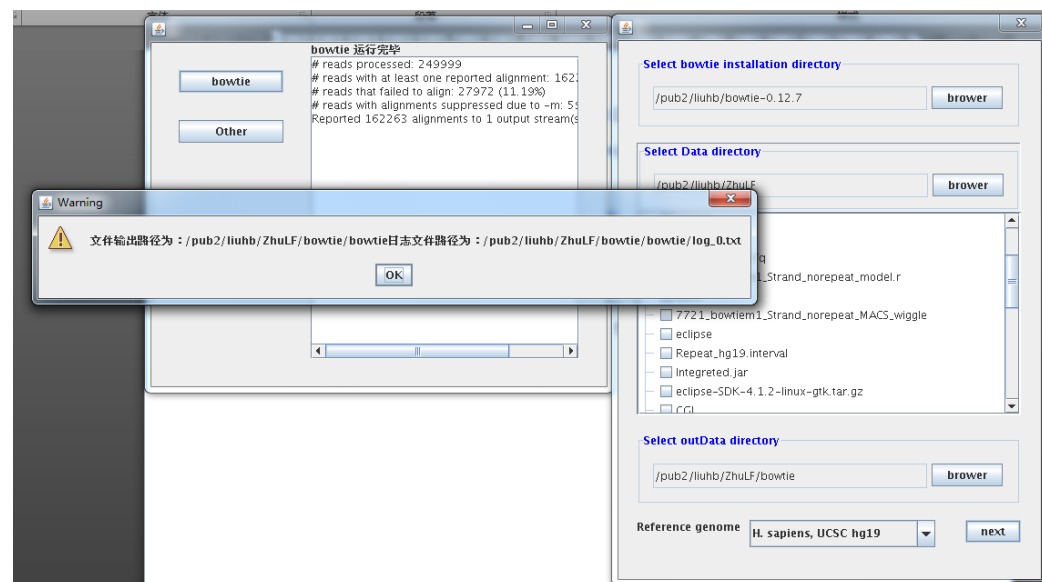
- 1、请先将对应物种的序列索引加入 bowtie 的相应目录 indexes 中；目前该软件支持 Bowtie 网站提供的所有 index 文件（请用默认文件名）。
- 2、Select Data directory 是您存放测序 fq 格式文件的文件夹，可在下面的框中选择您要进行分析的 fq 文件，可以多选几个，进行批量处理。
- 3、Select outData directory 设置输出目录，请尽量选择程序命名的 bowtie 文件夹

点击 next 开始进行序列比对！

数据分析过程可能会比较长，请耐心等待！！切勿关闭程序！！

运行完后结果如图所示：

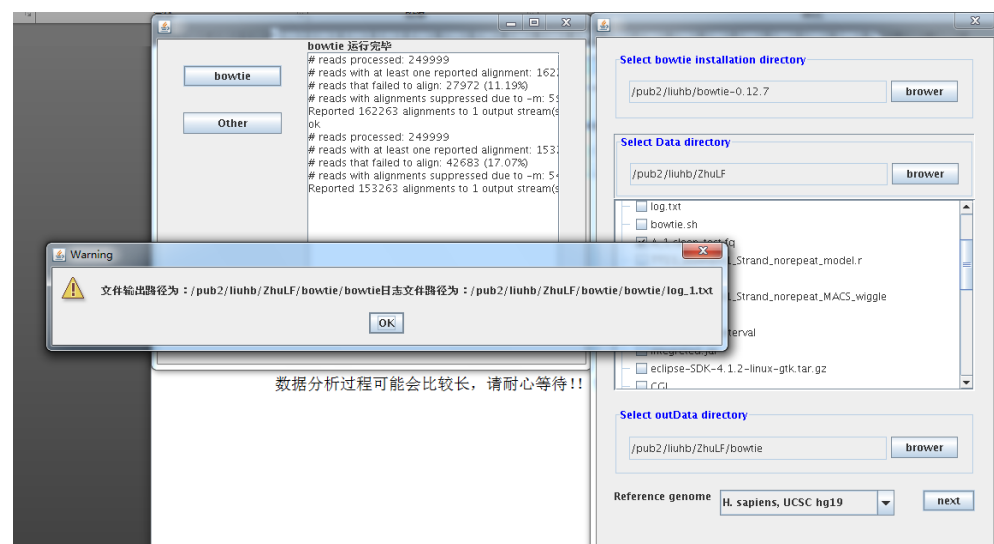
Bowtie 的命令行输出会在左侧面板的右侧区域显示，可以根据这个数据 MSAM 会自动计算 Bowtie 比对的统计表格，该结果会输出在屏幕以及日志文件中。



Bowtie 的输出文件在提示框指定的地址中。

点击 OK 后程序会自动处理下一个 fq 文件。

数据分析过程可能会比较长，请耐心等待！！切勿关闭程序！！

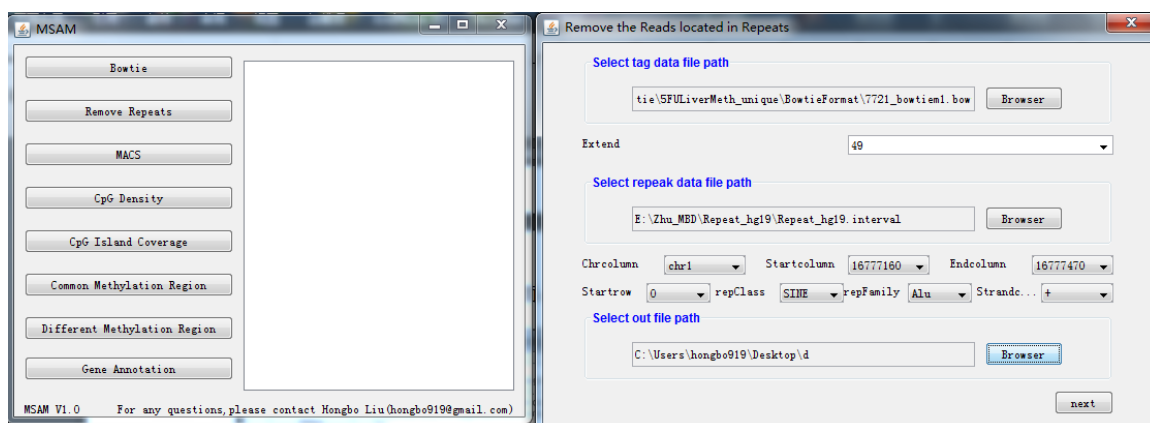


点击 OK，如果已经处理完所有 fq 文件，会出现下面的提示



如有要继续去除 repeat 相关的 reads，点击 Yes，如果不，点击 No。

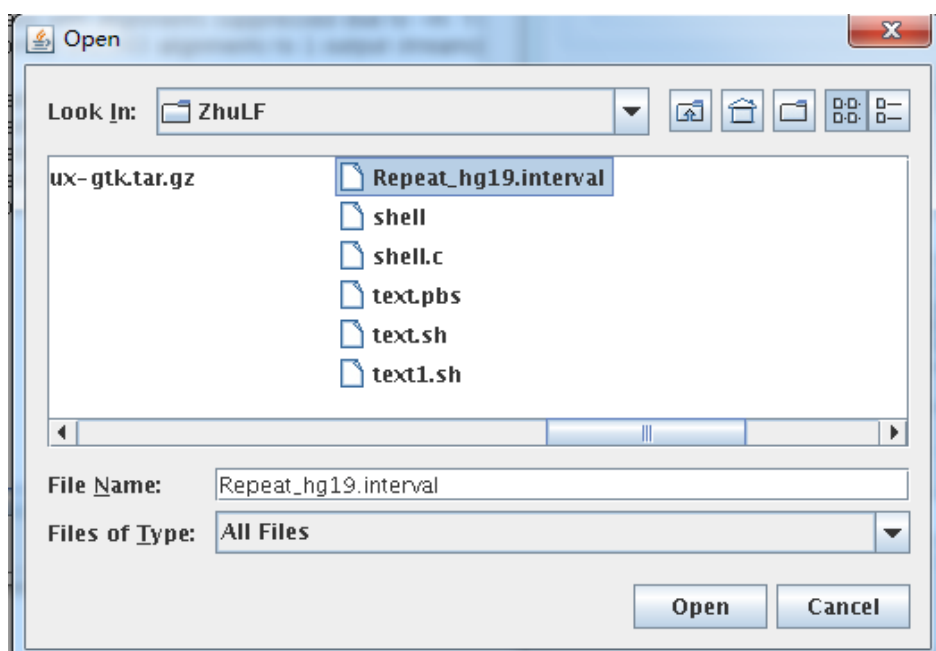
点击 Yes 后，出现去 Repeat 的界面：



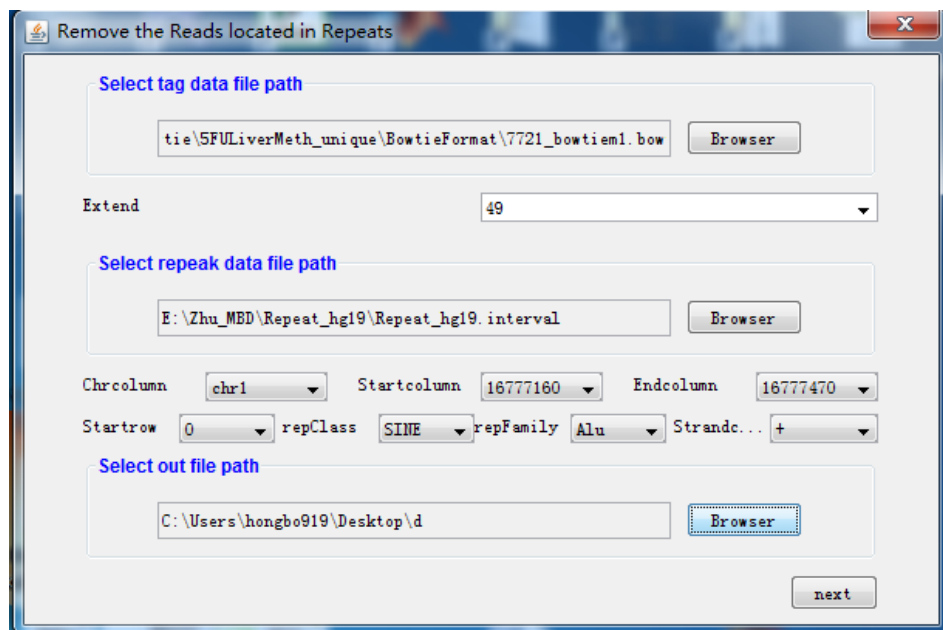
选择要去除 Repeat 的文件（即刚刚 Bowtie 的结果），程序会自动定向到刚刚的结果文件夹，选择其中一个以 bowtie 结尾的文件即可。

可以通过 Extend 来确定扩展的长度，多设置了几个参数，默认为 MBD-seq 片段长度 49。同时支持手写输入，如 50，51，以防没有选择项的长度出现。

接下来选择 repeat 的文件，注意下面的 Files of Type: All Files 如此设置，选择 Repeat\_hg19.interval.



打开后选择各列的属性：



特别注意的是：由于用户提供的 Repeat 文件可能多样化，这里设计的比较自由，需要用户自己选择列属性。

- 1、设置好 Repeat 的位置列 chrome、start、end；
- 2、如果文件中有一行列标题，请在 Stratrow 这里选择 1，如果没有默认为 0 即可；
- 3、Repeatc... SINE Alu 这两列是为了对 Repeat 进行分类用的，切记选对 repClass 和 repFamily 两列（定义见 Schema for RepeatMasker - Repeating Elements by RepeatMasker in UCSC）；
- 4、最后记得选择 strand 列，因为我们在进行这个工作的时候是考虑正负链的。

点击 Next 开始进行去 Repeat 相关的 Reads 工作！！

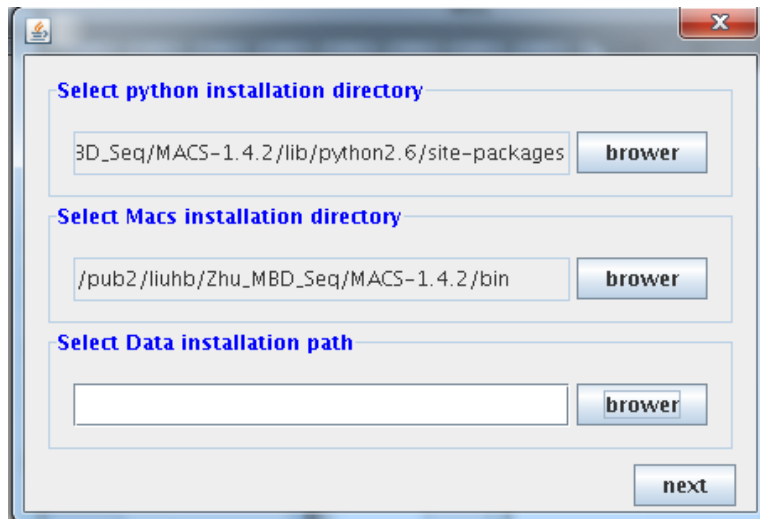
该过程比较耗时，请耐心等待！！

第一个结束后请更改输入文件，进行第二个文件的去 Repeat！

均结束后，会提示：



点击 Yes 继续进行 MACS 峰值探测：



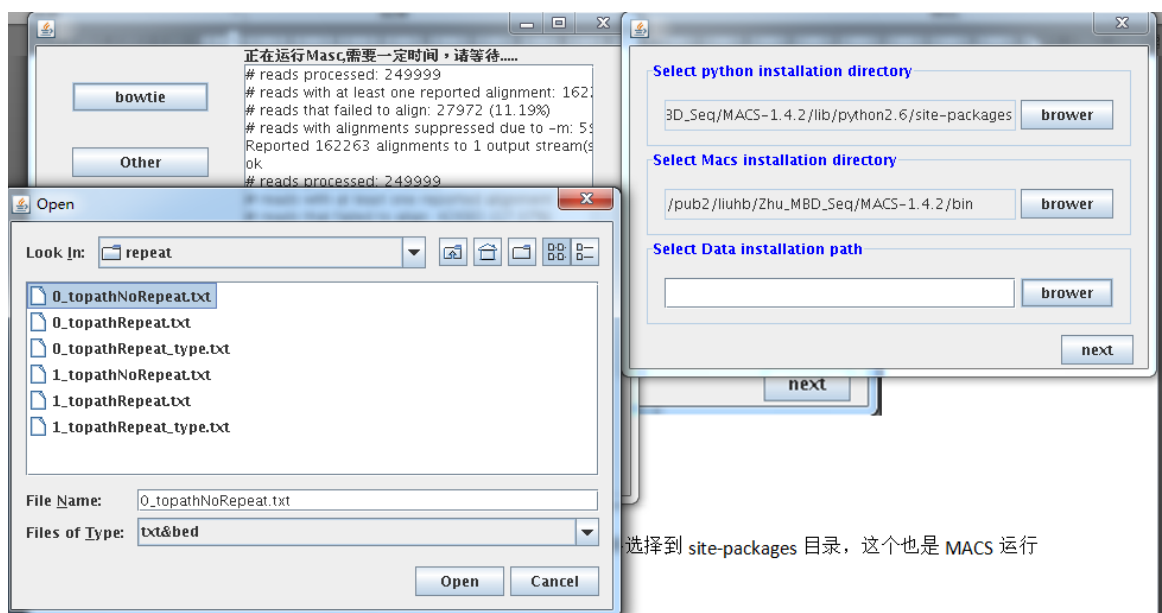
首先如上图设置 python 和 MACS 的路径。

这里要注意：不知怎么设，以下是我的 which 查询的结果

- 1、python 的地址务必按照上述地址安装，并选择到 site-packages 目录，这个也是 MACS 运行时所必须的。
- 2、MACS 的安装路径务必选择到 bin 目录。

注意：此处参数的设置可能会因各自 MACS 安装的不同而不同，但由于目前开发者对 MACS 的运行方式不是很明确，如果此步无法自动化处理，请用您本机 MACS 进行 peak calling，然后再用 MSAM 进行下一步的分析。

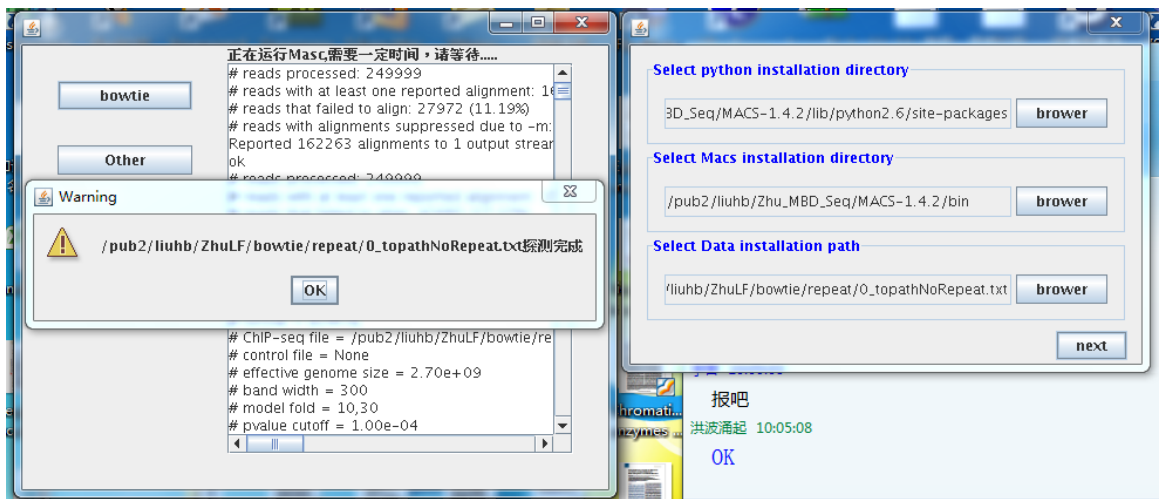
接下来输入要进行峰值探测的文件，及上一步的结果文件夹中的 Norepeat 的文件。



选择打开后，点击 next 开始 MACS 分析！

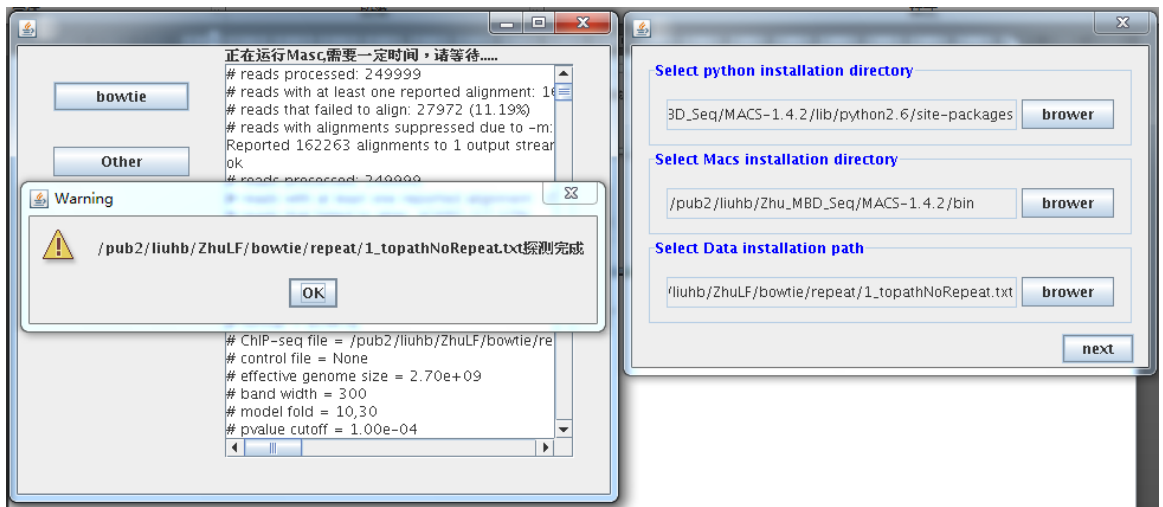
该过程比较耗时，请耐心等待！

第一个文件 MACS 完成后提示：

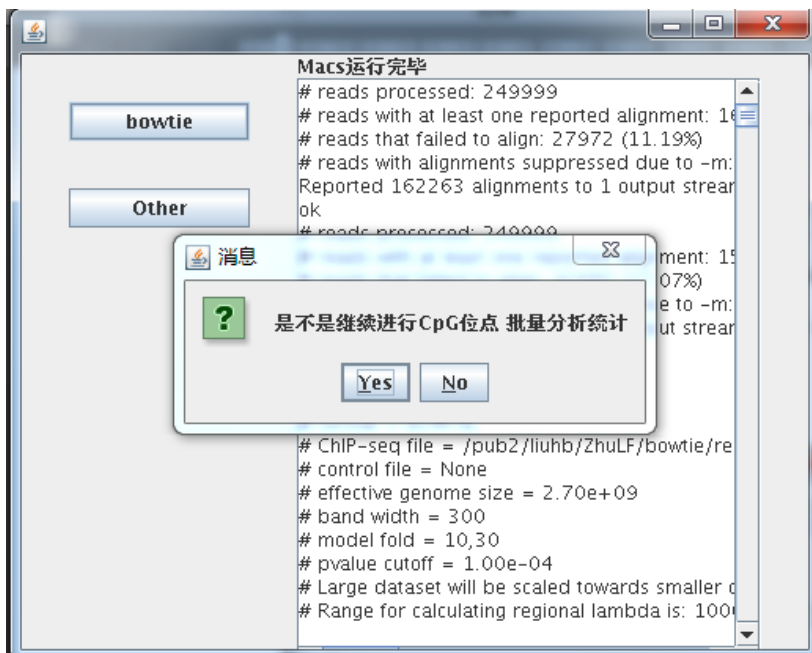


点击 OK 选择第二个文件，并进行 MACS 工作

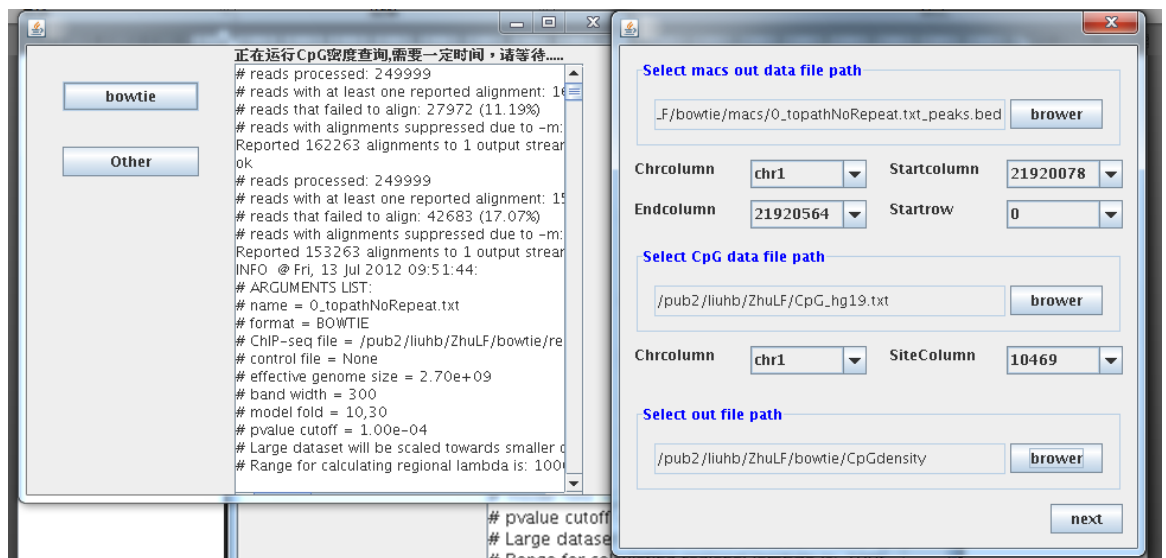
该过程比较耗时，请耐心等待！



完成所有 MACS 工作后，点击 OK，会提示是否进行 peak 内 CpG Density 的计算。



如下图设置



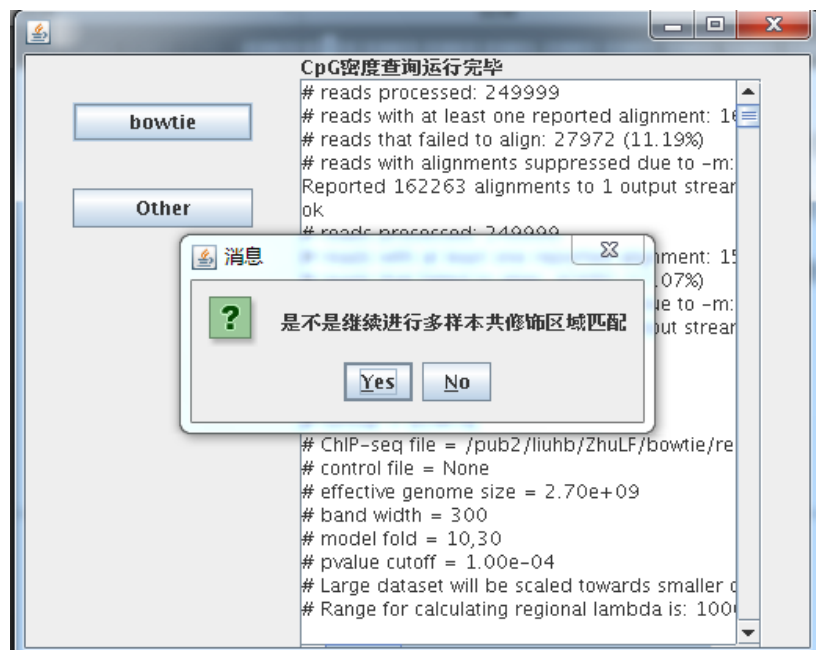
注意：

- 1、请选择 bed 格式 peaks 文件，并选择各列信息。
- 2、选择正确的 CpG 位点数据，不同的物种请更换之。(这个数据是要自己计算，还是可以下载？)
- 3、输出目录可能会默认到 CpGdensity 下，请返回上一层选择该文件夹

点击 next 进行 CpGdensity 计算

该过程比较耗时。请耐心等待！！

计算完后，会提示是否进行多样本间共甲基化 peak 的识别。



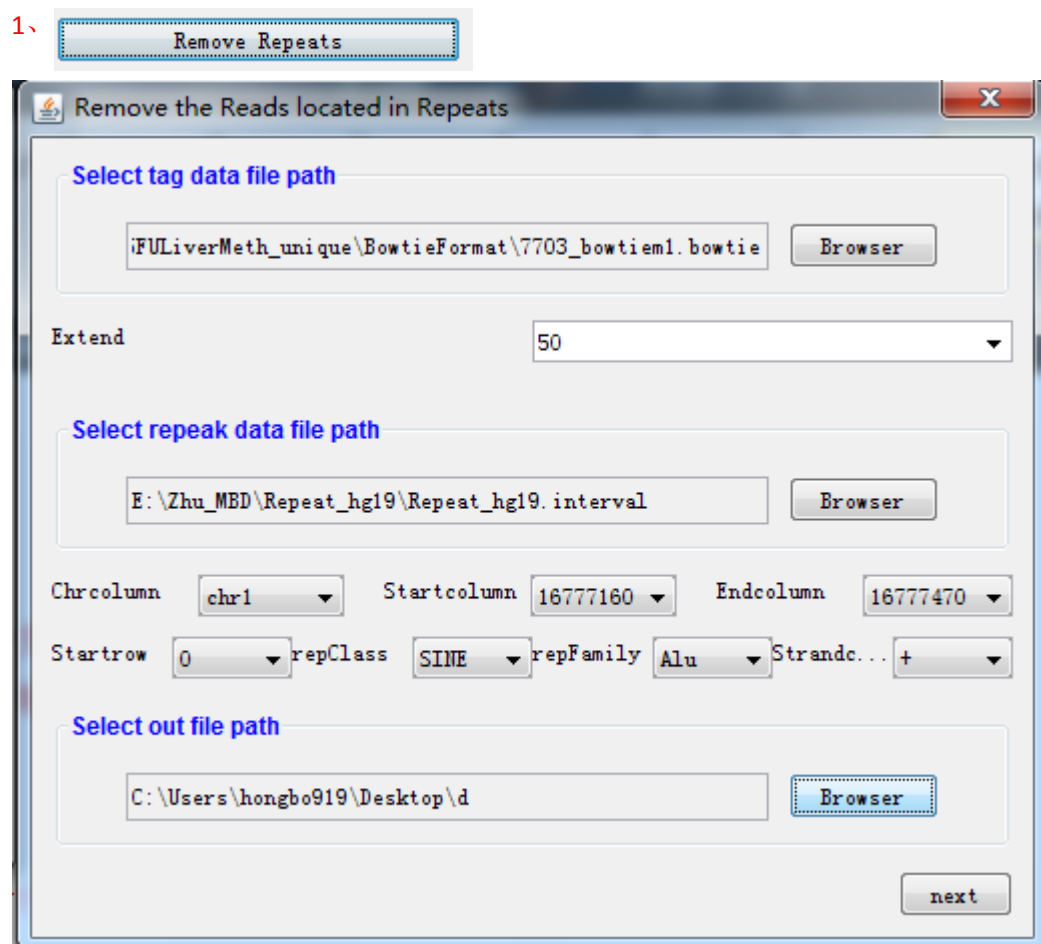
点击 yes 会程序会在后台自动处理。结果会存储在 command 文件夹下。



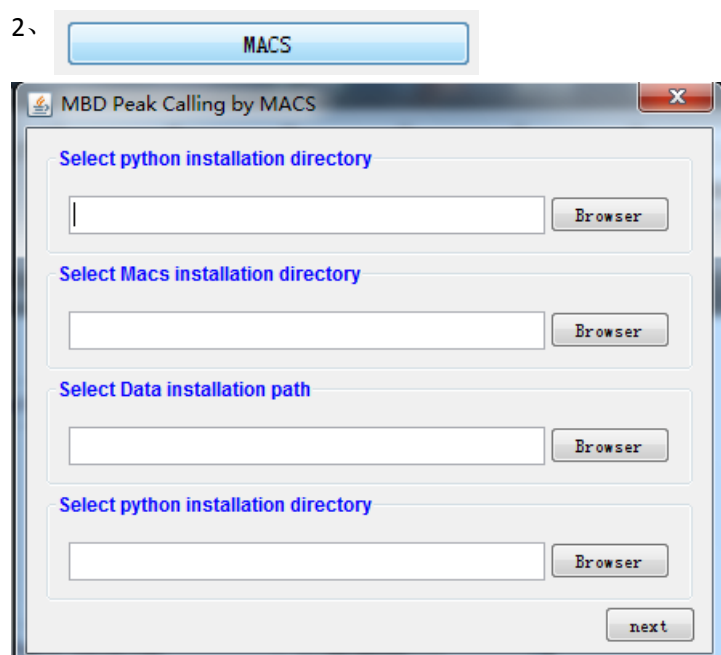
该过程并不耗时。完成后会提示：是否进行多样本间差异的筛选，点击 Yes 即可。结果会存储在 different 文件夹下。

特别注意：目前 1.0 版本的 MSAM 软件支持分模块处理数据。如果使用过程中某步有问题，可以利用其他程序完成工作后，继续利用本软件中提供的各个功能按钮完成相应功能。

1、



2、



- 3、 此处只支持 MACS 得到的 peak.bed 格式的文件

CpG Density

Compute CpG Density of MBD Peaks

Select macs out data file path

adspeaks\7703\_bowtiem1\_Strand\_norepeat\_peaks.bed Browser

Chrcolumn chr1 Startcolumn 564332

Endcolumn 570442 Startrow 0

Select CpG data file path

C:\Users\hongbo919\Desktop\CpGsite\_hg19.txt Browser

Chrcolumn chr1 SiteColumn 526389

Select out file path

C:\Users\hongbo919\Desktop\peak Browser

next

- 4、 此处只支持 MACS 得到的 peak.bed 格式的文件

CpG Island Coverage

Compute CpG Island coverage of MBD Peaks

Select macs out data file path

adspeaks\7703\_bowtiem1\_Strand\_norepeat\_peaks.bed Browser

Chrcolumn chr1 Startcolumn 564332

Endcolumn 570442 Startrow 0

Select CpG data file path

E:\Zhu\_MBD\CGI\CGIs\_hg19\_UCSC.txt Browser

Chrcolumn chr1 Startcolumn 28735

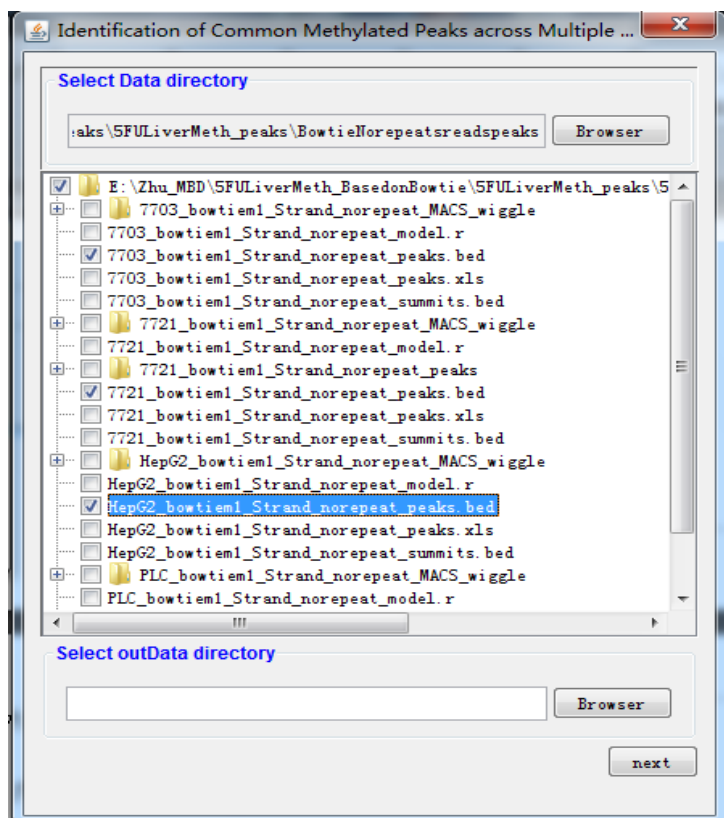
Endcolumn 29810 Startrow 0

Select out file path

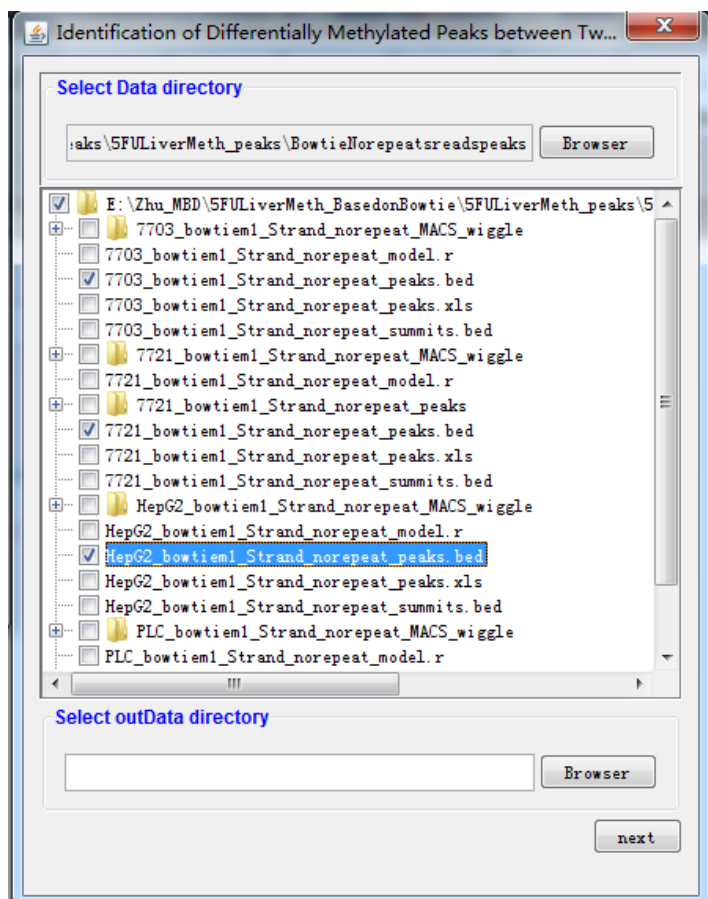
C:\Users\hongbo919\Desktop\CGI Browser

next

- 5、 **Common Methylation Region** 为了实现三个及其以上样本间 common 区域的筛选，目前只支持输入 MACS 输出的结果文件夹，内有要比较分析的 peak.bed 文件。并在下面的多选框中选择要分析的 peak.bed 文件



- 6、 **Different Methylation Region**



7、 Gene Annotation 此处只支持UCSC提供的Refseq基因列表

Refseq基因下载地

址: [http://genome.ucsc.edu/cgi-bin/hgTables?hgsid=287016677&clade=mammal&org=Mouse&db=mm9&hgta\\_group=genes&hgta\\_track=refGene&hgta\\_table=0&hgta\\_regionType=genome&position=chr12%3A57795963-57815592&hgta\\_outputType=primaryTable&hgta\\_outFileName=Refseq\\_hg19](http://genome.ucsc.edu/cgi-bin/hgTables?hgsid=287016677&clade=mammal&org=Mouse&db=mm9&hgta_group=genes&hgta_track=refGene&hgta_table=0&hgta_regionType=genome&position=chr12%3A57795963-57815592&hgta_outputType=primaryTable&hgta_outFileName=Refseq_hg19)

Gene Annotation of the interesting of peaks

Select macs out data file path

itsreadspeaks\7703\_bowtiem1\_Strand\_norepeat\_peaks.bed Browser

Chrcolumn chr1 Startcolumn 564332

Endcolumn 570442 Startrow 0

Select Gene data file path

C:\Users\hongbo919\Desktop\Refseq\_hg19.txt Browser

Chrom chr1 txStart 13... txEnd 13... Startrow 1

cdsStart 13... cdsEnd 13... Strand + GeneName NM...

exonStarts 134212701, 13... exonEnds 134213049, 13...

Select out file path

C:\Users\hongbo919\Desktop\Gene Browser

next

通过该模块可以对 MACS 得到的 peaks 进行注释，也可以对任意的有基因组位置的区域进行基因组注释。

如有不尽事宜，请联系刘洪波（[hongbo919@163.com](mailto:hongbo919@163.com)）。谢谢！