

# 数据完整性检验

为了确保下载的文件与我们释放的一致，在数据下载到本地之后，需要做一个完整性检验，确保下载到本地的文件与我们提供的是完全一致的。

人们通常采用MD5校验和来判断两个文件是否完全一致，因为两个文件略有差异就会导致它们的MD5校验和存在非常巨大的不同。

下载数据之后不要挪动这两个文件的位置，保持与其它文件的相对位置不变。

操作系统不同使用方法也不同：

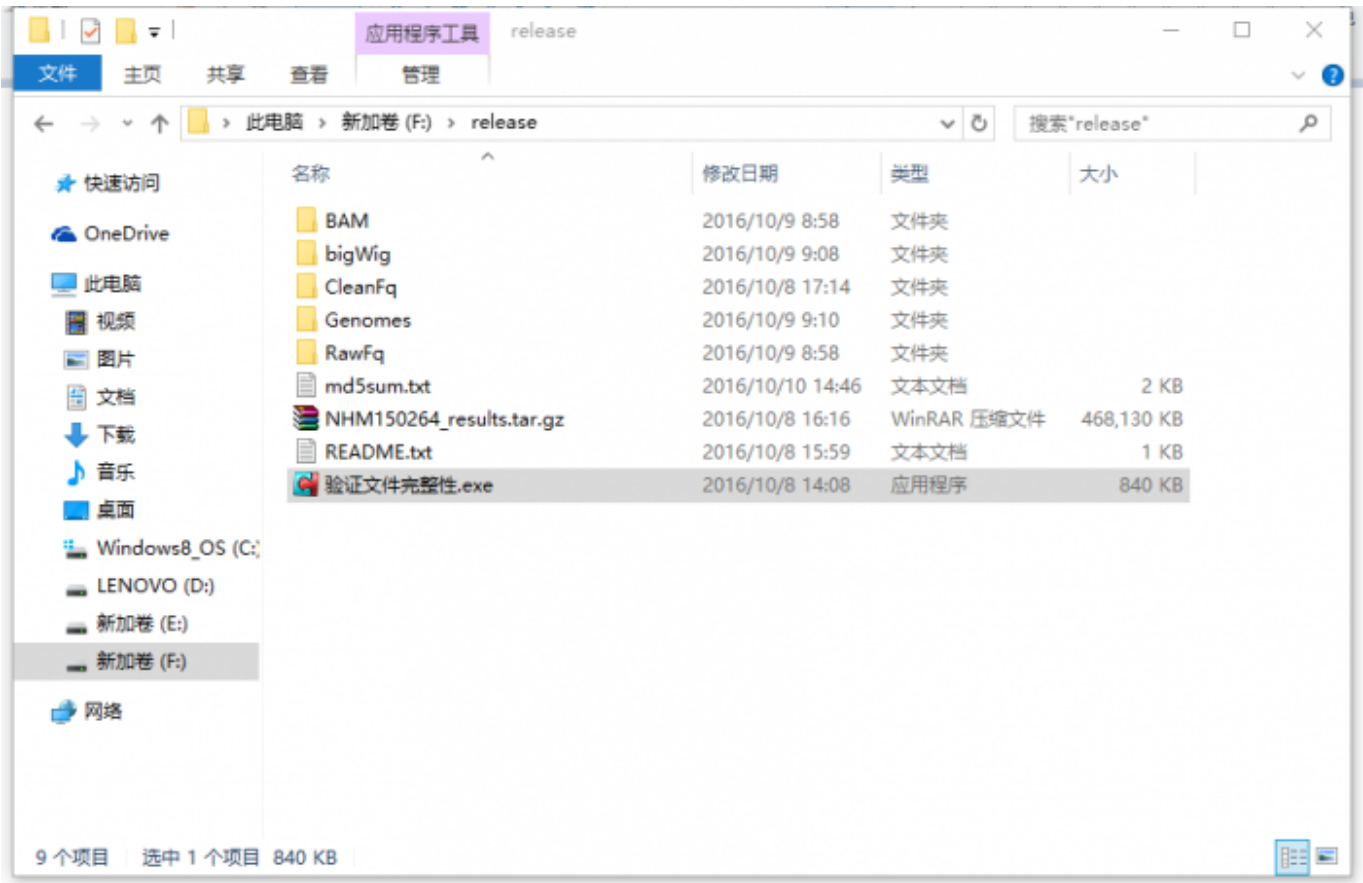
## Windows用户

这里我们提供需要校验的文件的MD5校验值文件 (md5sum.txt文件) 以及校验程序md5summer (验证文件完整性.exe文件)。

检验步骤如下：

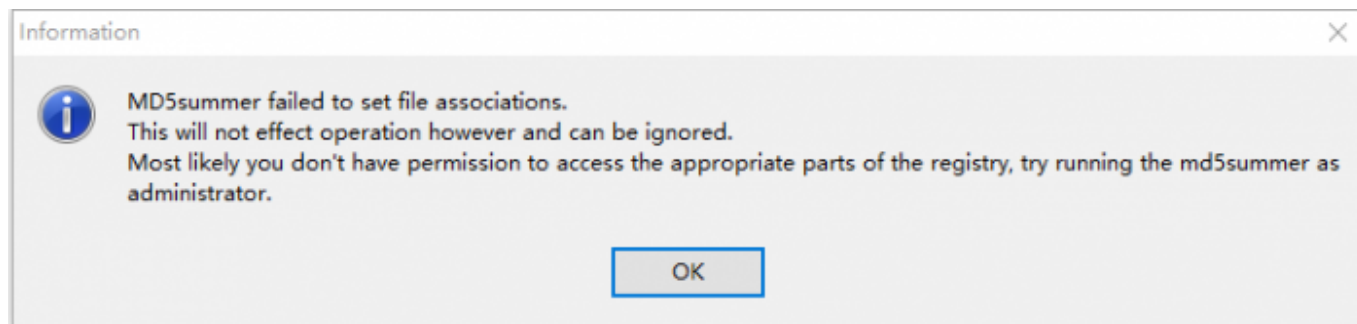
### 第一步: 打开数据文件

打开释放数据的文件夹, 不同产品的数据目录结构和文件夹名称可能不同，这不影响检验。 示例如下：

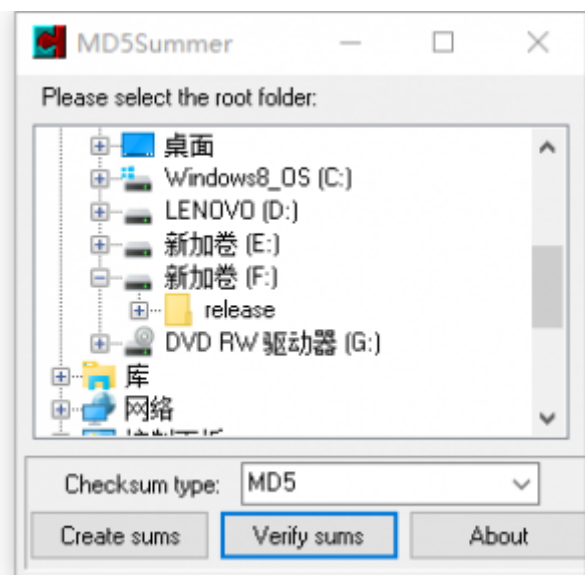


第二步：左键双击“验证文件完整性.exe”

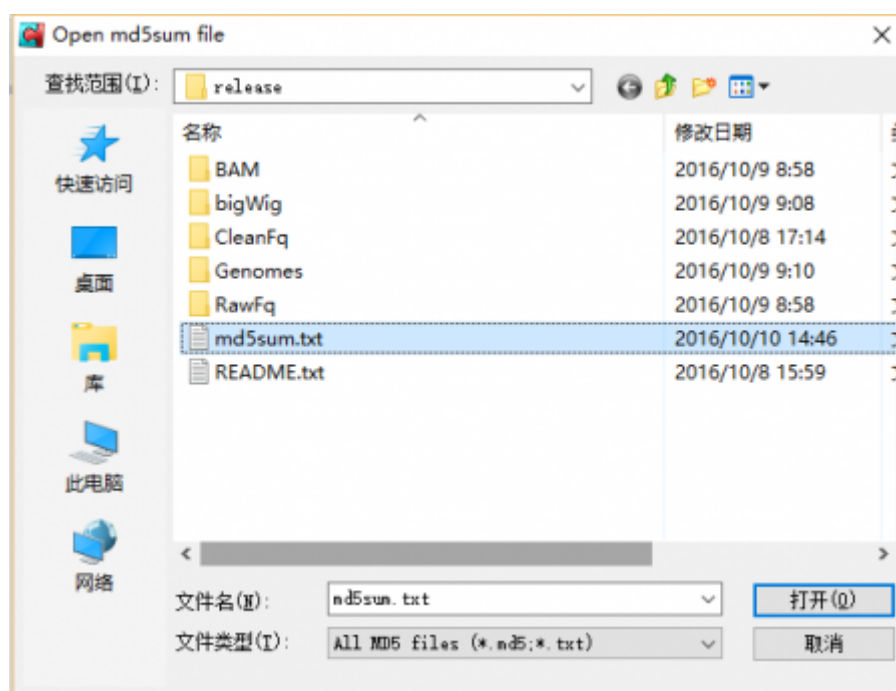
如果出现如下对话框，点击ok即可，不影响使用。



第三步步: 左键点击**Verify sums**

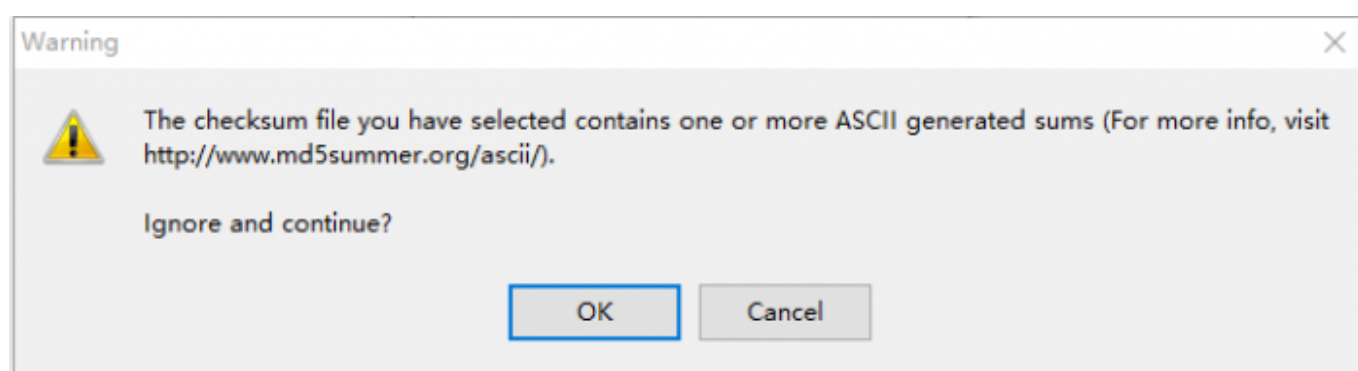


第四步: 左键双击**md5sum.txt**文件

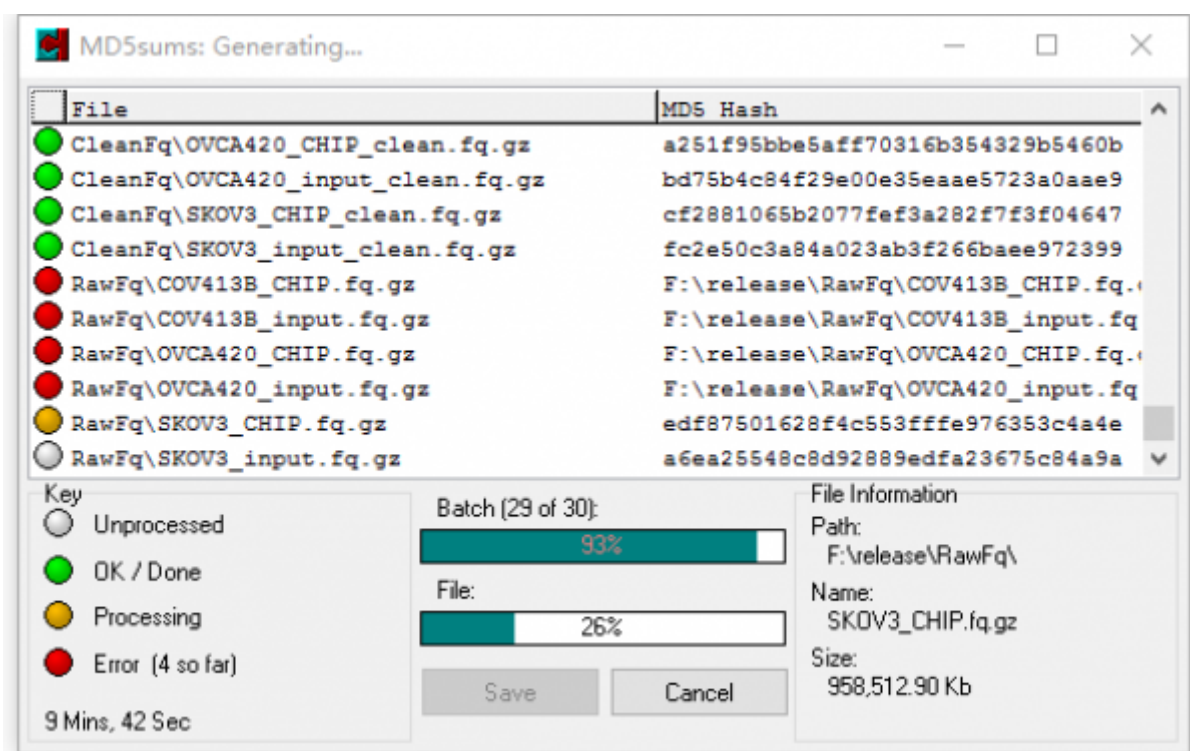


### 第五步：执行校验操作

如果出现此对话框，点击ok 即可：



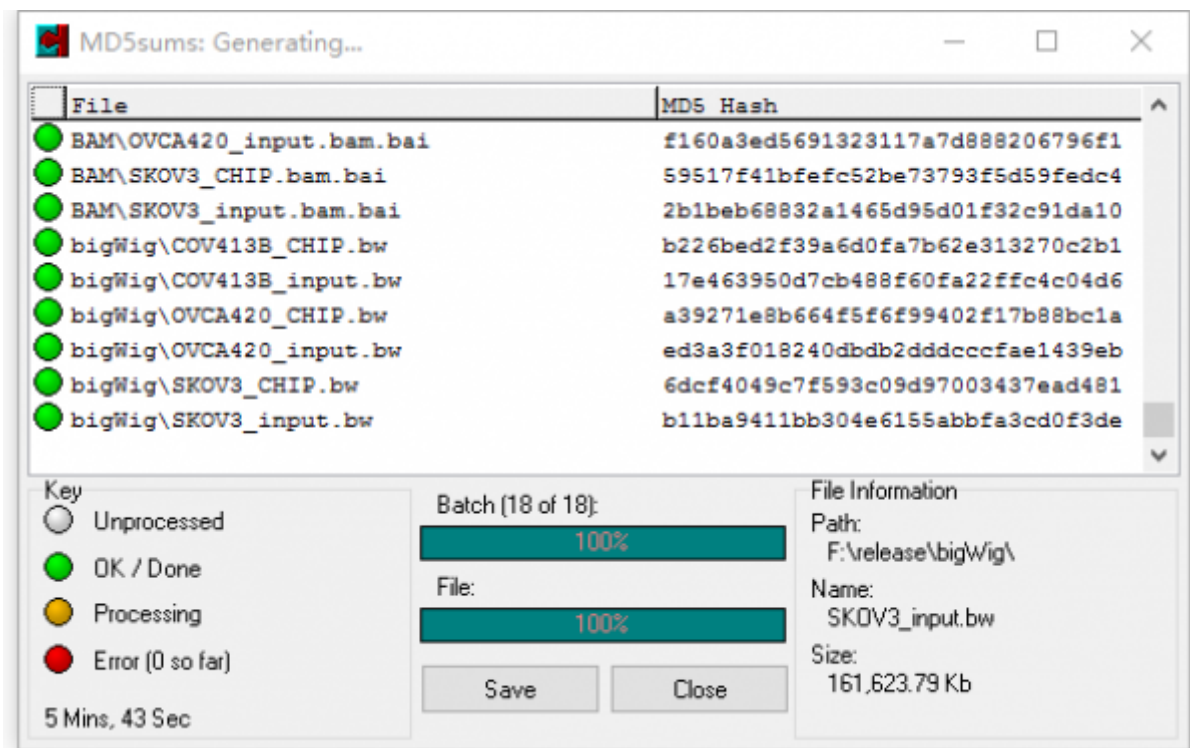
这时候需要耐心等待，所需要的时间和文件大小、所使用的计算机性能等有关系。如下图所示，文件名前边的Key的颜色表示所处状态，灰色表示等待处理，棕色表示处理中，绿色表示校验通过，而红色则表示没有通过校验。



## 第六步：校验结果检查

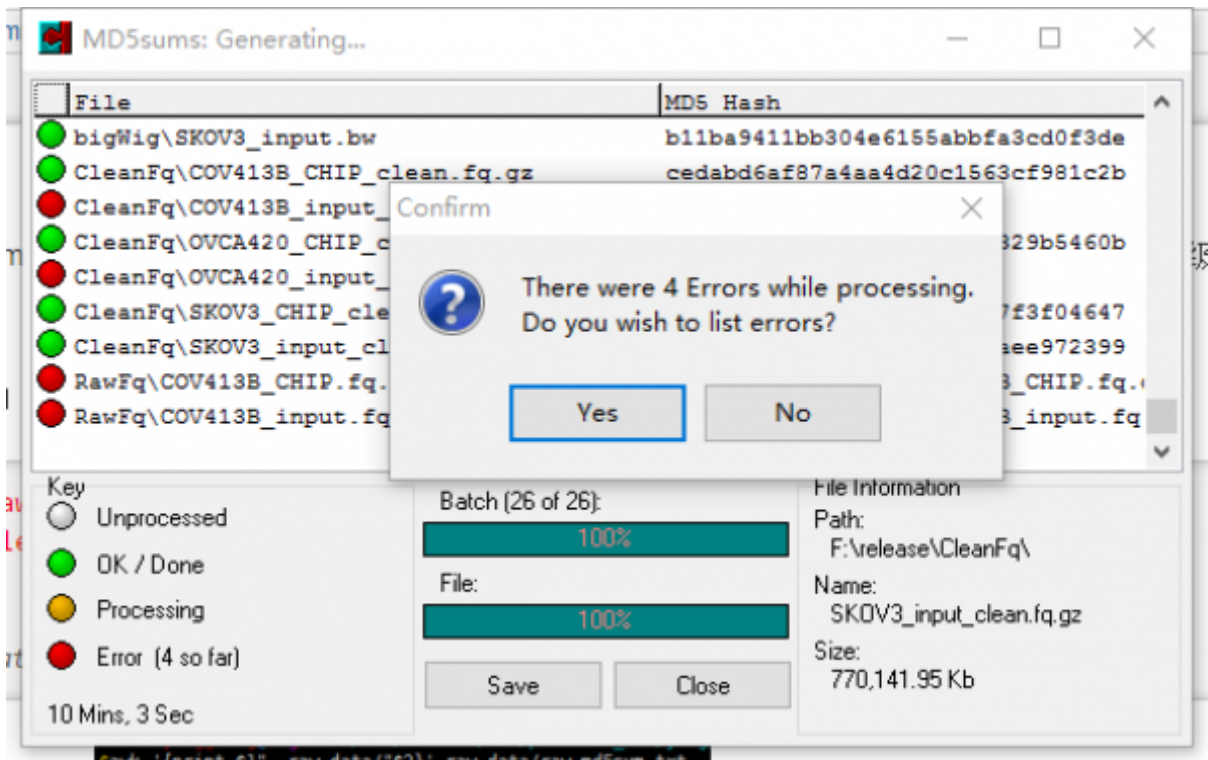
### 1、匹配结果示例

绿色表示文件MD5校验匹配，下载到本地的文件与我们提供的完全一致，下载过程未出现错误。



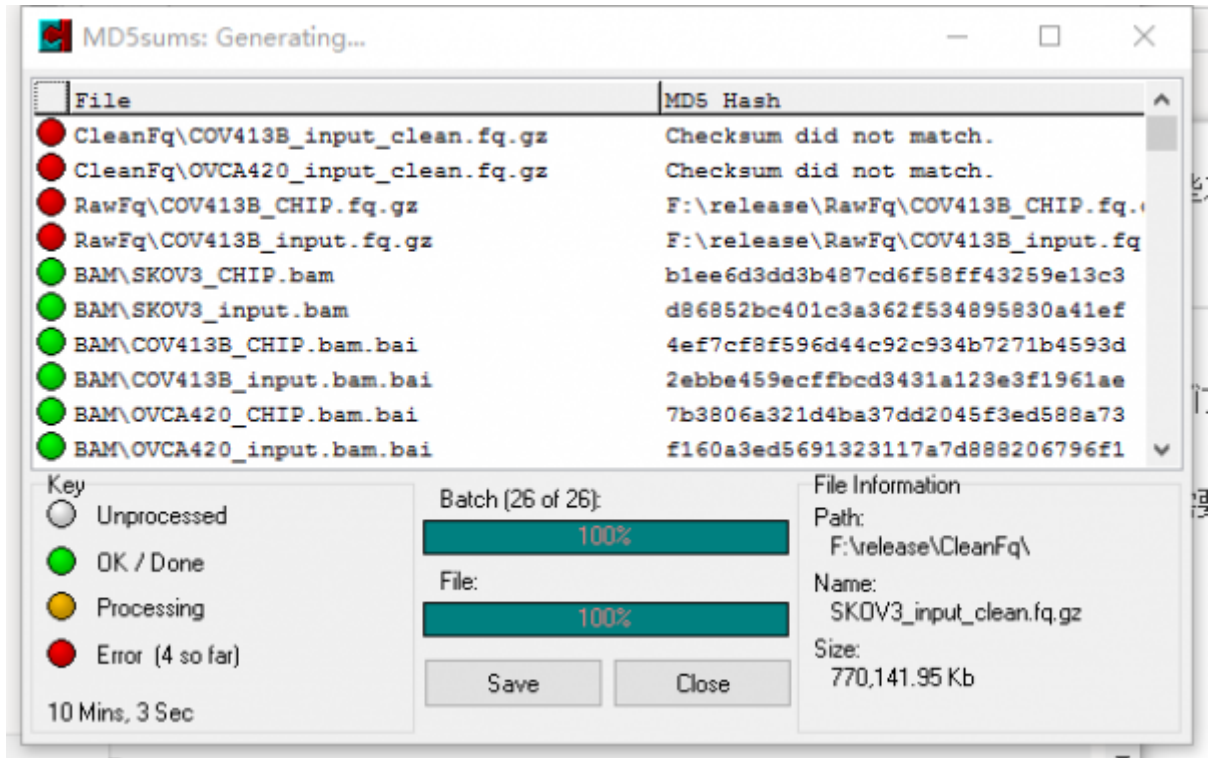
### 2、不匹配结果示例

红色表示文件MD5校验出错。



在校验结束后有提示是否将错误文件列出，若点击Yes则所有的不匹配文件会置顶。

错误信息若为“Checksum did not match.”则表示此文件与我们提供的不完全一致，可能是下载出错导致的；若错误信息为文件路径，例如下图中“F:\release\RawFq\COV413B\_CHIP.fq.gz”则表示在此路径下未找到该文件，此文件没有下载到本地。



如果出现上面的情况，则这些出错的文件需要重新下载，下载之后再进行验证。这次只需要验证重新下载的文件，故可以将md5sum.txt文件中的未通过验证的文件拷贝到新建文件中，然后用此新文件重新进行前面的步骤，也可以将md5sum.txt文件中的已通过验证的文件删掉，然后重复操作前面步骤。

如果再三下载之后都不能通过校验，请及时与我公司项目相关运营人员联系。

## Linux用户

### 第一步：进入释放数据目录

```
$cd data_release/
```

### 第二部 md5sum -c 命令验证

```
md5sum -c md5sum.txt
```

```
$md5sum -c md5sum.txt
```

### 第三步：等待验证完成

这时候需要耐心等待，所需要的时间和文件大小、所使用的计算机性能等有关系。如下图所示“OK”表示验证通过“FAILED”表示验证失败，下载文件不完整“No such file or directory”和“FAILED open or read”表示此文件没有被下载到本地或不在此路径下。如果最下面出现WARNING信息则表示有文件没通过验证，则这些出错的文件需要重新下载，下载之后再进行验证。这次只需要验证重新下载的文件，故可以将md5sum.txt文件中的未通过验证的文件拷贝到新建文件中，然后用此新文件重新进行前面的步骤，也可以将md5sum.txt文件中的已通过验证的文件删掉，然后重复操作前面步骤。

```
$md5sum -c md5sum.txt
rawdata/A11879JD_1.fq.gz: OK
rawdata/A11879JD_2.fq.gz: FAILED
cleandata/A11879JD_1.clean.fq.gz: OK
cleandata/A11879JD_2.clean.fq.gz: OK
IGV_data/bam/A11879JD.bam: FAILED
P2016070744-B1-4_Homo_sapiens_lncRNA_report.tar.gz: OK
P2016070744-B1-4_Homo_sapiens_lncRNA_result.tar.gz: OK
IGV_data/genome.fasta.gz: OK
IGV_data/genome.gtf.gz: OK
md5sum: WARNING: 2 of 9 computed checksums did NOT match
```

如图，若所有文件均为“OK”则认为本次验证成功。



```
$md5sum -c md5sum.txt.bak
rawdata/A11879JD_1.fq.gz: OK
rawdata/A11879JD_2.fq.gz: OK
cleandata/A11879JD_1.clean.fq.gz: OK
cleandata/A11879JD_2.clean.fq.gz: OK
IGV_data/bam/A11879JD.bam: OK
P2016070744-B1-4_Homo_sapiens_lncRNA_report.tar.gz: OK
P2016070744-B1-4_Homo_sapiens_lncRNA_result.tar.gz: OK
IGV_data/genome.fasta.gz: OK
IGV_data/genome.gtf.gz: OK
```

如果再三下载之后都不能通过校验，请及时与我公司项目相关运营人员联系。

注：截图系统内核版本为Red Hat 4.4.7-11

## Mac用户

### 第一步：打开终端

首先打开Launchpad像火箭一样的图标，然后找到“实用工具”，最后找到终端并且打开。

### 第二步：校验文件MD5值

首先在终端中输入md5然后再输入一个空格，最后将要验证的文件直接用鼠标拖到终端里，然后按回车键，等待（视文件大小和电脑性能而定），在下面会出现计算得到的md5值，将这个值与我们提供的md5值对照。若全部相同，则验证通过；若有不一样的，表示该下载文件与我们提供的不一致，请重新下载错误文件，并再次进行验证。若多次验证均不能通过，请及时与我公司项目相关运营人员联系。