# perl基础考核

## **考核要求**

1. 可以查阅书籍，不可以上网查询；

2. 考核时间4小时；

3. 需要提交结果及使用的命令或脚本；

4. 如果没有思路，可以提示，但不超过3次，且提示会扣分（2分/次），涉及到题目意思的问题不限次数且不扣分。

5. 70分为及格。

## 考核题目：

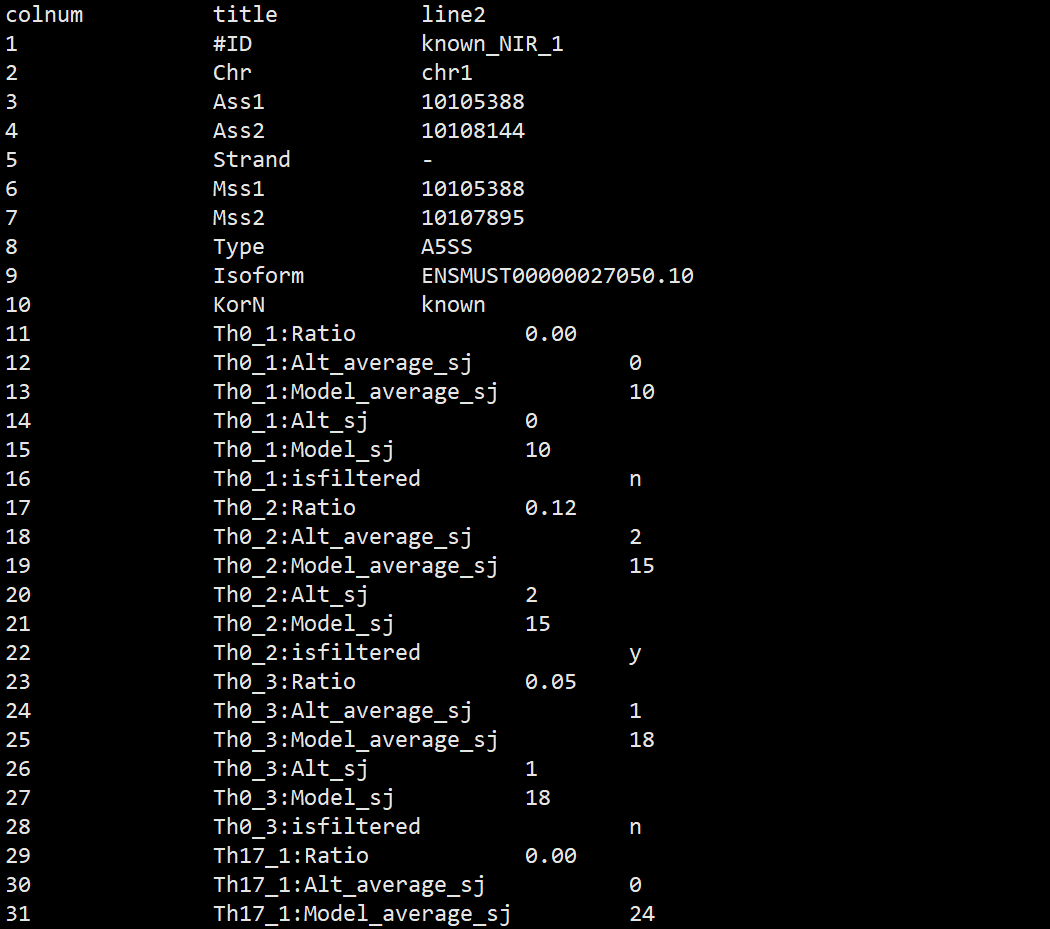
### 写一个脚本，计算n的阶乘。（5分）

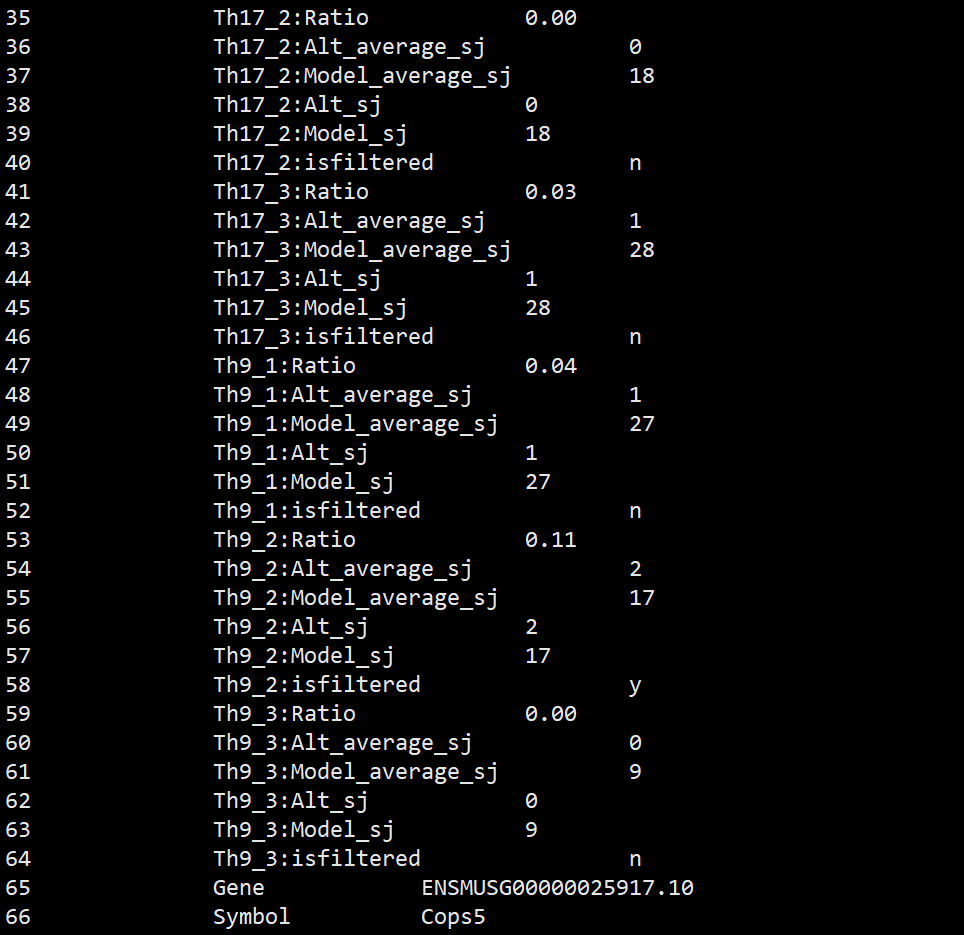
要求根据输入的数，判断是否合法（只能输入自然数），不合法则输出提示信息，合法则输出结果；

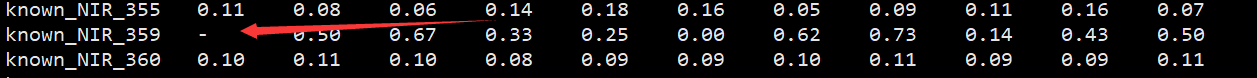
### 写一个脚本，根据基因ID列表（数目100），在表达量总表中提取出各样本中此基因的表达量。（10分）

### 写一个命令行根据基因Symbol列表（数目100），在表达量总表中提取出各样本中此基因的表达量，把Symbol这一列放第一列，同时去掉Chromosome Genetype Description后3列。（10分）

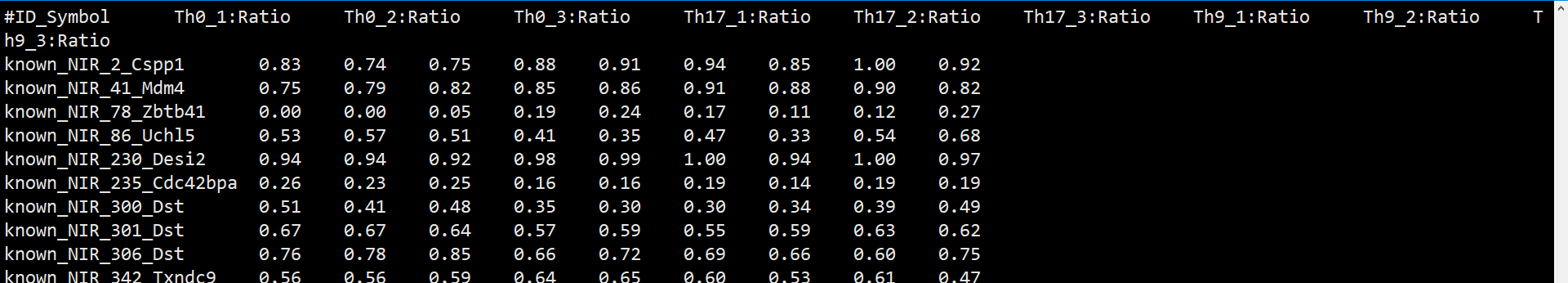
### 信息提取：写一个命令行（15分）：





在附件文件中根据第一列ID提取每个样本的ratio同时去掉含有-的行

格式如下：注意第一列ID+Symbol

写一个脚本（15分）：

上述有了结果后，Th0样本的ratio的平均值和Th17样本的ratio相减的差值添加到最后一列输出到一个文件，再根据此文件再输出差值>=0.1的输出一个文件

### 命令行或脚本：将fastq文件中的2端reads序列反向并且互补，质量序列反向，最终结果输出到新文件中；（5分）

fastq文件格式

· 第一行：记录测序时，序列的坐标信息

· 第二行：测序得到的序列信息（分ATCGN5种情况，N代表不确定碱基类型）

· 第三行：以“+”开始，可以储存一些附加信息

· 第四行：以ASCII码形式储存碱基的质量信息（与第二行的碱基序列一一对应）

只需要关注第二行和第四行

### 写一个脚本：统计上题反向互补后的fastq中reads的各碱基（ATCG）及N含量，将统计结果输出到新文件中；（10分）

统计A有多少，T有多少，C有多少，G有多少

最终输出的文件格式：

#base Number

A 1

T 1

C 1

G 1

N 1

### 写一个脚本：把基因组gff文件转换成bed文件（5分）

Human的gff文件：

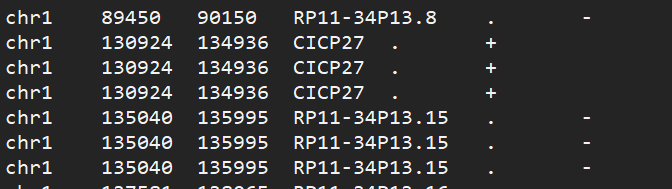
gff文件说明：



Bed的文件说明：

<https://www.jianshu.com/p/9208c3b89e44>

转换后的bed文件：



### 写一个脚本(5分)：

这是差异表达量deg\_genes\_fpkm\_of\_all\_samples\_in\_deg.txt，筛选出90%以上的样本的表达量>=1基因数目，90%以上的样本表达量>=0.5的基因数目，90%以上的样本表达量>=0.2的基因数目

最后输出文件格式参考：

*#90%DElncRNA\_filter\_parameter Number*

0.2 55  18

0.5 23  12

1   15  8

### 统计\*Sig\_DEG\_cell\_type\_c0.5\_p0.01.txt 每组中的每个细胞的基因数量（10分）

输出格式：

Type MP\_vs\_MA MA\_vs\_YC SP\_vs\_YC SP\_vs\_MP SP\_vs\_MA MP\_vs\_YC

Monocytes 13 1 7 0 54 1

### 10.根据通路基因：goterm\_symbol.txt，找到每个细胞上对应的基因的相关性系数DEG\_cor\_cell\_type\_c0.5\_p0.01\_ttest\_0.05.txt(第三列是相关性系数)（10分）

**输出格式：**

Term Gene T cells follicular helper Monocytes Macrophages M2 Macrophages M0

cell differentiation Caprin1 0 0 0 0.5299146664032

cell differentiation Cav1 0 0 0 0.576299507156627

cell differentiation Fgf16 0.505747027221823 0 0 -0.595172036989326

cell differentiation Fgf23 0 0 0 -0.505423671903459

cell differentiation Fli1 0 0 0 -0.527742041559994

cell differentiation Mdfi 0 0 0 -0.547605635286485

cell differentiation Notch3 0 0 0 -0.50679258216516

cell differentiation Nsun2 0 0 0 0.567042435721529

cell differentiation Pik3cd 0 -0.590958292053976 0 -0.505780092789869

cell differentiation Sh2b3 0 0 0 -0.551137353267428

cell differentiation Thrb 0 0 0 -0.516377664704311

immune response Ccl12 0.58754591155344 0 0 0

这是第10题，第一列是通路，第二列是基因，接下来是细胞，如果这个基因的细胞有相关性系数，就有值，没有就是0