perl上机练习2

目的:熟悉哈希表的操作

熟悉正则表达式

熟悉基本生物信息数据提取编程

1. 在miRNA调控靶基因的信息文件

/home/shenwk/perl\_practice/test2\_file/2.1Predicted\_Targets\_Info.txt中，找出/home/shenwk/perl\_practice/test2\_file/2.1name.txt文件中存在的基因名的被调控信息，即这些基因被miRNA所调控的信息，并输出到新文件(利用perl的正则表达式)。

1. 运用hash表：

文件 /home/shenwk/perl\_practice/test2\_file/2.2gene\_info中，第一列为基因名，第二列为该基因在疾病与正常样本中的表达量变化倍数取对数，试从中找出

/home/shenwk/perl\_practice/test2\_file/2.2gene\_id中的gene的表达量变化，并将其输出到输出文件，输出文件第一行为基因名，第二行为表达量变化倍数取对数。

1. 使用正则表达式从文件

/home/shenwk/perl\_practice/test2\_file/2.3fred.txt中找出匹配“fred”的行并输出行号，忽略大小写。

修改你的程序，使其只匹配fred，即fred前后不可以有别的字符。

4. 从NCBI上的Nucleotide数据库中查找这个目录下的Fasta格式文件中的基因

/home/shenwk/perl\_practice/test2\_file/2.4CALR.txt，并从中提取出对应的蛋白质编码区域，并以CDS命名该段序列。