**实习四 基因组序列（突变位点）注释实验**

**一、实习目的**

1. 掌握ANNOVAR的基本原理与使用方法，理解其在基因组变异注释中的作用。
2. 学习对21号染色体的突变位点进行功能注释，包括基因影响、人群频率、临床意义等。
3. 培养生物信息学数据分析能力，包括数据预处理、注释结果解读和过滤。
4. 熟悉变异注释在遗传病研究中的应用，为后续科研或临床分析奠定基础。

**二、实习题**

使用ANNOVAR将上一节实验课的突变位点结果（21号染色体）进行注释。

**三、实习内容和要求**

1. **软件安装**：安装ANNOVAR软件。

软件已安装好，路径为：

ANNOVAR: /home/sky/annovar​

2. **下载突变位点数据**：

上节课鉴定的人类21号染色体的突变位点，数据路径在Github上，需要自己下载：<https://github.com/zhangwei2015/Test/> 下面的[GRCh38\_chr21.out.vcf.gz](https://github.com/zhangwei2015/Test/blob/main/GRCh38_chr21.out.vcf.gz" \o "GRCh38_chr21.out.vcf.gz)文件，文件下载后，保存于自己建的目录下。

3. **下载基因注释数据库**:

1. 在自己的目录下建立目录：

mkdir HumDB

1. 下载四个数据库数据到新建的目录HumDB：
   1. RefSeq Gene：perl /home/sky/annovar/annotate\_variation.pl -buildver hg19 -downdb -webfrom annovar refGene HumDB/
   2. dbSNP: perl /home/sky/annovar/annotate\_variation.pl -buildver hg19 -downdb -webfrom annovar avsnp150 HumDB/
   3. ClinVar: perl /home/sky/annovar/annotate\_variation.pl -buildver hg19 -downdb -webfrom annovar clinvar\_20220320 HumDB/
   4. ExAC: perl /home/sky/annovar/annotate\_variation.pl -buildver hg19 -downdb -webfrom annovar exac03 HumDB/

4. **基因突变位点格式转换**: 将21号染色体的基因突变位点数据的格式转换，转换为ANNOVAR软件需要的格式。

perl /home/sky/annovar/convert2annovar.pl -format vcf4 GRCh38\_chr21.out.vcf.gz > input.avinput

**5. 对基因突变位点进行注释**: 基于上述4个数据库，使用ANNOVAR软件对21号染色体的基因突变位点进行注释。

perl /home/sky/annovar/table\_annovar.pl input.avinput HumDB/ -buildver hg19 -out output -remove -protocol refGene,exac03,avsnp150,clinvar\_20220320 -operation g,f,f,f -nastring .

**6. 解读注释结果文件**：解读结果文件output.hg19\_multianno.txt。在文件打开的情况下（less -S 文件名），可以搜索（/区域名称）位于不同区域的突变位点注释情况。

区域名称包括：exonic，intronic，UTR，upstream，downstream，intergenic，ncRNA等

四、**课后需提交作业：**

每一步运行的过程和结果截图，作简要说明，写成word文档提交。