

# 疾病SOP

---

- 1. 流程简介
- 2. 预先准备
  - 2.1. 下载流程
  - 2.2. 部署 nextflow config
- 3. 运行准备
  - 3.1. 运行变异检测流程
  - 3.2. disease.txt
  - 3.3. gene.txt
  - 3.4. sample.ped
  - 3.5. sample.cogped(特殊情况下需要), 格式与sample.ped相同
- 4. 运行流程

## 1. 流程简介

基于老师提供的信息, 对突变位点进行有害性注释。当老师提供符合标准的家系样品时

## 2. 预先准备

### 2.1. 下载流程

如果第一次使用该流程, 请git clone

```
git clone git@git.majorbio.com:long.huang/dna.git
```

如果已有repo, 请使用git pull。务必每次运行前运行git pull, 以确保代码为最新代码。

### 2.2. 部署 nextflow config

```
cd dna/Disease
cp -f nextflow.config.tmpl nextflow.config
```

nextflow.config.tmpl文件根据172环境开发, 如果搬迁到236服务器, 请自行调整对应参数。如果不会调整, 后续询问小汤。

## 3. 运行准备

### 3.1. 运行变异检测流程

具体步骤问zdd

### 3.2. disease.txt

该文件为老师感兴趣的疾病编号, 一行一个。示例如下:

```
[sanger-dev@login-0-10 test]$ cat disease.txt
C1290638
C0266878
C0035851
```

3.3. gene.txt

该文件为老师感兴趣的基因名称，一行一个。示例如下：

```
[sanger-dev@login-0-10 test]$ cat gene.txt
IRF8
FLNA
```

3.4. sample.ped

该文件为样品的家系信息，制表符分隔，根据分析附件填写，具体格式如下（文件本身不需要表头，以下表头仅用于方便理解）：

家系编号	样品编号	父亲编号	母亲编号	性别	是否患病
父亲母亲须和孩子在 同一家系中	与分析时使用的 编号相同	无对应样 品填0	无对应样 品填0	男性为1，女性为 2，未知为0	正常为1，患 病为2

示例如下：

```
[sanger-dev@login-0-10 test]$ cat sample.ped
fam1    mother  0      0      2      1
fam1    father  0      0      1      1
fam1    son     father  mother  1      2
```

3.5. sample.cogped(特殊情况下需要)，格式与sample.ped相同

当用户提供的家系信息不是严格的trios家系(父、母、患者)时，应当进行处理。如果用户提供的家系信息包含trios，则只保留trios样品对应的记录。如果用户提供的家系不包含trios，那么可以使用touch命令创建一个空文件。这样可以跳过新生突变检测，而保持共分离分析。

4. 运行流程

运行代码示例：

```
nextflow run /mnt/lustre/users/sanger-dev/sg-
users/yiwei.tang/offline_src/disease/human_disease.nf --wgsresult
/mnt/ilustre/isanger_workspaceWgsV4/20230721/WgsV4_cdlr_jmrnob8tfh71krrscvbcjv/out
put/ --cog --ped sample.ped --outdir disease_result -bg -resume --disease
disease.txt --report
```

命令参数说明:

```
Usage:
  The typical command for running the pipeline is as follows:
  nextflow run human_disease.nf --outdir hd_result --snpindel pop.vcf.gz --ped
  sample.ped

  --db          <file>  注释数据库所在文件夹，默认不写
  --wgsresult   <file>  wgs output dir [required]
  --vcf         <file>  vcf相对wgs_result的相对路径，自定义vcf时才修改该参数
  --ped         <file>  样品信息表 [required]
  --disease     <file>  目标疾病名，一行一个
  --outdir      <dir>   输出文件夹
  --norename    <bool>  不需要重命名染色体
  --wgs         <bool>  是否wgs测序，默认wes
  --cog         <bool>  是否进行样品遗传模式分析
  --cogped      <file>  新生突变使用的家系文件
  --gene        <file>  老师感兴趣的基因列表
  --report      <bool>  是否生成报告
  --noasso     <bool>  不运行基因疾病关联分析
  --help       <bool>  打开帮助
```

最终结果一般位于输出文件夹下的publish文件夹，输出文件夹组成如下:

```
disease_result/
├── annotation # 流程结果文件
├── publish    # 生成报告和可以交付的结果目录
└── vcf       # 流程中间文件
```