

## #版本: PGG.Panel

### 1.前期准备

---

基础软件安装:

在开始前, 请先安装以下工具: bcftools, vcftools, plink2, bgzip, king, shapeit5, minimac4

# 安装 miniconda (用户目录)

```
wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
```

```
bash Miniconda3-latest-Linux-x86_64.sh -b -p $HOME/miniconda
```

```
source $HOME/miniconda/bin/activate
```

# 通过 conda 安装

```
conda install -c bioconda bcftools vcftools plink -y
```

```
conda install bioconda::plink2
```

# king 安装

下载链接: <https://www.kingrelatedness.com/executables/Linux-king228.tar.gz>

# shapeit5 安装

# 将以下模块下载至 shapeit5 目录下

```
https://github.com/odelaneau/shapeit5/releases/download/v5.1.1/ligate\_static
```

```
https://github.com/odelaneau/shapeit5/releases/download/v5.1.1/phase\_common\_static
```

```
https://github.com/odelaneau/shapeit5/releases/download/v5.1.1/phase\_rare\_static
```

```
https://github.com/odelaneau/shapeit5/releases/download/v5.1.1/simulate\_static
```

```
https://github.com/odelaneau/shapeit5/releases/download/v5.1.1/switch\_static
```

# 此外将 GLIMPSE2 的以下模块下载至 shapeit5 目录下

```
https://github.com/odelaneau/GLIMPSE/releases/download/v2.0.0/GLIMPSE2\_chunk\_static
```

# 下载所需 map 文件 (需解压)

下载链接: <https://github.com/odelaneau/shapeit4/tree/master/maps>

# minimac4 安装

安装教程见 <https://github.com/statgen/Minimac4>

# 下载所需 map 文件 (GRCh38)

下载链接: [https://share.sph.umich.edu/minimac4/map\\_files/genetic\\_map\\_hg38\\_withX\\_3cols.tsv.gz](https://share.sph.umich.edu/minimac4/map_files/genetic_map_hg38_withX_3cols.tsv.gz)

## 2. 正式运算

---

### 二、流程步骤

```
# 创建项目目录并进入
mkdir PGG_Panel
cd PGG_Panel
```

#### 1. qc

在父目录 PGG\_Panel 下建立 frq 的工作目录

```
mkdir 01.qc; cd 01.qc
```

将以下脚本下载到工作目录中

```
https://github.com/Shuhua-Group/PGG.Panel/blob/main/qc.sh
```

修改脚本，将以下三行替换为你的 vcf 文件路径、结果文件前缀及采取的质控方案（VQSR 或 filtered 硬过滤）

```
vcf="/mnt/share9/gaoyang/20240807.Saile/14.POP_zyHan_HXDL/VQSR/190.variant.quality
.score.recal/POP2099_HXDL_zyHan.genomewide.VQSR.variants.vcf.gz"
prefix='POP_zyHan_HXDL'
qc='VQSR' #qc=filtered
```

对 chr1-22 执行脚本

```
for k in {1..22}
do
    sh qc.sh $k 1>qc.chr$k.log 2>&1
done
```

结果文件应包含 chr1-22 的 vcf 文件和 bfile 文件，以 chr22 为例：

```
POP_zyHan_HXDL.chr22.VQSR.PASS.biallelic.SNPs.FM10.AC1.bed
POP_zyHan_HXDL.chr22.VQSR.PASS.biallelic.SNPs.FM10.AC1.bim
POP_zyHan_HXDL.chr22.VQSR.PASS.biallelic.SNPs.FM10.AC1.fam
POP_zyHan_HXDL.chr22.VQSR.PASS.biallelic.SNPs.FM10.AC1.log
POP_zyHan_HXDL.chr22.VQSR.PASS.biallelic.SNPs.FM10.AC1.vcf.gz
POP_zyHan_HXDL.chr22.VQSR.PASS.biallelic.SNPs.FM10.AC1.vcf.gz.tbi
```

#### 2. relative

在父目录 PGG\_Panel 下建立 relative 的工作目录

```
mkdir 02.relative; cd 02.relative
```

将以下脚本下载到工作目录中

```
https://github.com/Shuhua-Group/PGG.Panel/blob/main/relative.sh
https://github.com/Shuhua-Group/PGG.Panel/blob/main/5011.relative.combinations.py
```

修改脚本，将以下一行替换为你的结果文件前缀及采取的质控方案（和 step1 保持一致）

```
prefix='POP_zyHan_HXDL'
qc='VQSR' #qc=filtered
```

执行脚本

```
sh relative.sh 1>relative.log 2>&1
```

结果文件应包含 chrA 的 bfile 文件及亲缘关系推断文件，如下所示：

```
POP_zyHan_HXDL.chrA.VQSR.PASS.biallelic.SNPs.FM10.AC1.bed
POP_zyHan_HXDL.chrA.VQSR.PASS.biallelic.SNPs.FM10.AC1.bim
POP_zyHan_HXDL.chrA.VQSR.PASS.biallelic.SNPs.FM10.AC1.fam
POP_zyHan_HXDLallsegs.txt
POP_zyHan_HXDL.kin0
rm.relative.king.list
no.relate.list
```

### 3. phase

在父目录 PGG\_Panel 下建立 phase 的工作目录并进入

```
mkdir 03.phase; cd 03.phase
```

将以下脚本下载到工作目录中

```
https://github.com/Shuhua-Group/PGG.Panel/blob/main/phase.sh
https://github.com/Shuhua-Group/PGG.Panel/blob/main/shapeit5.sh
```

对 phase.sh 进行编辑，将以下三行替换为你的结果文件前缀、采取的质控方案（和 step1 保持一致）及 GRCh38 版本的 shapeit map 文件路径

```
prefix='POP_zyHan_HXDL'
qc='VQSR' #qc=filtered
map="/home/sunyumeng/software/shapeit4-4.2.2/maps/b38"
```

对 shapeit5.sh 进行编辑，将下行替换为你的 shapeit5 安装路径

```
SHAPEIT_DIR="/home/sunyumeng/software/shapeit5.1.1"
```

对 chr1-22 执行脚本

```
for k in {1..22}
do
    sh phase.sh $k 1>phase.chr$k.log 2>&1
done
```

结果文件应包含 chr1-22 的 phased vcf 及其索引文件，以 chr22 为例：

```
POP_zyHan_HXDL.chr22.VQSR.PASS.biallelic.SNPs.FM10.AC1.shapeit5.final.rmPP.vcf.gz
POP_zyHan_HXDL.chr22.VQSR.PASS.biallelic.SNPs.FM10.AC1.shapeit5.final.rmPP.vcf.gz.tbi
```

#### 4. panel

在父目录 PGG\_Panel 下建立 panel 的工作目录并进入

```
mkdir 04.panel; cd 04.panel
```

将以下脚本下载到工作目录中

```
https://github.com/Shuhua-Group/PGG.Panel/blob/main/panel.sh
```

对 panel.sh 进行编辑，将以下三行替换为你的结果文件前缀、采取的质控方案（和 step1 保持一致）及 minimac4 路径

```
prefix='POP_zyHan_HXDL'
qc='VQSR' #qc=filtered
minimac4='/home/sunyumeng/software/minimac4-4.1.6-Linux-x86_64/bin/minimac4'
```

对 chr1-22 执行脚本

```
for k in {1..22}
do
    sh panel.sh $k 1>panel.chr$k.log 2>&1
done
```

结果文件应包含 chr1-22 的 msav 文件，以 chr22 为例：

```
POP_zyHan_HXDL.chr22.VQSR.PASS.biallelic.SNPs.FM10.AC1.shapeit5.msav
```

#####  
以下是用 reference panel 对 target vcf 进行 imputation 并获取后续 meta-imputation 所需文件的示例  
#####

#### 5. impute

在父目录 PGG\_Panel 下建立 impute 的工作目录并进入

```
mkdir 05.impute; cd 05.impute
```

以 chr22 为例，定义 minimac4 路径、minimac4 map 文件路径、panel 路径及 target 路径

```
i=22
minimac4='/home/sunyumeng/software/minimac4-4.1.6-Linux-x86_64/bin/minimac4'
minimac4_mapFile='/home/sunyumeng/data/genetic_map_hg38_withX_3cols.tsv.gz'
panel='./04.panel/POP_zyHan_HXDL.chr22.VQSR.PASS.biallelic.SNPs.FM10.AC1.shapeit5.msav'
target='/home/sunyumeng/AS/AS.chr$i.PASS.biallelic.FM1.SNPs.phased.vcf.gz'
```

进行 imputation 并通过 -e 获取后续 meta-imputation 所需文件

```
$minimac4 $panel $target -O vcf.gz -o imputed.chr$i.dose.vcf.gz -e imputed.chr$i.empirical_dose.vcf.gz -m  
$minimac4_mapFile --all-typed-sites -f GT,DS,HDS --min-ratio 0.000001
```

此外，还可以通过 --sample-ids-file 参数定义 panel 中用于 impute 的样本集，如使用无亲缘关系样本集进行 imputation:

```
--sample-ids-file ../02/no.relate.list
```