GWAS Tutorial

Author: Chioma Onyido

File Manipulation

In this document, we will demonstrate how to manipulate files in different formats

1. Set Up

    i. Create a folder "GWAS_practice"

```
mkdir GWAS_practice
```

    ii. Change directory to "GWAS_practice"

```
cd GWAS_practice
```

    iii. Download map and ped files (and plink 1.9 if you haven't) to this folder:

for hapmap1.map

```
wget http://www.bioinf.wits.ac.za/courses/sahgp/plinkdata/hapmap1.map
```

for hapmap1.ped

```
wget http://www.bioinf.wits.ac.za/courses/sahgp/plinkdata/hapmap1.ped
```

for plink (visit https://www.cog-genomics.org/plink/ and ensure to download for the appropriate OS).

```
wget https://s3.amazonaws.com/plink1-assets/plink_linux_x86_64_20230116.zip
```

Extract the downloaded file:

```
unzip plink_linux_x86_64.zip
```

    iv. Check the contents of your hapmap file

```
head hapmap1.ped
```

2. Plink

    v. Run plink and save log file to hapmap1 (the log file contains various information about errors or progress messages run by plink, it is helpful for documentation)

```
plink --file hapmap1 --out hapmap1
```

   vi. Convert ped and map files to plink binary files (.bed, .bim and .fam)

```
plink --file hapmap1 --make-bed --out hapmap1
```

   vii. To update the IDs to a naming convention of your choice (I saved mine as "PRAC")

```
awk '{$1 = "PRAC" NR; print}' OFS='\t' hapmap1.fam > updated_ID
```

   viii. Extract variant/ info of first 20 individuals

```
sed -n 1,20p updated_ID > extract_ID
```

   ix. See first 10 lines

```
head extract_ID
```

   x. To search for lines in a file

```
grep -w "PRAC20" updated_ID
grep rs307347 updated_ID.bim
```