

PRS calculation for GTEx SNPs:

Problems:

- 1- GTEx SNPs have different SNPs ID format than those used in GWAS and the LD reference panel (used for the calculation of PRS with the tool PRS-CS)
- 2- GTEx reference Genome is different than that used in GWAS and in the LD reference panel

Initial Solution:

Get RSIDs by querying some databases using the chromosome number and the SNPs position:

→ Very slow and there is still the problem of matching the reference genome eventually.

Second Solution:

Mapping to the correct SNPs ID format using the chromosome number and the position using LD reference panel SNPs:

→ Very few SNPs were mapped because of the difference in the reference genome version (GWAS_sumstats and LD reference panel are GR37 but GTEx SNPs are GR38)

-Original number of SNPs in GTEx: 69763935

-Mapped to RS-ID: 36980

-LD reference panel SNPs: 1120697

Final Solution:

Map (lift) the reference Genome to hg19 using the tool LiftOverPlink

1.Create map file from bim+bed+fam format using plink:

```
./plink --bfile GTEx --recode --out GTEx
```

2.Use the tool LifOverPlink to lift the reference genome from 38 to 37 matching the LD reference used in PRC-CS and the GWAS reference

```
python liftOverPlink.py -m GTEX.map -p GTEX.ped -o lifted -c  
hg38ToHg19.over.chain -e liftOver.exe
```

-Output three files: lifted.map, lifted.ped and lifted.bed.unlifted

Number of mapped SNPs to the new reference Genome in lifted.map: 64041302
original number before lifting is 69763935

3. Transfer the map file created back to bim+bed+fam format

```
./plink --file lifted --make-bed --out GTEX-lifted --memory  
120000 --allow-extra-chr
```

-Output: GTEX-lifted.bed, GTEX-lifted.bim, GTEX-lifted.fam, GTEX-lifted.nosex

4. Map the SNPs ID of GTEEx data to the correct format using the LD reference
panel SNPs: see script map_IDs.R

*** Now the number of SNPs mapped to RS ID is 1128658 → There are
duplicates because the LD reference has 1120697 SNPS →
Remove duplicates

```
awk '{seen[$2]++; if(seen[$2]==1){ print}}'  
mapped_GTEEx_SNP.s.bim > mapped_GTEEx_SNP.s.nodup.bim
```

-Final number of resulting SNPs mapped to LD reference after duplicate removal
is: 1119899 (difference is small only 796 SNPs= 1120697-1119899)

5. Remove header of bim file > mapped_GTEExSNPs.nodup.bim

6. Use PRS_CS: see script 01_run_PRScs.sh