1. Download discovery dataset from [PGC](https://www.med.unc.edu/pgc/results-and-downloads) or [ENIGMA](http://enigma.ini.usc.edu/research/download-enigma-gwas-results/)
2. Subset for SNPs that intersect between the discovery dataset and experiment dataset. Separate the SNPs by chromosome.
3. Clump SNPs
   1. Download the 1000G EUR data, Phase 3 (VCF format)
   2. Extract SNPs form 1000G EUR that was subset in step 2 (by chromosome) and remove related subjects

plink --vcf *1000Gfile* --extract *snplist\_step2* --remove *related\_subjects* --make- bed --out *1000Gfiles\_subset*

* 1. Remove duplicate SNPs

plink --bfile *1000Gfiles\_subset* --list-duplicate-vars --out *duplicated\_snps\_to\_remove*

* 1. Clump SNPs with the following parameters

plink --bfile *1000Gfiles\_subset* --exclude *duplicated\_snps\_to\_remove* --keep *n\_eur\_subjects\_only* --clump *discovery\_data\_file* --clump-p1 1 --clump-p2 1 --clump-r2 0.1 --clump-kb 1000 --out *clumped\_result*

1. Create polygenic risk score association file and calculate polygenic risk score:
   1. Subset of the discovery data to include only the SNPs that were clumped and format for plink
      1. Filtered all SNPs where P < 0.5 in the discovery dataset. If odds ratio (OR) is < 1, the alternative allele was used as the risk allele and OR was calculated as OR = log(1/OR). NOTE: ENIGMA provided beta values instead of OR. If the beta < 0, the alternative allele as the risk allele and the beta’s sign was flipped.
   2. Create the polygenic risk score with plink, using Northern Europeans only

plink --bfile *plinkformateddata* --keep *n\_european\_ids* --score *output\_step4A* --out *final\_prs*