

Session 10 Exercise: Inferring Population Structure with PLINK for HapMap populations

1. Use PLINK to perform a multi-dimensional scaling (MDS) analysis (which is similar to principal components analysis), for population structure inference using the PLINK file “plink_hapmap_CEU_YRI_ASW_CHB.bed.” You can use the following two PLINK commands for the MDS analysis:

```
./plink --bfile plink_hapmap_CEU_YRI_ASW_CHB --genome --  
out plink_hapmap_4pops --noweb
```

```
./plink --bfile plink_hapmap_CEU_YRI_ASW_CHB --read-  
genome plink_hapmap_4pops.genome --cluster --mds-plot 10 --  
out plink_hapmap_4pops --noweb
```

The first command above creates a genetic relatedness matrix (GRM) with PLINK. The second command performs the MDS analysis with PLINK using the GRM. See the PLINK website for more details on these commands.

2. Using R, read in the PLINK output file for the MDS analysis. Also read in the file “HapMap_4pops.txt” that is available on the course website. Inspect the files.
3. In R, make scatterplots of the top dimensions from the MDS analysis with PLINK to illustrate the population structure of the HapMap samples (Hint: Use the R-script that is available on the course website)