

4MDS analysis

Prerequisites

Your genotyping data in PLINK format: e.g. mydata.bed, mydata.bim, mydata.fam

**** make sure your data is aligned to human genome build 19 and to the + strand ****

Software: PLINK, R

Files: MDSplot.R

1KG reference data: refdata.bed, refdata.bim, refdata.fam

1000 Genomes Reference Populations include:

North-west Europe (**CEU**, n=81),

West Africa (**YRI**, n=85),

East Asian (**CHB**, n=97) and the Americas;

Puerto Rico (**PUR**, n=55),

Mexican from Los Angeles (**MXL**, n=59) and

Colombia (**CLM**, n=60) extracted from the 1000 Genomes Project integrated phase I release.

STEP1: Combine your dataset with reference dataset

```
plink --bfile mydata --bmerge refdata.bed refdata.bim refdata.fam --make-bed --out mydata.refdata
```

**** If there is a strand issue, please flip (--flip in Plink) your SNPs based on the file of mydata.refdata.missnp created in STEP2 and redo this step. ****

STEP2: Select variants with geno 0.05 and MAF > 0.05

```
plink --bfile mydata.refdata --maf 0.05 --geno 0.05 --make-bed --out mydata.refdata.QCed
```

Note: For big data sets ($N > 8000$), please use a callrate cut-off of 99% (--geno 0.01)

STEP3: Make genome file

```
plink --bfile mydata.refdata.QCed --Z-genome --out mydata.refdata.QCed.Z
```

**** this step may take very LONG! ****

STEP4: MDS

```
plink --bfile mydata.refdata.QCed --read-genome mydata.refdata.QCed.Z.genome.gz \
--cluster --mds-plot 10 --out mydata.refdata.QCed.MDS
```

STEP5: MDS-PLOT

Use MDSplot.R to create MDS plot for the first 2 components. Your dataset will be colored in red.

Input: MDSFILE: *.mds file from preceding step
 REFFILE: "refdata.fam" from reference data set (used to identify 1KGP populations)
 NAME: Prefix for Output-Name

Usage: R --slave MDSFILE REFFILE NAME < MDSplot.R

Example: R --slave mydata.refdata.QCed.MDS.mds refdata.fam MyOutput < MDSplot.R

Output: MyOutput.C1-C2.png

Figure 1: AA MDS PLOT based on ~4K SNPs

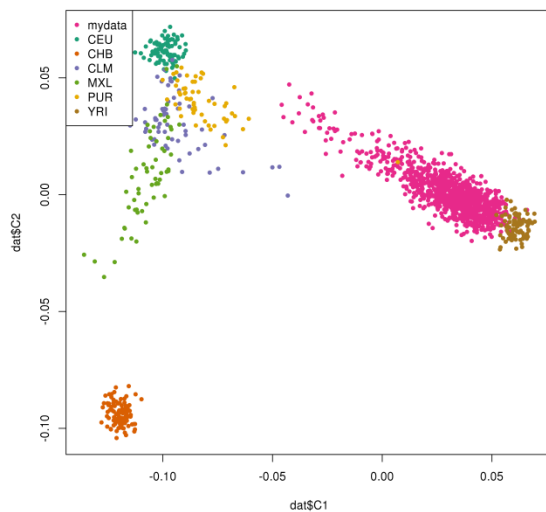


Figure 2: AA MDS plot based on 77.793 SNPs (higher resolution)

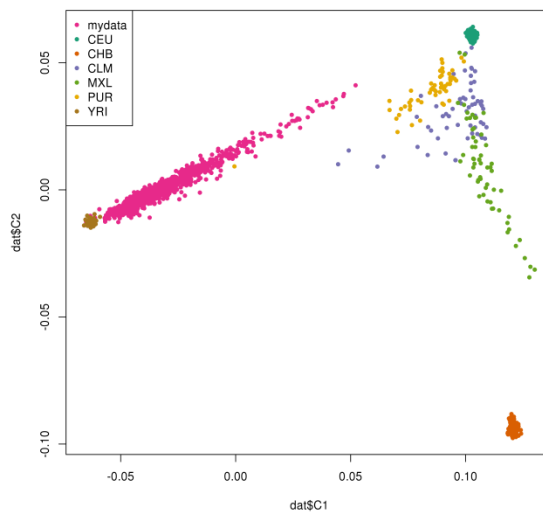


Figure 3: EA MDS PLOT based on ~4K SNPs

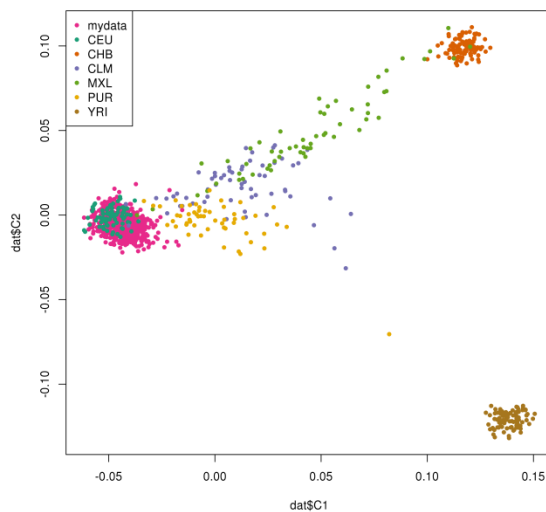


Figure 4: HA MDS PLOT based on ~4K SNPs

