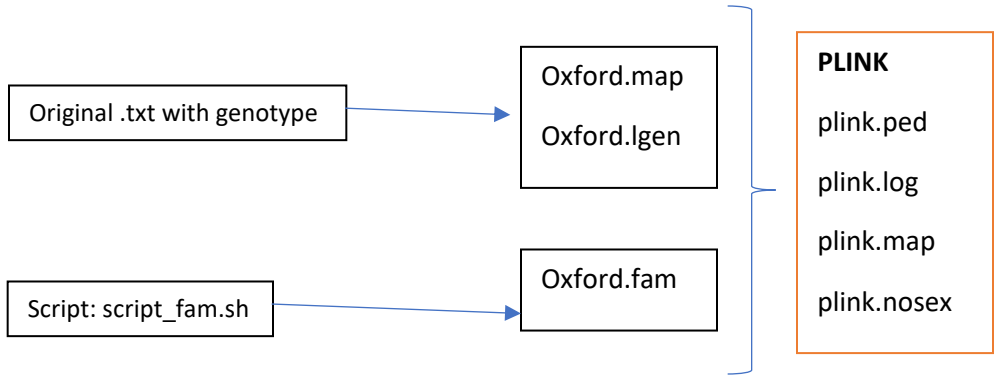
Step1. Obtain .txt file sets for PLINK



Take the SNPs information for map file:

```
tail -n +2 _Oxford_eQTL_cohort_genotype.txt | awk '{print $5 "\t" $1 "\t" $6}' | sort | uniq > Oxford.map

tail -n +2 ega-box-169_Oxford_eQTL_cohort_genotype2.txt | awk '{print $5 "\t" $1 "\t" $6}' | sort | uniq > Oxford2.map
```

Create LGEN files:

```
awk '{print $2, "\t", $2, "\t", $1, "\t", $3, "\t", $4}' _Oxford_eQTL_cohort_genotype.txt  > snps.lgen
tail -n +2 snps.lgen > Oxford.lgen
sed -i 's/-/0/g' MONOCYTE_eQTLs/genotypes/Oxford.lgen

awk '{print $2, "\t", $2, "\t", $1, "\t", $3, "\t", $4}' ega-box-169_Oxford_eQTL_cohort_genotype2.txt  > snps2.lgen
tail -n +2 snps2.lgen > Oxford2.lgen
sed -i 's/-/0/g' MONOCYTE_eQTLs/genotypes/Oxford2.lgen
```

Create FAM files:

```
bash script_fam.sh

bash script_fam2.sh
```

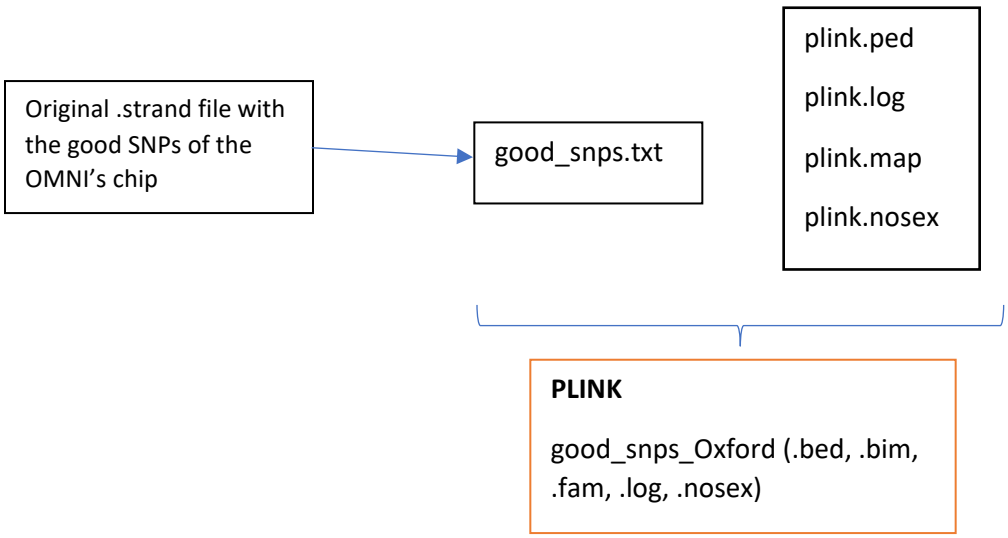
Create .ped files:

```
bin/plink --lfile MONOCYTE_eQTLs/genotypes/Oxford --recode

bin/plink --lfile MONOCYTE_eQTLs/genotypes/Oxford2 –recode
```

Step2. Filter by good SNPs

We pass from 733.202 SNPs in the original files, to 716.356 good SNPs.



Obtain good SNPs:

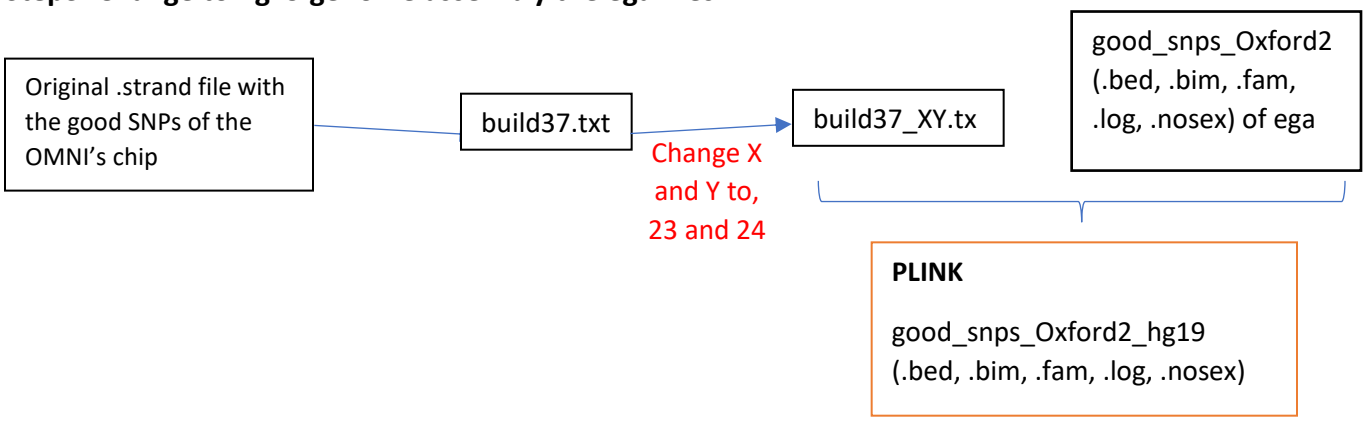
```
cut -f1 humanomniexpress-24v1-0_a-b37.strand > goodsnps.txt
```

Filter by good SNPs:

```
bin/plink.exe --file MONOCYTE_eQTLs/genotypes/Oxford/plink --extract MONOCYTE_eQTLs/genotypes/goodsnps.txt --make-bed --out good_snps_Oxford
```

```
bin/plink.exe --file MONOCYTE_eQTLs/genotypes/Oxford2/initial_files/plink --extract MONOCYTE_eQTLs/genotypes/goodsnps.txt --make-bed --out good_snps_Oxford2
```

Step3. Change to hg19 genome assembly the ega files



Extract hg19 annotation:

```
cut -f1,2,3 MONOCYTE_eQTLs/genotypes/humanomniexpress-24v1-0_a-b37.strand > build37.txt
```

Pass from X and Y to 23 and 24:

```
awk '{ if ($2 == "X") $2="23"; print $1, "\t", $2, "\t", $3 }' build37.txt  > build37_X.txt

awk '{ if ($2 == "Y") $2="24"; print $1, "\t", $2, "\t", $3 }' build37_X.txt  > build37_XY.txt
```

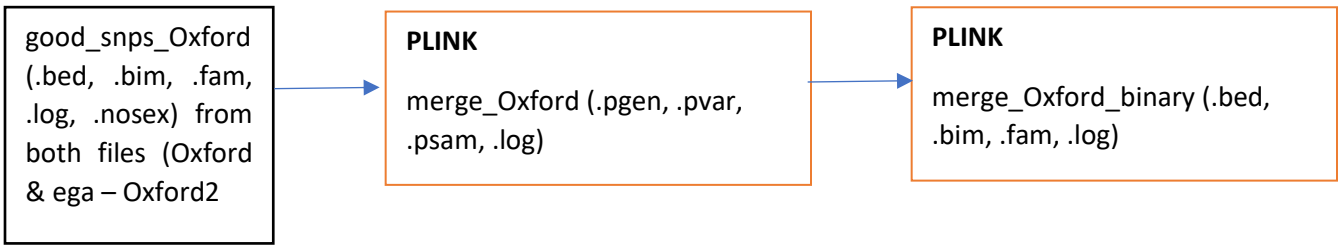
Change hg19 annotation

```
bin/plink.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford2/good_snps_files/good_snps_Oxford2 --update-map build37_XY.txt --make-bed --out good_snps_Oxford2_hg19
```

Step4. QC

1) Merge X Y chromosomes (PLINK 2)

First, we need to sort the chromosomes to avoid splitted chromosomes, but only .pgen format support it. So, we perform the sort and the merge in a .pgen, and then we pass the .pgen to .bed (binary format).



a) Oxford

```
bin/plink2.exe --bfile good_snps_Oxford -make-pgen --sort-vars --merge-x -out merge_Oxford
```

```
PLINK v2.00a3 32-bit (9 Apr 2020)          www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to merge_Oxford.log.
Options in effect:
  --bfile good_snps_Oxford
  --make-pgen
  --merge-x
  --out merge_Oxford
  --sort-vars
Start time: Thu May 07 17:30:37 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
Allocated 1296 MiB successfully, after larger attempt(s) failed.
Using up to 8 compute threads.
288 samples (0 females, 0 males, 288 ambiguous; 288 founders) loaded from
good_snps_Oxford.fam.
--merge-x: 515 chromosome codes changed.
716356 variants loaded from good_snps_Oxford.bim.
Note: No phenotype data present.
Writing merge_Oxford.pvar ... done.
Writing merge_Oxford.psam ... done.
Writing merge_Oxford.pgen ... done.
End time: Thu May 07 17:37:47 2020
```

```
bin/plink2.exe --pfile merge_Oxford -make-bed -out merge_Oxford_binary
```

```
PLINK v2.00a3 32-bit (9 Apr 2020)          www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to merge_Oxford_binary.log.
Options in effect:
  --make-bed
  --out merge_Oxford_binary
  --pfile merge_Oxford

Start time: Thu May 07 17:52:24 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
```

```
Allocated 1296 MiB successfully, after larger attempt(s) failed.
Using up to 8 compute threads.
288 samples (0 females, 0 males, 288 ambiguous; 288 founders) loaded from
merge_Oxford.psam.
716356 variants loaded from merge_Oxford.pvar.
Note: No phenotype data present.
Writing merge_Oxford_binary.fam ... done.
Writing merge_Oxford_binary.bim ... done.
Writing merge_Oxford_binary.bed ... done.
End time: Thu May 07 17:52:26 2020
```

b) Oxford2 (ega)

```
bin/plink2.exe --bfile good_snps_Oxford2_hg19 -make-pgen --sort-vars --merge-x -out merge_Oxford2
```

```
PLINK v2.00a3 32-bit (9 Apr 2020)          www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to merge_Oxford2.log.
Options in effect:
  --bfile good_snps_Oxford2_hg19
  --make-pgen
  --merge-x
  --out merge_Oxford2
  --sort-vars

Start time: Thu May 07 19:29:28 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
Allocated 1296 MiB successfully, after larger attempt(s) failed.
Using up to 8 compute threads.
144 samples (0 females, 0 males, 144 ambiguous; 144 founders) loaded from
good_snps_Oxford2_hg19.fam.
--merge-x: 457 chromosome codes changed.
716356 variants loaded from good_snps_Oxford2_hg19.bim.
Note: No phenotype data present.
Writing merge_Oxford2.pvar ... done.
Writing merge_Oxford2.psam ... done.
Writing merge_Oxford2.pgen ... done.
End time: Thu May 07 19:34:51 2020
```

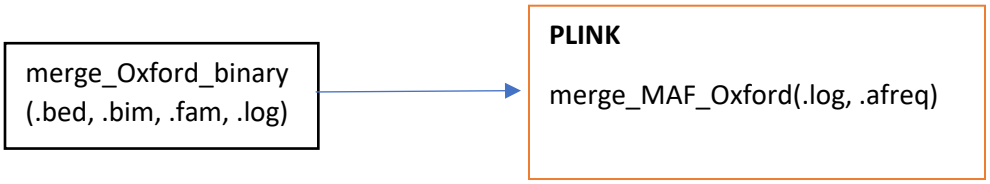
```
bin/plink2.exe --pfile merge_Oxford2 -make-bed -out merge_Oxford2_binary
```

```
PLINK v2.00a3 32-bit (9 Apr 2020)          www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to merge_Oxford2_binary.log.
Options in effect:
  --make-bed
  --out merge_Oxford2_binary
  --pfile merge_Oxford2

Start time: Thu May 07 19:36:55 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
Allocated 1296 MiB successfully, after larger attempt(s) failed.
Using up to 8 compute threads.
144 samples (0 females, 0 males, 144 ambiguous; 144 founders) loaded from
merge_Oxford2.psam.
716356 variants loaded from merge_Oxford2.pvar.
Note: No phenotype data present.
Writing merge_Oxford2_binary.fam ... done.
Writing merge_Oxford2_binary.bim ... done.
Writing merge_Oxford2_binary.bed ... done.
End time: Thu May 07 19:36:56 2020
```

2) ANALYZE MARKERS

2.1) Calculate MAF (PLINK2)



a) Oxford

```
bin/plink2.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford/files_XY_merge/merge_Oxford_binary --freq --out merge_MAF_Oxford
```

```
PLINK v2.00a3 32-bit (9 Apr 2020)          www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to merge_MAF_Oxford.log.
Options in effect:
 --bfile MONOCYTE_eQTLs/genotypes/Oxford/files_XY_merge/merge_Oxford_binary
 --freq
 --out merge_MAF_Oxford

Start time: Thu May 07 19:44:22 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
Allocated 1296 MiB successfully, after larger attempt(s) failed.
Using up to 8 compute threads.
288 samples (0 females, 0 males, 288 ambiguous; 288 founders) loaded from
MONOCYTE_eQTLs/genotypes/Oxford/files_XY_merge/merge_Oxford_binary.fam.
716356 variants loaded from
MONOCYTE_eQTLs/genotypes/Oxford/files_XY_merge/merge_Oxford_binary.bim.
Note: No phenotype data present.
Calculating allele frequencies... done.
--freq: Allele frequencies (founders only) written to merge_MAF_Oxford.afreq .
End time: Thu May 07 19:44:23 2020
```

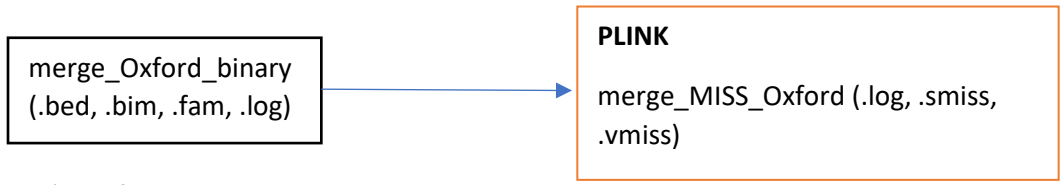
b) Oxford2 (ega)

bin/plink2.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford2/files_XY_merge/merge_Oxford2_binary --freq --out merge_MAF_Oxford2

```
PLINK v2.00a3 32-bit (9 Apr 2020)          www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to merge_MAF_Oxford2.log.
Options in effect:
 --bfile MONOCYTE_eQTLs/genotypes/Oxford2/files_XY_merge/merge_Oxford2_binary
 --freq
 --out merge_MAF_Oxford2

Start time: Thu May 07 19:51:08 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
Allocated 1296 MiB successfully, after larger attempt(s) failed.
Using up to 8 compute threads.
144 samples (0 females, 0 males, 144 ambiguous; 144 founders) loaded from
MONOCYTE_eQTLs/genotypes/Oxford2/files_XY_merge/merge_Oxford2_binary.fam.
716356 variants loaded from
MONOCYTE_eQTLs/genotypes/Oxford2/files_XY_merge/merge_Oxford2_binary.bim.
Note: No phenotype data present.
Calculating allele frequencies... done.
--freq: Allele frequencies (founders only) written to merge_MAF_Oxford2.afreq .
End time: Thu May 07 19:51:09 2020
```

2.2) Calculate missing genotypes (PLINK2)



a) Oxford

bin/plink2.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford/files_XY_merge/merge_Oxford_binary --missing --out merge_MISS_Oxford

```
PLINK v2.00a3 32-bit (9 Apr 2020)          www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to merge_MISS_Oxford.log.
Options in effect:
 --bfile MONOCYTE_eQTLs/genotypes/Oxford/files_XY_merge/merge_Oxford_binary
 --missing
 --out merge_MISS_Oxford

Start time: Thu May 07 19:56:30 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
Allocated 1296 MiB successfully, after larger attempt(s) failed.
Using up to 8 compute threads.
288 samples (0 females, 0 males, 288 ambiguous; 288 founders) loaded from
MONOCYTE_eQTLs/genotypes/Oxford/files_XY_merge/merge_Oxford_binary.fam.
716356 variants loaded from
MONOCYTE_eQTLs/genotypes/Oxford/files_XY_merge/merge_Oxford_binary.bim.
Note: No phenotype data present.
Calculating sample missingness rates... done.
Calculating allele frequencies... done.
```

```
--missing: Sample missing data report written to merge_MISS_Oxford.smiss .
--missing: Variant missing data report written to merge_MISS_Oxford.vmiss .
End time: Thu May 07 19:56:32 2020
```

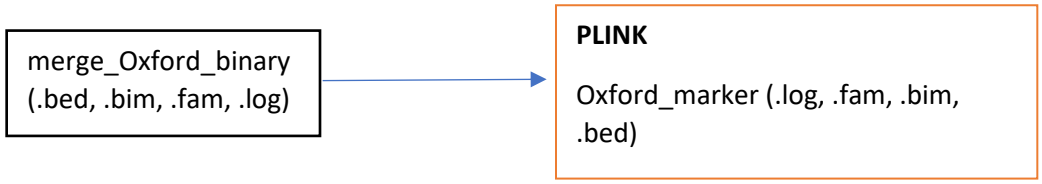
b) Oxford2 (ega)

```
bin/plink2.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford2/files_XY_merge/merge_Oxford2_binary --missing --out merge_MISS_Oxford2
```

```
PLINK v2.00a3 32-bit (9 Apr 2020)          www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to merge_MISS_Oxford2.log.
Options in effect:
  --bfile MONOCYTE_eQTLs/genotypes/Oxford2/files_XY_merge/merge_Oxford2_binary
  --missing
  --out merge_MISS_Oxford2

Start time: Thu May 07 20:00:17 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
Allocated 1296 MiB successfully, after larger attempt(s) failed.
Using up to 8 compute threads.
144 samples (0 females, 0 males, 144 ambiguous; 144 founders) loaded from
MONOCYTE_eQTLs/genotypes/Oxford2/files_XY_merge/merge_Oxford2_binary.fam.
716356 variants loaded from
MONOCYTE_eQTLs/genotypes/Oxford2/files_XY_merge/merge_Oxford2_binary.bim.
Note: No phenotype data present.
Calculating sample missingness rates... done.
Calculating allele frequencies... done.
--missing: Sample missing data report written to merge_MISS_Oxford2.smiss .
--missing: Variant missing data report written to merge_MISS_Oxford2.vmiss .
End time: Thu May 07 20:00:18 2020
```

2.3) Remove markers as per maf/geno/hwe thresholds (PLINK2)



a) Oxford

```
bin/plink2.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford/files_XY_merge/merge_Oxford_binary --geno 0.05 --hwe 1e-06 --maf 0.01 --make-bed --out Oxford_marker
```

```
PLINK v2.00a3 32-bit (9 Apr 2020)          www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford_marker.log.
Options in effect:
  --bfile MONOCYTE_eQTLs/genotypes/Oxford/files_XY_merge/merge_Oxford_binary
  --geno 0.05
  --hwe 1e-06
  --maf 0.01
  --make-bed
  --out Oxford_marker

Start time: Thu May 07 20:05:06 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
Allocated 1296 MiB successfully, after larger attempt(s) failed.
Using up to 8 compute threads.
288 samples (0 females, 0 males, 288 ambiguous; 288 founders) loaded from
MONOCYTE_eQTLs/genotypes/Oxford/files_XY_merge/merge_Oxford_binary.fam.
716356 variants loaded from
MONOCYTE_eQTLs/genotypes/Oxford/files_XY_merge/merge_Oxford_binary.bim.
Note: No phenotype data present.
Calculating allele frequencies... done.
--geno: 684 variants removed due to missing genotype data.
--hwe: 125 variants removed due to Hardy-Weinberg exact test (founders only).
76378 variants removed due to allele frequency threshold(s)
(--maf/--max-maf/--mac/--max-mac).
639169 variants remaining after main filters.
Writing Oxford_marker.fam ... done.
Writing Oxford_marker.bim ... done.
Writing Oxford_marker.bed ... done.
End time: Thu May 07 20:05:08 2020
```

b) Oxford2 (ega)

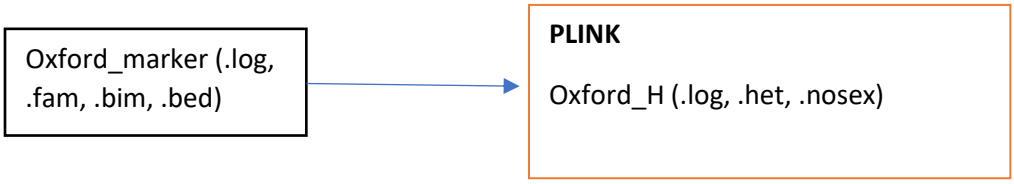
bin/plink2.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford2/files_XY_merge/merge_Oxford2_binary --geno 0.05 --hwe 1e-06 --maf 0.01 --make-bed --out Oxford2_marker

```
PLINK v2.00a3 32-bit (9 Apr 2020)           www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford2_marker.log.
Options in effect:
  --bfile MONOCYTE_eQTLs/genotypes/Oxford2/files_XY_merge/merge_Oxford2_binary
  --geno 0.05
  --hwe 1e-06
  --maf 0.01
  --make-bed
  --out Oxford2_marker

Start time: Thu May 07 20:06:36 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
Allocated 1296 MiB successfully, after larger attempt(s) failed.
Using up to 8 compute threads.
144 samples (0 females, 0 males, 144 ambiguous; 144 founders) loaded from
MONOCYTE_eQTLs/genotypes/Oxford2/files_XY_merge/merge_Oxford2_binary.fam.
716356 variants loaded from
MONOCYTE_eQTLs/genotypes/Oxford2/files_XY_merge/merge_Oxford2_binary.bim.
Note: No phenotype data present.
Calculating allele frequencies... done.
--geno: 1774 variants removed due to missing genotype data.
--hwe: 263 variants removed due to Hardy-Weinberg exact test (founders only).
76031 variants removed due to allele frequency threshold(s)
(--maf/--max-maf/--mac/--max-mac).
638288 variants remaining after main filters.
Writing Oxford2_marker.fam ... done.
Writing Oxford2_marker.bim ... done.
Writing Oxford2_marker.bed ... done.
End time: Thu May 07 20:06:37 2020
```

3) ANALYSE INDIVIDUALS

3.1) Analyse heterozygosity (PLINK)



a) Oxford

bin/plink.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford/Markers/Oxford_marker --het --out Oxford_H

```
PLINK v1.90b6.17 32-bit (28 Apr 2020)       www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford_H.log.
Options in effect:
  --bfile MONOCYTE_eQTLs/genotypes/Oxford/Markers/Oxford_marker
  --het
  --out Oxford_H

16168 MB RAM detected; reserving 2047 MB for main workspace.
Allocated 1535 MB successfully, after larger attempt(s) failed.
639169 variants loaded from .bim file.
288 people (0 males, 0 females, 288 ambiguous) loaded from .fam.
Ambiguous sex IDs written to Oxford_H.nosex .
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 288 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Warning: Nonmissing nonmale Y chromosome genotype(s) present; many commands
treat these as missing.
Total genotyping rate is 0.998693.
639169 variants and 288 people pass filters and QC.
Note: No phenotypes present.
--het: 622512 variants scanned, report written to Oxford_H.het .
```

b) Oxford2 (ega)

bin/plink.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford2/Markers/Oxford2_marker --het --out Oxford2_H

```
PLINK v1.90b6.17 32-bit (28 Apr 2020)        www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford2_H.log.
Options in effect:
  --bfile MONOCYTE_eQTLs/genotypes/Oxford2/Markers/Oxford2_marker
  --het
  --out Oxford2_H

16168 MB RAM detected; reserving 2047 MB for main workspace.
Allocated 1535 MB successfully, after larger attempt(s) failed.
638288 variants loaded from .bim file.
144 people (0 males, 0 females, 144 ambiguous) loaded from .fam.
Ambiguous sex IDs written to Oxford2_H.nosex .
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 144 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Warning: Nonmissing nonmale Y chromosome genotype(s) present; many commands
treat these as missing.
Total genotyping rate is 0.999379.
638288 variants and 144 people pass filters and QC.
Note: No phenotypes present.
--het: 622000 variants scanned, report written to Oxford2_H.het .
```

3.2) Plot missing vs. heterozygosity (R → imiss-vs-het.R script)

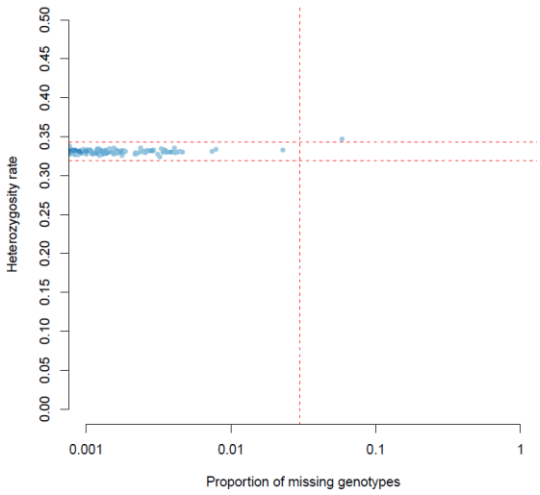
merge_MISS_Oxford.smiss

Oxford_H.het

R

imiss-vs-het.R

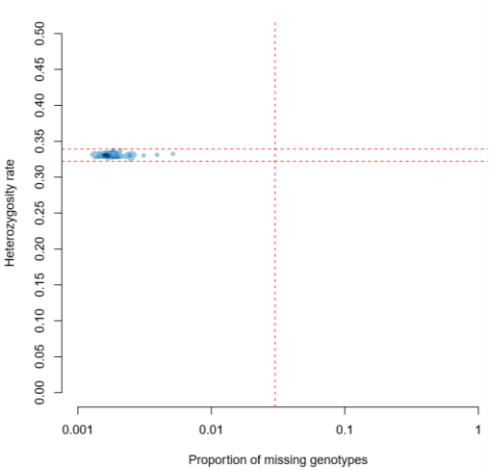
a) Oxford




```
imiss=read.table("MONOCYTE_eQTLs/genotypes/Oxford/Miss_genotype/merge_MISS_Oxford.smiss",h=F)
imiss$logF_MISS = log10(imiss[,5])
het=read.table("MONOCYTE_eQTLs/genotypes/Oxford/Heterozygosity/Oxford_H.het",h=T)
het$meanHet = (het$N.NM. - het$O.HOM.)/het$N.NM.
library("geneplotter")
colors  <- densCols(imiss$logF_MISS,het$meanHet)
pdf("imiss-vs-het.pdf")
plot(imiss$logF_MISS,het$meanHet, col=colors, xlim=c(-3,0),ylim=c(0,0.5),pch=20, xlab="Proportion of missing genotypes",
ylab="Heterozygosity rate",axes=F)
axis(2,at=c(0,0.05,0.10,0.15,0.2,0.25,0.3,0.35,0.4,0.45,0.5),tick=T)
```

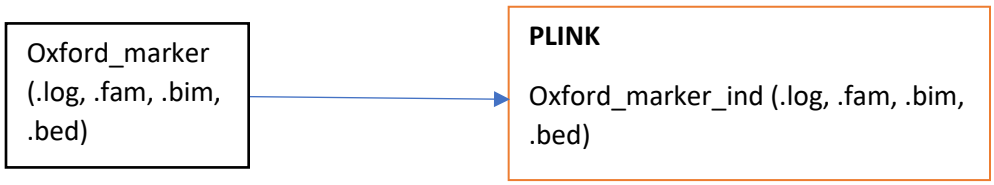
```
axis(1,at=c(-3,-2,-1,0),labels=c(0.001,0.01,0.1,1))
abline(h=mean(het$meanHet)-(4*sd(het$meanHet)),col="RED",lty=2)
abline(h=mean(het$meanHet)+(4*sd(het$meanHet)),col="RED",lty=2)
abline(v=-1.522879, col="RED", lty=2)
dev.off()
```

b) Oxford 2 (ega)




```
imiss=read.table("MONOCYTE_eQTLs/genotypes/Oxford2/Miss_genotype/merge_MISS_Oxford2.smiss",h=F)
imiss$logF_MISS = log10(imiss[,5])
het=read.table("MONOCYTE_eQTLs/genotypes/Oxford2/Heterozygosity/Oxford2_H.het",h=T)
het$meanHet = (het$N.NM. - het$O.HOM.)/het$N.NM.
library("geneplotter")
colors  <- densCols(imiss$logF_MISS,het$meanHet)
pdf("imiss-vs-het.pdf")
plot(imiss$logF_MISS,het$meanHet, col=colors, xlim=c(-3,0),ylim=c(0,0.5),pch=20, xlab="Proportion of missing genotypes",
ylab="Heterozygosity rate",axes=F)
axis(2,at=c(0,0.05,0.10,0.15,0.2,0.25,0.3,0.35,0.4,0.45,0.5),tick=T)
axis(1,at=c(-3,-2,-1,0),labels=c(0.001,0.01,0.1,1))
abline(h=mean(het$meanHet)-(4*sd(het$meanHet)),col="RED",lty=2)
abline(h=mean(het$meanHet)+(4*sd(het$meanHet)),col="RED",lty=2)
abline(v=-1.522879, col="RED", lty=2)
dev.off()
```

3.3) Remove individuals with > 0.03 missing markers (PLINK2)



a) Oxford

```
bin/plink2.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford/Markers/Oxford_marker --mind 0.03 --make-bed --out Oxford_marker_ind

PLINK v2.00a3 32-bit (9 Apr 2020)          www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford_marker_ind.log.
Options in effect:
```

```
 --bfile MONOCYTE_eQTLs/genotypes/Oxford/Markers/Oxford_marker
 --make-bed
 --mind 0.03
 --out Oxford_marker_ind

Start time: Thu May 07 22:52:56 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
Allocated 1296 MiB successfully, after larger attempt(s) failed.
Using up to 8 compute threads.
288 samples (0 females, 0 males, 288 ambiguous; 288 founders) loaded from
MONOCYTE_eQTLs/genotypes/Oxford/Markers/Oxford_marker.fam.
639169 variants loaded from
MONOCYTE_eQTLs/genotypes/Oxford/Markers/Oxford_marker.bim.
Note: No phenotype data present.
Calculating sample missingness rates... done.
1 sample removed due to missing genotype data (--mind).
ID written to Oxford_marker_ind.mindrem.id .
287 samples (0 females, 0 males, 287 ambiguous; 287 founders) remaining after
main filters.
Writing Oxford_marker_ind.fam ... done.
Writing Oxford_marker_ind.bim ... done.
Writing Oxford_marker_ind.bed ... done.
End time: Thu May 07 22:52:58 2020
```

Removed sample ID: 82

b) Oxford2 (ega)

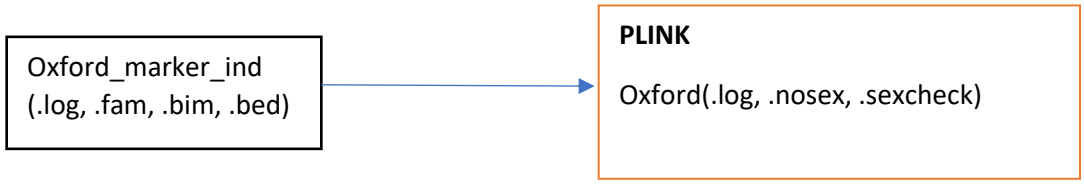
bin/plink2.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford2/Markers/Oxford2_marker --mind 0.03 --make-bed --out Oxford2_marker_ind

```
PLINK v2.00a3 32-bit (9 Apr 2020)           www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford2_marker_ind.log.
Options in effect:
 --bfile MONOCYTE_eQTLs/genotypes/Oxford2/Markers/Oxford2_marker
 --make-bed
 --mind 0.03
 --out Oxford2_marker_ind

Start time: Thu May 07 22:59:00 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
Allocated 1296 MiB successfully, after larger attempt(s) failed.
Using up to 8 compute threads.
144 samples (0 females, 0 males, 144 ambiguous; 144 founders) loaded from
MONOCYTE_eQTLs/genotypes/Oxford2/Markers/Oxford2_marker.fam.
638288 variants loaded from
MONOCYTE_eQTLs/genotypes/Oxford2/Markers/Oxford2_marker.bim.
Note: No phenotype data present.
Calculating sample missingness rates... done.
0 samples removed due to missing genotype data (--mind).
144 samples (0 females, 0 males, 144 ambiguous; 144 founders) remaining after
main filters.
Writing Oxford2_marker_ind.fam ... done.
Writing Oxford2_marker_ind.bim ... done.
Writing Oxford2_marker_ind.bed ... done.
End time: Thu May 07 22:59:01 2020
```

0 samples removed

3.4) Check sex concordance (PLINK2)



a) Oxford

bin/plink.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford/mind_0.03/Oxford_marker_ind --check-sex --out Oxford

```
PLINK v1.90b6.17 32-bit (28 Apr 2020)       www.cog-genomics.org/plink/1.9/
```

```
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford.log.
Options in effect:
  --bfile MONOCYTE_eQTLs/genotypes/Oxford/mind_0.03/Oxford_marker_ind
  --check-sex
  --out Oxford

16168 MB RAM detected; reserving 2047 MB for main workspace.
Allocated 1535 MB successfully, after larger attempt(s) failed.
639169 variants loaded from .bim file.
287 people (0 males, 0 females, 287 ambiguous) loaded from .fam.
Ambiguous sex IDs written to Oxford.nosex .
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 287 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Warning: Nonmissing nonmale Y chromosome genotype(s) present; many commands
treat these as missing.
Total genotyping rate is 0.998886.
639169 variants and 287 people pass filters and QC.
Note: No phenotypes present.
--check-sex: 15054 Xchr and 0 Ychr variant(s) scanned, 287 problems detected.
Report written to Oxford.sexcheck .
```

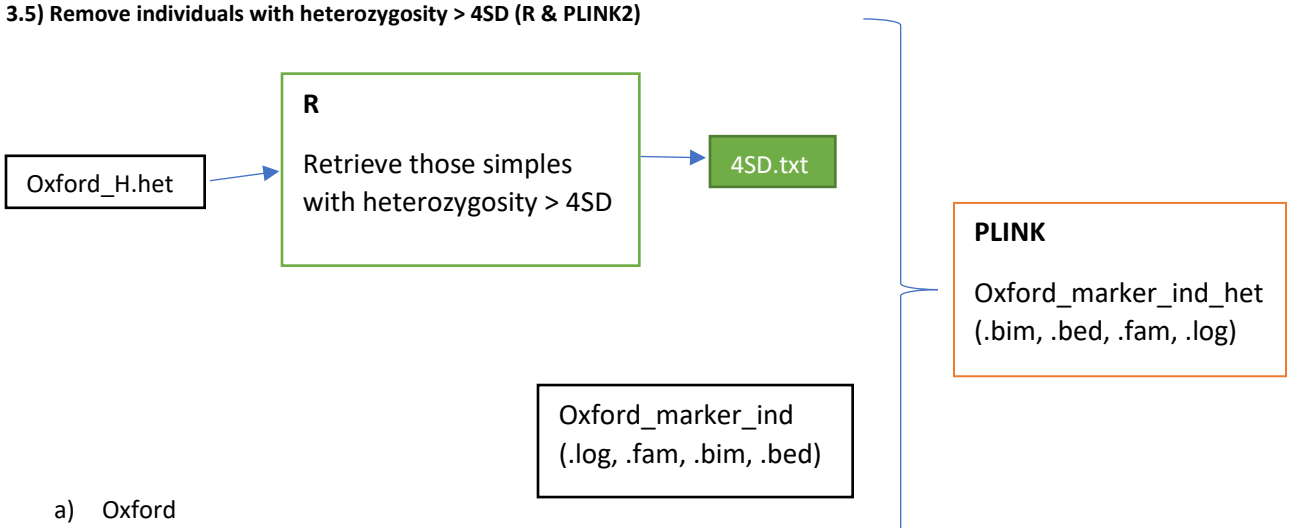
b) Oxford2 (ega)

```
bin/plink.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford2/mind_0.03/Oxford2_marker_ind --check-sex  --out Oxford2
```

```
PLINK v1.90b6.17 32-bit (28 Apr 2020)        www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford2.log.
Options in effect:
  --bfile MONOCYTE_eQTLs/genotypes/Oxford2/mind_0.03/Oxford2_marker_ind
  --check-sex
  --out Oxford2

16168 MB RAM detected; reserving 2047 MB for main workspace.
Allocated 1535 MB successfully, after larger attempt(s) failed.
638288 variants loaded from .bim file.
144 people (0 males, 0 females, 144 ambiguous) loaded from .fam.
Ambiguous sex IDs written to Oxford2.nosex .
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 144 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Warning: Nonmissing nonmale Y chromosome genotype(s) present; many commands
treat these as missing.
Total genotyping rate is 0.999379.
638288 variants and 144 people pass filters and QC.
Note: No phenotypes present.
--check-sex: 14922 Xchr and 0 Ychr variant(s) scanned, 144 problems detected.
Report written to Oxford2.sexcheck .
```

3.5) Remove individuals with heterozygosity > 4SD (R & PLINK2)



a) Oxford

First, we need to select those individuals, which have a heterozygosity mean higher than 4SD. For this I performed these commands in R:

```
het <- read.table("MONOCYTE_eQTLs/genotypes/Oxford/Heterozygosity/Oxford_H.het", header=TRUE)
het$meanHet <- (het$N.NM. - het$O.HOM.)/het$N.NM.
SD <- sd(het$meanHet)
mean_Het <- mean(het$meanHet)
upper <- (SD*4)+mean_Het
lower <- mean_Het - (SD*4)
which(lower > het$meanHet) #0 results
data <- data.frame(row.names=which(upper < het$meanHet), which(upper < het$meanHet))
write.table(x = data, file = "4SD.txt", sep = "\t", row.names=TRUE, col.names=TRUE)
```

```
bin/plink2.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford/mind_0.03/Oxford_marker_ind --remove 4SD.txt --make-bed --out Oxford_marker_ind_het
```

```
PLINK v2.00a3 32-bit (9 Apr 2020)          www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford_marker_ind_het.log.
Options in effect:
  --bfile MONOCYTE_eQTLs/genotypes/Oxford/mind_0.03/Oxford_marker_ind
  --make-bed
  --out Oxford_marker_ind_het
  --remove 4SD.txt

Start time: Fri May 08 10:21:23 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
Allocated 1296 MiB successfully, after larger attempt(s) failed.
Using up to 8 compute threads.
287 samples (0 females, 0 males, 287 ambiguous; 287 founders) loaded from
MONOCYTE_eQTLs/genotypes/Oxford/mind_0.03/Oxford_marker_ind.fam.
639169 variants loaded from
MONOCYTE_eQTLs/genotypes/Oxford/mind_0.03/Oxford_marker_ind.bim.
Note: No phenotype data present.
--remove: 284 samples remaining.
284 samples (0 females, 0 males, 284 ambiguous; 284 founders) remaining after
main filters.
Writing Oxford_marker_ind_het.fam ... done.
Writing Oxford_marker_ind_het.bim ... done.
Writing Oxford_marker_ind_het.bed ... done.
End time: Fri May 08 10:21:25 2020
```

Removed sample ID: 14, 82, 165, 215

a) Oxford2 (ega)

First, we need to select those individuals, which have a heterozygosity mean higher than 4SD. For this I performed these commands in R, but we saw that there were not any individual exceeding the threshold. Even this, we performed –make-bed step, to keep track with the name of the files.

```
het <- read.table("MONOCYTE_eQTLs/genotypes/Oxford2/Heterozygosity/Oxford2_H.het", header=TRUE)
het$meanHet <- (het$N.NM. - het$O.HOM.)/het$N.NM.
SD <- sd(het$meanHet)
mean_Het <- mean(het$meanHet)
upper <- (SD*4)+mean_Het
lower <- mean_Het - (SD*4)
which(lower > het$meanHet) #0 results
which(upper < het$meanHet) #0 results
```

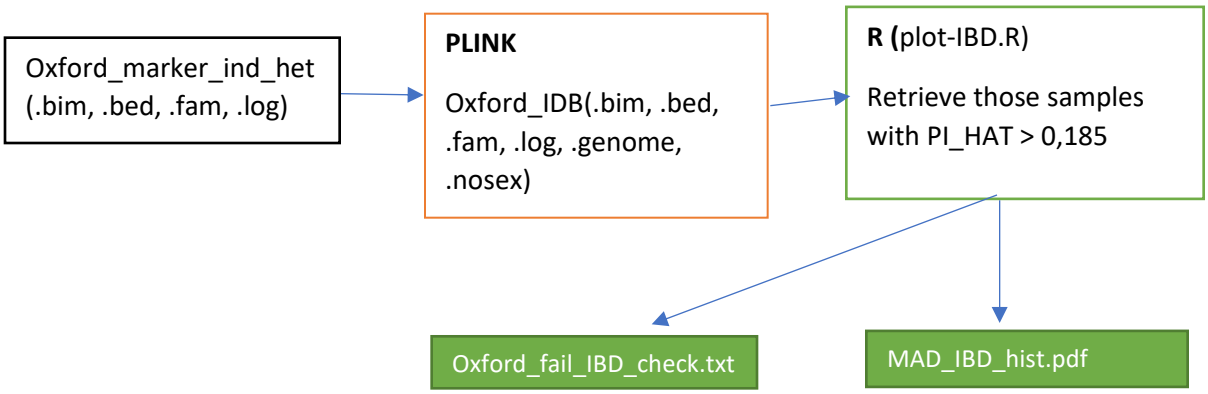
```
bin/plink2.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford2/mind_0.03/Oxford2_marker_ind --make-bed --out Oxford2_marker_ind_het
```

```
PLINK v2.00a3 32-bit (9 Apr 2020)          www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford2_marker_ind_het.log.
Options in effect:
  --bfile MONOCYTE_eQTLs/genotypes/Oxford2/mind_0.03/Oxford2_marker_ind
  --make-bed
  --out Oxford2_marker_ind_het

Start time: Fri May 08 10:42:53 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
Allocated 1296 MiB successfully, after larger attempt(s) failed.
Using up to 8 compute threads.
144 samples (0 females, 0 males, 144 ambiguous; 144 founders) loaded from
MONOCYTE_eQTLs/genotypes/Oxford2/mind_0.03/Oxford2_marker_ind.fam.
```

```
638288 variants loaded from
MONOCYTE_eQTLs/genotypes/Oxford2/mind_0.03/Oxford2_marker_ind.bim.
Note: No phenotype data present.
Writing Oxford2_marker_ind_het.fam ... done.
Writing Oxford2_marker_ind_het.bim ... done.
Writing Oxford2_marker_ind_het.bed ... done.
End time: Fri May 08 10:42:54 2020
```

3.6) Calculate relatedness by IDB (PLINK2 & R)



a) Oxford

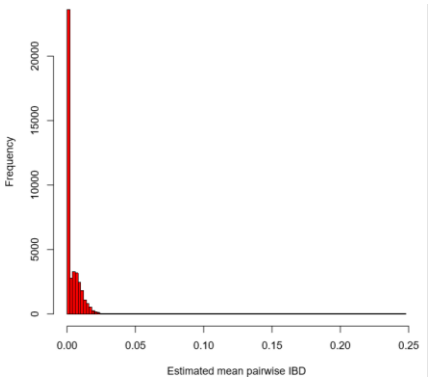
The first step was to calculate the relatedness by IDB with PLINK:

```
bin/plink.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford/Check_Het/Oxford_marker_ind_het --genome --make-bed --out Oxford_IBD
```

```
PLINK v1.90b6.17 32-bit (28 Apr 2020)         www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford_IBD.log.
Options in effect:
  --bfile MONOCYTE_eQTLs/genotypes/Oxford/Check_Het/Oxford_marker_ind_het
  --genome
  --make-bed
  --out Oxford_IBD

16168 MB RAM detected; reserving 2047 MB for main workspace.
Allocated 1535 MB successfully, after larger attempt(s) failed.
639169 variants loaded from .bim file.
284 people (0 males, 0 females, 284 ambiguous) loaded from .fam.
Ambiguous sex IDs written to Oxford_IBD.nosex .
Using up to 8 threads (change this with --threads).
Before main variant filters, 284 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Warning: Nonmissing nonmale Y chromosome genotype(s) present; many commands
treat these as missing.
Total genotyping rate is 0.998879.
639169 variants and 284 people pass filters and QC.
Note: No phenotypes present.
--make-bed to Oxford_IBD.bed + Oxford_IBD.bim + Oxford_IBD.fam ... done.
Excluding 16657 variants on non-autosomes from IBD calculation.
IBD calculations complete.
Finished writing Oxford_IBD.genome .
```

Once calculated the relatedness we plot the estimated mean pairwise IDB by R (plot-IDB.R)




```
data=read.table("MONOCYTE_eQTLs/genotypes/Oxford/IDB/Oxford_IBD.genome",h=T)
pdf("MAD_IBD_hist.pdf")
hist(data$PI_HAT,col="RED",breaks=100,xlab="Estimated mean pairwise IBD",main="")
dev.off()
table(data$PI_HAT>0,185) # we didn't obtain any sample
```

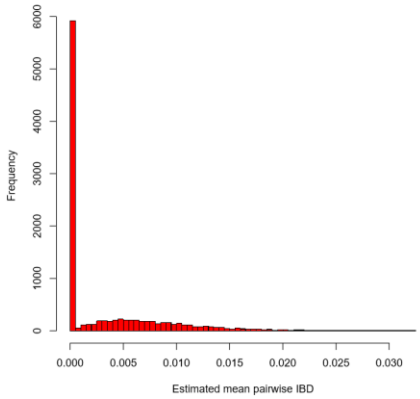
b) Oxford2 (ega)

bin/plink.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford2/Check_Het/Oxford2_marker_ind_het --genome --make-bed --out Oxford2_IBD

```
PLINK v1.90b6.17 32-bit (28 Apr 2020)          www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford2_IBD.log.
Options in effect:
  --bfile MONOCYTE_eQTLs/genotypes/Oxford2/Check_Het/Oxford2_marker_ind_het
  --genome
  --make-bed
  --out Oxford2_IBD

16168 MB RAM detected; reserving 2047 MB for main workspace.
Allocated 1535 MB successfully, after larger attempt(s) failed.
638288 variants loaded from .bim file.
144 people (0 males, 0 females, 144 ambiguous) loaded from .fam.
Ambiguous sex IDs written to Oxford2_IBD.nosex .
Using up to 8 threads (change this with --threads).
Before main variant filters, 144 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Warning: Nonmissing nonmale Y chromosome genotype(s) present; many commands
treat these as missing.
Total genotyping rate is 0.999379.
638288 variants and 144 people pass filters and QC.
Note: No phenotypes present.
--make-bed to Oxford2_IBD.bed + Oxford2_IBD.bim + Oxford2_IBD.fam ... done.
Excluding 16288 variants on non-autosomes from IBD calculation.
IBD calculations complete.
Finished writing Oxford2_IBD.genome .
```

Once calculated the relatedness we plot the estimated mean pairwise IDB by R (plot-IDB.R)




```
data=read.table("MONOCYTE_eQTLs/genotypes/Oxford2/IDB/Oxford2_IBD.genome",h=T)
pdf("MAD_IBD_hist2.pdf")
hist(data$PI_HAT,col="RED",breaks=100,xlab="Estimated mean pairwise IBD",main="")
```

```
dev.off()
table(data$PI_HAT>0,185) # we obtain one result: samples 16 and 111
out<-data[data$PI_HAT>0.185,]
write.table(out,"Oxford2_fail_IBD_check.txt",sep="\t", quote=F, row.names=F)
```

WE WILL NOT REMOVE THE SAMPLES RELATED 16 AND 111, UNTIL WE DON'T CALCULATE THE RELATNESS ACROSS THE TWO FILES.

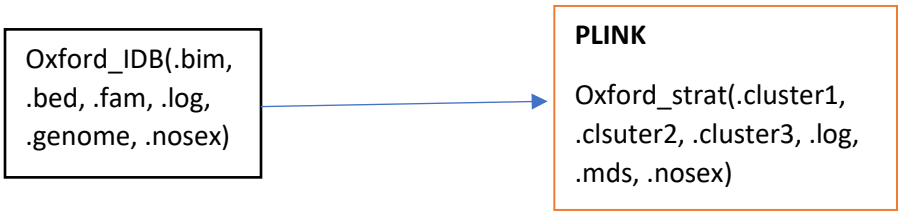
3.7) Ethnicity (R)

With the metadata file .txt from ArrayExpress, we saw that all the samples had the same ethnicity (European).

```
table <- read.table("E-MTAB-2232.txt", header=TRUE, sep = "\t")
table(table$Characteristics.ethnicity.)

European
   1449
```

3.8) Calculate 4PCs (MDS) for downstream analysis association studies (PLINK2)



a) Oxford

```
bin/plink.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford/IDB/Oxford_IBD --cluster --mds-plot 4 --K 2 --out Oxford_strat

PLINK v1.90b6.17 32-bit (28 Apr 2020)        www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford_strat.log.
Options in effect:
  --K 2
  --bfile MONOCYTE_eQTLs/genotypes/Oxford/IDB/Oxford_IBD
  --cluster
  --mds-plot 4
  --out Oxford_strat

16168 MB RAM detected; reserving 2047 MB for main workspace.
Allocated 1535 MB successfully, after larger attempt(s) failed.
639169 variants loaded from .bim file.
284 people (0 males, 0 females, 284 ambiguous) loaded from .fam.
Ambiguous sex IDs written to Oxford_strat.nosex .
Using up to 8 threads (change this with --threads).
Before main variant filters, 284 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Warning: Nonmissing nonmale Y chromosome genotype(s) present; many commands
treat these as missing.
Total genotyping rate is 0.998879.
639169 variants and 284 people pass filters and QC.
Note: No phenotypes present.
Excluding 16657 variants on non-autosomes from distance matrix calc.
Distance matrix calculation complete.
Clustering... done.
Cluster solution written to Oxford_strat.cluster1 , Oxford_strat.cluster2 , and
Oxford_strat.cluster3 .
Performing multidimensional scaling analysis (SVD algorithm, 4
dimensions)... done.
MDS solution written to Oxford_strat.mds .
```

Analysing the clusters created with these samples, we saw that the majority of them (276) were clustered together in cluster 0, while the rest (8 samples), were clustered in cluster 1. The samples from cluster 1 were: 5, 29, 32, 75, 83, 87, 175 and 213.

b) Oxford2 (ega)

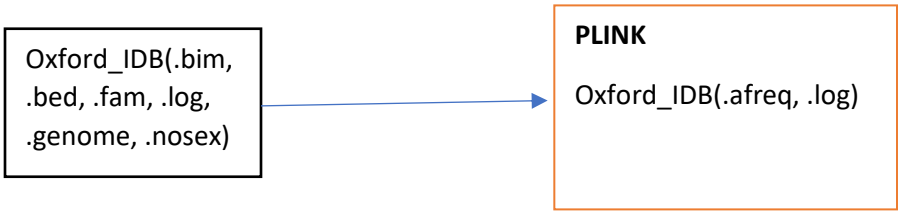
bin/plink.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford2/IDB/Oxford2_IBD --cluster --mds-plot 4 --K 2 --out Oxford2_strat

```
PLINK v1.90b6.17 32-bit (28 Apr 2020)        www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford2_strat.log.
Options in effect:
  --K 2
  --bfile MONOCYTE_eQTLs/genotypes/Oxford2/IDB/Oxford2_IBD
  --cluster
  --mds-plot 4
  --out Oxford2_strat

16168 MB RAM detected; reserving 2047 MB for main workspace.
Allocated 1535 MB successfully, after larger attempt(s) failed.
638288 variants loaded from .bim file.
144 people (0 males, 0 females, 144 ambiguous) loaded from .fam.
Ambiguous sex IDs written to Oxford2_strat.nosex .
Using up to 8 threads (change this with --threads).
Before main variant filters, 144 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Warning: Nonmissing nonmale Y chromosome genotype(s) present; many commands
treat these as missing.
Total genotyping rate is 0.999379.
638288 variants and 144 people pass filters and QC.
Note: No phenotypes present.
Excluding 16288 variants on non-autosomes from distance matrix calc.
Distance matrix calculation complete.
Clustering... done.
Cluster solution written to Oxford2_strat.cluster1 , Oxford2_strat.cluster2 ,
and Oxford2_strat.cluster3 .
Performing multidimensional scaling analysis (SVD algorithm, 4
dimensions)... done.
MDS solution written to Oxford2_strat.mds .
```

Analysing the clusters created with these samples, we saw that the majority of them (136) were clustered together in cluster 0, while the rest (8 samples), were clustered in cluster 1. The samples from cluster 1 were: 291, 319, 353, 372, 377, 382, 397 and 401.

3.9) Calculate frequency file for imputation (PLINK2)



a) Oxford

bin/plink2.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford/IDB/Oxford_IBD --freq --out Oxford_IBD

```
PLINK v2.00a3 32-bit (9 Apr 2020)          www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford_IBD.log.
Options in effect:
  --bfile MONOCYTE_eQTLs/genotypes/Oxford/IDB/Oxford_IBD
  --freq
  --out Oxford_IBD

Start time: Fri May 08 13:39:11 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
Allocated 1296 MiB successfully, after larger attempt(s) failed.
```

```
Using up to 8 compute threads.
284 samples (0 females, 0 males, 284 ambiguous; 284 founders) loaded from
MONOCYTE_eQTLs/genotypes/Oxford/IDB/Oxford_IBD.fam.
639169 variants loaded from MONOCYTE_eQTLs/genotypes/Oxford/IDB/Oxford_IBD.bim.
Note: No phenotype data present.
Calculating allele frequencies... done.
--freq: Allele frequencies (founders only) written to Oxford_IBD.afreq .
End time: Fri May 08 13:39:13 2020
```

b) Oxford2 (ega)

bin/plink2.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford2/IDB/Oxford2_IBD --freq --out Oxford2_IBD

```
PLINK v2.00a3 32-bit (9 Apr 2020)          www.cog-genomics.org/plink/2.0/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Oxford2_IBD.log.
Options in effect:
 --bfile MONOCYTE_eQTLs/genotypes/Oxford2/IDB/Oxford2_IBD
 --freq
 --out Oxford2_IBD

Start time: Fri May 08 13:42:00 2020
16168 MiB RAM detected; reserving 1728 MiB for main workspace.
Allocated 1296 MiB successfully, after larger attempt(s) failed.
Using up to 8 compute threads.
144 samples (0 females, 0 males, 144 ambiguous; 144 founders) loaded from
MONOCYTE_eQTLs/genotypes/Oxford2/IDB/Oxford2_IBD.fam.
638288 variants loaded from
MONOCYTE_eQTLs/genotypes/Oxford2/IDB/Oxford2_IBD.bim.
Note: No phenotype data present.
Calculating allele frequencies... done.
--freq: Allele frequencies (founders only) written to Oxford2_IBD.afreq .
End time: Fri May 08 13:42:01 2020
```

4) PREIMPUTATION CHECK (Will Rayner's perl script v4.2.11)

a) Oxford

perl HRC-1000G-check-bim.pl -b MONOCYTE_eQTLs/genotypes/Oxford/IDB/Oxford_IBD.bim -f MONOCYTE_eQTLs/genotypes/Oxford/Freq/Oxford_IBD.afreq -r HRC.r1-1.GRCh37.wgs.mac5.sites.tab -h -v

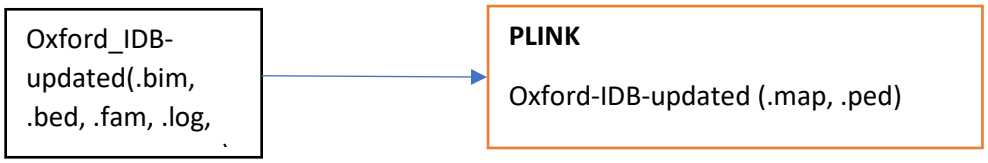
632444 variants and 284 people pass the filters and QC. → Oxford_IBD-updated.log

b) Oxford2 (ega)

perl HRC-1000G-check-bim.pl -b MONOCYTE_eQTLs/genotypes/Oxford2/IDB/Oxford2_IBD.bim -f MONOCYTE_eQTLs/genotypes/Oxford2/Freq/Oxford2_IBD.afreq -r HRC.r1-1.GRCh37.wgs.mac5.sites.tab -h -v

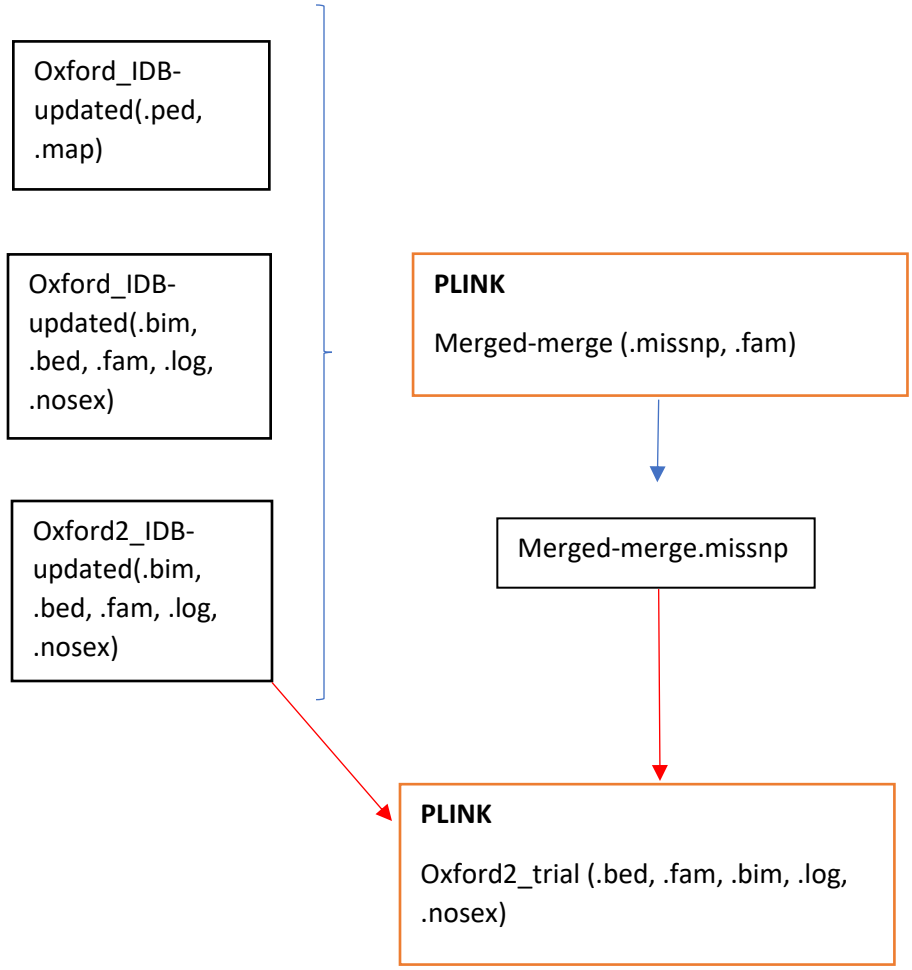
632488 variants and 144 people pass filters and QC. → Oxford2_IBD-updated.log

5) MERGE COHORT



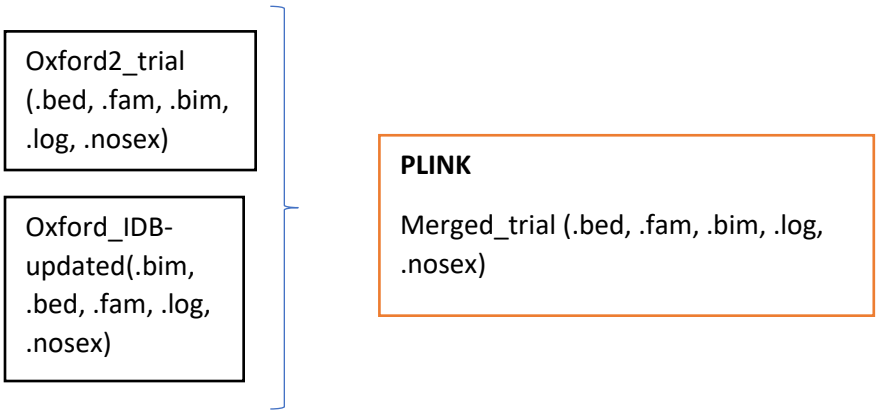
5.1) Convert one of the data sets to map/ped

bin/plink.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford/Will_Rayner/Oxford_IBD-updated --recode --out Oxford_IBD-updated



5.2) Due to strandness mismatches between the files, we take the Oxford2 (ega) and we changed it doing a flip of the SNPs with the error

bin/plink.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford2/Will_Rayner/Oxford2_IBD-updated –flip Merged-merge.missnp --make-bed --out Oxford2_trial

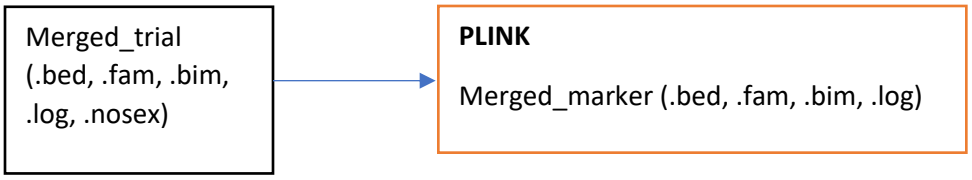


5.3) Merge files

bin/plink.exe --bfile MONOCYTE_eQTLs/genotypes/Oxford/Will_Rayner/Oxford_IBD-updated --bmerge Oxford2_trial --make-bed --out merged_trial

637935 variants and 428 people pass filters and QC.

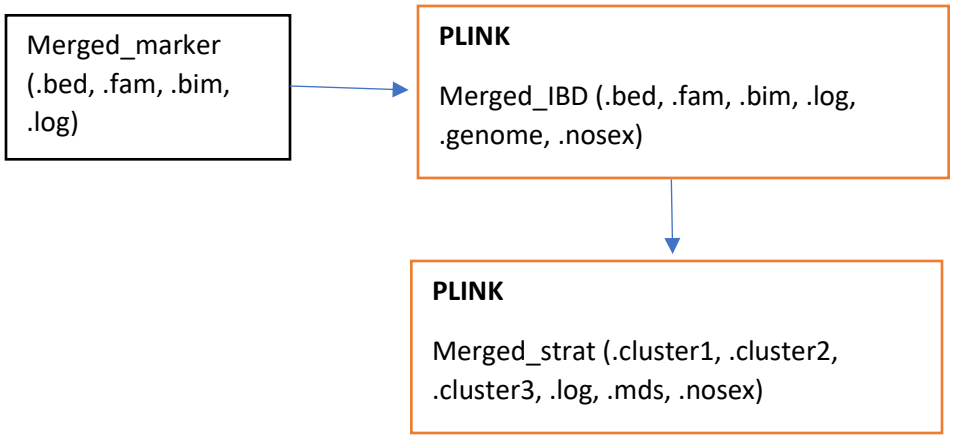
5.4) Remove markers according to QC thresholds



bin/plink2.exe --bfile MONOCYTE_eQTLs/genotypes/Merge_files/merged_trial --geno 0.05 --hwe 1e-06 --maf 0.01 --make-bed --out merged_marker

626858 variants remaining after main filters

5.5) Perform PCA and calculate relatedness by IBD



bin/plink.exe --bfile MONOCYTE_eQTLs/genotypes/Merge_files_markers/merged_marker --genome --make-bed --out Merged_IBD

bin/plink.exe --bfile MONOCYTE_eQTLs/genotypes/Merge_files_markers/merged_marker --cluster --read-genome Merged_IBD.genome --mds-plot 4 --K 2 --out Merge_strat

We found 12 samples related with a PI-HAT higher than 0,185. In front of this and to avoid the redundancy of the data, we delete one of each pair from the total set. To decide which one, I decided to look for the number of missing hardcalls in every sample in the files .smiss:

111 vs 16
awk '{if ($1 == 111) print $3}' MONOCYTE_eQTLs/genotypes/Oxford/Miss_genotype/merge_MISS_Oxfo
rd.smiss
600

awk '{if ($1 == 16) print $3}' MONOCYTE_eQTLs/genotypes/Oxford/Miss_genotype/merge_MISS_Oxfor
d.smiss
481

17 vs 320
awk '{if ($1 == 17) print $3}' MONOCYTE_eQTLs/genotypes/Oxford/Miss_genotype/merge_MISS_Oxfor

d.smiss
236

awk '{if ($1 == 320) print $3}' MONOCYTE_eQTLs/genotypes/Oxford2/Miss_genotype/merge_MISS_Oxf
ord2.smiss
1244

75 vs 291
awk '{if ($1 == 75) print $3}' MONOCYTE_eQTLs/genotypes/Oxford/Miss_genotype/merge_MISS_Oxfor
d.smiss
423

awk '{if ($1 == 291) print $3}' MONOCYTE_eQTLs/genotypes/Oxford2/Miss_genotype/merge_MISS_Oxf
ord2.smiss
1296

115 vs 375
awk '{if ($1 == 115) print $3}' MONOCYTE_eQTLs/genotypes/Oxford/Miss_genotype/merge_MISS_Oxfo
rd.smiss
369

awk '{if ($1 == 375) print $3}' MONOCYTE_eQTLs/genotypes/Oxford2/Miss_genotype/merge_MISS_Oxf
ord2.smiss
1386

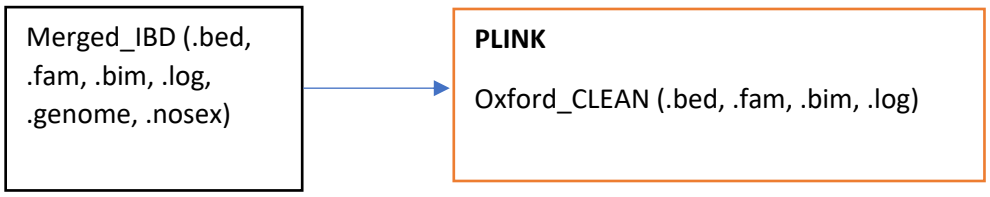
258 vs 386
awk '{if ($1 == 258) print $3}' MONOCYTE_eQTLs/genotypes/Oxford/Miss_genotype/merge_M
ISS_Oxford.smiss
850

awk '{if ($1 == 386) print $3}' MONOCYTE_eQTLs/genotypes/Oxford2/Miss_genotype/merge_MISS_Oxf
ord2.smiss
1321

275 vs 389
awk '{if ($1 == 275) print $3}' MONOCYTE_eQTLs/genotypes/Oxford/Miss_genotype/merge_MISS_Oxfo
rd.smiss
486

awk '{if ($1 == 389) print $3}' MONOCYTE_eQTLs/genotypes/Oxford2/Miss_genotype/merge_MISS_Oxf
ord2.smiss
1097

We removed samples 111, 320, 291, 375, 386 and 389



bin/plink2.exe -bfile MONOCYTE_eQTLs/genotypes/Merge_IBD/Merged_IBD --remove remove.txt --make-bed --out Oxford_CLEAN

The cluster of the samples (k=2), grouped the majority of them in group 0 (417 samples), and there was 5 of the total that were grouped in cluster 1: 28, 81, 210, 373 and 378.