# BICF Nano Course: Determining Population Structure and Genetic Relationships using PLINK

## Estimate Principal Components

1. Check the data

```
cd course2
ls
```

```
all.sh    gwa.hapmap.shared.bed  hapmap.bed
bin       gwa.hapmap.shared.bim  hapmap.bim
gwa.bed   gwa.hapmap.shared.fam  hapmap.fam
gwa.bim   gwa.sample             hapmap.sample.grp
gwa.fam   gwa.sample.grp         install.course2.sh
```

```
head gwa.bim
```

```
1       rs3934834       0       995669  T       C
1       rs3737728       0       1011278 A       G
1       rs6687776       0       1020428 T       C
1       rs9651273       0       1021403 A       G
1       rs4970405       0       1038818 G       A
1       rs12726255      0       1039813 G       A
1       rs2298217       0       1054842 T       C
1       rs4970357       0       1066927 C       A
1       rs4970362       0       1084601 A       G
1       rs9660710       0       1089205 A       C
```

```
head gwa.fam
```

```
0 A2001 0 0 1 2
1 A2002 0 0 1 2
2 A2003 0 0 1 2
3 A2004 0 0 1 2
4 A2005 0 0 1 2
5 A2006 0 0 1 2
6 A2007 0 0 1 2
7 A2008 0 0 1 2
8 A2009 0 0 1 2
9 A2010 0 0 1 2
```

2. Estimate principal component

```
bin/plink --bfile gwa.hapmap.shared --pca --out gwa.hapmap.shared
```

```
less -S gwa.hapmap.shared.eigenvec
```

```
0 A2001 0.0190776 0.0312478 -0.0251495 0.00739507 -0.00275138 -0.0388116 -
1 A2002 0.0196351 0.0304162 -0.0262818 0.00406776 -0.00237609 -0.0405913 0
2 A2003 0.0189446 0.0309258 -0.0253458 0.00758003 -0.00670025 -0.042192 0.
3 A2004 0.0199457 0.0305012 -0.0253393 0.00919678 -0.00604249 -0.0375134 -
4 A2005 0.0189858 0.030917 -0.0230897 0.00746736 -0.00350501 -0.0400103 0.
5 A2006 0.0196612 0.0312536 -0.0226719 0.0070045 -0.00159345 -0.0377338 -0
6 A2007 0.0193189 0.0310956 -0.025389 0.00726291 -0.00215933 -0.0364316 0.
7 A2008 0.0196107 0.030796 -0.0238714 0.00717875 -0.00175065 -0.0396749 -0
8 A2009 0.0199533 0.0309213 -0.0261653 0.00734643 -0.00254497 -0.0358469 0
9 A2010 0.0195248 0.0309607 -0.0284516 0.00513481 -0.000732567 -0.0384615
10 A2011 0.0197376 0.0314058 -0.0270225 0.00488164 -0.00246897 -0.0429084
11 A2012 0.0189865 0.0309724 -0.02227 0.00718675 -0.00227535 -0.0426462 -0
12 A2013 0.0201447 0.030883 -0.0277418 0.00743795 -0.00502048 -0.0400874 -
```

3. Combine PCs and ethnic groups

```
   sort -k2,2 gwa.hapmap.shared.eigenvec >
gwa.hapmap.shared.eigenvec.sort
```

```
   join -1 1 -2 2 gwa.hapmap.sample.grp gwa.hapmap.shared.eigenvec.sort >
gwa.hapmap.shared.eigenvec.sort.grp
```

```
   less -S gwa.hapmap.shared.eigenvec.sort.grp
```

```
A1901 1 1900 0.0195706 0.0309059 -0.025932 0.00599042 -0.00291579 -
A1902 1 1901 0.0194568 0.0313669 -0.0251413 0.00608022 -0.00185091
A1903 1 1902 0.0196991 0.0312037 -0.0272494 0.00550248 -0.00150931
A1904 1 1903 0.019516 0.0307047 -0.0219999 0.00584493 -0.00132622 -
A1905 1 1904 0.0195664 0.0308378 -0.024556 0.00572075 -0.00266382 -
A1906 1 1905 0.019653 0.0303152 -0.0265921 0.00623471 -0.00329396 -
A1907 1 1906 0.0196239 0.0313903 -0.0218344 0.00580631 -0.00213431
A1908 1 1907 0.0194955 0.0318596 -0.0222067 0.00679484 -0.00582083
A1909 1 1908 0.0194342 0.0310117 -0.025388 0.00929734 -0.00133433 -
A1910 1 1909 0.0198634 0.0306821 -0.0239332 0.00649145 -0.00382488
A1911 1 1910 0.0193187 0.0309021 -0.0258117 0.00647108 -0.00288309
A1912 1 1911 0.0196536 0.0307733 -0.0250979 0.00643052 -0.00116234
```

4. Plot PCs and ethnic groups for hapmap, cases and controls

```
   ./plot_pc_1.sh gwa.hapmap.shared.eigenvec.sort.grp
gwa.hapmap.shared.eigenvec.sort.grp.pdf
```

Population list:

http://www.internationalgenome.org/category/population/

5. Estimate PCs for gwa only

```
   bin/plink --bfile gwa --pca --out gwa
```

6. Combine PCs and case/control status

```
   sort -k2,2 gwa.eigenvec > gwa.eigenvec.sort
```

```
   sort -k1,1 gwa.sample.grp > gwa.sample.grp.sort
```

```
join -1 1 -2 2 gwa.sample.grp.sort gwa.eigenvec.sort > gwa.eigenvec.sort.grp
```

```
less -S gwa.eigenvec.sort.grp
```

```
A1901 1 1900 -0.0811105 -0.0457348 0.150845 -0.0819834 -0.111204
A1902 1 1901 0.202932 -0.0167122 -0.0437225 0.0374532 -0.0729355
A1903 1 1902 -0.0488828 -0.0866916 0.081276 -0.0481999 -0.039171
A1904 1 1903 -0.0208352 -0.0388757 -0.0526029 0.0631078 -0.06759
A1905 1 1904 0.0656299 -0.0827277 0.0373445 -0.0862979 -0.010340
A1906 1 1905 0.169268 0.0351219 0.0524578 0.0157024 -0.018702 -0
A1907 1 1906 0.00815148 -0.0576587 -0.0947834 -0.0431531 0.04542
A1908 1 1907 -0.0445181 0.114217 -0.0145049 -0.0675119 -0.080937
A1909 1 1908 -0.140595 0.127927 0.00427774 0.00921316 0.0346994
A1910 1 1909 0.0317326 0.089254 0.0367306 0.0362855 -0.0211359 -
A1911 1 1910 0.029694 -0.0635268 0.0387513 0.0759965 -0.0220129
A1912 1 1911 -0.0348471 0.110004 -0.014663 -0.136484 -0.0395563
```

7. Plot PCs for cases and controls

```
./plot_pc_2.sh gwa.eigenvec.sort.grp gwa.eigenvec.sort.grp.pdf
```

## Estimate Genetic Relationships

8. Estimate IBD for each pair of gwa samples

```
bin/plink --bfile gwa --genome --out gwa
```

```
less -S gwa.genome
```

| FID1 | IID1 | FID2 | IID2 | RT | EZ | Z0 | Z1 | Z2 | PI_HAT | PHE | DST |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | A2001 | 1 | A2002 | UN | NA | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 1 | 0.720609 |
| 0 | A2001 | 2 | A2003 | UN | NA | 0.9771 | 0.0229 | 0.0000 | 0.0115 | 1 | 0.721555 |
| 0 | A2001 | 3 | A2004 | UN | NA | 0.9720 | 0.0255 | 0.0025 | 0.0153 | 1 | 0.725143 |
| 0 | A2001 | 4 | A2005 | UN | NA | 0.9957 | 0.0000 | 0.0043 | 0.0043 | 1 | 0.722814 |
| 0 | A2001 | 5 | A2006 | UN | NA | 0.9770 | 0.0230 | 0.0000 | 0.0115 | 1 | 0.721296 |
| 0 | A2001 | 6 | A2007 | UN | NA | 0.9647 | 0.0353 | 0.0000 | 0.0176 | 1 | 0.723848 |
| 0 | A2001 | 7 | A2008 | UN | NA | 0.9987 | 0.0013 | 0.0000 | 0.0006 | 1 | 0.721635 |
| 0 | A2001 | 8 | A2009 | UN | NA | 0.9938 | 0.0062 | 0.0000 | 0.0031 | 1 | 0.722202 |
| 0 | A2001 | 9 | A2010 | UN | NA | 0.9857 | 0.0143 | 0.0000 | 0.0071 | 1 | 0.721925 |
| 0 | A2001 | 10 | A2011 | UN | NA | 0.9898 | 0.0000 | 0.0102 | 0.0102 | 1 | 0.723611 |
| 0 | A2001 | 11 | A2012 | UN | NA | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 1 | 0.721292 |
| 0 | A2001 | 12 | A2013 | UN | NA | 0.9677 | 0.0323 | 0.0000 | 0.0162 | 1 | 0.722853 |

9. Plot IBD

```
./plot_ibd.sh gwa.genome gwa.genome.png
```

10. Estimate IBD for each pair of hapmap samples

```
bin/plink --bfile hapmap --genome --out hapmap
```

11. Plot IBD

```
./plot_ibd.sh hapmap.genome hapmap.genome.png
```