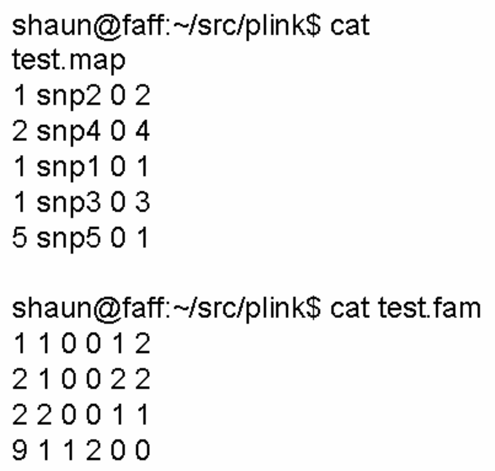
PLINK Support for Illumina WG Arrays
Date: May 23, 2007

Overview:
Described below is a process to input Illumina's HumanHap BeadChips data into PLINK to use in downstream WGA studies. The genotype data is originally created within BeadStudio and exported as a "Final Report" (default settings). The information provided below explains PLINK data architecture and how to convert the BeadStudio output into an input for PLINK.

The input data to PLINK consists of:

1) *.lgen file with genotypes: five columns, one genotype per line.
2) *.map file with SNP information: four columns, one SNP per line
3) *.fam file with person information: six columns, one person per line.

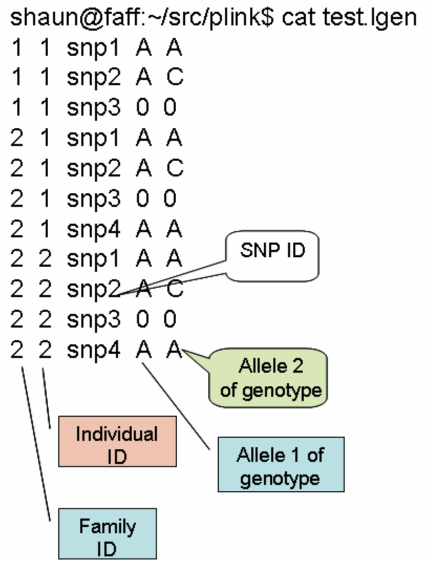
The MAP and FAM files are described on the main PLINK website (http://pngu.mgh.harvard.edu/~purcell/plink/). To create MAP and FAM files, remove the header rows and the extra columns (intensity values) – see examples below:



```
shaun@faff:~/src/plink$ cat
test.map
1 snp2 0 2
2 snp4 0 4
1 snp1 0 1
1 snp3 0 3
5 snp5 0 1

shaun@faff:~/src/plink$ cat test.fam
1 1 0 0 1 2
2 1 0 0 2 2
2 2 0 0 1 1
9 1 1 2 0 0
```

The LGEN file should be self explanatory (see example below)



```
shaun@faff:~/src/plink$ cat test.lgen
1  1  snp1  A  A
1  1  snp2  A  C
1  1  snp3  0  0
2  1  snp1  A  A
2  1  snp2  A  C
2  1  snp3  0  0
2  1  snp4  A  A
2  2  snp1  A  A
2  2  snp2  A  C
2  2  snp3  0  0
2  2  snp4  A  A
```

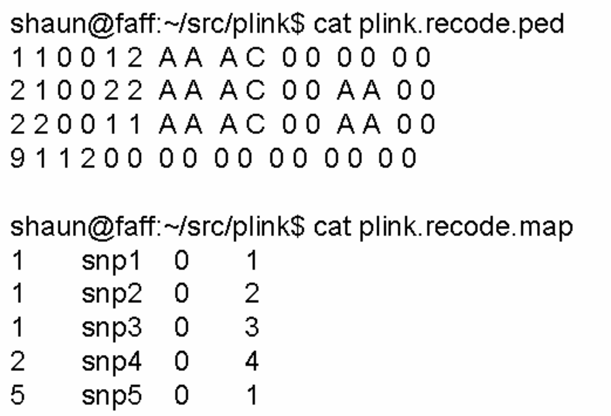
<u>Note:</u> all entries do not need to be present in the LGEN file (e.g. snp5 or person 9/1) or snp4 for person 1/1. These genotypes will be set to missing internally. The order also need not be the same in the LGEN file as for the MAP or FAM files. If a genotype is listed more than once, the final version of it will be used.

<u>Note:</u> All individuals must be uniquely identified by the combination of the family and individual IDs.

LGEN file can be reconstructed into a PED file using the following command:

```
plink --lfile test --recode
```

which creates these two files:



```
shaun@faff:~/src/plink$ cat plink.recode.ped
1 1 0 0 1 2  A A  A C  0 0  0 0  0 0
2 1 0 0 2 2  A A  A C  0 0  A A  0 0
2 2 0 0 1 1  A A  A C  0 0  A A  0 0
9 1 1 2 0 0  0 0  0 0  0 0  0 0  0 0

shaun@faff:~/src/plink$ cat plink.recode.map
1       snp1    0       1
1       snp2    0       2
1       snp3    0       3
2       snp4    0       4
5       snp5    0       1
```

<u>Note:</u> The --lfile command will be available in the next PLINK release, v1.00.