<http://zzz.bwh.harvard.edu/plink/dataman.shtml>

convert genotype file format

Try this first:

plink --file mydataname --recode12 --o mydataname\_allele12

try this second:

plink --file mydataname --allele1234 --make-bed

etc.

plink --file data --recode12

will recode the alleles as 1 and 2 (and the missing genotype will always be 0).

To recode SNP alleles from A,C,G,T to 1,2,3,4 or vice versa, use --allele1234 (to go from letters to numbers) and --alleleACGT (to go from numbers to letters). These flags should be used in conjunction with a data generation command (e.g. --make-bed), or any other analysis or summary statistic option. Alleles other than A,C,G,T or 1,2,3,4 will be left unchanged.

Unless manually specified, for all these options, the usual filters for missingness and allele frequency will be set so as not to exclude any SNPs or individuals. By explicitly including an option, e.g. --maf 0.05 on the command line, this behaviour is overriden (see [this page](http://zzz.bwh.harvard.edu/plink/thresh.shtml)).

Genotypes (column 7 onwards) should also be white-space delimited; they can be any character (e.g. 1,2,3,4 or A,C,G,T or anything else) except 0 which is, by default, the missing genotype character. **All markers should be biallelic**. All SNPs (whether haploid or not) must have two alleles specified. Either Both alleles should be missing (i.e. 0) or neither. No header row should be given. For example, here are two individuals typed for 3 SNPs (one row = one person):

FAM001 1 0 0 1 2 A A G G A C

FAM001 2 0 0 1 2 A A A G 0 0

...

The default missing genotype character can be changed with the --missing-genotype option, for example:

plink --file mydata --missing-genotype N

**A PED file must have 1 and only 1 phenotype in the sixth column.**

###### Different PED file formats: missing fields

Sometimes data arrive in a number of different formats: for example, where the genotype information just has a single ID column followed by all the SNP data, with the other family and phenotype information residing in a separate file. Rather than have to recreate new files, it is sometimes possible to read in such files directly. The standard behavior of PLINK when reading a PED file with --file or --ped can be modified to allow for the fact that one or more of the normally obligatory 6 fields are missing:

##### --no-fid

indicates there is no Family ID column: here the first field is taken to be individual ID, and the family ID is automatically set to be the same as the individual ID (i.e. obviously, all individuals would be treated as unrelated). In other files that require family and individual ID (e.g. alternate phenotype file and cluster files, for which this flag has no effect), the individual ID would need to be entered also as the family ID therefore.

##### --no-parents

indicates that there are no paternal and maternal ID codes; all individuals would be assumed to be founders in this case

##### --no-sex

indicates that there is no sex field; all individuals set to have a missing sex code (which also sets that individual to missing unless the allow-no-sex option is also used)

##### --no-pheno

indicates that there is no phenotype filed; all individuals are set to missing unless an [alternate phenotype file](http://zzz.bwh.harvard.edu/plink/data.shtml#pheno) is specified.