

	DESCRIPTION	SOFTWARE REQUIRED	INPUT FILE FORMAT	OUTPUT FILE FORMAT	CONTENT OUTPUT FILE
STEP 1	Exclude from the analysis SNPs that do not have IDs of the type 'rsXXXX' Imputation step	QCTOOL	.bim .cal .bgen .sample	.gen .log	Raw SNPs data, by chromosome
STEP 2	Convert .gen into .ped and .map	GTOOL	.gen	.ped .map	Raw SNPs data, by chromosome
STEP 3	Exclude duplicate SNPs, subjects without ambiguous sex and subjects not meeting inclusion criteria	PLINK	.ped .map list_ID_to_delete.txt (containing IDs of subjects not meeting inclusion criteria)	.bed .bim .fam	Duplicates-free raw SNPs data, by chromosome with only subjects included in the study
STEP 4	Filter SNPs based on MAF, call rate, LD and HWE Filter subjects based on call rate	PLINK	.bed .bim .fam	.bed .bim .fam	Duplicates-free raw SNPs data, by chromosome with only subjects included in the study
STEP 5	Merge chromosome files into one	PLINK	.bed .bim .fam	.bed .bim .fam	One file with duplicates-free raw SNPs data
STEP 6	Split data based on subjects' sex	PLINK	.bed .bim .fam	.bed .bim .fam	2 files (female and male) with duplicates-free raw SNPs data
STEP 7	Split data into train and test sets	PLINK	.bed .bim .fam list_ID_scotland.txt (containing IDs of subjects part of the test sets)	.bed .bim .fam	4 files (female train and test; and male train and test) with duplicates-free raw SNPs data
STEP 8	Filter subjects according to a .txt file (study dependent) Recode SNP data according to additive model	PLINK	.bed .bim .fam .txt (containing IDs of subjects to include in subsequent study)	.bed .bim .fam .raw	Duplicates-free recoded SNPs data
STEP 9	Convert Plink files into .hdf5 file	Python	.bed .bim .fam	.hdf5	Duplicates-free recoded SNPs data in .hdf5 format