

Data Summary

Jianhui Gao

18/02/2022

UK Biobank Population

The data release contains genotypes of 488,377 UK Biobank participants, and 825,927 markers are genotyped.

Quality control

The following QC steps are conducted

- Exclude individuals with $> 10\%$ missing genotypes.
- Exclude SNPs with $\text{maf} < 0.05$
- Include only SNPs with a 90% genotyping rate
- Exclude markers that failure the Hardy-Weinberg test at 0.001
- Include only White-British population (Code 1001 in Data field 21000, self reported)
- LD pruning. Window size = 50 SNPs, pairwise LD > 0.5 is removed.
- Exclude one member of each pair of samples with kinship coefficient greater than 0.117. (First degree relatives are removed)

After QC steps and removing individuals with missing age, sex, or phenotype height, 408207 samples and 211,587 genetic markers remained.

Basic Association Model

Linear regression, adjusted for age, sex and 10 genetic PC, where the PCA is performed by FastPCA software as per standard.

Height Prediction

In total, 25697 individuals were removed in the association analysis due to being first-degree relatives with other people. After excluding people with missing weight,

Total Number of Individuals: 25594

Total Number of Male: 11481

Total Number of Female: 14113

