This repository contains Shell scripts to run a genome-wide association analysis for type 2 diabetes in UK Biobank using BOLT-LMM and to generate a global polygenic score for type 2 diabetes using the LDpred algorithm. For more information on the implementation of BOLT-LM in the UK Biobank data set and the use of LDpred, please see <https://alkesgroup.broadinstitute.org/BOLT-LMM/BOLT-LMM_manual.html> and <https://github.com/bvilhjal/ldpred>.

Description:

This code has been used to investigate the interplay between genetics and diet quality on the development of type 2 diabetes among 35,759 adults from three US cohorts including data from the Nurses’ Health Study I and II and the Health Professionals Follow up Study.

To generate a global polygenic score for type 2 diabetes, we used external data from a random sample of 391,147 participants within UK Biobank with white genetically confirmed racial/ethnic background.

We conducted a genome-wide association analysis for type 2 diabetes (n= 17,403 cases) using linear mixed models implemented in BOLT-LMM. Type 2 diabetes cases were defined based on self-reported diagnosis and secondary care data according to ICD-10 code (E10, E11, E13, or E14). Models were adjusted for age, age-square, sex, study center, and the first 20 principal components.

We applied the LDPred algorithm using obtained summary statistics. To generate the global polygenic score, we excluded variants with minor allele frequency < 1%. We used a linkage disequilibrium reference panel of 503 European samples from 1000 Genomes phase 3 version 5, ranging causal fractions from 0.001 to 1. In total, ~850,000 independent genetic variants were used to calculate the global polygenic score.

We next tested for the predictive performance of the global polygenic score in an internal validation set within UK Biobank (n= 20,000 participants, 893 type 2 diabetes cases). For everyone in the UK Biobank, the number of associated alleles weighted by the log of the odds ratio was counted and summed across all genetic variants.

Information including the procedures to obtain and access the data and codes used in this study in the Nurses’ Health Study I and II, and the Health Professionals Follow-Up Study is described at http://www.nurseshealthstudy.org/researchers for the Nurses’ Health Study (contact: nhsaccess@channing.harvard.edu) or https://www.hsph.harvard.edu/hpfs/ for the Health Professionals Follow-up Study (contact: [hpfs@hsph.harvard.edu](mailto:hpfs@hsph.harvard.edu)).