# DIAMANTE multi-ancestry GWAS meta-analysis of type 2 diabetes
# 5 April 2022

**Reference.** Mahajan A, et al. (2022). Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat Genet XXX.

We accumulated association summary statistics from 122 GWAS in 180,834 T2D cases and 1,159,055 controls (effective sample size 492,191) across five ancestry groups: European ancestry (51.1% of the total effective sample size); East Asian ancestry (28.4%); South Asian ancestry (8.3%); African ancestry, including recently admixed African American populations (6.6%); and Hispanic individuals with recent admixture of American, African, and European ancestry (5.6%). Each ancestry-specific GWAS was imputed to reference panels from the 1000 Genomes Project, Haplotype Reference Consortium, or population-specific whole-genome sequence data. Subsequent association analyses were adjusted for population structure and relatedness. We considered 19,829,461 bi-allelic autosomal single nucleotide variants (SNVs) that overlapped reference panels with minor allele frequency (MAF) > 0.5% in at least one of the five ancestry groups.

We aggregated association summary statistics across GWAS via: (i) multi-ancestry meta-regression, implemented in MR-MEGA, which models allelic effect heterogeneity correlated with genetic ancestry; and (ii) fixed-effects meta-analysis. We also conducted ancestry-specific fixed-effects meta-analysis across GWAS from each of the five ancestry groups. Association summary statistics were double corrected for genomic control (before and after meta-analysis). We conducted fine-mapping within each locus attaining multi-ancestry genome-wide significance ($p<5x10^{-9}$). We first dissected distinct association signals through approximate conditional analysis. We then constructed 99% credible sets of SNVs for each distinct association signal that incorporated an annotation-informed prior model of causality.

The following datasets are available for download:

1. Multi-ancestry GWAS meta-analysis summary statistics: DIAMANTE-TA.sumstat.txt.gz. A space delimited file with one row per SNV including the following columns: chromosome and position (hg19, build 37); chromosome and position ID; rs ID; effect allele and other allele; MR-MEGA association *p*-value (corrected for genomic control); MR-MEGA ancestry-correlated heterogeneity *p*-value; MR-MEGA residual heterogeneity *p*-value; fixed-effects meta-analysis beta (log-odds ratio); fixed-effects meta-analysis standard error (corrected for genomic control); fixed-effects meta-analysis association *p*-value (corrected for genomic control); and total effective sample size.

2. Ancestry-specific GWAS meta-analysis summary statistics (East Asian, European, and South Asian): DIAMANTE-EAS.sumstat.txt.gz, DIAMANTE-EUR.sumstat.txt.gz, and DIAMANTE-SAS.sumstat.txt.gz. A space delimited file with one row per SNV including the following columns: chromosome and position (hg19, build 37); chromosome and position ID; rs ID; effect allele and other allele; effect allele frequency; fixed-effects

meta-analysis beta (log-odds ratio); fixed-effects meta-analysis standard error (corrected for genomic control); and fixed-effects meta-analysis association *p*-value (corrected for genomic control). **Please note that African and Hispanic ancestry-specific summary statistics will be made available on acceptance of manuscripts describing the results of these analyses**.

3. Annotation-informed fine-mapping for each distinct association signal from approximate conditional analysis: fine_mapping_upload.tar. A tar directory containing one space delimited file per distinct association signal, using the nomenclature finemap.LOCUS.indexSNV.txt. Each file includes one row per SNV in the locus, with the following columns: chromosome and position ID (hg19, build 37); MR-MEGA association *p*-value (corrected for genomic control); MR-MEGA $\log_{10}$ Bayes factor in favour of association; annotation-informed posterior probability of association.

4. Annotation-informed 99% credible sets of SNVs for each distinct association signal: credible_set_upload.tar. A tar directory containing one space delimited file per distinct association signal, using the nomenclature credset99.LOCUS.indexSNV.txt. Each file includes one row per SNV in the credible set, with the following columns: chromosome and position ID (hg19, build 37); MR-MEGA association *p*-value (corrected for genomic control); MR-MEGA $\log_{10}$ Bayes factor in favour of association; annotation-informed posterior probability of association.

The sample size and precision of the statistics presented should preclude de-identification of any individual subject included in the meta-analysis. However, in downloading these data, you undertake: (i) not to attempt to de-identify individual subjects; and (ii) not to repost these data to a third-party website.

Please refer any enquiries to Andrew Morris: andrew.morris-5@manchester.ac.uk.