DIAGRAM, GoT2D, and T2D-GENES joint effort

Fasting glucose and insulin exome-array meta-analysis

30th October 2014

The files contain association summary statistics for the fasting glucose (FG) and insulin (FI) exome-array (single-variant and gene-based) meta-analysis in Mahajan *et al*. (submitted).

The meta-analysis consists of 33,231 (FG) and 30,825 (FI) non-diabetic individuals from 14 studies of European ancestry, all genotyped with the Illumina HumanExome BeadChip. Sample and SNP quality control were undertaken within each study. Within each study, residuals for both traits were calculated after adjustment for BMI and other study-specific covariates. Study-specific inverse-rank normalized residuals were then tested for single-variant association using a linear mixed model. Allelic effect sizes were obtained from analysis using the untransformed residuals. The association summary statistics across studies were then combined using fixed-effect meta-analysis.

Gene-based meta-analysis was performed using the sequence kernel association test (SKAT) and a frequency-weighted burden test applying four alternate variant masks which combine functional annotation and allele frequency thresholds. The four masks are (1) protein-truncating variants only (PTV-only), (2) protein-truncating and missense variants (MAF < 1% in all ancestries) (PTV+missense), (3) protein-truncating and missense variants predicted to be damaging by all five algorithms (PTV+NSstrict), and (4) protein-truncating and missense variants (minor allele frequency (MAF) < 1% in all ancestries) predicted to be damaging by at least one algorithm (PTV+NSbroad). In silico algorithms: Polyphen2-HumDiv, PolyPhen2-HumVar, LRT, MutationTaster, and SIFT.

Single-variant analysis result files

For each SNP, we have provided the following information:

1. Chromosome and postion (build 37, bp)
2. Effect and non-effect allele (aligned to the forward strand)
3. P-value for association
4. Regression coefficient estimates and standard error for effect allele
5. Total sample size

Gene-based analysis result files

For each Gene, we have provided the following information:

1. Gene name
2. Trait
3. Mask
4. Burden test P-value
5. SKAT test P-value
6. Number of variants included
7. Total minor allele count
8. Variants tested in gene (Chromosome:position, build 37)

The sample size and precision of the data presented should preclude de-identification of any individual subject. However, in downloading these data, you undertake not to attempt to de-identify individual subjects.

Reference

Mahajan A. et al. Identification and functional characterization of *G6PC2* coding variants influencing glycemic traits define an effector transcript at the *G6PC2-ABCB11* locus (Submitted)

Please refer any queries to:

Anna Gloyn (anna.gloyn@drl.ox.ac.uk)
Anubha Mahajan (anubha@well.ox.ac.uk)