DIAGRAM Metabochip meta-analysis of type 2 diabetes (T2D) susceptibility in fine-mapping regions: 99% credible sets of variants

Prepared by Andrew Morris: 30 October 2015

This file contains association summary statistics for variants contained in 99% credible sets for each distinct association signal from the DIAGRAM Metabochip meta-analysis of T2D susceptibility across 39 fine-mapping regions in established loci, published in Gaulton *et al*. (2015).

We considered a total of 27,206 T2D cases and 57,574 controls from 23 studies from populations of European ancestry, all genotyped with the Metabochip. Sample and variant quality control was performed within each study. To improve the quality of the genotype scaffold in each study, variants were subsequently removed if: (i) allele frequencies differed from those for European ancestry haplotypes from the 1000 Genomes Project Consortium phase 1 integrated reference panel (March 2012 release) by more than 20%; AT/GC variants had MAF>40% because of potential undetected errors in strand alignment; or (iii) MAF<1% because of difficulties in calling rare variants. Each scaffold was then imputed up to up to the phase 1 integrated reference panel (all ancestries, March 2012 release) from the 1000 Genomes Project Consortium. Within each study, well-imputed variants were tested for T2D association under an additive model after adjustment for study-specific covariates, including principal components to adjust for population structure. Association summary statistics for each variant for each study were corrected for residual population structure using the genomic control inflation factor obtained from 3,598 independent (r^2<0.05) QT-interval variants, which were not expected to be associated with T2D.

We then combined association summary statistics for each variant across studies via fixed-effects inverse-variance weighted meta-analysis. The results of the meta-analysis were subsequently corrected by a second round of QT-interval genomic control (λ_{QT}=1.18) to account for structure between studies. Variants were excluded from downstream analyses if they were reported in less than 80% of the total effective sample size, defined as $N_{eff} = 4xN_{cases}xN_{controls}/(N_{cases}+N_{controls})$, thus removing those that were not well imputed in the majority of studies.

For each distinct signal, we calculated the posterior probability, π_{Cj}, that the *j*th variant is driving the association, given by

$$\pi_{Cj} = \frac{\Lambda_j}{\sum_k \Lambda_k},$$

where the summation is over all retained variants in the fine-mapping region. In this expression, Λ_j is the approximate Bayes' factor for the *j*th variant, given by

$$\Lambda_j = \sqrt{\frac{V_j}{V_j+\omega}} \exp\left[\frac{\omega\beta_j^2}{2V_j(V_j+\omega)}\right],$$

where β_j and V_j denote the estimated allelic effect (log-OR) and corresponding variance from the meta-analysis across Metabochip studies. In loci with multiple distinct signals of association, results are presented from exact conditional meta-analysis after adjusting for all other index variants in the fine-mapping region. In loci with a single association signal, results are presented from unconditional meta-analysis. The parameter ω denotes the prior variance in allelic effects, taken here to be 0.04. The 99% credible set for each signal was then constructed by: (i) ranking all variants according to their Bayes' factor, Λ_j; and (ii) including ranked variants until their cumulative posterior probability of driving the association attained or exceeded 0.99.

For each distinct association signal, we provide a file containing variants contained in the 99% credible set. For each variant, we have provided the following columns of information:

1. 1000 Genomes identifier (phase 1 integrated, March 2012 release).
2. Chromosome and position (build 37, base-pairs).
3. Approximate Bayes' factor.
4. Posterior probability of driving the association signal.
5. P-value for association.

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

Reference: Gaulton KJ, *et al*. (2015). Genetic fine-mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet* (in press).

Please refer any queries to: Andrew Morris (a.p.morris@liverpool.ac.uk).