Readme - Fine-mapping results of glycaemic trait loci with different fine-mapping approaches

Results are from:

Combining functional annotation and multi-trait fine-mapping methods improves fine-mapping resolution at glycaemic trait loci.

Jana Soenksen, Ji Chen, Arushi Varshney, Susan Martin, Stephen C J Parker, Andrew P Morris, Jennifer L Asimit, Inês Barroso, *Human Molecular Genetics*, Volume 35, Issue 2, 2026, ddaf164, https://doi.org/10.1093/hmg/ddaf164 PMID: 41251494.

https://academic.oup.com/hmg/advance-article/doi/10.1093/hmg/ddaf164/8326576?login=true#540440970

See the publication for detailed description of methods. Summary stats and some of the multi-ancestry fine-mapping results were taken from Chen J, Spracklen CN, Marenne G, et al. The trans-ancestral genomic architecture of glycemic traits. *Nat Genet*. 2021;53(6):840-860. doi:10.1038/s41588-021-00852-9.

See Zenodo (10.5281/zenodo.17371040) for code used to generate and analyse results.

There is a tsv file for each trait and fine-mapping method. Each file contains the 99% credible sets of locus-trait combinations analysed with the respective method. The naming structure is as follows: credible_sets_*fine-mapping-method*_*trait*.tsv

Fine-mapping approaches

Single trait agnostic fine-mapping (ST agno) was carried out with FINEMAP v1.4.

Single trait agnostic fine-mapping with FINEMAP v1.1 (ST v1 agno) correspond to the EUR-like results from Chen et al 2021. Some loci were rerun after re-inserting lead variants with FINEMAP v1.1 as for the original run.

Single trait annotation informed fine-mapping (ST anno) used fGWAS to calculate priors based on annotations and then used FINEMAP v1.4 to carry out fine-mapping with these priors.

Multi-trait agnostic fine-mapping (MT agno) used single-trait fine-mapping results from individual traits generated with FINEMAP v1.4 (ST agno) as input for multi-trait fine-mapping with Flashfm.

Multi-trait annotation informed fine-mapping (MT anno) used the single-trait annotation informed fine-mapping results from fGWAS + FINEMAP v1.4 (ST anno) as input for Flashfm.

Multi-ancestry agnostic fine-mapping (MA agno) correspond to the trans-ancestry results from Chen et al 2021. Some loci were rerun after re-inserting lead variants with FINEMAP v1.1 as for the original run.

Traits and sample sizes

For detailed explanation of phenotype definitions see Chen et al 2021.

Abbreviation	Description	Adjusted for BMI	EUR-like sample size	MA sample size
FG	Fasting glucose	Yes	165,515	242,353
FI	Fasting insulin	Yes	124,123	183,654
2hGlu	2hr Glucose post glucose challenge	Yes	44,208	66,726
HbA1c	Glycated haemoglobin	No	127,524	194,648

File columns cred_set files

chr – chromosome

pos – position of variant (hg19)

A1 – Allele 1

A2 – Allele 2

MPP – marginal posterior probability of variant, calculated by summing up the posterior probability of all models containing this variant

Credset_size – number of variants in the 99% credible set – determined by summing up all models from the largest to the smallest until they reached 99% posterior probability of containing or tagging the causal variants and counting the unique variants in it.

Credset_id – unique identifier for this method-locus-trait combination. Format: *method*_*trait*_*chromosome*_*start_position_locus*_*end_position_locus*. The

position refers to the coordinates of the whole locus, not just the coordinates of the credible set.

max_causal_variants – this corresponds to the maximum number of causal variants at this locus according to the fine-mapping method, i.e. the maximum number of variants in the models/configurations. If it is >1, the MPP of variants in the 99% credible set might add up to more than 1.

Partial summary stats files

The files called *trait*_sumstat.tsv are partial summary stats files generated with METAL that contain only loci that were significant $5x10^{-8}$ with at least one of the traits and 10^{-6} with the trait of interest in EUR-like ancestry. These differ from the full summary stats available for Chen et al 2021 because they do not include metabochip data and only contain significant loci (see above). For some loci we also reinserted some key variants that had been QC'd out, see methods of paper. These partial summary stats files are useful for fine-mapping of significant loci.

File columns summary stats files

Locus – locus identifier in format: Chr_startpos_endpos

Variant – variant identifier in format: Chr_Pos_Allele1_Allele2

Chr - chromosome

Pos - position of variant (hg19)

Allele1 - Effect allele

Allele2 – non-effect allele

Freq1 – frequency of allele 1

Effect – Effect size (of allele 1)

StdErr – Standard error

P.value – p value

z – Z score

TotalSampleSize – total sample size for this variant

Acknowledgements

When using the downloadable data please acknowledge the source of the data as follows: "Data on glycaemic traits have been contributed by MAGIC investigators and have been downloaded from www.magicinvestigators.org" and cite the paper.