GWAS summary statistics for 25 hydroxyvitamin D (25OHD)

3 March 2020

Below is a description of the summary statistics for the 25 hydroxyvitamin D (25OHD) genome-wide association study published in Revez et al. (2020) Nature Communications.

The summary statistics were generated for individuals of European ancestry from the UK Biobank (UKB) sample.

Revezetal2020_25OHD.gz: Summary statistics generated with rank-based inverse normal transformed (RINT) 25OHD levels from 417,580 individuals of European ancestry. Covariates were accounted for in the linear mixed model implemented in fastGWA[1]. Covariates were: age at assessment, sex, assessment month, assessment centre, first 40 principal components (PCs), genotyping batch, supplement intake (variable with four levels, namely: "no information", "never taken", "other supplements", "25OHD supplements").
Columns are:

- CHR: chromosome (chromosome 23 and 25 are chromosome X and chromosome X pseudo-autosomal region, respectively)
- SNP: SNP rs ID
- POS: base pair position (hg19)
- A1: effect allele
- A2: other allele
- N: sample size
- AF1: frequency of A1
- BETA: effect size of A1
- SE: standard error BETA
- P: *P*-value

Revezetal2020_25OHD_BMIcov.gz: Summary statistics generated as "Revezetal2020_25OHD.gz", but with BMI included as a covariate as well. Columns are as described for Revezetal2020_25OHD.gz.

Revezetal2020_25OHD_BMIcond.gz: Summary statistics generated with mtCOJO[2], a summary-data-based conditional-analysis approach that was shown in simulations to be robust to collider bias when conditioning on a correlated trait[3]. Specifically, 25OHD SNP effects were conditioned on BMI using the 25OHD SNP association results (Revezetal2020_25OHD.gz) and BMI summary statistics generated with the UKB sample [3]. Columns are as described for Revezetal2020_25OHD.gz.

Revezetal2020_25OHD_log: Summary statistics generated for meta-analysis with those made available from the SUNLIGHT consortium[4]. The phenotype was processed as Revezetal2020_25OHD.gz but 25OHD levels were natural-log transformed (instead of RINT), and BMI, but not supplement intake, was included as a covariate. These summary statistics were generated based on 416,247 individuals of European ancestry. Columns are as described for Revezetal2020_25OHD.gz.

Revezetal2020_25OHD_SUNLIGHTmeta: Summary statistics obtained from a sample size-based meta-analysis of the 25OHD SNP association results (Revezetal2020_25OHD_log) and SUNLIGHT summary statistics[4] imputed to 1000 Genome Project (1KGP). Of 8,546,067 SNPs available in the UKB GWAS (Revezetal2020_25OHD_log), 6,912,458 overlapped with the imputed summary statistics from the SUNLIGHT cohort. After meta-analysis and QC, 6,912,294 SNPs remained and are reported.

Columns are SNP, SNP rs ID; A1, effect allele; A2, other allele; freq, frequency of A1; b, se and p, effect size, standard error and *P*-value of A1; N, sample size.

References

1 Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of large-scale data. *bioRxiv*, 598110 (2019).

2 Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369-375, S361-363 (2012).

3 Xue, A. *et al.* Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun* **9**, 2941 (2018).

4 Jiang, X. *et al.* Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. *Nat Commun* **9**, 260 (2018).