FinnGen core analysis document- Kanta lab v2

February 2024

All unique lab measurements (OMOPID) with at least 1,000 individuals with at least one measurement were analyzed (n = 383 lab measurements) + 5 derived labs

QC and data preparation

(see Kanta_labs_over1000_individuals_summary_table.txt for further details)

- Clear outliers were removed, if any (see Kanta_labs_GWAS_results_v2_summary.txt for detailed outlier filters). For derived labs, outlier filters were inherited if any in the labs used to derive them.

- 347 of the labs were analyzed as quantitative. All of them were normalized (inverse-rank) and
- 36 of the labs were analyzed as binary (ever having measurement above a certain threshold or test outcome as 'abnormal' or 'high' vs never having measurement above a certain threshold or test outcome as 'abnormal' or 'high')

- Sample-wise median was calculated for all lab measurements that were analyzed as quantitative

- Inverse-rank normalisation (IRN) was done for all labs measurements that were analyzed as quantitative

GWAS

- GWAS was performed for all 388 lab measurements using REGENIE v3.3
- Same covariates (age at end of follow-up or death, sex, genotyping batch and PC1-10) were used as in core endpoint GWASes
- Linear mixed model was used for lab measurements that were analyzed as quantitative, and logistic mixed model was used for binary phenotypes

- Summary statistics were filtered for variants with MAC >30 (for quantitative only) and imputation info score > 0.6

FINEMAP

- All 388 summary statistics were finemapped using FINEMAP and SUSIE, using pipeline described here: https://github.com/FINNGEN/finemapping-pipeline
- GWS ($p < 5E-8$) regions were finemapped
- MHC region was excluded

Autoreporting

- For finemapped (GWS regions, MHC region excluded) regions, autoreporting was performed using pipeline described here: https://github.com/FINNGEN/autoreporting

Colocalization

- For all 388 lab measurements, all gws signals were colocalized with all R12 endpoints (2,469) using pipeline described here: https://github.com/FINNGEN/coloc.susie.direct

LDSC (heritability estimates and genetic correlations)

- For all 388 lab measurements
    1. Heritability estimates and
    2. Genetic correlations with all R12 endpoints (2,469) were calculated using LDSC software and pipeline described here: https://github.com/FINNGEN/LDSC