SCALLOP consortium analysis plan for INF panel proteins

Adapted from SCALLOP/CVD1 analysis plan, last updated 9/11/2018

Timeline for completing cohort-specific analyses and uploading the results for this project:

---

1. Overview

The SCAndinavian coLLaboration for Olink plasma Protein genetics (SCALLOP) consortium, https://www.olink.com/scallop/, is a collaborative framework for discovery and follow-up of genetic associations with proteins on the Olink Proteomics platform. A meta-analysis has been conducted on Olink CVD1 panel data from participating cohorts; consequent requests were sent and contributions made on the Olink INF panel. This document follows closely the SCALLOP/CVD1 analysis plan for the analysis, and in particular highlights relevant information required to facilitate the meta-analysis.

As with the CVD1 meta-analysis, the tasks will involve

- Identification of pQTLs in SCALLOP discovery cohorts
- Study of pQTLs in replication cohorts
- Investigation of the mechanistic basis of identified cis- and trans-pQTL by functional annotation
- Examination of pQTL pleiotropic effects
- Evaluation over the causal role of INF proteins disease outcomes such as CHD and stroke
- Other downstream analysis

2. Data and analysis

Proteins

The Olink INFlammation panel of 92 proteins, e.g, https://github.com/jinghuazhao/INF/blob/master/doc/olink.inf.panel.annot.tsv.

SNPs

- 1000 genomes imputation, build 37 (hg19) positions.
- SNPs filtering on imputation quality at time of meta-analysis.
- Quality control on aspects such as SNP/sample call rates, gender mismatch, abnormal inbreeding coefficient, failed cryptic relatedness test, ancestry outlier, heterozygosity and Hardy-Weinberg equilibrium test.

Association analysis

- Rank-based inverse normal transformation on the raw measurement of proteins including those below lower limit of detection, e.g., via `invnormal` function,

```
invnormal <- function(x)
  qnorm((rank(x,na.last="keep")-0.5)/sum(!is.na(x)))
```

- Multiple linear regression for all samples including sex, age, principal components and other cohort specific covariates.
- Additive genetic model
- For case-control data, cases and controls are analysed separately – results will be merged at meta-analysis stage

Software

It is preferable to use software which account for genotype uncertainty, such as SNPTEST, QUICKTEST, and BOLT-LMM.

3. Descriptive statistics

Please fill out the spreadsheet as with SCALLOP/CVD1 with naming convention:

- STUDY.descriptives.DATE.xls
- Where STUDY is a short (14 characters or less) identifier for the population studied, which is the same for all files provided by your study.
- DATE is the date on which the file was prepared, in the format "DDMMYYYY".

4. File formats for GWAS results

SNP table for GWAS results

Please include the following columns. Missing values are coded as "NA".

No	Variable name	Description of variable
1	SNPID	CHR:POS_A1_A2 (such that A1-10. The results will be replicated in independent cohorts.

6. Uploading of results

See the CVD1 analysis plan.

7. Contact information

For general questions about SCALLOP, please contact Anders Malarstig (anders.malarstig@ki.se). For technical issues about TRYGGVE, please contact Lasse Folkersen (lasfol@cbs.dtu.dk).

For questions regarding SCALLOP/INF, please contact Jing Hua Zhao (jhz22@medschl.cam.ac.uk) and James Peters (jp549@medschl.cam.ac.uk).